Vector fields as a framework for modelling the mobility of commodities

Sima Farokhnejad[1,*], Angélica S. da Mata[2], Mariana Macedo[3,4],
Ronaldo Menezes[1,*]

1 Department of Computer Science, University of Exeter, UK
2 Department of Physics, Federal University of Lavras, Minas Gerais, Brazil
3 Department of Data Science, Northeastern University, London, UK
4 Khoury College of Computer Sciences, Northeastern University, Boston, United States

* sfarokhnejad@biocomplexlab.org, r.menezes@exeter.ac.uk

Abstract

Commodities, including livestock, flow through trade networks across the world, with trajectories that can be effectively captured using mobility pattern modelling approaches similar to those used in human mobility studies. However, documenting these movements comprehensively presents significant challenges; it can be unrealistic, costly, and may conflict with data protection regulations. As a result, mobility datasets typically contain inherent uncertainties due to sparsity and limitations in data collection methods. Origin-destination (OD) representations offer a powerful framework for modelling movement patterns and are widely adopted in mobility studies. However, these matrices possess inherent structural limitations: locations absent from the OD framework lack spatial information regarding potential mobility directions and intensities. This spatial incompleteness creates analytical gaps across different geographical granularities, constraining our ability to fully characterise movement patterns in underrepresented areas. In this study, we introduce a vector-field-based method to address these persistent data challenges, transforming OD data into vector fields that capture spatial flow patterns more comprehensively, enabling us to also study mobility directions in a robust manner. We utilise cattle trade data from Minas Gerais, Brazil, as our case study for commodity flows. This region's extensive livestock trading network makes it an ideal test case for our vector-field methodology. Cattle movements are particularly significant as they directly impact disease transmission, including foot-and-mouth disease. Accurately modelling these flows enables more effective disease surveillance and control strategies, with implications for both animal health and economic stability. Our vector-field approach reveals fundamental patterns in commodity mobility and can infer movement information for unrepresented locations; a critical factor in modelling scenarios such as disease spread. Our approach offers an alternative to traditional network-based models, enhancing our capacity to infer mobility patterns from incomplete datasets and advancing our understanding of large-scale commodity trades.

Introduction

Modelling mobility patterns has long been the focus of many studies [1–5], with network-based methods playing a central role in capturing the flow of commodities across the world. While networks have proven useful in many contexts, they often miss key

nuances, particularly when applied to large and complex datasets. Mobility datasets of commodities, such as those tracking livestock trade, tend to be vast, containing many individual records [6], which makes network generation both computationally expensive and difficult to interpret [7]. These datasets are also subject to uncertainty [8,9], not only because of errors in data collection but also because it is practically impossible to capture every detail of large-scale movements completely. This uncertainty is inherent to the nature of mobility data [10], meaning that complete accuracy in tracking movement is unattainable. For instance, most work in human mobility rely on specific datasets that often capture only a fraction of reality, such as call detail records (CDRs), extended detail records (XDRs), and location-based social networks (LBSNs) [11,12]. As a result, modelling with incomplete data – often the only available option – can introduce inaccuracies and uncertainties in contexts such as mobility analysis or disease spread prediction. This highlights the importance of addressing the issue of missing data and exploring methods to enhance the accuracy of mobility modelling and its applications.

Another common limitation is that network models typically rely on origin-destination (OD) matrices, which focus on the starting and ending points of movements [13]. However, this representation is also incomplete by design. Locations that are not explicitly involved in the recorded trades are ignored, even though they may play significant roles in the overall flow of commodities. For example, intermediate locations through which commodities are transported can still influence the dynamics of the trade, yet they are not represented in OD matrices. This leads to gaps in our understanding of movement patterns, particularly in regions where data are missing or uncertain. If these unknown locations are cities or towns, for instance, we are unable to assess their importance in the holistic view of the system.

Take, for example, the scenario in Fig 1. The map in panel A represents the typical origin-destination (OD) model of commodity movements, where only the start and end points of movements are captured. In this representation, certain locations that may serve as critical intersections for multiple movements are excluded because they are neither origins nor destinations (illustrated as green nodes in Fig 1B). Even though these locations are not directly involved in the movements, they may play an essential role in the overall flow of commodities. In our modelling, we consider the movements as a series of vectors and calculate the resulting vector that represents that region. Once that is done, we can use interpolation to calculate all the missing vectors for regions that do not have information (filling the blue cells in Fig 1C). The interpolation approach is based on the reasonable assumption, particularly in real-world scenarios, that neighbouring locations are likely to participate in and influence similar activities. Thus, our approach aims to address these data gaps by inferring the likelihood of flow for unrepresented locations, offering a more comprehensive view of the system dynamics. Furthermore, interpolation accuracy and robustness depends on the amount of missing information as explained later in Fig 3.

Fig 1. Data sparsity can impact network modelling as it can disregard important locations. (A) Typical OD model of commodity movements where locations not included in the recorded movements are excluded entirely from the network. **(B)** Certain areas that serve as spatial junctions are left out simply because they are neither origins nor destinations, despite their potential importance in facilitating flow. **(C)** Data gaps result in missing regions (*e.g.*, blue cells), limiting our ability to assess their significance within the overall movement system. This toy example illustrates how OD models may fail to capture the full spatial complexity of commodity flows.

We propose a vector field-based method as an alternative to traditional network models for analysing commodity flows/mobility. By transforming OD data into vector fields, our approach captures likely movement directions even in unrepresented areas,

offering a more realistic view of spatial dynamics. This is particularly valuable because, in practice, it means that if future trade nodes are established in these previously unrepresented locations, we would already have a preliminary indication of the likely directions their commodities will flow towards, providing a foundation for further analysis.

