HYBRID APPROACH FOR PREDICTION OF TEMPERATURE AND MOISTURE IN GREENHOUSES USING ARIMA, ARTFIMA AND SVM METHODS

SEBA, M. R.[1*] – SEBA, D.[2] – BERKANI, A.[3] – MEKHLOUFI, B.[4]

[1]*Department of Mechanical engineering, Applied Mechanics Laboratory, University of Sciences and Technology of Oran Mohamed Boudiaf (USTO-MB), Oran, Algeria*

[2]*Department of Mathematics, Applied Mathematics Laboratory, University of Bejaia, Bejaia, Algeria*

[3]*Laboratory of L2GEGI, Faculty of Applied Sciences, University of Tiaret BP 78 size Zaaroura 14000, Tiaret, Algeria*

[4]*Department of Science and Technology, University of Tiaret BP 78 size Zaaroura 14000, Tiaret, Algeria*

[*]*Corresponding author*
e-mail: mohamedrida.seba@univ-usto.dz

(Received 21st Jul 2023; accepted 22nd Sep 2023)

Abstract. Although greenhouses are used in controlled environments, it can be difficult to maintain the right temperature and moisture inside, especially when the outdoor climate changes. To ensure that the requirements are met, this research aims to develop a model that can accurately predict the temperature and moisture inside the greenhouse. Accordingly, we have developed two hybrid models that can accurately predict the temperature and moisture inside the greenhouse by combining the advantages of Auto Regressive Integrated Moving Average (ARIMA) and Auto Regressive Fractionally Integrated Moving Average (ARTFIMA) on the one hand and the advantages of Support Vector Machine (SVM) on the other. These methods train the models and adjust their parameters using previous experimental data, allowing them to make accurate predictions even in complex scenarios. In this study, 1,987 measurements were used over two time series, one for temperature and one for moisture, with an interval of 10 minutes between each observation. The two hybrid models were then compared with the experimental results, which were found to give accurate predictions with an error rate close to zero. They were also compared with other single models. When results are compared based on statistical error, they are far more accurate in terms of the root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) than results based on a single model. The two hybrid model values, in general, are the most suited ones since they can accurately correct experimental data. In conclusion, it is possible to forecast greenhouse temperature and brick ratio using the ARIMA-SVM and ARTFIMA-SVM models.
Keywords: *advanced forecasting, controlled environment, statistical modelling, support vector modelling, time series analysis*

Introduction

A type of greenhouse structure known as a multi-span greenhouse has numerous linked spans that together provide a sizable growth area. The performance and atmosphere within a multi-span greenhouse are significantly influenced by its design. The quantity of light and heat that enters the building, as well as the effectiveness of the heating and cooling systems, are all influenced by the size, shape, and direction of the building. The greenhouses performance and capacity to maintain a stable atmosphere are greatly influenced by the sort of covering material chosen. Furthermore, the environment outside

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_57375751

might have a big impact on the environment within the multi-span greenhouse. Temperature, moisture, wind direction, and light levels all have an impact on how well the greenhouse can keep a stable growth environment. Yet, producers may manage and optimize the inside climate to promote optimal crop development and output by adopting sustainability control mechanisms, such as heating, cooling, and ventilation.

Owing to the significance of greenhouse structures, several research investigations, including those focusing on the mechanical and thermophysical properties, have been carried out. An analysis of multi-span greenhouse constructions using numerical simulation is provided in a study by Fernández-García et al. (2020). According to their explanation, numerical simulations may be used to forecast how greenhouse buildings would behave under various environmental factors. In a paper, (Kim et al., 2017) studied the pressure coefficient values while considering a number of greenhouse design elements. including the roof slope, the bend diameter, and the direction of the wind. Based on the defined parts of the greenhouse surface, various wind directions, and the number of spans. The assumption of a uniform distribution of heat and mass inside greenhouses was challenged due to various factors, as demonstrated in a study from Ogunlowo et al. (2021). A dynamic system modeling program was developed by Rasheed et al. (2020) for a construction energy simulation model for multi-span greenhouses, allowing for daily and seasonal adjustment of screens, roof ventilation, and heating set points. The influence of vent layout and opening size on greenhouse climate was evaluated through the creation, verification, and utilization of a 3D model, as evidenced in work carried out by He et al. (2015). In research from Lee and Short (2000), a computational fluid dynamics (CFD) system served to explore natural adequate ventilation in a commercial greenhouse, considering factors like wind direction, speed, and ventilation volumes. A modeling process was carried out on real plastic greenhouses as part of an investigation (Ha et al.,2017) to evaluate the dynamic properties of these structures. In addition, an updated finite element model was developed in order to investigate how connection and support .stiffness affected these structures' dynamic properties (Reichrath and Davies, 2002) validated stress distribution estimates produced by Fluent CFD software on the outside of the greenhouse roof using experimental results from the 52 span Venlo greenhouse. Based on time series models, some additional exploration (Tol and de Vos, 1993) provide statistical evidence that growing greenhouse gas concentrations cause an increase in average temperature.Another study (Khan et al., 2014) used multiple econometric tools to investigate the long-term link between energy usage and greenhouse gas emissions. The researchers as (Cai et al.) developed a Light Gradient Boosting Machine method (LGBM) to estimate greenhouse temperature using ambient and control data, taking physical aspects, seasonality, and behavioral correlation into consideration. Thermal losses through the polymeric cover of greenhouses were calculated, and techniques for calculating greenhouse dimensions and projecting future temperatures were provided, as shown by Leonidopoulos (2000).

