國立臺灣大學理學院地理環境資源所

碩士論文

Department of Geography

College of Science

National Taiwan University

Master Thesis

人工神經網絡和長短期記憶模型探討森林生態系通量特徵

Investigating fluxes characteristics of forest ecosystem by using artificial neural networks and long short-term memory models

蔡硯丞

Yan-Cheng Cai

指導教授：莊振義 博士

Advisor: Jehn-Yih Juang, Ph.D.

中華民國 112 年 2 月

February 2023

doi:10.6342/NTU202300497

謝誌

謝謝莊振義老師這兩年半的悉心教導
您在學術上面的提點
我才能夠站在巨人的肩膀看世界

謝謝地理環境資源研究所的同學、助教、朋友、研究室同伴
你們在孤單的研究之路上點亮一展路燈
我才能夠度過每一個研究的夜晚

謝謝在台北的同伴
你們在台北陪伴我、展覽、登山、旅遊、聚餐
我才能夠再次享受生活的美好

謝謝在遠在嘉義的家人
您們在金錢、精神上都無償的支柱我
我才能在研究所期間無後顧之憂

謝謝太保火把教會的同伴
你們在兩年半的六日許多的支持
我才能飛的更高更遠

謝謝台灣大學
您在每天都提醒我的不足
我才能督促自己不斷的往前邁進

一千天的夢要醒了，蛻去人生中最後的學生時光
餘生回頭來看這段青春歲月
並向每一個旅途上的同伴好好的道別

Psalms 23:4

Even though I walk through the darkest valley, I will fear no evil, for you are with me; your rod and your staff, they comfort me.

doi:10.6342/NTU202300497

摘要

在過去的幾十年裡，渦流相關係數技術已經被廣泛運用於觀測不同的地表-大氣交互作用，然而，由於設備故障或者低風速所造成的弱渦流現象，目前仍有大約 20~60%的通量測量數據的缺失，在所有資料補遺的方法當中，機器學習(machine learning, ML)是一個強大的工具，可以簡單地建立輸入和輸出之間的非線性關係，並在許多通量的研究上面被廣泛利用。

本研究中，選擇以台灣亞熱帶地區棲蘭通量站，以探討機器學習在該地區的效能，本研究以溫度，土壤溫度，相對溼度，淨輻射通量四個參數作為主要分析對象，但也加入風速、風向、能見度、光合有效輻射做進一步的討論，其中主要使用人工神經網絡(ANN)和長短期記憶(LSTM)來預測生態系統當中的 CO_2 交換和潛熱通量(LE)。並使用決定係數 R^2、平均絕對誤差 MAE、均方根誤差 RMSE 等指標對於機器學習進行分析，後續研究使用偏微分(PaD)來探討不同參數的貢獻，以了參數之間的關係，最後將缺漏值進行填補，並和未填補資料進行比較及分析，以了解是否在季節性或或者整體上面有落差。

ANN 和 LSTM 的初步結果顯示，在各種參數的組合之下使用八個參數可以得到較高的 R^2，CO_2 的部分分別為 0.74 和 0.71，LE 的部分為 0.71 和 0.67，在日夜分開測量可以發現夜間無法有效預測，然而在 PaD 當中顯示 Rn 對於兩個參數的貢獻量最大，分別為 62%和 35%，而在填補後可以發現在冬季缺值較多的季節四分位明顯縮小，機器學習對於棲蘭地區有良好的預測能力。

關鍵字:二氧化碳通量、潛熱通量、渦度相關係數、缺值填補、偏微分、機器學習

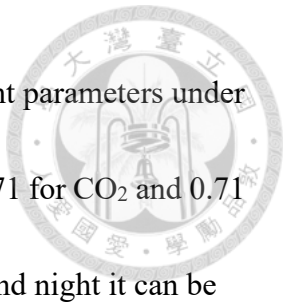
doi:10.6342/NTU202300497

Abstract

The eddy-covariance technique has been widely applied to quantify the surface-atmosphere interactions over different landscapes in the past decades. However, nowadays there are still about 20 to 60% missing data in the flux measurement because of equipment failure or the weak turbulence caused by low wind speed. Among all the gap-filling methods, machine learning (ML) is a powerful tool to simply establish the non-linear relationship between the input and output parameters and has been broadly utilized in many flux studies. However, very little attention to ML was given to the analysis of multi-landscape comparison.

In this study, the CLM flux station in the subtropical region of Taiwan is chosen to investigate the effectiveness of machine learning in this region. This study uses artificial neural networks (ANN) and long short-term memory (LSTM) to predict the CO_2 exchange and latent heat flux (LE) in the ecosystem. In the subsequent study, partial derivation (PaD) was used to investigate the contribution of different parameters to understand the relationship between parameters, and finally, the missing plants were filled and compared with the unfilled data to understand whether there is any seasonal or overall discrepancy.

doi:10.6342/NTU202300497



The preliminary results of ANN and LSTM show that using eight parameters under various combinations of parameters can obtain higher R^2, 0.74 and 0.71 for CO_2 and 0.71 and 0.67 for LE, respectively, and in separate measurements of day and night it can be found that nighttime is not effectively predicted, however, in PaD it shows that Rn has the largest contribution to two parameters, 62%, and the PaD shows that Rn contributes the most to both parameters, with 62% and 35% respectively, and after filling the quartiles, it can be found that the quartiles shrink significantly in the winter season when there are more missing values, and the machine learning has good prediction ability in CLM.

Keywords: Carbon dioxide flux, Latent heat flux, Eddy covariance, Gap filling, Partial derivation, Machine learning

doi:10.6342/NTU202300497

Content

doi:10.6342/NTU202300497

doi:10.6342/NTU202300497

Figure Content

doi:10.6342/NTU202300497

doi:10.6342/NTU202300497

doi:10.6342/NTU202300497

Table Content

doi:10.6342/NTU202300497

Chapter 1 Introduction

1.1 Background and motivation

For more than 30 years, the eddy covariance technology (**Figure 1**) has been used to measure the exchange of land-atmosphere, which includes greenhouse and energy that can track the atmospheric disturbance in the atmosphere every half minute (Pastorello et al., 2020); it is based on high-frequency (10-20 Hz) measurements of vertical wind speed and scalars (CO_2, H_2O, temperature), it provides an estimate of the scalar net exchange in the source footprint area (Aubinet et al., 2012).

To climate change, the hydrology of the atmosphere has changed drastically in recent years (Gleick, 1989). The atmosphere's interaction with fundamentals such as carbon dioxide (CO_2), latent heat (LH), and sensible heat(H) fluxes are indispensable. The data was calculated by the eddy covariance instrument, combined with a sonic anemometer and net radiometer; these instrument s can detect ambient air information.

Evapotranspiration is a combination of two independent processes in which water is lost from the soil surface through evaporation on the one hand, and crops are lost through transpiration on the other (Jensen et al., 1990), evaporation turns liquid water into vapor, which is removed from soil, lakes, rivers and the humid environment. There has a conversion process through solar radiation and temperature, which provide the energy.

doi:10.6342/NTU202300497

Insufficient vapor pressure and wind speed enhance the diffusion of water vapor in the atmosphere (Granata, 2019). Considering the above reasons, solar radiation, temperature, wind speed, and air humidity are the critical factors affecting the evaporation process.

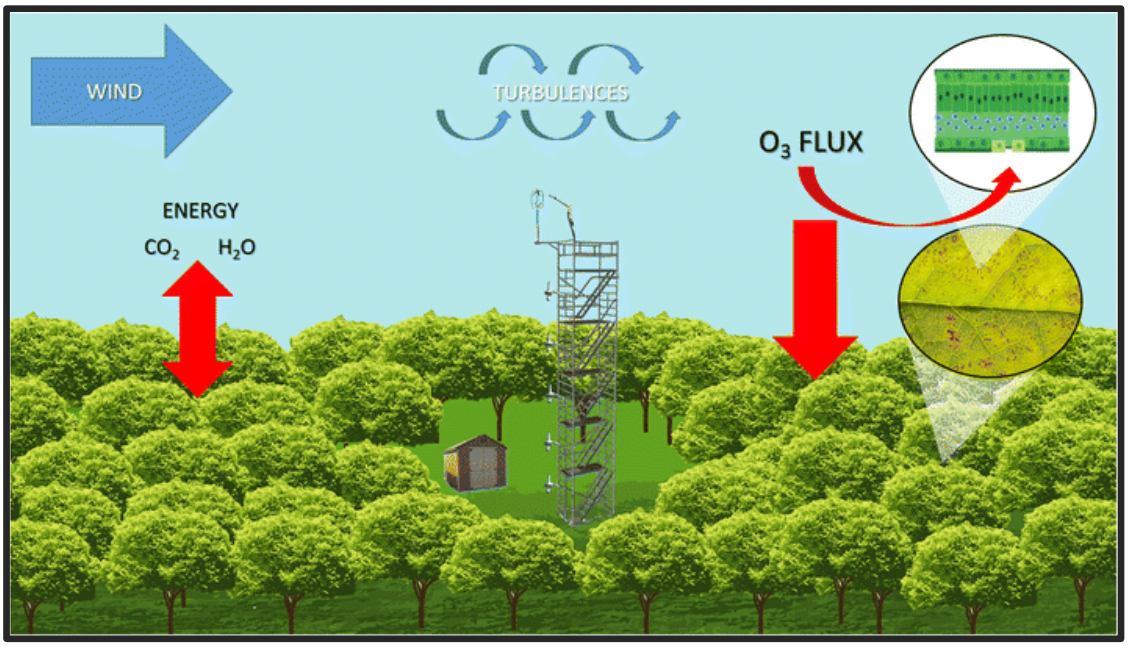


Figure 1. Eddy covariance equipment (Fares et al., 2018)

Unfortunately, sometimes the instruments could malfunction, and maintenance and calibration impact the operation of the instruments for a variety of reasons that caused the instrument's failures, for example about 22% of the total measurements were found with some gaps and poor-quality data at the Walker Branch Watershed (WBW) Amer Flux site (Wilson & Baldocchi, 2001), and 35% at the duke site (Baldocchi et al., 2001). So, there are so many types of gaps filling methods (**Figure 2**), such as nonlinear regression techniques (NLRs), 3 the look-up table (LUT), marginal distribution sampling (MDS),

doi:10.6342/NTU202300497

the semi-parametric model (SPM) (Moffat et al., 2007), show these model's advantages and disadvantages.