The application of vector fields to commodity flows is novel, constituting a significant methodological advance in the study of commodity movement. However, the concept of vector fields itself is not new [14]. Vector fields have been utilised in disciplines such as meteorology, where they are used to represent wind direction and velocity across geographical regions [15]. These fields allow meteorologists to visualise and analyse weather patterns, predict storm paths, and model airflow. By translating commodity trade movements into vector fields (which we refer to as commodity flow maps in this paper), we can build on the extensive foundation of methods developed for weather mapping and other fields that deal with spatial flows [16]. Established techniques such as vector interpolation, field integration, and streamline generation—routinely used in weather analysis—can be directly adapted to the study of commodity flows, enhancing both the accuracy and efficiency of our approach. This cross-disciplinary application enables a more dynamic and flexible analysis of trade data, where unrepresented areas can still contribute to the overall understanding of mobility patterns.

To demonstrate the utility of our approach, we apply this vector field methodology to cattle trade data from the Brazilian state of Minas Gerais. Minas Gerais is one of the largest cattle-trading regions in Brazil, with a complex and extensive trade network. The scale of this trade activity, coupled with the incomplete nature of available data, makes it an ideal test case for evaluating the effectiveness of our methodology in converting OD data into vector fields that represent commodity flows.

One key direct application of this method is in understanding the spread of diseases, such as those affecting livestock like foot-and-mouth disease [17–19]. In traditional models, the lack of data on certain locations limits our ability to predict how a disease might spread through a location that is not originally represented as origin or destination. Vector fields, however, allow us to estimate potential commodity (livestock) flows through these unrepresented areas, offering a more comprehensive tool for assessing risk. This can be crucial for decision-makers who need to predict and manage the spread of diseases across regions with incomplete data.

To analyse the flow of commodities over time, we rely on two essential parameters: entropy [20] and cosine similarity [21]. In this paper, entropy quantifies the diversity of trade directions throughout the year, helping to distinguish between regions with stable and fluctuating trade behaviours. Meanwhile, cosine similarity captures the temporal evolution of commodity flow directions across different timescales—monthly, weekly, seasonal, and yearly—allowing us to identify patterns of consistency or variability in trade routes. These measures provide a robust framework for understanding the broader structure of commodity flows and their temporal shifts within our vector field representation.

Beyond these statistical analyses, the vector field approach offers a comprehensive representation of commodity flows at multiple levels of granularity, enabling diverse experimental applications. A key aspect of our analysis involves identifying critical points within the vector field [22], which serve as focal points for understanding the local and global structure of commodity movements. By studying their behaviour, we gain insights into trade dynamics that are not captured by traditional network models. The classification of these critical points provides a structured understanding of the field's topology, and by assuming smooth transitions between regions, we can construct a simplified yet informative model of the overall commodity flow system.

Our results show that vector fields offer an efficient and adaptable framework for analysing large-scale commodity mobility datasets. This is demonstrated through our

successful interpolation of flow patterns in unrepresented areas and the identification of critical points that reveal underlying structural patterns in the cattle trade network of Minas Gerais. By focusing on spatial flows rather than discrete network nodes and edges, vector fields provide a more flexible approach to capture movement patterns. This method is particularly relevant in cases where uncertainty and missing data are unavoidable, offering a valuable alternative to traditional network-based models.

Materials and methods

Data formatting

Our approach to generating vector fields uses origin–destination (OD) data as input, which is the most common publicly available mobility data format. This data can be defined at different time windows and spatial resolutions. By segmenting the data into appropriate temporal intervals and spatial cells, we can project movements onto a grid and subsequently generate vectors that represent aggregate commodity flows between these defined areas. This process is not limited to any single commodity but is applicable to any scenario where movement data can be represented as OD pairs, making it a versatile tool for studying the flow of commodities, goods, or even people.

This general method is particularly valuable when dealing with imperfect or incomplete information. In many practical situations, data gaps and uncertainties exist. That is, certain locations might lack explicit records, or the data may not cover every possible route of movement. The vector field approach addresses these challenges by inferring likely commodity flows and filling spatial gaps using interpolation techniques. As a result, even with partial or uncertain OD data, this methodology offers a comprehensive view of movement patterns, aiding in the analysis and decision-making processes across diverse applications.

The Appendix provides a detailed description of how our vector field approach is applied to the movement of cattle as a commodity. While our methodology is general and applicable to any OD-defined movement data, we later focus on cattle trade in the Results section to illustrate the power and versatility of the approach in a concrete scenario.

Creating vector fields

We transform an origin-destination network into a vector field for the region of interest based on the following steps. The first step is to choose suitable temporal and spatial granularities, specifying both the timeframe for trades and how the region is partitioned. We then project the network (i.e. trades within the chosen timeframe) onto this spatial scheme so that each node lies within a particular cell (the first map on the left in Fig 2). For each cell, all outgoing edges are treated as vectors (shown in pink in the middle map in Fig 2), which are combined into a single resultant vector (shown in black in the middle map in Fig 2) whether through averaging, summation, or other methods. Repeating this process for every cell with outgoing edges produces a partial vector field. As some cells do not generate outgoing edges, we employ interpolation to fill in the missing vectors and complete the full vector field (shown in shades of orange in the rightmost map in Fig 2).

To assess the global behaviour of vector field patterns, it is necessary to interpolate values for locations without originating edges, thereby completing the field. Interpolation constructs vectors at new points based on known (or measured) vectors, enabling a continuous representation of commodity flows across the entire region. We opt for a triangle-based interpolation owing to its simplicity and suitability for our sparse

Fig 2. Generating a vector field from origin-destination data. Given locations where we have information (origins/destinations of commodity movements represented by directed arrows) shown in the map on the left, **(A)** we can interpret each origin-destination pair as a vector and calculate the resulting vector of all movements from each particular location. Once this is done for all locations with available data, **(B)** we can interpolate the missing information, illustrated in the rightmost figure as example arrows shown in shades of orange. The final result is a complete vector field where all locations (cities, for instance) have a vector with direction and magnitude. In this example, the arrows on the rightmost map represent direction and the shade represents the magnitude of the vector.

networks [23, 24]. This method estimates vector values in grid cells lacking *a priori* data, ultimately yielding a complete vector field (right map in Fig 2). For technical details on the interpolation method, see the Appendix.

Vector field analysis offers two complementary approaches for understanding commodity mobility patterns: global flow exploration and critical point identification.

