ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0015

Weather Conditions Prediction with Climate Change Dataset with Linear Regression Model, Autoregressive Model, LSTM Model

Yujia Jin[1], Caiwei Xiong[2], Jun Cao[3]

[1]Queen's University Belfast, Belfast, BT7 1NN, United Kingdom;

[2]School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China;

[3]Hailiang Foreign Language School, Shao Xing, 311800, China.

Abstract

Climate affects the whole earth. For ecology, climate affects biodiversity. For human societies, climate may affect the prosperity and decline of civilizations. The work uses the numerical weather prediction method, which based on a large number of climate data and mathematical models of machine learning method, specifically linear regression, time series vector autoregression, and neural network LSTM models to predict the climate elements mean temperature, mean vapor pressure deficit, and mean dew temperature. All machine learning methods used in this project have shown the higher degree of accuracy comparing to the basic models. LSTM model shows the best performance among all models, which due to the coverage of higher numbers of features and consideration of regularization. Among these climate elements, the prediction of average temperature has the highest degree of accuracy and that of mean vapour pressure deficit has the lowest degree of accuracy. Thus, An alternative methods of weather prediction on machine learning has shown effectiveness.

Keywords

Climate elements; Machine learning method; Linear regression; Time series vector autoregression, Neural network LSTM.

1. Introduction

Weather is a short-term manifestation of climate. Weather prediction shows a short period of future climate.

There are various methods to predict the weather. This follows the general process that weather information center takes the observation from satellites and input them into supercomputers. The supercomputers run the computational system several times a day to predict weather elements. The meteorologist will check the forecast along with time to see if it is accurate. [1] Numerical weather predictions were proposed as early as the early 20th century. Fluid mechanics and thermodynamics are mainly used in this method. Numerical weather prediction uses satellite observations, radar observations, and other data to predict the weather by computers. Detailed algorithm for systematic numerical weather prediction was first set out by Lewis Fry Richardson[2]. Most of today's methods of predicting the weather are based on his ideas. Nowadays, scientists are trying to predict weather conditions based on some computational methods by inputting the computational models and algorithms, for example, the sliding window algorithm[3]. More recently, machine learning methods have been used by Google to predict the short-term weather[4].

However, there are still some problems that the current weather forecast cannot solve. For some weather conditions that do not fit the linear equation, predictions are often inaccurate. Based on this,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0015

the weather conditions in previous years are used in this paper to train the model because the weather from year to year will not vary too much. The methods in these paper will improves the accuracy of the model's predictions.

In the work, three different machine learning models Linear Regression, vector auto-regression and LSTM are used to predict mean temperature, air pressure and mean dew temperature. Data from previous years were used to improve accuracy.

2. Data & Method

2.1 2.1 Data

This project used PRISM Spatial Climate Datasets for the Conterminous United State. PRISM datasets provide six climate indicators, precipitation (PPT), daily maximum temperature (T max), daily minimum temperature (T min), daily dew point temperature (Td mean), daily minimum vapour deficit (VPD_min) and daily maximum vapour deficit (VPD_max) and one derived variable daily mean temperature (T mean). This work derived a new indicator mean vapour pressure deficit (VPD_mean) which is the average of VPD_min and VPD_max. This project selects the data of a single location, Montgomery, Alabama, to predict the future weather conditions. Table 1 presents an overview of day wise data of climate elements between 1st Jan 2007 to 31st Dec 2017.

Table 1. Overview of day wise data of climate elements between 1st Jan 2007 to 31st Dec 2017

	mean	std	min	25% Q	50% Q	75% Q	max
ppt (inches)	0.150	0.412	0.000	0.000	0.000	0.050	4.310
tmin (degree F)	54.069	15.637	11.500	41.400	56.200	68.900	77.900
tmean (degrees F)	65.414	14.709	19.900	54.275	67.700	78.700	90.900
tmax (degrees F)	76.759	14.517	27.900	67.000	79.150	88.900	105.300
tdmean (degrees F)	54.892	15.580	4.400	43.700	58.400	68.600	76.300
VPD_min (hPa)	0.906	0.709	0.000	0.410	0.730	1.200	6.060
VPD_max (hPa)	18.446	9.641	0.550	10.720	18.100	25.410	59.040
VPD_mean	9.676	5.012	0.300	5.705	9.455	13.285	31.120