In scientific investigations involving temperature and moisture phenomena, data evaluation and estimation of changes in the examined phenomena have been conducted through various numerical methods (Arunsandeep and Chandramohan, 2018; dos Santos and Mendes, 2004; Xue et al., 2021) as well as statistical methodologies (Rasheed et al.,2022; Gorgin et al., 2020; Dimitriadis et al., 2021). Other methods of prediction exist through Temperature and moisture time series, but it can be difficult to analyze and predict in some complex cases. That is why researchers are striving to build hybrid approaches that integrate mathematical models with other techniques, such as deep

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_57375751

learning, neural networks, and other machine learning techniques, to increase the accuracy of analysis and prediction, and this strategy adopts on the advantages of many methods while minimizing the disadvantages, resulting in a more reliable and practical solution.

Greenhouses are utilized in controlled environments; yet, the task of regulating optimal temperature and moisture levels within them can be arduous, especially when faced with fluctuations in exterior climatic conditions. Drawing upon previous experimental findings, the main aim of this research is to expand the utilization of hybrid models, specifically Autoregressive Moving Average (ARIMA) - Support Vector Machines (SVM) and Auto Regressive Fractionally Integrated Moving Average (ARTFIMA) - Support Vector Machines (SVM), in various applications. The objective of these studies is to effectively forecast temperature and moisture levels within greenhouse environments in order to provide ideal cultivation conditions and optimize crop productivity. Our research aims to develop resilient prediction tools that can effectively adapt to diverse environmental conditions by using the capabilities of ARIMA, ARTFIMA, and SVM approaches. The aforementioned models will function as decision-support systems for the purpose of greenhouse management. Their primary objective is to facilitate proactive modifications, such as modifying ventilation or irrigation systems, in order to efficiently regulate interior climate conditions.

Phenomenon, materials and methods

Phenomenon

Three main factors affecting greenhouse production are the greenhouse itself, the external environment, and the internal environment, and these factors are linked in a sequential manner, as the external environment and the greenhouse affect the internal environment, specifically temperature degree and moisture levels, and the internal environment, in turn, affects the quality of the final product.

Since we have no control over the external environment, managing the internal environment, particularly regulating temperatures and moisture levels, is crucial to improving crop production.

These effects are as follows:

An increase in the external temperature with intense sunlight leads to an increase in the internal temperature and vice versa. As for the percentage of moisture, it increases with an increase in external moisture, and evaporation resulting from the increase in temperature and hot winds increases the temperature, and the opposite is always true. As for the percentage of moisture, it increases with increasing moisture. The externality and the evaporation caused by the increase in temperature and hot winds increase the temperature, and the opposite is always true.

A greenhouse with a glass body allows more sunlight in, which can warm the indoors. Conversely, a greenhouse with a plastic cover may not let in as much sunlight, resulting in cooler temperatures.

The materials utilized to construct the greenhouse also have an impact on the internal conditions. Certain materials, for example, have higher insulating characteristics and can assist in retaining heat and moisture more efficiently.

Figure 1 depicts a schematic of the examined phenomenon occurrence together with its controls and monitoring mechanisms.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_57375751

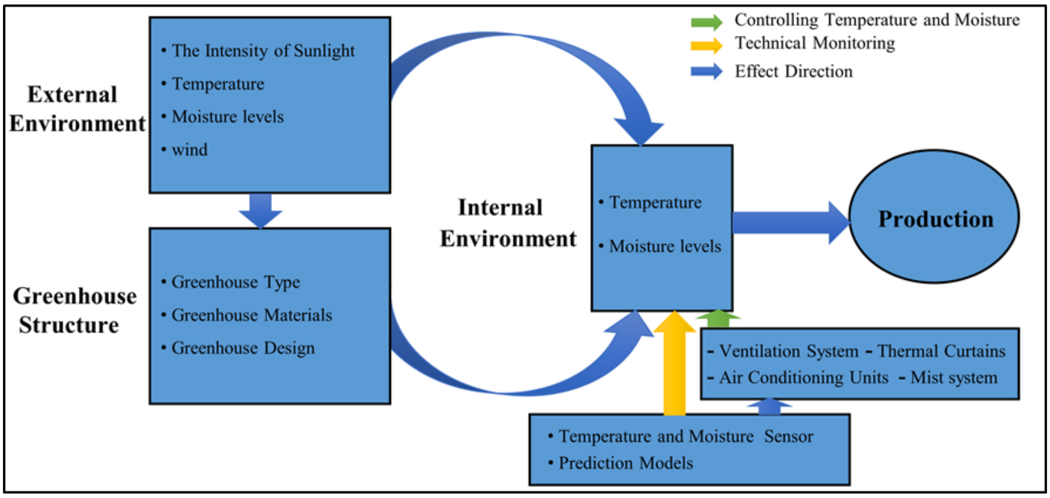
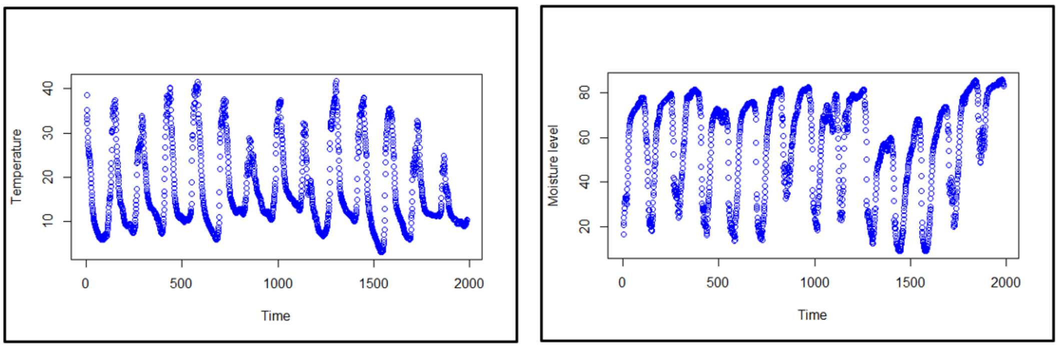