Global exploration of flow patterns focuses on how a region's mobility behaviour changes through its dominant flow vector, as we later demonstrate with cattle trade in Section Diversity and regularity of (cattle) commodity flows. This approach examines the vector computed for each region, following the method outlined in Fig 2. Changes in the dominant direction of a region's vector, as well as similarities between its vector dynamics and those of other regions, can be effectively observed through vector field representations. Temporal patterns can be quantified using vector directions by applying measures such as cosine similarity and entropy (see the Appendix for the methodological details), while spatial correlations based on vector magnitudes can highlight groups of cities with comparable trade behaviour, including clusters that reflect different trading distances—such as short-, medium-, and long-distance commodity flows. Examining these spatial patterns can identify areas with consistently high or low trade intensity. Comparing these characteristics across multiple time intervals allows clustering regions with similar dynamic behaviours, and identifying key roles in overall mobility patterns.

Critical points in vector fields, where the flow vanishes, reveal essential structural characteristics of mobility patterns. These points can be classified as attracting or repelling points based on the eigenvalues of the Jacobian matrix of the field [22]. We illustrate the presence of critical points within a vector field, highlighting regions of flow convergence (see Fig. A 2B in Appendix). These points are explored in our analysis of cattle trade (see Section Critical points in cattle vector field). However, their full interpretation in the context of commodity flows remains a question for future research.

In summary, transforming networks into vector fields offers a novel way to model the dynamics of commodity flows. In the context of cattle trade (see Section Results), this transformation helps reveal seasonal patterns, evolving critical points, and changes in flow dynamics over time.

Results

We apply our proposed vector field approach to cattle trade data to demonstrate how this method reveals complex mobility patterns. By transforming raw trade data into geographic vectors, we identify key insights such as dominant flow directions, trade hubs, and critical points. These insights are often missed by traditional network models.

Our analysis spans municipality and micro-region levels, using techniques like entropy, cosine similarity, and spatial autocorrelation (see Appendix for further details). We then showcase the method's ability to handle incomplete data, uncover commodity movement

patterns, and support applications such as planning, risk assessment, and disease control.

Robustness of vector fields

Before analysing the trade data, we assess the robustness of our vector field reconstruction approach by systematically removing spatial regions and evaluating the extent to which original flow directions are preserved. This procedure simulates the presence of missing trade data and demonstrates the capacity of vector fields to interpolate commodity flows in unobserved areas. As shown in Fig 3, we quantify robustness in terms of the proportion of vectors affected and the angular deviation in degrees. Three key findings emerge: *(i)* robustness remains consistent across years (fluctuations tend to be less than 3%), *(ii)* most deviations are below 15 degrees, and *(iii)* over 60% of the spatial information must be removed before more than half of the vectors experience any change. These findings indicate that our approach is highly resilient to spatial sparsity, with typical deviations constrained to minimal angular shifts.

Fig 3. Robustness of vector field estimations under spatial data removal. Proportion of regions exhibiting directional changes as a function of the number of randomly removed areas for 2014 (A, C) and 2015 (B, D). The results demonstrate the robustness of vector field estimations despite spatial data gaps, highlighting the reliability and efficiency of our approach across different years.

Cattle commodity flows and fields

We use the cattle trade data at two levels of granularity: municipalities and micro-regions. For the municipality level (Fig. A 1A in the Appendix), we begin by isolating a single municipality (shown in grey in Fig. A 1A in the Appendix) and considering all trades originating from it to other municipalities. We do not consider local trades that occur within the same municipality (self-loops). Each trade is transformed into a vector starting from the centre of a municipality and pointing towards the centre of the destination municipality. By averaging (or summing) these vectors, we obtain a single representative vector for the municipality, resulting in a partial vector field that captures the predominant flow direction and intensity for that location over a given time period. This process exemplifies how scattered trade data are aggregated into a coherent vector field at the municipal scale.

To move from this partial representation to a comprehensive view, we apply triangle-based interpolation as it preserves the spatial relationship (structure) and avoids artefacts by considering local patterns [23, 24] (further details in the Appendix). This interpolation fills in gaps where direct trade data are absent, creating a complete vector field over the entire municipal region that faithfully represents the continuous flow patterns of cattle movements.

Diversity and regularity of (cattle) commodity flows

We analyse the monthly vector fields from the trade data for two distinct spatial divisions: municipalities and micro-regions. Our goal is to assess how the directions of these commodity flows evolve over time and across different regions. To achieve this, we use two key metrics:

Entropy: measures the diversity of trade directions over the year. In particular, we use Shannon entropy, which quantifies how often trades occur across different directions (see the Appendix for the details of Shannon entropy).

Cosine similarity: captures the similarity (or dissimilarity) in commodity flow direction between pairs of consecutive months. A value near 1 indicates that two monthly vectors point in nearly the same direction (high regularity), while a value near -1 indicates they point in opposite directions (high variability).

During each year, every municipality (or micro-region) has 12 monthly vectors of trade flows. We categorize each vector into one of four quadrants based on its direction in the Cartesian plane, thus reducing the original set of 12 vectors to a sequence of 12 quadrant values. We then compute the Shannon entropy of this set, capturing the diversity or unpredictability of trade directions over the year.

Fig 4 presents entropy heatmaps of municipalities and micro-regions. Many municipalities (and micro-regions) in the North exhibit lower entropy (darker blue shades, closer to zero) indicating that trading directions tend to be more stable. Some regions with higher entropy (darker red shades) have more frequent shifts in commodity[1] flow direction, indicating greater unpredictability. As the highest inter-regional trade volume in Minas Gerais occurs between the North and West regions, the unpredictability is mostly observed in the South-east—regions with less established trading relationships.

Fig 4. Entropy of monthly vectors for (A) municipalities and (B) micro-regions. Each location's 12 monthly vectors are grouped into four directional clusters. Shannon entropy then yields a single value per municipality or micro-region, which is colour-coded on the map to reflect diversity in commodity flow directions over the year. The entropy is normalised for visualisation purposes using MinMax normalisation.

