Revista Brasileira de Estudos Regionais e Urbanos
v. 19, n. 2, p. 266-288, 2025
DOI: 10.54766/rberu.v19i2.1156

Brazilian exports imputation: A new algorithm for estimating the municipal production directed at the foreign market

Alan Leal[1] | Michelle Marcia Viana Martins[2]

[1] Ph.D. candidate in Economics at University of São Paulo. E-mail: prof@alanleal-econ.com
[2] Professor of Economics at Federal University of Viçosa. E-mail: michelle.viana@ufv.br

ABSTRACT
This paper aims to present a new algorithm to estimate Brazilian municipalities exports at the HS6 level. Currently, the information available to academics and policymakers are available at a HS4 level and more importantly it does not inform the municipalities producing and directing this producing to foreign markets, it only addresses the bureaucracy of the export activity. Beyond this, we also review, through LLama, the literature that mentions the words "municipal exports", finding that no particular work mentions this distinction. The method proposed in this paper can help many stakeholders to analyze and make more informed decisions regarding themes of international trade. Hence, it possesses academic and policy relevance within the type of data that is available in Brazil.

KEYWORDS
Exports data, Brazil, Spatial Methods

Imputação das exportações brasileiras: Um novo algoritmo para estimar a produção municipal destinada ao mercado externo.

RESUMO
Este paper apresenta um novo algoritmo para estimar as exportações municipais brasileiras no nível SH6. Atualmente, a informação disponível a acadêmicos e policymakers se encontra a nível SH4 e lida com os municípios exportando, não com os municípios produzindo para a exportação, ou seja, é uma informação de natureza mais burocrática que econômica. Além disso, o paper também emprega o Llama para avaliar a literatura que menciona as palavras "exportações municipais", encontrando nenhuma menção particular em relação a essa diferenciação. O método proposto neste paper pode ajudar agentes interessados a analisar e tomar melhores decisões em temas relacionados ao comércio internacional. Logo, este estudo possui relevância acadêmica e política, considerando o perfil de dados que o Brasil possui.

PALAVRAS-CHAVE
Dados de Exportação, Brasil, Métodos Espaciais

CLASSIFICAÇÃO JEL
C81, R12, F14

Recebido: 26/01/2025. Aceito: 22/04/2025
e-ISSN: 2447-7990 | https://revistaaber.org.br/rberu

1. Introduction

This article has a methodological nature concerning a highly relevant topic related to international trade data in Brazil. Brazilian international trade data primarily focuses on the origin of exports (or imports), meaning that the data provided by the Ministry of Development, Industry, and Trade (MDIC) does not focus on where export-oriented production occurs, but where the export operation occurred.

This presents some important challenges for articles analyzing trade by Brazilian municipalities. It is also difficult for policymakers, given their inability to assess how much municipalities are exporting. In the first case, articles analyzing exports at the HS4 level will tend to overestimate the effects of some policies or the opposite, given the fact that they are not dealing with export-oriented production, they are dealing with exports as bureaucratically registered in MDIC's export systems. In the latter case, policymakers are blind to the true effect of any policy on a municipality's exports. Another peculiarity of Brazilian data is the fact that municipal export data is only available at the HS4 level[1], while state-level exports are available at the HS6 level.

This article builds on previous work that attempts to estimate municipal exports, developing a generic method that can leverage research and policy knowledge to properly enable considerations about municipal exports. This analysis is used for soybean production[2] in Mato Grosso[3], as a way to test this method and also as a way to demonstrate its use.

As a literature exercise, we searched for Portuguese-language works in Brazilian Google Scholar using the query "municipal exports". We found 77 articles, of which 51 documents[4] were analyzed. Then, we employ a Large Language Model (LLM), Llama 3.2, a version of Meta LLM. The query used for Llama explicitly required that the analyzed article make a distinction between municipal export data originating from a municipality and data about production in a municipality aimed at the foreign market by a municipality[5].

[1]The Harmonized System is a comprehensive nomenclature for internationally traded goods. It works at different levels of detail, ranging from dozens of classes of goods with the Harmonized System 2 - two digits - to products with the Harmonized System 8 - eight digits.

[2]HS4:1201 for soybeans and HS6:120190 for soybeans, except for sowing.

[3]In 2023, about 62% of Mato Grosso municipalities did not officially export any soybeans, according to Comex Stat records.

[4]The articles analyzed in the study are: Sant'Anna (2015), da Silva et al. (2024), Pena et al. (2022), Sant'Anna and Júnior (2016), Leal (2024), Seibert et al. (2024), Barbosa (2024), Lacerda et al. (2013), de Oliveira et al. (2017), Rajão et al. (2020), Castriota (2024), Scazufca (2008), Saboia (2013), Castilho and Arrais (2017), Perpetua and Junior (2013), Luiz and de Oliveira (2020), Kuhn et al. (2017), Moreira et al. (2013), da Silva Araújo and da Silva (2018), dos Santos (2017b), Biavaschi (2006), Ponte (2022), Mendes et al. (2020), dos Santos (2017a), Cabral (2022), Hidalgo (2017), Souza and Raiher (2022), Metzdorff et al. (2015), Silveira and Braga (2023), Seibert and Perobelli (2022), Ferreira et al. (2019), Vivaldi (2021), Dorneles (2024), de Assis (2024), Perpetua et al. (2012), Da Silva (2013), Mendes et al. (2021), Vivaldi et al. (2020), De Almeida (2018), Mancini and Carneiro (2018), Kühn (2008), Corrêa et al. (2015), Souza and Raiher (2022), Cardoso (2018), Sachser (2015), Tatsch and Batisti (2013), Araujo (2017), Biavaschi (2005).

[5]See Appendix for more details.

We find that only two articles make this important data distinction, even though they do not propose another way to circumvent this difficulty present in Brazilian municipal export data. These articles are Leal (2024) and Biavaschi (2006). The latter article does not directly address municipal exports in terms of their productive origins. The first article, although dealing with municipal exports, these are not exports based on the municipality of production. Therefore, regarding the literature, this LLM experiment of the literature indicates that many articles either use Comex Stat municipal export data as synonymous with a municipality being both an exporter and producer of a certain good, or are not aware of this important data limitation.

In addition to this Introduction, this article is divided into a Method section, a Data description and analysis section, focusing on what is currently available and what this method promises; then it moves on to the application of this method to soybean production in Mato Grosso; a section that implements a Monte Carlo experiment and tests the algorithm with simulated export data for Brazilian states; a section detailing the relevance of this algorithm for academics and policymakers is the penultimate section of this work; a Final Considerations section emphasizes the advantages and limitations of this method.

2. Method

We propose a method that leverages two insights: (i) there is more knowledge of the spatial distribution of some economic activity than others Faria and Haddad (2014). For instance, we know municipal employment in any activity and the soybeans production by municipality, whereas data on municipal exports is lacking; (ii) Brazilian exports data has spatial anchors. We know its spatial distribution, imperfectly in terms of depth - product vs type of products - and spatial coverage - some municipalities officially exports, while others municipalities do not. Hence, this method uses these two insights to estimate the municipal exports of any Brazilian municipality and any goods.

Considering these two insights, let k be defined as the number of goods at the HS6 level present in a product category at the HS4 level, then the production directed to the foreign market in a given municipality at the HS6 level can be approximated by the following equation:

$$E_{i,\text{HS6}} = \underbrace{\alpha \mathbf{W} E_{i,\text{HS4}}}_{\text{Spatial Component}} + \underbrace{r E^{*} i \text{`HS6}}_{\text{Adjustment Component}} \tag{1}$$

where:

- i is a specific municipality.
- $E_{i,HS6}$ are the estimated municipal exports at the HS6 level.
- $E^{*}_{i*,HS6}$ is defined as $\sum_{i=1}^{I} \alpha(E_{i*,HS4} - \mathbf{W}E_{i*,HS4})$.

