DOI: https://doi.org/10.5281/zenodo.16943425

Simple linear regression model to evaluate the temperature gradient of Xalapa, Veracruz

Modelo de regresión lineal simple para evaluar el gradiente de temperatura de la ciudad de Xalapa, Veracruz

Juan Andrés Sánchez García[1] *y Marisol Carmona Ortega*[2]

Resumen: Este trabajo presenta un estudio en las islas de calor presentadas en la ciudad de Xalapa y que, basado en la teoría del gradiente de temperatura, se evalúa si la coordenada geográfica de altitud es una variable que determina el valor de la Temperatura Media del territorio. Con ello se tiene como objetivo medir la correlación estadística mediante el Modelo de Regresión Lineal Simple para validar los supuestos y, mediante el estudio de ANOVA, Residuos y la ecuación de Correlación, determinar un modelo que explique esta correlación, bajo la hipótesis a demostrar que entre más alta sea la altura media sobre el nivel del mar, mayor será el aumento de temperatura medida en grados centígrados. Dentro de los hallazgos se determinó que la altitud si impacta en la Temperatura Media pero no es capaz de construir un Modelo de Regresión Lineal que explique y prediga el comportamiento en la ciudad, dado que se encuentran muchos valores por debajo de la recta de normalidad y cuya gran distancia no determina un buen modelo, sin embargo, se concluye que aparecen las variables de longitud y latitud que también pueden aportar a la construcción de un buen modelo predictorio en futuros estudios.

Palabras clave: altitud, temperatura media, modelo de regresión lineal, gradiente de temperatura

Abstract: This work presents a study of the heat islands present in Xalapa City. Based on the temperature gradient theory, it evaluates whether the geographical coordinate of altitude is a variable that determines the value of the average temperature of the territory. The objective is to measure the statistical correlation using the Simple Linear Regression Model to validate the assumptions and, through the study of ANOVA, Residuals and the Correlation equation, determine a model that explains this correlation, under the hypothesis to be demonstrated that the higher the average height above sea level, the greater the increase in temperature measured in degrees Celsius. Among the findings, it is determined that altitude does impact the average temperature but is not capable of constructing a Linear Regression Model that explains and predicts behavior in the city, given that many values are below the normal line and their large distance does not determine a good model. However, it is concluded that the variables

1 Doctor en Arquitectura Diseño y Urbanismo. Universidad Veracruzana. juansanchez@uv.mx
2 Licenciada en Geografía. Universidad Veracruzana. marisolcarmonaortega@hotmail.com

of longitude and latitude also appear to contribute to the construction of a good predictor model in future studies.

Keywords: altitude, average temperature, linear regression model, temperature gradient.

Introduction

The world is immersed in the global warming problem, one of whose main consequences is the rise in temperatures in cities due to the loss of vegetation, leading to the phenomenon of heat islands.

Cities, within their vast territory, are considered heterogeneous surfaces whose temperature value may be due to the altitude of the colonies, based on the temperature gradient, so this chapter explains how a Simple Linear Regression Model determines the causal relationship between altitude and the average temperature of the Xalapa City, In this context, this chapter aims to determine a Regression Model that explains the relationship of the temperature gradient and that can be validated to generate a defined and evaluated Regression equation. This topic remains valid by being able to generate predictive models that help contrast certain areas where the average temperature has risen alarmingly and whose result shows that it is unlikely to be able to determine a single model given the heterogeneity of qualities of Xalapa City, which is why the question arises: What are the variables that can help or contribute to the study of temperatures in a city like Xalapa? Therefore, part of the residue analysis helps obtain spatial information on the conditions that determine the variability in temperatures.

Problematic

Global warming, or the increase in the Earth's average temperature, has been a topic of utmost relevance worldwide since the last decade of the 20th century. It is a phenomenon caused by the emission of greenhouse gases, which trap heat and cause the planet to warm more than it should. This suggests that several causes, such as greenhouse gas emissions, deforestation, and industrial activities, trigger climate change, temperature increases, and ecosystem alterations.

Given this approach, it must be recognized that:

> As cities grow, they replace areas of crops and vegetation with buildings, streets, and other types of infrastructure. These elements reach high temperatures due to their materials, resulting in greater heat inside cities compared to their surrounding areas (where vegetation is common). [conforme las ciudades crecen, reemplazan áreas de cultivos y vegetación por edificaciones, calles y otros tipos de infraestructura. Dichos elementos alcanzan temperaturas altas debido a sus materiales, lo que resulta en mayor calor dentro de las ciudades

> en comparación con sus alrededores (donde comúnmente hay vegetación)] (Lemoine Rodríguez & MacGregor Fors, 2024, párr. 1).
>
> This phenomenon has developed in different areas of the Mexican territory, especially in the Xalapa City which, as a central settlement of a Metropolitan Zone, maintains around 5 ° C more than the cities of its surroundings because it has less green areas in relation to its built area. It is assumed that the amount of heat can decrease by at least 2 ° C if the amount of plant tissue were greater to maintain shaded areas and that, combined with the air that cools by humidifying the environment, the effect of evapotranspiration is generated (Lemoine Rodríguez & MacGregor Fors, 2024).

The amount of vegetation in cities like Xalapa is a parameter for determining temperature increases. However, there is a certain temperature gradient phenomenon that also assumes that temperature decreases in relation to increasing altitude relative to sea level. Being a territory with an altitude ranging from 1,250 to 1,560 meters, it is assumed to be so heterogeneous that the temperature range must be adjusted to the altitude parameter, evaluated as a variable far from constant.

Although this work recognizes that there are temperature variations according to seasons, humidity, strong winds, among other factors, the main question focuses on evaluating and verifying whether the temperature change in the Xalapa City responds to the variation in its altitude relative to sea level in order to present the spatial features that help construct an approximation to the temperature gradient.

Theoretical foundation. Temperature under the behavior of geographical measurements

Today, the challenges posed by climate change research are so significant that they represent a global problem. Temperature increases impact the habitats of all cities due to a series of factors that merit attention in their causal relationships.

