Interpolation of methane emissions in a region of California

Sara Pinto

NOVA University of Lisbon – NOVA IMS, 20230915@novaims.unl.pt

Abstract: The objective of this work was to apply techniques and methodologies that allow interpolating methane emissions. The data used were obtained from a study conducted in California between 2016 and 2018, carried out using an unmanned aerial vehicle, and equipped with a spectrometer, which allowed to quantify the emissions of this pernicious gas into the atmosphere.

The methodology was initiated through exploratory data analysis using Excel software and ArcGIS Pro, and included methodologies such as Voronoi maps and Moran statistics (Local and Global). The analysis revealed an asymmetric distribution of emissions, with many low values and some very high values, indicating a marked variability. The absence of spatial bias and global autocorrelation limited the choice of interpolation methods. In this sense, two deterministic methods were performed, IDW and Kernel, with the latter demonstrating greater precision.

Despite the challenges in accuracy due to the variability of the data and the "bull eye" effect at some isolated points, an approximate surface area of methane emissions was obtained. The existence of local anisotropy was suggested in the south of the study area, but there was no evidence of global anisotropy. The application of deterministic interpolation techniques provided a basis for understanding the spatial distribution of methane emissions.

Key words: Methane emissions; Interpolation; Deterministic Methods

1. Introduction

Methane (CH4) is one of the seven gases responsible for the greenhouse effect, and although it occurs naturally in nature, its rapid increase in recent decades comes from human activities, namely the coal industry, oil, agriculture and the decomposition of waste in landfills (Maasakkers et al., 2022). Although this gas has a shorter life in the atmosphere than CO2, it is estimated that 30% of climate change is due to methane alone, due to its high efficiency in trapping heat. For this reason, targets for its reduction have been established in several strategies and international agreements, such as the Global Methane Pledge, signed by more than 150 countries, including Portugal (*Global Methane Pledge*, 2021)

The study on which the work is based was conducted between 2016 and 2018, in 5 sampling campaigns (Duren et al., 2019). A total of 272,000 infrastructures were surveyed using an unmanned aerial vehicle equipped with a spectrometer. As part of this research, super-emitters were identified – only 10% of the points contributed to 60% of the total emissions.

1.1 Goals

The main objective of this work is the application of the techniques and methodologies taught in the discipline of Spatial Statistics, namely, interpolation, in a relevant data set.

The specific objectives are: (a) to conduct exploratory data analysis and spatial exploratory data analysis; b) to ascertain the existence/absence of spatial trend of the data, as well as the global and local autocorrelation; c) create an interpolated surface.

2. Study Region & Data

The study region is located in the State of California, United States (Figure 1). The data used in this study were obtained from the supplementary materials of the article *California's methane super-emitters* (Duren et al., 2019). Originally published in a pdf table, the data was converted to an excel spreadsheet, and later to a shapefile. The data consisted of 1181 drone samples at 691 different geographic points, overlapping potential methane sources. In some places, sample collections were more recurrent (at successive times and days), while in other places, sampling was more sporadic. For this work, the data obtained during the month of October 2016 were selected, which correspond to 98 geographical points. The average methane emission values (Qplume (kg/hr)) obtained during that period was calculated and associated with the respective point.

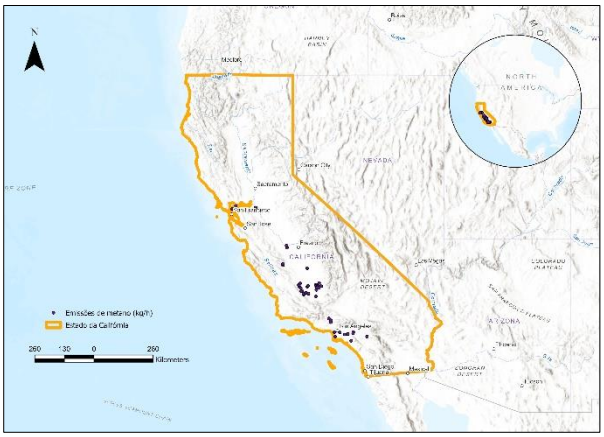


Figure 1 - Area of study and data used in the scope of this work.

3. Methodology

The methodology followed in this work was based on the materials available in the discipline of Spatial Statistics (Costa, 2023a) (Costa, 2023b) (Costa, 2023c). The workflow can be consulted in the Appendix B. The exploratory analysis of the data was performed in Excel software and the remaining analyses - Voronoi maps, Moran Local I statistics, Moran Global I statistics and interpolation - were performed in ArcGIS Pro.

3.1 Exploratory Interpolation Tool

For the configuration of the Exploratory Interpolation tool, the option *Hierarchical sorting with tolerances*, as a method of comparison, using the criteria listed in the Table 1.

Table 1 - Exploratory Interpolation tool settings.

Parameter	Tolerance Type	Tolerance value
Square Root of Mean Square Error (RMSE) (Precision)	Percent	5
Mean error (ME) (bias)	Percent	10

4. Results and Discussion

4.1 Exploratory Data Analysis

The sample size corresponds to 98 methane emission measurements, ranging from 30 kg/h to 1883 kg/h. The average emissions were high, 260.6 kg/h[1], and the median was 166.5 kg/h, indicating positive asymmetry, and thus suggesting the presence of many low values and some very high outliers. The standard deviation was particularly high, 278.6 kg/h, reflecting a large variability in the measurements. The mode of emissions was 139 kg/h, which is consistent with the positive asymmetry. In addition, the high kurtose values and asymmetry confirm the unequal distribution of values (Appendix C).

Confirming what was already anticipated through the parameters of descriptive statistics, the analysis of the histogram (Figure 2) shows a positive asymmetric distribution. This means that a significant part of the data is accumulated on the left side of the distribution, but the presence of *Outliers* causes a long tail to the right, influencing the mean and dispersion measurements

[1] This value gains more significance compared to other reference values. If the value for the month of October is representative of the year 2016, then approximately 2.3 million tons would be emitted annually. In Portugal, during the same year, approximately 372 thousand tons were registered, a considerably lower value (https://www.pordata.pt/portugal/emissoes+de+gases-1081-8827). It should be noted that the sampling area is smaller than the area of mainland Portugal.

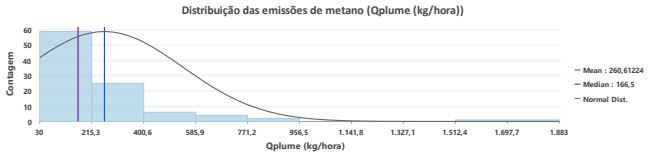


Figure 2 - Histogram of methane emission values obtained during October 2016.

4.2 Exploratory Spatial Data Analysis

There are no spatial trends in the distribution of the data (Figure 3).

In fact, there are low emissions near points where higher emission values have been recorded, indicating significant variability. In addition, the points are not evenly distributed throughout the study area, with areas where they are sparse, and others where they are highly concentrated.

