An Investigation of Canadian Greenhouse Climate Prediction using Time Series

Cheng Yuan[1, *, †], Jiaoyang Zhang[2, a, †], Leyan Zheng[3, b, †], Jingyu Zhu[4, c, †]

[1]McMaster University Hamilton, Canada

[2]University College London London, United Kingdom

[3]Jinling High school Hexi Campus Nanjing, China

[4]Chengdu Shude High School Foreign Language Campus Chengdu, China

*Corresponding Author Email: yuanc10@stu.cpu.edu.cn, [a]15036366813@163.com, [b]pikafishovo@163.com, [c]2561587229@qq.com

[†]These authors contributed equally.

Abstract. Due to the increasing global warming trend, the greenhouse effect is becoming more and more serious, which caused a very hot summer around the world. This study aims to identify the most important factors contributing to global warming by investigating and predicting greenhouse gas (GHG) emissions. There are not many studies to predict Canada's future GHG emissions. Therefore, It was decided to use the ARIMA model in a time series analysis to predict and simulate GHG emissions in Canada. A dataset containing seven different aspects of Canadian GHG emissions over the past 27 years was used. The result shows that overall emissions will continue to rise but the growth rate of GHG emissions will decline. Among them, Agriculture and Transportation are the two influencing factors that will increase GHG emissions the most in the future. In general, to reduce GHG emissions in the future, people need to live a low-carbon life and reduce unnecessary means of transportation or use more energy-saving resources such as electric cars. For agriculture, it takes less land to produce more food, such as hybrid rice in China. The fundamental reason is that the earth's population keeps increasing, people need more private cars to travel easily, and more food, thus increasing the area of arable land but reducing the area of green forest.

Keywords: ARIMA, greenhouse gas, time series

1. Introduction

In the era of the internet of things, theoretical and experimenting facilities have improved computational efforts by applying intelligent systems to the study topics [1, 2]. With climate change being a major concern for various nations, how maintaining the supply of food, fuel, and fiber for the world population has become a pressing need. Adverse weather conditions characterized by famine, irrational weather patterns, and poor agricultural yields made greenhouse production possibilities to be explored and expanded [3]. In order to ensure that these greenhouse productions are efficient, prediction models using time series have been utilized to ensure that the physical and physiological processes of greenhouse biophysics are understood. Considered a convenient method for maintaining the supply of food and fiber for life, intelligence agriculture shows a greater need for data collection.

According to Li et al., obtaining greenhouse sensors data in real-time therefore provides a clear prediction of the environment within the greenhouse so that the trends and development status can be used to correctly reinforce crop production and stable growth. Mathematical models used for simulation further reinforce this approach in that they predict climate control within the greenhouse with the primary focus being the determination of climate control requirements such as heating over a period [4, 5]. Speaking of which, this period can be predicted for a longer period of time meaning that a ten-year prediction can be done with simulations to show the efficiency of greenhouse production now and in the future. Monitoring and prediction further allow for the construction of effective greenhouses which align with the prediction model used [1, 2]. In essence, the population has to be fed with delicious and nutritious meals regardless of the season meaning that cultivating agriculture within a greenhouse has to take into account the different seasonal outlooks.

In addition, Chiu et al. while modeling and predicting the seasonal requirements, and statistical models are applied to improve crop outcomes by predicting resistance levels. In the recent past, Canadian agricultural research has focused on using time series data collected from greenhouses to advance integrated pest management practices, especially for whiteflies and aphids [6]. With time series, therefore, yields can be predicted depending on the resistance level of the various crops planted in the greenhouse. The technology of prediction also provides a clear insight into the evaporative cooling, natural ventilation, and irrigation control or shading processes in the greenhouse [1, 2]. For the analyzing user, therefore, considering any of these predictions for correlational purposes would provide multivariate time-series data for forecasting. In other words, the location within Canada, history, and check-in records of the different crops under greenhouse production in various conditions can be mapped. Mapping out these correlations gives access to regular or irregular crop growth patterns, growth stages, and need for different growth conditions [7]. At the end of the day, a nonlinear climatic prediction of the Canadian greenhouse can be obtained using time series data.

