財團法人國際合作發展基金會
第 19 屆外交替代役役男專題報告

題目：機器學習應用於田間微氣候之即時預報

役男：何宇峰
指導技師：黃凱煜技師
服勤單位：駐聖克里斯多福及尼維斯技術團
　　　　　農業因應氣候變異調適能力提升計畫

中文摘要

　　精準、智慧農業，是掌握作物栽培技巧以及充分利用田間各項農業信息，提高農業韌性及減緩天然災害所帶來的影響，使得農業系統的運轉更加有效，同時具有可持續性。農業氣象資料的獲取在其中扮演重要的角色，特別在聖克里斯多福及尼維斯，農民傾向栽植溫帶、熱敏感的作物如番茄及甜椒，田間溫度的監控尤其重要。由於當地基本建設不易，資材昂貴，廣設農業氣象站的效益不高。

　　因此，本研究引入機器學習機制，透過島上四座已建設的農業氣象站，搭配不同微氣候下設置的溫濕度感測器，建立機器學習模型以預估各個微氣候環境下的溫度，作為高溫預警的依據。實驗模型有三個，其一是預測有屋頂之鐵絲網室的溫度，其二是預測有屋頂之防蟲網室內的溫度，其三是預測半徑三公里內之無遮罩戶外環境的溫度。

　　估計模型顯示，依照三個不同的環境，模型的平均預測溫度誤差分別落在 ±1.1、±2.2 及 ±1.3，並且 R^2 分別為 0.78、0.61 及 0.69。較僅以溫度進行簡單線性迴歸的模型中，各個模型都有至少提升 10%的預測準確性，而在模型一和模型三中，高溫超過 40°C的情況下，預測值會有低估的情況發生。

　　本研究對於克島農業計畫有加分的效果，透過收取即時氣象資料進行田間實際氣溫的預測，結合農業信息的推播，提升當地農民對環境溫度的掌握度，達到有效的農業預警，減輕氣候變異對農業的影響。

Abstract

This study aims to utilize the statistical methods for the prediction of the realistic temperature in the target locations. It reaches the goals of promoting smart agriculture, costing down the expense of farmers. It introduces the multivariate polynomial regression and LOASS regression to predict the temperature of outdoor fields and the varied types of cultivated house. As a result, the predictive models succeed to improve grasping the microenvironment in farmers' fields.

Introduction

Precision, smart agriculture provides farmers with high control of input and output. It emphasizes to optimize resources, improve productivity, and reduce waste released in a clever way. To practice this type of cultivation, farmers have to know the parameters and planting methods related to their crop growth. On the one hand, the cultivated manuals are useful for farmers to follow for dealing with general problems. On the other hand, controlling environmental factors is prominently important. Climate is one of the most important environmental parameters. More precisely farmers could control, more effectively farmers could manage their fields. At this point, the microclimate real-time monitoring system is usually used as a tool for obtaining such information[1].

Nevertheless, the monitoring system basically correlated with high costs. The farmers who practice precision agriculture usually have to pay the relevant high price for apparatuses and simultaneously, they need to possess the professional skills about device maintenance. The high threshold erects the barrier for governments or agricultural organizations to promote and put precision, smart agriculture into practice in their country lands.

Generally, farmers are used to following the meteorological data from communal weather stations, but the variation of air temperature is influenced by other meteorological factors such as humidity and solar radiation. However, the correctness of air temperature reflecting the real-time temperature in the field is always doubtful because the hardware of the weather station is under radiation shield and the outdoor environment is not. Apparently, without shield, air temperature in the field increases more dramatically, especially at noon. With the high-intensity radiation, the temperature detected by the weather station would usually underestimate the realistic temperature in the field. Thus, little underestimation would not have a significant impact on most of the crops, but for the heat-sensitive crops growing in tropical areas like sweet pepper and tomato, the ignorance of heat-stress would lead to an unexpected outcome and directly cause productivity loss.

Machine learning is an application of artificial intelligence that provides systems to automatically improve from experience, in other words, data. The universality of the concept leads

it to apply in the most data-based domain including meteorology. Supervised machine learning tools nowadays are usually applied for weather forecasting in many aspects, such as the prediction of solar irradiation depending on weather variability[2]. Regarding the agricultural domain, the application of machine learning provides the farmers with empirical evidence to grasp the microenvironmental information in their fields.

Overall, precision agriculture required resources and funds to establish a data collecting system. Thus, in some countries, the establishment of infrastructure is not profitable due to the high price of maintenance and materials. In this situation, machine learning displays its talent for computationally cheap inference. This study would leverage supervised machine learning method, try to nowcast the realistic, microenvironmental temperature over outdoor fields and different kinds of cultivated house, and ultimately embrace the concept of precision, smart agriculture in an intellectual way.

Materials and methods

1.Meteorological apparatuses and Data

The meteorological data is collected by weather stations scattering throughout Saint Kitts and Nevis. The three main stations for this study are situated at Needmust farm, Ecopark, and Mansion and they are at different altitudes (Fig. 1). They represent climatic characteristics on the different sides of the island. Mansion station gets northern wind, Needmust station gives general meteorological information in the urban area, and Ecopark station covers the eastern area of the island. There is a temporary weather station established in La Guerite, Basseterre. Moreover, there are three types of cultivated house in the same place, which is installed mounted sensors. The information on the stations and sensors are listed in Table 1.

Fig. 1: The distribution of weather stations in Saint Kitts and Nevis. The three main stations are Needmust, Ecopark and Mansion ordered from altitude.

Table 1: All of the meteorological apparatuses in Saint Kitts. Agrometeorological weather stations (AWS) are from Campbell Scientific and are able to collect multiple measurements, and mounted sensors (MS) are from HOBO and are able to show the basic information in particular environments.