To assess how trade directions evolve over short time intervals, we use the cosine similarity between pairs of consecutive monthly vectors for each municipality or micro-region (e.g., from January 2013–February 2013 to November 2016–December 2016). This method allows us to focus on the structure and direction of commodity flows rather than their magnitude, making it well-suited for detecting shifts or redirections in trade patterns over time. We then use these cosine similarity values to cluster municipalities and micro-regions based on their month-to-month flow direction patterns (Fig 5). We identify four well-established clusters: from highly static (shades of blue) to highly dynamic (shades of orange). Different colours in the cluster maps indicate groups of regions that share similar directional dynamics over time. In essence, we group areas (cities or microregions) whose flow directions not only change month-to-month by similar angles, but also follow similar patterns of change throughout the year. This captures both the magnitude and rhythm of directional shifts. Fig 5 displays the average cosine similarity values of municipalities or micro-regions within each cluster. Detailed cosine similarity values for individual areas are provided in our previous study [25].

Fig 5. Cosine-based clustering of (A, B) municipalities and (C, D) micro-regions. Using K-medoids clustering (k=4), locations are grouped by their cosine similarity values. Clusters, indicated by distinct colours, reveal that in many peripheral locations the commodity flow directions remain stable over the years.

In both municipalities (Fig 5A-B) and micro-regions (Fig 5C-D), a large dominant cluster emerges, consisting of areas with relatively stable flow directions (cosine values close to 1), indicating regular and predictable cattle movement. However, there are cities and micro-regions where flow directions vary significantly—more than 45 degrees across several monthly intervals over the four-year period. Even in these more dynamic cases, the angular changes exhibit a periodic, sinusoidal-like pattern, with a noticeable drop in

---

[1]Note, the term 'commodity' refers to 'cattle as commodity' in the results section given that cattle trade has been used as the case study.

cosine similarity—indicating increased angular change—towards the end of each year (e.g., November–December), as shown in Fig 5B and 5D.

This observed rhythmicity reveals a deeper temporal structure in trade patterns, going beyond static flows or random fluctuations. Even if the peak of directional change shifts in time across regions (e.g., occurring earlier or later in the year), the underlying wave-like structure remains, highlighting the existence of region-specific temporal rhythms. These findings indicate a higher level of predictability in cattle movement dynamics and demonstrate that the combined use of entropy (to measure directional diversity) and cosine similarity (to assess regularity) can be effectively extended to analyse specific types of cattle trade transactions.

Going beyond direction in vector fields

In our previous findings, we focused primarily on the direction of vector fields. While this provided valuable insights, it overlooked a key dimension: magnitude. Two vectors may share the same direction yet represent very different dynamics—one indicating a short trade distance, the other a much longer one. By incorporating vector magnitudes into our analysis, we expand the study to capture both spatial and temporal variations in trade distances, offering a more comprehensive view of mobility patterns. Specifically, we investigate whether regions tend to have trade distances similar to those of their neighbouring areas. This spatial analysis sheds new light on the role of geographic proximity in shaping trade behaviour and commodity movement patterns.

Fig 6 illustrates the spatial variation in vector magnitudes across municipalities and the corresponding spatial lag values. These maps highlight municipalities with significantly higher or lower vector magnitudes than their neighbours, identifying notable patterns such as 'doughnuts' (low values surrounded by high values) and 'diamonds' (high values surrounded by low values). These features are particularly evident in the spatial lag map (Fig 6B), which provides a more distinct representation of clusters compared to the raw vector magnitude map (Fig 6A).

Fig 6. Maps of trading distances and spatial lag values. (A) Each municipality's trade distance is represented by summing its outgoing vectors, categorised into five distance clusters with distinct colours. **(B)** Spatial lag values, calculated with Queen-based spatial weights, highlight municipalities that differ markedly from their neighbours, revealing 'doughnut' patterns, where low values are encircled by higher ones, and 'diamond' patterns, where high values are encircled by lower ones. These maps are generated using the flow data from January 2013.

To quantify spatial relationships, we compute Moran's I, a global measure of spatial autocorrelation. The Moran scatter plot (Fig 7A) reveals a positive correlation between vector magnitudes and their spatial lags, indicating a tendency for similar values to cluster spatially. The observed Moran's I value (Fig 7B) is statistically significant, suggesting that vector magnitudes are not randomly distributed—also visually evident in Fig 6. This spatial clustering persists over time, as shown in the Moran's I heatmap (Fig 7C), which highlights consistent autocorrelation across monthly vector fields.

By analysing vector magnitudes alongside directions, we reveal deeper patterns in trade dynamics and regional differences. While demonstrated here with cattle trade, this vector field approach is broadly applicable to other commodity flows, offering valuable insights for logistics planning and policy development.

Fig 7. Global Moran's I calculation. (A) Scatter plot of vector magnitudes against spatial lag with a fitted line. **(B)** To evaluate the significance of the observed pattern, we compared the Moran's I value calculated from the vector field to those from 1000 simulations in which vector magnitudes were randomly shuffled among municipalities. The observed Moran's I value of 0.258 was significantly higher than the simulated values, indicating a non-random spatial structure. **(C)** The heatmap shows Moran's I values for monthly vector fields across different years. Since no blue shades appear in the heatmap, it indicates that all calculated Moran's I values are positive, suggesting positive spatial autocorrelation—meaning that municipalities with high vector magnitudes tend to be located near each other.

Critical points in cattle vector field

To uncover dynamic movement patterns, we use interpolation methods to estimate vectors between known points, transforming scattered data into continuous vector fields. These fields not only capture the flow of commodity trade but also reveal regions of attraction and repulsion, offering valuable insights into spatial dynamics—particularly in understanding the spread and containment of epidemics. By analysing our vector field data, we identify areas where trade concentrates (e.g., high-risk zones for epidemics) and where it diverges (e.g., key suppliers or hubs that can facilitate disease spread).

To quantify these patterns, we calculate the critical points in our seasonal vector fields over a four-year period, distinguishing attracting points (sinks, shown in Fig 8A) from repelling points (sources, shown in Fig 8B). The locations of these sinks and sources shift across seasons, suggesting that while some regions maintain directional stability over time (Fig 5), disease outbreak management requires careful consideration of the specific month, year, and trade dynamics as these critical points evolve throughout the period.

