

Article

Oil Extraction and Agricultural Storage Co-Location: A GIS Spatial Analysis in North Dakota

Edmond Loni M. Lisinge and Raj Bridgelall * 

Department of Supply Chain and Transportation, College of Business, North Dakota State University, Fargo, ND 58108-6050, USA; edmond.lisinge@ndsu.edu

* Correspondence: raj@bridgelall.com

Abstract

Oil extraction and agricultural production are central to North Dakota's economy, yet their spatial interactions remain poorly understood. This study conducts a statewide geospatial analysis integrating OpenStreetMap data, GIS processing, DBSCAN clustering, and spatial statistics to examine the colocation of 7102 oil well and 4277 grain silo sites. Hotspot and spatial heterogeneity tests using the Getis–Ord G_i^* statistic and local Moran's I reveal a pronounced spatial divide: oil activity is tightly clustered in the western Bakken region, whereas grain storage facilities concentrate across central and eastern counties. The limited geographic overlap suggests minimal systemic land-use conflict, though localized high-intensity interactions emerge in McKenzie, Dunn, and Mountrail counties. These patterns provide stakeholders with insight into potential shared logistics pressures and localized land-use tensions. More broadly, the study demonstrates the value of spatial data mining techniques applied to free, publicly available data for identifying intersectoral industrial patterns that inform policy and infrastructure planning across North Dakota.

Keywords: GIS; DBSCAN; cluster analysis; oil wells; grain silos; spatial interaction; sustainable industries

1. Introduction

Sustainable land-use planning and freight system resilience depend on understanding how multiple infrastructures interact within shared geographic space. In resource-intensive regions, overlap between industrial and agricultural activities generates localized pressure on transportation networks, environmental systems, and land allocation. North Dakota offers a clear case: rapid oil development associated with the Bakken formation has expanded alongside an established grain production system, creating counties where oil well sites and grain storage facilities are situated in close proximity. These co-located activities have the potential to increase competition for freight capacity, influence safety and environmental outcomes, and affect regional resilience during demand shocks.

During the 2013–2014 crude-by-rail surge, oil shipments displaced grain movements across the Upper Midwest, causing railcar shortages, wider basis levels, and higher storage pressures—clear indicators of constrained logistics capacity and shifting traffic [1,2]. These disruptions reduced producer revenues, raised operational risks, and complicated regional infrastructure coordination.

Despite the visible proximity of oil pads and grain silos on conventional maps, little is known about whether these sectors interact at spatial scales that meaningfully affect logistics performance, land-use dynamics, or safety outcomes. Though there is existing



Academic Editors: Felicia Di Liddo, Pierluigi Morano, Francesco Tajani and Paola Amoruso

Received: 12 June 2026

Revised: 9 July 2026

Accepted: 17 July 2026

Published: 19 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

research linking increased crude-by-rail volumes to agricultural bottlenecks, particularly for [3], many studies rely on aggregated regional data or sector-specific analyses that overlook fine-scale cross-sector interactions. Consequently, planners and policymakers lack the localized, empirical evidence needed to anticipate congestion risks, land-use conflicts, and infrastructure strain in regions experiencing concurrent industrial expansion.

The study aimed to identify statistically significant clusters where oil extraction and agricultural storage activities co-locate within spatial scales relevant to operational interactions in North Dakota. To achieve this goal, the study developed a geospatial workflow integrating free, publicly available data from OpenStreetMap—a workflow potentially transferable to regions with comparable spatial data availability. The workflow unified operational sites by using density-based clustering and applied spatial association techniques to detect localized co-location patterns relevant to agricultural and industrial operations.

The contributions of this work are threefold. First, the study provides a framework for converting heterogeneous spatial features into operational units to strengthen spatial analysis and decision-making, with potential transferability to comparable data environments. Second, the analysis identifies statistically significant oil–agriculture co-location hotspots that may warrant further investigation for infrastructure and risk-management planning. Third, the study presents a spatial cluster detection methodology adaptable to other regions and phenomena, with broader applications in land-use and logistics research. Collectively, these contributions advance the understanding of sectoral interactions in a state’s evolving landscape, supporting a data-driven approach to managing complex spatial relationships between energy and agriculture, with implications for safety, efficiency, and economic sustainability. The specific marginal contribution relative to prior North Dakota energy–agriculture research is not the individual analytical techniques, each of which (DB-SCAN, Getis–Ord G_i^* , local and bivariate Moran’s I) is established in the spatial statistics literature, but their integration into a single open-data, site-level pipeline that (i) converts heterogeneous point, line, and polygon OSM geometries for two structurally different infrastructure types into a common operational-site unit, and (ii) applies FDR-corrected, multi-scale bivariate cluster detection to that unit. Prior North Dakota studies examined energy–agriculture interactions at the county or regional scale (Section 2); to the authors’ knowledge, this is the first application of site-level, statistically validated co-location detection in the state using exclusively free, publicly available data, which is what makes the workflow directly reusable by state or county planning offices without proprietary GIS licenses or paid data subscriptions. A further contribution, developed in Section 5 to interpret why HH nodes form where they do, is a county-level statistical test of joint-sector concentration (an oil $\times$ silo interaction term) alongside an exploratory Agriculture–Energy Competition Index (AECI); both are offered as descriptive, hypothesis-generating tools for identifying candidate locations for closer study, rather than as validated measures of infrastructure strain.

The organization of the rest of this paper is as follows: Section 2 reviews the literature on energy–agriculture spatial interactions in North Dakota; Section 3 describes the geospatial workflow; Section 4 presents and interprets the results; Section 5 discusses implications and limitations; Section 6 concludes the study.

2. Literature Review

Since the early 2000s, rapid expansion of unconventional oil and gas development in North Dakota’s Bakken Shale has fundamentally transformed the state’s rural landscape. This expansion generated unprecedented conflicts between energy infrastructure and agricultural systems [4]. This energy boom, fueled by horizontal drilling and hydraulic fracturing technology, has industrialized traditional rural, agricultural counties in western

North Dakota, resulting in complicated interactions between competing land-uses [5]. The scale and intensity of this development make it a unique case study to understand how energy infrastructure interacts with established agricultural systems in rural America.

North Dakota's shift from a predominantly agricultural state to a major oil producer reflects broader conflicts between traditional rural economies and extractive industries [6]. North Dakota now accounts for roughly 11% of American oil production, with rural regions such as Williams County facing fast in-migration of oilfield workers and acute infrastructure challenges [7]. This section examines how energy development affects agricultural infrastructure and related rural systems, drawing on empirical research that documents land-use changes, transportation disputes, resource competition, and socioeconomic impacts in North Dakota's rural communities.

2.1. Direct Land-Use Conflicts and Agricultural Displacement

Fitzgerald et al. (2020) found that drilling reduced crop cover and increased fallow land, with effects that were spatially and temporally heterogeneous across cropland classes and frequently temporary [4]. This suggests that previous estimates have overvalued the net impact of drilling by as much as twice. Preston and Kim (2016) further quantified 12,990 hectares of land converted to well pads, primarily from agriculture (49.5%) and prairie (47.4%) [8]. Projected development is expected to affect 2.3% of the Bakken drilling landscape and adjacent grasslands. In addition, Alfred et al. (2015) referred to the infrastructure requirements of horizontal drilling and hydraulic fracturing as reducing large areas of the Great Plains into industrialized land [6]. This pattern reflects the broader transformation of agricultural landscapes driven by energy development.

Rakitan (2018) examined the economic repercussions of land displacement and its influence, particularly on surface rent in North Dakota [9]. Rakitan revealed that surface tenants' willingness to pay for land parcels near oil facilities had not changed significantly. This conclusion draws on approximately 17,000 land rental transactions obtained from State Trust Lands data. Although many scholars argue that energy development can reduce agricultural profitability, recent evidence shows that proximity to oil and gas activity is linked to measurable declines in agricultural output. Spatial analyses of unconventional oil and gas expansion consistently find reductions in crop acreage and increases in fallow land as drilling intensifies [10]. Similarly, studies in shale regions also reported farm-level adjustments including consolidation, altered production decisions, and reduced crop cover associated with drilling footprints [11]. Smith et al. (2019) further confirmed these patterns, highlighting recurring proximity effects such as land-use displacement and operational disruptions [12].

Taken together, these studies show that drilling-related land displacement is real but heterogeneous, varying by cropland class, time since drilling, and parcel-level lease terms. However, estimates of its magnitude differ substantially, ranging from Fitzgerald et al.'s [4] conclusion that net impacts may be overstated by as much as half to Preston and Kim's [8] direct measurement of thousands of hectares converted. Despite these advances, no previous study has statistically evaluated whether oil infrastructure and agricultural land co-locate at spatial scales relevant to shared operations. Addressing this gap is the focus of the present study.

2.2. Transportation Infrastructure Competition and Damage

Energy expansion has resulted in fierce rivalry for transportation infrastructure, particularly in the transfer of agricultural commodities. Villegas (2016) examined how growing rail demand from the energy sector has impacted wheat transportation in the Upper Midwest [13]. In this region, rail service is the most cost-effective shipping option for

agricultural commodities. As oil has taken up freight capacity on trains, farmers in regions like North Dakota have had to pay more to reach grain markets, resulting in financial losses. Similarly, Ortiz (2016) found that railroad congestion from energy development impaired wheat markets, creating price differentials across Midwest producers relative to Gulf Coast markets [13]. Historically, oil price declines exacerbated farmers' shipping challenges as energy firms used railcars to store excess crude oil inventory. This further impeded agricultural access to transportation infrastructure.

Heavy truck traffic from energy development caused substantial damage to rural roads and bridges. A study by McCarthy et al. (2015) highlighted the economic effects on roads and bridges in multiple states, especially in North Dakota [14]. This finding indicates a need for infrastructure upgrades and new bridge construction to address increased traffic volumes. Furthermore, Dharmadhikari et al. (2016) reported on rural bridge construction needs in North Dakota [15]. They revealed that oil exploration escalates vehicle traffic and heightens maintenance demands on infrastructure.

Rahm et al. (2015) used comparative data from Texas to reveal alarming safety trends [16]. These included a 26% increase in crash rates and a 49% rise in fatalities and severe injuries, alongside deteriorating road conditions and escalating maintenance costs. While the Texas study emphasized a regional context, the pattern of infrastructure strain it documented—increased crash rates and maintenance costs—has been similarly reported in North Dakota counties experiencing high drilling activity [17].

Across these transportation studies, a consistent mechanism emerges regardless of mode: energy development competes with agriculture for fixed-capacity infrastructure—rail cars in Villegas [13] and Ortiz [13], road and bridge capacity in McCarthy et al. [14] and Dharmadhikari et al. [15]—and the severity of impact scales with drilling intensity, as the North Dakota–Texas comparison in Rahm et al. [16] illustrates. What this literature does not establish is where, spatially, oil and grain-handling infrastructure are close enough to plausibly compete for the same local road segment or rail siding, which motivates the site-level co-location analysis developed in this study.

2.3. Environmental Impacts: Water Resource and Ecological Interference

Water resource competition represents a critical but often overlooked form of infrastructure interference. Horner et al. (2016) provided a comprehensive analysis of water use in the Bakken Shale and documented that water use for hydraulic fracturing grew five-fold, from 770 million gallons in 2008 to 4.3 billion gallons in 2012. [18]. The study found that existing groundwater resources are inadequate to meet hydraulic fracturing demand. This creates competition with agricultural water needs.

A study by Hearne & Fernando investigated water management strategies in North Dakota's oil-producing region. They observed that while the Missouri River and Lake Sakakawea supply abundant water, infrastructure expansion during oil boom periods proved challenging [19]. Sales from drilling activities have helped support the building of regional water systems. This resulted in a complicated interdependence between agricultural and energy water infrastructure. Furthermore, Fernando & Cooley (2016) discovered that transportable on-site treatment facilities could be cost-effective while reducing competition for freshwater resources [20].

Smith and Haggerty (2020) studied how public infrastructure investments in water systems contain “exploitable ambiguities” that various stakeholders use to pursue divergent agendas [21]. Their research of a large regional water supply project demonstrated how infrastructure decisions in energy-dependent societies have confusing implications for agricultural customers.

Beyond water resource competition, energy development generates broader ecological and environmental interference with agricultural landscapes, extending from soil and air quality to wildlife habitat.