Device ID	Company	Place	Category	Measurements
1	Campbell Scientific	Needmust (altitude: 0m)	Agrometeorological weather station	Air temperature, humidity, atmospheric pressure, volumetric water content, soil temperature, soil electrical conductivity, reference evapotranspiration, clear-sky radiation, solar radiation (kW), solar radiation (MJ), wind speed, wind direction, and standard deviation of wind direction
2	Campbell Scientific	Ecopark (altitude: 100m)	Agrometeorological weather station	Same as 1
3	Campbell Scientific	Mansion (altitude: 200m)	Agrometeorological weather station	Same as 1
4	Campbell Scientific	La Guerite	Agrometeorological weather station	Air temperature, humidity, solar radiation, solar radiation, rainfall, volumetric water content and soil temperature
5	HOBO	Needmust: Outdoor	Mounted sensor	Air temperature, humidity and dew-point temperature
6	HOBO	La Guerite: Wire-net house	Mounted sensor	Same as 5
7	HOBO	La Guerite: shading-wire-net house	Mounted sensor	Same as 5
8	HOBO	La Guerite: Insect-screen-net house	Mounted sensor	Same as 5

2. Methodology and anaylsis processes

This study is separated into two parts: (1) data collecting and pre-processing, and (2) the different steps of machine learning.

The diagram of the data source, pre-processing, and pretest is given by Fig. 2. Data collection is done by the stations and sensors mentioned in the former paragraph. Pre-processing creates a readable data frame for the following analysis and reduces the data size. The purpose of the pretest is to prove the difference between temperature collected from weather stations and sensors.
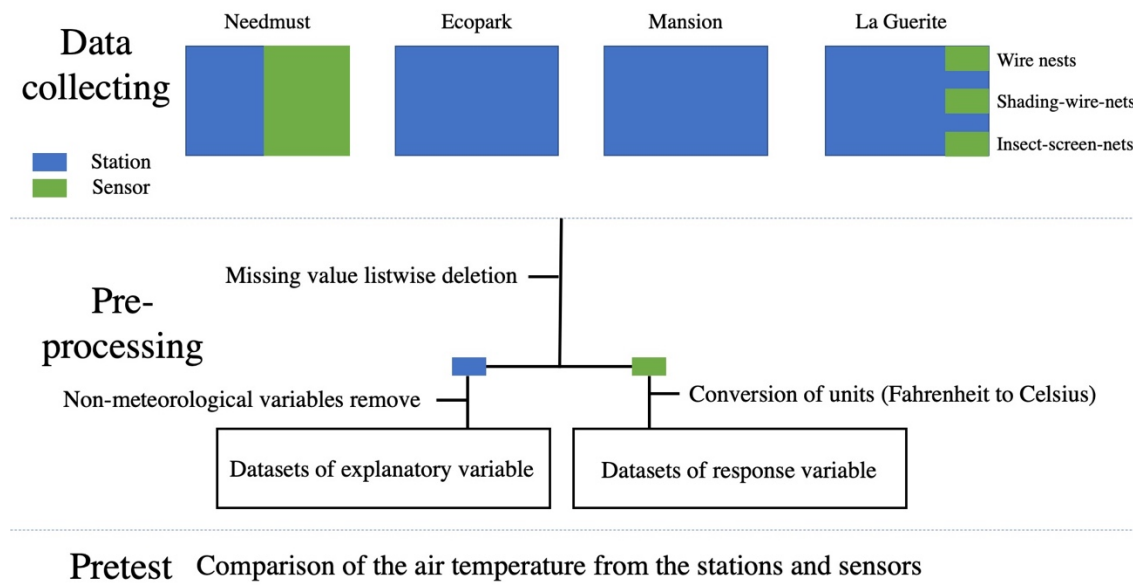

Fig. 2: The digram of the preparation of machine learning

Fig. 3 summarizes the outline of the machine learning process and the detail of the used methodologies is in the following paragraphs. The process begins with merging agrometeorological weather stations' data with mounted sensors' data in the same or different locations. Forward selection generates the preliminary models, and LASSO regression is introduced for model regularization if collinearity exists. Afterward, the generated models would be evaluated the predictive power. The final models can be implemented, and use the real-time information form the three agrometeorological weather stations to nowcast the microenvironmental temperature.
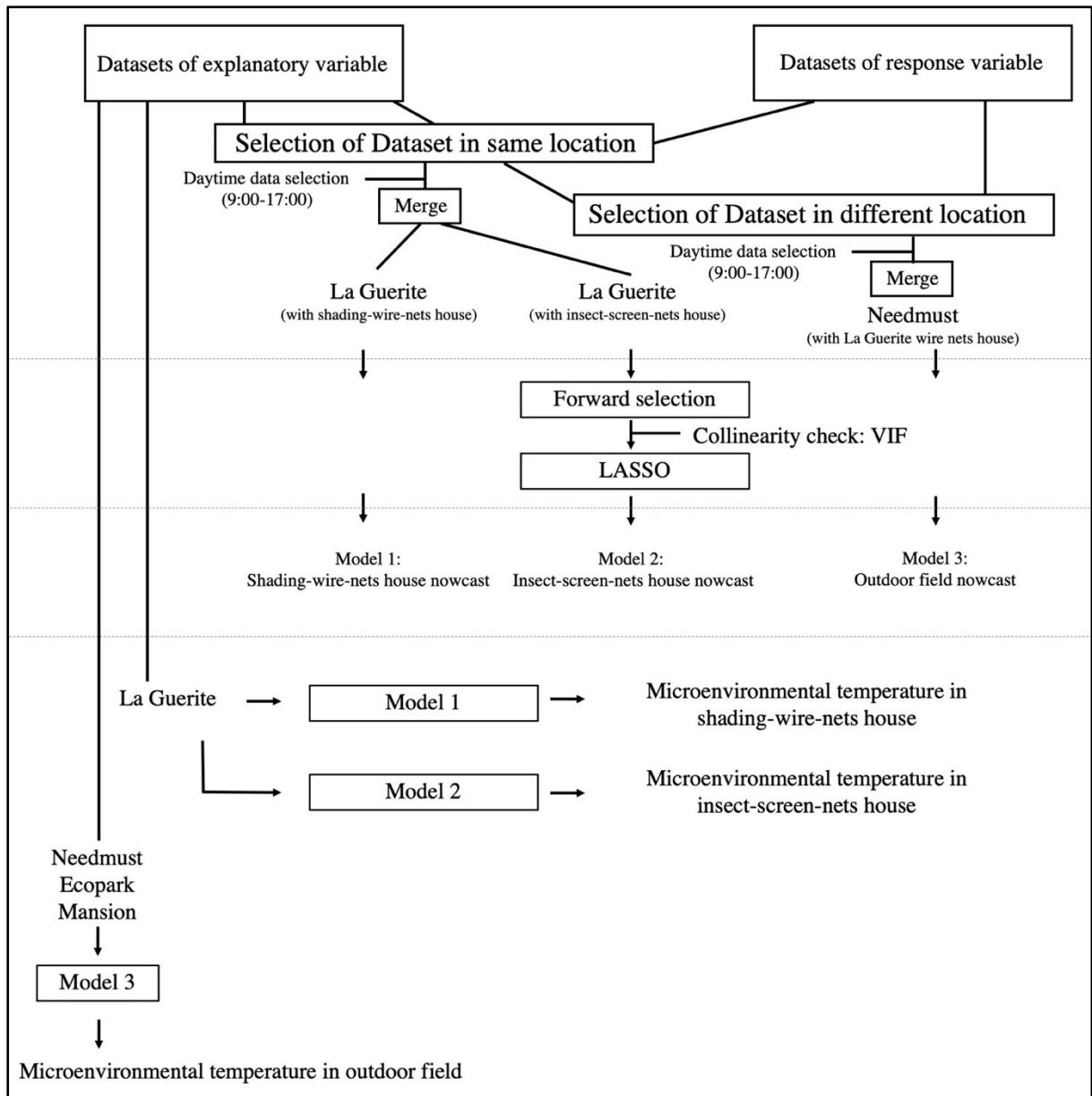
Fig. 3: The process flow digram of machine learning