Fig 8. Sinks (A) and sources (B) in the cattle trade vector field. Critical points are shown in the vector field for four seasonal periods, aggregated over a four-year span. The initial vector fields are derived from cattle trades occurring in each season. A triangle-based interpolation method is used to generate a continuous vector field from these initial vectors. Critical points within the field are identified, distinguishing between sinks and sources. These heatmaps indicate the density of sink and source points, with darker shades representing higher concentrations—darker red for sinks and darker green for sources.

The movement of cattle follows a structured network, where seasonal shifts and geographical factors play a role in shaping the distribution of sources and sinks. This variability is shaped by a combination of factors, including the geography of the region, the type of premises (e.g., farms, slaughterhouses, or rearing centres), the purpose of the trades, and the broader seasonal trends in cattle trade. For instance, source areas become more prominent during breeding seasons such as spring and summer [26], while a high density of sink areas corresponds to periods of higher cattle slaughter, such as the months leading up to the New Year. In addition, we observe that regions with a high density of slaughterhouses (reported for 2016 [27]) overlap with high-density sink areas in our vector field data during spring.

These patterns highlight not only the mobility within the trade system but also the changing role of different regions in the commodity flow network. Over time, the distribution and strength of sinks and sources may be tied to evolving logistical strategies, shifting market demands, and the structural role of specific regions. The central and western regions frequently act as sources, likely due to their role as intermediate hubs

in long-distance transportation. The seasonal vector fields represent cattle movements rather than fixed farm locations. Given the considerable distance between the northern and southern parts of the state, cattle are typically routed through these central source regions before reaching their final destinations. A comparison of the maps in Fig 8 and Fig 6B reveals that sink areas often coincide with short-distance transport zones, while medium- to long-distance transport regions are typically associated with source areas.

Discussion

This study introduces a vector-field-based framework for analysing commodity mobility, which bridges gaps in traditional origin-destination (OD) approaches by incorporating unrepresented locations into a comprehensive view of mobility patterns. Using cattle trade in the state of Minas Gerais, Brazil, as a case study, we demonstrate that vector fields are an effective tool for revealing dynamic trade patterns, including: *(i)* clusters of stable trade direction over time, *(ii)* spatial autocorrelation in commodity flow magnitudes, and *(iii)* key attractors and repellers within the region.

Vector fields provide a comprehensive framework for representing flows, capturing both intensity and direction simultaneously, enabling interpolation in regions with missing data, and supporting advanced analyses such as critical-point identification, temporal patterns, and spatial autocorrelation. While alternative methods (such as origin-destination matrices or trajectory-based models) can represent flow direction, they often cannot reconstruct flows in unobserved regions or require detailed individual-level data. Our approach combines these strengths in a continuous, scalable, and theoretically grounded framework. Limitations include sensitivity to the density and quality of input data, potential artefacts in sparsely observed regions, and the focus on a single commodity in this study. Although based on widely used origin–destination data, the method can be readily adapted to other sources, such as CDR or XDR records. Future work could explore higher-resolution data, multi-commodity flows, and applications to human mobility, further extending the versatility and impact of the vector-field framework.

It should be noted that the application of vector fields to human mobility is not entirely novel. Previous works, such as Mazzoli [28] and Liu [29], introduced and generalised vector-field frameworks for characterising recurrent commuting flows and broader mobility patterns from individual trajectories. Our approach builds on this foundation but focuses on commodity mobility using OD matrices, incorporating robust interpolation to infer flows in unrepresented locations, and leveraging entropy and critical-point analysis (sources and sinks) to capture the diversity, stability, and structural patterns of trade flows.

A key contribution of this work is the development of robust interpolation techniques that address incomplete and sparse datasets, enabling continuous mobility analysis across spatial regions. Even when more than 50% of the data is removed, the deviation in vector directions remains below 15 degrees, demonstrating the method's resilience and robustness to sparsity. Additionally, the introduction of entropy and cosine similarity analyses derived from vector fields offers novel ways to measure trade diversity, predictability, and month-to-month flow dynamics. Entropy quantifies the variation in trade directions, identifying regions with stable or unpredictable patterns, while cosine similarity captures temporal changes in flow dynamics, enabling the clustering of regions with similar trade behaviours. Together, these analytical tools enhance the understanding of spatial and temporal dimensions of commodity flows.

A key contribution of this work is the development of robust interpolation techniques that address incomplete and sparse datasets, enabling continuous mobility analysis across spatial regions. Our robustness analysis demonstrates that the method maintains

reliable performance even under substantial data loss. When more than 50% of the data is removed, the deviation in vector directions remains below 15 degrees and sink displacement remains below 85 km, indicating resilience to data sparsity. These findings suggest that the framework can be applied confidently in scenarios with incomplete data coverage, such as regions with limited surveillance infrastructure or privacy-restricted datasets. Sources and magnitude show greater sensitivity to area removal, as they are directly affected by data reduction. This highlights that the primary utility of vector flow analysis lies in identifying directional patterns and the structural organisation of flows, rather than providing precise quantification of trade volumes. For applications requiring accurate magnitude estimation, denser data coverage is recommended.

Another important contribution of this work is its compatibility with established vector field theories, which open avenues for advanced analyses. Techniques such as critical point dynamics analysis could be explored to understand the temporal evolution of sinks and sources, or streamline visualisation could be applied to trace continuous trade routes. These analytical approaches enable a more granular understanding of mobility patterns and facilitate predictions about unobserved areas based on inferred commodity flows.

While our vector field method offers significant advantages, certain limitations remain. The accuracy of interpolated vectors is influenced by the quality and density of the input data, with sparser regions showing higher uncertainty. In addition, this study focuses on a single commodity, leaving open the question of how well the method generalises to other trade systems or integrated datasets. Although our modelling is based on widely used origin–destination flow data, the approach is easily adaptable to other sources such as CDR and XDR records. These limitations point to promising directions for future research, including the use of higher-resolution data, the integration of multi-commodity flows, and applications to human mobility [1]. Addressing these challenges will further strengthen the versatility and impact of the vector field framework.