> It's important to highlight that a major problem facing most Latin American cities is that their natural capital is not leveraged, much less connected. One of the serious drawbacks of urban greenery is its deteriorating state, whether due to abandonment, pollution, or resource exploitation. [Es importante destacar que un problema importante que presenta la mayoría de las ciudades latinoamericanas, es que el capital natural no se encuentra potenciado y mucho menos se encuentra conectado. Uno de los graves inconvenientes del verde urbano es su estado en deterioro, ya sea por abandono, contaminación o explotación del recurso] (Avid Nava & Winfield Reyes, 2023, pág. 101).

Climate change is a process that establishes phenomena that impact cities, such as more frequent extreme weather events, rising sea levels with the risk of flooding, heat waves, threats to food and water, and increasing inequalities that establish the imperative to manage some climate change actions (Climate Promise, 2024). One of the greatest concerns is the emergence of heat waves in urban areas, increasing the temperature load which, combined with the increase in urbanization, causes cities to begin to concentrate excessive heat within their interiors.

Within cities, the urban climate has been characterized by an increase in temperature, an effect known as a heat island. In other words, it refers to a warmer environment within the city compared to the surrounding areas of the city itself, and this has triggered some phenomena such as torrential rains or power outages. Therefore, the study of the urban climate is a priority to improve the quality of human life (Conde Álvarez & Luyando López, 2021).

Therefore, temperature becomes an interesting concept to study in the face of the challenge of global warming processes. Temperature, as a magnitude that indicates the temperature level of the environment, is a variable used to measure the heat of an object or environment in cities. The variation concentrated in temperature varies according to certain factors within the territory that contribute to the greenhouse effect, such as carbon dioxide, methane, nitrous oxide, and fluorinated gases.

However, other factors are attributed to temperature changes, such as prevailing winds, ocean currents, distance from the sea, latitude, relief, and a very important factor, geographical altitude, which is simply the vertical distance between a point on the surface and sea level. This influences the temperature; as higher altitudes lead to lower temperatures.

This relationship is what, in theory, should or can explain the hot zones in a city, based on a phenomenon called temperature gradient, which is called the “rate of change in air temperature per unit of distance, normally referred to with respect to height” [razón del cambio de la temperatura del aire por unidad de distancia, normalmente referido con respecto a la altura] (eltiempo, 2025, par. 1), with which it is worth mentioning that there is also a negative thermal gradient known as thermal inversion.

From a mathematical perspective, temperature is also an intensive magnitude that takes on different values for each point in space and can be described as a scalar field (which represents the spatial distribution of a scalar magnitude, associating a value with each point in space) by the following equation.

Equation 1. Temperature by scale magnitudes

$$T = T\ (x, y, z)$$

Where x, y, and z are the coordinates of the point in question.

It should be noted that this thermal gradient is calculated by dividing the temperature by the distance between the two points, using the formula:

Equation 2. Thermal gradient

$$GT = \frac{(TB - TA)}{DX}$$

TB: Is the temperature at point B
TA: Is the temperature at point A
DX: Is the distance between points A and B

From a geometric perspective, this temperature gradient represents a rate of temperature increase per unit of depth. Therefore, in theory, the relationship between height and temperature forms a suitable link for predicting heat islands in cities. It should not be overlooked that altitude, along with latitude and longitude, is a parameter that forms part of the geographic coordinates used to locate points on Earth and can determine the determining factors for assessing temperature in a city.

For this reason, the geographic coordinate of altitude is considered an essential element for evaluating or predicting temperature behavior in cities using a statistical procedure that helps determine the magnitude of the relationship between both variables.

Some studies, such as those by Ojeda Misses & González López (2023) and Boyles et al. (2019), have developed models and simulations to interpret temperature behavior, in this example using Newton's law of cooling. Other studies have referred to the measurement of models to minimize error (Wang, Kochan, & Su, 2019), and others have focused on thermoregulation with ecological impacts (Rezende & Bacigalupe, 2015). The truth is that modeling allows for the evaluation of procedures related to temperature under certain phenomena. Therefore, validating the theory depends not only on performing a paired relationship procedure but also on validating the models that help determine relationships and spatial patterns that determine the level of relationship between temperature based on the city's altitude. Therefore, the following methodology is used.

In other studies, such as the one presented by Soto Soto, Garzon Barrero, & Jiménez Cleves (2020), they have also developed correlation models between the LST and the NVDI and suggested that green areas manage to buffer and inhibit excess temperature and heat in the territory. In the work carried out by Lemoine Rodríguez, et. al (2022), it is concluded that in their study, in April 2019, Xalapa is a neotropical green city that shows that the relationship between the size of green spaces and the glass transition temperature in the city, indicates the need for green spaces with minimum areas of 2

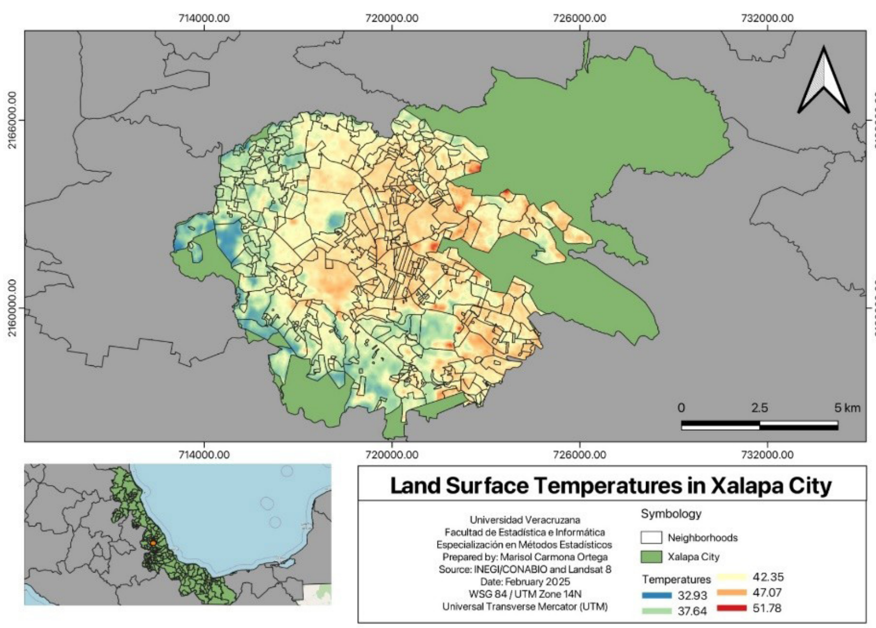
Ha to mitigate heat by ~2 °C, so the latent relationship of temperatures in green areas reflects a significant articulation.