3. Analysis tools

3.1 Forward selection

The forward selection as a model selection approach is a type of stepwise regression that begins with an empty model and adds in variables one by one. In each iteration, once the variable has been selected, the new model would be evaluated on the basis of certain criteria and compare with the former model. The procedure stops, when no other variables are left that meet the entry criterion. In this study, we used Bayesian Information Criterion (BIC) as our main criterion and complexity parameter (Cp) as our auxiliary.

3.1.1 Bayesian Information Criterion (BIC)

A criterion for model selection among a finite set of models; the model with the parameters by introducing a penalty term for the number of parameters in the model.

$$-2 \cdot logL\left(\hat{\beta}_{ML}, \hat{\sigma}^2{}_{ML}\right) + (\log n)p \tag{1}$$

$L\left(\hat{\beta}_{ML}, \hat{\sigma}^2{}_{ML}\right)$ represents likelihood function, $\hat{\beta}$ is the estimated beta coefficients from each iteration of regression, and $\hat{\sigma}^2$ is the unbiased estimate of residual sum of square (RSS). For the penalty term, n represents the number of observations and p is the number of parameters in the model, namely selected explanatory variables. By integrating a penalty term depending on the number of independent parameters, BIC tends to choose a parsimonious model[3].

3.1.2 Mallow's complexity parameter (Mallow's Cp)

When using the Cp criterion, the subset of explanatory variables for which

(1) the Cp value is small and (2) the Cp is close to p should be chosen. It compares the full model and the models with some absent predictors. It could help to choose the optimal selection of explanatory variables with less bias and precision, and it usually used in stepwise regression as a stopping rule.

$$C_p = \frac{SSE_p}{MSE_{full}} - (n - 2p) \tag{2}$$

This measure assumes that MSE_{full} is an unbiased estimator of σ^2, and SSE_p is the error sum of squares for the model with p parameters. The criterion gives a compromise influenced by the number of observations n and parameters p.

3.2 Least absolute shrinkage and selection operator, LASSO

Least absolute shrinkage and selection operator (LASSO) used in this study to generalize the selected best subset model, in order to lower the impact of collinearity and simultaneously deal with overfitting. After selecting the regressors by best subset selection, LASSO is introduced, adding a factor of sum of absolute value of coefficient for slight adjustment of residual sum of square.

$$\widehat{\beta}^{\,lasso} = arg \min_{\beta}\{RSS(W) + \lambda \times (sum\ of\ absolute\ value\ of\ coefficients)\}$$

$$= arg \min_{\beta}\left\{\sum_{i}^{n}\left(y_i - \beta_0 - \sum_j x_{ij}\beta_j\right)^2 + \lambda \sum_{j=1}^{p}|\beta_j|\right\}^2 \tag{3}$$

LASSO is combined by the formula of RSS and penalty term, called L1 regularization. The tuning parameter lambda (λ) controls the strength of the penalty. When λ increases, the bias increase, and coefficients would be shrunk more and even eliminated. In less important variables, the coefficients will go all the way to zero. Through cross-validation, the best λ would be calculated before going to minimize $\widehat{\beta}^{\,lasso}$ for each regressor, and the predictive model finally adopts $\widehat{\beta}^{\,lasso}$ as its coefficients.

3.3 Verification

To evaluate the adequacy of models or the performance of accuracy, every process requires different statistics.

3.3.1 Root Mean Square Error (RMSE)

RMSE is the standard deviation of the residual which is commonly used in forecasting and regression analysis.

$$RMSE = \sqrt{\frac{1}{n}\sum e_t^2},\ t = 1, 2 \dots \dots n, n = observation\ number \tag{4}$$

It has a direct relationship with the correlation coefficient. If the correlation coefficient is 1, the RMSE will be 0. Which means all the point is on the regression line.