- α is the proportion coefficient at the state level defined as:

$$\alpha = \frac{E_{state,HS6}}{E_{state,HS4}} \tag{2}$$

  where E means Exports, therefore $E_{state,HS6}$ means state exports of a given good at the HS6 level, while $E_{state,HS4}$ means state exports of the good category at the HS4 level.

- **W** is a row-normalized $n \times n$ spatial weights matrix where n is the number of municipalities, such that:

$$\sum_{j=1}^{n} w_{ij} = 1 \quad \forall i \tag{3}$$

  And $w_{ij} \geq 0 \quad \forall i, j$

- **r** is an $n \times 1$ vector that sums to 1. In this work, it is considered to be the relevance of each municipality in the state's formal workforce in the soybean sector.

- $E_{i*,HS4}$ is an $n \times 1$ vector of observed municipal exports at the HS4 level, where * indicates that some entries may be zero

2.1 Properties

The method satisfies several key properties:

1. **Dimensional Consistency:**

$$\mathbf{W}(n \times n)E_{i*,HS4}(n \times 1) - r(n \times 1) \tag{4}$$

2. **Separability and Aggregation Properties:** The spatial and adjustment components sum (across all possible k) to a value identical to the correct HS4 exports for the entire state. This happens because the Adjustment component is calculated based on the difference between product export values and their estimates. This allows us to calculate a more direct value for exports of a given good at the HS6 level. As an example, Mato Grosso produces crushed soybeans for sowing and not for sowing. The algorithm proposed in this paper imputes municipal exports at this level (soybeans not for sowing), such that the sum of these imputed municipal exports for the two goods equals Mato Grosso's crushed soybean exports. This configures the aggregation property.

3. **Zero Treatment:** The method naturally accommodates municipalities with zero exports through the matrix multiplication structure.

2.2 Specification of W

The only requirement postulated about W is that it be row-normalized and have only non-negative elements. This allows for a myriad of matrix options and different relationships to be postulated in the algorithm. Therefore, one can use spatial matrix theories that postulate spatial dependence between different spatial units. This is an interesting interpretation and matrix definition. We employ this method for soybean production and compare the results as a way to assert which approach can produce the most interesting results.

It is relevant to emphasize that different methods are well-suited for different goods. Goods whose production appears spatially concentrated and correlated will have spatial weights matrix producing the best results. As a way to see how the spatial solution behaves compared to the HS4 solution, we employ spatial descriptive analysis to emphasize our findings. Therefore, one can use contiguity matrix (more precisely here Queen) and inverse distance matrix. We employ these two methods to deepen our analysis. But before detailing their application, we explain these forms of measuring spatial relationships. A more in-depth discussion about spatial weights matrix can be found in Kelejian and Piras (2017); Arbia et al. (2014).

- Queen:

Figure 1. Queen-type Contiguity Spatial Weights Matrix: On the left, we see the location units positioned in numbered cells, while on the right we see an $n \times n$ matrix, detailing whether the unit in row i is contiguous to the unit in column j.

Spatial Layout

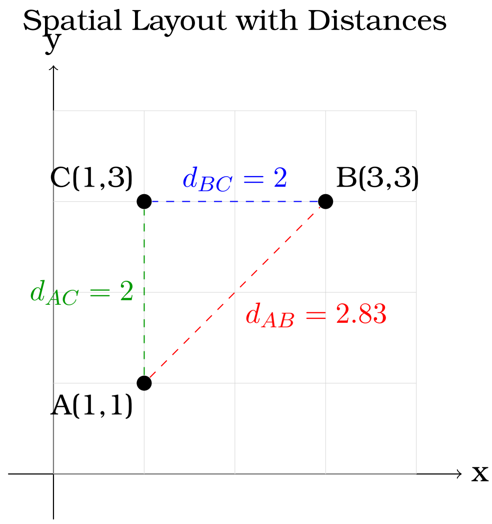
1	2	3
4	5	6
7	8	9

Queen Contiguity Matrix

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

- Inverse Distance:

Figure 2. Inverse Distance Spatial Matrix. The figure at the top considers the distance between three locations, points A, B, and C. In the bottom left, we have the distance between points A, B, C among themselves. The main diagonal is 0, as it measures the distance from a point to itself. The inverse distance matrix, on the bottom right, is the element-by-element inverse of the distance matrix D. The main diagonal is also 0, which is consistent with other spatial weight matrices, including the Queen spatial matrix.



Distance Matrix D

$$\begin{bmatrix} 0 & 2.83 & 2.00 \\ 2.83 & 0 & 2.00 \\ 2.00 & 2.00 & 0 \end{bmatrix}$$

Inverse Distance Matrix $W = 1/D$

$$\begin{bmatrix} 0 & 0.35 & 0.50 \\ 0.35 & 0 & 0.50 \\ 0.50 & 0.50 & 0 \end{bmatrix}$$

Note: This type of matrix frequently operates with locations as points, not polygons, in contrast to the Queen spatial weights matrix

2.3 Specification of r

$\mathbf{r}$ must also have strictly non-negative elements and sum to 1. This matrix exists to consider the weight of a single municipality's exports in terms of HS4 in the information of its HS6 level exports. Although $\mathbf{W}$ has interesting properties, as outlined earlier, its main diagonal is always 0 by construction. In terms of the task at hand, this means that the spatial matrix captures spatial dependency aspects of production well among municipalities, but does not capture spatial self-dependency aspects of production well, which necessitates the use of the adjustment component.

Therefore, to reassess the weight of a municipality's export data as presented in Comex Stat[6] as its actual production destined for the foreign market, municipal-level labor data is used (from the Ministry of Labor RAIS). By combining labor data with

[6]This is the web platform through which MDIC releases Brazilian international trade data.

export production, we can derive the weight of a municipality in the formal workforce of a particular sector for the entire state. This weight is used to create the matrix $\mathbf{r}$. After matching, formal workforce numbers are used to derive these municipal weights. This is similar to the procedure of Faria and Haddad (2014), who employs it using RAIS wages to resize state accounts; although the present work has employed RAIS data on number of jobs.

This procedure meets some good requirements for our demands: (i) first, it establishes municipal export data as a good indication for their true production values destined for foreign markets; (ii) it uses established and reliable data (RAIS) to measure the relevance of a sector in a given municipality; the greater the relevance of a sector in a municipality, the more it can be expected that this municipality exports part of its production[7]; (iii) it has mathematical properties - row normalization - that makes the algorithm self-contained and algebraically consistent.

There are limitations in the specifications of $\mathbf{r}$. Only formal labor data is considered, ignoring informal labor market movements. It can be argued that this limitation will be more relevant for some situations than others. First, sectors differ in the level of informality and more formal sectors will likely be better considered by the algorithm proposed here than others. However, if the formal and informal markets are similar across the state for the same sector, this issue should not be very important, since we are concerned with employment shares, not employment levels. See more details of matrix construction in the Appendix.