Environmental interference between energy and agricultural systems poses significant challenges for rural communities. A study by Spiess et al. (2020) explored the impact of fracking traffic on bird and invertebrate communities [22]. The study revealed that dust emissions led to substantial deposition 180 m into adjacent crop fields. While wildlife in agricultural landscapes displayed some resilience to these impacts, the findings highlighted the environmental pathways through which energy development can affect agricultural areas.

Boslett et al. (2021) documented increased light pollution associated with shale oil and gas development in rural regions [23]. Their analysis of nationwide data from 2000 to 2012 indicated strong correlations between increased drilling activities, insufficient sleep, and deteriorating health outcomes. These findings highlighted a quality-of-life interference that adversely affects agricultural communities. Additionally, Van der Burg et al. (2023) examined the responses of grassland bird populations to the expansion of energy portfolios [24]. Their findings indicated that birds reacted more negatively to biofuel feedstocks, particularly corn and soybeans, compared to oil and gas development. This research suggested that renewable energy policies encouraging biofuel production may inadvertently create broader ecological disruptions within grassland systems than fossil fuel extraction. This is because both forms of energy development contribute to shifts in wildlife population distributions.

2.4. Economic and Social Dimensions of Interference

Socioeconomic impacts of energy–agriculture interactions introduce complex challenges beyond land-use conflict alone. A framework by Fernando & Cooley (2016) identified five critical issues arising from oil booms: the need for affordable housing, improvements in community infrastructure, enhancement of public services, attraction of new businesses, and strategies for community integration [20]. These elements significantly influence the social dynamics and economic sustainability of agricultural communities.

Addey (2019) applied autoregressive distributed lag models to investigate the effects of the oil boom on North Dakota’s agricultural sector, providing evidence that the oil industry adversely affected agriculture, but did not significantly affect manufacturing [25]. The study recorded an 80% adjustment speed coefficient, indicating that imbalances may rectify themselves within approximately three months. Additionally, their study advocated for structural spending policies aimed at mitigating the negative influences on sectors competing for labor.

Recent studies indicate that the social effects of energy development are complex. They extend well beyond the physical structures associated with energy development, in a manner not captured by measures of distance alone. For instance, McGranahan et al. (2017) discovered that the observed effects of energy development are correlated to the affected area in a weak fashion [5]. The study revealed the importance of various social factors. These include views on development, compensation levels, and levels of disruption, as important considerations in community responses to energy development. A study of Williams County, North Dakota, found that the influx of energy development contributed to “rural gentrification” [7]. In this phenomenon, oil field employees, who are generally wealthier, displace the existing agricultural populations in the region. The study revealed instability of the social structures of the region, influenced by the associated markets of production.

2.5. Policy Responses and Mitigation Strategies

Many studies have examined governance responses and mitigation measures addressing energy–agriculture conflicts. For instance, Smith et al. (2019) examined the community reactions towards the adaptation of shared services for dealing with service crises during the energy boom [12]. They found that although the innovations have demonstrated resourcefulness for the communities, they have triggered unrealistic boom periods and contributed towards the development of potentially unsustainably operating projects.

Fernando and Goreham (2018) examined the community capital perspectives in two different rural cities [26]. They discussed the optimization of social capital investment, identification of cultural capital alterations, and the utilization of political capital for effectively dealing with boom effects. In related work using predictive modeling tools to aid transportation infrastructure planning in areas experiencing rapid energy expansion, Lee et al. (2021) demonstrated that by adapting the susceptible–infected–recovered (SIR) epidemic framework, drilling locations with significant traffic implications can be forecasted ahead of time [27]. Forecasting allows transportation agencies to better anticipate and mitigate emerging infrastructure conflicts.

2.6. Spatial Analytical Methods for Infrastructure Co-Location Detection

The literature reviewed above establishes energy–agriculture interaction in North Dakota as a substantive concern but also reveals a near-complete absence of studies applying formal spatial statistical clustering methods to detect where these interactions concentrate. The relevant spatial statistics methods were developed and validated primarily in other application domains. Getis and Ord [28] introduced the G_i^* statistic for local spatial cluster detection, a foundational hotspot analysis tool since applied broadly across public health, criminology, and economic geography, but rarely to energy–agriculture infrastructure specifically. Anselin [29] extended this logic to local indicators of spatial association (LISA), enabling classification of high–high, low–low, high–low, and low–high spatial regimes. This study adopts the LISA method for bivariate co-location detection. Density-based clustering algorithms such as DBSCAN, used in the present study to convert heterogeneous point, line, and polygon OpenStreetMap features into discrete operational sites, are similarly well established in the broader spatial data mining literature but have seen limited application to infrastructure co-location problems of this kind.

Within the North Dakota energy–agriculture context specifically, prior quantitative work has favored predictive and simulation-based approaches over descriptive spatial-cluster detection. Lee et al. [27] adapted a susceptible–infected–recovered (SIR) epidemic framework to forecast the spread of drilling-related traffic impacts, demonstrating that quantitative spatial predictive methods can usefully inform transportation infrastructure planning in this domain, though at a network forecasting scale rather than the site-level statistical detection pursued in the present study. More broadly, the reliability of any site-level spatial analysis in this setting depends on the completeness of the underlying geospatial data; a recent national assessment of OpenStreetMap data quality [30] found that completeness and positional accuracy are systematically lower in the nonmetropolitan counties that characterize most of North Dakota’s oil- and agriculture-producing regions, a consideration that directly informs the validation approach adopted later in Section 4.1.

Taken together, this body of work establishes the individual analytical building blocks used in this study, but none of it has previously been integrated into a reproducible, site-level, statistically validated co-location-detection workflow for energy–agriculture infrastructure interaction of the kind developed here. Overall, the literature on energy development and agricultural infrastructure in North Dakota reveals pressures that extend beyond simple land displacement impacts. Impacts vary in time and space, with some

disruptions being short and others lasting. Energy-related traffic impedes agricultural product movement, posing logistical issues. As the competition for resources like water grows, agricultural operations become strained. Economic research yields conflicting results on land rental values and productivity losses. However, the social consequences, such as housing relocation and demographic shifts, are more visible. These changes suggest that energy growth fundamentally alters rural systems through interwoven infrastructural, economic, and social dynamics.

Critically, no prior study has examined oil–agriculture spatial interactions at the site level using a transferrable, open-data geospatial workflow that enables statistical detection of fine-scale co-location patterns across North Dakota. This gap motivates the present analysis.

3. Materials and Methods

The study conducted all spatial data processing and analysis in Python (version 3.12) within a transferrable computational workflow. The study obtained and preprocessed OpenStreetMap (OSM) data externally as GeoPackage (GPKG) files and imported them into the Python environment using GeoPandas (version 1.1.4) to perform spatial data manipulation, including coordinate transformation and geometric operations. The workflow clustered oil well and silo features using the DBSCAN algorithm from the scikit-learn (version 1.9.0) library. The workflow projected all spatial data to NAD83/UTM Zone 14N (EPSG:26914) prior to analysis to ensure accurate distance-based computations.

Spatial weights for local and global statistics were constructed using the DistanceBand class from the PySAL (libpysal version 2.6.0) library, with row-standardization applied. Bivariate spatial autocorrelation was computed using PySAL's esda (version 2.10.0) package, specifically the Moran_BV (global) and Moran_Local_BV (local) implementations. Statistical significance was assessed using 999 Monte Carlo permutations for both global and local Moran's I statistics. This permutation count follows conventional practice in conditional randomization inference for local spatial statistics: the smallest achievable pseudo p -value from n conditional permutations is $1/(n + 1)$, so 999 permutations is the minimum count that allows pseudo p -values to resolve down to the 0.001 significance threshold used downstream in the FDR-corrected analysis (Section 4.2), while remaining computationally tractable across the full bandwidth $\times$ DBSCAN-parameter sensitivity grid reported. All analyses were conducted in a Python environment using standard scientific computing libraries, including NumPy (version 2.5.1) and Pandas (version 3.0.4), with visualization and mapping performed using Matplotlib (version 3.11.0) and GeoPandas (version 1.1.4). Figure 1 illustrates this study's workflow with the procedures, as coded in software.

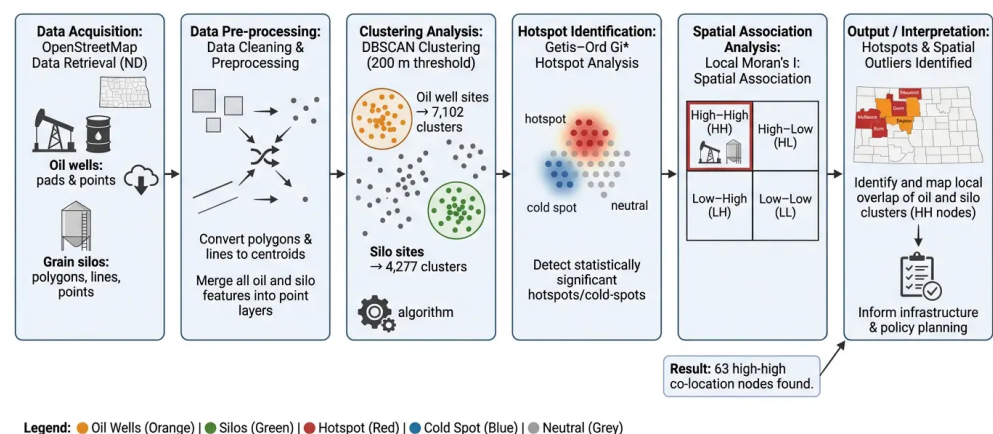


Figure 1. Methodological workflow of this study.

3.1. Data Source

All spatial operations were conducted using a projected coordinate reference system (CRS), specifically NAD83/UTM Zone 14N (EPSG:26914). This projection preserves spatial relationships in metric units and is appropriate for distance-based analyses such as DBSCAN clustering and neighborhood definition. All distance thresholds were computed directly in meters within this projected coordinate system. Degree-based equivalents are reported only for reference. “Well sites” and “silo sites” refer to DBSCAN-derived site centroids representing aggregated groups of spatially proximate features. Individual “wells” and “silos” refer to the underlying OSM features contained within each site. The following OSM searches retrieved site data for oil wells:

KEY ‘industrial’ = ‘oil’ IN ‘North Dakota’

KEY ‘man made’ = ‘petroleum well’ IN ‘North Dakota’

Table 1 summarizes the results of the search for oil wells in North Dakota. The search returned polygons that reflect the outline of oil pads. A GIS procedure converted the polygons to centroid objects for point-based spatial analysis. Another GIS procedure merged 5640 well pad centroids and 7315 well points to produce a total of 12,955 well points.

Table 1. Results of OSM Search for Oil Wells in North Dakota.

Returned Objects	Objects	Action
Oil Well Pads	5640	Replaced with centroid
Oil Substations	4	Deleted
Oil Substation Gates	30	Deleted
Oil Well Points	7315	Retained

Cluster size, defined as the number of OSM features aggregated within each DBSCAN site, served as the spatial intensity attribute for G_i^* and Moran’s I . This metric reflects the density of mapped features per site. Geometry processing followed an explicit, type-specific rule set applied uniformly to both infrastructure layers: polygon features (well pads, silo footprints) were replaced with their centroids, line features (silos mapped as elongated structures) were replaced with their midpoint centroids, and point features were retained as originally digitized. All resulting centroids were merged into a single point layer prior to DBSCAN clustering.

To remove possible duplicate representations of the same object, the workflow conducted a nearest-neighbor distance audit comparing OSM Node, Line, and Polygon primitives. For silos, Node-to-Line and Polygon-to-Line distances were computed. For wells, Node-to-Line distances were computed, at threshold distances of 1, 5, 10, 25, 50, and 100 m. Candidate duplicate well Node–Line pairs within 10 m were inspected by satellite imagery and removed if obviously duplicated.

Figure 2 shows an example of the two cases of cluster centroid labeling: (1) the centroid of a well site containing a well pad with no wells, and (2) the centroid of a well site containing wells on two nearby well pads. Hence, this example shows two well sites with one containing no wells (left) and the other (right) containing a total of 12 well points on two nearby well pads.

The study established silo sites using the same clustering approach applied to well sites. The OSM search used:

KEY ‘man made’ = ‘silo’ IN ‘North Dakota’

The search returned polygon, line, and point representations of silos as shown in the example of Figure 3. Table 2 summarizes the results of the search for silos in North Dakota.