2.2 2.2 Method

The elements selected as the label for this project are mean temperature (Tmean), Daily mean dew point temperature (Tdmean), and the average between maximum vapor pressure deficit and the minimum vapor pressure deficit (VPD_mean). The work uses the following methods to predict the value of these labels in year 2018. For all the methods used in this project, the training data are data from 1st January 2007 to 31st December 2017 and the test data are data start from 1st January 2018 to 31st December 2018. The main goals behind trying different models aim to find the best model to predict the weather conditions with the lowest inaccuracy.

2.2.1 Baseline model

The baseline model used in this project is linear regression with single variable. In this method, the work assign one weather condition as the label, set the data of this element from the previous day as the sole feature, and predict the value of the label base on the data from the previous day via building up a linear relationship between this two values, $y = kx + b$. Use package linear_module from scikits.learn library to build up the model. Take the temperature baseline model as an example, according to table 2, the feature column 'Actual T mean' is the actual mean temperature of that day. The label column 'Forecasting Tmean' takes actual mean temperature as the forecasting values. The project forecast the mean temperature of the next day only based on the actual temperature on the present day.

ISSN: 2414-1895 | DOI: 10.6919/ICJE.202101_7(1).0015

Table 2. first five rows of temperature baseline model data

Date	Actual Tmean (degrees F)	Forecasting Tmean (degrees F)
2007/1/1	60.1	44.7
2007/1/2	44.7	43.9
2007/1/3	43.9	46.4
2007/1/4	46.4	62.2
2007/1/5	62.2	56.7

2.2.2 Multivariable Linear regression with multiple days as features

This model differs from the baseline model that set the same weather element from several previous days as the features to predict the value of a specific day. The linear relationship can be present as $y = \sum_i^n k_i x_i + b$. This method builds the model via the same package as method 2.2.1. This project build the models on previous seven days, fourteen days and thirty days for the label Tmean, Tdmean and VPD separately. For instance, table 3 illustrates the first five rows of data used in temperature forecast with previous seven days. Columns 'Actual Tmean' is the actual mean temperature of that day, the rest of columns shows the Tmean of each seven previous days.

Table 3. First five rows of temperature data for linear regression with multiple days

Date	Actual Tmean	Day-1	Day-2	Day-3	Day-4	Day-5	Day-6	Day-7
2007/1/8	60.2	60.8	56.7	62.2	46.4	43.9	44.7	60.1
2007/1/9	45.6	60.2	60.8	56.7	62.2	46.4	43.9	44.7
2007/1/10	43.8	45.6	60.2	60.8	56.7	62.2	46.4	43.9
2007/1/11	41.3	43.8	45.6	60.2	60.8	56.7	62.2	46.4
2007/1/12	47.1	41.3	43.8	45.6	60.2	60.8	56.7	62.2

Table 4. First five rows of multivariable linear regression with multiple elements as features for temperature forecasting

Date	ppt (inches)	tmin (degrees F)	tmax (degrees F)	tdmean (degrees F)	VPD_min (hPa)	VPD_max (hPa)	Forecasting tmean (degrees F)
2007/1/1	0.78	49.7	70.6	59.3	0.14	4.68	44.7
2007/1/2	0	30	59.4	38	0.42	8.09	43.9
2007/1/3	0	30.8	57.1	34.6	0.09	8.89	46.4
2007/1/4	0	35.3	57.5	46.4	0.42	5.06	62.2
2007/1/5	0.04	55.7	68.8	60.9	0.83	4.82	56.7

2.2.3 Multivariable linear regression with multiple elements as features

This model selects one element as label and set the remaining elements from the one previous day as features, and build up the same linear relationship as method 2.2.2, $y = \sum_i^n k_i x_i + b$. According to table4, the first six columns are features of the corresponding data of that date. The column 'Forecasting t mean' is the mean temperature of the next day what the work is going to predict later.

