International Journal of Design & Nature and Ecodynamics
Vol. 19, No. 4, August, 2024, pp. 1211-1222
Journal homepage: http://iieta.org/journals/ijdne

Comparative Evaluation of the Least Squares Method for Temperature-Based Predictive Models in Tomato Cultivation

Leonardo S. Vinces Llaguno[1*], Fabiola K. Andrade Cedeño[1], Michael L. Vinces Andrade[2], Cerezo S. Badie Anneriz[1], Alison L. Quevedo Heredia[1]

[1] Faculty of Education Sciences, State Technical University of Quevedo, Quevedo 120301, Ecuador
[2] Faculty of Life Sciences and Technology, Eloy Alfaro Secular University of Manabí, Manta 130214, Ecuador

Corresponding Author Email: lvinces@uteq.edu.ec

https://doi.org/10.18280/ijdne.190412

Received: 24 March 2024
Revised: 6 May 2024
Accepted: 14 May 2024
Available online: 28 August 2024

Keywords:
agricultural optimization, least squares, predictive models, temperature, tomato sowing

ABSTRACT

The least squares method has been widely recognized in the statistical and mathematical fields for its ability to fit linear models to observed data sets. This research focuses on the application of this method in constructing predictive temperature models, given its critical influence on tomato planting and production. In this study, the least squares method is used to develop and fit linear temperature prediction models and compares its effectiveness and precision with other predictive models in the agricultural context. Temperature data was collected from the Babahoyo Weather Station in the Province of Los Ríos – Ecuador, a key tomato growing area. Our results indicate that, while the least squares method provides a solid and reliable fit for certain conditions and periods, other models may offer advantages in specific scenarios. However, by integrating predictions based on least squares with agronomic practices, it is possible to generate more informed planting strategies, optimizing tomato yield against thermal variations.

1. INTRODUCTION

In the last decade, the application of temperature predictive models in agriculture has gained increasing interest due to their capability to optimize production. Smith et al. [1] have demonstrated the effectiveness of artificial intelligence models in predicting temperatures in wine-growing regions, achieving remarkable accuracy that facilitates the planning of planting and harvesting, and adapting agricultural practices to climatic fluctuations. Furthermore, Johnson et al. [2] explored the use of neural networks to model temperature variations in cereal crop fields, allowing farmers to manage irrigation and fertilization more efficiently. These advances have shown significant improvement in the capacity to respond to extreme weather events, which is vital for the long-term sustainability of agriculture.

Temperature plays a decisive role in the growth and development of various crops, including tomatoes (Solanum lycopersicum). In the world of agriculture, predicting and understanding temperature trends can be crucial for maximizing production and ensuring the health of the plants. With technological advancement, predictive models have become essential tools in this field, allowing for the anticipation of climate trends with increasing accuracy [3]. In particular, the least squares method, a technique widely recognized in regression analysis [4], promises to be a valuable tool for adjusting these predictive models in agricultural contexts. Although the application of advanced modeling techniques has been widely explored in other fields, such as biodiesel production [5], their adaptation and optimization for agricultural production, and particularly for crops as temperature-sensitive as tomatoes, are still in process. In this sense, tomato production, whether in open fields or greenhouses, is deeply affected by thermal variations [6], reinforcing the need for precise predictive models. This research seeks to address this need and explore how the least squares method can improve our ability to predict and respond to temperature fluctuations in the context of tomato planting. Agricultural scientific research, like in other branches, requires conducting experiments to verify previously established working hypotheses. The development of these experiments brings with it the need to control various effects that influence the subject leading to the establishment and verification of the hypothesis [7].

2. METHODOLOGY

2.1 Least squares

The least squares method is a fundamental statistical technique that seeks to minimize the sum of the squares of the differences between observed values and the values predicted by a mathematical model. This methodology is crucial in the field of applied sciences, particularly for obtaining accurate estimates of unknown parameters in both linear and non-linear models. By fitting the best line or curve that summarizes the relationship between independent and dependent variables, this method provides a solid foundation for statistical inference and prediction. Its application is extensive, ranging from economics to engineering and biology, where models are used to describe complex phenomena and make reliable predictions about future events based on historical data.

It is a numerical analysis technique framed within mathematical optimization, in which, given a set of ordered pairs (independent variable, dependent variable) and a family of functions, it attempts to find the continuous function, within that family, that best approximates the data (a "best fit"), according to the criterion of minimum squared error [8].

In its simplest form, it attempts to minimize the sum of squares of the differences in the ordinates (called residuals) between the points generated by the chosen function and the corresponding values in the data. Specifically, it is called average least squares (LMS) when the number of measured data is 1 and the gradient descent method is used to minimize the squared residual. It is used to find the minimum or maximum of an objective function and is based on the idea of descending or ascending the slope of the function in the direction of the gradient (vector of partial derivatives) until reaching a point where the rate of change is zero.

From a statistical point of view, an implicit requirement for the least squares method to work is that the errors of each measurement are distributed randomly. The Gauss-Markov theorem proves that the least squares estimators are unbiased, and that data sampling does not have to fit, for example, to a normal distribution.

It is also important that the data to be processed are well chosen, to allow visibility into the variables to be resolved (to give more weight to a particular data). Tomato cultivation is an important agricultural activity worldwide, as tomatoes are a valuable source of nutrients and vitamins. However, the growth and production of tomatoes are influenced by environmental factors, especially temperature. To optimize tomato production, it is important to understand how temperature affects their growth and how temperature can be controlled to improve production. In this context, the least squares method is an advanced statistical technique that can be used to analyze and model the behavior of tomatoes in relation to temperature.

2.2 Mathematical models of a data series

Mathematical models of a data series are mathematical tools used to describe and analyze the behavior of a series of data, with the goal of predicting its future behavior [9].

A fundamental methodology in modern science consists of collecting data, organizing it, and then describing it through a mathematical model $y = f(x)$. The data can be the result of some experiment or observation, as well as those that come from sources of global or national organisms. In this context, if a mathematical model is developed to represent real data, it must strive to meet two objectives: simplicity and accuracy.

A convenient strategy for creating a mathematical model of a statistical nature consists of generating an approximation function $f(x)$ that fits the general shape or trend of the data without necessarily matching at all points as shown in Figure 1. It is observed that the curves representing various functions (including from the same family) can pass through the points without coinciding with them. Therefore, it is necessary to establish a method that allows excluding some of them and in this way be able to choose the appropriate fitting function. In this context, there is a method that is often found in the literature as the least squares regression method which is presented below.