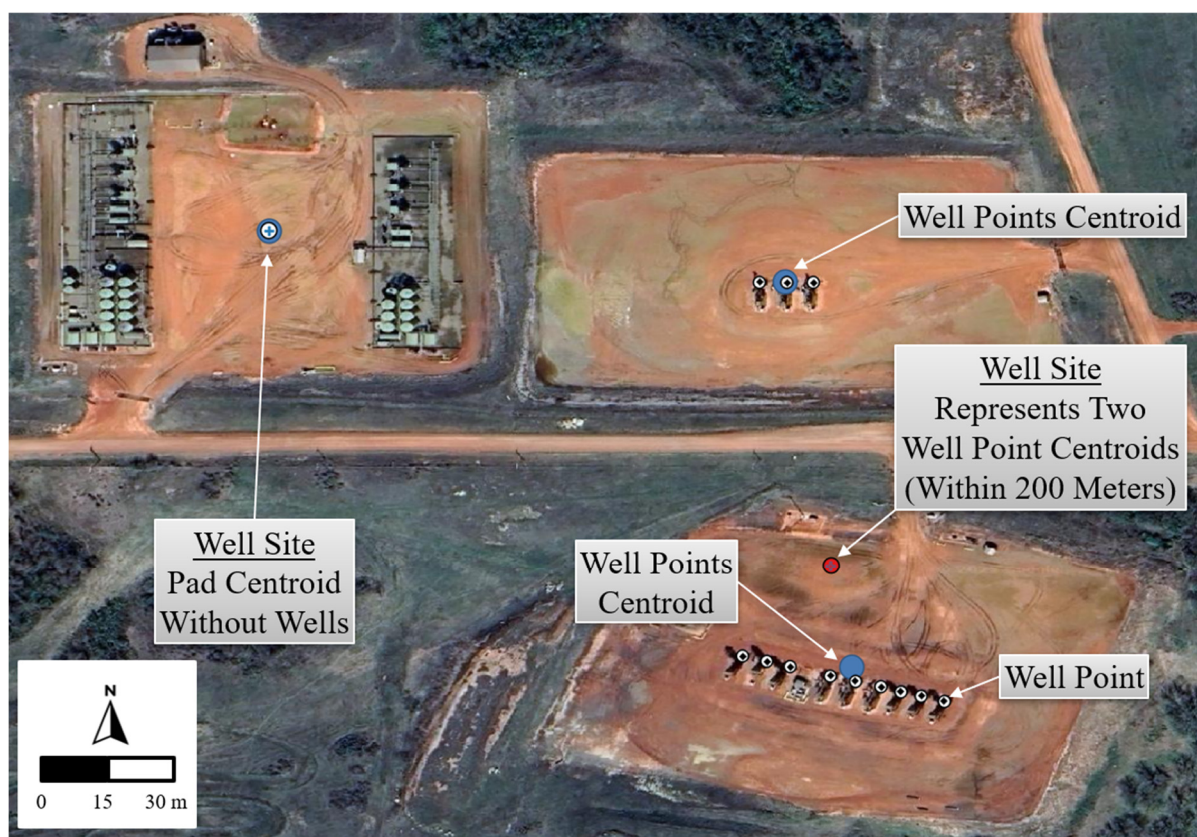


Figure 2. Clustering of Well Point and Well Pad Centroids.

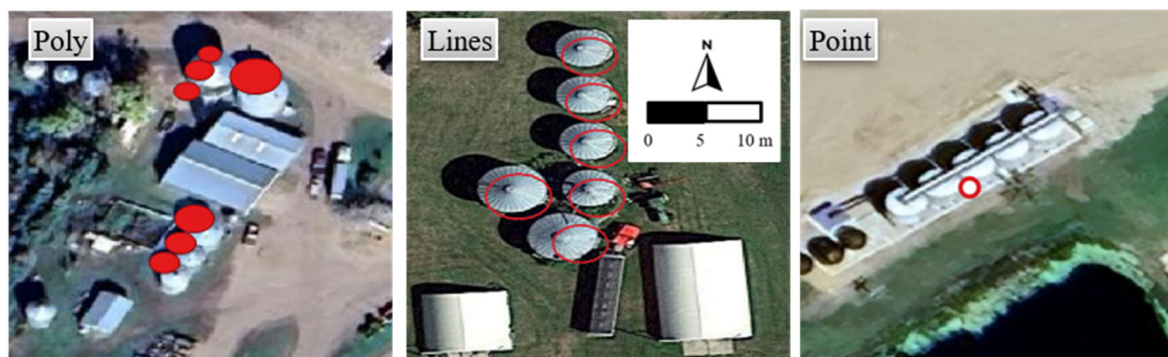


Figure 3. Polygon, Line, and Point representation of Silos in the OSM Database.

Table 2. Results of OSM Search for Silos in North Dakota.

Returned Objects	Objects	Action
Polygons	26,131	Replaced with centroid
Line	618	Replaced with centroid
Points	627	Replaced with centroid

A GIS procedure merged the total of 27,376 centroid points into a single layer. DBSCAN with a threshold of 200 m produced 4277 centroids representing silo sites. Figure 4 shows an example silo site (centroid) containing numerous silo points.



Figure 4. Cluster centroid representing a site containing numerous silo points.

3.2. Cluster Analysis

The study applied the Getis–Ord G_i^* statistic to identify local clusters of high and low values in the spatial distribution of oil wells and grain silos [28]. The G_i^* measure evaluates whether each site and its surrounding neighbors exhibit values that jointly differ from a spatially random pattern. That is, the statistic is considered significant only when its deviation from a spatially random pattern is large enough that the probability of observing it under the null hypothesis (p -value) is 0.05 or less. Hence, the statistic is a LISA that captures the intensity of clustering by considering the attribute value of the focal location together with those of nearby sites. This formulation makes G_i^* appropriate for detecting hotspot centers rather than broad regional trends.

Let x_j denote the attribute value at site j , which is the number of wells or silos. Let w_{ij} represent the spatial weight between sites i and j . Spatial weights define the structure of spatial interaction by specifying which locations influence each other and the strength of that influence. The weights matrix $W = [w_{ij}]$ in this analysis identifies n neighbors of each location i that are within a predefined distance band. The weight assigned to those neighbors is unity while all other entries are zero. A positive weight indicates that the attribute value at location j contributes to the local statistic for location i . The weights are row-standardized so that each row sums to one. This ensures that the influence of the neighborhood is comparable across locations.

The G_i^* statistic for location i is given by

$$G_i^* = \frac{\sum_{j=1}^n w_{ij}x_j - \bar{X}\sum_{j=1}^n w_{ij}}{S\sqrt{\frac{\left(\sum_{j=1}^n w_{ij}^2\right) - \left(\sum_{j=1}^n w_{ij}\right)^2/n}{n-1}}}, \quad (1)$$

where

$$\bar{X} = \frac{1}{n}\sum_{j=1}^n x_j \quad (2)$$

and

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - \bar{X}^2} \quad (3)$$

The numerator measures how much the weighted sum of values around location i deviates from the global mean value. The denominator standardizes this deviation, allowing significance testing under randomization. Large positive G_i^* values indicate significant hotspots where the focal location and its neighbors display unusually high counts. Large negative values indicate cold spots where the location and its neighbors show unusually low counts.

The G_i^* statistic was well suited for this study because oil extraction sites and grain storage facilities form spatially discrete clusters that reflect geological constraints, land-use practices, transportation access, and industrial zoning. Identifying significant hotspots highlights areas where concentrated energy or agricultural activity may impose infrastructure demands, environmental burdens, or opportunities for co-located development. Detecting cold spots identifies sparsely developed zones where expansion, buffering, or coordinated planning may be warranted. Therefore, the G_i^* framework provided an explicit and statistically rigorous method for detecting localized spatial patterns that support evidence-informed policy discussions related to energy and agricultural production in North Dakota.

3.3. Spatial Association

The local Moran's I statistic identifies spatial clusters and spatial outliers in a spatial distribution of sites [29]. This statistic evaluates whether the value at a focal site is similar to or different from the values observed among its neighbors. It measures the contribution of each location to the global spatial autocorrelation structure. Hence, this formulation is a complementary tool to the Getis–Ord G_i^* statistic. Whereas G_i^* detects hotspot centers formed by jointly high or low values at a focal location and its neighbors, local the Moran's I distinguishes five outcomes at each site: high values surrounded by high values (HH), low values surrounded by low values (LL), high values surrounded by low values (HL), low values surrounded by high values (LH), and non-significant (NS). These categories reveal both local clusters and local spatial outliers. Hence, it provides insight into spatial heterogeneity that G_i^* cannot capture.

Let x_i denote the attribute value at location i , which is the number of wells or silos, and let $\bar{x}$ represent the mean across all n locations. As before, the spatial weight w_{ij} defines whether location j is a neighbor of location i . The local Moran's I for location i is given by

$$I_i = \frac{(x_i - \bar{x})\sum_{j=1}^n w_{ij}(x_j - \bar{x})}{S^2} \quad (4)$$

where

$$S^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2 \quad (5)$$

is the global variance of the attribute. Positive and significant I_i values indicate locations that form HH or LL clusters, depending on whether the focal value is above or below the mean. Negative and significant values identify HL or LH spatial outliers, where the focal site differs sharply from its surrounding neighbors. Local Moran's I therefore complements the G_i^* statistic by characterizing the direction and type of local spatial association. It offers a more detailed understanding of localized structure in the distribution of oil wells and silos.

To evaluate cross-variable spatial dependence between silo and well distributions, this study employed the bivariate Moran's I statistic, which extends the local Moran's I framework by assessing whether the value of one variable at location i is associated with the spatial lag of another variable in its neighborhood. The bivariate local Moran's I is defined as

$$I_i^{xy} = x_i \sum_{j=1}^n w_{ij} y_j$$

where x_i and y_j are standardized values of the two variables (e.g., silo and well counts), and w_{ij} denotes spatial weights.

Inference relied on conditional randomization using 999 Monte Carlo permutations to generate pseudo p -values for both global and local statistics. The study constructed spatial weights using a distance band approach with row-standardization, ensuring comparability across locations with varying neighborhood sizes. The study assessed local significance using permutation-based pseudo p -values without formal multiple comparison correction. While this may increase the potential for false positives, robustness was supported through a multi-scale sensitivity framework (DBSCAN $\times$ bandwidth) to determine if key spatial patterns—particularly the prevalence of HL relationships and limited HH clusters—were stable across cluster threshold and neighborhood distance band combinations.

4. Results

The total oil well and silo sites identified by clustering were $7102 + 4277 = 11,379$. The following subsections discuss the results of G_i^* cluster analysis and the spatial association analysis using local Moran's I indicators.

4.1. OSM Versus NDIC Validation

A comparison between the OpenStreetMap (OSM) dataset and the North Dakota Industrial Commission (NDIC) records (Table 3) indicated that OSM captured approximately 29.7% of all wells reported by NDIC statewide (12,955 vs. 43,598 wells). This difference largely reflected the inclusion of inactive, decommissioned, and historical wells in NDIC records, as well as incomplete coverage in the crowd-sourced OSM dataset. When the comparison was restricted to active wells only, OSM coverage improved significantly to 58.6% (12,955 vs. 22,107 wells). This represented a substantial increase relative to total-well coverage and suggested that OSM more accurately reflected currently operational infrastructure.

Li et al. (2025) found that nonmetropolitan counties across the contiguous United States systematically exhibit lower OSM completeness and accuracy than metropolitan counties, a pattern attributed to lower volunteer-contributor density in rural areas [30]. Since North Dakota's oil- and agriculture-producing counties are predominantly rural, this volunteer-mapping bias mechanism is a plausible contributor to the coverage gap reported above, in addition to the temporal factors (inactive and historical wells) already discussed.

Table 3. OSM vs. NDIC Validation.

Metric	Value (Total Wells)	Value (Active Wells Only)
OSM Wells	12,955	12,955
NDIC Wells	43,598	22,107
Coverage (%)	29.71	58.60
County-level correlation	0.929	0.927

Furthermore, the county-level correlation of $r = 0.927$ (active wells) and $r = 0.929$ (total wells) between OSM and NDIC counts indicated that OSM reliably reproduced the spatial distribution of wells. The marginally higher correlation for total wells likely reflects NDIC's more complete historical record of established well locations, including inactive wells that appear consistently across both datasets. While some counties showed coverage values exceeding 100%, likely due to differences in feature representation, mapping practices, or temporal dataset updates, the overall pattern demonstrated that OSM provided a spatially consistent and reliable representation of active oil infrastructure.