Applied Methodology

Los datos obtenidos para este trabajo fueron obtenidos de the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) [National Commission for the Knowledge and Use of Biodiversity] and the Instituto Nacional de Estadística y Geografía (INEGI) [National Institute of Statistics and Geography]. Data were also obtained through the Landsat 9 satellite image (ID: LC09_L2SP_025046_20240604_20240605_02_T1) acquired on June 4, 2024, which corresponds to Path 025 and Row 046 of the Landsat system.Fue mediante la imagen satelital Landsat 9 (ID: LC09_L2SP_025046_20240604_20240605_02_T1) adquirida el 4 de junio de 2024, que corresponde al Path 025 y Row 046 del sistema Landsat.[1]

For this work, 478 colonies that make up the town of Xalapa, Veracruz, are taken as a study population, assuming that the temperature level is dispersed throughout the urban area, detecting certain spatial patterns as shown in Figure 1.

Figure 1 Temperature map for the Xalapa City.

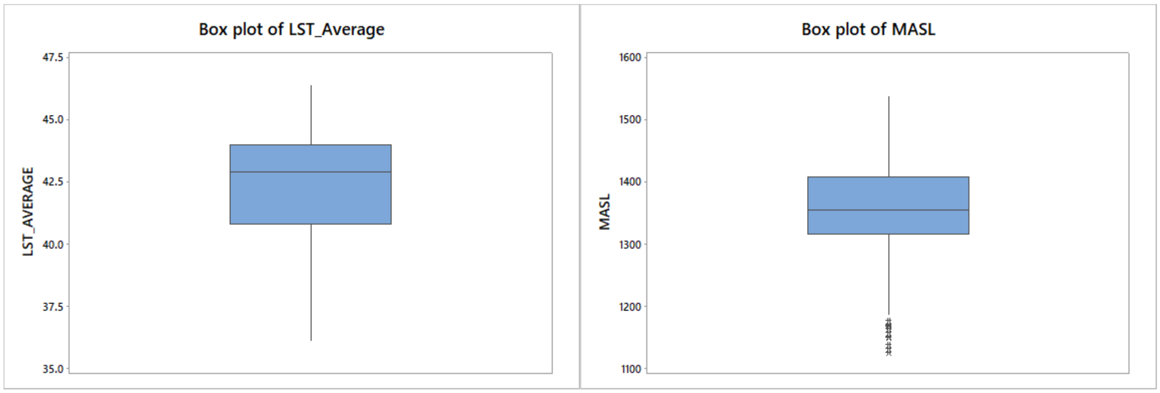


The Xalapa town, divided into colonies, assumes a heterogeneous distribution in the parameter of altitude and average temperature due to its peculiar topography, which represents in a first study the dispersion of data as shown in the following figures.

Figure 2 shows that, when there are atypical data in one of the two variables, it cannot

1 Data obtained from https://earthexplorer.usgs.gov/

Figure 2 Comparison of data dispersion. Average Temperature and Meters Above Sea Level (MASL)



grow in the same proportion when subjected to a paired relationship, however the following table 1 is presented where it is shown that, although the standard deviations are heterogeneous, the coefficient of variation remains stable as a statistical measure that compares the variability of a set of data, it is expressed as a percentage and is calculated by dividing the standard deviation by the arithmetic mean and is generally used to compare the dispersion of data sets with different populations.

Table 1 Descriptive statistics for MASL and average temperature

Variable	N	Media	Standard deviation	Variance	Coefficient of variation	Minimun	Maximum	Range
MASL	478	1357.8	80.1	6413.0	5.90	1126.1	1538	412.2
Average temperature	478	42.325	2.116	4.477	5.00	36.1	46.4	10.3

Therefore, it is recognized that the causal correlation expressed by the thermal gradient theory cannot be determined by a descriptive analysis, since the box graph presents atypical data in the MASL variable and contrasts with the descriptive statistics that show a similar Coefficient of Variation between both variables.

Faced with this dilemma, it is decided to establish the *Pearson Correlation Coefficient,* which is obtained through a *Linear Regression Model* whose objective "is to try to explain the relationship that exists between a dependent variable (response variable) Y and a set of independent variables (explanatory variables) X1,..., Xn."[es tratar de explicar la relación que existe entre una variable dependiente (variable respuesta) Y un conjunto de variables independientes (variables explicativas) X1,..., Xn.] (Carrollo Limeres, 2020, p. 2), so when there are two variables, a *Simple Linear Regression Model* is performed.

The only requirement for using this model is that the variables must be in continuous data. That said, the Simple Linear Regression model is expressed as follows:

Equation 3 Simple Linear Regression Model

$$Y = \beta_0 + \beta_i X_i + e$$

Where Y is the variable to be predicted,[2] β_0 is the intercept or alpha, β_i is the slope or coefficient, X_i is the variable that causes the change in variable Y[3], and e is the average of the residuals.[4] It should be noted that the random component e of this statistical model assumes a normally distributed random variable with mean zero and variance σ^2; that is; $e \sim N(0, \sigma^2)$. This assumption also links the errors to being independent with the same distribution. The general hypothesis under which the model is built is that X contributes significantly to explaining Y.

This regression hypothesis implies the rejection of the null hypothesis, in favor of the postulated alternative (Ojeda Ramírez, 2000), which is as follows:

H_o: $\beta j = 0$ (X_j Doesn't influence)
vs
H_a: $\beta j \neq 0$ (X_j Does influence)

It should be remembered that the null hypothesis (Ho) generally "refers to the statement contrary to the one reached by the researcher (...). It is the hypothesis that the researcher intends to reject" [refiere a la afirmación contraria a la que ha llegado el investigador (…). Es la hipótesis que el investigador pretender rechazar] (Marco San Juan, 2020, p. 1), while the alternative hypothesis (Ha) is "the conclusion that the researcher has reached through his research" [la conclusión a la que el investigador ha llegado a través de su investigación] (Marco San Juan, 2020, p. 1). That said, the value supported by which the assumptions are evaluated is through the p-value, which when less than .05, it can be said that the null hypothesis is rejected as long as we work with a 95% confidence level, otherwise it is not rejected and this will give evidence of the existence of correlation or the appearance of independence (Marco San Juan, 2020), so under this multivariate technique the relationship between the average temperature of the colonies and the altitude in each of them is made.