3.3.2 Mean absolute error (MAE)

By taking absolute value of the errors, it measures the average magnitude of the errors in a set of predictions.

$$MAE = \frac{1}{n}\sum|e_t|\ \ t = 1,2 \dots \dots n, n = observation\ number \tag{5}$$

To sum up the difference between RSME and MAE. RMSE is inclined to aim at arithmetic mean gives the weight to errors which would lead to sensitivity to outliers. MAE aims at the median and it is able to ignore outliers. Nevertheless, if the predictive model has an obvious bias, MAE will relatively lack adequacy.

3.3.3 variance inflation factor (VIF)

VIF quantifies how severe multicollinearity in an ordinary least square regression analysis. Which tests the single variable with the other covariates using auxiliary regression.

$$(VIF)_i = \frac{1}{1-R_i^2} \ , \quad (VIF)_i = 1, \ R_i^2 = 0, \tag{6}$$

where R_i^2 denotes the coefficient of multiple determination when X_i is regressed on remaining explanatory variables.

For serious collinearity:

$$\max_i (VIF)_i > 10 \tag{7}$$

Based on the results of VIF, for serious collinear cases, we would use LASSO to update the predictive model.

4. Others

4.1 Type of cultivated house

There are three types of cultivated houses in La Guerite. Owing to the difference of architecture and covering material, this study assumes they might contain a unique microenvironment inside. The information lies in Table 2 and the outward appearances are in Fig. 4.

Table 2: The information of structure and the expected situation inside the three cultivated houses

Type	Surrounding	Roof	Expected Temperature (comapre with outdoor)
Wire-net house	barbed wire fence	No	Similar
shading-wire-net house	barbed wire fence	Yes	Similar
Insect-screen-net house	Insect net	Yes	Higher

(a) Wire-nets house   (b) Shading-wire-nets house   (c) Insect-screen-nets house
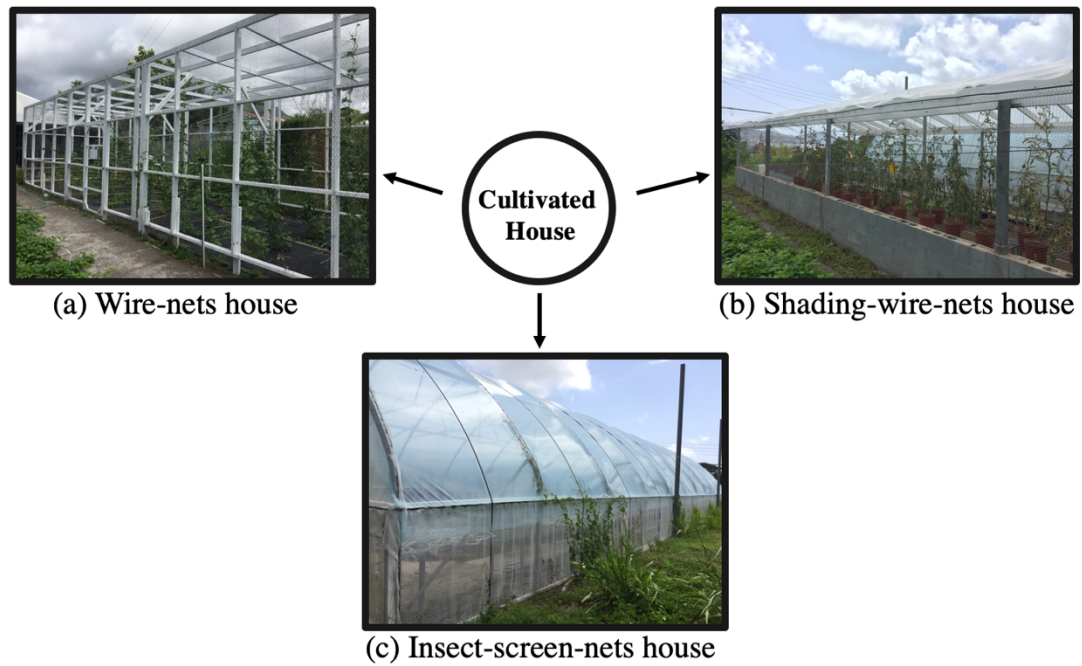
Fig. 4: The outward of three cultivated houses

Result

1. Pretest

1.1 Difference of detected temperature from apparatuses

The data collected from the agrometeorological weather station and the mounted stations have different patterns. Influence by factors such as solar radiation, humidity, and rainfall, the temperature collected from two apparatuses turned out to be different. This study choose 2020/02/02 as the example and compare the fluctuation of air temperature in two apparatuses (Fig. 5). As a result, since 6:00 in the morning, the air temperature started to increase. From 9:00 to 15:00, mounted sensors report higher temperatures than agrometeorological weather station. Moreover, the sensors return the temperature above 40°C sporadically which is regarded as heat stress for heat-sensitive crops. From the overall perspective, MSs detected higher temperatures. The graphs also indicate that the MSs returned rather different highest temperature at noon. From (a) to (c) is about 36°C, 40°C, and 45°C respectively.
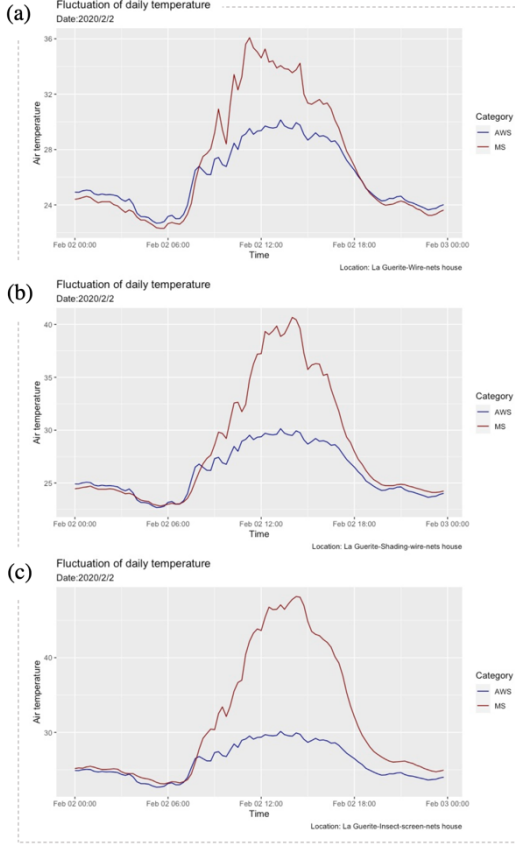
(a)

(b)

(c)

Fig. 5: Line graphs of the fluctuation of daily temperature compared between agrometeorological weather station (AWS) and mounted sensor (MS) among three cultivated houses in La Guerite: (a) Wire-nests house, (b) insect-screen-net house (c) shading-wire-net house.