Figure 1. *Scheme of the studied phenomenon*

The data set used in this study was selected from the experimental work of Petrakis et al. (2022), where we used 1,987 observations with an estimated time difference of 10 minutes from the temperature and moisture data inside the greenhouse, and we divided this data into two parts, the first with a total of 1700 data for training and a total of 287 data for testing. *Figure 2* shows the data of temperature and moisture .As for the greenhouse targeted in the study, it consists of 4 units featuring a glass cover, the size of each unit is 199.68 cubic meters, with a height of 3.9 meters above the ground.



(a) (b)

Figure 2. *Temperature (C°) (a) and Moisture (g/m³) (b) time series outside the greenhouse*

Method

ARIMA and ARTFIMA models

Autoregressive integrated moving average (ARIMA) models predict future values based on past values, it gauges the strength of one dependent variable relative to other changing variables (Box and Jenkins, 1976).

A stochastic process $(X_t)_{t\geq 0}$ is said to be an ARIMA (p, d, q) an integrated mixture autoregressive moving average model if it adheres to *Equation (1)*, the elements of which are shown in *Equations (2) and (3)*:

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_57375751

$$\phi(L)(1-L)^d X_t = \theta(L)\varepsilon_t \quad \forall t \geq 0 \qquad \text{(Eq.1)}$$

where $d \in \mathbb{N}$, L is lag operating, $\varepsilon_t \sim \mathcal{N}(0,\sigma^2)$ i.i.d. errors, with $\sigma^2 < \infty$.

$$\phi(L) = \left(1-\phi_1 L - \ldots - \phi_p L^p\right) \text{ with } \phi_p \neq 0 \qquad \text{(Eq.2)}$$

$$\theta(L) = \left(1-\theta_1 L - \ldots - \theta_q L^q\right) \text{ with } \theta_q \neq 0 \qquad \text{(Eq.3)}$$

In the case of $d = 0$, we obtain *Equation (4)* of ARMA (p, q) process:

$$\phi(L) X_t = \theta(L)\varepsilon_t \quad \forall t \geq 0 \qquad \text{(Eq.4)}$$

According to Sabzikar et al. (2015) if $d \in \mathbb{R}$, we get Fractional ARIMA (ARFIMA). ARTFIMA is defined by *Equation (5)* as follows:

$$\phi(L)\left(1-e^{-\lambda}L\right)^d X_t = \theta(L)\varepsilon_t \quad \forall t \geq 0 \qquad \text{(Eq.5)}$$

SVM model

Support vector machine (SVM) analysis is a popular machine learning tool for classification and regression, it is considered a nonparametric technique because it relies on kernel functions (Vapnik, 1995).

Given deviation data of training $x_i(1,2,\ldots\ldots,m)$ where $x_i \in \mathbb{R}^n$is the input vector with n −dimension, $y_i \in \mathbb{R}^n$ is the associated desired output value of x_i. Then the SVMs model is formulated in *Equation (6)* as follows:

$$f(x) = w\phi(x) + b \qquad \text{(Eq.6)}$$

where $(\phi(x))$ is called the feature that is not linearly mapping from the input space x. The w and b are coefficients that are estimated by minimizing the regularized risk function shown in *Formula (7)*:

$$\min \frac{1}{2} w^T w + C \sum_{i=1}^{m} L(y - f(x), x) \qquad \text{(Eq.7)}$$

where C is the regularized constant determining the trade-off between the empirical error and the regularization term, The larger the constant C is, the more the minimum experience risk is emphasized, and the lower the generalization of function f.

Using the Lagrange function and duality theory, and with the kernel function $k(x, x)$ introduced, the function given in *Formula (7)* can be transformed into a quadratic programming problem can be transformed into a quadratic programming problem in *Equation (8)* as follows: as follows:

$$\min_{\alpha_i,\alpha_j} \sum_{i,j=1}^{m} \left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right) k\left(x_i, x_j\right) + \check{n} \sum_{i=1}^{m} \left(\alpha_i + \alpha_i^*\right) \sum_{i=1}^{m} y_i \left(\alpha_i - \alpha_i^*\right) \qquad \text{(Eq.8)}$$

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_57375751

$$0 \le \alpha_i, \alpha_i * \le C, \quad i = 1, 2, ..., m$$

where α_i and $\alpha_i *$ are Lagrange multipliers. They are obtained by solving this quadratic programming problem, and the input vector x_i corresponding to the nonzero α_i and $\alpha_i *$ is the support vector. Thus, As may be seen below, *Equation (6)* is transformed into *Equation (9)*:

$$f\left(x, \alpha_i, \alpha_i^*\right) = \sum_{i=1}^{m} \left(\alpha_i - \alpha_i^*\right) k\left(x_i, x_j\right) + b \qquad \text{(Eq.9)}$$

Now, we can forecast using the equation above.