Notwithstanding, the long short-term memory (LSTM) time series prediction model can also be used in this case considering that the greenhouse climate does not only affect crop growth but also the cost of production cultivation and energy consumption forming the econometric aspect of time series investigations [8]. With these dynamics, Liu et al. assert that bootstrap resampling of data reinforces that paying attention to the changes in temperature is essential as this directly affects crop growth and the expected final yield from the greenhouse during that planting period or season. Basically, prediction models are collaborative in nature and since time series is used in this case, different matrix factorizations can be used depending on the end goal of the research or analysis. This simply indicates that the flow of information from the sensors encompassing time series data must be accurate so that the predictions can meet the consumption and social behavior of the Canadian people as proposed by the predictions under the captured environmental factors.

Since climate change can be diverse and dynamic, yield change is expected in some cases which would create an expected margin of error for different crops [9]. In as much as capturing the dynamics is essential in understanding and investigating greenhouse climate, optimization for proper process implementation cannot be overemphasized. Moreover, the period allocated for the investigation is also crucial because time will determine the historical data and contexts that will be taken into consideration and how far the predictions can be forecasted. It is clear therefore that time series investigation is essential in understanding greenhouse in Canada by capturing environmental factors and convolutions essential in predicting crop growth and stability in feeding the population while aligning with the consumption behavior of the people [8].

The primary aim of this investigation is to use time series sensor data in the greenhouse to monitor climate for better control of the present and future production under the greenhouse. The rationale behind this investigation is based on environmental stability maintenance and artificial control of climate change for the benefit of the Canadian population. Concurrently, the proposed model data would help in adjusting the greenhouse climate to ensure that different varieties of crops can be produced efficiently without necessarily altering the environmental factors while ensuring that soil moisture calibration data collected is suitable for crop production, growth, and development.

2. Methodology

An autoregressive integrated moving average (ARIMA) is a statistical analysis model that uses time-series data to understand data sets better or predict future trends. Future forecasts are obtained from historical data by fitting time series data into an ARIMA model. Theoretically, an ARIMA model consists of three components: the autoregressive (AR) model, the moving average (MA) model and the integrated (I) term, where the AR model has the order p and the MA model has the order q. An autoregressive moving average (ARMA) model is a combination of AR and MA, which is formulated as (1),

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + e_t - \theta_1 e_{t-1} \\ -\theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \tag{1}$$

where $\varnothing$ is a number strictly between −1 and +1, and θ is the weights. Series $\{Y_t\}$ is a mixed ARMA process of orders *p* and *q*, respectively, abbreviated as ARMA(*p*,*q*). If the time series $\{Y_t\}$ needs to be made stationary by *d*th difference $W_t = \nabla^d Y_t$, then the series $\{Y_t\}$ follows the ARIMA(*p*,*d*,*q*) process. In this study, *d* = 1 is sufficient to make the series stationary. Therefore, the ARIMA(*p*,1,*q*) process with $W_t = Y_t - Y_{t-1}$is formulated as (2).

$$W_t = \emptyset_1 W_{t-1} + \emptyset_2 W_{t-2} + \cdots + \emptyset_p W_{t-p} + e_t - \theta_1 e_{t-1} \\ -\theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \tag{2}$$

Firstly, through the analysis of the ARIMA model, the dataset that can be used for time series analysis needs to be established. So it was decided to collect data from the Meteorological Service of Canada database. The Meteorological Service of Canada has very detailed data on Canada. Data on climate temperature change, seasonal temperature change, precipitation, surface temperature, and greenhouse gas emissions were collected. These aspects can be used for time series analysis to predict the trend changes of various factors in the future and are very suitable for AR or ARIMA model after preliminary investigation. After comparison, finding GHG emissions is very interesting and fits the ARIMA model better than other aspects. Concerned that the world has been experiencing sweltering summers in recent years, Canada's GHG emissions are being examined to see if they could lead to a more severe greenhouse response in the future. Therefore, in this study, data on GHG emissions were used to predict changes in GHG emissions over the next few years. This dataset contains GHG emissions from 1990 to 2017 for seven different aspects, which are agriculture, buildings, electricity, heavy industry, oil and gas, transportation, waste and other. Seven different models were built to predict each of these seven aspects.

Then Use Eviews for time series analysis, prediction, and modeling. Through the study of references, the ARIMA model is very suitable for the impact of GHG on future climate. Therefore, the ARIMA model is established to predict GHG emissions of various sectors. Modeling and forecasting greenhouse Whitefly Incidence [10] used a similar ARIMA model, using similar judgment methods such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Hannan-Quinn Information Criterion (HQIC) [4].