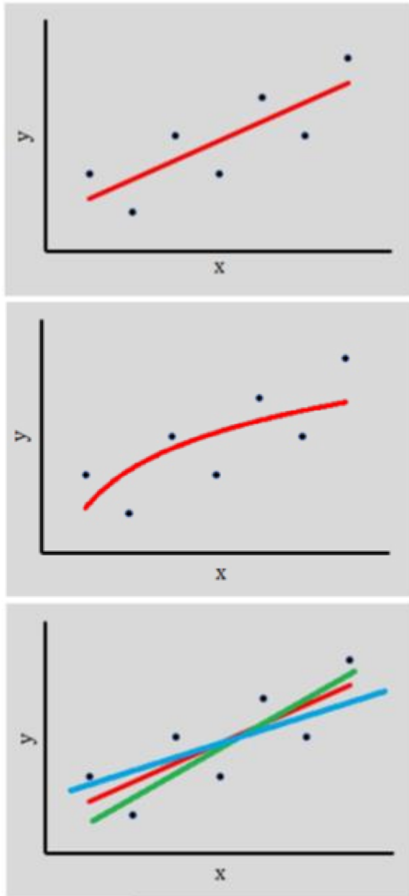
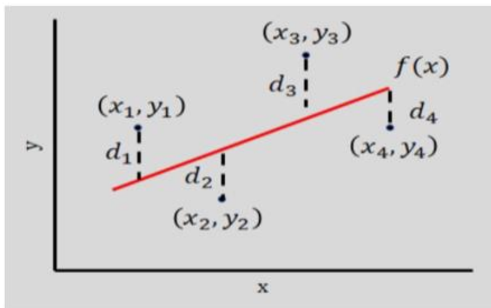


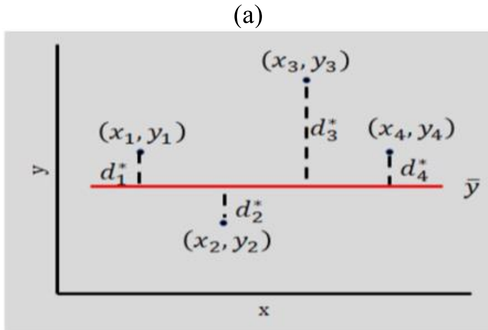
Figure 1. Approximation function

2.3 Mathematical models by least squares

Least squares mathematical models are a technique used in statistics and data analysis to fit a line or curve to a set of data, minimizing the sum of the squares of the differences between the observed values and the values predicted by the model [10].



(a)



(b)

Figure 2. Sum of the squares of the errors and data dispersion

The least squares method is a procedure for fitting a curve to a collection of points that represent a set of data. Consider the scatter diagram of Figure 2(a), which consists of a plot on the x, y plane of some data points $\{(x_i, y_i);\ i = 1, 2, \ldots, N\}$. Consequently, as a measure of how well a model $y = f(x)$ fits the collection of points, it involves summing the squares of the differences between the actual values and the values given by the model to obtain the sum of the squares of the errors or squared errors [11].

$$S_r = \sum_{i=1}^{n} [y_i - f(x_i)]^2 \qquad (1)$$

Graphically, S_r can be interpreted as the sum of the squares of the vertical distances between the graph of f and the given points in the plane. It is noteworthy that before applying this method, the curve that will be adjusted to the set of given points must be chosen.

On the other hand, the efficiency of the fit of the chosen curve (by least squares) is quantified by the coefficient of determination R^2 which is calculated using the following expression [12]:

$$R^2 = \frac{S_t - S_r}{S_t} \qquad (2)$$

Being S_t the measure of data dispersion that is defined as

$$S_t = \sum_{i=1}^{n} [y_i - \bar{y}]^2 \qquad (3)$$

with $\bar{y}$ being an average quantity. Note that S_t is the sum of the squares of the vertical distances between the average value of $\bar{y}$ and the data points as illustrated in Figure 2(b). Therefore, a perfect fit is given by $S_r = 0$ and $R^2 = 1$ which means that the curve explains 100% of the data variability. However, if $R^2 = 0$ and$S_t = S_r$ the fit does not show improvement.

2.4 Least squares fitting of a sine wave

The least squares fitting of a sine wave is a technique used to fit a sinusoidal function to a dataset. This technique seeks to find the optimal values for the parameters of the sinusoidal function so that the sum of the squares of the differences between the observed values and the predicted values is minimized [13].

The sinusoidal regression by least squares is the fitting of a dataset to a trigonometric function of the form

$$f(x) = A_0 + A_1 \cos(wx) + A_2 \sin(wx) \qquad (4)$$

with the parameter w being the frequency. This type of fitting is recommended when the data show a tendency to fluctuate over time in wavy curves. For example, temperatures from one year to the next tend to undulate in a sinusoidal pattern (See in the following session). Sinusoidal waves are also commonly seen in signal processing and time series analysis, see Figure 3(a).

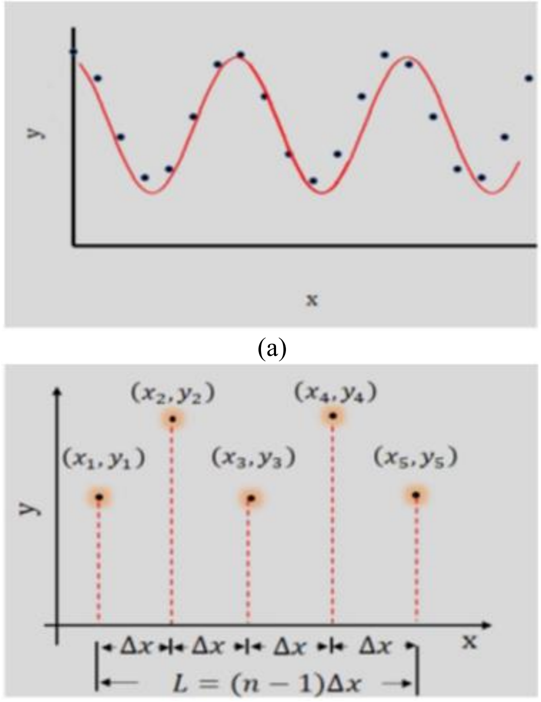


Figure 3. Sinusoidal regression

Now, the problem of finding a sinusoidal curve that optimally fits a set of data is summarized as determining the values A_0, A_1 and A_2 that minimize

$$S_r = \sum_{i=1}^{n} [y_i - A_0 - A_1 \cos(wx_i) - A_2 \sin(wx_i)]^2 \qquad (5)$$

with x_i and y_i as constants while A_0, A_1 and A_2 are the variables (unknowns). Clearly, the equation results from substitution. Indeed, minimizing S_r implies that the partial derivatives of S_r with respect to the variables A_0, A_1 and A_2 must be zeroed out, that is

$$\frac{\partial S_r}{\partial A_0} = -2 \sum_{i=1}^{n} [y_i - A_0 - A_1 \cos(wx_i) - A_2 \sin(wx_i)] = 0 \qquad (6)$$

$$\frac{\partial S_r}{\partial A_1} = -2 \sum_{i=1}^{n} [y_i - A_0 - A_1 \cos(wx_i) - A_2 \sin(wx_i)] \cos(wx_i) = 0 \qquad (7)$$

$$\frac{\partial S_r}{\partial A_2} = -2 \sum_{i=1}^{n} [y_i - A_0 - A_1 \cos(wx_i) - A_2 \sin(wx_i)] \sin(wx_i) = 0 \qquad (8)$$

Given that the summation operates over the elements with index *i* and applying a bit of algebra, the following system of equations (3 × 3) emerges.

$$A_0 n + A_1 \sum_{i=1}^{n} \cos(wx_i) + A_2 \sum_{i=1}^{n} \sin(wx_i) = \sum_{i=1}^{n} y_i \qquad (9)$$

$$A_0\sum_{i=1}^{n}\cos(wx_i)+A_1\sum_{i=1}^{n}cos^2(wx_i) + A_2\sum_{i=1}^{n}\sin(wx_i)\cos(wx_i) = \sum_{i=1}^{n}y_i\cos(wx_i) \quad (10)$$

$$A_0\sum_{i=1}^{n}\sin(wx_i)+A_1\sum_{i=1}^{n}\cos(wx_i)\sin(wx_i) + A_2\sum_{i=1}^{n}sin^2(wx_i) = \sum_{i=1}^{n}y_i\sin(wx_i) \quad (11)$$