2.4 Global Moran's I

Moran's I is an established metric frequently used to verify if there is spatial correlation (and dependence) between different spatial units. Its use requires a spatial dataset. The global Moran's I statistic Moran (1948) is defined as:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \cdot \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{5}$$

where:

- n is the number of spatial units (municipalities of Mato Grosso here)
- x_i is the variable of interest for municipality i (soybean exports)
- $\bar{x}$ is the mean of the variable

[7]There are studies, such as Nonnemberg et al. (2024) and Faria and Haddad (2014), that use the production area (or production itself) allocated to an agricultural good to perform the downscaling of exports at the municipal level. This approach, while quite sensible for agricultural crops, tends to be impractical for industrial goods. For these types of goods, the specification of $\mathbf{r}$ proposed here is quite reasonable. The major attractive aspect of the method proposed here as opposed to that employed by Nonnemberg et al. (2024) is that it is more generic. Researchers and policymakers using any method to calculate $\mathbf{r}$ should use Monte Carlo experiments and simulation to ensure that their method of imputing municipal exports produces the most accurate results.

- w_{ij} are the elements of the spatial weights matrix $\mathbf{W}$

The expected value of Moran's I under the null hypothesis of no spatial autocorrelation is:

$$E(I) = -\frac{1}{n-1} \tag{6}$$

The variance can be calculated and used to test statistical significance.

2.5 Local Moran's I

The local version of Moran's I (LISA - Local Indicators of Spatial Association) for location i is:

$$I_i = \frac{(x_i - \bar{x})}{\sum_{j=1}^{n}(x_j - \bar{x})^2/n} \sum_{j=1}^{n} w_{ij}(x_j - \bar{x}) \tag{7}$$

Properties:

- The sum of local Moran's I is proportional to the global Moran's I:

$$\sum_{i=1}^{n} I_i = \gamma \cdot I \tag{8}$$

  where γ is a proportionality constant

- Local Moran's I can identify clusters (high-high, low-low) and spatial outliers (high-low, low-high)

These two methods are employed to analyze whether Mato Grosso's soybean production appears to be different from a random distribution in space. This can be used as an argument for adjusting the use of the spatial weights matrix to calculate municipal exports.

2.6 Monte Carlo Experiment

A Monte Carlo experiment is implemented and aims to apply the proposed algorithm to simulated export data for all Brazilian states. More clearly, a Monte Carlo experiment[8] is implemented for every Brazilian state, aiming to qualify which spatial matrix implies the lowest error levels. The vector r is also generated randomly, such that it is non-negative and has a normalized sum equal to 1. The Monte Carlo experiment is run 10,000 times for each state and each spatial weight matrix. Its result can be informative in terms of indicating which matrix tends to perform better in which scenarios - state sizes. This information is crucial to inform researchers and policymakers in terms of which spatial matrix is most appropriate for the local reality of each Brazilian state. Ideally, whenever an export imputation procedure is performed, this type of Monte Carlo simulation is necessary to attest that the errors are minimal.

[8] The R code used for this Monte Carlo experiment is present in the Appendix.

The error metric used in this experiment is the Mean Absolute Percentage Error (MAPE), whose formula is given by:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{9}$$

However, as states will be compared under the same metric, there must be that there be some normalization, and this is done by the MAPE of the queen matrix. The normalization formula is given by the following equation:

$$\frac{MAPE_i^{Inverse\ Distance} - MAPE_i^{Queen}}{MAPE_i^{Queen}} \tag{10}$$

This ensures that the errors of Brazilian states are minimally comparable. The objective of this experiment is to make explicit which type of matrix is most appropriate for the simulated data.

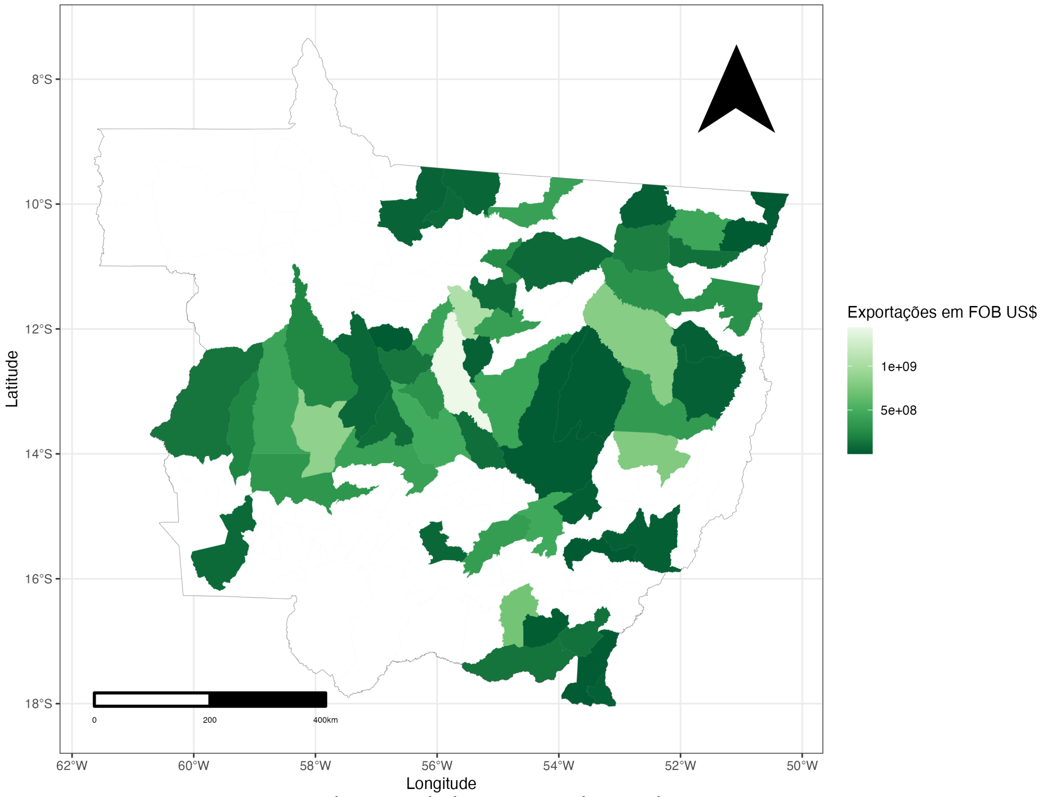
3. Data Description and Analysis

This section presents the spatial analysis of export data as presented by Comex Stat. The presence of spatial correlation and dependence can be seen as a strong indication for using spatial weight matrix to impute municipal exports.

3.1 Descriptive State of Mato Grosso Soybeans in 2024

First, we present the data in HS4 format, focusing on soybeans at the HS6 level after our procedure has been implemented. Figure 3 presents soybean exports (at the HS4 level) from Mato Grosso municipalities, as provided by Comex Stat. This is municipal data detailing which municipalities export soybeans, not municipalities that produce soybeans for export.

Figure 3. Mato Grosso soybean exports at HS4 level



Source: Authors' elaboration based on Comex Stat.

3.2 Spatial Correlation and Dependence of Mato Grosso Soybeans in 2024

As an initial exploration of this analysis, we motivate the relevance of using spatial weights matrix as the matrix weighting municipal exports, considering Moran's I statistic for Queen and inverse distance spatial weight matrices. In this case, Table 1 below presents this statistic for these two types of spatial weights matrices.