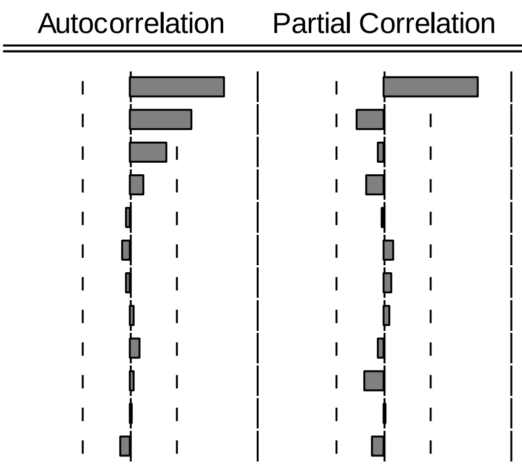
3. Results And Discussion

When the ARIMA model was built, the first step was to ensure the stationary of the series. This is because stable means and variances allow the data to be predicted more easily. Augmented Dickey-Fuller (ADF) test is used to test whether the series is stationary or not. The results show the existence of unit roots for all the series except agriculture, which means that none of them are stationary. After obtaining the first order difference of the six series, the absence of unit roots for all series means that they are all stationary now and meet the requirements for building an ARIMA model.

The second step is to determine which model to use by conducting drawing autocorrelation function (ACF) and partial autocorrelation function (PACF) of the series. The ACF ρ_k for a variety of lags *k* = 1, 2, …. shows the correlation among$(Y_1,\ Y_{1+k}),\ (Y_2, Y_{2+k}),\ (Y_3,\ Y_{3+k}), \cdots, and\ (Y_{n-k},\ Y_n)$, which is formulated as (3).

$$\rho_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2} \tag{3}$$

The PACF at lag k shows the correlation between Y_t and Y_{t-k} after removing the effect of the intervening variables $Y_{t-1},\ Y_{t-2},\ Y_{t-3}, \ldots,\ Y_{t-k+1}$. If PACF cuts off after lag *p* and ACF tails off to zero, then AR model should be adopted. On contrast, MA model should be adopted. If PACF and ACF all have their tails being cut down at a point, then ARIMA model is not appropriate for the series. If PACF and ACF all tails off to zero, then ARMA model should be established. If meanwhile the series needs to be differenced to achieve stationary, then ARIMA model is built up. An example of the ACF and PACF of agriculture is shown in Fig. 1, where the ACF trails off and the PACF cuts off. Therefore, the AR model should be used for prediction for the series of agriculture.



	AC	PAC	Q-Stat	Prob
1	0.776	0.776	18.734	0.000
2	0.516	-0.217	27.337	0.000
3	0.302	-0.046	30.392	0.000
4	0.109	-0.125	30.810	0.000
5	-0.021	-0.009	30.826	0.000
6	-0.054	0.087	30.939	0.000
7	-0.022	0.072	30.959	0.000
8	0.043	0.063	31.037	0.000
9	0.083	-0.043	31.338	0.000
10	0.043	-0.153	31.425	0.000
11	0.000	0.021	31.425	0.001
12	-0.070	-0.088	31.679	0.002

Figure 1. ACF and PACF of agriculture

The same steps are used for the remaining six series. The ARIMA model was used for all the series except for the agriculture and electricity series, where the AR model was used.

After determining the type of model, the order *p* and *q* of the model needs to be defined. AIC, BIC and HQIC were used to select the best *p* and *q* values. The first approach to determine the order is to select a model that minimize the AIC, which is defined as (4),

$$AIC = -2\log(maximum\ likelihood) + 2k \tag{4}$$

where $k = p + q + 1$ if the model contains an intercept or constant term and $k = p + q$ otherwise. Another approach is to minimize the BIC that is defined as (5),

$$BIC = -2\log(maximum\ likelihood) + klog(n) \tag{5}$$

where *n* is the sample size.

The results of the parameters are shown in Table I, where A means agriculture, B means buildings, E means electricity, HI means heavy industry, OG means oil and gas, T means transportation, W means waste and others. The letter D in front of some series means that these series have taken a first order difference to achieve stationarity (d = 1). The number in brackets after the model represent the order of this model, where the number after AR represents order *p* and the number after MA represents order *q*. For example, the model of buildings is ARIMA (3, 1, 3). From the *p* value, it is obvious that all the estimated coefficients of the selected models are significant, indicating that the models are all appropriate.