2.2.4 Polynomial linear regression

This method run on the same data as the method 2.2.3 but apply different degrees on features during data processing. It employs package PolynomialFeatures to add the polynomial degree. This project runs the models on several degrees and find the best degrees are between 2-3 for three labels.

2.2.5 Lasso Regression

Lasso regression performs on feature selection that encourage the simple and sparse models to enhance the accuracy of the prediction.[5] This method run on the same data as method 2.2.3. The work uses Lasso package and apply GridSearchCV package from scikits.learn library to find the best parameter for the model. The best parameters found for Tmean are 0.001, 0.01 for VPD_mean and 1×10^{-15} for Tdmean.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0015

2.2.6 Vector autoregression

Vector autoregression (VAR), a stochastic process model, is capable of capturing the linear interdependencies between many time series. Contrasting to Multivariable Linear regression with multiple days (method 2.2.2), this method differs in the two main aspects. On the one hand, VAR contains all variables in the data in the form of vectors. Each vector represents the same variable on certain days. Its matrix notation could be written as Y=BZ+U, where Z stands for the variable matrix. On the other hand, VAR contains the predictor value of the previous days as the feature while method 2.2.2 did not account the predictor value of the previous day as feature. As using this model requires some conditions, this work carry out some tests beforehand.

The work did ADF test(Augmented Dickey Fuller Test) at first, which can test whether the time series of variables are stationary or not. If they are not stationary, the difference or the logarithm should be taken. Fortunately, it has shown that the time series of variables are stationary without any needs of transformation.

Then, the work did the cointegration test to explore whether there is cointegration relationship between variables which required to be predictd and other variables. Since there is a cointegration relationship between them at a significance level of 0.05, this work include all variables into the model.

In order to maintain a high accuracy rate, the AIC rule be chosen to filter models, which means to choose the number of lag days providing the lowest AIC. The result has shown that the lag of six days is most appropriate.

Table 5. VAR Order Selection (* highlights the minimums)

Lag days	AIC
0	4.143
1	0.7004
2	0.4656
3	0.3969
4	0.3857
5	0.3640
6	0.3539*
7	0.3584

The results of the model will be shown in the part of results. After modeling, this project passed residual stationarity test and residual correlation test and performed variance decomposition. Finally, the work made predictions on related variables one day at a time.

Since VPD_mean is the variable average of VPD_min and VPD_max, the work also average the predict VPD_min and VPD_max to get the predict VPD_mean.

2.2.7 LSTM

Since VAR may be biased due to abnormal points, and for weather, the occurrence of abnormal weather is likely to be affected by the surrounding weather. For example, as monsoon influence may come from the seaside towards the inland, the cities close to the seas would the first area to perceive changes in the weather. In this case, if the weather conditions of the cities near the seas couple of days ago is known, it is possible to forecast weather for inland cities in a more accurate way. Since the dataset have 48 regions in total, if each region takes 7 features, it will be too many variables for the VAR model. So, this work tries LSTM (Long short-term memory), which was used to deal with time series data in the paper, *Neural Granger Causality for Nonlinear Time Series*[6].

The raw data needs to be processed before entering LSTM. Seven variables in different regions this year and last year are chosen as features, thus, there are 7*48*2=672 features in total, which is the number of columns. The number of rows is the lag days entered into the model, which be chosen as 7, 14 and 30. Thus, the total days entering the model could be 14, 28 and 60. The feature matrix of

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0015

past days and a label of the current day each time the model enters to train the parameters. The 7-day feature matrix diagram is as follows. Briefly show the case with only 2 features and 2 regions.

Table 6. 7-day feature matrix diagram

days	A1	A2	B1	B2	A1'	A2'	B1'	B2'
1								
2								
…								
7								

In the form, A and B represents the two regions. A1 means the first features of A, so A2 means the second. A1'means the first features of A in the previous year.