By transitioning from static network representations to dynamic vector fields, this research offers a versatile and scalable framework for studying commodity mobility. The approach not only advances our understanding of spatial flows but also establishes a foundation for future studies to refine and expand its applications in diverse fields, from public health to economic geography.

Critical points, such as sinks, sources, and saddle points, can provide deeper insights into the structural dynamics of commodity flows by identifying regions where trade activity converges, diverges, or transitions. Incorporating the identification of these points into the vector field framework enhances the understanding of how mobility patterns evolve over time and space, offering more detailed information about trade hubs, bottlenecks, and areas of influence. This analysis is particularly relevant for understanding how trade dynamics may impact broader systems, such as disease spread, resource allocation, or regional economic stability. The study of critical points also offers opportunities to integrate advanced vector field theories and analytical approaches for dynamic systems analysis. For instance, tracking the temporal evolution of these points could provide valuable insights into the stability and resilience of trade networks under changing conditions, such as seasonal variations or economic disruptions.

Our findings highlight a strong temporal variability in the spatial distribution of sink and source regions within the cattle trade system. Not only do the locations of these critical points shift across seasons, but their intensity also varies, as seen in the fluctuating density patterns in Fig 8. This temporal variation likely reflects the evolving roles of different premises and regions across seasons, influenced by market-driven dynamics and biological cycles such as breeding [26]. For instance, some premises may act as sources during active breeding or fattening periods, while others, such as slaughterhouses, consistently act as sinks due to receiving cattle without sending them onward. Notably,

these critical points may sometimes emerge in regions not directly present in the observed origin–destination data, suggesting that depending on the commodity type or trade characteristics, influential areas may arise that are not explicitly recorded in the original dataset.

This seasonal variability could also reflect larger economic or logistical factors—such as transport infrastructure or policy changes—that influence the movement of cattle over time. Future research could explore how different modes of transportation or changes in trade regulations might influence the emergence, persistence, or reconfiguration of these sinks and sources throughout the year. Future work also could explore how the characteristics of critical points (*e.g.*, their location, and movement) relate to the underlying drivers of trade, enabling targeted interventions or policy decisions.

Declaration

Acknowledgment

Author Contribution Statement

SF and RM developed the original ideas; All authors designed the study; RM supervised the development of the experiments; ASM wrote the formalisms; SF and MM collected, curated, and integrated the raw data; SF and MM performed the analysis; All authors analysed the results; SF and RM wrote the manuscript; SF and MM prepared the graphics; MM and ASM reviewed and revised the final version of the manuscript. All authors read, reviewed, and approved the final manuscript.

Appendix

Dataset and code availability This study examines cattle movement in Minas Gerais, Brazil, using data from the Institute of Agriculture (IMA) covering 2013–2016. The dataset includes around 420,000 locations (e.g., farms, slaughterhouses, and markets) and records 3.78 million transactions for purposes like fattening, breeding, and trade. Each movement details the origin, destination, species, gender, age, and number of animals. Import and export data (less than 5% of transactions) were excluded to focus on internal movement patterns. For the analyses presented in this study, origin-destination (OD) matrices (both monthly and seasonal) were used. These, along with all the code for analysis and visualisation, are available at `https://github.com/Sima-Far/Flow_vector_field`.

Interpolation of vector fields. There are multiple interpolation methods available for developing a continuous vector field. Among these, we opt for a triangle-based interpolation due to its computational efficiency and its suitability for handling the sparsity of our dataset's networks. Triangle-based interpolation enables us to estimate

Fig. A 1. Generating vector fields from cattle trade data for (A) municipalities and (B) micro-regions. (A-1) and **(B-1)** illustrate trade flows originating from a grey part (municipality or micro-region) to others. **(A-2)** and **(B-2)** represent the transformation of these trades into vectors, drawn from the centre of the grey area to the centres of destination areas, and then aggregated into a single resultant vector (shown in red). **(A-3)** and **(B-3)** demonstrate complete vector fields, where interpolation has been used to estimate vectors for municipalities or micro-regions lacking trade data during the selected time window. Different colours indicate varying vector magnitudes. This approach is applied for a specific time window and geographic division but can be adapted to different temporal or spatial granularities depending on analytical needs.

vector values at grid points where no initial vectors are present, while Delaunay triangulation efficiently manages sparse and irregularly spaced data by leveraging the geometric relationships between existing points [23], ensuring a smooth and coherent representation of commodity flows.

To perform the interpolation, we use the centres of cells with initial vectors, along with selected boundary points assigned zero vectors to mitigate boundary effects. First, we triangulate the region using Delaunay triangulation on these points, ensuring that each point of interest lies within a well-defined triangular structure. Assume point p falls inside a triangle with vertices having initial vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, as illustrated in Fig. A 2A. If h_1, h_2, and h_3 represent the distances from point p to the sides opposite to the angles at vertices with vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, respectively, the interpolated vector at point p, denoted as $\mathbf{v}_p$ is defined as [23, 24]

$$\mathbf{v}_p = \frac{h_1}{h_1 + h_2 + h_3}\mathbf{v}_1 + \frac{h_2}{h_1 + h_2 + h_3}\mathbf{v}_2 + \frac{h_3}{h_1 + h_2 + h_3}\mathbf{v}_3 \, . \tag{1}$$

Fig. A 2. Critical points in the vector field. (A) Triangle-based interpolation method. To estimate vectors at specific locations within the triangulated mesh (generated using Delaunay triangulation on points with known vectors), we use a triangle-based interpolation technique. This method calculates the vector at a point p by using the vectors at the three surrounding vertices. **(B)** Interpolated vector field. The resulting field displays interpolated vectors, with grey regions indicating critical points, which represent areas of attraction or repulsion. Colour variations reflect differences in vector magnitudes.

This interpolation method ensures that the resulting vector field is both continuous and adaptive to the underlying network structure. The final vector field representation, illustrated in Fig. A 2B, captures both the main flows for each cell and the transitions across the entire field, offering a more detailed spatial understanding of commodity movements.