Table 1. ARIMA Establishment

Variable	Lag	Coefficient	Std. Error	t-Statistic	Prob.	R-squared	Adjusted R-squared
A	AR(1)	1.007	0.004	231.055	0.000	0.815	0.815
DB	AR(3)	0.557	0.157	3.553	0.002	0.328	0.297
	MA(3)	-0.903	0.040	-22.638	0.000		
DE	AR(3)	0.424	0.180	2.356	0.027	0.183	0.183
DHI	AR(2)	-1.044	0.047	-22.165	0.000	0.142	0.104
	MA(2)	1.000	0.084	11.838	0.000		
DOG	AR(1)	0.825	0.133	6.190	0.000	0.024	-0.017
	MA(2)	-0.601	0.187	-3.208	0.004		
DT	AR(2)	0.340	0.189	1.801	0.085	-0.296	-0.352
	MA(1)	0.429	0.197	2.176	0.040		
DW	AR(3)	-0.402	0.177	-2.271	0.033	0.342	0.312
	MA(3)	0.910	0.048	18.996	0.000		

The seven different ARIMA models with parameters (*p, d, q*) identified above were built and used to forecast the GHG emissions of seven different aspects over the next eight years. The results of the predictions are shown in Table II. (GHG emissions are measured in MtCO2e).

Table 2. Forecasting Results

Year	Agriculture	Buildings	Heavy Industry	Electricity	Oil And Gas	Transportation	Waste And Others
2018	72.535	85.515	74.273	75.040	203.731	173.856	41.600
2019	73.074	86.622	72.151	78.174	204.727	173.856	41.587
2020	73.617	85.199	71.302	77.089	205.548	173.807	42.649
2021	74.164	85.486	71.843	73.815	206.225	173.807	42.810
2022	74.715	86.104	70.942	74.949	206.784	173.790	42.815
2023	75.270	85.310	70.582	78.367	207.244	173.790	42.388
2024	75.829	85.470	70.811	77.183	207.624	173.785	42.324
2025	76.393	85.815	70.429	73.614	207.937	173.785	42.322

Combining the previous known data from 1990 to 2017, a graph of the overall trend of GHG emissions in seven sectors is presented in Fig. 2. As can be seen from the graph, oil & gas and transport will remain the two largest contributors to GHG emissions in our forecast years, far outstripping emissions from other sectors. Whereas emissions from oil & gas will continue to rise, emissions from transport will remain the same as in 2017. For the remaining five emissions, emissions from agriculture are expected to increase slightly, emissions from heavy industry will decrease slightly, while emissions from buildings, electricity and waste & others will continue to fluctuate with no significant increase or decrease.

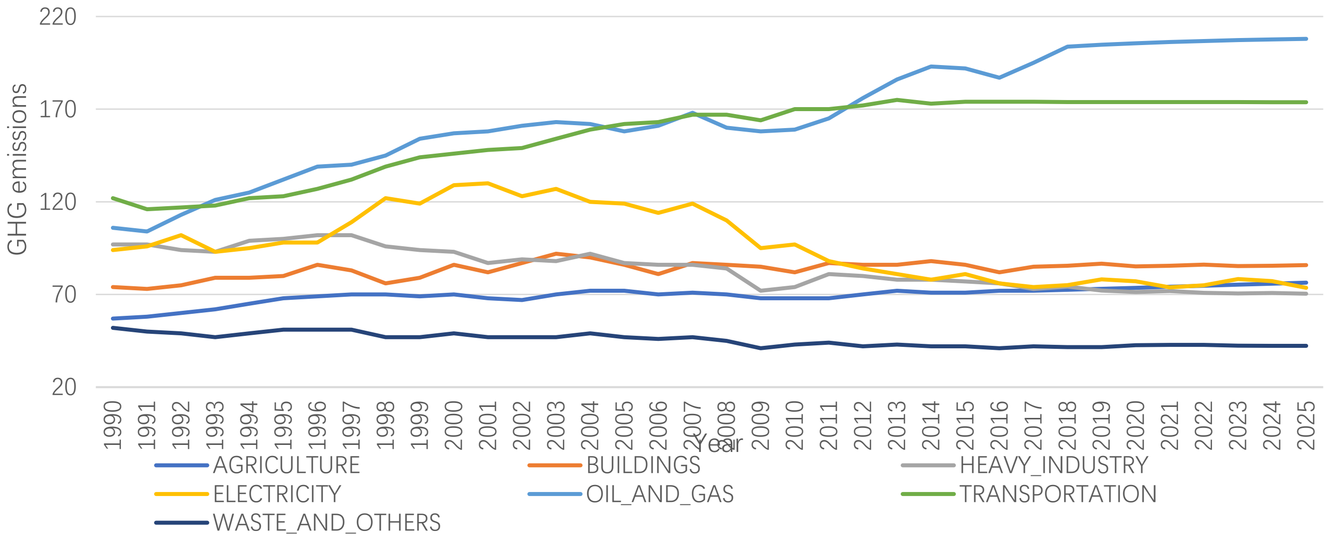


Figure 2. GHG emissions trend of 7 sectors (1990-2025)