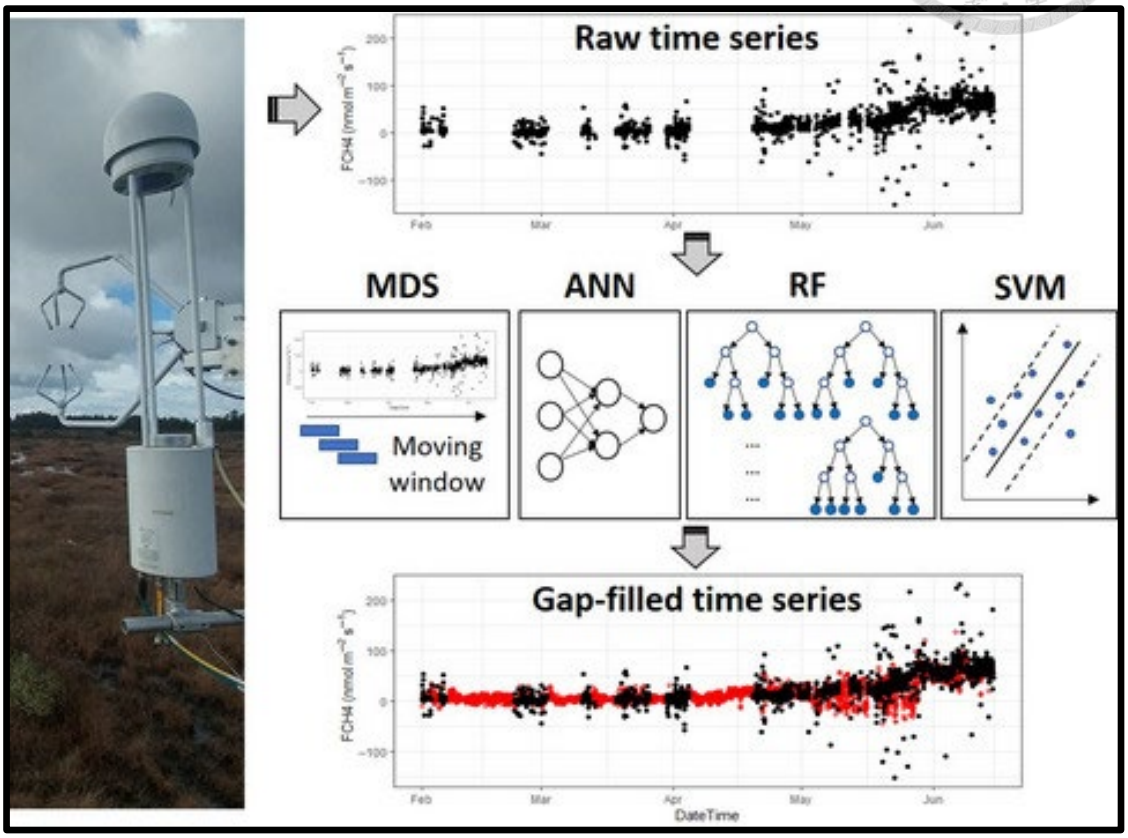


Figure 2. The gap-filling method through machine learning (Kim Y. et al., 2020)

Aside from the classic calculation method, gap-filling methods are based on data calculation, so the situation is very suitable for machine learning because the machine learning has an excellent algometric at the nonlinear data, like Dario Papale used artificial networks to estimate carbon and water fluxes at six different forest station in Europe (Papale et al., 2006), (Dou & Yang, 2018) used five types of machine learning methods to estimate carbon and water fluxes (Dou & Yang, 2018), Goodrich estimated agricultural nitrous oxide (N_2O), and the research used the machine learning method like neural 4

doi:10.6342/NTU202300497

networks and locally-weighted k-nearest neighbors (KNN) regression in the three footprint areas, the NN and KNN performed similarly well.

There are many fluxes tower that was established around the world (Baldocchi et al., 2001), and the tower record many data like Carbon Dioxide, Water Vapor, Energy flux densities, including global radiation (Rg) photosynthetic, Temperature (Ta), Temperature of soil (Ts) relative humidity (Rh), precipitation (P) and soil water content (SWC).

The eddy covariance method is the primary monitoring tool, and the instrument can measure the environmental parameters; the measurements are reported on a half-hourly. The instruments were used in many situations, such as in China there are many environments type; the environment parameters were compared in those situations (Yu et al., 2006), but because calibration or equipment failures result in occasional gaps in those periods. A significant limitation of the eddy covariance technique is the need for turbulent atmospheric conditions. Reject data was collected under low turbulence conditions based on the friction velocity threshold (u^*) (Papale et al., 2006). Data quality is critical to scientific results.

In the past, the study used nonlinear regression techniques (NLRs), the look-up table (LUT), marginal distribution sample (MDS), and the semi-parametric model (SPM), which generally showed good performance, in recent years, many technologies have been

doi:10.6342/NTU202300497

invented and applied (Wang H. J. et al., 2015) This paper uses a residual bootstrap method to estimate annual NEE, which provides good estimates for more frequent and less frequent gaps. And then the method is Machine learning, such as artificial neural networks (ANNs), was slightly superior to the other techniques (Moffat et al., 2007).

In recent years, different extreme events and disturbances, such as droughts, precipitation, hurricane, and heat waves, would change the environment's biodiversity and growth patterns (Alkama & Cescatti, 2016). Therefore, the environmental factors fluctuate very drastically, which may cause the traditional empirical model cannot to be applied effectively. The neural network was inspired by the human brain, where hundreds of billions of interconnected neurons deal with many types of information (Wang, 2003). The model used many functions to reach the actual value that is not only used in the gap filling but also the daily estimate (Kim et al., 2013), estimating the state-dependent biases of a land surface model (Wang et al., 2012) or using in the satellite remote sensing data (Ueyama et al., 2013). There are many different methods of machine learning, so this study will compare different machine learning methods in the CLM stations.

However, in this paper (Khan et al., 2021) various machine learning methods, including SVM, CNN, RF, and LSTM, are also proposed in this article to predict LE and

doi:10.6342/NTU202300497

it also includes remote sensing detection for the prediction of LE. The results show that machine learning has a good predictive ability.

1.2 Research objectives

From the current global warming data (**Figure 3**), it can be found that the carbon flux, evapotranspiration, and precipitation of the ecosystem will all be affected by climate change, which will affect the development of the entire ecosystem because the rapid climate changes (Hung & Li, 2020) conditions in Taiwan like an unusual typhoon event (Webster et al., 2005) that often cause previous experience function cannot use effectively.

The eddy covariance instruments are often damaged by the environment or malfunction by themselves. However, the current method of machine learning can effectively fill it up.

doi:10.6342/NTU202300497

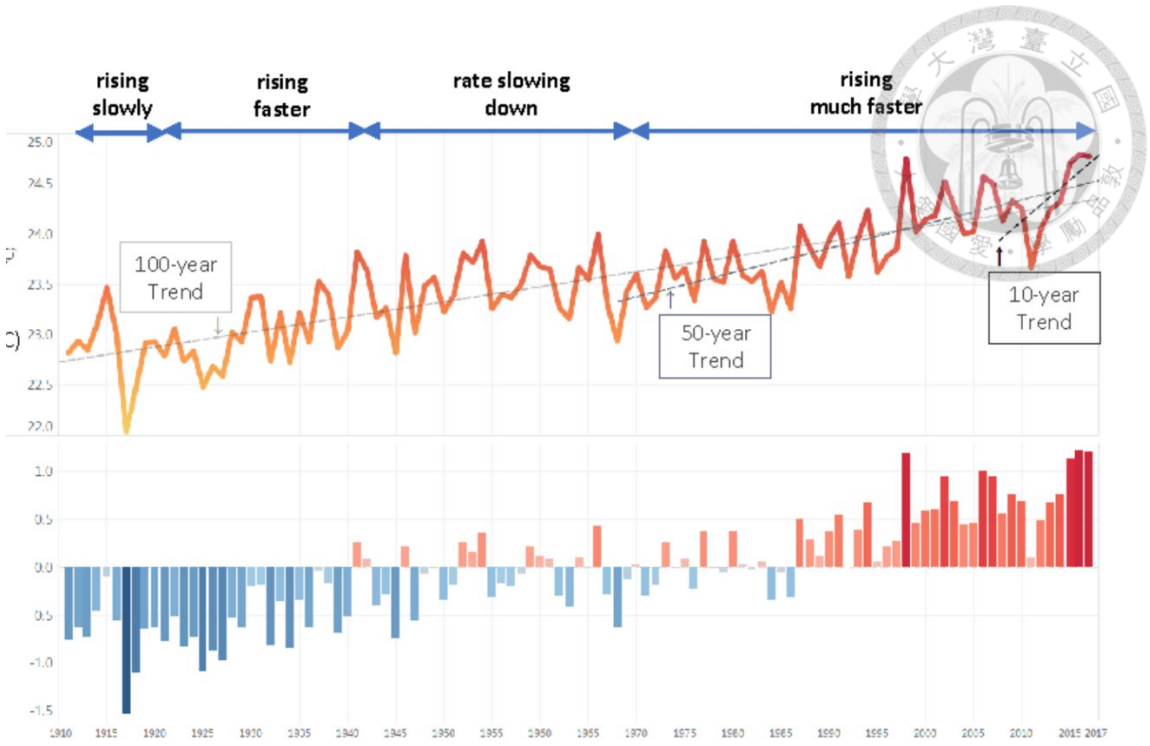


Figure 3. Climate change figure (TCCIP)

This research will use machine learning methods to estimate and calculate the CO_2 flux and latent heat flux in the Chi-Lan mountain (CLM) site, through some of the more commonly used models. By then, it is expected that the following goals can be achieved:

1. Establish climate models for flux predictions for CLM.

2. Explore the correlation coefficients between different inputs and outputs.

3. Filling the gap of the fluxes data in CLM.

doi:10.6342/NTU202300497

Chapter 2 Literature Review

2.1 Gap-filling techniques

In the study conducted by Moffat et al. (2007), the data from 10 stations were used, and the R^2 and RMSE on the day-night and half-hour and daily scales between the various gap-filling methods (**Figure 4**) were compared, respectively. It can be found that (**Figure 5**) the half-hour and R^2 during daytime were high and R^2 at nighttime was low, but the RMSE was the lowest for day-scale nighttime.