Hybrid methodology forecasting

Based on the relationship (10), we divide the time series into training set which is composed of 1700 observations and test set composed of 287 observations. We use the test set to validate the results:

$$X_t = Y_t + U_t \qquad \text{(Eq.10)}$$

where X_t is decomposed into two parts linear component Y_t and nonlinear component U_t, we apply statistical models (ARIMA and ARTFIMA) which are more suited to linear pattern $\hat{Y}(t)$, then the difference $\varepsilon_t = X_t - \hat{Y}(t)$ is the residual. The residual series ε_t contains the nonlinear parts, thus we use SVM and to fit residuals gives the predicted values $\widehat{U}(t)$, finally we combine the two predictive results in *Equation (11)*:

$$\hat{X}(t)_t = \hat{Y}(t)_t + \hat{U}(t)_t \qquad \text{(Eq.11)}$$

Augmented Dickey-Fuller model

The Augmented Dickey-Fuller test is a statistical test that is utilized for the purpose of validating or invalidating the null hypothesis that an autoregressive model that is applied to time-series data contains a unit root. The existence of a unit root suggests that the data from the time series is non-stationary, which means that it exhibits some kind of structure that is reliant on the passage of time, such as a trend or seasonality.

The fundamental model of the Augmented Dickey-Fuller test can be seen below in the form of *Equation (12)*, which reads as follows:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p-1} + \varepsilon_t \qquad \text{(Eq.12)}$$

where $\Delta Y_t = Y_t - Y_{t-1}, Y_t$ is the time series and Y_{t-1} is lag 1 of time series , α is a constant, β is the coefficient on a time trend, and ε is white noise.

The null hypothesis of the test is that the time series can be represented by a unit root, and that it is not stationary (it has some time-dependent structure). The alternative hypothesis (rejecting the null hypothesis) is that the time series is stationary. If the p-value obtained from the test is less than the critical value, usually 0.05, you reject the null hypothesis in favor of the alternative hypothesis.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_57375751

Practical implementation of the proposed hybrid ARTFIMA-SVM and ARIMA-SVM models

The hybrid ARTFIMA-SVM and ARIMA-SVM models have demonstrated their potential in practical applications by providing improved predictive capabilities. These models are particularly valuable in sectors and fields where accurate forecasting of environmental parameters, such as temperature and moisture, is essential. The following is a sequential breakdown of the practical implementation of the proposed approach:

Data Collection: The initial stage would entail acquiring an ample amount of historical data pertaining to the specific variable under investigation, such as temperature, moisture, stock prices, or any other form of time-series data.

Data Pretreatment: Data preprocessing is an essential step in preparing data for analysis, particularly in the context of time-series analysis. Prior to applying any model, it is often necessary to do various operations such as data cleaning, normalization, and transformation. These processes are undertaken to ensure that the data is appropriately formatted and meets the requirements of the research being conducted.

Initial Projection Utilizing ARIMA or ARTFIMA: The ARIMA or ARTFIMA model employed to analyze the preprocessed data and produce an initial forecast. The aforementioned models possess the ability to effectively represent the fundamental patterns, cyclicality, and additional linear elements that are inherent in the given dataset.

Determine Residuals: Following the first forecast, compute the residuals, which denote the disparities between the observed values and the values anticipated by the ARIMA or ARTFIMA model.

Training for SVM on Residuals: The residuals can be utilized as the input for training the Support Vector Machine (SVM) model. The support vector machine (SVM) will attempt to capture the non-linear patterns present in the residuals, which were not adequately represented by the ARIMA or ARTFIMA models.

Final Projection: The final forecast entails integrating the original forecast generated by either the ARIMA or ARTFIMA model with the predictions derived from the support vector machine (SVM) on the residuals. The resulting forecast will be optimized to account for both linear and non-linear tendencies.

Effectiveness Evaluation: The performance evaluation of the model can be conducted by utilizing metrics such as Root Mean Square Error (RMSE), Mean Square Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). A more favorable model fit would be indicated by lower values on these metrics.

Iterative Modification: Iterative refinement involves the process of iteratively adjusting the model parameters to enhance performance, taking into consideration performance measures and potentially additional domain-specific factors.

Results

Analyzing temperature and moisture time series

To analyses the temperature and moisture time series, we go through four steps. First, we perform the stability test of the statistical series, then we examine the degree of correlation between the data, then we proceed to the construction of prediction models using the methods mentioned in section (2.2), and finally these models are evaluated by a set of indicators.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_57375751

Stationarity test

The first step in the process is to determine whether the time series is stationary or not. Analysis of the behavior of the data and prediction of the future are more accurate when the time series are stationary or converted from non-stationary to stationary using statistical methods and procedures.

The Augmented Dickey-Fuller model Dickey-Fuller test (Mushtaq, 2011) was used in our survey to check the stationarity of temperature and moisture time series, and the results are as follows:

For temperature time series, Dickey-Fuller = -6.7067, p-value = 0.01.

For moisture time series, Dickey-Fuller = -7.4446, p-value = 0.01.

The p-value is lower than the significance level and the test statistic is more negative from the critical value at this significance level, you can reject the null hypothesis (not stationary) and conclude that the time series is stationary.

Autocorrelation and partial autocorrelation function

Time series analysis uses the autocorrelation function (ACF) and partial autocorrelation function (PACF) to find patterns in the data. Both ACF and PACF measure the correlation between a time series and its lagged values. However, PACF only measures A after removing the influence of the intermediate lags. *Figures 3, 4, 5* and *6* show the correlation coefficients for each lag in two time series related to temperature and moisture using ACF and PACF function.