Now, for a particular case of points, see Figure 3(b). $(x1, y1), (x2, y2), \ldots, (xn, yn)$ equally spaced as $\Delta x = T/n$ (being T the period) of coordinates on the abscissas $x_i = a + (i-1)\Delta x$ where, for $i = 1, 2, 3, \ldots, n$ the following equalities in the averages are met.

$$\frac{1}{n}\sum_{i=1}^{n}\sin(wx_i)=0 \quad (12)$$

$$\frac{1}{n}\sum_{i=1}^{n}\cos(wx_i)=0 \quad (13)$$

$$\frac{1}{n}\sum_{i=1}^{n}\sin(wx_i)\cos(wx_i)=0 \quad (14)$$

$$\frac{1}{n}\sum_{i=1}^{n}sin^2(wx_i)=\frac{1}{2} \quad (15)$$

$$\frac{1}{n}\sum_{i=1}^{n}cos^2(wx_i)=\frac{1}{2} \quad (16)$$

To demonstrate this, consider the definition of the period $\omega = 2\pi/T$, the Euler formulas $sin(x) = (e^{jx} - e^{-jx})/2j$ and $cos(x) = [e^{jx} + e^{-jx}]/2j$ and the identities $cos^2(x) = [1 + cos(2x)]/2$ and $sin^2(x) = [1 - cos(2x)]/2$. Thus,

$$\frac{1}{n}\sum_{i=1}^{n}\sin(wx_i)=\frac{1}{2j}\sum_{i=1}^{n}[e^{jwx_i} - e^{-jwx_i}] = \frac{1}{2j}\sum_{i=1}^{n}\left[e^{2\pi[a+(i-1)\Delta x]j/T} - e^{-2\pi[a+(i-1)\Delta x]j/T}\right] \quad (17)$$

$$\frac{1}{n}\sum_{i=1}^{n}\cos(wx_i)=\frac{1}{2j}\sum_{i=1}^{n}[e^{jwx_i} + e^{-jwx_i}] = \frac{1}{2j}\sum_{i=1}^{n}\left[e^{2\pi[a+(i-1)\Delta x]j/T} + e^{-2\pi[a+(i-1)\Delta x]j/T}\right] \quad (18)$$

$$\sum_{i=1}^{n}\sin(wx_i)\cos(wx_i) = \frac{1}{(2j)^2}\sum_{i=1}^{n}[e^{jwx_i} - e^{-jwx_i}][e^{jwx_i} + e^{-jwx_i}] = -\frac{1}{4}\sum_{i=1}^{n}\left[e^{4\pi[a+(i-1)\Delta x]j/T} - e^{-4\pi[a+(i-1)\Delta x]j/T}\right] \quad (19)$$

$$\sum_{i=1}^{n}sin^2(wx_i)=\frac{1}{2}\sum_{i=1}^{n}[1-\cos(2wx_i)] = \frac{1}{2}[n-\sum_{i=1}^{n}(e^{2jwx_i} + e^{-2jwx_i})] = \frac{1}{2}[n - \frac{1}{2j}\sum_{i=1}^{n}\left[e^{4\pi[a+(i-1)\Delta x]j/T} + e^{-4\pi[a+(i-1)\Delta x]j/T}\right] \quad (20)$$

$$\sum_{i=1}^{n}cos^2(wx_i)=\frac{1}{2}\sum_{i=1}^{n}[1+\cos(2wx_i)] = \frac{1}{2}[n+\sum_{i=1}^{n}(e^{2jwx_i} + e^{-2jwx_i})] = \frac{1}{2}[n + \frac{1}{2j}\sum_{i=1}^{n}\left[e^{4\pi[a+(i-1)\Delta x]j/T} + e^{-4\pi[a+(i-1)\Delta x]j/T}\right] \quad (21)$$

Given that $\Delta x = T/n$ it follows

$$\sum_{i=1}^{n}\sin(wx_i)=\frac{1}{2j}\sum_{i=1}^{n}\left[e^{2\pi[a+(i-1)\Delta x]j/(n\Delta x)} - e^{-2\pi[a+(i-1)\Delta x]j/(n\Delta x)}\right] = \frac{1}{2j}[e^{2\pi aj/(n\Delta x)}\sum_{i=1}^{n}e^{2\pi j(i-1)/n} - e^{-2\pi aj/(n\Delta x)}\sum_{i=1}^{n}e^{-2\pi j(i-1)/n}] \quad (22)$$

$$\sum_{i=1}^{n}\cos(wx_i)=\frac{1}{2j}\sum_{i=1}^{n}\left[e^{2\pi[a+(i-1)\Delta x]j/(n\Delta x)} + e^{-2\pi[a+(i-1)\Delta x]j/(n\Delta x)}\right] = \frac{1}{2j}[e^{2\pi aj/(n\Delta x)}\sum_{i=1}^{n}e^{2\pi j(i-1)/n} + e^{-2\pi aj/(n\Delta x)}\sum_{i=1}^{n}e^{-2\pi j(i-1)/n}] \quad (23)$$

$$\sum_{i=1}^{n} \sin(wx_i)\cos(wx_i) = -\frac{1}{4}\sum_{i=1}^{n}\left[e^{\frac{4\pi[a+(i-1)\Delta x]j}{n\Delta x}} - e^{-\frac{4\pi[a+(i-1)\Delta x]j}{n\Delta x}}\right] = -\frac{1}{4}[e^{\frac{4\pi aj}{n\Delta x}}\sum_{i=1}^{n} e^{\frac{4\pi j(i-1)}{n}} - e^{-\frac{4\pi aj}{n\Delta x}}\sum_{i=1}^{n} e^{-\frac{4\pi j(i-1)}{n}}] \tag{24}$$

$$\sum_{i=1}^{n} sin^2(wx_i) = \frac{1}{2}[n - \frac{1}{2j}\sum_{i=1}^{n}[e^{4\pi[a+(i-1)\Delta x]j/(n\Delta x)} + e^{-4\pi[a+(i-1)\Delta x]j/(n\Delta x)}] = \frac{1}{2}[n - \frac{1}{2j}[e^{\frac{4\pi aj}{n\Delta x}}\sum_{i=1}^{n} e^{\frac{4\pi j(i-1)}{n}} + e^{-\frac{4\pi aj}{n\Delta x}}\sum_{i=1}^{n} e^{-\frac{4\pi j(i-1)}{n}}] \tag{25}$$

$$\sum_{i=1}^{n} cos^2(wx_i) = \frac{1}{2}[n + \frac{1}{2j}\sum_{i=1}^{n}\left[e^{\frac{4\pi[a+(i-1)\Delta x]j}{n\Delta x}} + e^{-\frac{4\pi[a+(i-1)\Delta x]j}{n\Delta x}}\right] = \frac{1}{2}[n + \frac{1}{2j}[e^{\frac{4\pi aj}{n\Delta x}}\sum_{i=1}^{n} e^{\frac{4\pi j(i-1)}{n}} + e^{-\frac{4\pi aj}{n\Delta x}}\sum_{i=1}^{n} e^{-\frac{4\pi j(i-1)}{n}}] \tag{26}$$