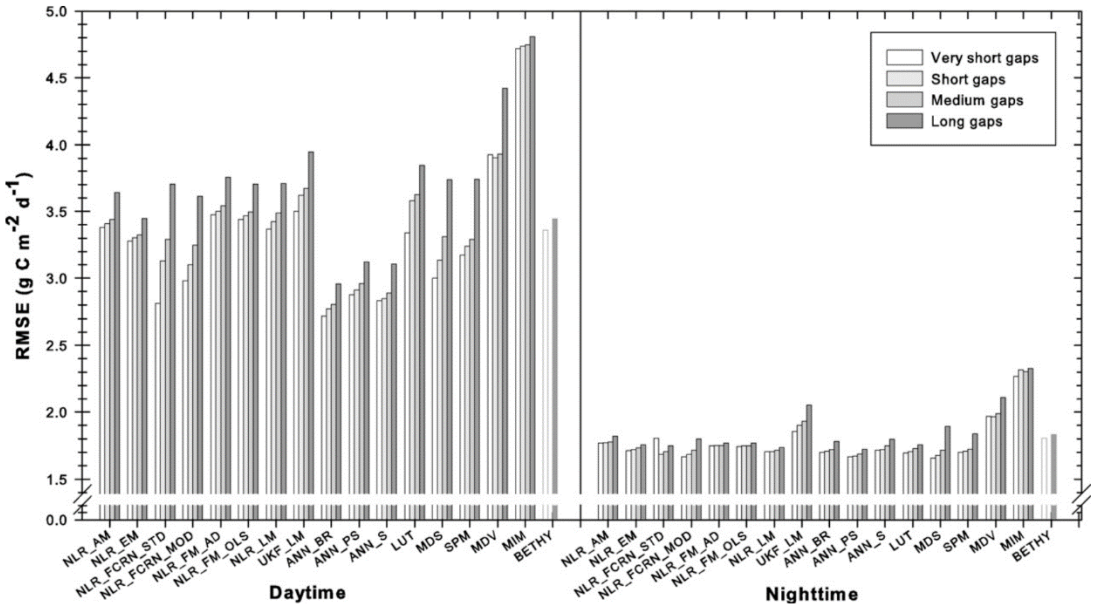


Figure 4. Comparison of the Gap filling method (Moffat et al., 2007)

doi:10.6342/NTU202300497

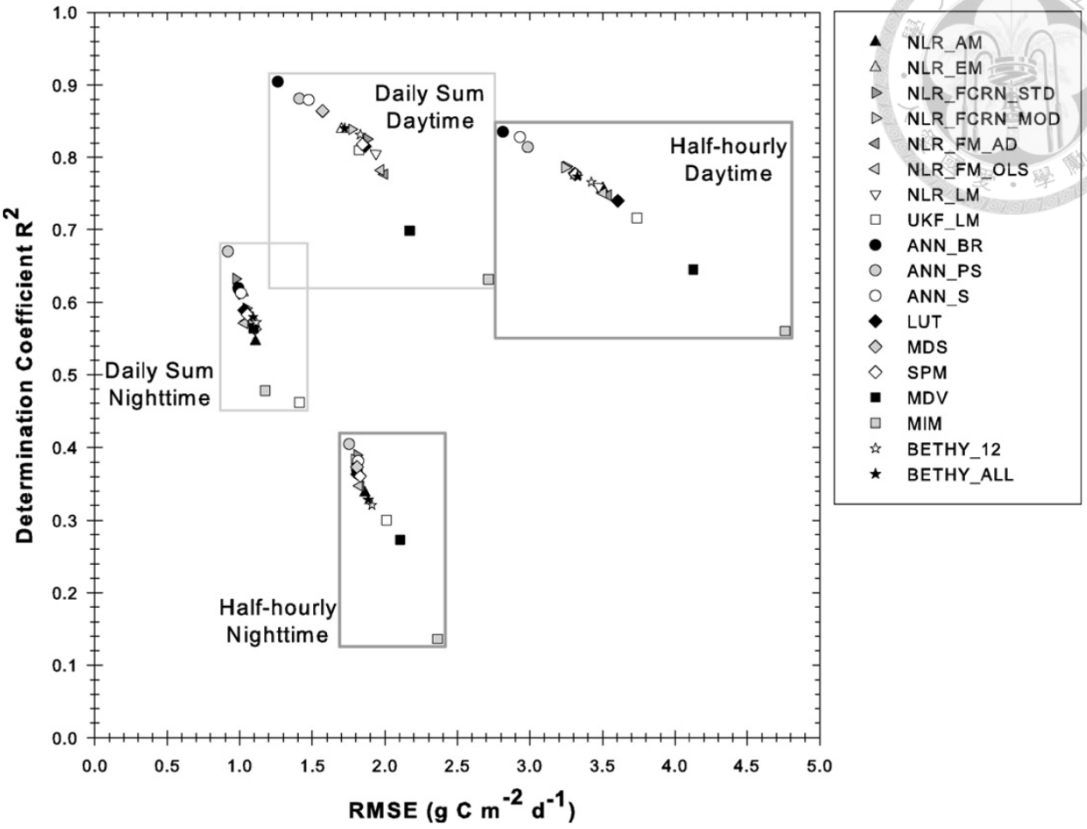


Figure 5. Comparing R^2 and RMSE (Moffat et al., 2007)

2.1.1 Non – linear regressions (NLRs)

Falge et al. (2001) summarized that many studies used one equation for ecosystem respiration (ER) and used the other equation for the light response of the ecosystem to quantify the gross primary production (GPP). For the condition during the nighttime, there was no sunshine. Therefore, through the equation NEE = GPP - ER with GPP = 0 in the night, that can fill the missing NEE values during the nighttime.

doi:10.6342/NTU202300497

2.1.2 Unscented Kalman filter (UKF)

The Kalman filter has been widely used and has provided an excellent method to simulate nonlinear datasets, and the unscented Kalman filter (UKF) was developed for time series where the data are autocorrelated (Gove & Hollinger, 2006). The method has two steps. The first step is that use noisy measurement data to continuously update nonlinear process model predictions, and the second step is to use the filter regression equations to predict the subsequent desired data. Finally, combine the predicted data and the observed data with improving the previous parameters. UKF was wildly used in many kinds of forest situations, different observed parameters, and different types of climates (Desai et al., 2008). However, the UKF has a severe disadvantage; if the dimensionality needs to be increased, the calculation time will become very long.

2.1.3 Look-up tables and further developments

A look-up table (LUT) method can memorize the data, and the method can provide estimates of model factors for each boom (Peng et al., 2012). In this case, the critical factors were stored in a table, and then when encountered the instrument malfunction or failure, searched the LUT and found the comparative figures filling the gap.

doi:10.6342/NTU202300497

2.1.4 The semi-parametric model technique(SPM)

The semi-parametric model technique (SPM) used a three-dimensional hypersurface for net CO_2, a nonlinear look table with light, temperature, and time (**Figure 6**); there are three values in three-dimensional, S_0 represents incident solar radiation, and T represents temperature. F_N is calculated by the following formula (Stauch & Jarvis, 2006):

$$F_n = f\{S_0, T, t\} + e \quad (1)$$

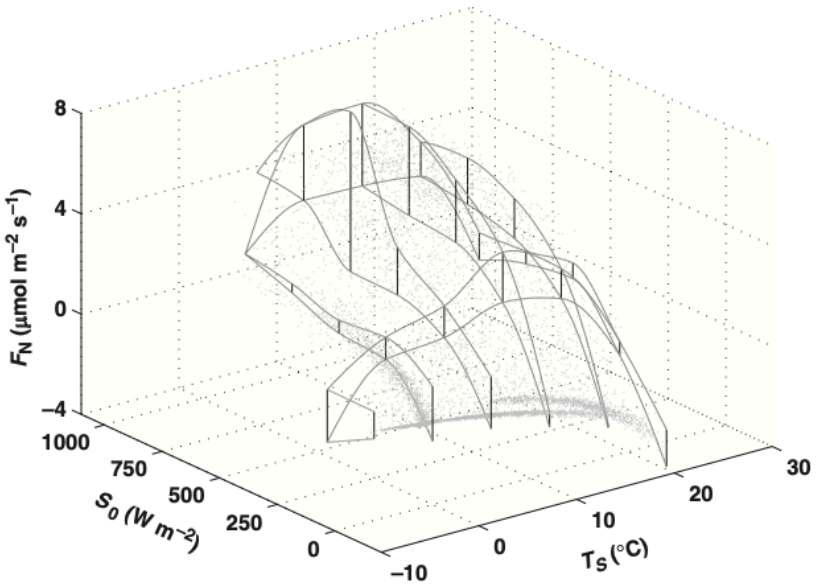


Figure 6. The concept of structure from SPM (Stauch & Jarvis, 2006)

doi:10.6342/NTU202300497

2.1.5 Mean diurnal variation (MDV)

Mean diurnal variation (MDV) is an interpolation technique. The missing data will be replaced with nearby day's data which were averaged at that time of day (Falge et al., 2001). In this article, it is shown that MDV performs better to fill in missing data for a 7-day interval during nighttime and a 14-day interval during daytime. However, when MDV is compared with the regression method, the performance of the annual sum is poor, but if other meteorological variables cannot be used, MDV is a good filling method.