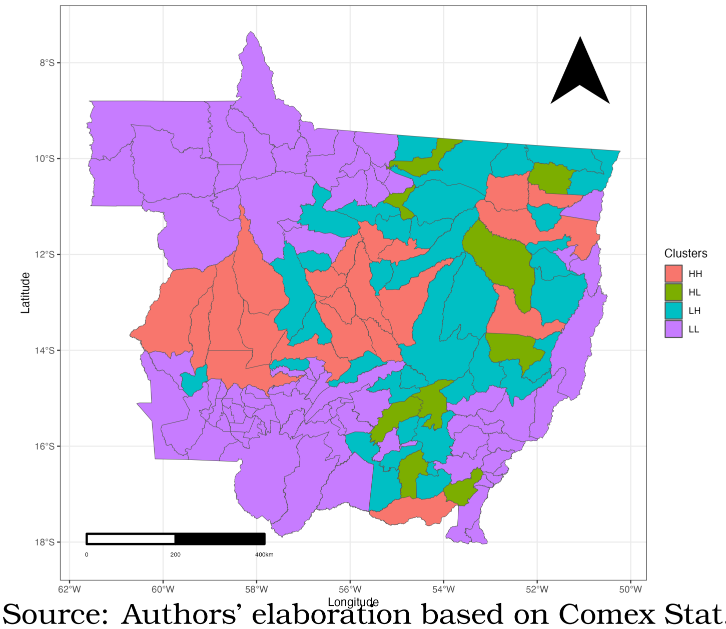
Table 1. Moran's I Test Results Under Different Spatial Weights Matrices

Metric	Queen	Inverse Distance
Moran's I Statistic	0.2053	0.0234
Expectation	-0.0071	-0.0071
Variance	0.0025	8.21×10^{-5}
Standard deviation	4.2861	3.3723
p-value	9.093×10^{-6}	0.0004

Note: Analysis performed on Soybean Exports (US$ FOB)

Table 1 indicates that in both cases, soybean exports appear to be spatially correlated. This is our argument for using a spatial matrix to distribute export data across Mato Grosso municipalities.

Figure 4. LISA of Mato Grosso soybean exports (HS4) - Queen

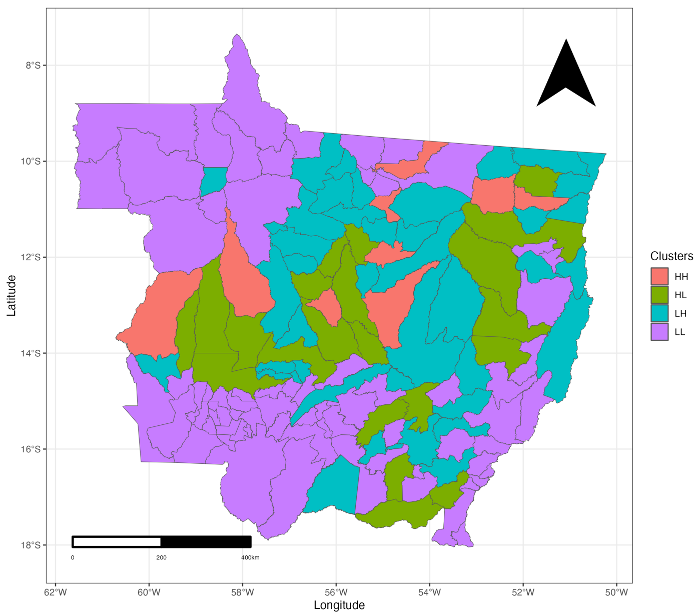


Source: Authors' elaboration based on Comex Stat.

In terms of spatial clusters, an analysis considering both Queen and inverse distance spatial weights matrices leads to the conclusion that there are export clusters in the case of soybeans (HS4)[9]. This can be seen in Figures 4 and 5:

Now, using the inverse distance spatial weights matrix:

Figure 5. LISA of Mato Grosso soybean exports (HS4) - Inverse Distance



Source: Authors' elaboration based on Comex Stat.

[9]In the Appendix, we present these maps considering the statistical significance of these clusters.

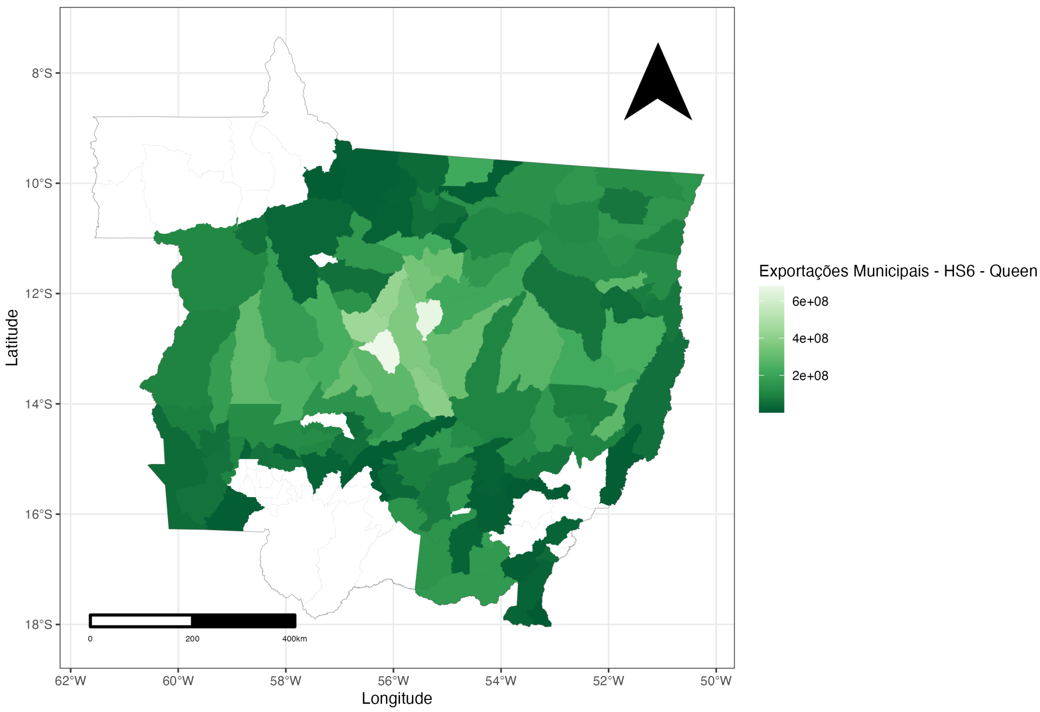
Both figures indicate that LL (low-low) clusters are more common than HH (high-high) clusters. This indicates that municipalities exporting soybeans surrounded by other exporters occur less frequently than low exporters surrounded by other low-exporting municipalities.

4. Municipal Soybean Exports from Mato Grosso in 2024

Finally, we present in the following analysis the implementation of our algorithm. We apply the algorithm proposed here using both Queen and inverse distance spatial weights matrices. The parameter α was calibrated as defined in previous sections and resulted, in the case of soybeans (HS6), in 0.9999936, which indicates that in the composition of soybean exports in the state of Mato Grosso, soybeans, except for sowing, is the most relevant product.