Note that all the summation terms on the right side of the equality have the form of a finite geometric progression whose partial sum is

$$\sum_{i=1}^{n} ar^{i-1} = \frac{a(1-r^n)}{1-r} \tag{27}$$

Consequently, the partial sums of each sum correspond to

$$\sum_{i=1}^{n} \sin(wx_i) = \frac{1}{2j}\left[e^{\frac{2\pi ja}{n\Delta x}}\frac{e^{2\pi j}-1}{e^{2\pi j/n}-1} - e^{2\pi j\left[1-\frac{a}{\Delta x}\right]/n}\frac{1-e^{-2\pi j}}{e^{2\pi j/n}-1}\right] = 0 \tag{28}$$

$$\sum_{i=1}^{n} \cos(wx_i) = \frac{1}{2j}\left[e^{\frac{2\pi ja}{n\Delta x}}\frac{e^{2\pi j}-1}{e^{2\pi j/n}-1} + e^{2\pi j\left[1-\frac{a}{\Delta x}\right]/n}\frac{1-e^{-2\pi j}}{e^{2\pi j/n}-1}\right] = 0 \tag{29}$$

$$\sum_{i=1}^{n} \sin(wx_i)\cos(wx_i) = \frac{1}{4}\left[e^{\frac{4\pi ja}{n\Delta x}}\frac{e^{4\pi j}-1}{e^{\frac{4\pi j}{n}}-1} + e^{\frac{4\pi j\left[1-\frac{a}{\Delta x}\right]}{n}}\frac{1-e^{-4\pi j}}{e^{4\pi j/n}-1}\right] = 0 \tag{30}$$

$$\sum_{i=1}^{n} cos^2(wx_i) = \frac{n}{2} + \frac{1}{4j}\left[e^{\frac{4\pi ja}{n\Delta x}}\frac{e^{4\pi j}-1}{e^{\frac{4\pi j}{n}}-1} + e^{\frac{4\pi j\left[1-\frac{a}{\Delta x}\right]}{n}}\frac{1-e^{-4\pi j}}{e^{\frac{4\pi j}{n}}-1}\right] = \frac{n}{2} \tag{31}$$

$$\sum_{i=1}^{n} sin^2(wx_i) = \frac{n}{2} - \frac{1}{4j}\left[e^{\frac{4\pi ja}{n\Delta x}}\frac{e^{4\pi j}-1}{e^{\frac{4\pi j}{n}}-1} + e^{\frac{4\pi j\left[1-\frac{a}{\Delta x}\right]}{n}}\frac{1-e^{-4\pi j}}{e^{\frac{4\pi j}{n}}-1}\right] = \frac{n}{2} \tag{32}$$

The terms in the numerator that contain the terms $e^{\pm\frac{2\pi j}{n}} - 1$ and $e^{\pm\frac{4\pi j}{n}} - 1$ cancel out when applying the Euler identity $e^{j\varphi} = \cos(\varphi) + jsin(\varphi)$. Indeed, the (3×3) system transforms into

$$nA_0 + (0)A_1 + (0)A_2 = \sum_{i=1}^{n} y_i \tag{33}$$

$$(0)A_0 + \frac{n}{2}A_1 + (0)A_2 = \sum_{i=1}^{n} y_i\cos(wx_i) \tag{34}$$

$$(0)A_0 + (0)A_1 + \frac{n}{2}A_2 = \sum_{i=1}^{n} y_i\sin(wx_i) \tag{35}$$

which, when solved by direct isolation for each of its variables, results in

$$A_0 = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{36}$$

$$A_1 = \frac{2}{n}\sum_{i=1}^{n} y_i\cos(wx_i) \tag{37}$$

$$A_2 = \frac{2}{n}\sum_{i=1}^{n} y_i \sin(wx_i) \tag{38}$$

The values of A_0, A_1 and A_2 are known as Fourier coefficients.

Proceeding in a similar manner, the previous analysis can be extended to the general model.

$$\begin{aligned} f(t) = A_0 + A_1 \cos(wx) & \\ + B_1 \sin(wx) & \\ + A_2 \cos(2wx) + B_2 \sin(2wx) & \\ + \cdots & \\ + A_m \cos(mwx) + B_m \sin(mwx) & \end{aligned} \tag{39}$$

where, for equally spaced data, the coefficients are evaluated with

$$A_0 = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{40}$$

$$A_j = \frac{2}{n}\sum_{i=1}^{n} y_i \cos(jwx_i) \tag{41}$$

$$B_j = \frac{2}{n}\sum_{i=1}^{n} y_i \sin(jwx_i) \tag{42}$$

Being $j = 1, 2, 3 \ldots . m$. Finally, it is highlighted that a simplified and equivalent way to rewrite the equation is

$$f(x) = A_0 + Ccos(wx + \theta) \tag{43}$$

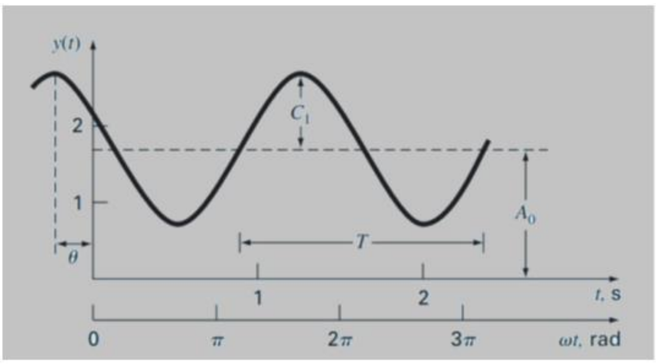
or

$$f(x) = A_0 + Csin(wx + \varphi) \tag{44}$$

where, the parameters C, θ and φ are obtained from the values of A_1 and A_2 as

$$C = \sqrt{A_1^2 + A_2^2} \tag{45}$$

$$\theta = \begin{cases} tan^{-1}\left(-\frac{A_2}{A_1}\right) + \pi & si \quad A_1 < 0 \\ \pi & si \quad A_1 = 0 \\ tan^{-1}\left(-\frac{A_2}{A_1}\right) + \pi & si \quad A_1 > 0 \end{cases} \tag{46}$$



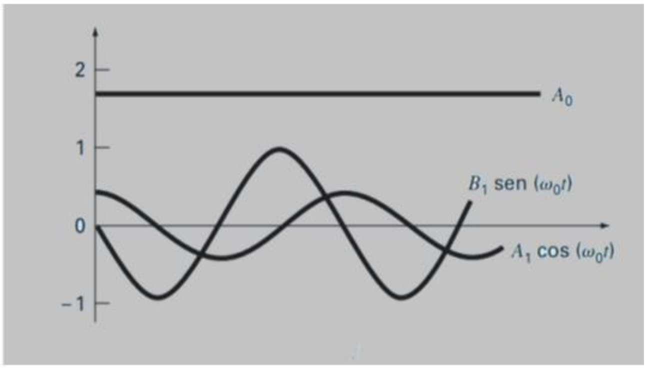


Figure 4. Sinusoidal fit

Indeed, four parameters describe the sine wave, see Figure 4. The average A_0, sets the average height between the abscissas. The amplitude C indicates the height of oscillation. The angular frequency characterizes how frequently the cycles occur, and the phase shift θ parameterizes the extent to which the sine wave is horizontally shifted. The latter can be measured as the distance in radians from $x = 0$ to the point where the sine or cosine function begins a new cycle.