2.1.6 Machine learning

machine learning (ML) techniques have overgrown in recent years (Braswell et al., 2005; Menzer et al., 2013; Schmidt et al., 2008); taking ANN as an example, this tool has good regression capabilities and has a good performance in nonlinear perdition (Huang & Hsieh, 2020). A neural network obtains the predicted output by inputting the previous data and approximating the observed value through the function after training (Wang, 2003).

Machine learning is not only applied to CO_2 and LE but is also widely used in FCH_4 (Kim et al., 2020). It can be seen from the figure (**Figure 7**) that machine learning has

doi:10.6342/NTU202300497

greater advantages over other traditional filling methods, no matter whether in long or short gaps, On Bias, can get good results.

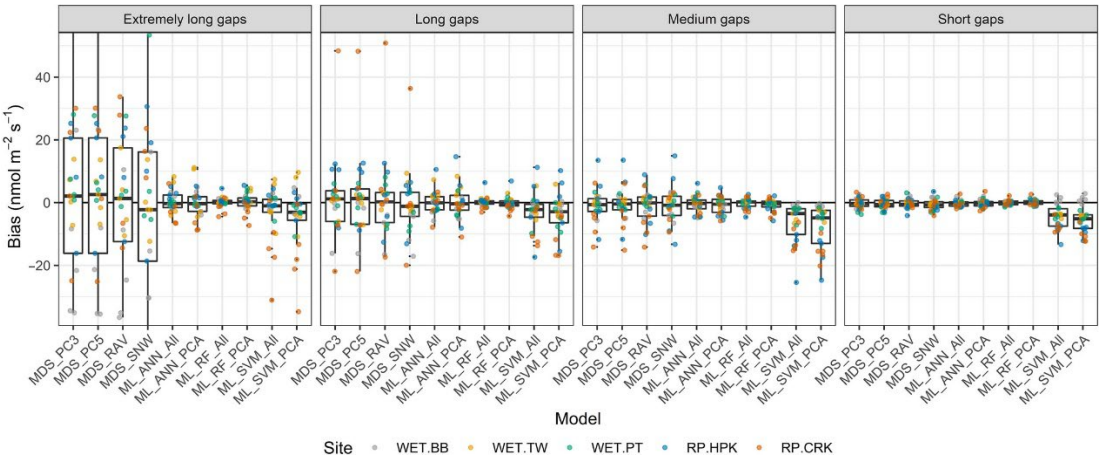


Figure 7. Half-hourly bias error (Kim et al., 2020).

This study (**Figure 8**) is mainly located in Canada, including three different ecosystem environments, the authors used four machine learning methods and a multiple regression model MLR, machine learning models including ANN, GRNN, ANFIS, SVM, it can be found that the predicted results of the machine learning models are better than MLR, but each model has its advantages in different stations. It can be found that ANN can get better results in CA-Obs, while ANFIS can get good results in CA-Oas, GRNN, and ANFIS can get good results in CA-Gro, so different models may have their suitable fields, but in general, machine learning can get good results in different testing stations.

doi:10.6342/NTU202300497

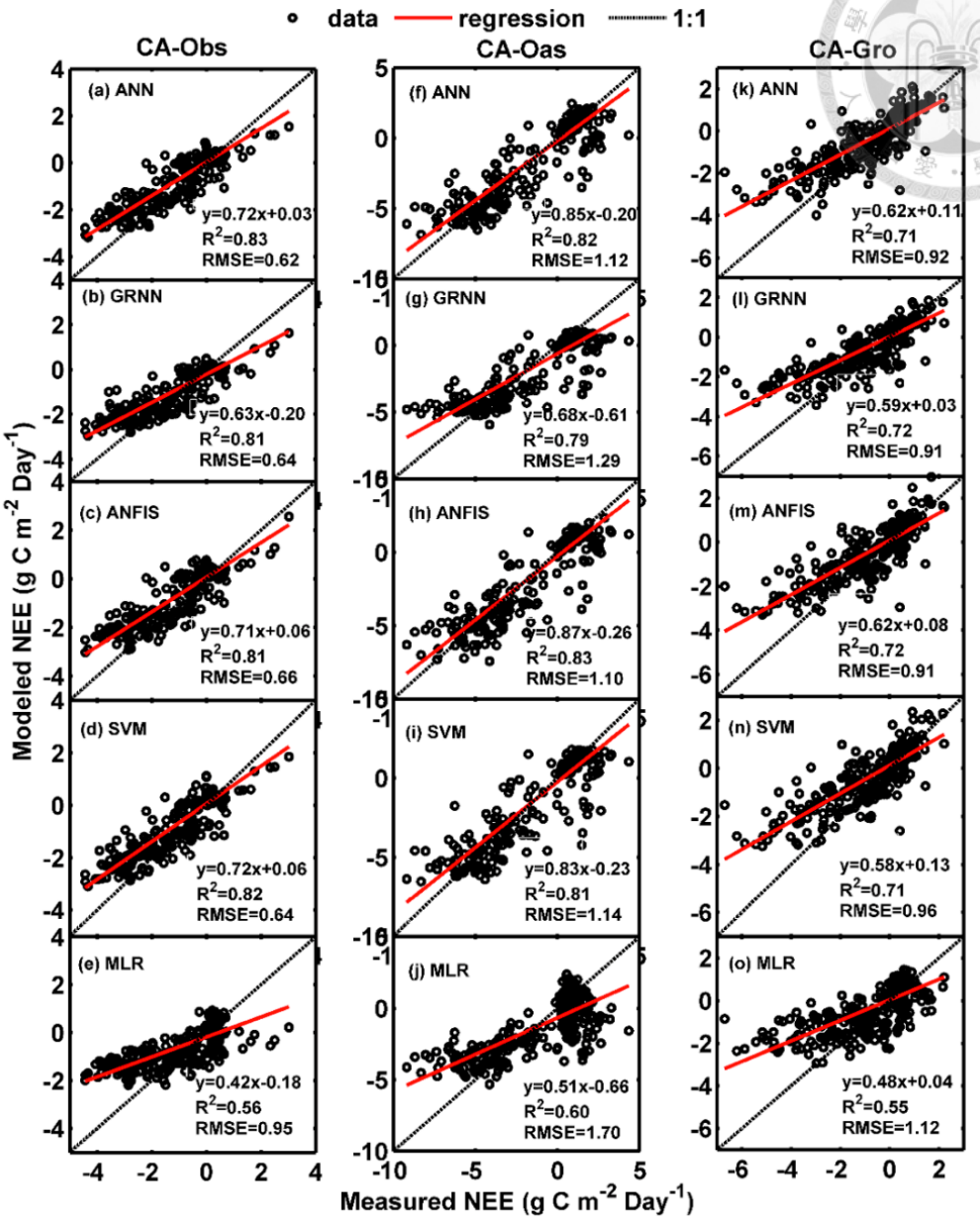


Figure 8. The results of the machine learning model and multiple linear regression model (Dou & Yang, 2018)

doi:10.6342/NTU202300497

2.2 CO_2 gap-filling method

2.2.1 The concept of the CO_2 gap-filling method

Correction approach method is based on the value of the friction velocity(u*) and selects a suitable friction velocity as an indicator for filtering data criteria and turbulence strength or weak, the method can be found in the following articles (Zhao et al., 2006), (Moureaux et al., 2006).

The filtered approach method is based on the reasonableness of the data quality and the elimination of the bad data, in which the environmental factors are matched and build their relationships model, finally, filling in the missing data.

2.2.2 Friction Velocity Correction (u_* correction)

It can be decomposed by the following equation, the vortex flow in the vertical direction can be expressed as (Stull 1988):

$$\tau_{xz} = -\rho\overline{u'w'} \quad (2)$$

$$\tau_{yz} = -\rho\overline{v'w'} \quad (3)$$

In the formula, τ means the momentum of the fluxes(kg/m/s^2), ρ means the air density(kg/m^3), u、v、w represents the x、y、z direction of the coordinate system.

doi:10.6342/NTU202300497

Momentum fluxes are also known as Reynolds stress ($\tau_{Reynolds}$, kg/m/s^2), and the formula can be rewritten as:

$$\left|\tau_{Reynolds}\right| = ((-\rho\overline{u'w'})^2 + (-\rho\overline{v'w'})^2)^{\frac{1}{2}} = \left(\tau_{xz}{}^2 + \tau_{yz}{}^2\right)^{\frac{1}{2}} \quad (4)$$

Reynolds defines the velocity of the eddy flux as the coefficient of friction, which is expressed as follows:

$$u_* = (\overline{u'w'}^2 + \overline{v'w'}^2)^{1/4} = (\frac{\left|\tau_{Reynolds}\right|}{\rho})^{1/2} \quad (5)$$

Friction speed is an indicator of the strength of turbulent flow development. If the friction coefficient is greater than the critical value, the value of carbon dioxide will usually be relatively stable. When there is a stable friction coefficient, the measured value will be more reliable.

doi:10.6342/NTU202300497

Chapter 3 Material and Method

3.1 Workflow

The flowchart of this study is shown in Figure 9. The procedures can be divided into the following steps. First, we processed the flux observation data and then conducted the data filtering and correction. To put data into machine learning, we must first process the data conversion and standardization before applying ANN and LSTM. After adjustment and training, we can build a working model, and analyze the results for daytime and nighttime trends of different input parameters. The results of the ANN models are then compared and analyzed for sensitivity, and finally, the model performance is checked by using different statistical processes.