1.2 Local temperature in different environments

Concerning different locations, they present individual characteristics of air temperature. This study chooses 2020/03/10 as the example and display the fluctuation of air temperature in different environments (Fig. 6). The insect-screen house shows an extraordinarily high temperature at noon. Furthermore, the mounted sensor inside returns temperature above 45°C which we infer that it might be caused by the greenhouse effect.
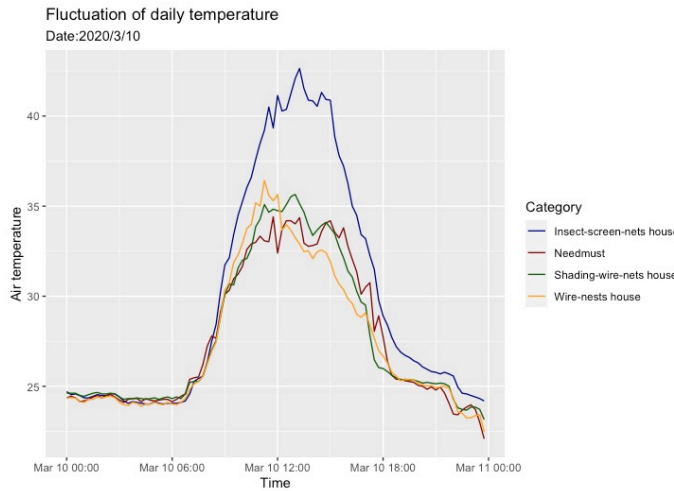

Fig. 6: Line graph of the fluctuation of daily temperature in four places. It displays the collected temperature from mounted sensors in four places. The temperature in the insect-screen-net house exceeds the others by approximately 5°C at noon and the others had irregular fluctuation during the day.

Overall, these pretests inform that the microenvironments are significantly different attributed to apparatuses and also surroundings. Therefore, our subsequent machine learning processes would be important and helpful.

2. Establishment of prediction model

2.1 Check of data size for prediction models

Dataset has two types: agrometeorological weather station and mounted sensor. Thus, the establishment of models only used datasets from two agrometeorological weather stations and four mounted sensors, and the total training data set depending on the period of collection. Furthermore, owing to the processes of missing value deletion and different dataset merging, the different models have a slight difference among data size. Table 3 summarizes the information of data size for each

prediction model.

Table 3: After pre-processing, the cleaned data for three models have different data sizes

| Model	Data collecting period	Number of validated data
Shading-wire-net house nowcast	2020/01/29-2020/06/02	3971
Insect-screen-net house nowcast	2020/01/29-2020/05/03	3012
Outdoor field nowcast	2020/06/11-2020/06/02	8484

2.2 Model selection: Forward selection

2.2.1 Selection of explanatory variables

Based on the result of pretest, the data used for prediction is determined to select from 9:00 to 17:00 every day. The forward selection method is conducted to select the explanatory variables from weather stations. BIC and Mallow's Cp are used as criteria and the selecting processes of three models is displayed in Fig. 7

(a) Model: shading-wire-nets house nowcast

BIC
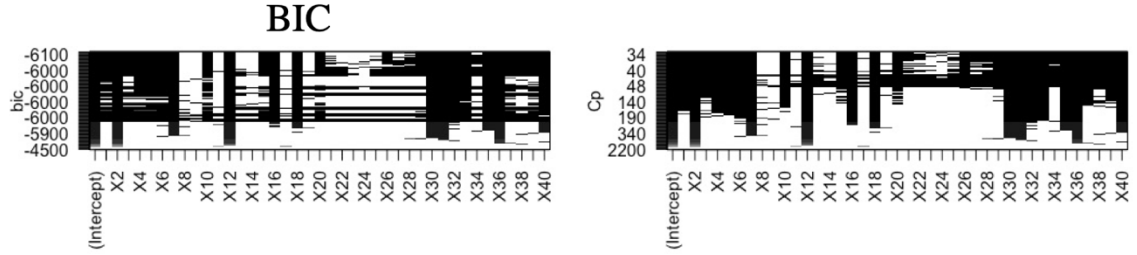
(b) Model: insect-screen-nets house nowcast
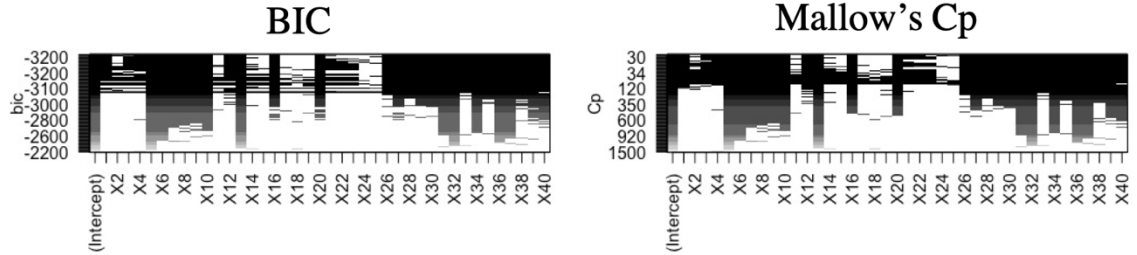
BIC   Mallow's Cp