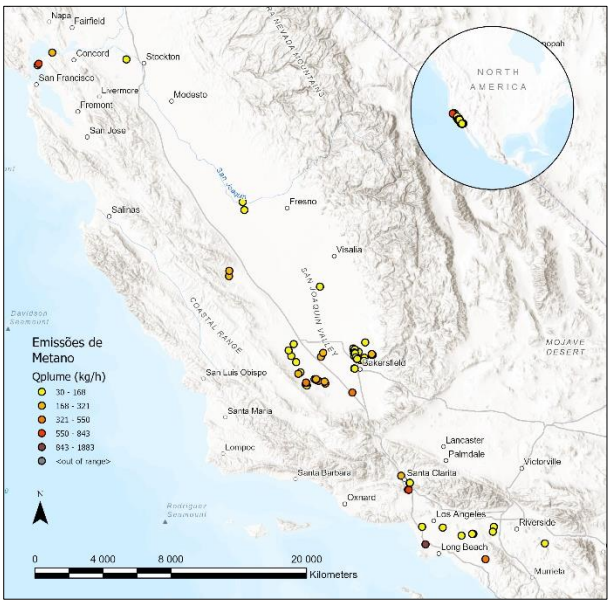


Figure 3 - Spatial representation of methane emissions (Qplume) using graded color symbology.

In view of the existence of overlapping points that prevent the observation of all the data, a histogram was created, selecting the extreme values to determine their location. It was found that there are points with extreme emissions throughout the study area, namely in the North, Central and South (Appendix D).

In order to investigate the variability of the data in more detail, the determination (Appendix E) and spatial representation of quartiles (Appendix F). However, the analysis of the indicator maps did not reveal any new information. In fact, the data do not show a clear spatial trend, with high and low values in all areas of the study area.

4.2.1 Maps of Voronoi

The analysis of the simple Voronoi map (Figure 4) reveals that the areas near Bakersfield and Los Angeles were sampled with greater intensity, since in these places the polygons are smaller. On the other hand, the areas located to the North and East showed a lower density of measurements, given the large size of the polygons. It should be noted that the interpolation of these areas may not be reliable, especially for sites that are further away from the sampling point.

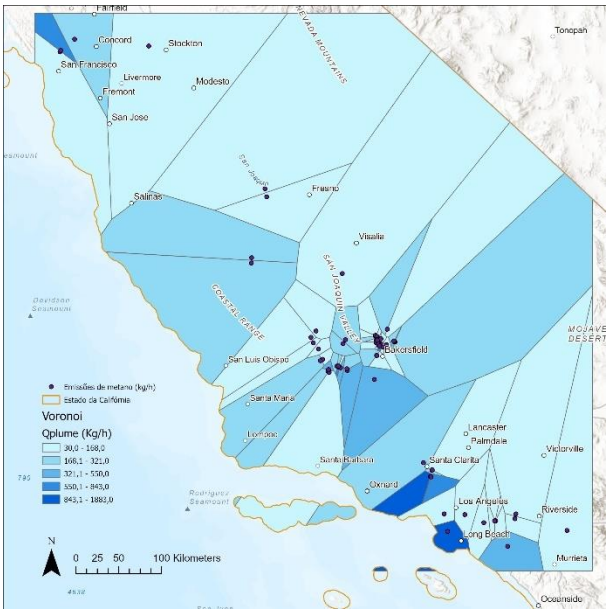


Figure 4 - Simple Voronoi map of methane emissions (Qplume (kg/h)).

This map reveals, once again, the absence of a spatial trend in the data. In fact, there are areas of high emissions in the north, near San Francisco, in the central zone, near Bakersfield, and in the south, near Los Angeles, surrounded by areas that showed low emissions, and this appears to be a random pattern.

The spatial representation of the data also does not seem to suggest a clear ellipse or orientation of the polygons that indicates a direction of greater continuity, therefore there does not seem to be anisotropy.

Similar to the simple Voronoi map, the Voronoi maps representing the mean and standard deviation

of methane emissions (Appendix G) indicate that the highest values are near the city of San Francisco, Bakersfield and Los Angeles. This may be indicative of a proportional effect in these 3 areas, since in these places where the mean is higher, the standard deviation also assumes high values. It should be noted that in areas with a high standard deviation (i.e. with high local variability), forecasts are less accurate.

4.2.2 Local Moran Statistics I

Moran's Local I statistic revealed that autocorrelation is not significant for most data, which may be indicative of a random distribution (Appendix H). However, in the Los Angeles area, some clusters (5 points) and 1 outlier were identified (Figure 5). The outlier is located near Loma Ridge and has a significantly high methane value when compared to its neighboring points. The clusters are of the low-low type, and are close points whose values are low. In this sense, there is positive local autocorrelation in this specific area.

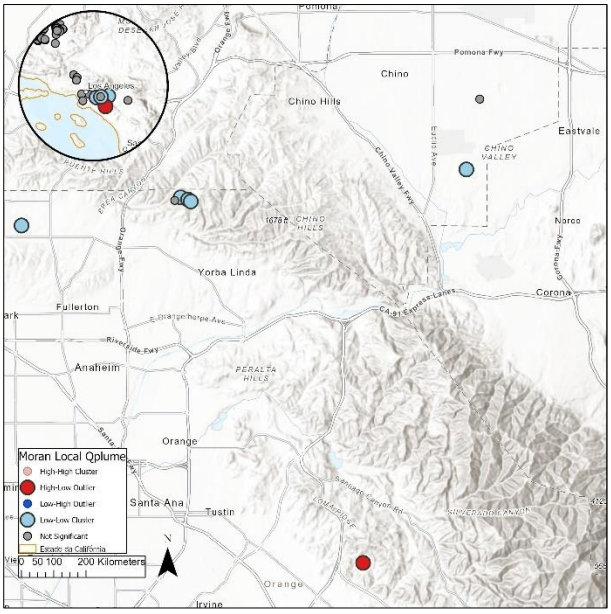


Figure 5 - Detail of the clusters and outlier identified through Local Moran's Statistic I.

4.2.3 Global Moran's Statistics

The variable under study is continuous – it is a gas that is emitted into the atmosphere – so the closest points will display more similar values than the more distant ones. However, this effect is expected to wane rapidly. In this sense, the conceptualization of the Inverse *Distance Squared spatial relationship was selected* and the tool to determine the Global Moran statistic was executed.

The value of the Moran index (0.11) is positive, which could indicate that the data are more spatially grouped than would be expected if the underlying spatial process were random (Table 2). However, the p-value is > 0.05 (the determined value was 0.14), and therefore the index is not statistically significant, indicating that the underlying spatial process is random.

Table 2 - Parameters determined using Global Moran's Statistics I.

Parameter Determined	Value
Moran's Index	0,106223
Expected Index	-0,10309
Variance	0,006246
Z-Score	1,474540
P-value	0,140336

4.3 Interpolation

Given the nature of the data and the study area (few data for a considerable area), and the lack of global autocorrelation, only deterministic interpolation methods were selected, since with these constraints, kriging methods do not bring advantages.