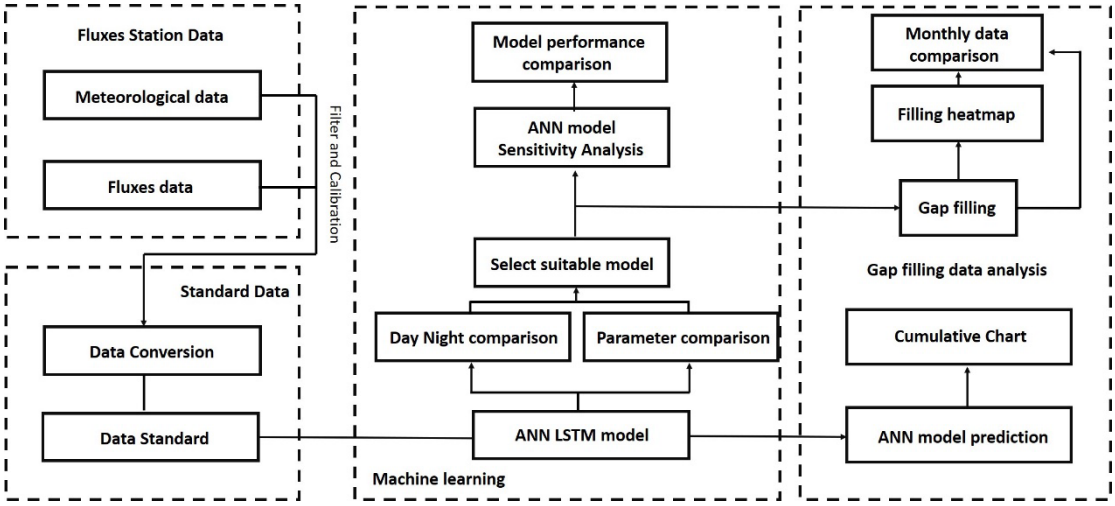


Figure 9. Workflow

doi:10.6342/NTU202300497

The CLM flux stations are then filled with data. The heat maps and monthly quartile comparison maps are plotted for analysis, and the best model is predicted for the full-time period and a four-year cumulative map via data analysis.

3.2 Study Site and Observation Data

The Chi-Lan Mountain (CLM) (**Figure 10**) site in northeastern Taiwan (24°35'N, 121°25'E), The elevation of the observation flux tower is 1650 meters, and it is a southeast-facing direction with an average slope of about degrees. The vegetation type is a warm-temperature coniferous and broad-leaved forest, with yellow cypress as the main dominant species, and the average tree height is about 11 to 14 meters (Chang et al., 2006). The meteorological parameters measured by the CLM flux tower include wind speed, air temperature, wind direction, air pressure, relative humidity, net radiant flux, sensible heat flux, latent heat flux, soil temperature, soil content, and soil heat flux.

doi:10.6342/NTU202300497

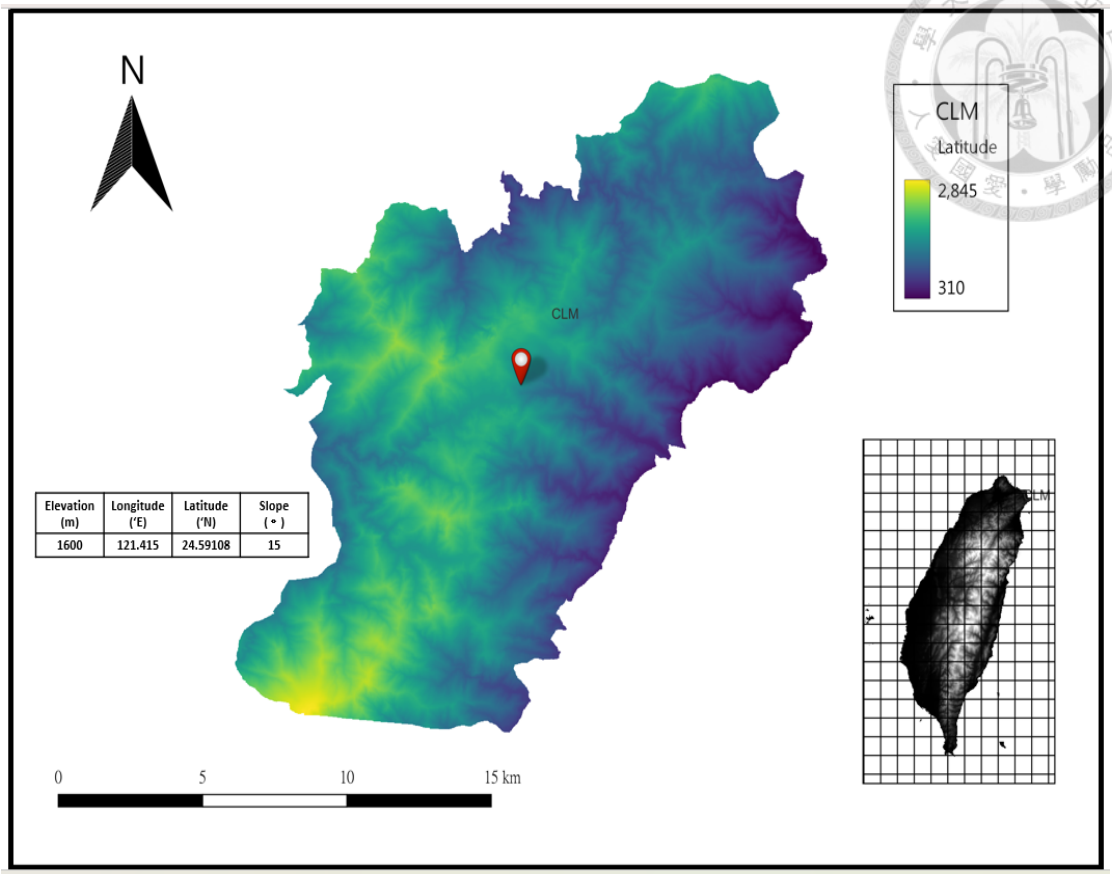


Figure 10. CLM station DEM and basic information.

CLM station height of about1600 meters, a latitude-longitude (121.412, 24.59108), and a slope is about 15 degrees, located in Yilan County, Taiwan

Eddy current correlation measuring instruments are installed at 24 meters above the ground surface, including an infrared gas analyzer (LI-7500 CO_2/H_2O analyzer, LI-COR Co., Lincoln, Nebraska, USA) and a three-dimensional ultrasonic anemometer. The raw data is sampled at 10Hz, and the data is transmitted to the data processor for storage. Raw data acquired at 10 Hz were processed using postprocessing, including spike removal, frequency response correction, sonic virtual temperature correction, the performance of the planar fit coordinate rotation, and corrections for density fluctuation (WPL correction).

doi:10.6342/NTU202300497

The temperature profile measurement system is a thermocouple thermometer (Thermocouple, Omega T-type) made by Donhua University's Hydrology and meteorology Laboratory to measure the vertical profile. It is divided into nine layers, the sampling frequency is 1Hz, and the average data of 2 minutes is stored in the data processor (CR23X, Campbell Sci., Logan, Utah, USA). The carbon dioxide and water vapor concentration measurement system are analyzed by a closed-circuit infrared gas analyzer at 11 meters, and the air from the 8-story air is pumped to 11 meters by an air extraction motor for analysis. The sampling frequency is 1 Hz, and it takes 2 minutes for data to reach the data processor (CR1000, LI-COR Co., Lincoln, Nebraska, USA).

Other meteorological data were set up with a visibility meter at 22 meters, a net radiometer at 22.5 meters, and a soil heat flux meter at a depth of 0.1 meters below the sample area.

CLM station is strongly affected by fog and clouds, (**Table 1**) and the missing data (data gaps) of meteorological factors are less than 2%, especially for wind speed, which is due to the face that the wind speed is weak at night, so the instruments may not be able to make accurate measurements.

doi:10.6342/NTU202300497

Table 1. Percentage of missing data during the entire study period

Gap%	Ta	Tsoil	Rh	Rn	Visibility	Wd	Ws
Day	1.3	1.3	1.3	1.3	1.4	1.3	6
Total	1.5	1.5	1.5	1.5	1.7	0.1	6
Night	1.8	1.8	1.8	1.8	1.9	1.8	7.4

The exchange of CO_2 in the ecosystem is mainly obtained through the following formula, where NEE is the net ecosystem exchange, and GPP is the gross primary production, which is mainly related to the ecological volume, and the ecosystem respiration, R_{eco}, which is mainly the function of biomass and the environmental parameters.

$$\mathrm{NEE} = \mathrm{GPP} + R_{eco} \quad (6)$$

The latent heat flux is mainly related to the evapotranspiration in the atmosphere, which can be derived through the following formula, where ET is evapotranspiration, ρ_w is the density of water vapor, and L is the latent heat of vaporization.

$$\mathrm{LE} = \mathrm{ET}\ \mathrm{x}\ \frac{1}{86400}\ \mathrm{x}\ \frac{1}{1000}\ \mathrm{x}\ \rho_w\ \mathrm{x}\ \mathrm{L} \quad (7)$$

doi:10.6342/NTU202300497

It can be found that CO_2 is negative during the day (**Figure 11**), and its time is mostly from 6:00 am after the sun rises to 5:30 pm, which has a great relationship with the sun's sunrise and sunset times because the exchange of carbon dioxide is mainly It is formed through photosynthesis, so the ecosystem absorbs carbon dioxide during the day, so that CO_2 enters the ecosystem during the day, so its value is negative, while it is positive at night because carbon dioxide respires at night. Therefore, carbon dioxide is discharged out of the ecosystem.