Next comes the modeling part. Here is a brief introduction to the basic principles of LSTM.

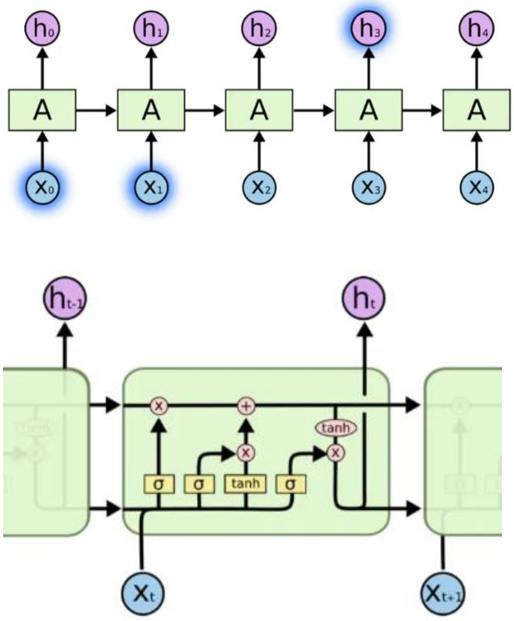

Figure 1. LSTM diagram [7]

LSTMs have a chain structure with repeating cells. The algorithm of each cell can be written as follows.[8]

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b) \quad (1)$$

$$\widetilde{C_t} = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (2)$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t} \quad (3)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (4)$$

$$h_t = o_t * \tanh(C_t) \quad (5)$$

The work use pytorch framework. The model structure in this project is roughly as follows.

LSTM + ReLu + LSTM+ ReLu + LSTM + ReLu

ReLu can be the activation function, and each LSTM has 2 layers.

The loss function is MSE. As for the gradient processing, SGD (Stochastic Gradient Descent) optimizer is chosen, which may work well on continuous data. This work set the learning rate to 1e-5. Considering there are too many features in the model, this work set weight_decay =1 , which can be seen as L2-norm. In order to avoid the gradient explosion, the gradient is cleared if the gradient is greater than 0.5. The work trains the entire data about 7 to 12 times each time.

ISSN: 2414-1895
DOI: 10.6919/ICJE.202101_7(1).0015

3. Result

As the paper mentioned earlier, this project assigns the data from 1st January 2007 to 31st December 2017 as training data and predict the weather conditions of the whole year of 2018 on all models. All models select the Tmean, Td mean and VPD_mean these three elements as predictors. This project compares the mean squared error of testing data from each model and evaluate the accuracy of the prediction based on that.

Figure 2, 3 and 4 shows the variation of actual and predicted climate elements from baseline models for year 2018 day wise, the climate elements are average mean temperature, mean dew point temperature and mean vapor pressure deficit, respectively. These figures clearly present the least differences betweem the predicted and actual weather elements. Generally, the predicted average between maximum and minimum vapour pressure deficit presents the best fit to actual value relative to other weather conditions. Mean temperature ranks the second, and the mean dew temperature has the least fit on graph.