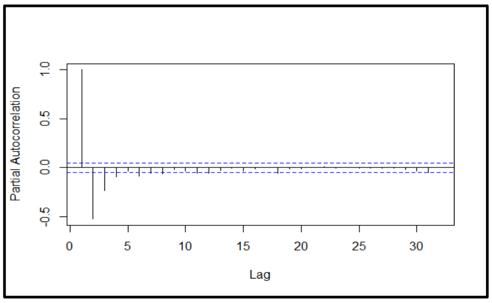


Figure 3. *Correlation coefficients for each lag time in temperature time series using ACF function*

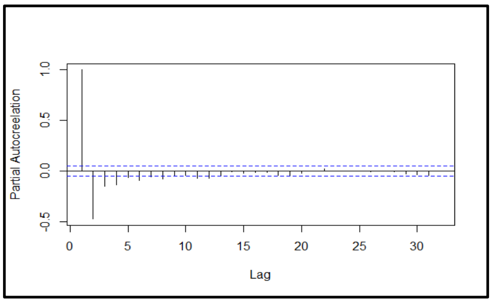


Figure 4. *Correlation coefficients for each lag time in moisture time series using ACF function*

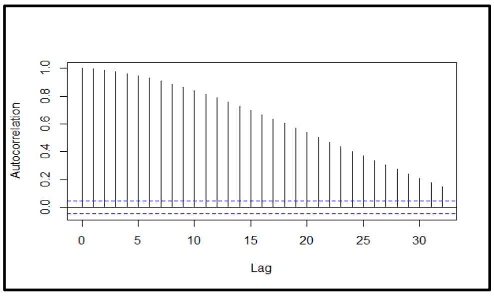


Figure 5. *Correlation coefficients for each lag time in temperature time series using PACF function*

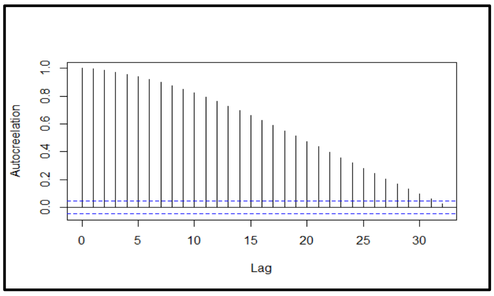


Figure 6. *Correlation coefficients for each lag time in moisture time series using PACF function*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_57375751

The greater the lag time in the values for the temperature and moisture time series. Correlation strength decreased. There is also a direct correlation when using the ACF function, and an inverse correlation when using the PACF function.

Construction of the models

Using historical data, we developed models in this research study and used the approaches and procedures described in section (2.2). In particular, we analyzed two independent time series data sets, one for temperature and the other for moisture, using the ARIMA model. We were able to establish the proper ordering for the moving average and autoregressive components by training the models with the data. The models that emerged were determined to be non-zero mean ARIMA (5; 0; 1) and ARIMA (4; 0; 4) respectively. The following are the *Equations (13) and (14)* for these models that were provided in the study.

$$\left(1+0.1117L-0.5972L^2-0.0195L^3+0.0083L^4-0.1316L^5\right)X_t=17.5988+(1-0.7457L)\varepsilon_t \quad \text{(Eq.13)}$$

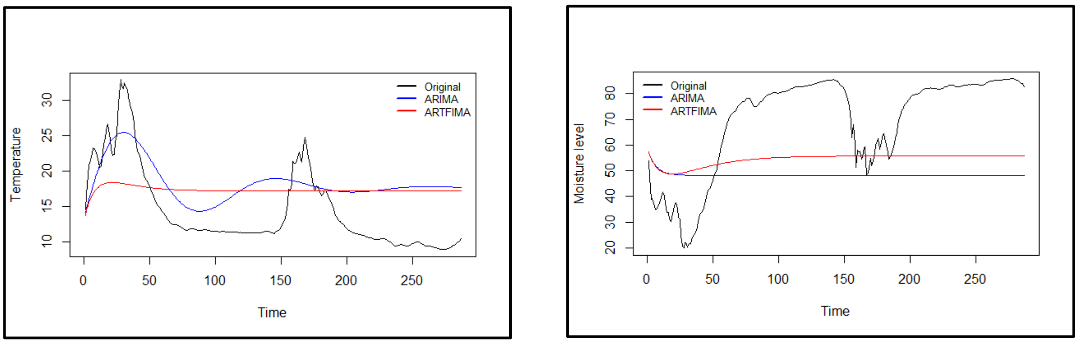
$$\left(1-1.1092L-0.2946L^2+0.0007L^3+0.4101L^4\right)X_t=17.5988+(1-0.6625L+0.0868L^2+ +0.6104L^3+0.3609L^4\right)\varepsilon_t \quad \text{(Eq.14)}$$

Now, we implement the ARTFIMA (0;1.828;0) and ARTFIMA (0;1.679;0) model for both time series of temperature and moisture which is a semi-long-term dependence, and the model is expressed in *Formulas 14 and 15* as follows:

$$\left(1-e^{-0.04}L\right)^{1.828}X_t=55.746+\varepsilon_t \quad \text{(Eq.15)}$$

$$\left(1-e^{-0.07}L\right)^{1.679}X_t=17.207+\varepsilon_t \quad \text{(Eq.16)}$$