Results. Linear correlation of temperature under the altitude parameter MASL

Once the simple linear regression model was carried out, the following results were obtained, where it is observed that the R-squared, or coefficient of determination, is 23.73% which, as a statistical measure that evaluates the accuracy of the regression model, shows that there is little correlation between the temperature and altitude of the Xalapa City. Likewise, the adjusted R-squared, which compares models that have

2. Also called dependent, response, predicted or endogenous variable.
3. Also called independent, explanatory, control, predictor, exogenous, stimulus or regressor variable.
4. Also called error, disturbance, random variable, or random deviation.

different predictor numbers, and the predicted R-squared, as a measure that indicates the model's ability to predict the response of new observations, reinforce the fact that it is not a reliable model to be able to predict the construction of a model in the face of new data.

Table 2 Summary of SLRM temperature vs. MASL

Model Summary			
S	**R-Square**	**R-Square (adjusted)**	**R-Square (Predicted)**
1.84982	23.73%	23.57%	23.15%

Figure 3 SLRM graph of average LST vs. MASL

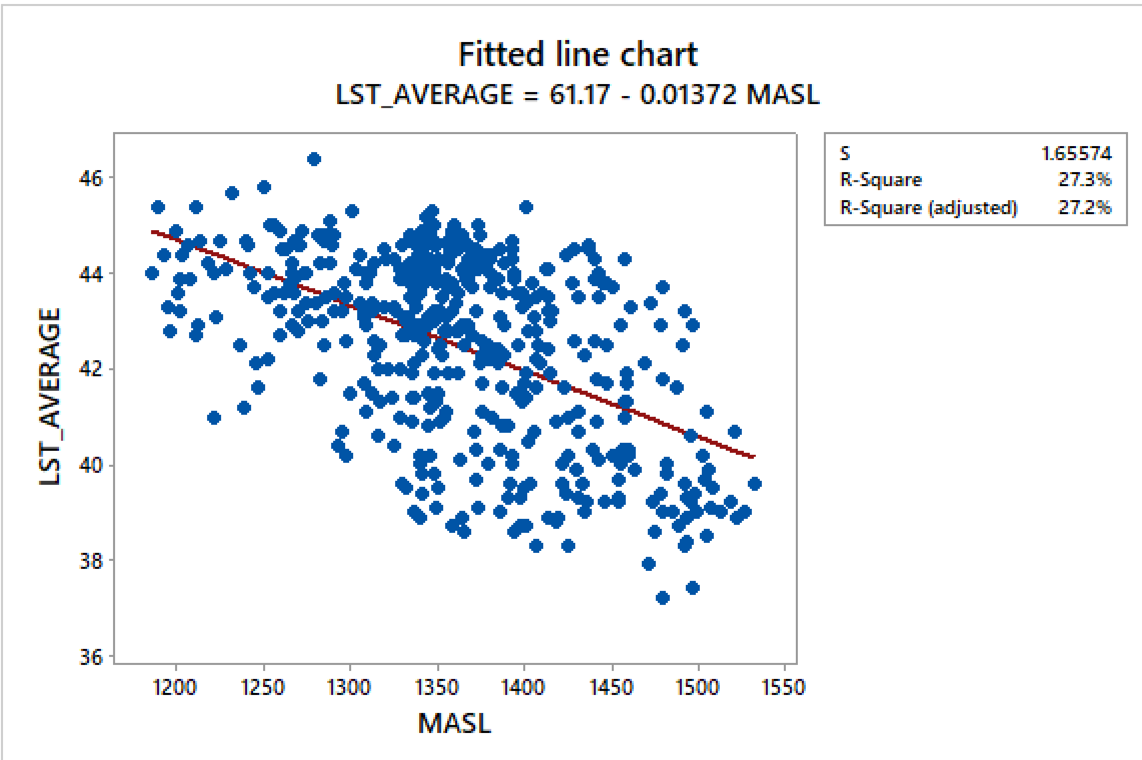


Figure 3 graphically shows a negative correlation. This trend assumes a relationship with the temperature gradient. In observations, the lower the altitude, the higher the average temperature in a city. However, it is observed that within this gradient, there are scattered values that do not approach the normal line.

The values of the model coefficients are also presented. Based on a P value of 0.00, it is concluded that, with a 95% confidence level (1-α) and an α of .05, the null hypothesis is rejected. Therefore, the independent variable MASL influences the temperature value. However, the Coefficient of Determination value explains only 23% of the data, making the model unreliable. This discrepancy is poorly explained using a purely statistical approach.

Table 3. Table of coefficient values

Coefficients					
Term	**Coefficient**	**Coefficient**	**T Value**	**P Value**	**Variance Inflation Factor**
Constant	59.80	1.44	41.57	0.000	
MASL	-0.01287	0.00106	-12.17	0.000	1.00

Finally, the multiple linear regression equation is presented, which represents the model on which the behaviors of the average temperature in Xalapa explained by the altitude can be predicted.

Table 4. Simple linear regression equation valor

Regression equation
LST Promedio = 59.80 - 0.01287 MASL

Likewise, the results of the Analysis of Variance of the model are presented by means of the sum of the adjusted squares, adjusted mean square, and the F Value to determine the equality in the values of the means.

Table 5. Variance Analysis of Average LST vs. MASL

Variance Analysis					
Source	**Degrees of Freedom**	**Adjusted sum of squares**	**Adjusted least squares**	**F Value**	**P Valor**
Regression	1	506.66	506.66	148.07	0.000
MASL	1	506.66	506.66	148.07	0.000
Error	476	1628.00	3.422		
Lack of fit	470	1614.52	3.435	1.44	0.344
Pure error	6	14.27	2.378		
Total	**477**	**2135.45**			

Within the regression analysis, 30 observations were determined that are considered atypical within the average LST vs. MASL model, based on the type of waste that is classified into large waste (R) and uncommon waste (X) as shown in table 6.