In the part of LE (**Figure 11**), the situation is the opposite. The peak of LE is around 10:00 am. This period is the time when the evapotranspiration in the CLM area is relatively strong. LE is not high, which can be verified from previous journals (Gu et al., 2021) there are many gaps at night, and the remaining ratio of day and night can be known from the figure (**Figure 12**), in the CO_2 part of the day, there are about 29.61% of the annual data, while the remaining 26.88% at night, and 29.61% and 26.88% respectively in the LE part, it can be speculated that there are more missing values at night, mainly because the turbulent flow is weaker at night, resulting in the value at night It is easy to be shaved off, so the missing values of CO_2 and LE account for 43.51% and 40.83% respectively.

doi:10.6342/NTU202300497

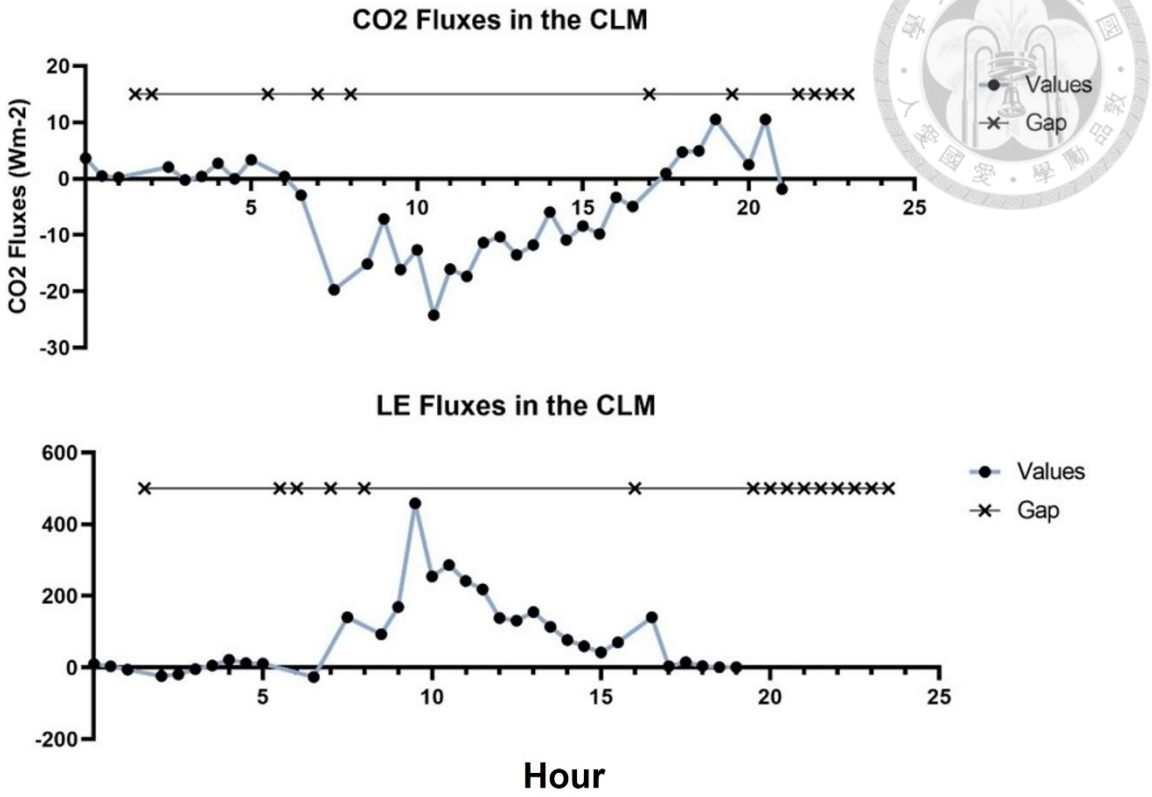


Figure 11. The ensemble average of CO_2 and LE time series in CLM

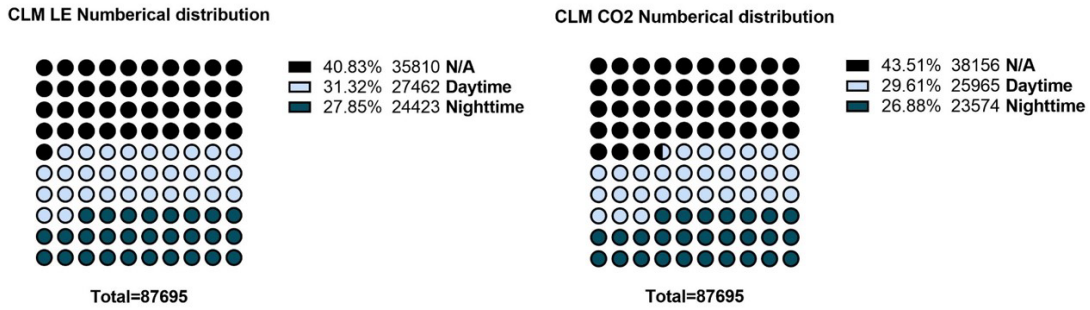


Figure 12. The gap precents of the CO_2 and LE in the CLM

The following is the heat map with missing data(**Figure 13**) in CLM. The X-axis represents the hour, and the Y-axis represents the day. It can be found that during the daytime, CO_2 presents a relatively deep negative value from 6:00 AM to 4:00 PM, and the more extreme value can be as high as – 40 ($\mu mol/m^2$), and in the daytime, it is mostly

doi:10.6342/NTU202300497

around 0 ($\mu mol/m^2$), and the highest can be as high as 40($\mu mol/m^2$), but it can be found that the larger positive values are mostly sporadic, which may be because there are many missing values in this period, so relatively large values are generated, while in LE Partly, it can be found that most of the values from 6:00 to 16:00 are positive, and the values are deeper and wider in summer and have a more obvious trend than CO_2. However, in 2011 The data errors in the first half of the year may be due to the mis planting of data at other times.

In summary, CO_2 has a larger range in the middle of the year, while LE has a larger range in summer, and the higher values between the two are concentrated at noon in summer, but most of the missing values can be found in the years are concentrated in autumn and winter, which may be affected by the weak turbulence in this season, while there are long gaps in summer, which may be due to incomplete data caused by typhoons or heavy rainfall.

doi:10.6342/NTU202300497

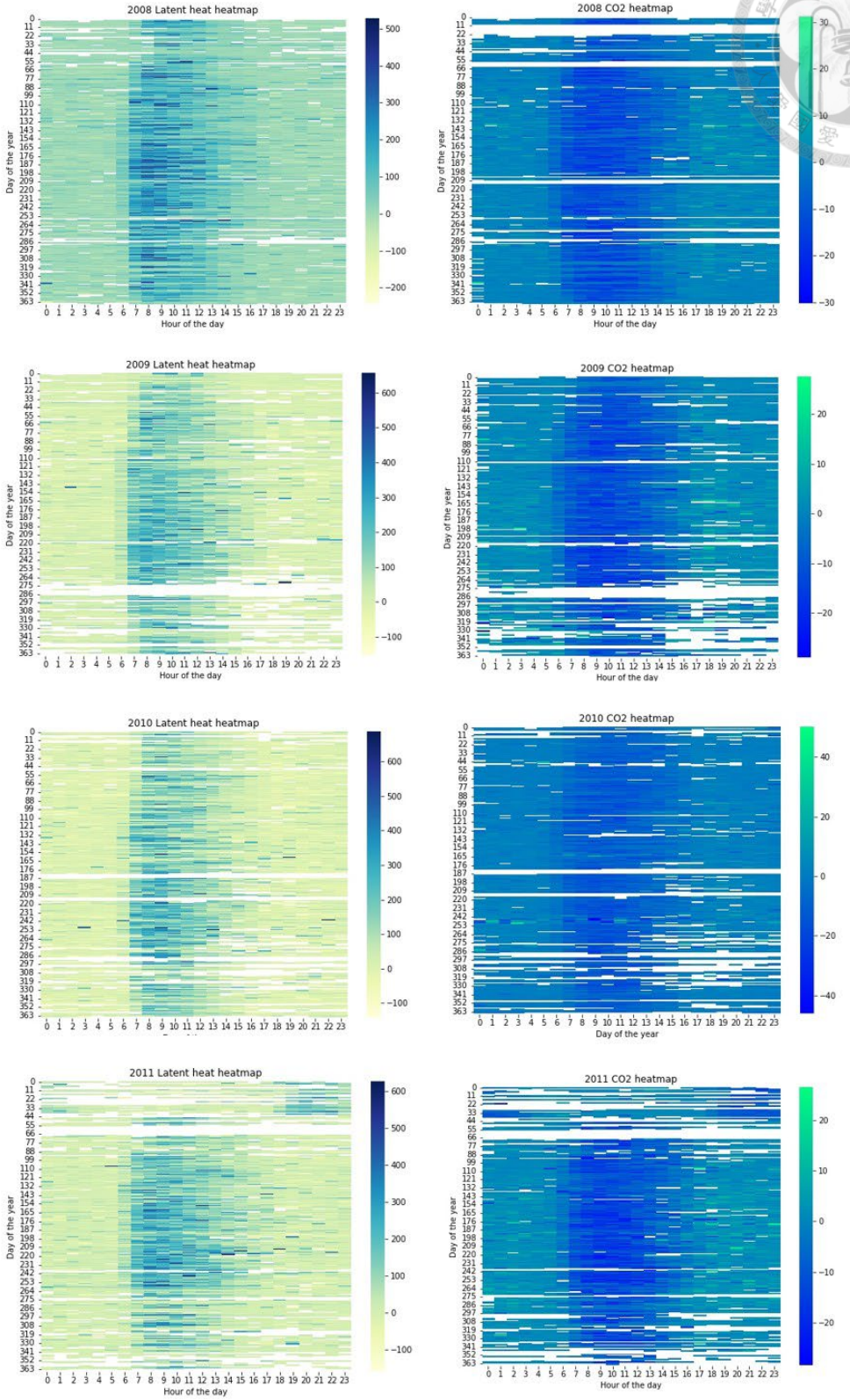


Figure 13. CLM latent heat heatmap

doi:10.6342/NTU202300497

3.3 Machine learning algorithms

3.3.1 Artificial neural network

In this study, the ANN architecture will be used to test different layers, activation functions, and various parameter adjustments to obtain a model suitable for CLM. However, because the training time of the ANN architecture is much shorter than that of the LSTM, the research selects a model that has a faster training speed and random distribution of data and a time series that requires more continuous data for comparison with LSTM.

ANN method (**Figure 14**) is a powerful tool to deal with such complex and challenging problems and has been widely used to study the mechanism of climate change and predict the trend of climate change(Zhang et al., 1998). The network structure

doi:10.6342/NTU202300497

includes three different types of layers. Arranged in order are the input layer, hidden layer, and output layer.