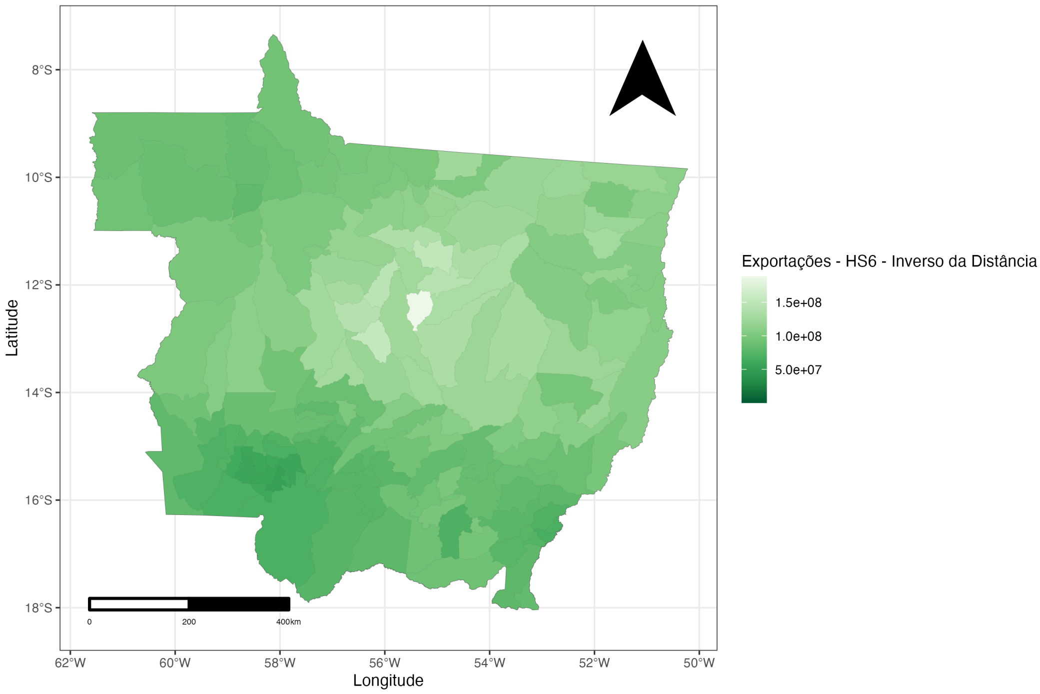
Figures 6 and 7 present the municipalities that produce for export according to our algorithm.

Figure 6. Mato Grosso soybean exports (HS6) - Queen



Source: Authors' elaboration based on Comex Stat.

Figure 7. Mato Grosso soybean exports (HS6) - Inverse Distance



Source: Authors' elaboration based on Comex Stat.

Interesting patterns emerge from these two choices of production destined for soybean exports: (i) the Queen spatial weights matrix, which postulates discrete relationships between local units, also produces a discrete-type distribution of production, oriented towards exports; (ii) the inverse distance matrix, which has non-zero relationship for any pair of locations, indicates, on the other hand, a continuous distribution of soybean production (HS6) destined for exports. This is not unexpected given how these two spatial weights matrices are defined, however they already indicate that an appropriate choice by the researcher, with field knowledge, is necessary to enable appropriate imputation of municipal production of a given good destined for exports[10].

5. Monte Carlo Experiment: Algorithm Error by State Size and Spatial Matrix

This section displays the result of the Monte Carlo experiment[11]. Figure 8 indicates the differential of the Mean Absolute Percentage Error. Positive values of this variable indicate that the queen matrix tended to produce more accurate results in

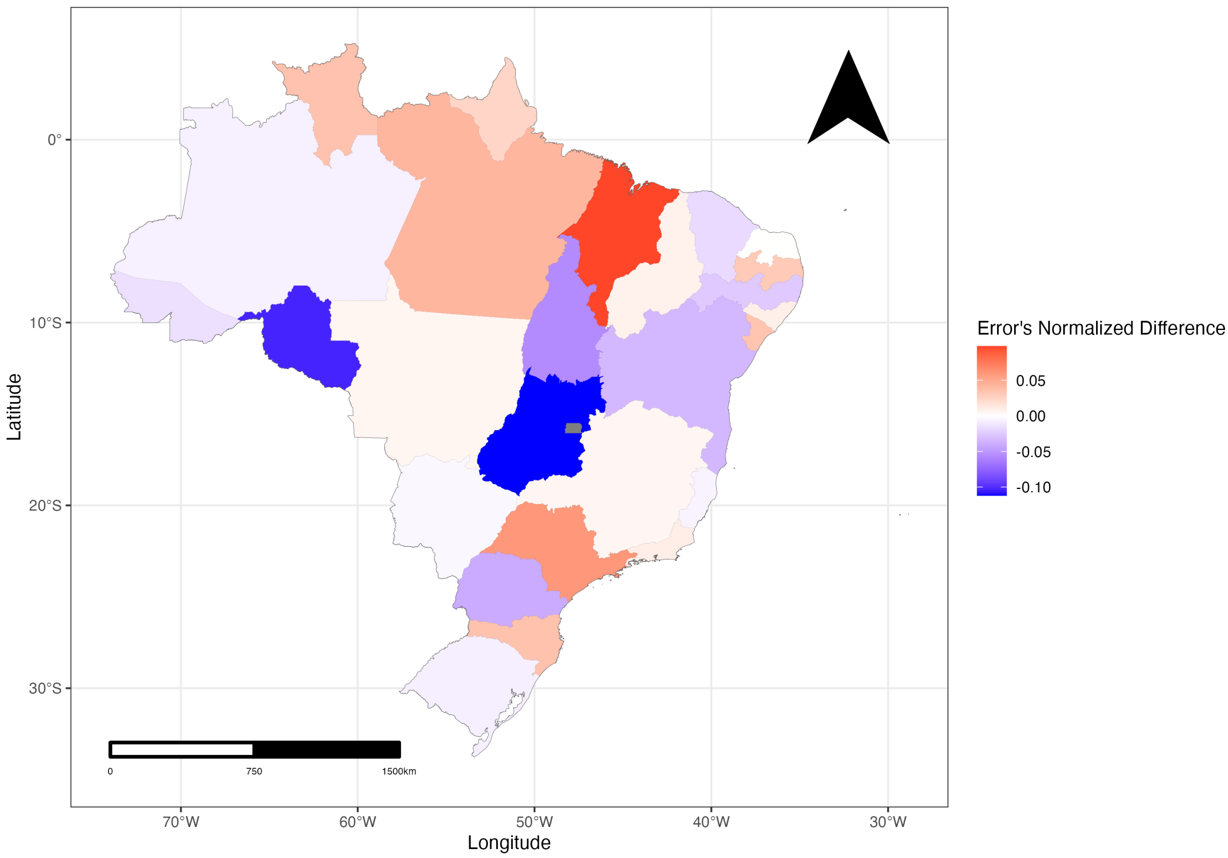
[10]Data used in this paper is available upon request and will also be part of a growing and soon-to-be public repository of the application of the algorithm proposed here for several goods, years, and Brazilian municipalities.

[11]Code available in the Appendix.

terms of predicting municipal exports in the simulation exercise. At the same time, negative values indicate that the inverse distance spatial matrix produced more accurate results.

Figure 8 indicates that for states like Rondônia and Goiás, the inverse distance type spatial matrix tended to produce more precise results in terms of imputing municipal export values. At the same time, for states like Maranhão, Pará, Amapá, and Goiás, the queen type spatial matrix tended to allow the algorithm better prediction of the true values of municipal exports.

Figure 8. State Monte Carlo Experiment for Different Spatial Matrices



Source: Authors' elaboration based on simulated data.

6. Academic and Public Policy Relevance

The algorithm proposed here is an evolution of previous works, such as those carried out in Nonnemberg et al. (2024), given that the exercise performed here is an advancement in relation to downscaling. Contrary to that study, however, the aim was to create: (i) a generic method of export imputation; (ii) a method that leverages field knowledge by researchers and stakeholders. These two requirements fulfilled by the method proposed here indicate a first success in this type of task.

It is relevant to state that the type of information this algorithm provides is not known by the public sector, given that this type of information is not collected in any way. This makes this effort even more important as a way to inform policymakers and academics interested in analyzing the relevance of export production in a given municipality.

7. Concluding Remarks

This article aimed to propose an algorithm to solve a problem that belongs to Brazilian international trade data. Currently, it is known which regions originated the export movement, however it is not known which regions produced a good destined for the foreign market.