This county-level correlation establishes agreement in the broad spatial distribution of wells but does not, by itself, establish accuracy at the level of the individual site pairs that form the analytical unit of the co-location analysis. Two implications follow. First, because NDIC coverage is not uniform across the state, under-mapping in OSM is unlikely to be spatially random; if omissions were concentrated in the same western counties where silo density is comparatively lower, the observed HH co-location count would tend to be conservative (an undercount) rather than inflated, whereas omissions concentrated near silo-dense counties would have the opposite effect. To test this directly a county-level Pearson correlation was computed between OSM well-coverage rate (OSM well count divided by NDIC well count, from the values in Table 4) and county-level silo density. The results were $r = 0.331$ (well coverage rate vs. raw silo count) and $r = 0.332$ (well coverage rate vs. silo share of the statewide total), both $p \approx 0.18$ across the 53 counties with at least one active well (Table 4). Neither correlation reached conventional statistical significance, so no systematic relationship between OSM well coverage and silo presence can be confirmed at this sample size. It is worth noting explicitly that the direction of the (non-significant) relationship is positive rather than negative—that is, counties with higher OSM well coverage tend, if anything, to also have somewhat more silos, which runs counter to the conservative estimate scenario rather than supporting it.

Table 4. OSM Coverage Bias Validation.

Comparison	r	p-Value
OSM well coverage rate versus raw silo count	0.331	0.180
OSM well coverage rate versus silo share (%)	0.332	0.179

Given the lack of statistical significance, this should not be interpreted as evidence of bias in either direction; the appropriate conclusion is that this analysis neither confirms nor rules out coverage-related bias in the HH count, and the screening-level framing of the reported estimate (rather than a claim of conservatism) is retained accordingly. Second, silo sites were not independently validated against USDA storage facility records because USDA maintains facility counts as tabular, county-level attribute data without a publicly available point or polygon layer suitable for a site-level spatial join; this is a genuine data-availability constraint rather than a design choice. The reported HH node count and its spatial extent should therefore be treated as a screening-level estimate of co-location rather than a precise inventory—not necessarily a conservative lower bound, since the

coverage-bias check above did not find a statistically significant directional bias. Hence, future work should pursue a parcel- or site-level match between OSM-derived clusters and NDIC well-permit coordinates (which, unlike NDIC county tallies, are individually geocoded) and, where available, county assessor or USDA-NASS elevator facility lists, to directly quantify the spatial bias introduced by incomplete OSM coverage. Taken together, these results suggested that although OSM underestimated total well counts, it effectively captured the core spatial patterns of active oil production, making it suitable for spatial analysis focused on structural patterns rather than absolute inventory completeness.

4.2. Duplicate Record Sensitivity Analysis

The nearest-neighbor distance audit described in Section 3.1 indicated that duplication was limited. Table 5 reports the number of Node–Line and Polygon–Line candidate pairs occurring within successive distance thresholds for each dataset. Among silos, only five Node–Line pairs and six Polygon–Line pairs occurred within 100 m, and no Node–Line pairs occurred within 10 m. Visual inspection of the closest candidate pairs showed that even silos separated by as little as 7–11 m represented distinct grain bins within the same facility rather than duplicate mappings of the same structure, and this distinction became increasingly obvious as separation distance increased. The well dataset exhibited a different pattern: seven Node–Line pairs occurred within 1 m, 157 within 5 m, and 417 within 10 m. The closest well pairs (0.006–0.92 m) frequently occupied the same well pad location and were difficult or impossible to distinguish using satellite imagery, representing the strongest duplicate candidates identified in the analysis. As separation distances increased to approximately 10–15 m, features increasingly represented different locations or components within the same well site rather than duplicate representations of the same object.

Table 5. Nearest-Neighbor Duplicate Distance Audit (candidate pairs within each distance threshold).

Dataset	≤1 m	≤5 m	≤10 m	≤25 m	≤50 m	≤100 m
Wells: Node→Lines	7	157	417	1345	1877	1997
Silos: Node→Lines	0	0	0	1	2	5
Silos: Polygon→Lines	0	0	1	2	5	6

As a follow-up robustness check, the 417 candidate duplicate well Node–Line pairs occurring within 10 m were removed and the full analytical pipeline—DBSCAN clustering, bivariate Moran’s I, Local Moran’s I, and FDR correction—was rerun on the deduplicated dataset across all tested DBSCAN and bandwidth combinations (Table 6). Global bivariate Moran’s I remained statistically significant across every combination tested ($p = 0.001$), and the HH hotspot structure was largely preserved: at broader bandwidths (10–20 km), HH retention rates after FDR correction generally exceeded 90% and reached 100% for several DBSCAN configurations. Smaller bandwidths (1–2.5 km) showed greater sensitivity to record removal, but the regional-scale hotspot pattern remained highly consistent. At the primary specification (DBSCAN = 200 m, bandwidth = 2.5 km), the deduplicated dataset yields 32 HH sites before FDR correction and 14 after correction, compared with the non-deduplicated values reported elsewhere in this manuscript; this comparison is discussed further in Section 5. Overall, these findings indicate that the potential duplicate records identified in the nearest-neighbor audit do not materially affect the primary conclusions of the spatial autocorrelation analysis.

Table 6. Bandwidth and Sensitivity Analysis (deduplicated wells dataset, 12,538 wells; see Section 4.7 for the duplicate-record audit and robustness re-analysis). HH values shown before and after Benjamini–Hochberg FDR correction.

DBSCAN (m)	BW (km)	Moran's I	<i>p</i> -Value	HH Before	HH After	HL After	LH After	LL After	HH Retained (%)
100	1.0	−0.2442	0.001	5	2	2441	544	2384	40.0
100	2.5	−0.2787	0.001	32	14	2453	1152	1456	43.75
100	5.0	−0.2798	0.001	67	55	2538	2172	1565	82.09
100	10.0	−0.2758	0.001	100	95	2574	3036	1707	95.0
100	20.0	−0.2704	0.001	117	109	2586	3783	1821	93.16
200	1.0	−0.2084	0.001	8	2	2079	322	2018	25.0
200	2.5	−0.2497	0.001	32	14	2152	903	1063	43.75
200	5.0	−0.2506	0.001	74	58	2276	1857	1210	78.38
200	10.0	−0.2482	0.001	103	100	2336	2672	1384	97.09
200	20.0	−0.2438	0.001	111	111	2364	3301	1505	100.0
300	1.0	−0.1734	0.001	8	1	1682	236	2012	12.5
300	2.5	−0.2203	0.001	28	10	1592	543	678	35.71
300	5.0	−0.2222	0.001	70	51	2042	1525	1009	72.86
300	10.0	−0.2207	0.001	97	87	2154	2299	1211	89.69
300	20.0	−0.2174	0.001	105	105	2204	2838	1366	100.0
400	1.0	−0.125	0.001	6	1	1673	186	2738	16.67
400	2.5	−0.1916	0.001	23	5	1542	366	1076	21.74
400	5.0	−0.1948	0.001	50	29	1660	1153	1258	58.0
400	10.0	−0.1943	0.001	76	70	1744	1931	1448	92.11
400	20.0	−0.1913	0.001	86	83	1762	2527	1542	96.51
500	1.0	−0.0663	0.001	5	1	1481	78	2727	20.0
500	2.5	−0.1515	0.001	14	4	1189	140	602	28.57
500	5.0	−0.1587	0.001	35	15	1453	683	934	42.86
500	10.0	−0.1588	0.001	70	59	1606	1505	1176	84.29
500	20.0	−0.1562	0.001	84	80	1676	2058	1379	95.24

4.3. Sensitivity Analysis

A sensitivity analysis was conducted to determine the bandwidth and threshold to select for the spatial cluster analysis. The sensitivity analysis evaluated five neighborhood bandwidths (BW) of {1, 2.5, 5, 10, 20} km and five DBSCAN thresholds (TH) of {100, 200, 300, 400, 500} m. The 25 parameter combinations collectively demonstrated a robust and consistent spatial relationship between silo and well sites. Table 6 shows that across all parameter combinations, Moran's I (MI) remain negative and statistically significant ($p = 0.001$), confirming a persistent pattern of spatial separation rather than co-location. The magnitude of Moran's I decreased systematically as the DBSCAN epsilon increased, indicating that larger clustering radii progressively smoothed spatial structure while the overall pattern remained stable, with HL clusters dominating and HH clusters remaining sparse. Spatial association metrics remained relatively stable between 2.5 km and 10 km bandwidths; larger bandwidths produced higher HH counts due to expanded neighborhood definitions without altering the underlying spatial configuration. The intermediate DBSCAN range (200–300 m) yielded the most stable results, avoiding both over-fragmentations observed at 100 m and over-aggregation at 500 m.

A persistence analysis confirmed that HH clusters remained geographically limited across parameter choices. Figure 5 illustrates the spatial distribution of HH persistence across DBSCAN and bandwidth sensitivity configurations. The results indicate that only a limited number of counties exhibit consistent HH clustering across the full range of parameter settings. Persistent HH signals are concentrated primarily in a small group of western counties, notably McKenzie, Williams, and Mountrail, where moderate persistence values

reflect repeated co-occurrence under multiple parameter configurations. In contrast, most counties across central and eastern North Dakota—including Burleigh, Cass, Stutsman, and Ramsey exhibit near-zero persistence, indicating that HH classification remains rare with parameter variation. Within the western region, HH persistence remained spatially fragmented rather than forming continuous zones. This pattern indicates that HH persistence is spatially limited and fragmented, with higher values concentrated in a small number of western counties. The sensitivity analysis identified the 200 m DBSCAN threshold as the most stable and interpretable configuration, with clustering patterns converging at this distance. Although larger thresholds (e.g., 300 m) yielded comparable outcomes, the 200 m threshold was selected as the minimum value at which stability is achieved. This choice minimized artificial aggregation of spatially distinct infrastructure while preserving locational precision. Furthermore, it aligned with the physical scale of oil well pads and grain storage facilities, effectively capturing features belonging to the same operational site while maintaining separation between neighboring installations.

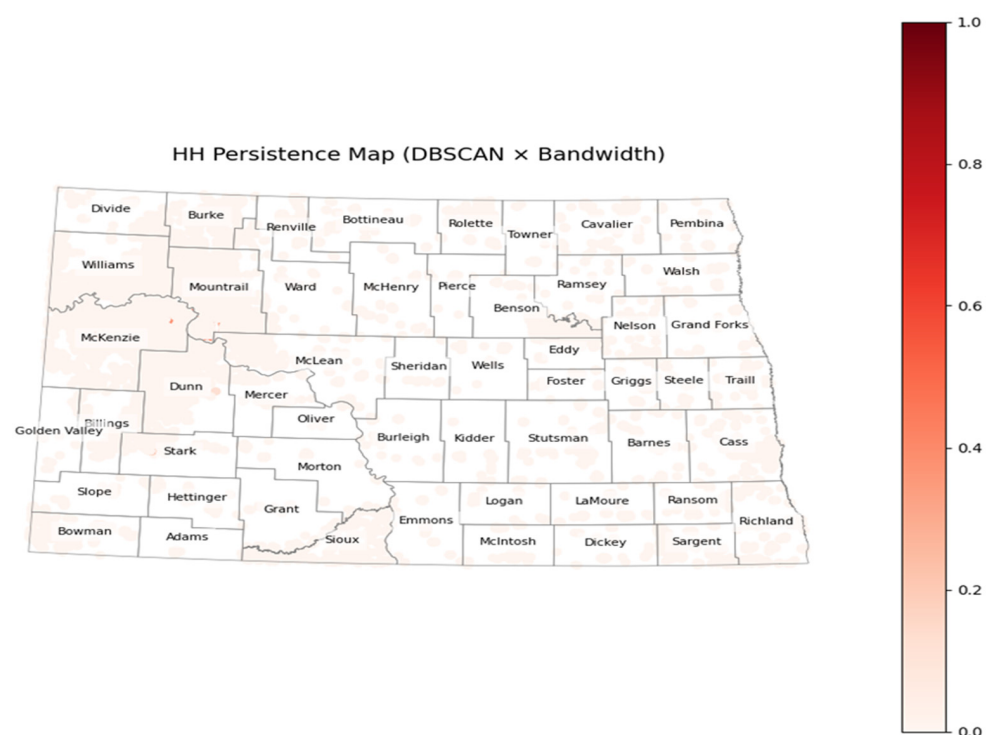


Figure 5. HH Persistence Map.