Table 6. Analysis of residuals for outliers in the Linear Regression Model*

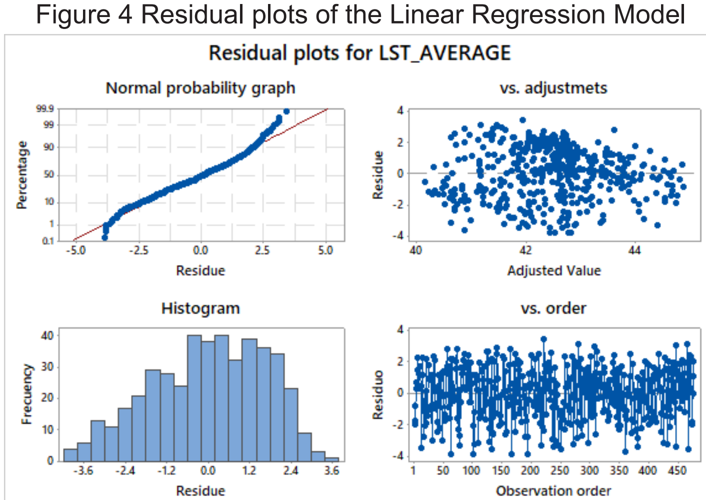
Observation	Average LST	Adjustment	Residue	Standardized residue	Waste type R X
17	44.50	45.222	-0.722	-0.39	X
38	37.60	41.593	-3.993	-2.16	R
49	44.30	45.1453	-0.853	-0.47	X
106	36.10	42.290	-6.190	-3.35	R
117	38.10	42.997	-4.897	-2.65	R
146	42.50	44.821	-2.321	-1.26	X

* R= Large Residue. X= Uncommon Residue

183	36.40	42.428	-6.028	-3.26	R
186	43.20	44.754	-1.554	-0.85	X
209	37.90	42.206	-4.306	-2.33	R
212	38.70	42.688	-3.988	-2.16	R
227	37.60	42.045	-4.445	-2.41	R
247	42.20	44.888	-2.688	-1.46	X
248	44.10	44.756	-0.656	-0.36	X
249	43.50	44.993	-1.493	-0.81	X
254	36.20	40.405	-4.205	-2.28	R
286	38.80	42.570	-3.770	-2.04	R
300	43.90	45.308	-1.408	-0.77	X
322	36.50	41.450	-4.950	-2.68	R
357	39.400	40.003	-0.603	-0.33	X
374	45.10	44.772	0.328	0.18	X
380	41.90	44.654	-2.754	-1.50	X
396	36.20	41.087	-4.887	-2.65	R
414	42.50	44.726	-2.226	-1.21	X
420	37.50	40.004	-2.504	-1.36	X
425	38.50	42.491	-3.991	-2.16	R
429	38.60	42.790	-4.190	-2.27	R
455	36.60	41.226	-4.626	-2.51	R
461	43.40	44.982	-1.582	-0.86	X

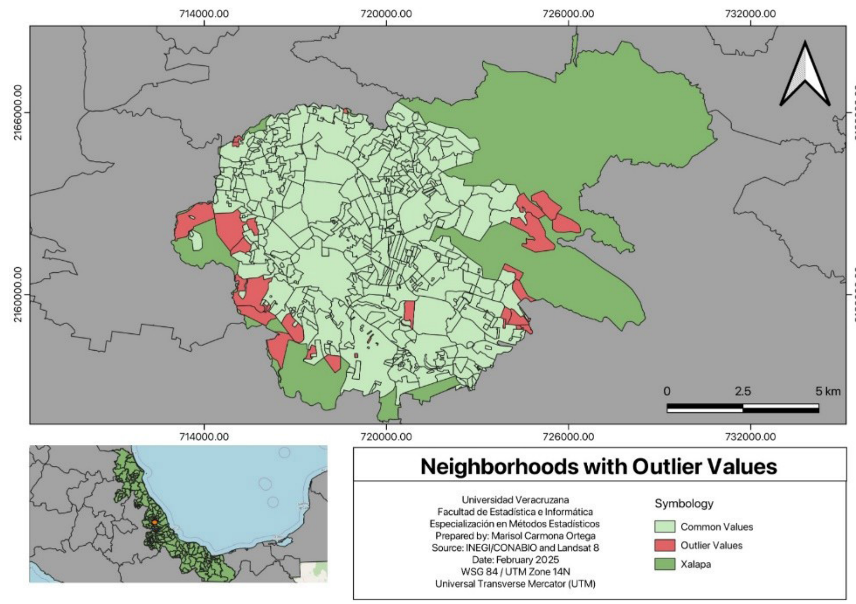
These residues are observed in the residual analysis graphs of the Regression Model under the principles of normality, homoscedasticity and independence, where the values that do not fit the normality of the model are also observed and cause the prediction to not fit a significant value.

Figure 4 Residual plots of the Linear Regression Model



This residual analysis establishes a map showing the 30 observations that differ from normality. Spatially, it is observed that they are located in the peri-urban areas of the urban sprawl. Geographically, these are areas with different characteristics than the central areas due to the amount of vegetation that can influence this arrangement, even though the altitude responds to the theory of temperature gradients.

Figure 5. Map of neighborhoods with outliers



By observing these spatial characteristics, the SLRM is run again, extracting the 30 atypical observations, obtaining the following regression equation and a coefficient of determination value that rose to 27.33%.

Table 7. Simple linear regression equation value

Regression equation
LST Promedio = 61.17- 0.01372 MSNM

Table 8. Summary of SLRM temperature vs. MASL

Model Summary			
S	**R-Square**	**R-Square (adjusted)**	**R-Square (Predicted)**
1.65574	27.33%	27.17%	26.74%

Based on this model, the results obtained by the model with only 448 observations are placed, both from the analysis of variance and from the coefficients of the second model.

Table 9. Analysis of Variance of Average LST vs. MSNM

Variance analysis					
Source	Source	Source	Source	Source	Source
Regression	1	459.88	459.884	167.75	0.000
MASL	1	459.88	459.884	167.75	0.000
Error	446	1222.69	2.741		
Lack of fit	441	1208.83	2.741	0.99	0.590
Pure error	5	13.87	2.773		
Total	**448**	**1682.58**			

Table 10. Table of coefficient values

Coefficients					
Term	**Coefficient**	**Coefficient**	**T Value**	**P Value**	**Variance Inflation Factor**
Constant	61.17	1.44	42.33	0.000	
MASL	-0.01372	0.00106	-12.95	0.000	1.00

Finally, the appearance of 21 new atypical data based on the second linear regression model is observed in the residuals.