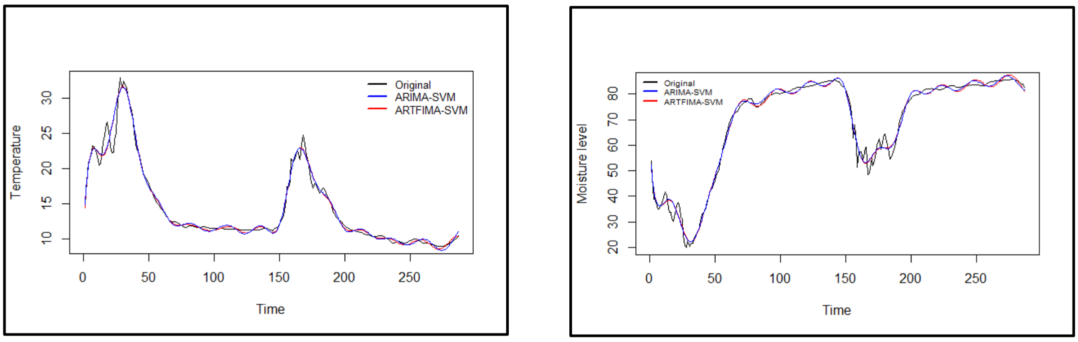
We estimate the parameters in ARIMA and ARTFIMA using Maximum likelihood estimation. Then we present the forecasting results in the following *Figures 7,8*.



(a) (b)

Figure 7. *Prediction results of the ARIMA and ARTFIMA model for temperature (a) and moisture (b)*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_57375751



(a) (b)

Figure 8. *Prediction results of the ARIMA-SVM and ARTFIMA-SVM model for temperature (a) and moisture (b)*

Following the prediction provided by the ARIMA and ARTFIMA models, we implement the hybrid prediction methodology using SVM-ARIMA and SVM-ARTFIMA, in which the non-linear patterns in the residual errors are captured and then re-predicted by SVM and these predictions are added to the ARIMA and ARTFIMA predictions to obtain final time series forecasts for text data set, as shown in the *figures* below.

Accuracy of the models

Common indicators utilized evaluate the effectiveness of a predictive model include mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean percentage error (MAPE). The *Formulas (17) to (20)* are regarded as mathematical expressions used for the computation of indicators.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{X}_t - X_t\right)^2 \tag{Eq.17}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{X}_t - X_t\right)^2} \tag{Eq.18}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{X}_t - X_t\right| \tag{Eq.19}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\hat{X}_t - X_t\right|}{X_t} \tag{Eq.20}$$

where n is the total data set, X_t represents the real value, and $\hat{X}_t$ represents the predicted value .The MSE computes the mean squared difference between the expected and actual values. A more understandable statistic, expressed in the same units as the target variable, is given by the RMSE, which is the square root of the MSE. The mean absolute difference between the expected and actual values is measured by MAE. In comparison to MSE and RMSE, it is less susceptible to outliers. The average relative error between the anticipated and actual values is measured using MAPE.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_57375751

Tables 1 and *2* present the indicators used to evaluate the effectiveness of predictive models for time series temperature and moisture in greenhouses.

Table 1. *Error values for temperature forecasting models with four indicators*

Models \ Indicators	RMSE	MSE	MAE	MAPE
ARIMA	5.58	31.17	5.01	0.42
ARTFIMA	6.16	38.01	5.62	0.44
ARIMA-SVM	0.80	0.64	0.50	0.03
ARTFIMA-SVM	0.79	0.62	0.44	0.02

Table 2. *Error values for moisture forecasting models with four indicators*

Models \ Indicators	RMSE	MSE	MAE	MAPE
ARIMA	28.84	808.96	26.33	0.39
ARTFIMA	22.72	516.40	20.92	0.32
ARIMA-SVM	2.01	4.04	1.55	0.02
ARTFIMA-SVM	1.94	3.79	1.46	0.02

Discussion

The lack of prediction and prior detection of the temperature and moisture level inside the greenhouse can cause significant damage to the product. Therefore, establishing an effective monitoring and warning system to make intelligent decisions and manage the indoor environment is a major scientific and technical challenge that must be addressed immediately. As temperature and moisture content are not always stable, the traditional statistical models used for forecasting often fail to provide reliable results. To overcome this issue, we employed hybrid prediction models based on data analysis and machine learning techniques to estimate temperature and moisture from sensors within the greenhouse in real time.

The model suggested in this paper can be used in the following aspects:

- Examining the performance of control systems and detecting any trends or anomalies in the data by analyzing time series of temperature and moisture changes inside the greenhouse. This can help optimize greenhouse design, management, resource utilization, and production conditions.
- The analysis of time series data can assist forecast temperatures and moisture levels, allowing for proactive greenhouse management to minimize any negative impacts on production.

Based on the preceding findings and the associated error metrics, the hybrid models ARIMA-SVM and ARTFIMA-SVM demonstrate a sophisticated methodology for time-series prediction, particularly in complex domains such as environmental science, where variables such as temperature and moisture levels are subject to both linear and non-linear dynamics. Traditional approaches such as ARIMA (AutoRegressive Integrated Moving Average) and ARTFIMA (AutoRegressive and Fractionally Integrated Moving Average) have been developed to effectively address linear correlations and accurately capture trends and seasonality. Nevertheless, their performance may be compromised in

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_57375751

situations where data demonstrate nonlinear phenomena, a common occurrence in several natural systems.

In contrast, Support Vector Machines (SVM) are designed with the explicit purpose of detecting and representing intricate, non-linear associations within datasets. The combination of ARIMA or ARTFIMA with SVM results in the creation of a multi-layered model. The initial layer of analysis involves the utilization of either the ARIMA or ARTFIMA model, which effectively captures the linear trends and autocorrelations present in the data. The residuals, which refer to the discrepancies between the predicted values and the actual values, are subsequently transmitted to the Support Vector Machine (SVM) layer. In this context, the Support Vector Machine (SVM) aims to minimize the residuals, with a particular emphasis on capturing the non-linear characteristics that may have been overlooked by the ARIMA or ARTFIMA models.