This alignment is supported by direct measurement rather than visual impression alone: a GIS-based digitization study of 3577 Bakken well pads [31] found average footprints of 5.26 acres (21,286 m²) for single-well pads and 8.60 acres (34,803 m²) for multi-well pads, corresponding to equivalent square-footprint linear dimensions of approximately 146 m and 187 m, respectively. The 200 m threshold therefore falls just above the upper end of this measured range, consistent with capturing a full well-pad footprint (including surrounding access infrastructure) while remaining well below the multi-kilometer scale of the surrounding land parcels, which further supports the threshold as a deliberate, evidence-based choice rather than an arbitrary round number.

To ground these parameter choices in the operational scale of the phenomenon rather than in statistical stability alone, the study defines “spatial co-location” operationally as the co-occurrence, within a single shared-access footprint, of a well site and a silo site close enough that the two could plausibly draw on the same local road segment, rail siding, or section-line access point. The 200 m DBSCAN threshold corresponds to the

physical footprint of a single well pad or elevator complex and is therefore appropriate for defining what counts as one operational site, not for defining cross-sector interaction distance. The 2.5 km neighborhood bandwidth used for the local Moran's I analysis was selected as the smallest bandwidth at which HH counts and Moran's I stabilize (Table 7), and it approximates a short local-haul radius (a few minutes of truck travel) rather than the longer-distance rail corridor or basin-wide scale (tens of kilometers) at which aggregate freight capacity competition described in the Introduction would actually operate.

Table 7. Marginal HH Cluster Growth Rate by Bandwidth Interval.

Bandwidth Interval	HH per km
1–2.5 km	8.0
2.5–5 km	17.6
5–10 km	7.2
10–20 km	1.7

To make this scale-dependence explicit rather than descriptive, the authors computed the marginal growth rate of the HH cluster count with respect to bandwidth (ΔHH per km) across the tested range in Table 6 and identify the bandwidth at which this marginal rate declines most sharply as the empirically supported transition from a site access scale to a corridor access scale. The sensitivity results therefore support two complementary readings: the 32 HH nodes identified at 2.5 km represent the most spatially precise, conservative estimate of immediate physical co-location, while the marginal rate analysis shows how the count of significant clusters grows at broader bandwidths (up to 112 at 20 km), indicating that a larger set of locations could plausibly experience shared logistics pressure if the relevant operational scale is the rail corridor rather than the individual road segment. Since the study cannot independently verify which scale governs actual freight or land-use interactions, the 2.5 km results are reported as the primary, spatially conservative finding, and the full multi-scale results in Table 5 are retained so that readers whose operational scale of interest differs from ours can identify the corresponding cluster counts.

4.4. Neighborhood Distribution

Guided by the sensitivity analysis, the Local Moran's I analysis used a neighborhood distance of 2.5 km (0.030 degrees). This threshold resulted in 356 sites without neighbors; these sites were excluded from the local Moran's I computation, as row-standardization is undefined for locations with no spatial connections. The sensitivity analysis also revealed that a much larger threshold of 0.214 degrees produced zero isolates but was too coarse to preserve the fine-scale spatial heterogeneity between oil well and silo site densities that is central to detecting localized co-location. Figure 6 shows the distribution of the number of neighbors, with the median of 8. The long tail indicates that a few sites have an exceeding large number of neighbors.

The above distribution establishes the spatial context for the inferential analyses that follow. Although it does not identify statistically significant clustering, it provides initial visual evidence of agricultural–energy spatial coincidence across counties, motivating the local cluster analysis in Section 4.5.

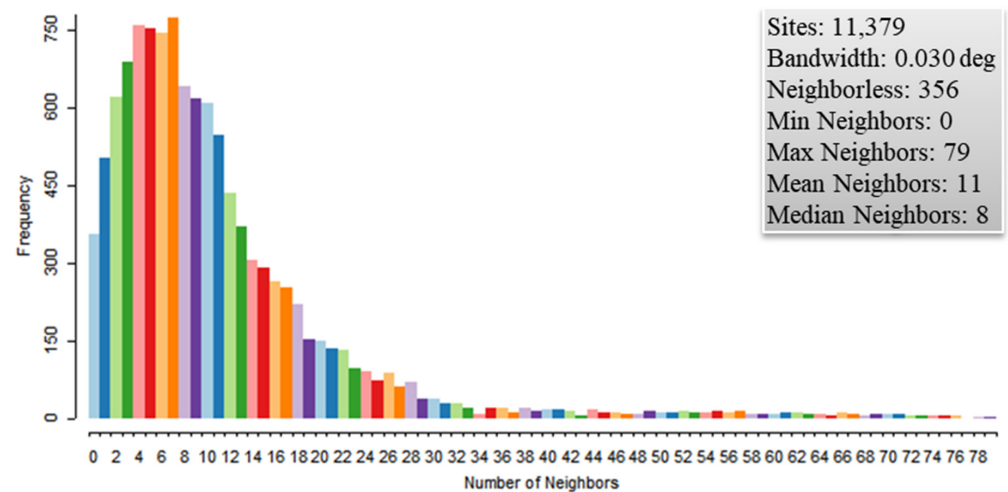


Figure 6. Distribution of the number of neighbors for each oil well or silo site.

4.5. Spatial Distribution

Figures 7 and 8 provide a county-level view of oil well counts and silo counts, respectively, revealing that counties such as Williams, McKenzie, and Mountrail contain large numbers of both wells and silos. However, the extent to which those locations exhibit significant spatial interaction with neighboring sites requires statistical methods of cluster analysis, covered in the following subsections.

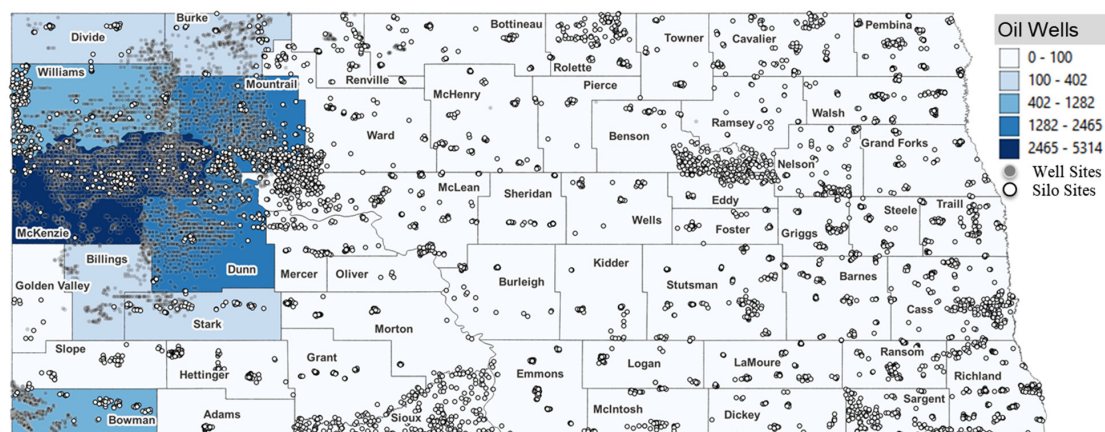


Figure 7. Oil well counts by county and silo sites relative to oil well sites.

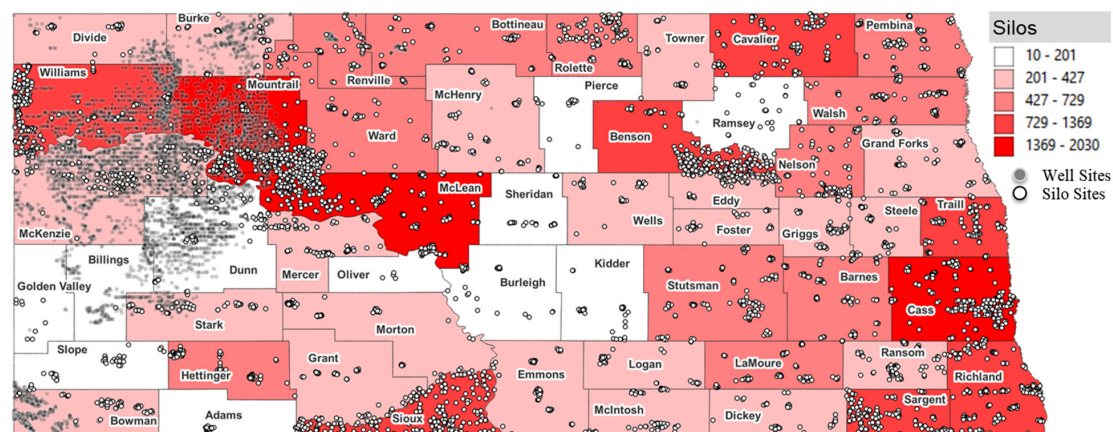


Figure 8. Silo counts by county and silo sites relative to oil well sites.

4.6. Cluster Identification

The global Moran's I for the combined infrastructure distribution was statistically significant and positive ($I = 0.1414$, $p < 0.01$), confirming non-random spatial clustering and providing the prerequisite foundation for local cluster analysis.

Figure 9a shows the results of the G_i^* cluster analysis. Red dots indicate hotspots where both a location and its neighbors have high values. Cold spots, shown as blue dots, require both the location and neighbors to have values that are low. Gray dots mark sites outside the threshold distance and therefore excluded from neighborhood analysis. The gray lines represent population tract boundaries within counties. The maps reveal clear and asymmetric clustering patterns for oil wells and grain silos across North Dakota.

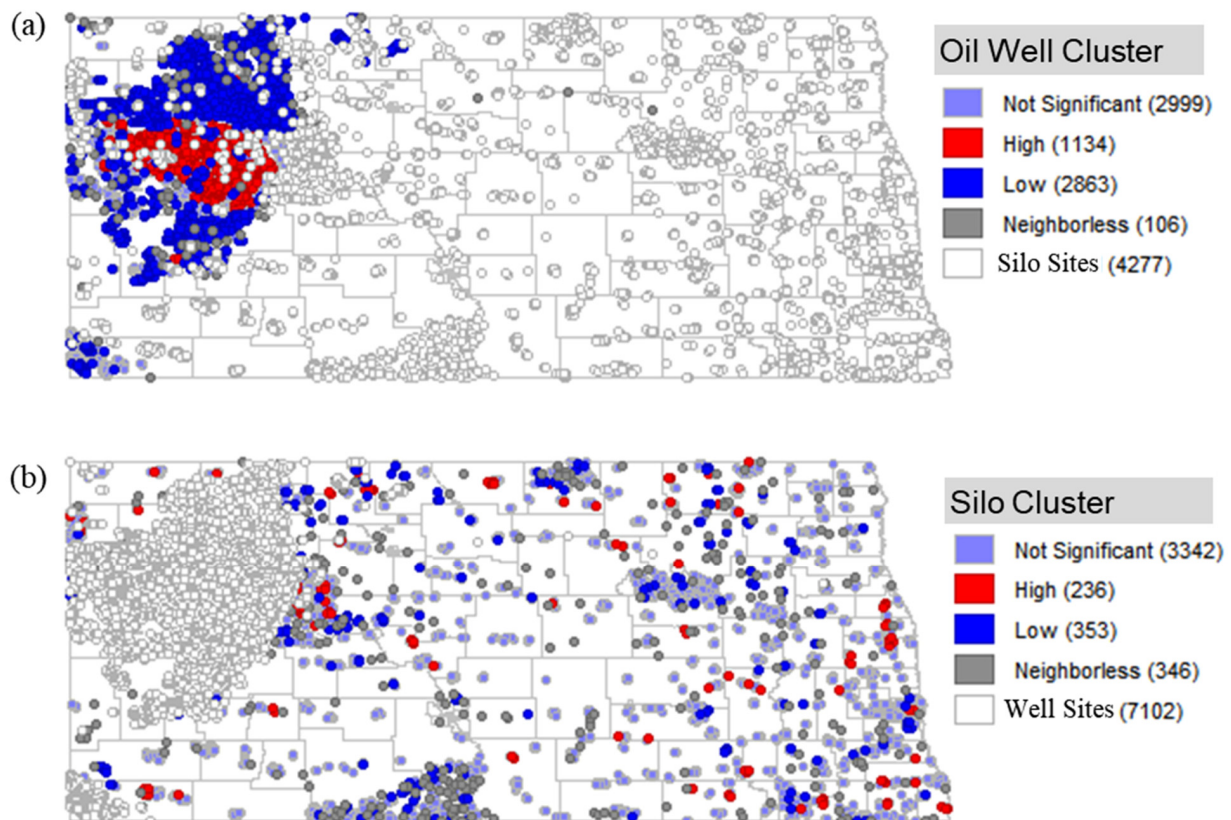


Figure 9. Significant clusters with high and low densities of (a) oil wells and (b) silos.