4.3.1 Exploratory Interpolation Tool

The results obtained through the Exploratory Interpolation tool (Annex I), indicate that the model with the best ranking was the Kernel. In fact, this model was the one with the lowest RMSE, which is a good indicator of the accuracy of the model's predictions. The efficiency of the kernel may be related to the fact that it does not estimate values for the entire study area (Appendix J). All other models come with the same ranking, although there are some differences in RMSE and ME.

4.3.2 Inverse Distance Weighting – IDW

As mentioned earlier, the data exhibits a random spatial pattern, so there are no advantages to using stochastic methods. In this sense, the IDW appears to be a deterministic methodology to be explored. Despite numerous attempts to define the number of neighbors, to select sectors and angles (in the Appendix K only a few of the attempts are found), the values obtained during cross-validation, namely the mean error (ME=2.86) and the square root of the

mean squared error (RMSE=305.08) are very high. These values indicate that the interpolation performance is compromised, and that the model has a low accuracy in predicting methane emissions. It should be noted that the values obtained are higher than those found by Exploratory Interpolation, because the tool used the value of 1 in the Power parameter in the configurations of its model, and in the present work the value 2 was used.

The methane emissions surface obtained (Figure 6) shows that the most problematic areas, i.e. the sites with the highest emission levels, are located near the city of Los Angeles (coastal area to the south in Figure 6) and to the city of San Francisco (coastal area to the north in Figure 6). The central zone of the study area is characterized by some variability, in which emissions range from low to moderate. Another noteworthy aspect is the low emissions detected in an area close to the highest emissions in the study area, near Los Angeles, demonstrating once again the lack of autocorrelation of the data. Although methane is often associated with rural areas, and in particular, livestock, the phenomenon of super emitters, i.e., specific sources that emit abnormally high amounts of gas, is well documented in California, and elsewhere in the world (Duren et al., 2019) (Elser et al., 2021) (Schuit et al., 2023) (Maasakkers et al., 2022). Usually, these super-emitters are infrastructure from the oil industry, natural gas, or are landfills, as is the case with the Los Angeles outlier. Landfill managers are often able to mitigate the problem of methane leakage when they are alerted, demonstrating the immediate usefulness of studies like this one.

Another problematic aspect is the isolated spots, which produced the "bull eye" effect. Although we tried to eliminate this effect by widening the search radius, the effect remained, so we opted for the first version in which there is less smoothing of the interpolated surface.

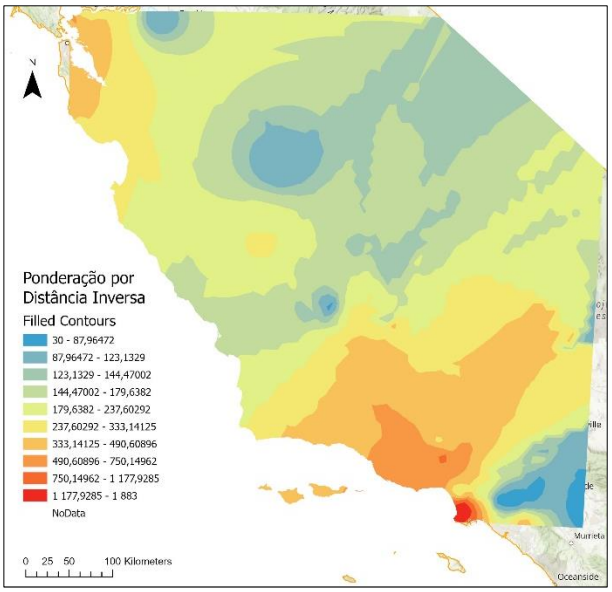


Figure 6 - Interpolated surface using the deterministic Inverse Distance Weighting (IDW) method of methane emissions produced during the month of October 2016, between San Francisco and Los Angeles (California).

4.3.3 Map of Isolines

It can be seen that there is an irregular distribution of the isolines, which is an indicator of variations in methane emissions (Figure 7).

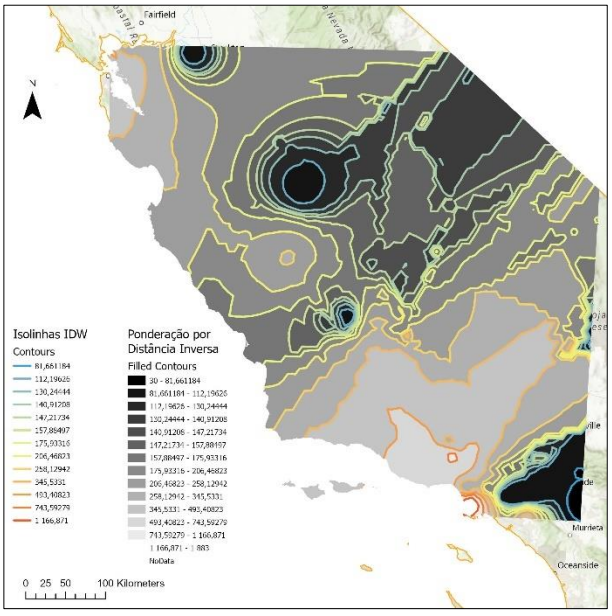


Figure 7 - Isoline map of a surface interpolated using the deterministic Inverse Distance Weighting (IDW) method of methane emissions produced during the month of October 2016, between San Francisco and Los Angeles.

4.3.4 Mobile Windows Statistics

Although the Voronoi maps may be indicative, there are more appropriate methods to investigate the existence of proportional effect, namely the

Statistics of Moving Windows, with and without overlapping neighborhood. The results (Appendix L and Appendix M) indicate that there is a proportional effect in the South zone (Los Angeles area), since the local mean and the local standard deviation are both high. The North zone, next to San Francisco, also seems to indicate a proportional effect, although the effect is more difficult to assess, because it is more subtle.

5. Conclusion

Through this study, it was intended to interpolate a set of methane emission points in California. The results showed the high variability in the emission values of this gas, with a positive asymmetric distribution and a dispersion of the data, reflecting the heterogeneity of the emission sources and the presence of super-emitters (in fact, detected in the study of Duren et al. (2019). Exploratory spatial analysis did not reveal strong spatial trends, and the overall autocorrelation was not statistically significant, reinforcing the random nature of the spatial distribution of methane emissions.

The application of deterministic interpolation methods, specifically IDW and Kernel, was adequate, given the constraints identified, such as the absence of global autocorrelation. The IDW method, although limited by its nature and by the data themselves, allowed a reasonable approximation for the representation of methane emissions, even if the accuracy was compromised by the high local variability and the "bulls eye" effect at some isolated points. The Kernel interpolation showed higher accuracy (lower RMSE), providing a useful alternative model to represent emissions.

Finally, this work illustrates the challenges associated with the interpolation of a complex variable of anthropocentric origin, i.e., that does not necessarily follow the patterns of natural environmental variables (such as rainfall or temperature), where autocorrelation is stronger and more frequent.