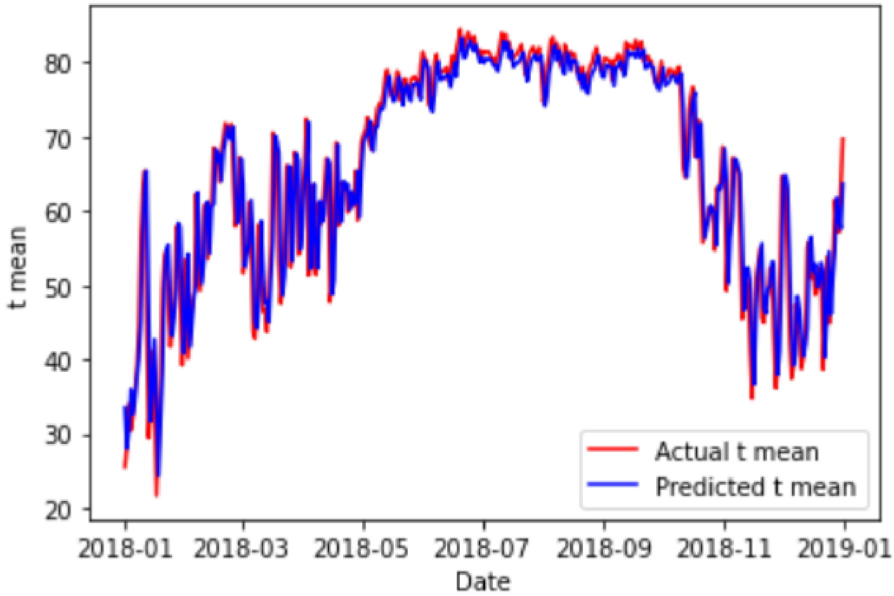

Figure 2. baseline graph of predicated average temperature versus actual mean temperature for year 2018

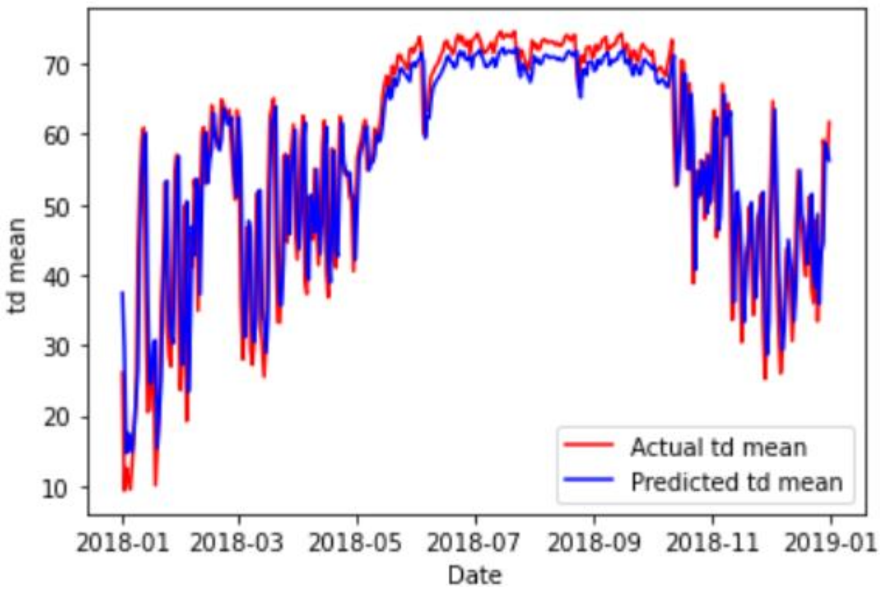

Figure 3. baseline graph of predicated mean dew point temperature versus actual mean dew temperature for year 2018