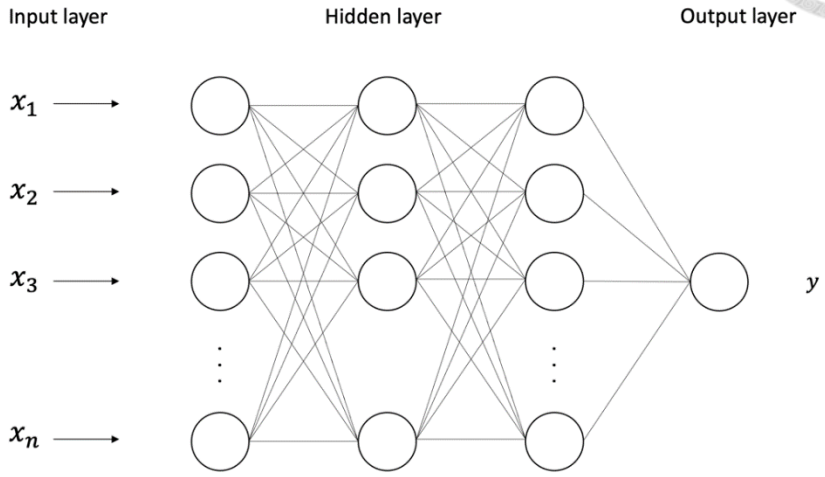


Figure 14. ANN structure

3.3.2 Long short-term memory (LSTM)

LSTM is a technology extended from RNN and invented by Hochreiter (Hochreiter & Schmidhuber, 1997). LSTM mainly improves some of the previous RNN problems (Ex: memory design) (Sherstinsky, 2020).

The technology can more effectively predict the long-term situation than other machine learning algorithms.

doi:10.6342/NTU202300497

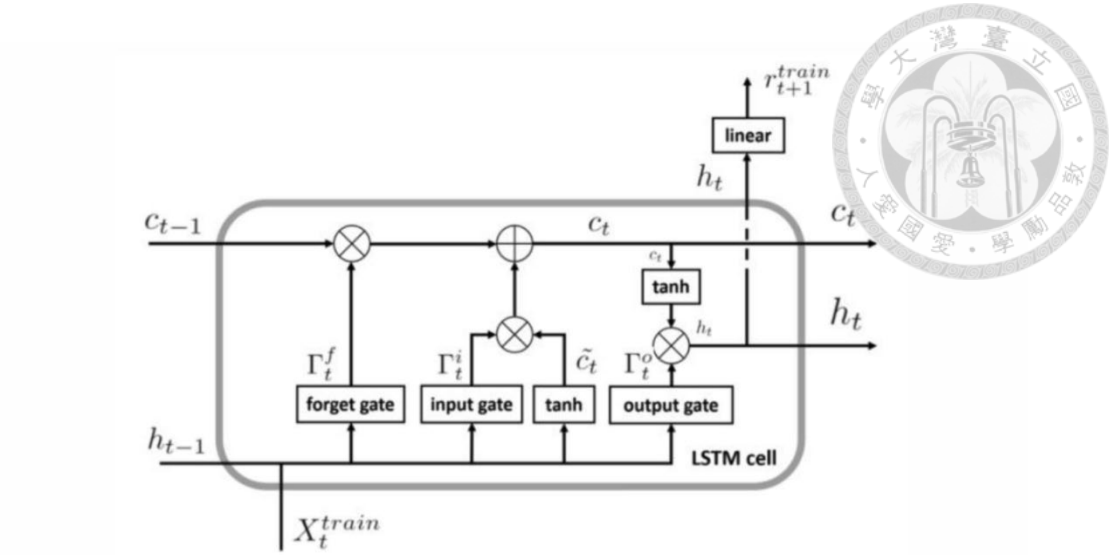


Figure 15. Long short-term memory

LSTM can predict trends in long-term gaps more effectively than other ML algorithms. The structure of the LSTM storage unit is shown in (**Figure 15**). The memory unit includes a forget gate, an input gate, and an output gate. The forget gate is used to filter whether the data memorized in the model is still valuable. Valuable data will be used for this round of predictions and retained for the next round. This process will be implemented through a Sigmoid function filter. The input gate and output gate are used to evaluate whether the new input data is valuable and serve as the output value of the neuron. In addition to the three-gate unit, there is a candidate value that will be used to determine the size of the updated neuron. X_t is the input vector to the LSTM model in this round of prediction; h_{t-1} and h_t are the prediction results of the previous and current round, $\oplus$ is a sequence summation, $\otimes$ is a matrix computing, google test the LSTM find the most crucial gate is forgotten gate, the second is input gate, the least is output gate.

doi:10.6342/NTU202300497

Each neural has an activation function, which is mainly used for nonlinear transformation and input to the next layer, where the feature in the data can be obtained in the process, and it is easier to obtain a nonlinear line.

In this study, the types of the activation function can see in (**Figure 16**)**.** The study mainly uses linear and tanh activation functions (Ramachandran et al., 2018). The linear function is more straightforward in terms of computation and partial derivation, but the Sigmoid and other activation functions are not excellent in this case.

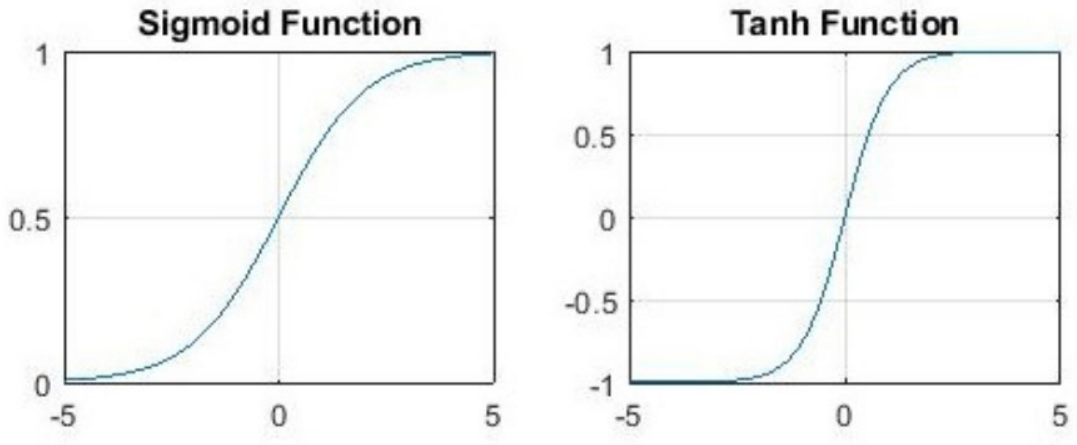


Figure 16. Activation function types ((Nwankpa et al., 2018))

3.3.3 Performance Metrics

The performance of the techniques was evaluated by comparing observed fluxes and predicted fluxes. The performance measures (Janssen & Heuberger, 1995) included the root mean square error (RMSE), R-Square, and mean absolute error (MAE). Those methods help us to know the performance of the prediction.

doi:10.6342/NTU202300497

R-squared is also known as the coefficient of determination. This method is used to measure the performance of the regression model and can represent the ratio of the independent variable x to the dependent variable y. The R-square can be calculated from the following function:

$$R^2 = \left(\frac{\sum_{i=1}^{n} (X_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{x})^2 (y_i - \overline{y})^2}} \right)^2 \quad (8)$$

Where X_i and y_i are the observed and estimated values, $\overline{x}$ and $\overline{y}$ are the averages of X_i and y_i, and n is the total of the parameters.

RMSE can be defined as the square root of the average of two differences. The lower the RMSE, the more accurate the prediction model. The RMSE value of the perfect prediction model will be 0. RMSE can be calculated from the following function:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2} \quad (9)$$

MAE is also suitable for expressing the error between the predicted value and the observed value. MAE can be calculated from the following function:

doi:10.6342/NTU202300497

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}| \quad (10)$$

Where y_i is predicted value of y.

3.4 Input Variables for Training the ML Models

The model used those data (**Figure 17**) including air temperature, soil temperature, relative humidity, net radiation, photosynthetically active radiation, visibility, wind direction, and wind speed. The difference in the amount of data is large, and the visibility values can be as high as around 3500(m), while other values are mostly less than 500.

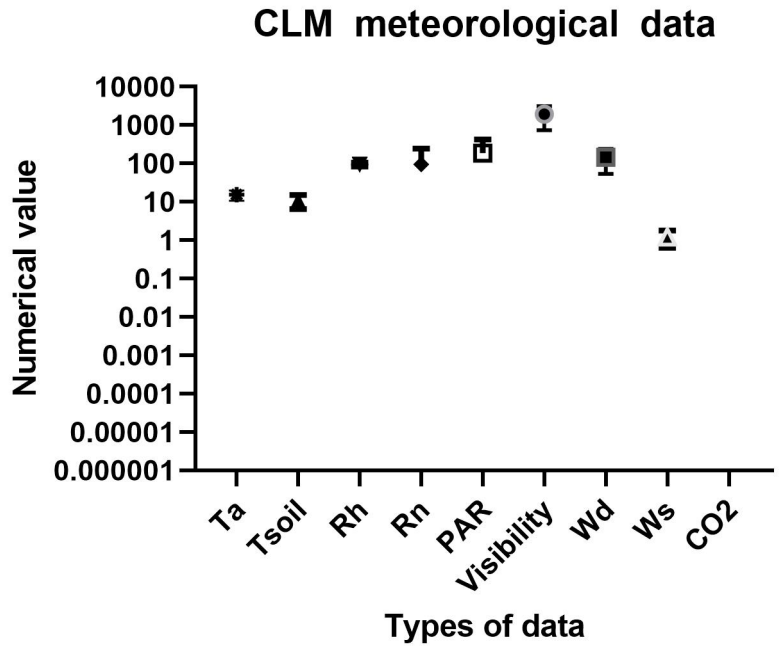


Figure 17. CLM meteorological data

doi:10.6342/NTU202300497

(**Table 2**) is the monthly average of all meteorological data in all CLM stations, among which Ta, Tsoil, Rn, PAR, Visibility, and Wd have the obvious seasonal change, the phenomenon is most obvious in the summer, which product a peak in that season.