Cosine similarity. To quantify temporal patterns in flow directions, we use the cosine similarity measure which quantifies the directional consistency of commodity flows over time. For region k between time intervals i and j, the similarity is given by [21]:

$$S_C(\mathbf{v}_k^i, \mathbf{v}_k^j) = \frac{\mathbf{v}_k^i \cdot \mathbf{v}_k^j}{\|\mathbf{v}_k^i\|\|\mathbf{v}_k^j\|} \, , \tag{2}$$

which represents the dynamic auto-correlation of the flow direction for region k. A value close to 1 indicates high directional stability over time, while lower values suggest greater variability.

The procedure consists of the following steps:

1. Generating multiple vector fields over predetermined time intervals (*e.g.*, months over four years).

2. Computing cosine similarity values between consecutive intervals for each region.

3. Constructing feature vectors for each region based on these similarity values.

4. Applying clustering methods to group regions with similar temporal mobility patterns.

This approach allows us to categorise regions based on the stability and evolution of their commodity flow directions, offering insights into the persistence and variability of trade movement patterns.

Shannon entropy. Shannon entropy is a measure of uncertainty or randomness associated with a set of possible outcomes. In a physical sense, it quantifies how dispersed or unpredictable a system is, akin to the concept of disorder in thermodynamics. It can be applied to assess the variability within a vector of values, capturing how evenly distributed different outcomes are [30].

Given a vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ where each x_i represents a category, event, or observed outcome, we first convert these values into a probability distribution. Let p_i denote the probability (or relative frequency) of outcome i occurring in the vector. The Shannon entropy $H(\mathbf{x})$ is then defined as [20]:

$$H(\mathbf{x}) = -\frac{1}{\log n} \sum_{i=1}^{n} p_i \log p_i \,, \tag{3}$$

where the logarithm is typically taken in base 2 (yielding entropy in bits). This equation sums over all possible outcomes, weighting each by the negative logarithm of its probability. Higher entropy indicates greater diversity or unpredictability in the outcomes represented by the vector, meaning that commodity flow directions are more evenly distributed across different routes. Conversely, lower entropy suggests a more regular or predictable pattern, where flows concentrate in fewer, more predictable pathways. This process yields a single scalar value $H(\mathbf{x})$ that summarises the diversity of the vector's components.

Critical points. Critical points in a vector field are essential for understanding the field's structure and dynamics. These are points where the magnitude of the vector field becomes zero, i.e., $\mathbf{v}(\mathbf{x}) = 0$, meaning all components of the vector vanish. In two-dimensional fields, these points are locations where field lines asymptotically converge or diverge, making the vector field's direction indeterminate.

A simple critical point is a special case where the vector field magnitude is zero at the point, but it does not vanish in its immediate surroundings. The behaviour near critical points can be studied using local linear approximations, assuming the vector field is smooth and differentiable. By employing a Taylor series expansion around a critical point $\mathbf{x}_0$, the local behaviour can be expressed as [22]:

$$\mathbf{v}(\mathbf{x}) = \mathbf{v}(\mathbf{x}_0) + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}(\mathbf{x} - \mathbf{x}_0)\,. \tag{4}$$

Because $\mathbf{x}_0$ is a critical point, $\mathbf{v}(\mathbf{x}_0) = 0$. Equation 4 can then be rewritten in matrix form as follows:

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} \frac{\partial v_x}{\partial x}(x_0, y_0) & \frac{\partial v_x}{\partial y}(x_0, y_0) \\ \frac{\partial v_y}{\partial x}(x_0, y_0) & \frac{\partial v_y}{\partial y}(x_0, y_0) \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}\,, \tag{5}$$

$$\mathbf{v} = \mathbf{J} \cdot (\mathbf{x} - \mathbf{x}_0) \,. \tag{6}$$

The Jacobian matrix, $\mathbf{J}$, is a mathematical representation that captures how a vector field changes near a critical point. It plays a key role in classifying critical points by analysing their local behaviour. By studying the eigenvalues and eigenvectors of $\mathbf{J}$, we can understand how tangent curves behave around the critical point.

Moran's I. The Moran's I statistic quantifies the degree of spatial autocorrelation by analysing both the values and locations of observations. It assesses whether a variable exhibits spatial clustering, dispersion, or randomness, summarising spatial patterns within a dataset. The Global Moran's I index is calculated as follows [31]:

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \times \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2} \,, \tag{7}$$

where:

- n is the number of observations,

- z_i is the standardised value of the observation at location i (commonly $z_i = y_i - \bar{y}$),

- y_i and y_j are values at locations i and j, respectively,

- $\bar{y}$ is the mean of all values,

- w_{ij} represents the spatial weights matrix, defining the spatial relationship between pairs of locations.

Moran's I values range from -1 to +1:

- **Positive values** indicate spatial clustering, where similar values are near each other.

- **Negative values** suggest spatial dispersion, where high values are situated near low values, and vice versa.

- **Values close to zero** imply randomness, with no discernible spatial pattern in the data.

As a graphical tool, the Moran Plot provides a valuable resource for understanding both global and local spatial autocorrelation. These plots take the form of scatter plots, where values of interest are plotted against spatial lag (as shown in Fig. A 7A) for each location [32].

Spatial lag variable represents weighted sums or averages of neighbouring values. The spatial lag characterises how a variable behaves near each location, essentially acting as a local smoother. This can be expressed in matrix notation as follows [31, 33, 34]:

$$y_{sl-i} = \sum_j w_{ij} y_j \,. \tag{8}$$

In Equation 8, y_{sl-i} represents the spatial lag for location i, calculated as the weighted sum of values at all other locations. As mentioned before, y_j is the value in location j, and w_{ij} is the entry in the i-th row and j-th column of a W spatial weights matrix. Since non-neighbours receive a weight of zero, y_{sl-i} effectively captures the weighted average of values at observation i's neighbours. Spatial lag of the vector magnitudes is calculated using spatial weights based on Queen Weights [35]. This weight reflects adjacency relationships using a binary variable that indicates whether a polygon shares

an edge or a vertex with another polygon. The matrix expression for this weight is as follows:

$$W_{ij} = \begin{cases} 1, & \text{if regions } i \text{ and } j \text{ share an edge or vertex,} \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

References

1. Barbosa H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, et al. Human mobility: Models and applications. Physics Reports. 2018;734:1–74.

2. Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, et al. Mobility network models of COVID-19 explain inequities and inform reopening. Nature. 2021;589(7840):82–87.

3. Verginer L, Riccaboni M. Cities and countries in the global scientist mobility network. Applied Network Science. 2020;5:1–16.

4. Wang J, Kong X, Xia F, Sun L. Urban Human Mobility: Data-Driven Modeling and Prediction. SIGKDD Explor Newsl. 2019;21(1):1–19. doi:10.1145/3331651.3331653.

5. Soriano-Paños D, Cota W, Ferreira SC, Ghoshal G, Arenas A, Gómez-Gardeñes J. Modeling communicable diseases, human mobility, and epidemics: A review. Annalen der Physik. 2022;534(6):2100482.

6. Bajardi P, Barrat A, Savini L, Colizza V. Optimizing surveillance for livestock disease spreading through animal movements. Journal of the Royal Society Interface. 2012;9(76):2814–2825.

7. Lentz HH, Koher A, Hövel P, Gethmann J, Sauter-Louis C, Selhorst T, et al. Disease spread through animal movements: a static and temporal network analysis of pig trade in Germany. PloS one. 2016;11(5):e0155196.

8. Farokhnejad S, Cardoso D, Rocha C, da Mata AS, Menezes R. A Data-driven Approach to Cattle Epidemic Modelling under Uncertainty. In: CompleNet. Springer Nature; 2022.

9. Zheng Y. Trajectory data mining: An overview. ACM Transactions on Intelligent Systems and Technology (TIST). 2015;6(3):1–41.

10. Gonzalez MC, Hidalgo CA, Barabasi AL. Understanding individual human mobility patterns. nature. 2008;453(7196):779–782.

11. Wang D, Pedreschi D, Song C, Giannotti F, Barabasi AL. Human mobility, social ties, and link prediction. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining; 2011. p. 1100–1108.

12. Blondel VD, Decuyper A, Krings G. A survey of results on mobile phone datasets analysis. EPJ data science. 2015;4:1–55.

13. Mamei M, Bicocchi N, Lippi M, Mariani S, Zambonelli F. Evaluating origin–destination matrices obtained from CDR data. Sensors. 2019;19(20):4470.

14. Liu Y, Weisberg RH. Patterns of ocean current variability on the West Florida Shelf using the self-organizing map. Journal of Geophysical Research: Oceans. 2005;110(C6).

15. Wittenbrink CM, Pang AT, Lodha SK. Glyphs for visualizing uncertainty in vector fields. IEEE transactions on Visualization and Computer Graphics. 1996;2(3):266–279.

16. Andrienko N, Andrienko G. Visual analytics of movement: An overview of methods, tools and procedures. Information visualization. 2013;12(1):3–24.

17. Tomley FM, Shirley MW. Livestock infectious diseases and zoonoses; 2009.

18. Volkova V, Howey R, Savill N, Woolhouse M. Potential for transmission of infections in networks of cattle farms. Epidemics. 2010;2(3):116–122.

19. Ferguson NM, Donnelly CA, Anderson RM. The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. Science. 2001;292(5519):1155–1160.

20. Shannon CE. A mathematical theory of communication. The Bell system technical journal. 1948;27(3):379–423.

21. Beggs JM, Timme N. Being critical of criticality in the brain. Frontiers in physiology. 2012;3:163.

22. Smolik M, Skala V. Vector field interpolation with radial basis functions. In: Proceedings of SIGRAD 2016, May 23rd and 24th, Visby, Sweden. 127. Linköping University Electronic Press; 2016. p. 15–21.

23. Watson D, Philip G. Triangle based interpolation. Journal of the International Association for Mathematical Geology. 1984;16(8):779–795.

24. Watson DF, Philip G. A refinement of inverse distance weighted interpolation. Geo-processing. 1985;2(4):315–327.

25. Farokhnejad S, Oliveira M, Ribeiro E, Menezes R. Using Vector Fields in the Modelling of Movements as Flows: A Case Study with Cattle Trade Networks. In: International Workshop on Complex Networks. Springer; 2023. p. 155–167.

26. Paulino P, Duarte M, Kahn L, Cottle D. Brazilian beef production. Beef cattle production and trade. 2014;107.

27. Vale R, Vale P, Gibbs H, Pedrón D, Engelmann J, Pereira R, et al. Regional expansion of the beef industry in Brazil: from the coast to the Amazon, 1966–2017. Regional Studies, Regional Science. 2022;9(1):641–664.

28. Mazzoli M, Molas A, Bassolas A, Lenormand M, Colet P, Ramasco JJ. Field theory for recurrent mobility. Nature Communications. 2019;10(1):3895. doi:10.1038/s41467-019-11841-2.

29. Liu E, Mazzoli M, Yan XY, Ramasco JJ. A generalized vector-field framework for mobility. Communications Physics. 2024;7(1):190. doi:10.1038/s42005-024-01672-z.

30. Tsallis C. Introduction to nonextensive statistical mechanics: approaching a complex world. vol. 1. Springer; 2009.

31. Rey S, Arribas-Bel D, Wolf LJ. Geographic data science with python. CRC Press; 2023.

32. Kawano Y. workshop-python-spatial-stats; 2021. `https://github.com/yohman/workshop-python-spatial-stats`. Accessed 2025-05-10.

33. Anselin L. Local indicators of spatial association—LISA. Geographical analysis. 1995;27(2):93–115.

34. Darmofal D. Spatial analysis for the social sciences. Cambridge University Press; 2015.

35. Berry BJL, Marble DF. Spatial analysis: a reader in statistical geography. Englewood Cliffs, NJ, Prentice-Hall. 1968;.