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0015

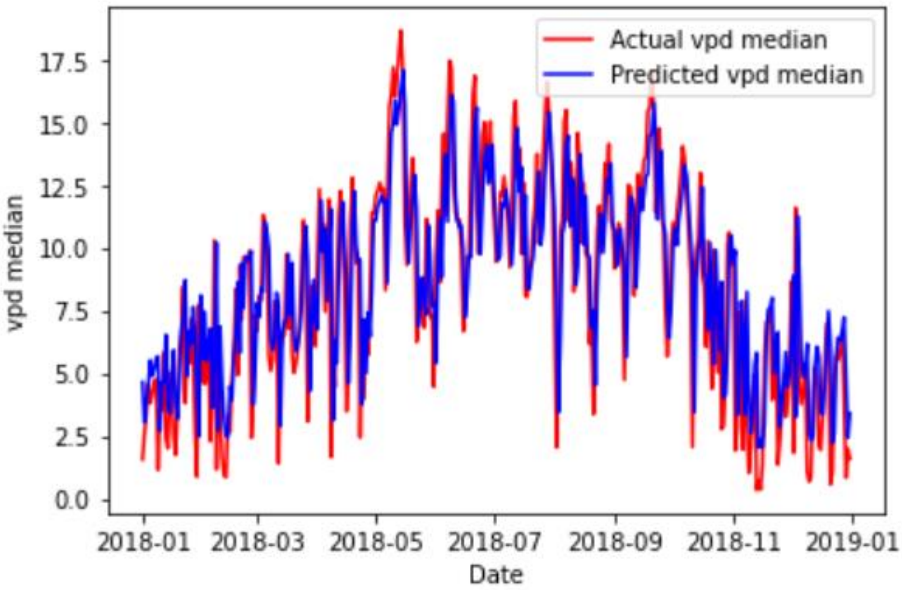

Figure 4. baseline graph of predicated mean vapor pressure deficit versus actual vapor pressure deficit for year 2018

Table 7. Mean square errors (MSE) and R square value of different linear regression models

Method	Predicator	MSE training	MSE test	R^2
2.2.1 Baseline	Tmean	26.199	26.434	0.881
	VPD_mean	8.377	7.0585	0.6011
	Tdmean	53.486	52.583	0.799
2.2.2 Single weather condition with multiple days	Tmean (day=7)	23.113	22.841	0.899
	Tmean (day=14)	22.695	22.919	0.898
	Tmean (day=30)	22.707	23.086	0.898
	VPD_mean (day=7)	7.747	6.622	0.626
	VPD_mean(day=14)	7.635	6.501	0.633
	VPD_mean(day=30)	7.661	6.527	0.631
	Tdmean (day=7)	45.969	45.568	0.826
	Tdmean (day=14)	44.750	45.149	0.828
	Tdmean (day=30)	44.309	45.947	0.825
2.2.3 Multiple condition with 1 day	Tmean	23.576	23.897	0.893
	VPD_mean	8.023	7.070	0.599
	Tdmean	42.318	43.204	0.834
2.2.4 polynomial regression multiple weather conditions with 1 day	Tmean (degree=2)	21.001	22.447	0.899
	Tmean (degree=3)	20.509	20.874	0.906
	VPD (degree=2)	7.381	6.459	0.634
	Tdmean(degree=2)	38.898	40.038	0.846
2.2.5 Lasso	Tmean	23.576	23.890	0.893
	VPD_mean	33.026	27.879	0.590
	Tdmean	42.374	43.259	0.833

Table 7 presents the MSE and R^2 values of different linear regression models employed in this project. For the same variable, the smaller the MSE, the better the performance. According to this table 7, the baseline model shows the lowest MSE for VPD_ mean as 7.059, Tmean in the middle which is 26.434 and the highest for Tdmean, 52.583. While for R^2 value, Tmean has the highest value 0.881, Tdmean has the second highest 0.799, and VPD_ mean ranks the lowest 0.601. Contrasting to the baseline model, all the other methods present a lower value of MSE corresponding to the specific weather conditions. In these regression models, the method 2.2.4, polynomial regression of multiple weather conditions with data from one previous day, has lowest MSE on test data for all three weather

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0015

conditions. The predictor Tmean have its least MSE (20.874) at polynomial degree of 3 for Tmean, while both VPD_ mean and Tdmean have the least value at polynomial degree of 2, 6.459 and 40.038, respectively. The prediction based on method 2.2.2, single weather condition with multiple days shows lower testing MSE for all three predictors comparing to the baseline model, but there are not much variation among data of day 7, 14 and 30. The MSE of method 2.2.3 multiple condition within one previous days shows the more or less same value as that in lasso methods.