References

Costa, A. (2023a). *Tutorial: Exploratory spatial data analysis of the R5D index.* Spatial Statistics, NOVA Information Management School.

Costa, A. (2023b). *Tutorial: Spatial interpolation of the R5D index using Empirical Bayesian Kriging*. Spatial Statistics, NOVA Information Management School.

Costa, A. (2023c). *Tutorial: Spatial interpolation of the R5D index using Inverse Distance Weighting*. Spatial Statistics, NOVA Information Management School.

Duren, R. M., Thorpe, A. K., Foster, K. T., Rafiq, T., Hopkins, F. M., Yadav, V., Bue, B. D., Thompson, D. R., Conley, S., Colombi, N. K., Frankenberg, C., McCubbin, I. B., Eastwood, M. L., Falk, M., Herner, J. D., Croes, B. E., Green, R. O., & Miller, C. E. (2019). California's methane super-emitters. *Nature*, *575*(7781), Article 7781. https://doi.org/10.1038/s41586-019-1720-3

Elser, H., Morello-Frosch, R., Jacobson, A., Pressman, A., Kioumourtzoglou, M.-A., Reimer, R., & Casey, J. A. (2021). Air pollution, methane super-emitters, and oil and gas wells in Northern California: The relationship with migraine headache prevalence and exacerbation. *Environmental Health*, *20*(1), 45. https://doi.org/10.1186/s12940-021-00727-w

Global Methane Pledge. (2021). Global Methane Pledge. https://www.globalmethanepledge.org/

Maasakkers, J. D., Varon, D. J., Elfarsdóttir, A., McKeever, J., Jervis, D., Mahapatra, G., Pandey, S., Lorente, A., Borsdorff, T., Foorthuis, L. R., Schuit, B. J., Tol, P., van Kempen, T. A., van Hees, R., & Aben, I. (2022). Using satellites to uncover large methane emissions from landfills. *Science Advances*, *8*(32), eabn9683. https://doi.org/10.1126/sciadv.abn9683

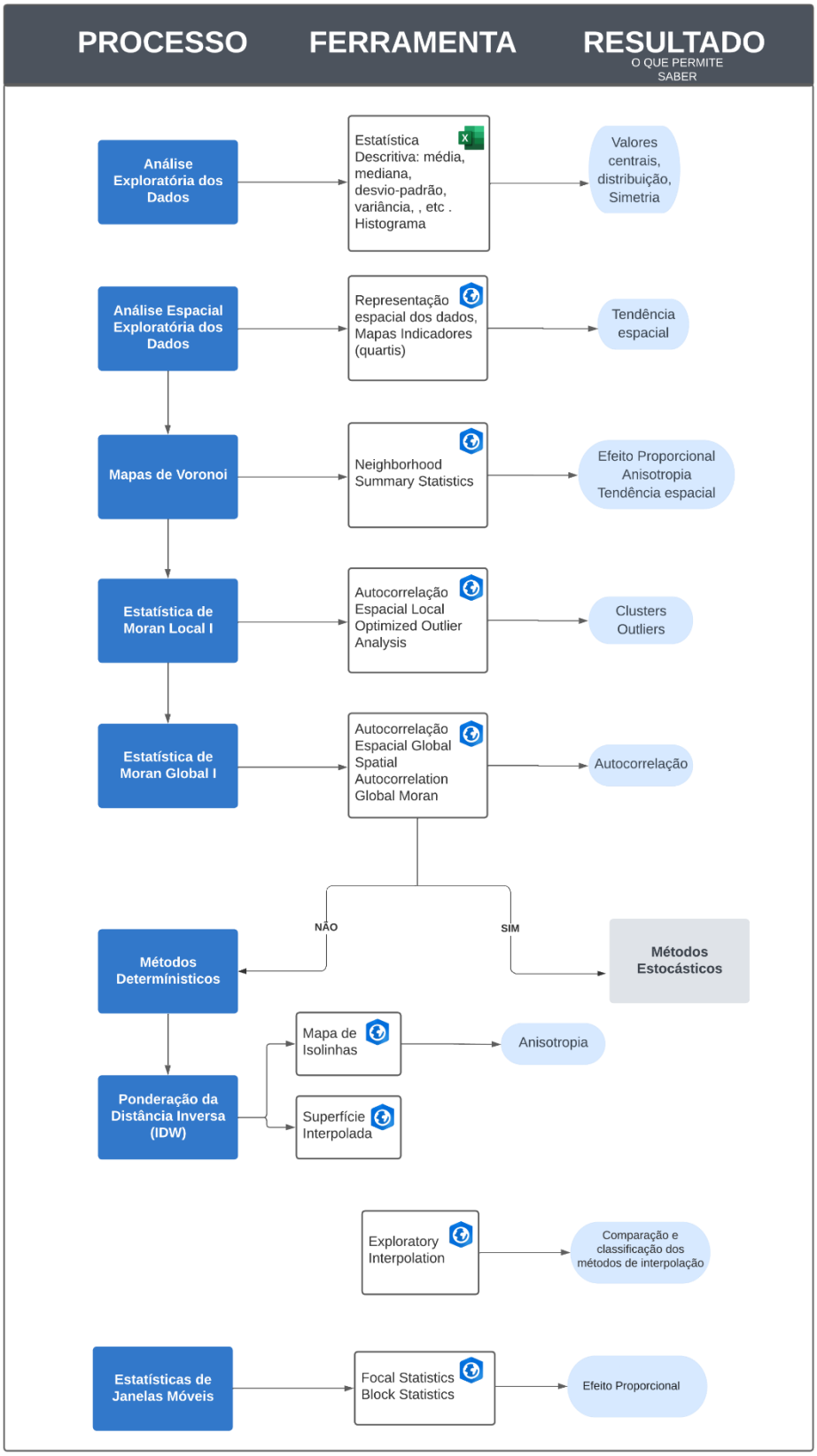
Schuit, B. J., Maasakkers, J. D., Bijl, P., Mahapatra, G., van den Berg, A.-W., Pandey, S., Lorente, A., Borsdorff, T., Houweling, S., Varon, D. J., McKeever, J., Jervis, D., Girard, M., Irakulis-Loitxate, I., Gorroño, J., Guanter, L., Cusworth, D. H., & Aben, I. (2023). Automated detection and monitoring of methane super-emitters using satellite data. *Atmospheric Chemistry and Physics*, *23*(16), 9071–9098. https://doi.org/10.5194/acp-23-9071-2023