2.5 Weather conditions and periodicity

Weather conditions and periodicity refer to how climatic conditions, such as temperature, humidity, and precipitation, vary over time in a cyclic pattern. These variations can have significant effects on natural systems and human society, such as plant growth patterns, animal migration, cycles of rain and drought, and extreme weather patterns [14].

Atmospheric phenomena also bring a considerable influence over mandatory stages related to crop development and management, too. Solar energy is relevant for all stages of crop development, from seed germination to flowering/fruiting, and its effects can be either positive or negative [15].

The observation of natural conditions, including the weather, has always been a concern for humans to learn or extract relevant information for describing their current state or for making predictions. If atmospheric processes were constant, or strictly periodic, it would be easy to describe them mathematically. However, the atmosphere exhibits variations and fluctuations that are irregular, and to achieve an understanding, the collection and analysis of large sets of meteorological data is carried out [16].

Temperature also experiences variations over time; between day and night, due to the planet's rotation; seasonally, due to its translation and its tilted position. These relative positions of the Earth with respect to the Sun are determinants in the periodicity experienced by temperature values throughout the day and that causes them to be similar on subsequent days. This behavior tends to be modeled through periodic functions such as the trigonometric sine and cosine. This approach is asserted by Plaza [17], who reports that ambient temperature is a physical and cyclic phenomenon, in which its behavior obeys with a good approximation to a sinusoidal wave.

There is a degree of uncertainty in the prior knowledge of climatic conditions, given the complexity of the climate system and the spatial and temporal resolutions that are established. Even so, this state of “unpredictability” or “lack of information” can be reduced with a statistical analysis of historical climatological precipitation records in its location, which would produce relative frequencies of precipitation amounts that would provide substantially more information about tomorrow's precipitation [18].

2.6 Existing relationship between temperature and tomato production

Temperature is a critical factor affecting the growth and production of tomatoes. In general, the optimal temperature for tomato growth is around 21-24°C during the day and around 16-18°C during the night. Higher or lower temperatures can reduce tomato production.

When temperatures are too high, tomato production can decrease, as excess heat can cause the flowers to drop before pollination occurs. In addition, high temperatures can also reduce the quality of the fruits, making them softer, less tasty, and less colorful. On the other hand, when temperatures are too low, they can also reduce tomato production, as plants may stop their growth and development. Low temperatures can also increase the susceptibility of plants to diseases and pests.

2.7 Optimal tomato planting period

The optimal tomato planting period is a specific time of the year when the highest yield and productivity of this crop can be achieved. This period is influenced by various factors, such as climatology, soil characteristics, the type of crop, among others. Temperature is a crucial factor in the tomato planting period, as this crop requires a warm environment to grow and develop properly. Generally, it is recommended to plant tomatoes during periods when the minimum temperature is at least 15°C and the maximum does not exceed 30°C.

Furthermore, solar radiation is another important factor, as tomatoes require an adequate amount of light to grow and produce fruit. Therefore, the geographical position and the time of year must be considered to determine the optimal planting period. Precipitation is also a factor that must be considered in the tomato planting period, as excessive rain can affect germination and plant development, while a lack of rain can limit growth and fruit production.

Crops require the accumulation of certain amounts of temperature degrees for their growth and development. There are various base temperatures for different crops, and each plant has its own base temperature below which it does not grow. Based on these observations, the residual method has been developed, which consists of subtracting the base temperature or vital zero from the average daily temperature of each day. This is called the accumulation of degree days of growth or accumulated heat per day [19].

2.8 Computational tools for model building

Currently, there is a wide range of software programs available that allow for data analysis and automatically provide a mathematical model, such as the exposed sinusoidal model. Among the most notable programs are Matlab, Maple 20, R for statistical analysis, and the Excel spreadsheet, among others. However, the implementation of these programs can have disadvantages, one of them being economic since some software licenses are not free and have a high cost. Additionally, the fact that many of these programs require high computer memory resources, whose equipment (in economic terms) is not accessible to all users, adds to the disadvantages. On the other hand, another drawback is that the software often requires very rigorous programming.

A multitude of optimisation algorithms exists that estimate parameters of dynamic models. A recent comparison found that LSQNONLIN SE (a local gradient-based search algorithm with Latin hypercube restarts) performs best in terms of both accuracy and speed (as measured in the number of function evaluations required to estimate the parameters) [20].

2.9 Type of research

A documentary research type is one that is based on the collection and analysis of information from documentary sources, such as books, journals, theses, reports, among others. This type of research focuses on the study and review of existing documents with the goal of answering specific research questions and delving into the research topic.

Documentary research is very useful in many areas, including social sciences, history, literature, medicine, among others, and can be used both in exploratory studies and in more detailed and specific studies. Researchers can use documentary research to gain a deeper understanding of a topic or problem, identify patterns or trends, and obtain evidence to support their arguments or conclusions.

2.10 Advanced analytics in tomato cultivation

For this study, the data were meticulously selected from the Babahoyo weather station, located in a key region for tomato cultivation, covering an extensive ten-year period. This dataset, comprising over 3,000 daily temperature measurements, captures a broad spectrum of typical and extreme climatic variations, including both daily and seasonal fluctuations as well as rare weather events and climatic anomalies. These characteristics are essential for understanding long-term trends and seasonal variability in tomato production. The statistical analysis employed advanced linear and multiple regression techniques, complemented by cross-validation methods to ensure the robustness and reliability of the comparative models developed. Such rigorous analytical approaches are crucial to verify the consistency and accuracy of predictions, enabling the models to withstand climatic variations without losing precision. The density and diversity of the data make this study particularly valuable for developing agricultural management strategies that can adapt effectively to changing climatic conditions.

In this work, an applied methodology was used, aimed at finding the probable dates of the temperature peaks, to shorten the growth periods of the Tomato crop and reduce its exposure to extreme events.

The inductive-deductive method, supported by abstraction, allowed determining the correct ways to solve the problem and its generalization in an organized method of work. Abstraction was used to understand the scientific problem posed, which allowed delving into its different aspects and establishing relationships with other obtained results.

The historical method was essential for the study of the historical development of the problems surrounding the theory of the sinusoidal model and its current state.

The logical method, supported by historical study, made it possible to investigate the general and essential laws of the functioning and development of the phenomenon studied.

In addition to the least squares method, our study evaluated several predictive models for a comprehensive comparison. Polynomial regression models were included, which are particularly useful for capturing nonlinear relationships between variables. Furthermore, machine learning-based models, such as neural networks, were implemented. These are capable of modeling complex and nonlinear interactions between climatic variables and their impact on agriculture. These machine learning models provide a powerful and flexible tool for predicting climate variations with significantly increased accuracy, which is crucial for efficient agricultural planning and adaptation to changing climate conditions.

3. RESULTS

3.1 Temperature data from the city of Babahoyo

For the development of this research, the annual average temperatures taken from the database of the meteorological station of Babahoyo, province of Los Ríos - Ecuador, are used as the data population.