Furthermore, the ARTFIMA model introduces an additional level of intricacy by possessing the capability to accommodate long-memory characteristics. Long-memory refers to the capacity of a system to retain or be impacted by its past values over an extended duration. The long-term effects of some factors are particularly significant in the context of environmental data analysis.

The root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are all important metrics for evaluating the performance of models in the context of time series forecasting. Each serves a distinct purpose in evaluating differences between predicted and actual data, with lower values indicating greater model accuracy.

The ARTFIMA-SVM hybrid model shows good performance in analyzing temperature data, as evidenced by exceptionally low values of root mean square error (RMSE) at 0.79, mean square error (MSE) at 0.62, mean absolute error (MAE) at 0.44, and The absolute percentage error (MAPE) is set at 0.02.

The humidity dataset also shows comparable advantages. The hybrid model, known as ARTFIMA-SVM, shows root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), and mean absolute error (MAPE) values of 1.94, 3.79, 1.46, and 0.02 over straight. While these values differ slightly compared to the ARMA-SVM model, whether in temperature or humidity.

The comparative analysis of the ARIMA model and its hybrid ARIMA-SVM counterpart demonstrates a significant disparity in performance. The root mean square error (RMSE) values for temperature and moisture, obtained solely by the application of the ARIMA model, are 5.58 and 28.84, correspondingly. The observed values exhibit a notable level of elevation in comparison to the ARIMA-SVM model, which demonstrates considerably superior accuracy rates. The observed pattern remains consistent for mean squared error (MSE) values, with temperature and moisture exhibiting values of 31.17 and 808.96 respectively. Similarly, mean absolute error (MAE) values for temperature and moisture are 5.01 and 26.33 respectively. Additionally, the mean absolute percentage error (MAPE) values for temperature and moisture are 0.42 and 0.39 respectively. The observed discrepancy in error rates highlights the intrinsic limits of the ARIMA model in accurately representing the intricate intricacies and non-linear patterns present in real-world time-series data. The inherent limitations in its capacity to comprehensively encompass intricate phenomena are highly evident, and the deficiencies in its ability to accurately forecast outcomes are not trivial, but rather substantial.

In a similar vein, it is worth noting that the ARTFIMA model, when considered independently, exhibits certain drawbacks. The model's root mean square error (RMSE)

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_57375751

values for temperature and moisture are 6.16 and 22.72, respectively. The mean square error (MSE) values are 38.01 and 516.40, while the mean absolute error (MAE) values are 5.62 and 20.92. Additionally, the mean absolute percentage error (MAPE) values for temperature and moisture are 0.44 and 0.32, respectively. Although ARTFIMA demonstrates the capability to incorporate long-memory and fractional differencing, it exhibits limitations in capturing non-linear dynamics, as evidenced by its comparatively higher error metrics in relation to the ARTFIMA-SVM model.

The large variation in observed error measures between the hybrid model and individual models, such as ARIMA or ARTFIMA in isolation, provides compelling evidence that the integration of linear and nonlinear modeling techniques results in a powerful forecasting tool. The ARTFIMA-SVM hybrid model is a very reliable choice for applications that prioritize accurate time series forecasting. It offers a powerful alternative to models that focus only on linear or nonlinear features, due to its remarkably low error rates.

It is known that the goal of SVM in a hybrid ARIMA-SVM and ARTFIMA-SVM model are to increase forecast accuracy by utilizing the SVM method to anticipate the residual of ARIMA and ARTFIMA errors, this can increase the predictability of complicated and unstable time series. Is clearly visible from the forecast comparison between *Figure 7* and *Figure 8*.

In contrast, the two proposed models in this paper rely on hybridization between a statistical method for predicting statistical chains and machine learning technology, and a method for predicting temperature and moisture level inside greenhouses, which is one of the methods for predicting nonlinear data. And then deduce the direction of changes in temperature and moisture. The benefits are clear when compared to automated approaches for estimating greenhouse temperature and moisture levels. First, reduced modeling data needs lead to cheaper modeling costs. Hence, in locations where there is a paucity of data for greenhouses, this approach can be used to predict the temperature and moisture level inside greenhouses.

Due to the strong application of ARIMA-SVM and ARTFIMA-SVM models for analyzing and predicting nonlinear problems in an uncertain environment, the prediction reliability of the two models is good. As a result, the accuracy of quality prediction is increased.

Conclusion

This study presented two hybrid models, ARIMA-SVM and ARTFIMA-SVM, to increase the precision of forecast data for temperature and moisture in inner greenhouse. Where the both models ARIMA and ARTFIMA model the trend and seasonality of the time series with the calculation of the residual errors between the observed values, where the residual errors are used as input data for the SVM algorithm to predict the residual errors in the future, and then additional error predictions are added to the trend of the model and the seasonal forecast to obtain the final prediction of the time series.

The result indicated that among the hybrid models better performance than single models, ARTFIMA-SVM also showed the best performance in terms of low RMSE, MSE, MAE and MAPE value compared to ARTFIMA-SVM.

We may utilize the two models ARTFIMA-SVM and ARTFIMA-SVM as a preventative step to employ the required tools to manage the temperature and moisture

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_57375751

within the glass houses in order to acquire the suitable environment for the products quality.