Appendix A

Area of Study

Appendix B

Workflow adopted in the context of this work

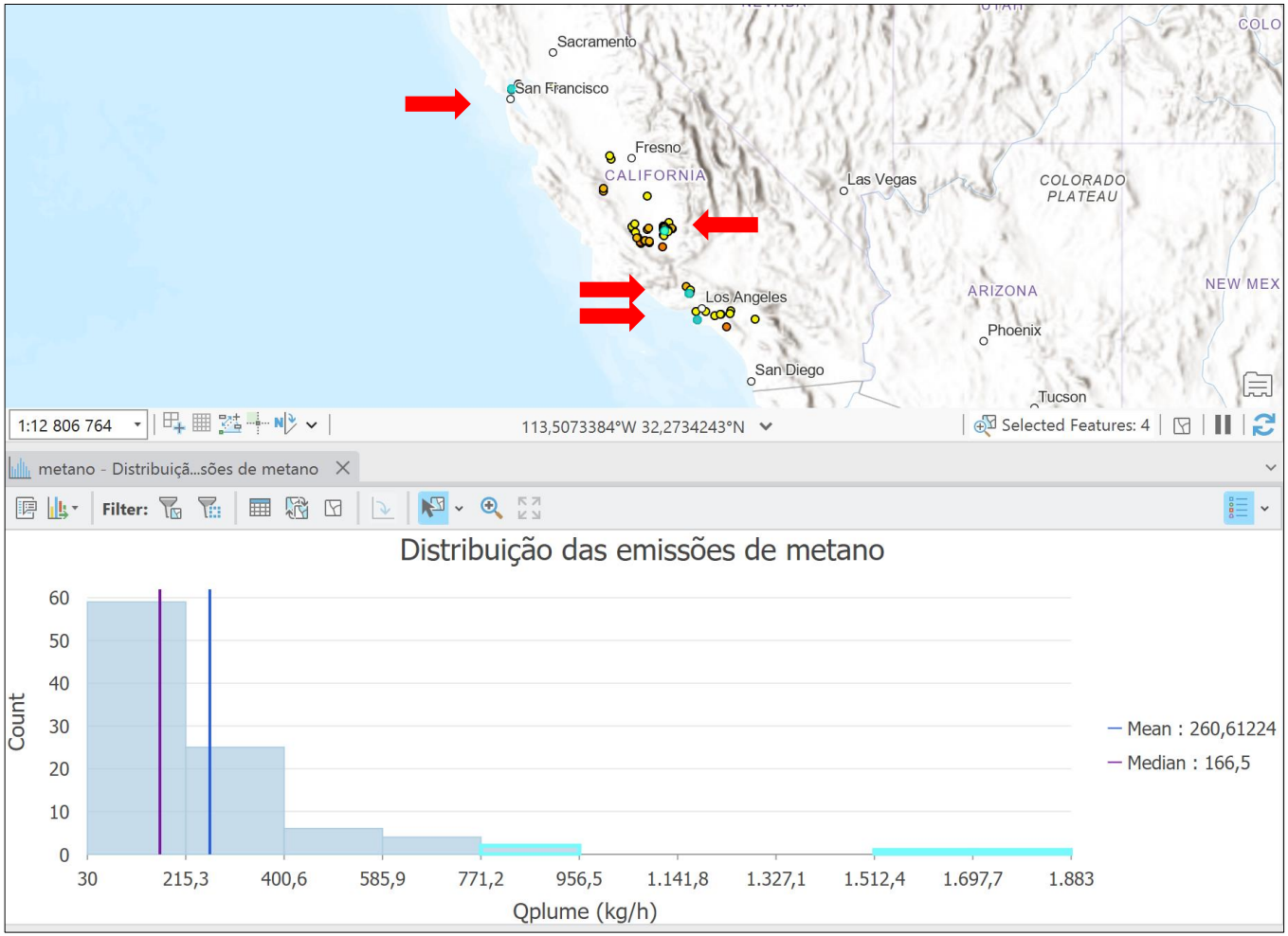


Appendix C

Parameters of the Descriptive Statistics of methane emissions data and respective value, obtained for 98 points, using the XLMiner Analysis Toolpak Excel add-in

Descriptive Statistics Q_{plume} (kg/hr)	**Value**
Average	260,6122
Standard Error	28,1405
Median	166,5
Fashion	139
Standard deviation	278,5768
Variance	77605,02
Kurtose	17,34261
Asymmetry	3,718775
Range	1853
Minimum	30
Maximum	1883
Sum	25540
Count	98
Bigger	1883
Minor	30
Confidence interval (95%)	55,85111

Selection of extreme values (represented by the turquoise blue color and evidenced by the red arrows) using the histogram in ArcGIS Pro.



Appendix E

Determination of quartiles and interquartile range

To determine the quartiles, a box plot chart was selected in ArcGIS Pro, which revealed the following information:

- 1st quartile is 128.75, i.e. 25% of the data are equal to or less than 128.75 kg/h;
- The 2nd quartile is 166.5, i.e., 50% of the data (corresponding to the median, which had already been determined) are equal to or less than 166.5 kg/h;
- The 3rd quartile is 279, i.e. 75% of the data is equal to or less than 279 kg/h;

The interquartile range, which corresponds to the difference between the 3rd and 1st quartile, is 150.25, which means that 50% of the data values have a variability of 150.25 kg/h