Figure 9a shows a dense and contiguous hotspot of wells sites in the western counties. The cluster is centered on the Bakken formation, which aligns with geological expectations. This region also contains a large cold-spot fringe where neighboring sites show significantly low well counts relative to the hotspot core. In contrast, Figure 9b shows that silo site hotspots occur primarily in central and eastern counties. This result also shows a limited geographic overlap between the two hotspot systems. That is, sites with intense drilling activity exhibit very few significant silo clusters. Similarly, agricultural regions with substantial silo concentration show almost no well-related hotspots. A few scattered points in both maps indicate isolated high-value sites surrounded by low-value neighbors. These sites represent localized industrial outposts rather than corridor-scale clustering. The results indicated substantial spatial separation between oil well and silo site hotspots, with limited geographic overlap and only minor zones of spatial interaction.

4.7. Spatial Interaction

Figures 10 and 11 show the significant clustering of oil well and silo sites, respectively.



Figure 10. Significant clustering of oil well sites.

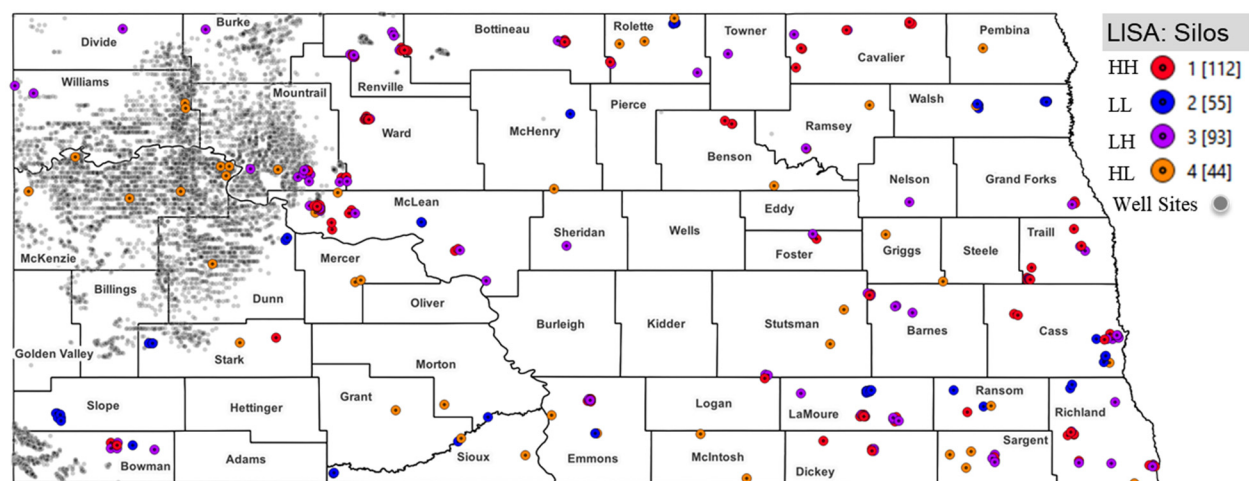


Figure 11. Significant clustering of silo sites based on local Moran's I analysis.

Figure 11 shows that the counties with significant silo clustering in high density oil well locations are McKenzie, Dunn, Mountrail, Williams, and McLean.

The local Moran's I analysis identified 44 silo sites exhibiting HL spatial heterogeneity. These HL sites represent locations with high silo counts surrounded by neighboring sites with relatively low silo counts in that part of the county. From the figure, it is evident that a few of these HL sites are located within major oil development areas, indicating locations where further investigation of shared infrastructure demand may be warranted.

Figure 12a shows the bivariate Moran's I clustering, where silo counts at focal locations (x) are evaluated against spatially lagged well counts in neighboring sites (y). This approach identifies locations where high silo density is associated with high surrounding well density, highlighting localized pockets of spatial co-location within an overall pattern of negative spatial association between wells and silos. Figure 12b highlights the HH cluster of 32 sites for clearer visualization.

The analysis identifies 32 High–High (HH) nodes, representing statistically significant areas of spatial co-location between infrastructure types. Although limited in number, these nodes constitute the primary locations where strong spatial clustering patterns occur. In terms of infrastructure representation, HH nodes encompass approximately 1.51% of silo sites and 0.29% of well sites, indicating that only a small fraction of total infrastructure is located within these statistically significant clusters. This suggests that co-location is highly selective, occurring in specific locations rather than across the entire study area. From a spatial perspective, HH zones occupy only 0.02% of the total state area, highlighting an extremely high degree of geographic concentration. In other words, a very small portion of

land contains the most statistically significant clustering patterns, emphasizing the localized nature of these zones. Despite their limited spatial footprint, HH nodes are distributed across approximately 5.6% of counties, demonstrating that these clusters are not confined to a single region but instead appear across multiple parts of the state, albeit sparsely.

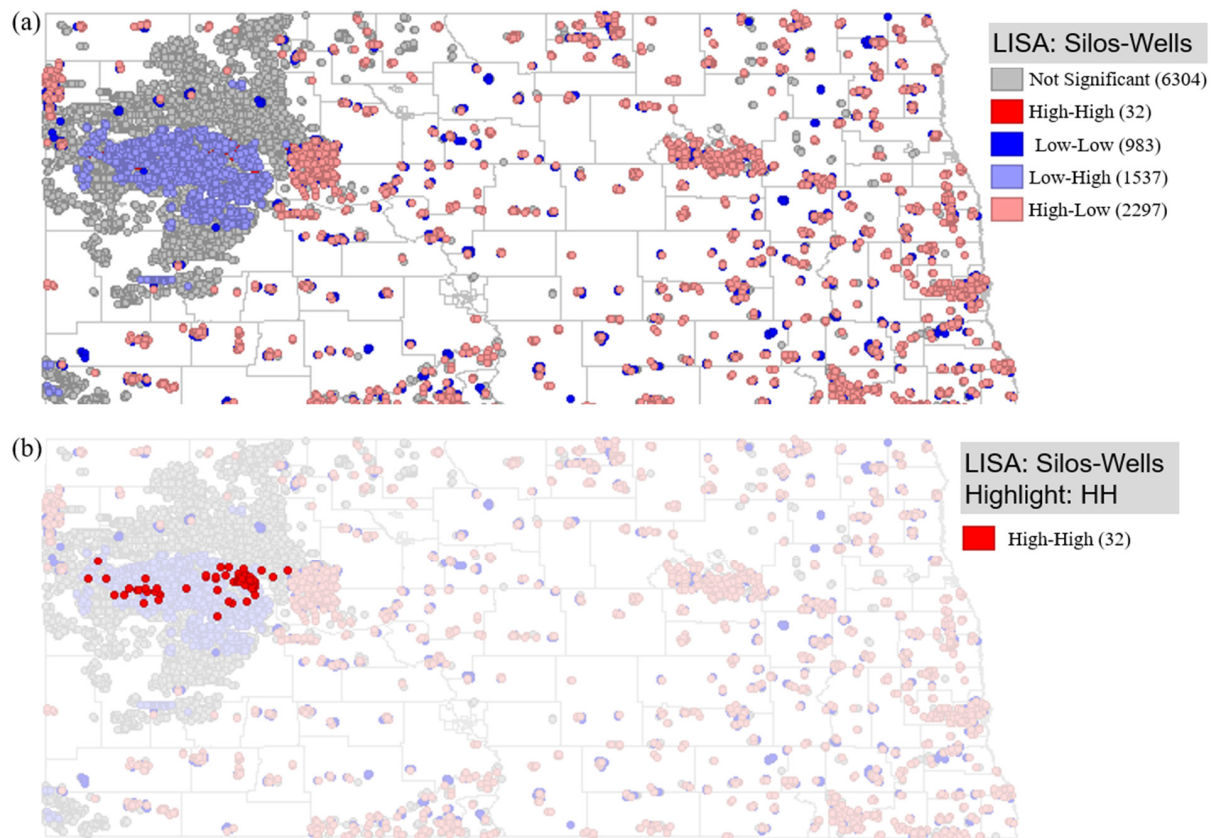


Figure 12. Bivariate Moran's for (a) Silos with Well neighbors and (b) a highlight of the HH cluster.

Figure 13a shows an exploded view of the HH cluster location and their counties.

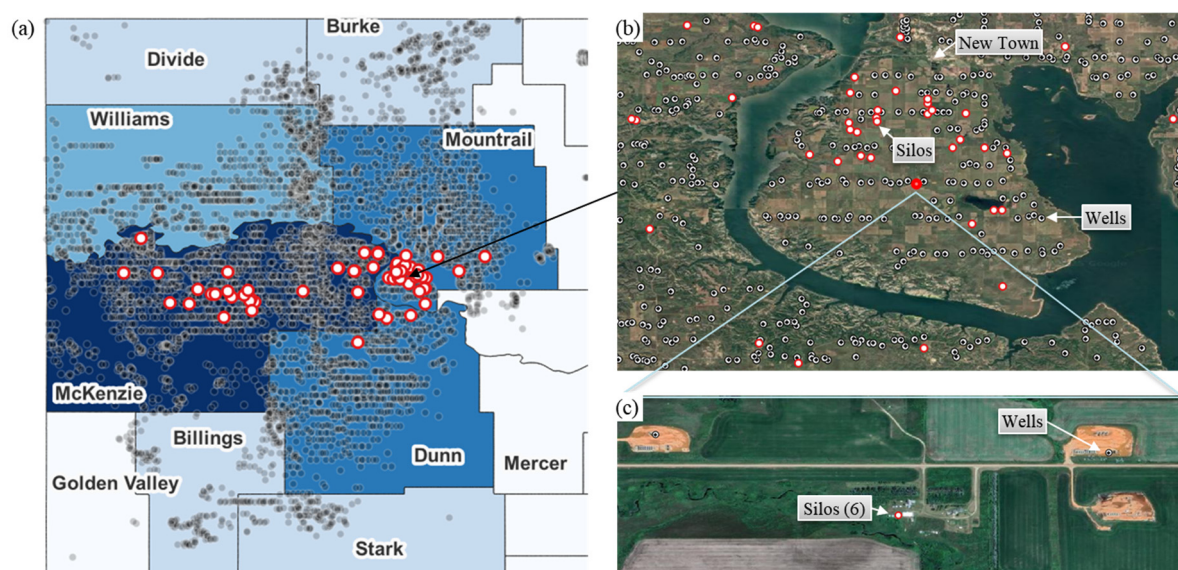


Figure 13. Counties with HH silos-wells. Source: Satellite Imagery by Google Maps (2026). (a) GIS representation. (b) Area panned in GIS with satellite basemap. (c) Area zoomed in GIS with satellite basemap.

These 32 silo sites, each associated with significantly high oil well neighbor counts, collectively contain 233 silos spanning across McKenzie, Mountrail, and Dunn counties. Figure 13b shows a satellite view of the highest intensity location, which is at the southwestern region of Mountrail County, directly south of New Town city. Figure 13c shows an exploded view of one of those high interaction locations that contains six silos surrounded by three well pads. To account for multiple testing, the Benjamini–Hochberg False Discovery Rate (FDR) correction [32] was applied to the permutation-based pseudo p -values from the local Moran’s I analysis. This step ensures that the identified clusters represent statistically reliable patterns rather than random noise.

The west–east divide, and the specific concentration of HH nodes in McKenzie, Mountrail, and Dunn counties, are consistent with, and can be explained by, three converging factors rather than an arbitrary artifact of the clustering procedure. Geologically, the productive extent of the Bakken–Three Forks petroleum system is confined to the Williston Basin in the western part of the state, which mechanically limits well site density to western counties regardless of agricultural conditions there. Agronomically and logistically, the highest-yield small-grain and oilseed production, and the elevator and rail siding network built to move it, are concentrated in the central and eastern Drift Prairie and Red River Valley, where soils and rail branch line density favor grain storage investment; this is why silo hotspots and well hotspots occupy largely disjoint zones. McKenzie, Mountrail, and Dunn counties emerge as the principal exception because they sit at the eastern and southern margin of the Bakken play, where productive well density is still high but overlaps a transition zone into rangeland and cropland historically served by county grain elevators along branch rail lines that also carry oilfield traffic; this geographic overlap of an active drilling margin with an established grain-handling corridor is the most plausible mechanism generating the observed HH nodes. As a quantitative complement to this qualitative interpretation, a county-level Poisson regression was fitted relating HH node incidence to silo count, active well count, rail density, and an oil $\times$ silo interaction term (Table 8).

Table 8. County-Level Driving Factor Poisson Regression Results.