(c) Model: outdoor field nowcast

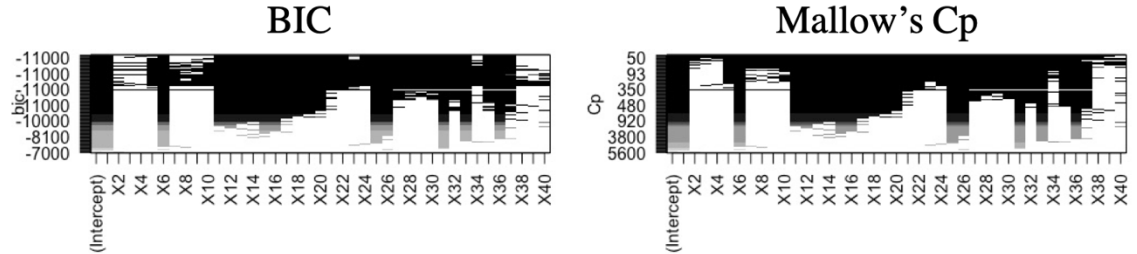
BIC   Mallow's Cp

Fig. 7: A plot of selecting process. The left-hand side is based on BIC for selecting, and the right-hand side is based on Mallow's Cp. The black bar presents that this variable has been included.

2.2.2 Cross validation for candidate models

Using the forward selection determines the top three candidate models. Cross-validation separates the dataset into 10 folds and applied RMSE, R-square and MAE to evaluate the candidate models. The result is sorted out in Table 4.

Table 4: The standard of verification in candidate models for three respective prediction models. The row named formula indicates the regression formula according to the best candidate models. The yellow labelling columns are belonged to the best models among the candidate models.

	Shading-wire-net house nowcast			Insect-screen-net house nowcast			Outdoor field nowcast		
RMSE	1.428	1.418	1.420	18.24	17.78	17.14	1.533	1.532	1.532
Rsq	0.7881	0.7912	0.7907	0.5909	0.6024	0.5951	0.7340	0.7341	0.7340
MAE	1.090	1.088	1.089	2.941	2.911	2.876	1.173	1.173	1.175
Formula	See Appendix (a)			See Appendix (b)			See Appendix (c)		

2.3 Model modification

After forward selection, the models are powerful to predict, and cross-validation already resolves the big part of the overfitting problem. Nevertheless, collinearity is also an issue contributing to overfitting in regression analysis models.

2.3.1 Collinearity check

Variance inflation factor (VIF) can determine the extent of explanatory variables that correlate to each other. The result of applying VIF to the models is sorted out in Table 5.

Table 5: The result of VIF test for all explanatory variables in the models. * symbol notes the variables which may exist collinearity.

Model	Explanatory variables	VIF (*: collinearity)
Shading-wire-net house nowcast	Air temperature	4.91
	Humidity	2.67
	Solar radiation (kW)	1200 *
	Solar radiation (MJ)	1199 *
	Volumetric water content	4.27
	Soil electrical conductivity	3.00
	Soil temperature	2.72
Insect-screen-net house nowcast	Air temperature	4.02
	Humidity	2.79
	Solar radiation (kW)	919 *
	Solar radiation (MJ)	918 *
	Rainfall	1.07
	Volumetric water content	2.00
	Soil electrical conductivity	1.65
	Soil temperature	2.19
Outdoor field nowcast	Air temperature	3.18
	Humidity	1.84
	Volumetric water content	11.10 *
	Soil electrical conductivity	10.28 *
	Soil temperature	2.50
	Solar radiation (kW)	4.17
	Solar radiation (MJ)	4.83
	Wind speed,	1.09

According to Table 5, the presence of collinear dependencies is detected in all models. Therefore, the model regularization is necessary to be used for eliminating the variance inflation of regression coefficients.

2.3.2 Regularization of regression model: LASSO regression

As the situation of collinearity, there are two ways to cope with it. One is to remove the inflating variables and the other is to add regularization. However, to avoid removing too much variables from the dataset, this study introduces LASSO regression for eliminating the unuseful variables.

By cross-validation, the best λ has been calculated as about 0.0025 in shading-wire-net nowcasting model, 0.0016 in insect-screen-net nowcasting model and 0.001 in outdoor field

nowcasting model. Taking advantage of λ, this study generates the new prediction model (Appendix (d)(e)(f)).

3. Predictive power evaluation

The nowcasting models now are able to generate predictive values representing air temperature in a certain microenvironment, and the standard of verification towards the models are summarise in Table 6.

Table 6: The standard of verification in the final nowcasting models.

	Shading-wire-net house nowcast	**Insect-screen-net house nowcast**	**Outdoor field nowcast**
RMSE	1.459	2.89	1.669
Rsq	0.779	0.606	0.685
MAE	1.113	2.23	1.270

We choose the same day as a pretest (2020/02/02) to draw a new line based on predictive values, in order to demonstrate the improvement of leveraging the nowcasting models (Fig. 8).
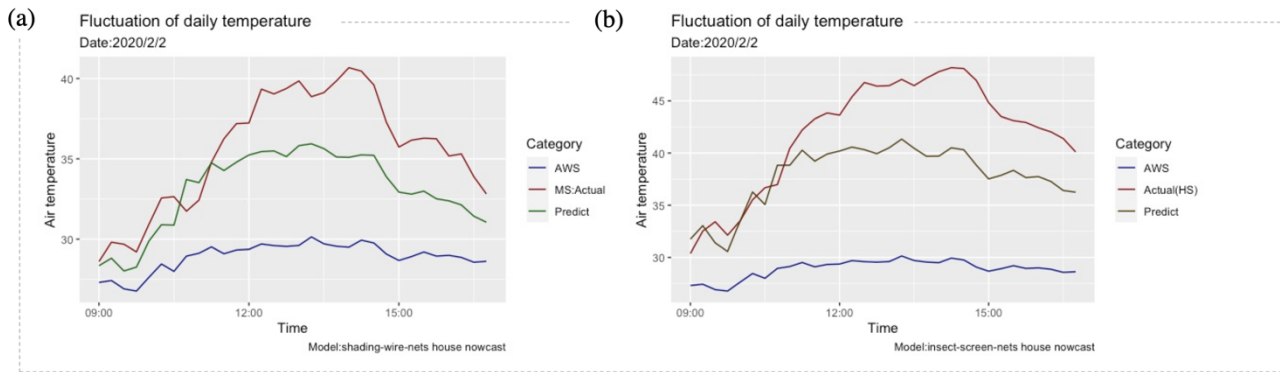

Fig. 8: Line graphs of fluctuation of daily temperature. Based on the pretest result, the figure zooms in the day period of 9:00-17:00 and adds a new line representing predictive values for demonstrating the improvement of using predictive models.