This means that this work is unique in its approach. Leveraging international trade and regional science methods requires some ingenuity, however allowing field and regional knowledge to act in the method's application seems to be a way to assert its interest for various types of goods.

This analysis can also be applied to exports directed to a single country or a collection (bloc) of countries. The method is well-suited for any consistent destination, dealing with destinations other than global.

This paper deals directly with trade in goods, not considering trade in services. Brazilian Trade Ministry (MDIC) does not publicize the trade in service numbers with the same granularity as the trade in goods. Although our method is agnostic in the type of trade considered, the spatiality considerations here assume the goods as not as mobile in space as services. This mobility makes the locality attribution to the exports much more complicated than in the case of the trade in goods. Hence, we do not recommend using this method for trade in services, without some adaptations.

References

Araujo, R. (2017). Uso/ocupação do solo em área de preservação permanente do rio paraguai entre os bairros cavalhada i e centro, cáceres - mato grosso. `https://www.researchgate.net/publication/321070677`. Monografia de graduação, Universidade do Estado de Mato Grosso.

Arbia, G. et al. (2014). *A primer for spatial econometrics*, volume 230. Springer.

Barbosa, A. M. (2024). Política de desmatamento zero: Uma estimativa de perdas potenciais de exportação de soja do cerrado mato-grossense.

Biavaschi, M. A. C. (2005). Borgismo, coronelismo e dissidências políticas nas páginas dos jornais santa-marienses. *BIBLOS - Revista do Instituto de Ciências Humanas e da Informação*, 17:165–178.

Biavaschi, M. A. C. (2006). Coronelismo e conflitos políticos em santa maria (1889-1930). *Revista Sociais e Humanas*, 19(1):61–77.

Cabral, G. d. S. A. (2022). A paradiplomacia de goiânia e aparecida de goiânia (2017-2020): um estudo comparado.

Cardoso, D. M. (2018). Mineração e subdesenvolvimento: Impactos da atividade mineradora nos municípios de canaã dos carajás, marabá e parauapebas (2004–2015). Dissertação de mestrado, Universidade Estadual de Campinas. Repositório da UNICAMP.

Castilho, D. and Arrais, T. A. (2017). A ferrovia norte-sul ea economia regional do centro-norte do brasil. *Sociedade & Natureza*, 29(02):209–228.

Castriota, R. (2024). “aqui a vale é o estado”: neoextrativismo e autoritarismo na cidade, no campo e na floresta na região de carajás. *revista brasileira de estudos urbanos e regionais*, 26:e202408.

Corrêa, J. J. L. et al. (2015). *Copal do Brasil: ocorrência e caracterização físico-química da resina jutaicica de Santarém*. PhD thesis, Universidade Federal do Oeste do Pará.

da Silva, J. M., Matsuoka, S. T., Costa, S. R., Alva, Y. K., Tozi, L. A., and do Nascimento Nascimento, M. V. (2024). Mapeamento do potencial de movimentação de cargas: Estudo de caso no aeroporto de são josé dos campos (sjk) por meio de dados de exportações de municípios: Artigo inédito. foi revisado por pares e aprovado para apresentação oral no congresso cimatech da fatec de são josé dos campos em 12/dezembro/2024. *Revista Ciências Exatas*, 30(2).

Da Silva, L. P. B. (2013). Redes de transporte de mercadorias e fronteira: a logística nas cidades gêmeas de corumbá (brasil) e porto suárez (bolívia). *Revista transporte y territorio*, (9):141–160.

da Silva Araújo, D. and da Silva, R. M. (2018). O fenômeno da zona de processamento de exportação no rio grande do norte. *Revista de Economia Regional, Urbana e do Trabalho*, 7(1):56–82.

De Almeida, L. A. (2018). Impacto do capital humano no crescimento regional: Um estudo sobre os municípios do sudeste brasileiro.

de Assis, T. R. (2024). Terá consciência a europa do que esta fazendo por não fazer? *Nova Hileia | Revista Eletrônica de Direito Ambiental da Amazônia. ISSN: 2525–4537*, 16(1).

de Oliveira, L. R., Brene, P. R. A., do Nascimento, S. P., and Ferreira, C. R. (2017). Importância do pib e das exportações sobre o coeficiente de atratividade nos dez maiores municípios do estado do paraná. *Revista de Economia Regional, Urbana e do Trabalho*, 6(2):30–51.

Dorneles, C. P. (2024). Análise dos fatores críticos de sucesso nas exportações pela ótica dos técnicos do programa peiex atuantes na região nordeste do estado do rio grande do sul.

dos Santos, H. F. (2017a). Competitividade regional do setor sucroenergético no triângulo mineiro/alto paranaíba. *Formação (online)*, 1(25).

dos Santos, H. F. (2017b). A dinâmica da recente expansão do setor sucroenergético no triângulo mineiro/alto paranaíba: Da internacionalização à formação dos grandes grupos. *Estudos Geográficos: Revista Eletrônica de Geografia*, 15(1):106–132.

Faria, W. R. and Haddad, E. A. (2014). Estimação das elasticidades de substituição do comércio regional do brasil. *Nova Economia*, 24(1):141–168.

Ferreira, B. M., Missio, F. J., and Sonaglio, C. M. (2019). Notas sobre as etapas de desenvolvimento de três lagoas (ms): um exercício rostowiano. *Revista GEOSUL*.

Hidalgo, Á. B. (2017). Comércio internacional, salários e prêmios por qualificação nos municípios brasileiros no decênio.

Kelejian, H. and Piras, G. (2017). *Spatial econometrics*. Academic Press.

Kühn, D. D. (2008). Oportunidades, ruralidade e pobreza no rio grande do sul: as relações apresentadas pela abordagem das capacitações nos municípios gaúchos.

Kuhn, E. A., Sattler, M. A., and Magnus, L. D. (2017). Contribuições do conceito e da abordagem de metabolismo urbano para a avaliação do custo das decisões ambientais. *urbano*, 282:298.

Lacerda, F. C. C., Pessoti, G. C., and Jesus, J. A. (2013). Transformações estruturais,(des) concentração espacial e inserção internacional: Uma análise para a economia baiana com base na teoria da base produtiva. *Nexos Econômicos*, 7(1):141–178.

Leal, A. M. M. (2024). Trade share setorial médio para municípios.

Luiz, B. M. d. C. and de Oliveira, E. C. (2020). Uma análise das vantagens comparativas das exportações do estado do amazonas com o brasil, no período de 2000-2018. *RDE-Revista de Desenvolvimento Econômico*, 2(46).

Mancini, R. M. and Carneiro, M. S. (2018). A construção do mercado de trabalho de carvão vegetal na amazônia oriental: estratégias corporativas e crítica social. *Estudos de Sociologia*, 23(45).

Mendes, K., Hidalgo, A. B., and Luchine, A. A. (2020). O comércio internacional dos municípios brasileiros entre 2000 e 2010: Uma aplicação do modelo gravitacional. In *Anais do 48º Encontro Nacional de Economia (ANPEC)*.

Mendes, P. S., Amaral, P. V., de Carvalho, L. R., and Brito, G. (2021). Condicionantes territoriais do salário industrial brasileiro: uma análise municipal para a indústria de transformação-2006 a 2016. *Revista Econômica do Nordeste*.

Metzdorff, D. d. S. et al. (2015). Uma análise empírica sobre a importância do comércio internacional para o crescimento dos municípios brasileiros.

Moran, P. A. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2):243–251.