The two models put forward in this study still have much room for improvement. Both ARTFIMA-SVM and ARTFIMA-SVM models proposed in this paper ignore the external environment, which has a greater influence on changes in temperature and moisture and may interfere with the prediction process, thus reducing accuracy. It uses only the previous data on the temperature and moisture inside the greenhouse. Ways to minimize the involvement of external elements or to take into account the effects on the internal environment still need to be sought.

REFERENCES

[1] Arunsandeep, G., Chandramohan, V. P. (2018): Numerical Solution for Determining the Temperature and Moisture Distributions of Rectangular, Cylindrical, and Spherical Objects During Drying. – Journal of Engineering Physics and Thermophysics 91(4): 895-906.

[2] Box, G. E. P., Jenkins, G. M. (1976): Time series analysis, forecasting and control. – The Clarendon Press, San Francisco: Holden-Day, 2nd edition.

[3] Cai, W., Wei, R., Xu, L., Ding, X. (2022): A method for modelling greenhouse temperature using gradient boost decision tree. – Information Processing in Agriculture 9(3): 343-354.

[4] Dimitriadis, P., Koutsoyiannis, D., Iliopoulou, T., Papanicolaou, P. (2021): A Global-Scale Investigation of Stochastic Similarities in Marginal Distribution and Dependence Structure of Key Hydrological-Cycle Processes. – Hydrology 8(2): 59.

[5] dos Santos, G. H., Mendes, N. (2004): Analysis of numerical methods and simulation time step effects on the prediction of building thermal performance. – Applied Thermal Engineering 24(8-9): 1129-1142.

[6] Fernández-García, M. S., Vidal-López, P., Rodríguez-Robles, D., Villar-García, J. R., Agujetas, R. (2020): Numerical Simulation of Multi-Span Greenhouse Structures. – Agriculture 10(11): 499.

[7] Gorgin, R., Luo, Y., Wu, Z. (2020): Environmental and operational conditions effects on Lamb wave based structural health monitoring systems: A review. – Ultrasonics 105: 106114.

[8] Ha, T., Kim, J., Cho, B.-H., Kim, D.-J., Jung, J.-E., Shin, S.-H., Kim, H. (2017): Finite element model updating of multi-span greenhouses based on ambient vibration measurements. – Biosystems Engineering 161: 145-156.

[9] He, K. S., Chen, D. Y., Sun, L.-J., Liu, Z. L., Huang, Z. Y. (2015): The effect of vent openings on the microclimate inside multi-span greenhouses during summer and winter seasons. – Engineering Applications of Computational Fluid Mechanics 9(1): 399-410.

[10] Khan, M. A., Khan, M. Z., Zaman, K., Naz, L. (2014): Global estimates of energy consumption and greenhouse gas emissions. – Renewable and Sustainable Energy Reviews 29: 336-344.

[11] Kim, R. W., Hong, S. W., Lee, I. B., Kwon, K. S. (2017): Evaluation of wind pressure acting on multi-span greenhouses using CFD technique, Part 2: Application of the CFD model. – Biosystems Engineering 164: 257-280.

[12] Lee, I.-B., Short, T. H. (2000): Two-Dimensional Numerical Simulation of Natural Ventilation in a Multi-Span Greenhouse. – Transactions of the ASAE 43(3): 745-753.

[13] Leonidopoulos, G. (2000): Greenhouse dimensions estimation and short time forecast of greenhouse temperature based on net heat losses through the polymeric cover. – Polymer Testing 19(7): 801-812.

[14] Mushtaq, R. (2011): Augmented dickey fuller test. – SSRN Electronic Journal.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_57375751

[15] Ogunlowo, Q. O., Akpenpuun, T. D., Na, W. H., Rabiu, A., Adesanya, M. A., Addae, K. S., Kim, H.-T., Lee, H.-W. (2021): Analysis of Heat and Mass Distribution in a Single- and Multi-Span Greenhouse Microclimate. – Agriculture 11(9): 891.
[16] Petrakis, T., Kavga, A., Thomopoulos, V., Argiriou, A. A. (2022): Neural Network Model for Greenhouse Microclimate Predictions. – Agriculture 12(6): 780.
[17] Rasheed, A., Kwak, C. S., Na, W. H., Lee, J. W., Kim, H. T., Lee, H. W. (2020): Development of a Building Energy Simulation Model for Control of Multi-Span Greenhouse Microclimate. – Agronomy 10(9): 1236.
[18] Rasheed, S., Khan, S. M., Ahmad, Z., Mustafa, G., Haq, Z. U., Shah, H., Ansari, L., Jatt, T. (2022): Ecological assessment and indicator species analyses of the Cholistan desert using multivariate statistical tools. – Pakistan Journal of Botany 54(2).
[19] Reichrath, S., Davies, T. W. (2002): Computational fluid dynamics simulations and validation of the pressure distribution on the roof of a commercial multi-span Venlo-type glasshouse. – Journal of Wind Engineering and Industrial Aerodynamics 90(3): 139-149.
[20] Sabzikar, F., Meerschaert, M. M., Chen, J. (2015): Tempered fractional calculus. – Journal of Computational Physics 293: 14-28.
[21] Tol, R. S. J., de Vos, A. F. (1993): Greenhouse statistics-time series analysis. – Theoretical and Applied Climatology 48(2-3): 63-74.
[22] Vapnik, V. N. (1995): The Nature of Statistical Learning Theory. – Springer, New York.
[23] Xue, Y., Fan, Y., Ge, J. (2021): Moisture distribution in wall-to-floor thermal bridges under the analysis of coupled heat and moisture transfer. – Proceedings Article.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_57375751