Table 8. P value & Coefficient of Method 2.2.3 Multivariable Linear Regression with Multiple Elements

		coefficient	p-value
VPD	const	-0.2852	0.61
	ppt	-0.838	0
	Tmin	-0.4873	0.439
	Tmean	1.128	0.371
	Tmax	-0.5412	0.39
	Tdmean	-0.0389	0.03
	VPD_min	-0.1396	0.11
	VPD_max	0.3526	0
Tmean	Const	9.7654	0
	ppt	-2.1274	0
	Tmin	0.7836	0
	Tmax	0.3251	0
	Tdmean	-0.2401	0
	VPDmin	-1.42	0
	VPDmax	0.1688	0
Tdmean	const	3.9037	0
	ppt	-1.7031	0
	Tmin	4.4892	0.002
	Tmean	-7.2977	0.012
	Tmax	3.7298	0.01
	VPD_min	-2.3149	0
	VPD_max	0.0885	0.001

Table 9. Test MSE of all models

Method	days	T_mean	Vpd_mean	Tdmean
2.2.1 baseline	1	26.4336	7.0585	52.583
2.2.2 Multivariable lr	1	23.8972	7.0695	43.2036
2.2.3 Single feature for multiple days	7	22.841	6.6223	45.5678
2.2.3 Single feature for multiple days	14	22.9187	6.5009	45.1487
2.2.3 Single feature for multiple days	30	23.0864	6.5271	45.9471
2.2.4 polynomial lr	1	20.7375	6.3059	39.1347
2.2.5 Lasso	1	23.8902	7.0504	43.259
2.2.6 VAR	6	21.6263	6.1896	36.5113
2.2.7 LSTM	30	15.8762	3.9593	22.6184
2.2.7 LSTM	14	42.3184	2.7117	32.1293
2.2.7 LSTM	7	10.1683	1.3004	30.2709

Table 8. and the results of VAR model have shown that precipitation, temperature, and vapor pressure in the past has great impact on t_mean in both models. While dew point temperature is not obvious in VAR but obvious in Multivariable Linear Regression. As for vapor pressure, dew point temperature and the max vapor pressure in the past can influence it obviously. When it comes to dew point temperature, almost all the variables have significant impact on it. Among them, precipitation is

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0015

negatively correlated with vapor pressure and temperature, and positively correlated with dew point temperature. Temperature is positively correlated with vapor pressure, and dew point temperature is negatively correlated with vapor pressure.

4. Discussion

Table 9 contrast the MSE and R^2 on test set for models all above. Contrasting to all of other method, generally LSTM presents the lowest MSE for all three climate elements which indicate the highest degree of accuracy. This fact could be possibly explained by the variation of feature selection to other methods. The feature capacity varies with different models, and the results of prediction can also reflect the variation of different feature capacities. The difference in feature selection is mainly reflected in whether to use all weather features, how many days' data are used, and whether to include other regions for training. The difference in model optimization is mainly reflected in whether to use regularization.

One of the reasons that average VPD_mean has the lowest MSE among these predictors is that it has the least variations on day wise data relative to other weather conditions. For method 2.2.5 Lasso, as the paper discussed earlier, it has parameters close to 0. This fact is the main reasons that why it has the nearly same results as method 2.2.3 multiple conditions with one previous day. LSTM has the best overall performance among the above models, probably because it covers the most features and also considers regularization.

Method 2.2.7 LSTM with 7 days(7 days for the current year and 7 days for the previous year) has the best performance on T_mean (10.1683) and Vpd_mean(1.3004). For Tdmean, LSTM with 30 days presents the best performance with MSE of 22.6184. This could be interpreted as that the dew point temperature is related to humidity, and the humidity of a certain condition lasts longer than other variables.

Table 10. R^2 from different models

Method	days	T_mean	Vpd_mean	tdmean
baseline	1	0.8811	0.6011	0.7992
Multivariable lr	1	0.8925	0.599	0.8335
Single feature for multiple days	7	0.8987	0.6258	0.826
Single feature for multiple days	14	0.8983	0.6327	0.8276
Single feature for multiple days	30	0.8976	0.6312	0.8245
polynomial lr	1	0.9067	0.6621	0.8492
Lasso	1	0.8925	0.6	0.8333
VAR	6	0.904	0.6502	0.8605
LSTM	30	0.9221	0.771	0.9037
LSTM	14	0.7924	0.8431	0.8632
LSTM	7	0.9501	0.9248	0.8711

Since R^2 and MSE have a linear relationship, they are the same when comparing single variables in different models. In terms of these three climate elements, T_mean and tdmean are relatively easy to predict, while VPD_mean is more difficult to predict accurately. For VPD_mean, because of its smallest variance, a little error will cause R^2 to vary greatly. Hence, its R^2 is more difficult to reach a higher level. As the result shown LSTM with 7 days shows much better prediction than other days, one of the main reason for this fact could be the good fitting ability of LSTM. Another possible reason is that vapor pressure is related to regions, and LSTM is the only model that may infer region information.