Variable	Coefficient	p -Value
Intercept	1.553	0.021
Silos	0.0025	0.078
NDIC_active	0.0003	0.040
rail_density	−216.8494	<0.001
oil_silo_interaction	0.000001	0.001

Since FDR-corrected HH counts at the primary 2.5 km specification were too sparse for county-level regression (14 nodes statewide), this analysis instead used the broader 20 km bandwidth as a robustness check, where the FDR-corrected count reaches 112 HH nodes concentrated in the same three counties (Mountrail, 59; McKenzie, 40; Dunn, 13) identified at the primary scale. Following the deduplication described in Section 4.7, this 20 km robustness count changed only marginally, from 112 to 111 statewide; the county-level breakdown of this deduplicated 20 km result has not yet been separately verified and is noted here for transparency. The regression found a small but significant positive effect of active well count (coefficient = 0.0003, p = 0.040) and, most notably, a statistically significant oil $\times$ silo interaction term (p = 0.001), indicating that HH node incidence is better explained by the joint presence of both sectors than by either sector’s density alone; the silo count main effect was only marginal (p = 0.078) and rail density was not a dominant driver in this specification.

As a further descriptive check, an exploratory Agriculture–Energy Competition Index (AECI, defined as silo count $\times$ active well count/rail density) was computed for each county (Table 9). The three highest-AECI counties are McKenzie, Mountrail, and Williams, and Dunn ranks fourth; three of these four (McKenzie, Mountrail, Dunn) are also the counties with detected HH nodes, but Williams—despite ranking third by AECI—has zero detected HH nodes at this bandwidth (see Figure 14). This partial rather than complete overlap is informative: it shows that a high AECI value is not, by itself, sufficient for statistically significant co-location to emerge, and it illustrates precisely why the index is treated below as an exploratory screening tool rather than a confirmed predictor. These county-level results are consistent with, and provide corroborating quantitative support for, the geological-and-infrastructure explanation offered above, though as a broader-bandwidth robustness check rather than a test of the primary 2.5 km specification, and as an exploratory, descriptive index rather than a validated measure of infrastructure competition, the AECI pattern should be treated as suggestive of where joint oil–agriculture activity concentrates rather than as confirmed evidence of operational strain on shared logistics infrastructure.

Table 9. Top Agriculture–Energy Competition Index (AECI) Counties.

County	Competition Index	HH Nodes
McKenzie	1.87×10^8	40
Mountrail	1.50×10^8	59
Williams	9.67×10^7	0
Dunn	7.59×10^7	13

Agriculture–Energy Competition Index (AECI) North Dakota Counties

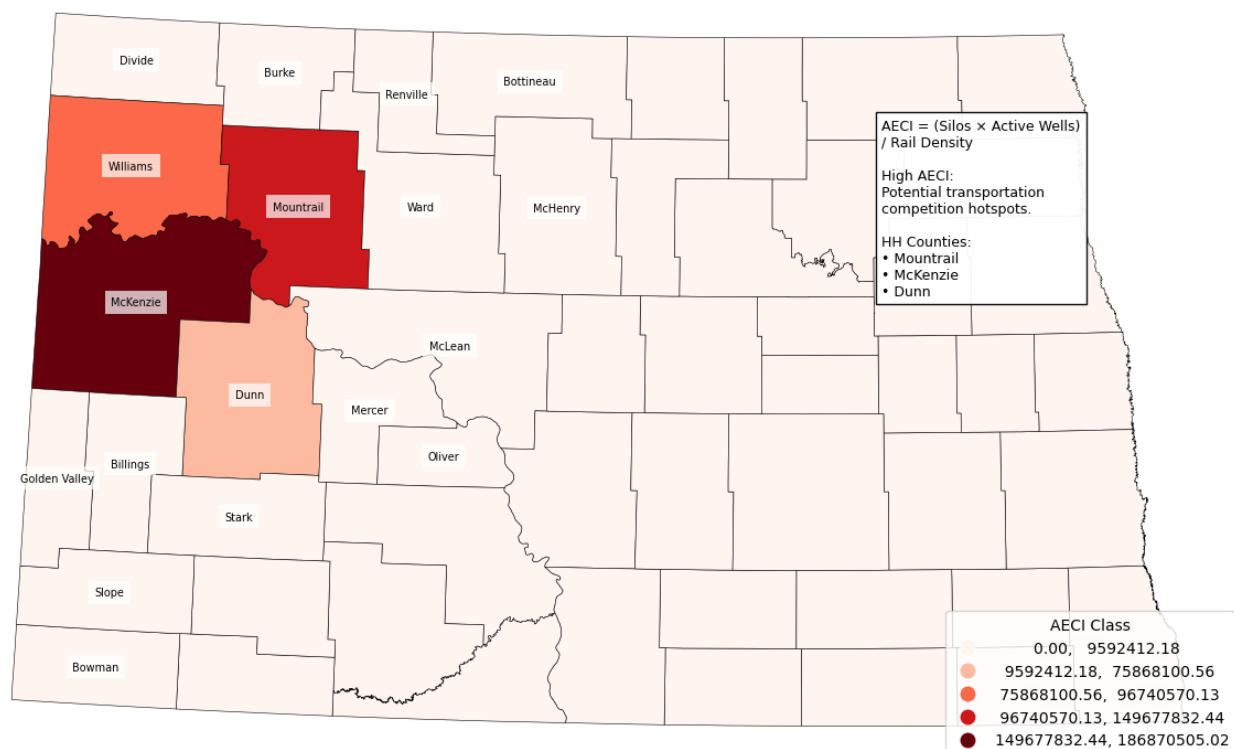


Figure 14. Agriculture–Energy Competition Index (AECI) map by North Dakota county. Darker shading indicates higher AECI values; McKenzie, Mountrail, Williams, and Dunn counties show the highest values.

Table 6 compares HH clusters before and after FDR correction across varying spatial bandwidths for DBSCAN = 200 m. The results show substantial reductions at smaller bandwidths. Following the duplicate record audit and deduplication described in Section 4.7, HH clusters decreased from 8 to 2 at 1 km and from 32 to 14 at 2.5 km. At intermediate scales, the reduction becomes less pronounced, with counts declining from 74 to 58 at 5 km. In contrast, only minor changes are observed at larger bandwidths, where clusters decrease slightly from 103 to 100 at 10 km, and essentially no change is observed at 20 km, where 111 of 111 HH clusters are retained (100%).

These findings indicate that a significant proportion of HH clusters detected at smaller spatial scales are likely driven by statistical noise, whereas clusters identified at larger bandwidths remain largely intact after correction. This confirms that spatial co-location patterns between wells and silos are more robust and reliable at regional scales, while fine-scale clustering is more susceptible to statistical artifacts (see Figure 14).

5. Discussion

The spatial patterns identified in this study provide a clearer understanding of how oil extraction and agricultural storage systems coexist within North Dakota. Interpreting these patterns in relation to the study's central goal shows that the two sectors interact only in narrowly defined locations rather than across the broader landscape. Oil wells form a concentrated western corridor shaped by the Bakken petroleum system, while grain silos dominate central and eastern regions where agronomic conditions and haul corridors support storage infrastructure [31]. This spatial separation indicates that broad, system-wide overlap is not the prevailing condition. The key finding is a limited set of HH co-location nodes where dense grain storage and well site clusters coincide within operational distances. This concentration highlights the highly localized nature of cross-sector spatial interaction.

These localized intersections complement prior research by identifying where potential interactions occur at finer spatial scales. Earlier studies documented regional-scale land-use displacement, agricultural disruption, and transportation congestion associated with drilling activity [4,6], but they relied on aggregated county-level data that obscure fine-scale dynamics. By contrast, this study demonstrates that meaningful interactions occur only within small, statistically identifiable corridors. These findings suggest that potential cross-sector interactions are spatially constrained and are more clearly observable using site-level analysis. The presence of HL patterns within the core drilling region further illustrates how agricultural storage activity diminishes rapidly in areas dominated by oil development, reinforcing the localized nature of these interactions and aligning with proximity-based effects reported in studies of agricultural adjustment near drilling sites [11,12].

The implications of these findings are most relevant for stakeholders responsible for managing freight movement and regional infrastructure. Since interactions are concentrated rather than widespread, planning efforts can focus on the specific corridors where shared transportation demand is most likely to intensify. These areas may experience simultaneous trucking, railcar, and storage demands from both sectors, consistent with prior reports of congestion and shipment delays during crude by rail [13,14]. Identifying these locations supports more efficient monitoring and resource allocation, consistent with sustainability principles that emphasize targeted intervention where system interactions are most likely to influence operational performance.

The results also contribute to the literature by providing empirical evidence that complements and sharpens earlier work. Studies examining the effects of drilling on crop cover, transportation networks, and rural infrastructure have emphasized broad regional impacts [4,5,16]. The present findings suggest that the spatial conditions associated with

such pressures are not uniform but instead emerge only where infrastructure footprints intersect at operational distances. This refinement helps reconcile inconsistencies in prior research by demonstrating that both widespread separation and localized interaction can occur simultaneously, depending on spatial scale. It also aligns with work highlighting the importance of fine-scale heterogeneity in understanding community-level responses to energy development [7].

The results show that oil–agriculture interactions in North Dakota are not uniform across space but are concentrated in a small number of statistically significant regions. These locations identify where shared transportation systems may be more susceptible to localized strain. By specifying where interactions occur and where they do not, the study provides a clearer basis for interpreting cross-sector dynamics and supports location-specific, data-driven assessment of regional infrastructure systems.

Limitations: The analysis relied on open-source geospatial data that may contain positional inaccuracies and inconsistent feature representations. North Dakota Industrial Commission (NDIC) tracking of approximately 19,500 active and inactive wells statewide suggests that the publicly available OSM data can lead to partial underrepresentation. However, NDIC reports strongest coverage in the Bakken region where core patterns are concentrated. Silo distributions broadly align with the National Agriculture Statistics Service (NASS) indicators from the United States Department of Agriculture (USDA), though on-farm storage is likely underrepresented. These limitations may influence the identification of HH co-location nodes, particularly in sparsely mapped areas; however, sensitivity analysis indicates that the primary spatial patterns are robust. Future research should integrate datasets such as NDIC well records and USDA storage capacity data to improve spatial accuracy and completeness, and extend the framework to related applications such as supply chain analysis for perishable and high-value commodities.

The cross-sectional design of the workflow does not capture temporal changes in infrastructure development or freight demand, and the study does not incorporate transportation network performance data that could further clarify how spatial proximity translates into operational strain. These limitations highlight opportunities for future research to integrate longitudinal datasets, administrative records, and network-level performance measures.

As a further limitation, this study identifies statistically significant spatial proximity between mapped infrastructure sites; it does not measure freight congestion, land values, environmental exposure, or any other operational outcome, and it does not establish that co-located sites are experiencing elevated cross-sectoral stress. Spatial proximity is a necessary but not sufficient condition for the operational interactions (freight competition, land-use conflict, environmental effects) discussed in the Introduction, and the findings should not be interpreted as direct evidence that such interactions are occurring at the 32 identified HH nodes. This caveat applies equally to the county-level oil $\times$ silo interaction effect and the exploratory AEI reported above: both are statistically significant, correlational patterns describing where oil and agricultural infrastructure co-occur, and neither measures, nor should be read as demonstrating, actual freight congestion, land-use conflict, or infrastructure strain. Confirming real-world operational effects would require linking these locations to independent records such as rail siding utilization, permit or complaint filings, or remote-sensing-based land-use change, which is beyond the scope of the present spatial pattern analysis and is proposed as a natural extension of this work.

The transferability of the proposed framework depends on several conditions that have not been empirically validated beyond this case study. The approach requires sufficiently dense and consistent spatial data, making it most reliable in regions with strong OpenStreetMap coverage comparable to North Dakota. DBSCAN clustering depends on adequate feature density, and results may become unstable in sparsely mapped environ-

ments. Computational demands scale with dataset size but remain manageable for typical regional analyses. Under these conditions, the workflow is potentially transferable to other regions and infrastructure co-location problems, though such transferability remains a direction for future empirical validation.

6. Conclusions

This study examined how oil extraction and agricultural storage systems are distributed across North Dakota, addressing the broader question of where these two major land-use activities intersect within a shared regional landscape. The spatial analysis revealed a clear geographic divide: oil well hotspots are concentrated in the western Bakken region, while silo site hotspots dominate central and eastern counties. Within this broader separation, the 32 statistically significant high–high (HH) co-location nodes identified by the analysis isolate the limited set of locations where both activities intensify in close proximity. These nodes represent the primary points at which the two sectors are spatially proximate at scales consistent with shared local infrastructure access. The study identifies where such proximity occurs; it does not, on its own, confirm that freight competition, land-use conflict, or environmental effects are actually taking place at these nodes, and the conclusions below are bounded accordingly.