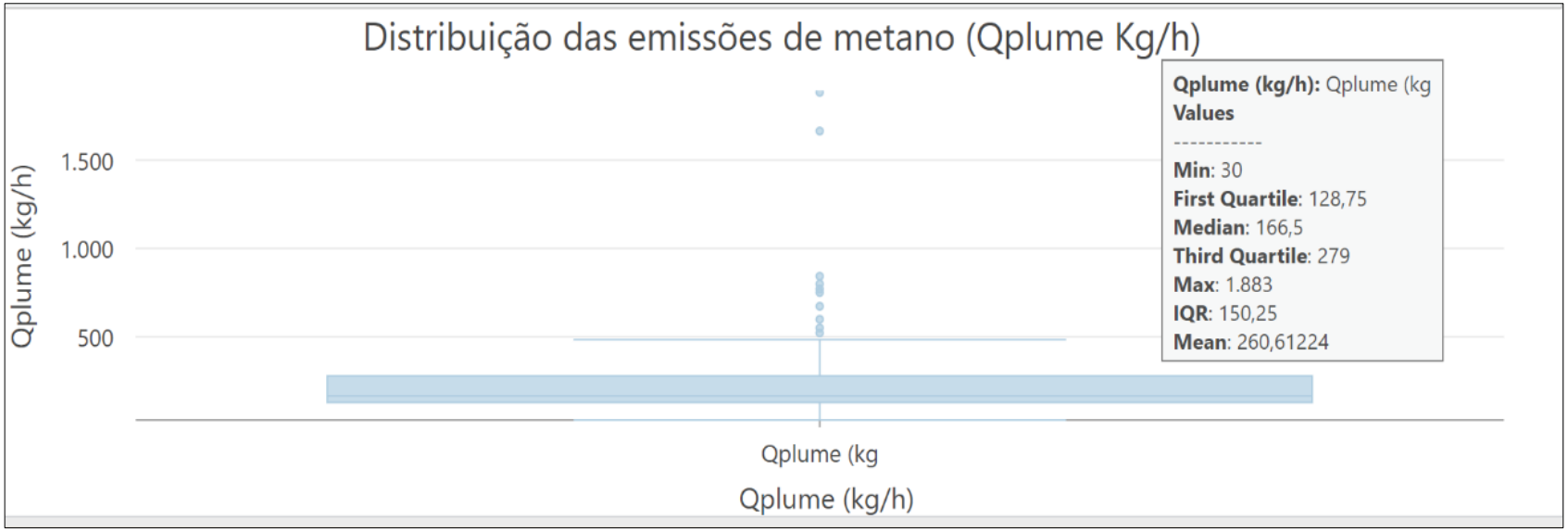
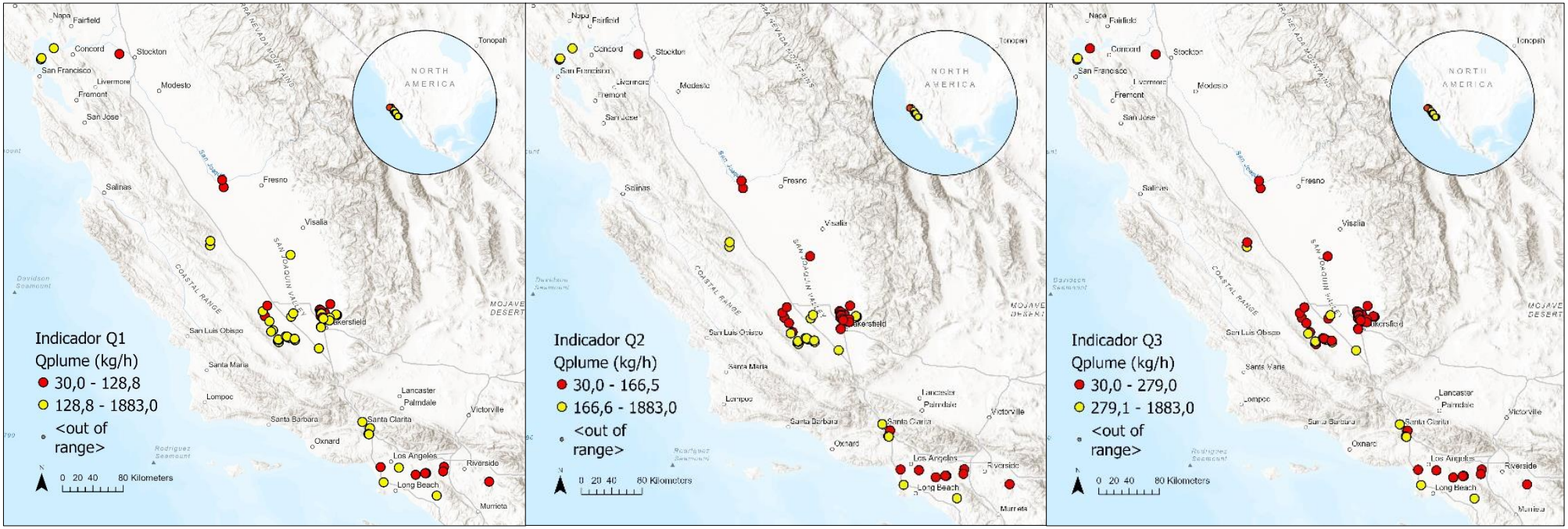


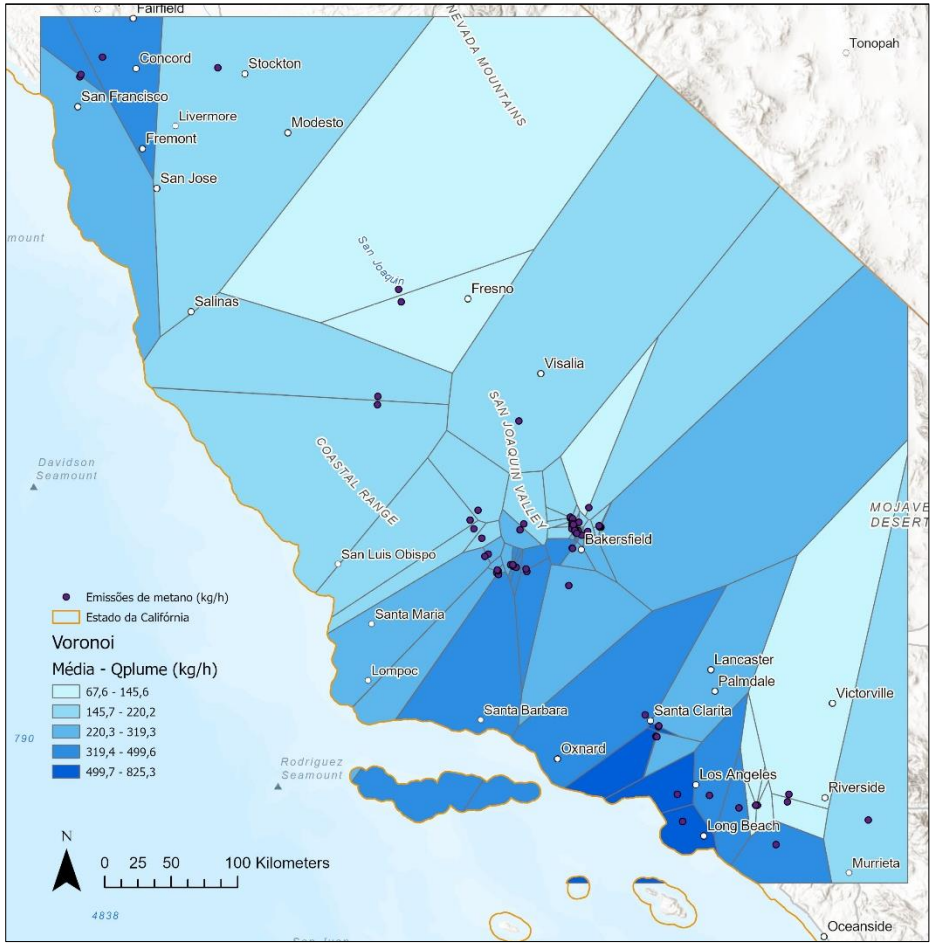
Figure 8 - Distribution of methane emissions, obtained by visualizing a box plot generated in ArcGIS Pro. The considerable number of outliers is also visible in this graphical representation.