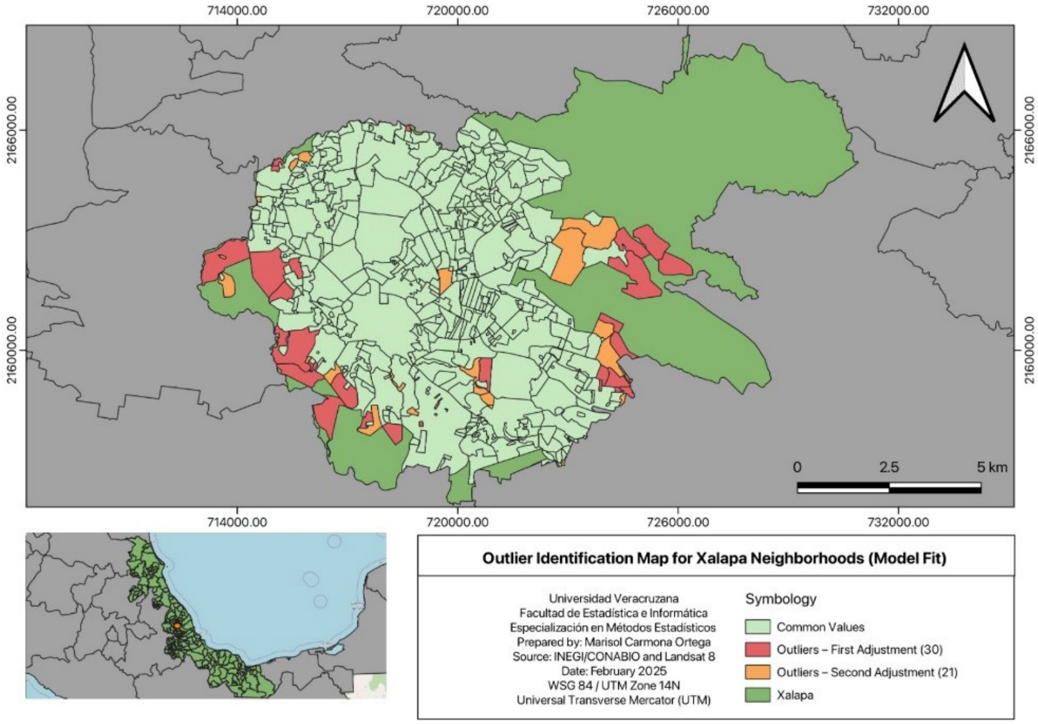
Table 11. Analysis of outliers based on the second regression model

Observation	Average LST	Adjustment	Residue	Standardized Residue	Waste type R X
13	38.600	42.034	-3.434	-2.08	R
59	44.000	44.906	-0.906	-0.55	X
64	38.600	42.441	-3.841	-2.32	R
94	39.600	40.137	-0.537	-0.33	X
102	38.700	42.523	-3.823	-2.31	R
103	39.000	40.215	-1.215	-0.74	X
132	38.300	41.867	-3.567	-2.16	R
167	39.000	42.827	-3.827	-2.31	R
195	38.900	42.790	-3.890	-2.35	R
218	39.600	42.925	-3.325	-2.01	R
222	45.400	41.949	3.451	2.09	R
239	41.000	44.408	-3.408	-2.07	R
252	38.900	42.449	-3.549	-2.15	R
276	44.400	44.818	-0.418	-0.25	X
361	37.200	40.873	-3.673	-2.23	R
381	39.500	42.894	-3.394	-2.05	R
391	45.400	44.866	0.534	0.32	X
398	43.300	44.779	-1.479	-0.90	X
447	39.400	42.764	-3.361	-2.03	R
469	39.100	42.664	-3.564	-2.15	R
474	42.800	44.768	-1.968	-1.20	X

Table 11 shows that the outliers are represented by large, negative residuals, suggesting that they are observations located far below the mean of the group analyzed. This result is mapped again in the town of Xalapa, and it is observed that the majority maintain the spatial pattern of peripheral areas, while 7 values are located within the urban sprawl, corresponding to parks and wooded areas, as shown in figure 6.

Given these results, the MRLS is run once more, extracting the 51 atypical data obtained in the two previous models, showing that the Coefficient of Determination rose to

Figure 6. Outlier identification map for Xalapa neighborhoods (model fit)



29.91% but no longer by the same amount as before, obtaining the following regression equation.

Table 12. Summary of SLRM temperature vs. MASL

Model Summary			
S	R-Square	R-Square (adjusted)	R-Square (predicted)
1.55504	29.91%	29.74%	29.27%

Table 13. Value of the simple linear regression equation

Regression equation
LST Promedio = 61.84- 0.01413 MASL

As can be seen in figure 7, the normal line remains negative and in this third model the atypical data that were below the arithmetic mean disappear, specifically in high areas whose temperature values did not fit the gradient expected in the model.

The ANOVA analysis of the 427 data in the Model is performed again and it is also observed that 15 atypical observations appear based on the model residuals whose indication continues to show large distances with respect to the mean.