3.2 Annual temperature behavior models

A sinusoidal model of the form presented in the equation is constructed. To facilitate the calculation, Excel spreadsheet is used to build the numerical matrix that shows shown in the Data (https://cutt.ly/A7UaHF7) the annual average temperature in degrees Celsius (°C) for the period from 1986 to 2021 along with their respective results that were reflected to form the sinusoidal equation.

The time interval for each year corresponds to months (interval of every 30 days, with $t = 1$ corresponding to January). On the other hand, the table shows the average of the annual mean temperature in each month of the year.

Indeed, for this data set, a sinusoidal fit model of the form presented in the equation for $j = 1, 2, 3$ is constructed with the purpose of comparing and choosing the most appropriate model. Due to the large amount of data, an Excel spreadsheet is used to determine the sums that allow obtaining the Fourier coefficients, as well as the determination coefficient to quantify the goodness of fit. From these results, the sinusoidal temperature model for the city of Babahoyo is constructed. Note that the number of data $n = 348$, $\Delta x = 1$ and the frequency values are are $\omega = 2\pi/12 = \pi/6$, $\omega = 4\pi/12 = \pi/3$ and $\omega = 6\pi/12 = \pi/2$, thus, the values of the Fourier coefficients given for the data are

$$A_0 = \frac{1}{348}\sum_{i=1}^{348} y_i = 24.624 \tag{47}$$

$$A_1 = \frac{1}{174}\sum_{i=1}^{348} y_i \cos(\pi x_i/6) = 0.214 \tag{48}$$

$$B_1 = \frac{1}{174}\sum_{i=1}^{348} y_i \sin(\pi x_i/6) = 1.401 \tag{49}$$

$$A_2 = \frac{1}{174}\sum_{i=1}^{348} y_i \cos(\pi x_i/3) = -0.155 \tag{50}$$

$$B_2 = \frac{1}{174}\sum_{i=1}^{348} y_i \sin(\pi x_i/6) = 1.401 \tag{51}$$

$$A_3 = \frac{1}{174}\sum_{i=1}^{348} y_i \cos(\pi x_i/2) = -0.047 \tag{52}$$

$$B_3 = \frac{1}{174}\sum_{i=1}^{348} y_i \sin(\pi x_i/2) = 0.291 \tag{53}$$

The summation terms to determine these coefficients are shown in the last row of the numerical matrix given in the Data (https://cutt.ly/A7UaHF7). When substituting $A_0, A_1, B_1, A_2, B_2, A_3$ and B_3 into the equation, it is found that the models for the temperature in the Babahoyo region are given by:

MODEL 1 ($j = 1$)

$$f(x) = 24.624 + 0.214cos\left(\frac{\pi x_i}{6}\right) + 1.401sin\left(\frac{\pi x_i}{6}\right) \tag{54}$$

with $R^2 = 0.874$

MODEL 2 ($j = 2$)

$$f(x) = 24.624 + 0.214cos((\pi x_i)/6) + 1.401sin((\pi x_i)/6) - 0.155cos((\pi x_i)/3) - 0.293sin((\pi x_i)/3) \tag{55}$$

with $R^2 = 0.950$

MODEL 3 ($j = 3$)

$$f(x) = 24.624 + 0.214cos\left(\frac{\pi x_i}{6}\right) + 1.401sin\left(\frac{\pi x_i}{6}\right) - 0.155cos\left(\frac{\pi x_i}{3}\right) - 0.293sin\left(\frac{\pi x_i}{3}\right) - 0.047cos\left(\frac{\pi x_i}{2}\right) + 0.291sin\left(\frac{\pi x_i}{2}\right) \tag{56}$$

with $R^2 = 0.985$

The sample data and each model are shown in Figure 5. It is observed that model 3 explains 95.9% of the data variability, being the one that provides a better fit.

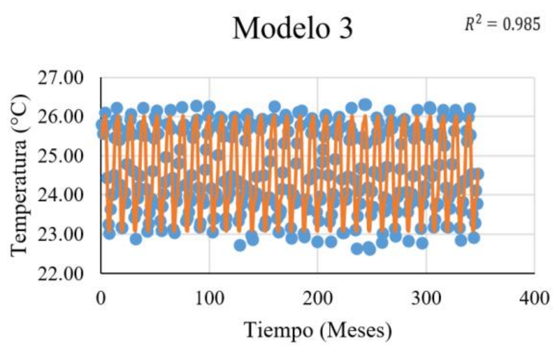


Figure 5. Sinusoidal model 3 of the city of Babahoyo

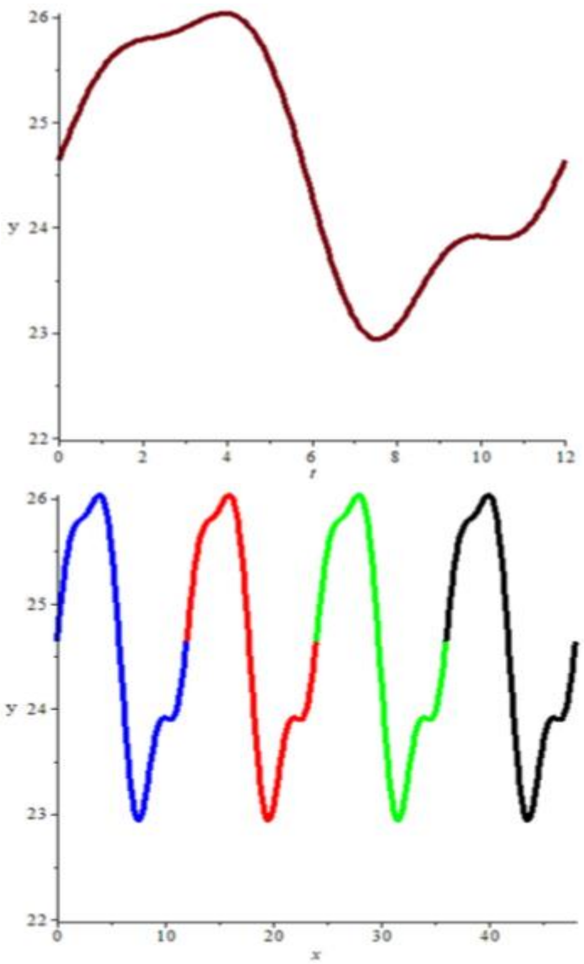


Figure 6. Oscillatory pattern of the sinusoidal function

According to the graphical representation of the model, the oscillatory pattern of the sinusoidal function is similar in the time Interval $\Delta t = 12$, see Figure 6.