Moreira, M. M., Blyde, J., Volpe, C., and Molina, D. (2013). Longe demais para exportar.

Nonnemberg, M. J. B., Martins, M. M. V., Cechin, A., da Silva Vianna, R., Cruz, C. C. P., Silva13, F. A., Bispo, S. Q. A., de Luna, F. E., and Santos, A. (2024). Efeitos da lei europeia contra desmatamento sobre as exportações brasileiras. *Publicação Expressa*.

Pena, H. W. A., Vilhena, F. C. F., Dallemole, D., Lisbôa, E. G., Neto, L. L. d. S., and Otobo, A. O. (2022). Inserção internacional: Uma aplicação regional da matriz de competitividade na amazônia. In *CRESCIMENTO E DESENVOLVIMENTO NUMA PERSPECTIVA INTERDISCIPLINAR: ENSAIOS SOBRE O CRESCIMENTO ECONÔMICO BRASILEIRO*, volume 1, pages 239–260. Editora Científica Digital.

Perpetua, G. M. et al. (2012). A mobilidade espacial do capital e da força de trabalho na produção de celulose e papel: um estudo a partir de três lagoas (ms).

Perpetua, G. M. and Junior, A. T. (2013). Dinâmica geográfica da mobilidade do capital na produção de celulose e papel em três lagoas (ms). *Revista da ANPEGE*, 9(12):55–69.

Ponte, A. G. R. (2022). Impactos econômicos do complexo industrial e portuário do pecém-cipp.

Rajão, R., Soares-Filho, B., Nunes, F., Borner, J., Machado, L., Assis, D., Oliveira, A., Pinto, L., Ribeiro, V., Rausch, L., et al. (2020). Maçãs podres do agronegócio brasileiro. *Science*, 369(6501):246–248.

Saboia, J. (2013). *Desconcentração Regional da Indústria de Transformação, Exportações e Intensidade Tecnológica–2003/2013*. PhD thesis, Instituto de Economia, Universidade Federal do Rio de Janeiro.

Sachser, L. E. (2015). Estudo exploratório para a melhoria das estimativas das receitas do issqn para o município de ivoti: período 2010-2015.

Sant'Anna, V. P. (2015). *Infraestrutura portuária no Brasil: uma análise do impacto do tempo dos procedimentos portuários sobre as exportações brasileiras*. PhD thesis, Universidade de São Paulo.

Sant'Anna, V. and Júnior, S. K. (2016). Port efficiency and brazilian exports: A quantitative assessment of the impact of port procedures time. In *Anais do XLIII Encontro Nacional de Economia [Proceedings of the 43rd Brazilian Economics Meeting]*, number 125. ANPEC-AssociaÃ §.

Scazufca, P. S. (2008). *Determinantes das exportações industriais: evidência empírica dos municípios paulistas*. PhD thesis, Universidade de São Paulo.

Seibert, C. E. and Perobelli, F. S. (2022). Estrutura da cadeia produtiva do agronegócio nas regiões imediatas do centro-oeste brasileiro: Uma análise de insumo-produto a partir do método iioas. *50º Encontro Nacional de Economia (ANPEC). Anais eletrônicos [. . .]. Fortaleza, CE.*

Seibert, C. E., Perobelli, G. S., Carrara, A. F., and Brito, M. S. (2024). Acordo mercosul-união europeia: Desafios e oportunidades para as exportações de carne brasileira. In *Anais do XXII Encontro Nacional da Associação Brasileira de Estudos Regionais e Urbanos (ENABER)*, Vitória, ES.

Silveira, C. E. B. and Braga, H. (2023). A relevância e os desafios do setor extrativista mineral para o desenvolvimento regional do quadrilátero ferrífero em minas gerais. *Cuadernos de Educación y Desarrollo*, 15(9):8044–8060.

Souza, V. and Raiher, A. (2022). O auxílio emergencial e seu efeito na dinâmica do mercado de trabalho dos municípios de são paulo: uma análise espacial. *Revista Economia Política do Desenvolvimento*, 13(29).

Tatsch, A. L. and Batisti, V. (2013). Os fluxos de comércio do rio grande do sul: trocas interestaduais e com o resto do mundo. *Revista Economia Ensaios*, pages 27–46.

Vivaldi, G. A. D. (2021). A dinâmica do café e suas implicações no desenvolvimento regional: uma análise na microrregião de varginha.

Vivaldi, G. A. D., Júnior, P. d. S. P., and Alves, A. F. (2020). A cafeicultura na microrregião de varginha-mg. *Interação-Revista de Ensino, Pesquisa e Extensão*, 22(2):143–159.

A. Appendix

A.1 Llama 3.2

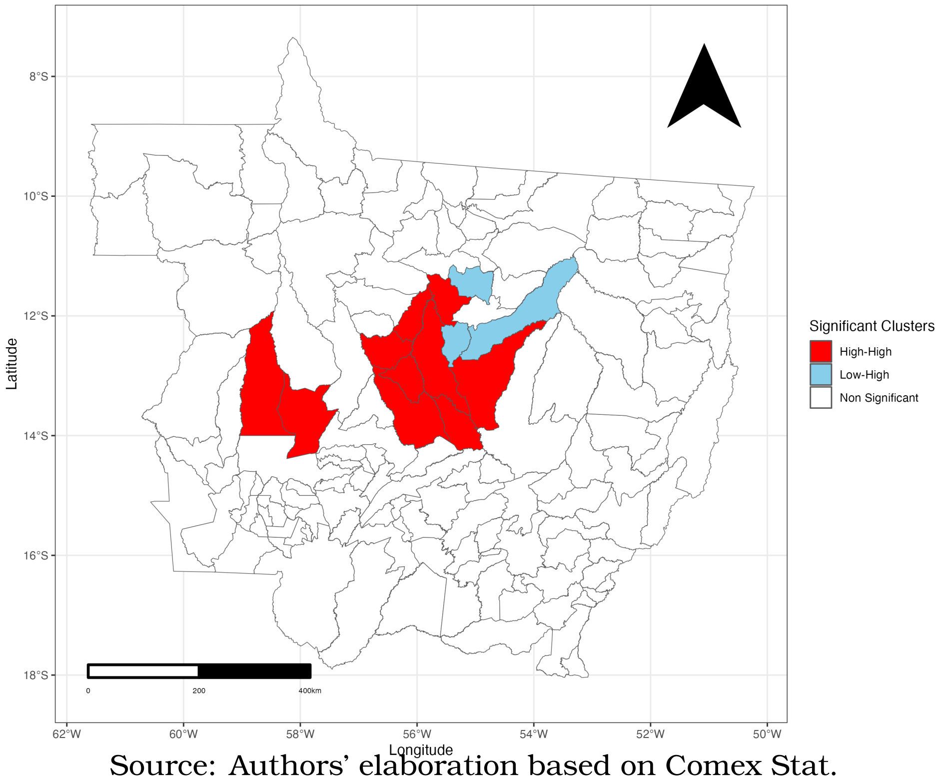
This brief Appendix aims to explain the parameters used in the Llama 3.2 model employed here. Llama is a free-use LLM developed by Meta. Contrary to paid LLMs such as Open AI's Chat GPT, Anthropic's Claude.AI, Microsoft's Copilot, Google's Gemini, and X's Grok, Llama can be run locally on a machine, without the need for internet and the common limitations faced by paid tiers in the previously listed LLMs. Llama was employed using the following computer specifications: MacBook Air M1 (2020), with 8GB RAM and a 256GB NVMe SSD. The query asked if Llama was able to perceive a distinction in the pdf, which was whether the authors consider Comex Stat municipal exports different from municipally produced exports. Then, Llama was asked to produce a 150-word text, justifying its decision about whether an article documented this difference or not. As the main text indicates, this is not something that most academic articles using Comex Stat municipal export data are concerned with.