Several localized intersections occur within regions previously associated with elevated transportation demand. Their spatial configuration indicates that cross-sector pressures arise not as a statewide condition, but as concentrated, place-specific interactions shaped by the alignment of drilling activity, storage density, and freight routing patterns. This distinction clarifies the nature of oil–agriculture coexistence in North Dakota by separating widespread spatial separation from a focused set of meaningful overlaps.

The results provide several potential applications for researchers and practitioners. For analysts and planners, the study provides a clearer empirical basis for identifying where infrastructure systems are most likely to experience localized strain, allowing attention to be directed toward the specific corridors where oil extraction and agricultural storage coincide. For the academic community, the findings advance current knowledge by demonstrating that cross-sector interactions are highly concentrated and statistically bounded rather than broadly distributed across the landscape. These insights support practical applications such as prioritizing monitoring efforts in HH nodes, refining regional assessments of freight vulnerability, and improving situational awareness in areas where multiple land-use systems converge.

These recommendations can be differentiated by the type of county identified through the spatial analysis rather than applied uniformly statewide. In oil-dominated western counties with little silo presence, planning attention is best directed at road and bridge maintenance funding tied to well pad traffic, since Section 2 documents pavement and bridge damage from drilling-related truck traffic that is independent of any grain-storage overlap. In agriculture-dominated central and eastern counties with little well presence, the findings indicate that rail capacity and elevator-siting decisions can proceed largely without accounting for oilfield freight competition, since the spatial evidence shows negligible well presence in these areas. In the small set of HH co-located counties—McKenzie, Mountrail, and Dunn—the recommended action is targeted rather than general: county road departments and rail branch line planners should prioritize joint capacity assessments specifically at the 32 identified node locations (e.g., shared-access road segments and grain elevator rail sidings within the Bakken transition zone), and land-use permitting in these counties can use the HH node map as a screening layer to flag new well pad or elevator siting applications for closer review, rather than applying additional review statewide. This targeting is largely, though not completely, corroborated by the county-level AEI analysis

in Section 4: three of the four highest-AECI counties (McKenzie, Mountrail, Dunn) match the HH node counties, while the fourth (Williams) has a high AECI value but no detected HH nodes, highlighting that AECI is used here as a supporting screening signal rather than as independent proof of the targeting. This differentiation replaces a single generic recommendation with three zone-specific, directly implementable actions.

The methodological approach may have broader relevance in regions with comparable spatial data availability and infrastructure characteristics. The integration of open geospatial data with clustering and spatial association techniques can be adapted to examine other forms of infrastructure co-location, including interactions between energy development, transportation assets, and environmental features. This generalizability highlights the long-term value of the workflow for studies seeking to understand how multiple land-use systems overlap at operational scales and how spatial intensity patterns shape regional infrastructure dynamics.

Building on the limitations discussed in Section 5, three directions for future research follow directly from this study: (1) site-level validation of both well and silo locations against official NDIC well permit coordinates and USDA-NASS facility records, to move beyond the county-level and screening-level validation used here; (2) empirical linkage of the identified HH nodes to independent operational indicators such as rail siding utilization, permit filings, or remote-sensing-based land-use change, to test whether spatial co-location corresponds to measurable freight or land-use impacts; and (3) extension of the driving factor regression introduced in Section 4 with additional covariates (e.g., land-use zoning, parcel-level ownership) and finer temporal resolution, to better isolate the mechanisms generating the observed oil $\times$ silo interaction effect.

Overall, the study contributes to a more precise understanding of where and how oil extraction and agricultural storage intersect within North Dakota. By identifying the specific spatial contexts in which overlap occurs and where it does not, the analysis supports more informed interpretation of cross-sector dynamics without extending beyond the evidence presented.

Author Contributions: Conceptualization, E.L.M.L. and R.B.; methodology, E.L.M.L. and R.B.; software, E.L.M.L. and R.B.; validation, E.L.M.L. and R.B.; formal analysis, E.L.M.L. and R.B.; investigation, E.L.M.L. and R.B.; resources, R.B.; data curation, E.L.M.L. and R.B.; writing—original draft preparation, E.L.M.L. and R.B.; writing—review and editing, E.L.M.L. and R.B.; visualization, E.L.M.L. and R.B.; supervision, R.B.; project administration, R.B.; funding acquisition, R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Office of the Chief Economist; Agricultural Marketing Service. *Rail Service Challenges in the Upper Midwest: Implications for Agricultural Sectors—Preliminary Analysis of the 2013–2014 Situation*; U.S. Department of Agriculture: Washington, DC, USA, 2015.
2. Frittelli, J.; Andrews, A.; Parfomak, P.W.; Pirog, R.; Ramseur, J.L.; Ratner, M. *US Rail Transportation of Crude Oil: Background and Issues for Congress*; Congressional Research Service: Washington, DC, USA, 2014.
3. Bushnell, J.B.; Hughes, J.E.; Smith, A. Food versus fuel? Impacts of the North Dakota oil boom on agricultural prices. *J. Assoc. Environ. Resour. Econ.* **2022**, *9*, 79–112. [[CrossRef](#)]

4. Fitzgerald, T.; Kuwayama, Y.; Olmstead, S.; Thompson, A. Dynamic Impacts of U.S. Energy Development on Agricultural Land Use. *Energy Policy* **2020**, *137*, 111163. [CrossRef]
5. McGranahan, D.A.; Fernando, F.N.; Kirkwood, M.L. Reflections on a Boom: Perceptions of Energy Development Impacts in the Bakken Oil Patch Inform Environmental Science & Policy Priorities. *Sci. Total Environ.* **2017**, *599–600*, 1993–2018. [CrossRef] [PubMed]
6. Allred, B.W.; Smith, W.K.; Twidwell, D.; Haggerty, J.H.; Running, S.W.; Naugle, D.E.; Fuhlendorf, S.D. Ecosystem services lost to oil and gas in North America. *Science* **2015**, *348*, 401–402. [CrossRef] [PubMed]
7. Gershenson, C.; Jin, O.; Haas, J.; Desmond, M. Fracking Evictions: Housing Instability in a Fossil Fuel Boomtown. *Soc. Nat. Resour.* **2024**, *37*, 347–364.
8. Preston, T.M.; Kim, K. Land cover changes associated with recent energy development in the Williston Basin; Northern Great Plains, USA. *Sci. Total Environ.* **2016**, *566–567*, 1511–1518. [CrossRef] [PubMed]
9. Rakitan, T.J. The Impact of the U.S. Shale Boom on Agriculture: Evidence from North Dakota. *SSRN Electron. J.* 2018. Available online: <https://ssrn.com/abstract=3135519> or <http://dx.doi.org/10.2139/ssrn.3135519> (accessed on 1 June 2026). [CrossRef]
10. Xu, Y.; Elbakidze, L.; Etienne, X. Unconventional oil and gas development and agricultural land-use in the US. *J. Agric. Resour. Econ.* **2023**, *48*, 219–236.
11. Hoy, K.A.; Xiarchos, I.M.; Kelsey, T.W.; Brasier, K.J.; Glenna, L.L. Marcellus shale gas development and farming. *Agric. Resour. Econ. Rev.* **2018**, *47*, 634–664. [CrossRef]
12. Smith, K.K.; Haggerty, J.H.; Kay, D.L.; Coupal, R. Using Shared Services to Mitigate Boomtown Impacts in the Bakken Shale Play: Resourcefulness or Over-adaptation? *J. Rural Community Dev.* **2019**, *14*, 66–86.
13. Ortiz, L.V. No Train No Gain: The Impact of Increased Demand for Rail Services By the Energy Sector on Wheat Prices—A Preliminary Analysis. *Int. J. Food Agric. Econ.* **2016**, *4*, 103–125.
14. McCarthy, L.A.; Park, S.; Casazza, P.; Giancola, A.R. *Impacts of Energy Developments on U.S. Roads and Bridges*; The National Academies Press: Washington, DC, USA, 2015.
15. Dharmadhikari, N.; Lee, E.; Kayabas, P. The Lifecycle Benefit-Cost Analysis for a Rural Bridge Construction to Support Energy Transportation. *Infrastructures* **2016**, *1*, 2. [CrossRef]
16. Rahm, D.; Fields, B.; Farmer, J.L. Transportation Impacts of Fracking in the Eagle Ford Shale Development in Rural South Texas: Perceptives of Local Government Officials. *J. Rural Community Dev.* **2015**, *10*, 78–99. [CrossRef]
17. Kubas, A.; Vachal, K. *Impact of Energy Sector Growth on Perceived Transportation Safety in the Seventeen-County Oil Region of Western North Dakota: A Three-Year Case Study*; Mountain-Plains Consortium: Fargo, ND, USA, 2015.
18. Horner, R.M.; Harto, C.B.; Jackson, R.B.; Lowry, E.R.; Brandt, A.R.; Yeskoo, T.W.; Murphy, D.J.; Clark, C.E. Water Use and Management in the Bakken Shale Oil Play in North Dakota. *Environ. Sci. Technol.* **2016**, *50*, 3275–3282. [CrossRef] [PubMed]
19. Hearne, R.R.; Fernando, F. Strategies for Community and Industry Water Management in the Oil Producing Region of North Dakota. *Water* **2016**, *8*, 331. [CrossRef]
20. Fernando, F.N.; Cooley, D.R. Socioeconomic System of the Oil Boom and Rural Community Development in Western North Dakota. *Rural Sociol.* **2016**, *81*, 407–444. [CrossRef]
21. Smith, K.K.; Haggerty, J.H. Exploitable ambiguities & the unruliness of natural resource dependence: Public infrastructure in North Dakota’s Bakken shale formation. *J. Rural Stud.* **2020**, *80*, 13–22. [CrossRef]
22. Spiess, J.; McGranahan, D.A.; Whippo, C.; Poling, B.; Daigh, A.L.; Hovick, T. Bird and Invertebrate Communities Appear Unaffected by Fracking Traffic along Rural Roads Despite Dust Emissions. *Ambio* **2020**, *49*, 605–615. [PubMed]
23. Boslett, A.; Hill, E.; Ma, L.; Zhang, L. Rural Light Pollution from Shale Gas Development and Associated Sleep and Subjective Well-being. *Resour. Energy Econ.* **2021**, *64*, 101220. [CrossRef] [PubMed]
24. Burg, V.D.; Post, M.; Otto, C.; MacDonald, G. Trending against the Grain: Bird Population Responses to Expanding Nergy Portfolios in the US Northern Great Plains. *Ecol. Appl.* **2023**, *33*, e2904. [CrossRef] [PubMed]
25. Addey, K.A. A Diagnostic Test for the ‘Dutch Disease’ in the U.S.A using the ARDL Bounds Testing Technique. *Res. Appl. Econ.* **2019**, *11*, 32–46. [CrossRef]
26. Fernando, F.N.; Goreham, G.A. A Tale of Two Rural Cities: Dynamics of community Capitals during a North Dakota Oil Boom. *Community Dev.* **2018**, *49*, 274–291. [CrossRef]
27. Lee, E.; Chakraborty, D.; McDonald, M. Predicting Oil Production Sites for Planning Road Infrastructure: Trip Generation Using SIR Epidemic Model. *Infrastructures* **2021**, *6*, 14. [CrossRef]
28. Getis, A.; Ord, J.K. The analysis of spatial association by use of distance statistics. *Geogr. Anal.* **1992**, *24*, 189–206. [CrossRef]
29. Anselin, L. Local indicators of spatial association—LISA. *Geogr. Anal.* **1995**, *27*, 93–115. [CrossRef]
30. Li, D.; Gamage, M.M.; Cao, J.; Jiang, Y.; Li, Y.; Zhang, L.; Nguyen, Q.C.; Nguyen, T.T.; Siebeneck, L.K. Mapping the completeness and positional accuracy of OpenStreetMap road data at the county level in the contiguous United States. *Trans. GIS* **2025**, *29*, e70077. [CrossRef]

31. Sonnenberg, S.A. The Bakken–Three Forks super giant play, Williston Basin. *AAPG Bull.* **2020**, *104*, 2557–2601. [[CrossRef](#)]
32. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **1995**, *57*, 289–300. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.