When further observation is made of the future trends of each indicator, similarities and differences will be found. To be more specific, as shown from Fig. 3, GHG emissions from heavy industry and transport are both expected to decline gently between 2018 and 2025, albeit with small fluctuations, that is, from 74.273 to 70.429 and from 173.856 to 173.785 separately, which may mean that in the future these two GHG will have less and less effect on temperature.

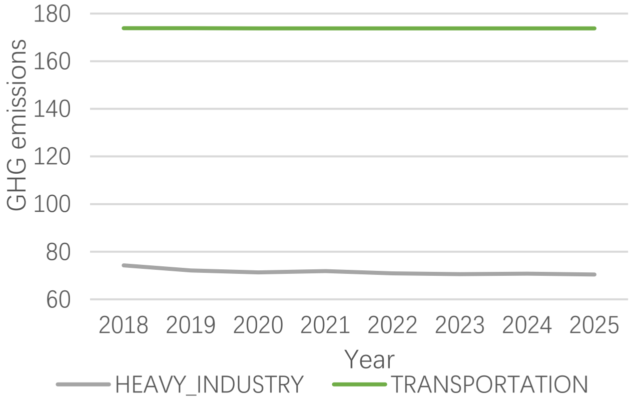


Figure 3. Comparison of GHG emissions from heavy industry and transportation

However, the trend of agriculture and oil and gas is in reverse. Fig.4 shows the gentle upward trend of agriculture and oil and gas in the next eight years. Both emissions of GHG will increase by about 4, which may indicate that agriculture and oil and gas are still increasing impact on temperature.

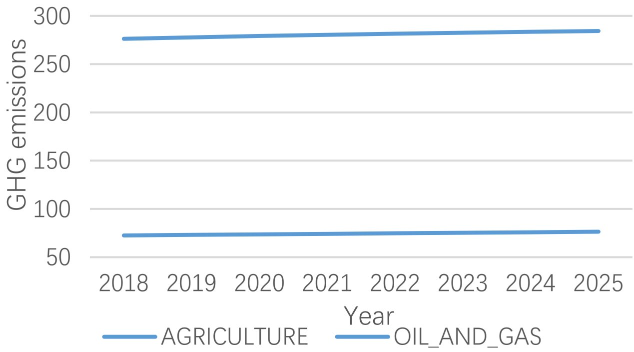


Figure 4. Comparison of GHG emissions from agriculture and oil & gas.

As to buildings, electricity and garbage and other sectors (Fig. 5), instead of a clear upward or downward trend, it is replaced by some fluctuations from 2018 to 2025. Even so, the changes in emissions from all three indicators are modest. It seems that the effect of buildings, electricity and garbage and other sectors on temperature is essentially constant in the future.

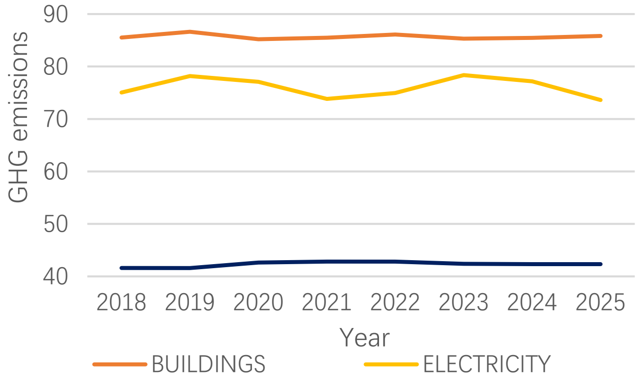


Figure 5. Comparison of GHG emissions from buildings, electricity and waste & others.