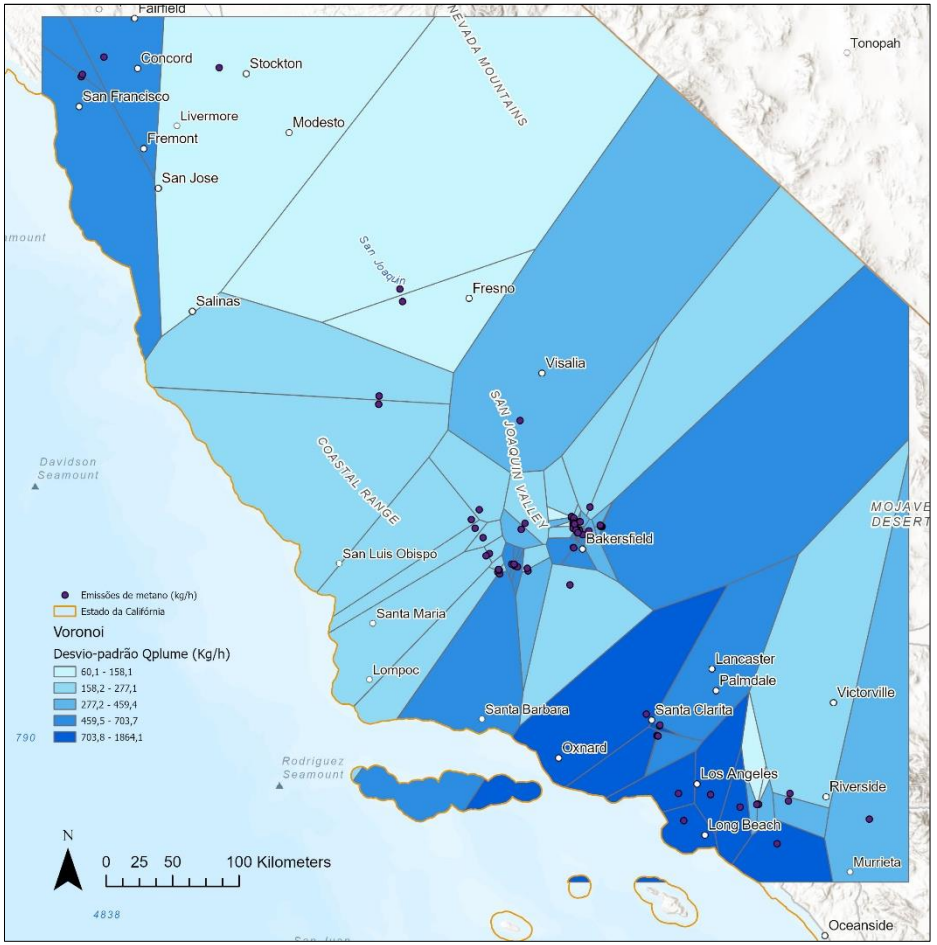
Indicator Maps, which spatially represent the data from the 1st, 2nd and 3rd quartile (from left to right).



Appendix G

Voronoi maps representing the mean (figure left) and standard deviation (figure right) of methane emissions.





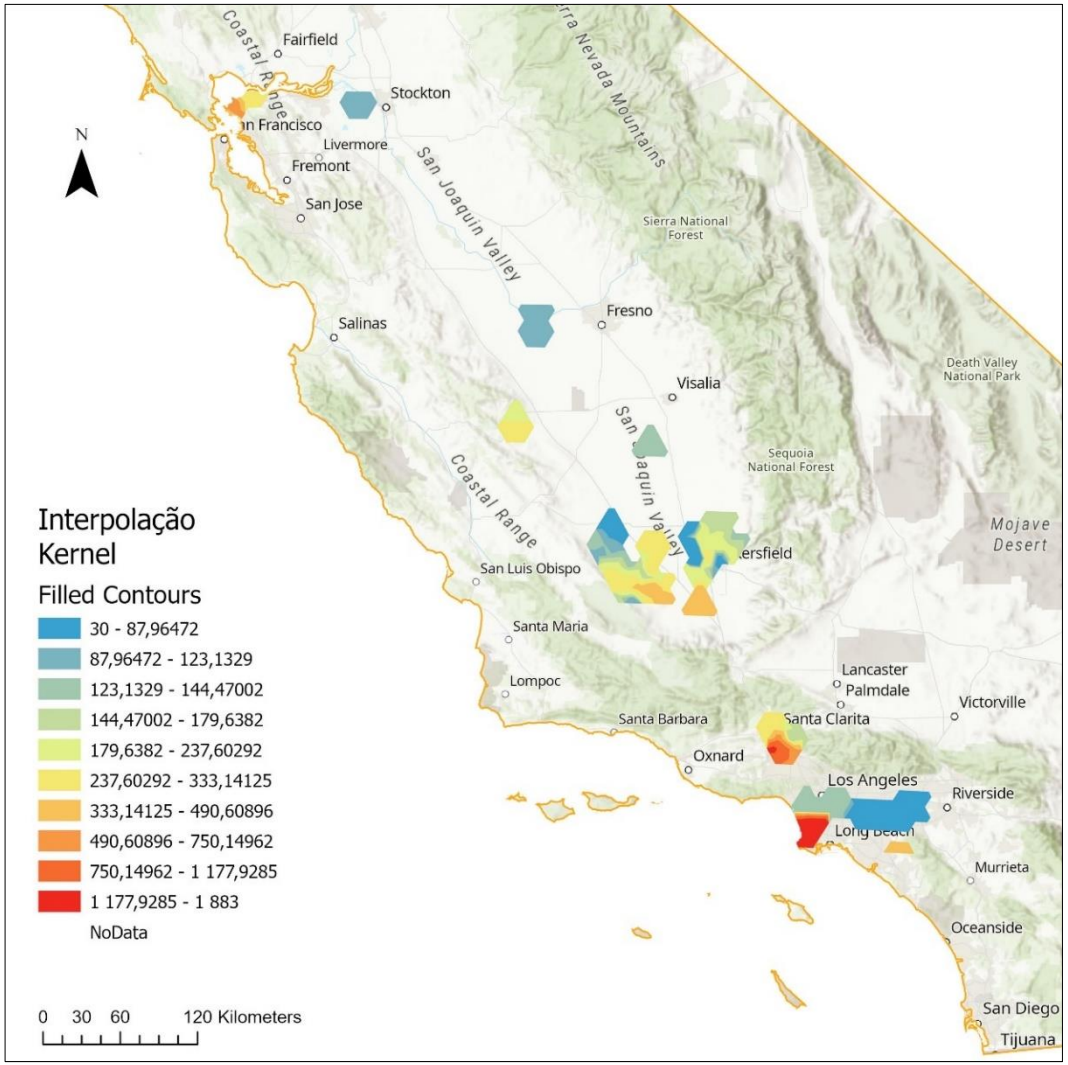
Appendix H

Local Moran statistic for methane emissions (Qplume) over the sampling area.

Classification of deterministic interpolation models obtained through the *Exploratory Interpolation tool* of ArcGIS Pro software.