5. Conclusion and future works

This work has completed models based a series of on linear regression, vector autoregression and neural network LSTM, which has shown the effectiveness of weather prediction based on machine

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0015

learning methods. Linear regression reflects the positive impact on the results when the linear regression is improved in different aspects. All of the methods tried in the project have shown an increased degree of accuracy comparing to the baseline models of the relative climate elements.

Overall, LSTM has the best performance among the all models used, which can probably explained as the coverage of the most features in data and the consideration of regularization. By comparing the models above, this article believes that different models have their own strengths. Neural networks are the best in terms of accuracy, while traditional machine learning based on regression has advantages in speed and interpretability. Contrasting with the speed of LSTM, traditional machine learning models run faster and are more explanatory. Because not only LSTM, but also all other methods have a lower value of MSE and higher score than its relative value in the baseline model, the work conclude that weather conditions of the previous day is not the only impact on the day.

In terms of the predicted three variables, even though vapor pressure deficit presents the lowest MSE value due to its relatively small daily variation, it shows the lowest degree of accuracy relative to other weather condition. This fact result from its small variance, little error will lead to significant changes of values. The mean dew temperature has the second highest accuracy, and the mean temperature ranks the highest.

This project remains some areas for improvement. For Linear regression, the future work can add the feature selection process, which selects the features with lower collinearity and low p value, or construct new variables to reduce multicollinearity, to obtain more stable and meaningful coefficients of models. In terms of VAR model, the future work can think about BIC to adjust parameters, which fixing them to zero or not, and handle outliers to make the model more robust. When it comes to LSTM, it would be better to show the influence of different regions on the predicted region before constructing the model.

References

[1] “How weather forecasts are created -Met Office.” [Online]. Available: https://www.metoffice. gov. uk/ weather/ learn-about/how-forecasts-are-made. [Accessed: 03-Sep-2020].

[2] “Weather Prediction by Numerical Process - Lewis Fry Richardson - Google Books.” [Online]. Available: https://books.google.co.uk/books?hl=en&lr=&id=D52d3_bbgg8C&oi=fnd&pg=PA3&dq=weather+pred iction +by +numerical +process &ots =60Seg -_HQr &sig =21lOK7tYAp0GUKzEDp5yXPt _t7g #v= onepage & q =weather prediction by numerical process&f=false. [Accessed: 07-Sep-2020].

[3] P. Kapoor and S. S. Bedi, “Weather Forecasting Using Sliding Window Algorithm,” ISRN Signal Process., vol. 2013, pp. 1–5, 2013.

[4] J. Hickey, “Google AI Blog: Using Machine Learning to ‘Nowcast’ Precipitation in High Resolution,” 2020. [Online]. Available: https://ai.googleblog.com/2020/01/using-machine-learning-to-nowcast.html. [Accessed: 07-Sep-2020].

[5] “Lasso Regression: Simple Definition -Statistics How To.” [Online]. Available: https:// www. statisticshowto.com/ lasso-regression/. [Accessed: 14-Aug-2020].

[6] A. Tank, I. Covert, N. J. Foti, A. Shojaie, and E. B. Fox, “Neural Granger Causality for Nonlinear Time Series,” 2018.

[7] “Understanding LSTM Networks -- colah’s blog.” [Online]. Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/. [Accessed: 07-Sep-2020].

[8] “LSTM — PyTorch 1.6.0 documentation.” [Online]. Available: https://pytorch.org/docs/stable/generated/ torch. nn.LSTM.html. [Accessed: 07-Sep-2020].