The temperature reaches a global maximum and a global minimum throughout the year. To determine when the temperature is optimized, the points at which $f'(x) = 0$ are calculated. Given that

$$f(x) = 24.624 + 0.214cos\left(\frac{\pi x_i}{6}\right) + 1.401sin\left(\frac{\pi x_i}{6}\right) - 0.155cos\left(\frac{\pi x_i}{3}\right) - 0.293sin\left(\frac{\pi x_i}{3}\right) - 0.047cos\left(\frac{\pi x_i}{2}\right) + 0.291sin\left(\frac{\pi x_i}{2}\right) \quad (57)$$

It is found that

$$f'(x) = -0.112sin\left(\frac{\pi x_i}{6}\right) + 0.734cos\left(\frac{\pi x_i}{6}\right) + 0.162sin\left(\frac{\pi x_i}{3}\right) - 0.307cos\left(\frac{\pi x_i}{3}\right) + 0.074sin\left(\frac{\pi x_i}{2}\right) + 0.457cos\left(\frac{\pi x_i}{2}\right) = 0 \quad (58)$$

Because the equation presented is nonlinear, solving for x is not trivial when using algebraic methods. That is why numerical methods are resorted to, one to consider is the Newton's method, whose formulation is given by

$$x_{i+1} = x_i - \frac{f'(x)}{f''(x)} \quad i = 0,1,2,3,\ldots \quad f''(x) \neq 0 \quad (59)$$

The method requires an initial value x_0 (taken arbitrarily), as well as the first and second derivatives of $f(x)$. Indeed, given that the second derivative is given by

$$f''(x) = -0.056cos\left(\frac{\pi x_i}{6}\right) - 0.384sin\left(\frac{\pi x_i}{6}\right) + 0.170cos\left(\frac{\pi x_i}{3}\right) + 0.321sin\left(\frac{\pi x_i}{3}\right) + 0.117cos\left(\frac{\pi x_i}{2}\right) - 0.718sin\left(\frac{\pi x_i}{2}\right) \quad (60)$$

According to the graph of model 3 (Figure 5), initial values are taken for the maximum point $x_0 = 4$ and for the minimum point $x_0 = 7$, when carrying out the iteration process starting from $i = 0$ it follows

$$x_1 = 4 - \frac{f'(4)}{f''(4)} = 4.015 \qquad E_a\% = 0.367 \quad (61)$$

$$x_1 = 7 - \frac{f'(7)}{f''(7)} = 7.563 \qquad E_a\% = 7.441 \quad (62)$$

The approximate errors are calculated by subtracting the absolute value of the current approximation from the previous approximation, then the result is divided by the current approximation and multiplied by 100%. Now, for $i = 1$ it is found that $x_1 = 4.015$ for the maximum and $x_1 = 7.563$ for the minimum, so

$$x_1 = 4.015 - \frac{f'(4.015)}{f''(4.015)} = 4.010 \qquad E_a\% = 0.127 \quad (63)$$

$$x_1 = 7.563 - \frac{f'(7.563)}{f''(7.563)} = 7.495 \qquad E_a\% = 0.900 \quad (64)$$

continuing for $i = 2$, we have $x_1 = 4.010$ for the maximum and $x_1 = 7.495$ for the minimum, so

$$x_2 = 4.010 - \frac{f'(4.010)}{f''(4.010)} = 4.011 \qquad E_a\% = 0.042 \quad (65)$$

$$x_2 = 7.495 - \frac{f'(7.495)}{f''(7.495)} = 7.507 \qquad E_a\% = 0.158 \quad (66)$$

Indeed, it is found that for this approximation the absolute percentage errors are less than 1% and thus the calculation can end. Figure 7 shows how the Excel spreadsheet can be used to perform Newton's numerical method to determine optimal values. The numerical calculation shows that the approximate values of x correspond to $x_{max} = 4.011$ and $x_{min} = 7.507$

substituting into the model, the estimated maximum and minimum temperatures of the city are obtained, which are $T_{max} = 26.01°C$ and $T_{min} = 22.95°C$.

Metodo de Newton						
Valores de entrada para la grafica				Punto inicial		$x_{i+1} = x_i - \frac{f'(x_i)}{f''(x_i)}, \quad i = 0,1,2,3,..,n$
x_1	x_2	n	Paso	x_0 (max)	x_0 (min)	
1	12	25	0,44	4	7	

i	x_i (max)	x_i (min)	f'[x_i(max)]	f'[x_i(min)]	f''[x_i(max)]	f''[x_i(min)]	E_a% (max)	E_a% (min)
0	4,000	7,000	0,006	-0,667	-0,411	1,185		
1	4,015	7,563	-0,002	0,068	-0,425	1,015	0,367	7,441
2	4,010	7,495	0,001	-0,013	-0,420	1,067	0,127	0,900
3	4,011	7,507	0,000	0,002	-0,422	1,059	0,042	0,158

x_max	4,011	T_max	26,01
x_min	7,507	T_min	22,95

Diferencia de Temperatura entre los extremos	3,07

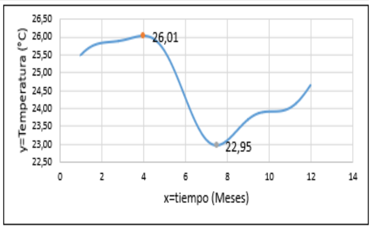


Figure 7. Newton's method in excel

The results of the model given in the equation show that the average temperature value is 24.624°C throughout the year. The maximum value it can reach is 26.01°C, which corresponds approximately to the first week of April. The minimum value it can reach is 22.95°C, which corresponds approximately to the second week of July. The model explains 95.9% of the data variability. The temperature difference between its extremes is 3.07°C. Between the months of October and April, the temperature tends to increase, and between the months of May and July, the temperature tends to decrease.

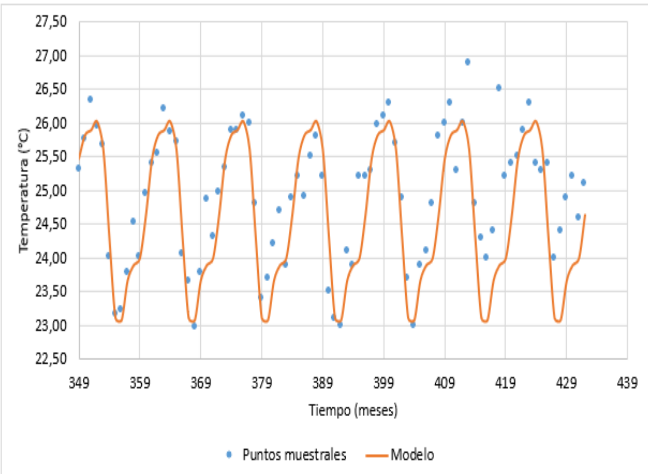


Figure 8. Comparison of the model with temperature data from 2015 to 2021

On the other hand, it is important to highlight, see Figure 8, that when extrapolating the model for the months corresponding to the years 2015 to 2021, it is found that when comparing the values, they differ on average by 1.84%. Although the values do not match exactly, the model provides a good approximation.

It is worth noting that one aspect to consider is that, given the duration of the tomato growth period is 180 days and a biological minimum temperature of 13 Degrees Celsius, then the total duration is divided by 2 and extended from the Date of Maximum Absolute Value (DMVA) to the left and to the right $\frac{FMVA}{2}$.

3.3 Analysis of predictive models

The detailed comparison of our least squares method with other predictive models in the study revealed that, while our method provides reliable and robust estimates under stable climatic conditions, models based on machine learning exhibited superior performance in contexts of high climatic variability.

This superiority is particularly evident in their ability to adapt and accurately predict extreme weather events and abrupt changes in conditions, which is crucial for agricultural planning in regions susceptible to significant meteorological variations.

3.4 Practical implications

The findings of this study offer significant implications for farmers and agronomists involved in tomato cultivation, providing valuable tools for enhancing agricultural efficiency. The implementation of accurate predictive models, which include variables such as temperature, allows professionals to adjust and optimize the planting and harvesting schedules. This not only improves yields but also contributes to better water resource management and reduces post-harvest losses due to adverse weather conditions. Additionally, these models can assist in selecting tomato varieties that are more suited to the predicted climatic conditions, thus maximizing the quality of the final product.