Figure 7. Average LST residue graph vs. MASL

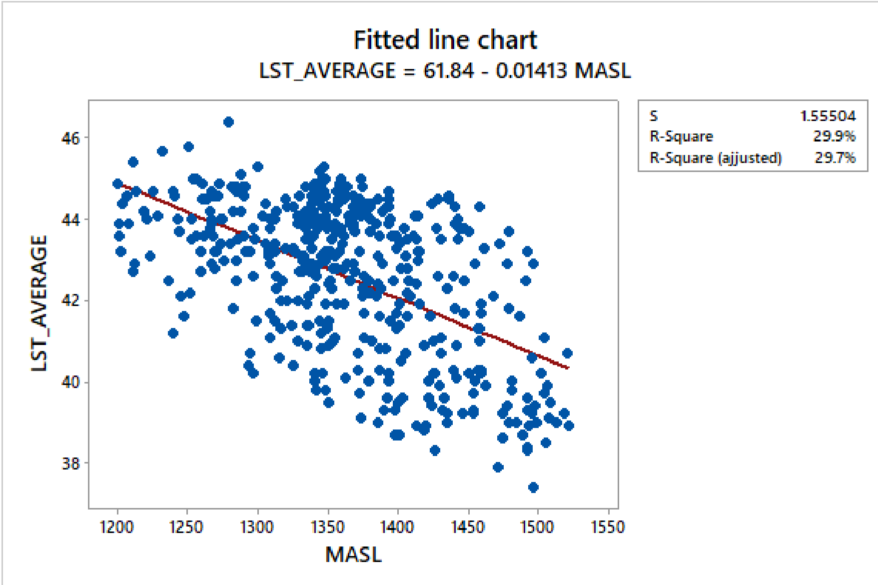


Table 14. Table of coefficient values

Coeficientes					
Term	**Coefficient**	**Coefficient**	**T Value**	**P Value**	**Variance inflation factor**
Constant	61.84	1.43	43.18	0.000	
MASL	-0.01413	0.00105	-13.47	0.000	1.00

Table 15. Analysis of variance for average LST vs. MASL

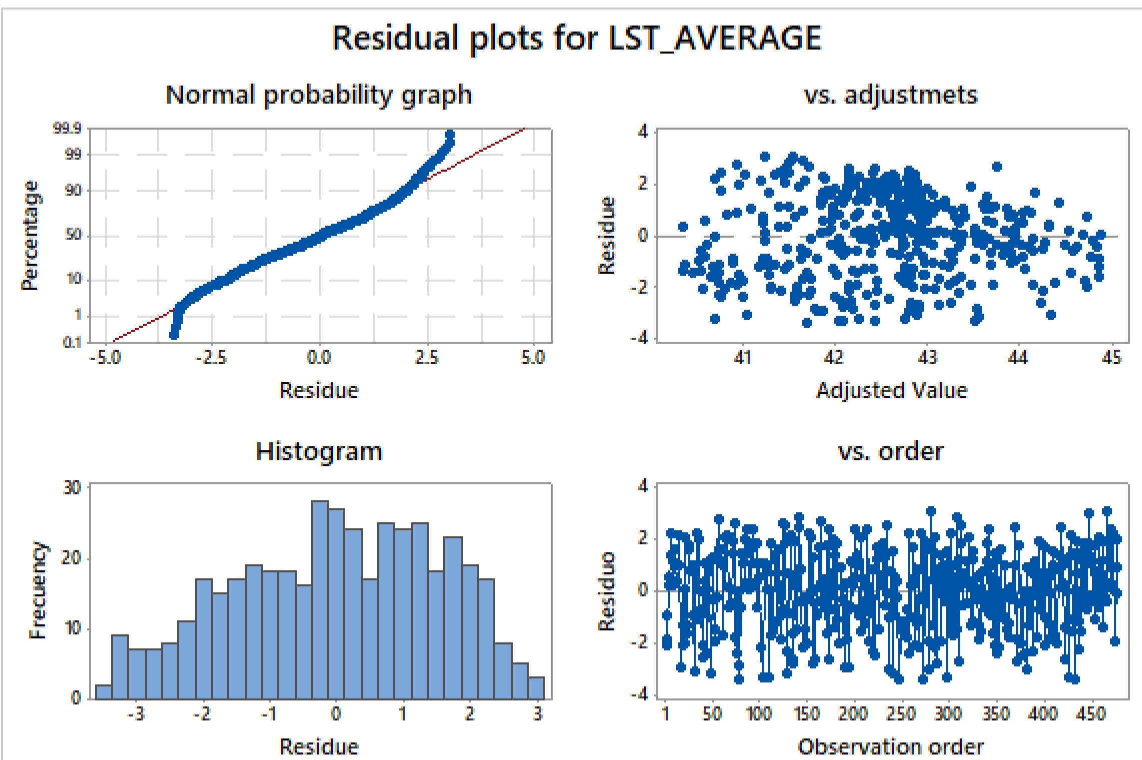
Variance analysis					
Source	**Source**	**Source**	**Source**	**Source**	**Source**
Regression	1	438.50	438.503	181.34	0.000
MASL	1	438.50	438.503	181.34	0.000
Error	425	1027.71	2.418		
Lack of fit	420	1013.84	2.414	0.87	0.666
Pure error	5	13.87	2.773		
Total	427	1466.21			

Table 16. Residual analysis of the 3rd simple linear regression model

Observation	Average LST	Adjustment	Residue	Standardized Residue	Waste type R X
48	40.400	43.570	-3.170	-2.04	R
76	44.900	44.898	0.002	0.00	X
77	38.700	42.087	-3.387	-2.18	R
104	37.400	40.693	-3.293	-2.13	R
110	40.200	43.516	-3.316	-2.14	R
128	37.900	41.051	-3.151	-2.03	R
240	41.200	44.336	-3.136	-2.03	R
246	38.300	41.694	-3.394	-2.19	R
259	43.900	44.870	-0.970	-0.63	X
273	38.700	32.040	-3.340	-2.15	R
298	39.000	42.255	-3.255	-2.10	R
304	43.200	44.858	-1.658	-1.07	X
384	43.600	44.877	-1.277	-2.83	X
426	39.500	42.754	-3.254	-2.10	R
433	39.100	42.436	-3.336	-2.15	R

Finally, the graph of normality, independence and homoscedasticity of the residuals is presented, which shows that the atypical patterns in the model are maintained.

Figure 8. MRLS graph of average LST vs. MASL



Discussions

The Simple Linear Regression Model was only run three times because a pattern was observed in the analysis where data exclusion did not significantly improve the presented model. In the third model, 51 of the 478 data points were omitted, equivalent to 10.61% of the observations, and the Coefficient of Determination increased from 23% to 29%, still showing little correlation in MRLS.

It is observed that the 15 observations from the third model that appeared in the residuals can be further excluded, which would represent 13.8% of the observed data. This could increase the value of the Coefficient of Determination, but it is assumed that it would not increase exponentially, and that, as outliers are continued to be excluded, the model begins to lose significance due to the number of statistical subjects excluded.