Secondly, the outdoor nowcasting model uses the AWS data from Needmust to predict the microenvironmental temperature in La guerite wire-net house regarded as an outdoor environment. Referring to the MS data from La guerite wire-net house, Fig. 9 demonstrates the predictive power of the model.

Fluctuation of daily temperature
Date:2020/2/2
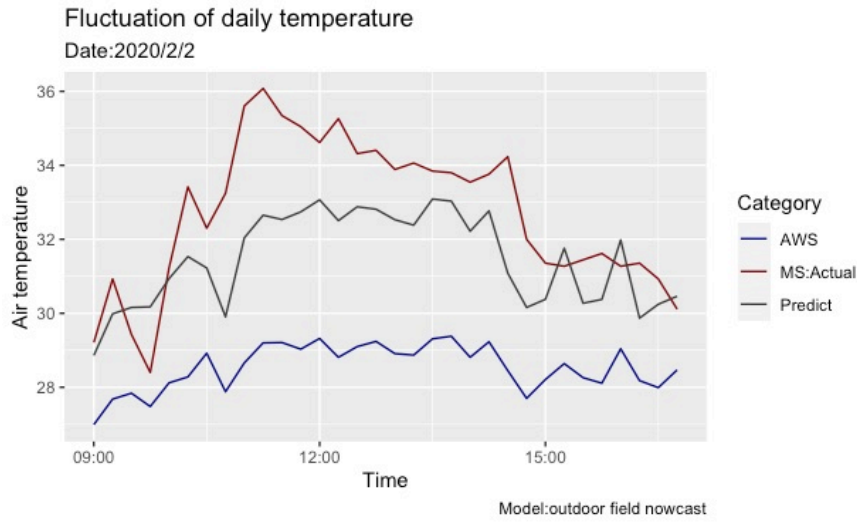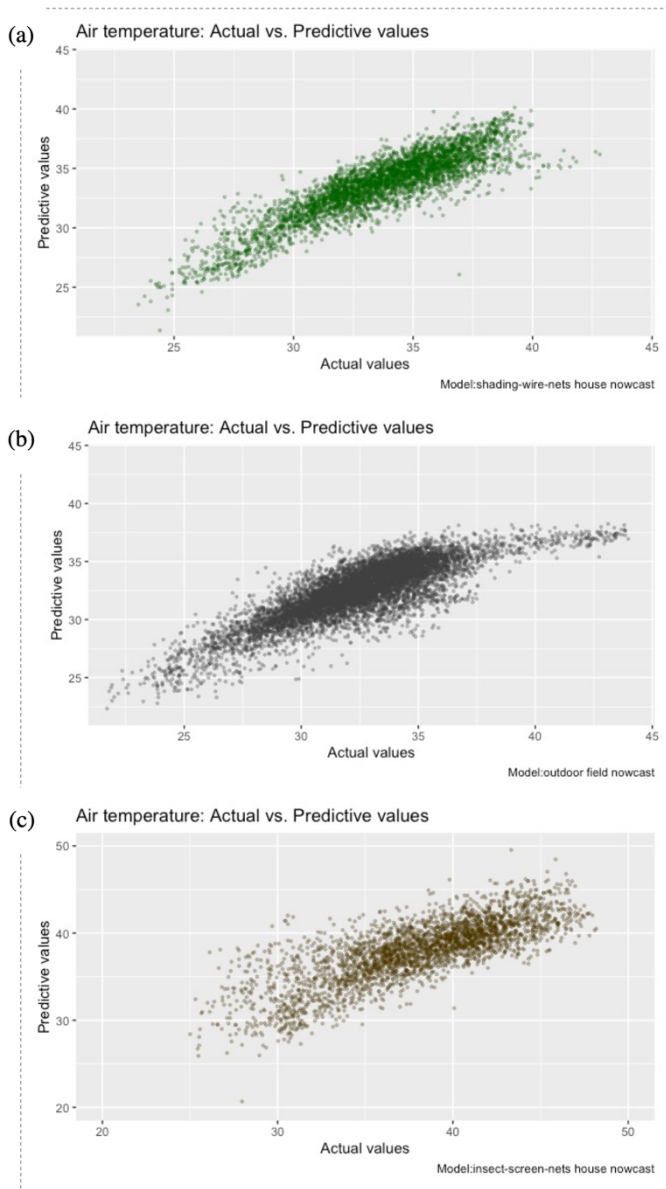
Model:outdoor field nowcast

Fig. 9: A line graph of fluctuation of daily temperature. The blue line represents the collected temperature from agrimeteorological weather station in Needmust, and the red line indicates the microenvironmental temeprature in La guerite. Then, the predictive values from outdoor nowcasting model is drawn as gray line.

The figure of predictive models in same location shows that the predictive value line is closer to the actual value line and with a similar pattern. Towards the outdoor nowcasting model, although the predicitve value line is still closer to the actual value line, the pattern is moderately different from actual data. Furthermore, the models would underestimate during the period of the extremely high temperature.

Finally, owing to every single predictive value has its corresponding actual air temperature, this study introduces the scatter plots for each model (Fig. 10).