Model	Rank	RMSE	ME	ME_STD	RMSE_STD	ASE
Kernel Interpolation	1	232,4086	2,998644	0,010133	0,973925	234,6889
Radial Basis Functions - Spline with tension	2	280,3732	3,348737	<Null>	<Null>	<Null>
Radial Basis Functions - Completely regularized spline	2	281,3892	3,357795	<Null>	<Null>	<Null>
Radial Basis Functions - Inverse multiquadric	2	283,7489	3,596595	<Null>	<Null>	<Null>
Inverse Distance Weighted - Optimized	2	287,8446	12,08594	<Null>	<Null>	<Null>
Global Polynomial Interpolation – Second order	2	301,2452	2,258682	<Null>	<Null>	<Null>
Inverse Distance Weighted - Default	2	305,5923	5,012732	<Null>	<Null>	<Null>
Radial Basis Functions - Multiquadric	2	318,3748	-0,97157	<Null>	<Null>	<Null>
Global Polynomial Interpolation – Third order	2	326,9026	-17,9684	<Null>	<Null>	<Null>
Radial Basis Functions - Thin plate spline	2	4288,031	439,4958	<Null>	<Null>	<Null>

Interpolation of methane emissions, using the Kernel model, generated by the *Exploratory Interpolation tool* of the ArcGIS Pro software.

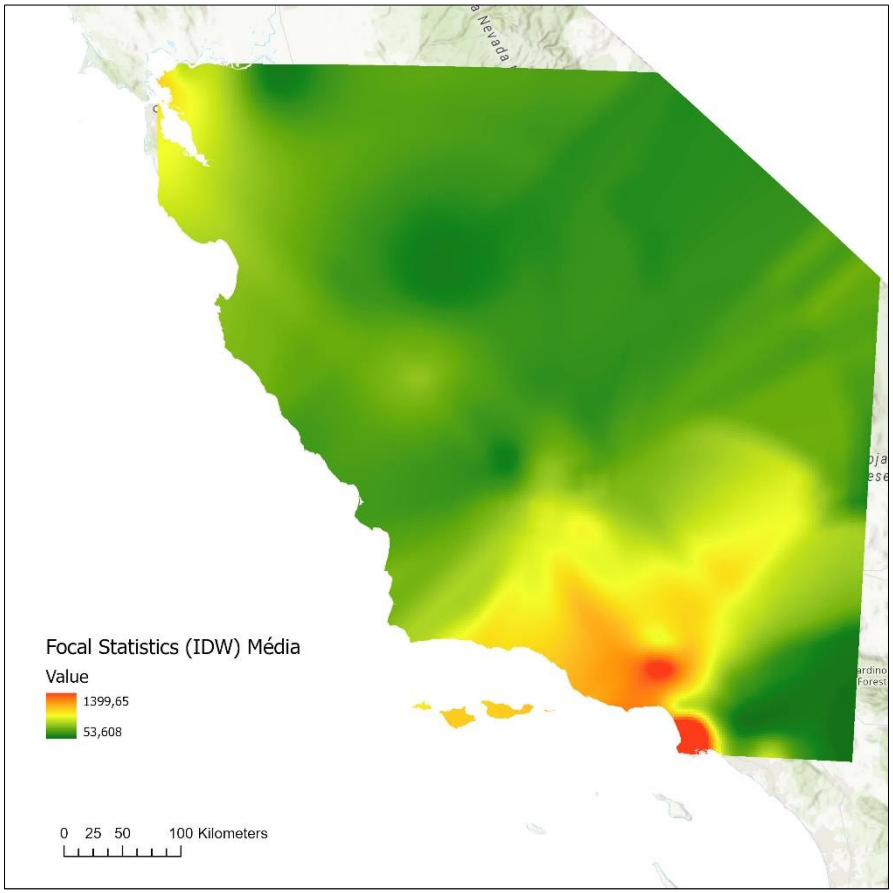


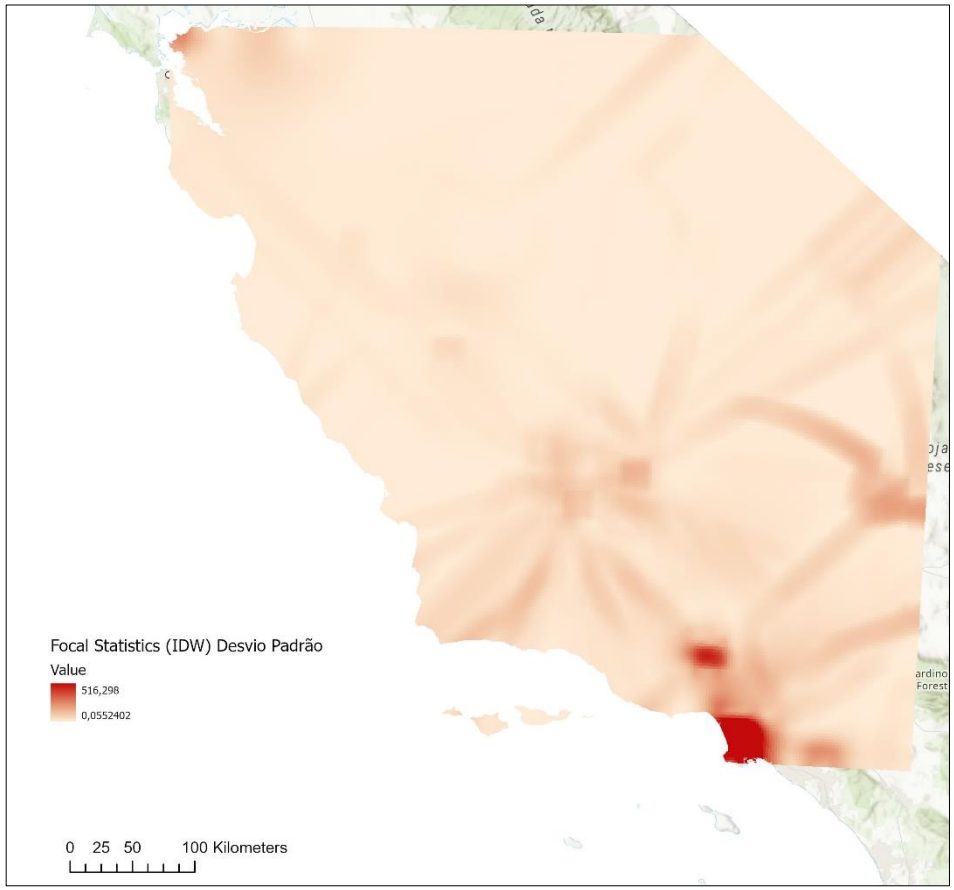
Local Neighborhood Definitions and Error Statistics Generated in IDW Cross-Validation

Type of sector	Max. Neighbors	Min. Neighbors	Average Error	Square Root of Mean Square Error (RMSE)
1	10	5	4,84	307,85
1	20	10	4,7	304,88
4	20	10	4,16	304,48
8	20	10	3,94	304,33
4 w/ angle 45o	6	3	2,86	305,08
4 w/ angle 45o	16	5	4,36	304,54

Appendix L

Representation of the local mean (map on the left) and standard deviation (map on the right), obtained through the Moving Windows statistic with overlay of the IDW forecast map of methane emissions





Representation of the local mean (map on the left) and standard deviation (map on the right), obtained through the Moving Windows statistic without overlapping the IDW forecast map of methane emissions