It is assumed that the model can be improved by segmenting the strata into specific areas with similar characteristics, and perhaps each model, by altitude stratum, can demonstrate greater significance within each interval of meters above sea level. However, the model's value is considered low due to the dispersion of the data below the arithmetic mean. However, spatial characteristics within the city can contribute to the improvement of the model if the geographic measurements of longitude and latitude are considered to center the neighborhoods within the Xalapa City.

It should be noted that, although the Coefficient of Determination is low, which does not explain a good model, the null hypothesis is rejected, so correlation is assumed. However, this cannot be explained in detail with a Simple Linear Regression equation. This can determine certain patterns within the scatter plot that can be analyzed independently.

Conclusions

Cities have increased their average temperatures as a result of climate change. However, heat stratification is not homogeneous in each neighborhood within the city itself, so temperatures vary depending on different variables. One of these is altitude, which represents a temperature gradient based on the fact that the lower the altitude, the hotter the area, and the higher the altitude, the lower the temperature.

This theoretical approach has been studied in most inhabited areas, but in this case, it shows that, although the behavior tends to occur in the manner mentioned above, it is difficult to systematize the prediction of a model that represents all the thermal qualities of the neighborhoods in the Xalapa City into an equation. The Simple Linear Regression Model, as a statistical tool, is an adequate way to model the behavior of a paired correlation between Average Temperature and Altitude, however, based on its P-Value, ANOVA and Residual Analysis validations, a series of outliers are expressed that make the Coefficient of Determination low and make it difficult to predict a single model.

This work meets the stated objective since it was possible to model, represent and validate the existing elevation between height and temperature, although the proposed hypothesis is rejected since the model also shows little elevation, so in future works other variables will be available that will help define the temperature change in Xalapa. The study of assumptions represents a statistical way of determining the model's conditions; however, two criteria are questioned. The first is that when mapped in a cartogram, conditions are determined that can influence temperature through other geographic coordinates, which can influence the construction of a Multiple Linear Regression Model that helps determine a better expression of the model. The second criterion is to be able to segment the heights of the scatter plot into three strata in order to evaluate each segment of the model and thereby search for microbehaviors that can better define each surface in the study area, thus finding possible spatial patterns that determine causes of data behavior; and this, in turn, can improve regression models in future studies.

References

Avid Nava, E. M., & Winfield Reyes, F. N. (2023). Evaluar la resiliencia climática desde su correlación con las redes verdes urbanas. Aplicación en Zona Coyol, Veracruz, México. E-RUA, 100-111. doi:https://doi.org/10.25009/e-rua.v15i03.198

Boyles, J. C., Levesque, D. L., Nowack, J., Wojciechowski, M. S., Stawski, C., Fuller, A., & Smit, B. (2019). Efficiency and its bounds for thermal engines at maximum power using Newton's law of cooling. Journal Ecology & Evolution, 12020-12025. doi: http://doi.10.1002/ece3.5721

Carrollo Limeres, C. (03 de Mayo de 2020). Departamento de estatistica e investigación operativa. Obtenido de http://eio.usc.es/eipc1/BASE/BASEMASTER/FORMULARIOS-PHP-DPTO/MATERIALES/Mat_50140116_Regr_%20simple_2011_12.pdf

Climate Promise. (06 de Junio de 2024). Las ciudades tienen un rol clave en la lucha contra el cambio climático: he aquí el porqué. Obtenido de Climate Promise: https://climatepromise.undp.org/es/news-and-stories/las-ciudades-tienen-un-rol-clave-en-la-lucha-contra-el-cambio-climatico-he-aqui-el#:~:text=Aumento%20del%20nivel%20del%20mar%20y%20mayor%20riesgo%20de%20inundaciones,ciudades%20en%20todo%20el%20mundo.

Conde Álvarez, C., & Luyando López, E. (06 de Septiembre de 2021). Ciudades y cambio climático. Obtenido de https://www.atmosfera.unam.mx/ciudades-y-cambio-climatico/#:~:text=El%20clima%20urbano%20se%20caracteriza,comparaci%C3%B3n%20con%20los%20alrededores%20rurales.

eltiempo. (02 de febrero de 2025). Obtenido de https://www.eltiempo.es/noticias/meteopedia/gradiente-de-temperatura

Lemoine Rodríguez , R., & MacGregor Fors, I. (12 de abril de 2024). Ciencia y Luz | Los desafíos climáticos de Xalapa. Diario de Xalapa. Recuperado el 28 de Febrero de 2025, de https://oem.com.mx/diariodexalapa/ciencia-y-salud/ciencia-y-luz-los-desafios-climaticos-de-xalapa-13419047

Lemoine Rodriguez, R., Inostroza, L., Falfán, I., & McGregor Fors, I. (2022). Too hot to handle? On the cooling capacity of urban green spaces in a Neotropical Mexican city. Urban Forestry & Urban Greening. doi:10.1016/j.ufug.2022.127633

Marco San Juan, F. J. (28 de abril de 2020). economipedia.com. Obtenido de https://economipedia.com/definiciones/hipotesis-nula.html

Ojeda Misses, M. A., & González López , J. C. (2023). Simulación numérica del control de temperatura para un modelo de parámetros concentrados obtenido de la ley de enfriamiento de Newton. Ongeniería Investigación y tecnología, XXIV, 1-9. doi:https://doi.org/10.22201/fi.25940732e.2023.24.3.020

Ojeda Ramírez, M. M. (2000). Modelación de Regresión. Xalapa Veraruz: Universidad Veracruzana.

Rezende, E., & Bacigalupe, L. (2015). Thermoregulation in endotherms: physiological principles and ecological consequences. Journal of Comparative Physiology

B: Biochemical, Systemic & Environmental Physiology, 709-772. doi:http:// doi.10.1007/ s00360-015-0909-5

Soto Soto, J. E., Garzon Barrero, J., & Jiménez Cleves, G. (2020). Análisis de islas de calor urbano usando imágenes Landsat: caso de estudio Armenia-Colombia 1996-2018. Espacios, 9-23.

Wang, J., Kochan, O., & Su, J. (2019). Information-measuring System to Study the Thermocouple with Controlled Temperature Field. Measurement Science Review, 161-169. doi: https://doi.org/10.2478/msr-2019-0022