Fig. 10: Scatter plots of the actual values versus predictive values. (a) The shading-wire-net nowcasting model. (b) The outdoor field nowcasting model. The data points include all collected data from cleaned datasets.

As the actual air temperature exceeds 40°C, Fig. 10 (a)(b) show more gentle slopes

which means the underestimation occurs. In Fig. 10 (c), the points of temperature prediction are more scattered than the others, especially when the actual temperature below 35°C. Nevertheless, the figures display the prediction is without significant bias in main intervals (30°C-40°C for (a), 30°C-37°C for (b) and 35°C-40°C for (c))

Discussion

The machine learning process can be looked into in many aspects. The purpose of machine learning is to better predict than the former methods. The limitations of the models are comparable important and need to be discussed. Last but not the least, the substantial contributions decide the whole machine learning process is useful or useless, which is concerned the most by farmers.

1. Comparing with simple linear regression model

In addition to machine learning, the simple linear regression might be the first idea, the straightforward one to deal with this subject. Using the big dataset of air temperature from AWS, it is able to explain the main variance and return the predictive value as well. However, how much accuracy machine learning can improve is always the concerned point. Therefore, this study recruits MAE, RMSE, and R square again for evaluating the difference between simple linear regression models and machine learning models (Table 7).

Table 7: The comparison of RMSE, R-square and MAE between machine learning (ML) models and simple linear regression (SLR) models of AWS temperature.

	Shading-wire-net house nowcast		Insect-screen-net house nowcast		Outdoor field nowcast	
	ML	SLR	ML	SLR	ML	SLR
RMSE	1.459	1.755	2.89	3.242	1.669	2.861
Rsq	0.779	0.680	0.606	0.504	0.685	0.566
MAE	1.113	1.343	2.23	2.508	1.270	3.088

2. Limitation of model usage

There are many limits for the prediction of microclimate due to different locations, outdoor or indoor environment. For example, the difference in altitude influences the precision of prediction. While the altitude increases 100 meters, the temperature ought to drop 0.6℃. Furthermore, the architecture of cultivated houses might be the problem as well. In this study, we tried three types of

cultivated house, but practically, farmers built cultivated houses in their ways. As long as the surrounding changed, the predictive power might lose credibility. The distance between the farmer's field and agrometeorological weather station also has an impact on the accuracy of prediction. According to the results, the outdoor field nowcasting model even possesses lower accuracy than the shading-wire-net house nowcasting model and also causes the problem of pattern altering. Moreover, although the outdoor nowcasting model selects AWS data and MS data in two places, the distance between them is only about 3 km (Fig. 6). It could not guarantee that the accuracy would not decline due to the longer distance.
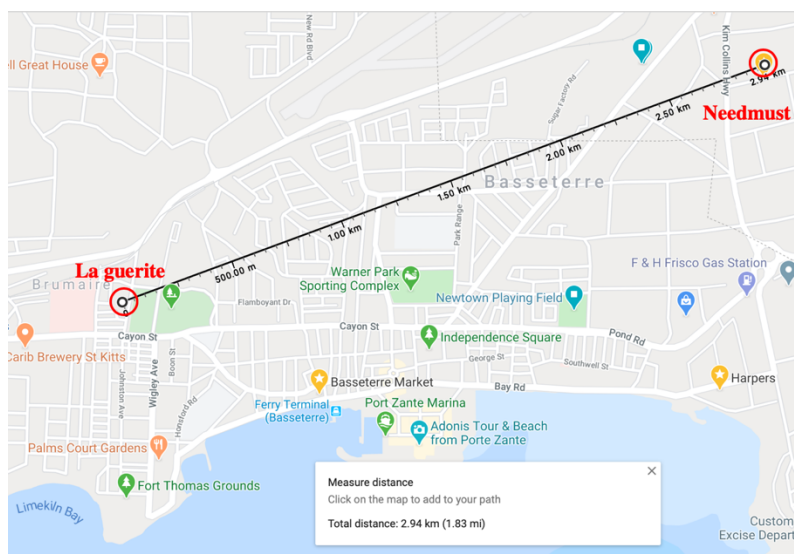

Fig. 11: The map indicates the rhumb line of La guerite and Needmust. The measure distance is 2.94 kilometer.

In contrast, the limitations constraint the versatility of models, but simultaneously they could be regarded as criteria. These criteria define the proper way to apply these models. For instance, there are three agrometeorological weather stations in Saint Kitts and they are delegations reporting the weather in different places. Needmust, Ecopark, and Mansion stations respectively represent the weather in 0m, 100m, and 200m height, and their data could be implemented with outdoor field models to predict the field temperature in the near areas with similar altitude. Moreover, the limitation of predicting distance help farmers choose the right AWS information to follow. At least with a radius of 3 kilometers, the outdoor field model has the confidence to provide accurate predictive values.

3. Contribution

This study has established three models for different microenvironment predictions. Although the models cannot do nowcasting perfectly, they already take a big step for smart agriculture in Saint Kitts. Through the empirical modeling, agricultural resilience, effective disaster warning system, the capability to mitigate the impact of increasing climate variability have all been reinforced.

Leveraged on the machine learning models and linked up with the dissemination of agricultural information, the real-time, accurate, and helpful information could be accessible for the local farmers and practically assist them to manage their fields.

Reference

1.   Watthanawisuth, N., A. Tuantranont, and T. Kerdcharoen. *Microclimate real-time monitoring based on ZigBee sensor network*. in *SENSORS, 2009 IEEE*. 2009.

2.   Fouilloy, A., et al., *Solar irradiation prediction with machine learning: Forecasting models selection method depending on weather variability*. Energy, 2018. **165**: p. 620-629.

3.   Berchtold, A., *Sequence Analysis and Transition Models*☆, in *Encyclopedia of Animal Behavior (Second Edition)*, J.C. Choe, Editor. 2019, Academic Press: Oxford. p. 506-517.