All in all, some GHG emissions are going to rise in the future, and some are going to fall. Research results show that oil and gas's GHG emissions in the past and future are high and there is a rising trend so Canada was proposed to diversify its energy sources by using more clean energy such as biofuels, solar and wind, which would greatly reduce the use of oil and gas to reduce its GHG emissions. Similarly, although GHG emissions from transportation are expected to decrease slightly in the future, they are still high. To mitigate this, Canada needs to improve the fuel efficiency of all types of transportation, for instance, vehicles, aircraft and so on, improve the efficiency of the transportation system and invest in public transportation. What's more, as to electricity, its greenhouse-gas emissions have fallen in recent years and Canada may need to establish electricity capacity markets that provide revenue for standby fossil fuel capacity without the need for continuous generation or use hydroelectric power and wind power to generate electricity to bring its emissions down even further. There are also some policies that Canada could adopt to reduce the GHG emissions, such as encouraging people to live in more energy-efficient houses, charging for "emissions" and so on. In addition, the GHG emissions of a city depend largely on its location, shape and the technology it possesses. However, this paper only considers the GHG emissions of the whole country of Canada in recent years and the prediction of the future, which lacks the comparison between its cities. So further subdivision and research are needed in the future.

4. Conclusion

Overall, Canada's total GHG emissions will show an increasing trend until 2025, but the growth rate of total emissions will slowly decline. Gas, oil, and transportation are among the biggest GHG emitters. For oil and natural gas, people need to find clean energy such as wind and solar energy that can replace oil and natural gas. Improving the use of oil and gas will help reduce GHG emissions. For transportation, people need to try to integrate low-carbon travel into their lives, reduce the use of fuel vehicles and vigorously develop new energy vehicles powered by clean energy. To sum up, one of the reasons for this phenomenon and result is the increasing global population. As the population of each country increases, the use of private cars will increase. As the population increases, both the use of electricity and fossil fuels will increase greatly. All these contribute to the increase in GHG emissions and global warming. The study could provide insight into the factors contributing to more GHG increases in Canada over the next three years. This can also provide the Canadian government with specific requirements on how to adjust the use of transportation vehicles, the arrangement of industrial plants, and the protection of green forests and plants to reduce GHG emissions in the future. In this study, prediction modeling was only conducted from the ARIMA model, without modeling from various aspects such as ARMA, AR, and MA models. If seasonal projections can be made, it is possible to know which quarter in Canada produces the most emissions.

References

[1] Y. Liu, D. Li, S. Wan, F. Wang, W. Dou, X. Xu, S. Li, R. Ma, and L. Qi, "A long short-term memory-based model for greenhouse climate prediction," International Journal of Intelligent Systems, 2022, 37(1), pp.135-151.

[2] J. Li, L. Sun, Q. Yan, Z. Li, W. Srisa-An and H. Ye, "Significant permission identification for machine-learning-based android malware detection," IEEE Transactions on Industrial Informatics, 2018, 14(7), pp.3216-3225.

[3] P.E.N.M. Momani and P.E. Naill, "Time series analysis model for rainfall data in Jordan: Case study for using time series analysis," American Journal of Environmental Sciences, 2009, 5(5), p.599.

[4] A. Hasni, Z. Chikr-el-Mezouar, B. Draoui and T. Boulard, "Applying time series analysis model to temperature data in greenhouses," Sensors & Transducers, 2011, 126(3), p.119.

[5] A. Hasni, B. Draoui, T. Boulard, R. Taibi, B. Dennai, "A particle swarm optimization of natural ventilation parameters in a greenhouse with continuous roof vents," Sensors & Transducers, 2009, 102(3), p.84.

[6] A. Al-Sakkaf, and G. Jones, "Comparison of time series models for predicting campylobacteriosis risk in New Zealand," Zoonoses and public health, 2014, 61(3), pp.167-174.

[7] J.D. Cryer and K.S. Chan, "Time series analysis: with applications in R," (Vol. 2). New York: Springer. 2008.

[8] S. Khaki, L. Wang, and S.V. Archontoulis, "A cnn-rnn framework for crop yield prediction," Frontiers in Plant Science, 2020, 10, p.1750.

[9] D.J.A. Rustia and T.T. Lin, "An IoT-based wireless imaging and sensor node system for remote greenhouse pest monitoring," Chemical Engineering Transactions, 2017, 58, pp.601-606.

[10] L.Y. Chiu, D.J.A. Rustia, C.Y. Lu and T.T. Lin, "Modelling and forecasting of greenhouse whitefly incidence using time-series and ARIMAX analysis," IFAC-PapersOnLine, 2019, 52(30), pp.196-201.