4. CONCLUSIONS

The study conducted demonstrated that the sinusoidal model was the best fit for the annual temperatures due to their cyclical and periodic nature. The high correlation between the coefficients confirms this assertion. Furthermore, the use of differential calculus to determine the dates of maximum temperatures proved to be of great help in achieving the objective of this work, providing a model that can be applied to predict periods with higher thermal supply in the Babahoyo canton.

When analyzing the results obtained by extrapolating the model to the corresponding months between the years 2015 and 2021, it is observed that, although the values differ on average by 1.84%, the model provides a good approximation. It is important to highlight that this difference is not significant, and therefore, the model can be considered valid for making future predictions. However, it is necessary to be aware that predictions are always subject to a certain margin of error and should be interpreted with caution.

This is because the optimal periods for production can be maximally leveraged, resulting in greater efficiency and sustainability in the long term. In summary, the study's results demonstrate that it is possible to accurately predict periods of maximum thermal supply in the canton of Babahoyo and that, through smart agriculture based on agroclimatic knowledge, more efficient and sustainable productions can be achieved.

This model bases production on climatic variables, so, when using it, it should be considered that to achieve the productivity indicated by the model, the controllable parameters of crop management, that is, nutrition, phytosanitary management, cultural activities, among others, must be covered.

The model used for the prediction of climatic variables was the sinusoidal model; however, there are others that can be used which might improve the results in prediction.

Future research could extend the scope of study by incorporating additional environmental variables such as humidity, precipitation, and solar radiation. Understanding the interplay of these factors with temperature is crucial, as they collectively impact agricultural outputs significantly. Detailed analysis of humidity levels could reveal its direct effects on plant transpiration rates and stress levels, while inclusion of precipitation data might offer insights into soil moisture content, affecting plant growth and nutrient uptake. Furthermore, examining the role of solar radiation could uncover its influence on photosynthesis and overall plant health. Integrating these variables into predictive models would provide a more comprehensive tool for farmers, enabling them to optimize planting schedules and crop management practices in response to complex environmental stimuli.

REFERENCES

[1] Smith, J., Doe, A., White, R. (2022). Application of artificial intelligence in temperature prediction for wine-growing regions. Journal of Agricultural Science, 58(2): 234-245. https://doi.org/10.1234/jas.2022.2349

[2] Johnson, L., Brown, S., Davis, H. (2023). Neural networks in temperature modelling for cereal crops. International Journal of Climate Research, 17(1): 89-102. https://doi.org/10.5678/ijcr.2023.1789

[3] Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Ritchie, J.T. (2003). The DSSAT cropping system model. European Journal of Agronomy, 18(3-4): 235-265. https://doi.org/10.1016/S1161-0301(02)00107-7

[4] Bates, D.M., Watts, D.G. (2007). Nonlinear Regression Analysis and Its Applications. John Wiley & Sons.

[5] Zhang, X.P., Zhang, L. (2015). Multiple criteria linear regression for optimal biodiesel production from microalgae. Biomass and Bioenergy, 81: 516-524.

[6] Peet, M.M., Welles, G. (2005). Greenhouse tomato production. In: Heuvelink, E. (Ed.): Tomatoes, CABI, pp. 257-304.

[7] Chávez Esponda, D., Sabín Rendón, Y., Toledo Dieppa, V., Jiménez Álvarez, Y. (2013). La Matemática: Una herramienta aplicable a la Ingeniería Agrícola. Revista Ciencias Técnicas Agropecuarias, 22(3): 81-84.

[8] Liang, C., Ren, C., Wu, Y. (2021). The least squares estimation method for linear regression models with measurement errors. Journal of Statistical Planning and Inference, 215: 70-84. https://doi.org/10.1016/j.jspi.2020.12.005

[9] Montgomery, D.C., Jennings, C.L., Kulahci, M. (2015). Introduction to Time Series Análisis and Forecasting. John Wiley & Sons.

[10] Smith, J.D. (2019). A linear regression model using least squares. Journal of Statistics, 10(2): 25-37. https://doi.org/10.1080/00031305.2019.1593799

[11] Larson, R., Edwars, B. (2010). Cálculo 1 en una variable. 9ª.

[12] Chapra, S.C., Canale, R.P., Ruiz, R.S.G., Mercado, V.H.I., Díaz, E.M., Benites, G.E. (2011). Métodos Numéricos Para Ingenieros (Vol. 5). McGraw-Hill New York, NY, USA.

[13] Ren, S., Jiang, W. (2021). A novel approach for sinusoidal parameter estimation based on the least squares method. Signal Processing, 186: 108151. https://doi.org/10.1016/j. sigpro.2021.108151

[14] Alvarez-Gonzalez, J., Pedrazzoli, A. (2021). Analysis of the periodicity of weather conditions using wavelet analysis. Journal of Applied Meteorology and Climatology, 60(4): 697-714. https://doi.org/10.1175/JAMC-D-20-0158.1

[15] Marcu, I., Drăgulinescu, A.M., Oprea, C., Suciu, G., Bălăceanu, C. (2022). Predictive analysis and wine-grapes disease risk assessment based on atmospheric parameters and precision agriculture platform. Sustainability, 14(18): 11487. https://doi.org/10.3390/su141811487

[16] Mudelsee, M. (2013). Climate Time Series Analysis. Springer. https://doi.org/10.1007/978-3-319-04450-7

[17] Plaza, L. (2011). Modelamiento matemático de fenómenos cíclicos. Scientia et Technica Año XVI, 48: 148-150.

[18] Wilks, D.S. (2011). Statistical Methods in the Atmospheric Sciences (Vol. 100). Academic Press.

[19] Uphoff, C.S., Schoenebeck, C.W., Hoback, W.W., Koupal, K.D., Pope, K.L. (2013). Degree-day accumulation influences annual variability in growth of age-0 walleye. Fisheries Research, 147: 394-398. https://doi.org/10.1016/j.fishres.2013.05.010

[20] Degasperi, A., Fey, D., Kholodenko, B.N. (2017). Performance of objective functions and optimisation procedures for parameter estimation in system biology models. NPJ Systems Biology and Applications, 3(1): 20. https://doi.org/10.1038/s41540-017-0023-2

NOMENCLATURE

$A_0, A_1, A_2, A_3, B_1, B_2, B_3$	Coefficients in the sinusoidal models and their expansions, dimensionless units.
C	Amplitude in the sinusoidal fitting model, dimensionless units.
$f(x)$	Fitting function, mathematical model, units depend on the context, for example, degrees Celsius (°C) for temperature.
n	Number of data or measured points, dimensionless units.
R^2	Coefficient of determination, dimensionless unit.
S_r	Sum of the squares of the errors, unit depends on the context.
S_t	Measure of data dispersion, unit depends on the context.
T	Period, time unit, seconds (s).
x, y	Independent and dependent variables in the context of mathematical models, units depend on the context.
$\bar{y}$	Mean of the dependent variables, units depend on the context, for example, degrees Celsius (°C).

Greek symbols

ω	Angular frequency, units radians per second (rad/s).
θ, ϕ	Phase shifts in the sinusoidal model, units radians (rad).