A.2 The construction of vector r

Vector r is crucial for allowing the consideration of self-production based on Comex Stat trade data. By combining formal labor data with trade data, we have an important mapping about the descriptive and relational nature that involves labor and trade. Further elaboration on how this mapping is possible deserves consideration. First, RAIS is a formal labor dataset released annually, containing information about each formal job that exists in Brazil. It contains information about wages, education, job locality, employment sector, and much more information. Regarding the employment sector, RAIS has a bottom-up approach, allowing sector identification in a very disaggregated way. This disaggregated sector is used in our analysis. Municipal trade data from Comex Stat, as previously established, has the disaggregation level of HS4, that is, Harmonized System 4. Moreover, as argued, this is not truly municipal data, it is data attributed to a municipality based on international trade bureaucracy. The mapping between these labor and trade data occurs based on officially released mapping of labor data to Mercosur goods coding (NCM-Mercosur Common Nomenclature). As this Mercosur nomenclature is derived from HS6, this triple mapping is straightforward and without need for imputation or deviations.

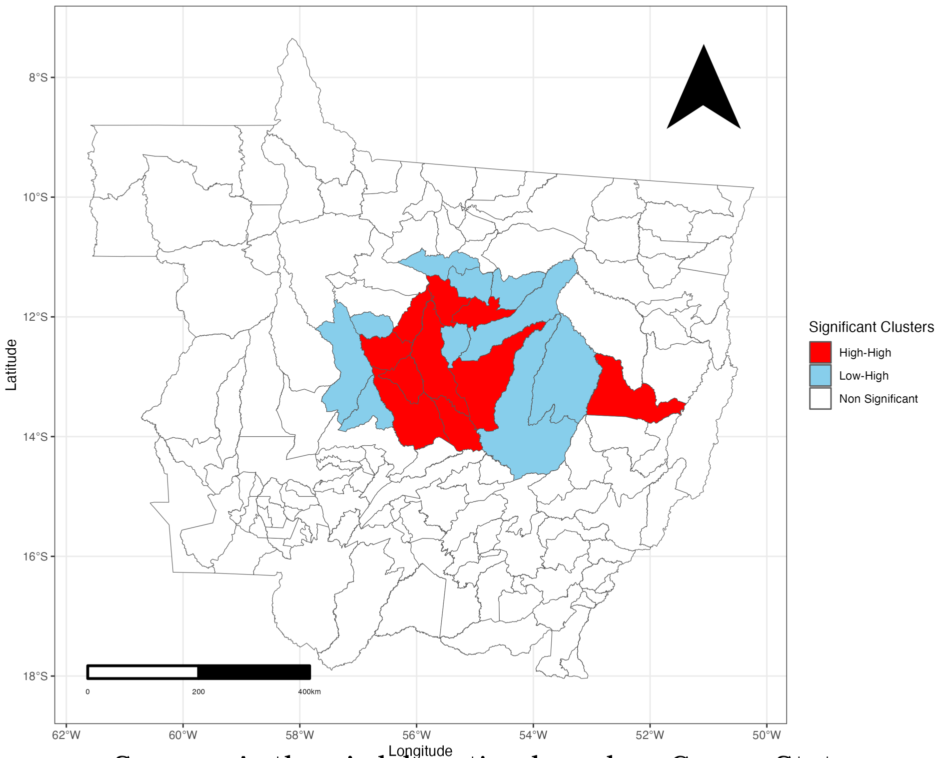
A.3 Significant Clusters of the Mato Grosso's soybeans exports in 2023

A.1. LISA of Mato Grosso soybean exports (HS4) - Queen - Significant Clusters



Source: Authors' elaboration based on Comex Stat.

A.2. LISA of Mato Grosso soybean exports (HS4) - Inverse of Distance - Significant Clusters



Source: Authors' elaboration based on Comex Stat.

A.4 Monte Carlo Experiment

The Monte Carlo experiment simulates the errors (MAPE) of using different spatial weights matrices (queen and inverse distance) about 10,000 times for each Brazilian state. The objective here is to verify which matrix tended to produce the most accurate results in terms of predictions of exports at the HS6 level, based on simulated true values.

The vectors of municipal exports at HS4, HS6 level and municipal employment in a given sector are iteratively simulated in the Monte Carlo simulation in order to consolidate the understanding of which matrix tends to produce the best results in terms of export imputation. The R code follows below:

```
library(tidyverse)
library(spdep)
library(sf)
library(geobr)
library(progress)
set.seed(12345678)
# Load municipalities data
municipios <- geobr::read_municipality(year = 2020)
estados <- unique(municipios$abbrev_state) %>% setdiff("DF")
# Function to calculate spatial matrix
calcular_matriz_espacial <- function(municipios, coords, perfil_matriz) {
  if (perfil_matriz == "queen") {
    nb <- poly2nb(municipios, queen = TRUE)
    lw <- nb2listw(nb, style = "W", zero.policy = TRUE)
  } else if (perfil_matriz == "inversa_distancia") {
    inv_dist_nb <- dnearneigh(coords, 0, Inf)
    lw <- nb2listw(
    inv_dist_nb,
    style = "W",
    glist = lapply(nbdists(inv_dist_nb, coords), function(x) 1 / x)
    )
  }
  return(spdep::listw2mat(lw))
}
# Function to simulate errors
simular_erros <- function(W, expsh6, expsh4, emprego) {
  r <- emprego / sum(emprego)
  alpha <- sum(expsh6) / sum(expsh4)
  soja_mun_total <- alpha * (W %*% expsh4)
  Diferencial <- sum(alpha * expsh4) - sum(soja_mun_total)
  ajuste <- soja_mun_total + r * Diferencial
  mape <- mean(abs((expsh6 - ajuste) / expsh6)) * 100
  return(mape)
}
# Function to simulate by state
simular_estado <- function(estado, num_simulacoes, municipios) {
  municipios_estado <- municipios %>% filter(abbrev_state == estado)
  coords <- st_coordinates(st_centroid(st_geometry(municipios_estado)))

```

```
    W_queen <- calcular_matriz_espacial(municipios_estado, coords, "queen")
    W_inversa <- calcular_matriz_espacial(municipios_estado, coords,
        "inversa_distancia")

    pb <- progress_bar$new(
    format = paste0("  Simulating for ", estado, " [:bar] :percent in :elapsed
        s"),
    total = num_simulacoes, clear = FALSE, width = 60
    )

    resultados <- vector("list", num_simulacoes)
    for (i in seq_len(num_simulacoes)) {
      pb$tick()
      expsh6 <- runif(nrow(municipios_estado), 0, 12345666)
      emprego <- runif(nrow(municipios_estado), 0, 1)
      expsh4 <- expsh6 + runif(nrow(municipios_estado), 0, 591029286)

      resultados[[i]] <- data.frame(
      perfil_matriz = c("queen", "inversa_distancia"),
      mape = c(
      simular_erros(W_queen, expsh6, expsh4, emprego),
      simular_erros(W_inversa, expsh6, expsh4, emprego)
      ),
      tamanho_estado = sum(expsh4)
      )
    }

    bind_rows(resultados) %>% mutate(estado = estado)
  }
  # Run simulations
  resultados_todos_estados <- map_dfr(estados, ~ simular_estado(.x, 10000,
      municipios))
```