Table 2. Input monthly average data in 2008

Year	Month	Ta	Tsoil	Rh	Rn	PAR	Vis	Wd	Ws
	Jan	8.98	4.72	96.04	65.08	102.59	1401.79	123.39	1.23
	Feb	6.17	2.96	97.7	54.88	74.6	1049.44	136.31	1.2
	Mar	10.66	4.55	86.91	77.89	138.27	1969.55	151.28	1.31
	Apr	15.28	8.15	91.75	96.17	173.25	1882.91	109.63	1.09
	May	16.44	9.26	91.38	81.83	211.41	2083.56	113.25	1.29
2008	Jun	19.37	11.68	91.13	89.46	217.97	2469.48	98.15	1.4
	Jul	20.1	12.71	92.02	78.4	220.5	2518.22	134.76	1.5
	Aug	20.08	12.74	90.49	101.76	237.62	2376.38	209.22	1.62
	Sept	18.84	12.43	95.31	82.32	171.25	2061.17	200.7	1.54
	Oct	16.78	11.1	97.56	70.06	136.48	1407.32	180.58	1.29
	Nov	12.16	7.95	96.05	54.71	115.12	1611.37	194.02	1.28
	Dec	8.67	4.3	88.35	61.97	141	1831.41	203.6	1.41
	Average	14.46	8.54	92.89	76.21	161.67	1888.55	154.57	1.346

Through the chart the study can find the relationship between the 8 parameters (**Figure 18**), most of the parameters increase in July and August, and the temperature has a high correlation with the soil temperature, but in the Rh, there is a peak value between

doi:10.6342/NTU202300497

September and October, and its value is close to 100%, while the value of Rn will decrease a lot between September and October, from about 100 (w/m^2) to about 50 (w/m^2). It is speculated that it should be related to the cloud and fog during this period. The formation is related, from the perspective of visibility, it can also be found that visibility drops very seriously during this period. In particular, the value of wind speed in winter is usually low, and it is concentrated at around 1.2 (m/s) in four years in October. It is speculated that this may impact the gap The main reason is that the wind direction shows that the value in 2008 is much lower than other years on average, which may be a problem with the instrument.

doi:10.6342/NTU202300497

Table 3. Input monthly average data in 2009

Year	Month	Ta	Tsoil	Rh	Rn	PAR	Vis	Wd	Ws
	Jan	6.17	1.72	95.43	60.71	124.3	1556.73	190.93	1.38
	Feb	13.25	5.24	87.19	97.04	202.66	1996.56	207.88	1.15
	Mar	12.44	10.2	89.86	111.112	206.91	1812.58	194.59	1.2
	Apr	12.31	10.49	96.54	71.66	131.35	1608.78	202.34	1.52
	May	16.34	12.25	82.88	124.68	249.37	2429.13	212.23	1.43
	Jun	19.32	14.97	90.07	98.56	199.21	2431.35	220	1.54
2009	Jul	20.52	16.78	91.97	109.43	216.31	2345.49	201.14	1.52
	Aug	20.59	17.91	92.96	99.77	197.11	2379.45	205.2	1.55
	Sept	18.8	16.73	95.08	87.5	163.55	1917.06	191.97	1.55
	Oct	15.41	14.71	96.91	56.3	100.37	1146.22	171.54	1.32
	Nov	12.76	12.14	96.18	64.5	118.5	1487.59	189.82	1.41
	Dec	7.93	8.55	95.55	55.41	108.11	1408.22	186.09	1.23
	Average	14.65	11.80	92.55	86.38	168.14	1876.59	197.81	1.4

doi:10.6342/NTU202300497

Table 4. Input monthly average data in 2010

Year	Month	Ta	Tsoil	Rh	Rn	PAR	Vis	Wd	Ws
2010	Jan	8.83	7.25	90.67	61.72	127.8	1818.24	200.37	1.31
	Feb	11.28	8.37	87.34	68.47	147.49	1976.3	201.28	1.28
	Mar	12.23	9.7	89.62	97.75	181.84	1999.27	209.97	1.31
	Apr	13.94	10.89	93.65	93.99	156.61	1845.54	199.47	1.21
	May	17.78	13.58	93.26	89.67	155.61	2044.51	208.18	1.18
	Jun	19.06	15.22	90.42	82.79	143.56	2301.68	210.02	1.24
	Jul	20.87	16.87	88.11	105.32	195.62	2634.99	225.82	1.55
	Aug	20.29	17.09	92.26	89.67	177.68	2476.45	214.73	1.6
	Sept	19.07	16.71	94.94	94.19	161.33	2215.84	199.38	1.51
	Oct	15.54	14.93	98.48	57.41	86.68	1411.37	178.91	1.25
	Nov	11.78	11.32	96.87	52.88	80.57	1310.19	186.63	1.32
	Dec	8.62	8.36	85.69	54.72	116.36	1936.78	204.56	1.4
	Average	14.94	12.52	91.77	79.04	144.26	1997.59	203.27	1.34

doi:10.6342/NTU202300497

Table 5. Input monthly average data in 2011

Year	Month	Ta	Tsoil	Rh	Rn	PAR	Vis	Wd	Ws
2011	Jan	5.03	5.35	100	47.7	60.96	985.53	178.29	1.09
	Feb	8.92	6.37	93.24	76.27	121.52	1762.07	195.76	1.16
	Mar	7.77	7.01	99.39	67.22	92.73	1115.32	177.79	1.02
	Apr	13.58	9.17	85.1	103.56	161.89	2145.35	213.44	1.33
	May	17.7	13.41	92.26	91.1	147.57	2187.84	208.78	1.22
	Jun	20.22	15.98	92.24	98.94	166.61	2432.27	215.37	1.45
	Jul	20.43	16.83	91.46	97.64	169.59	2447	222.87	1.51
	Aug	20.45	16.86	87.94	118.95	200.03	2533.12	214.35	1.68
	Sept	18.23	15.69	90.12	102.84	170.65	2036.28	203.01	1.53
	Oct	15.16	14.35	98.84	51.93	73.71	1137.77	177.19	1.13
	Nov	14.35	13.01	98.06	55.65	82.95	1473.9	184.5	1.29
	Dec	8.15	8.95	100	42.23	45.51	368.57	142.91	1.28
	Average	14.16	11.91	94.05	79.05	124.47	1718.75	194.52	1.3

doi:10.6342/NTU202300497

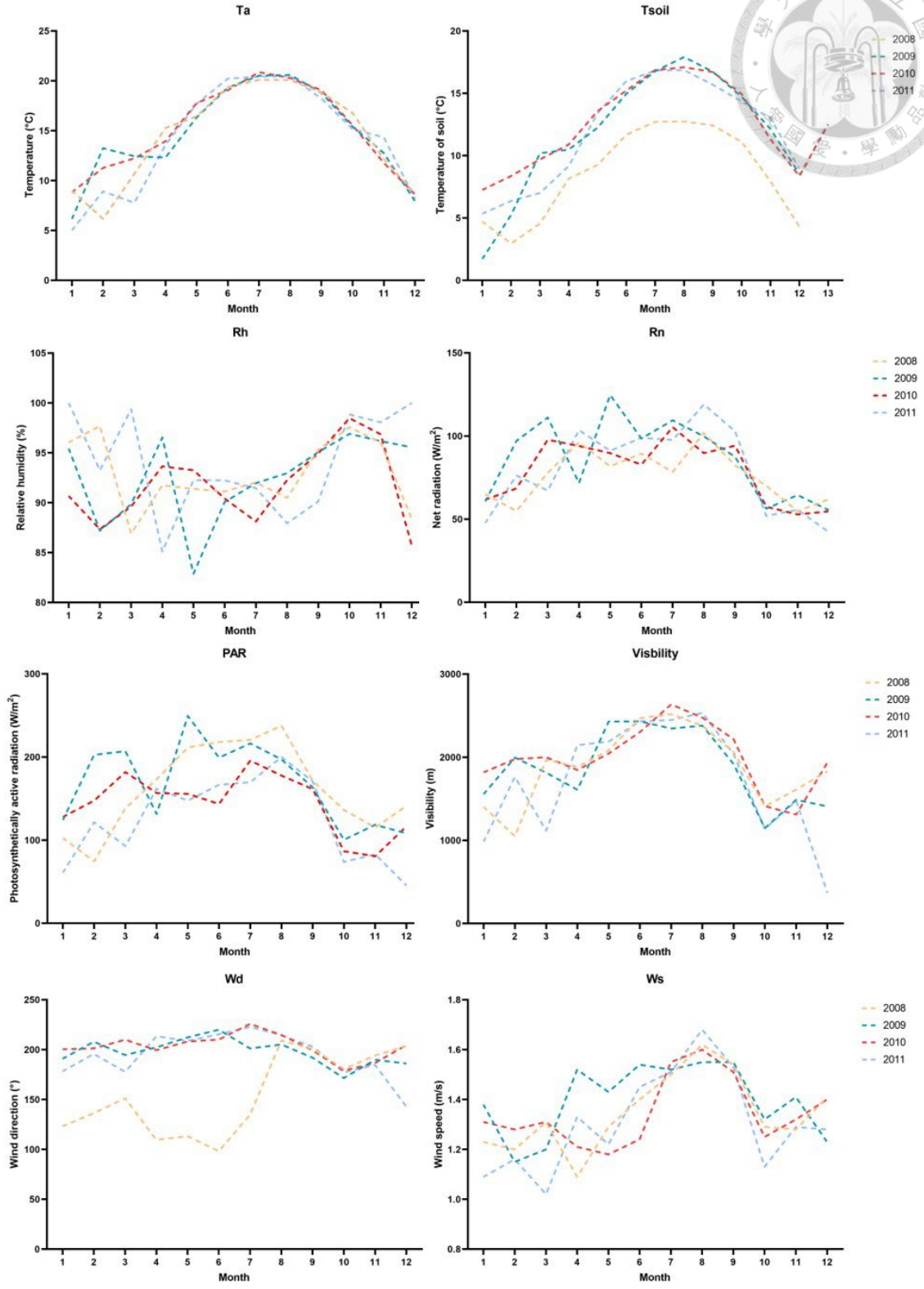
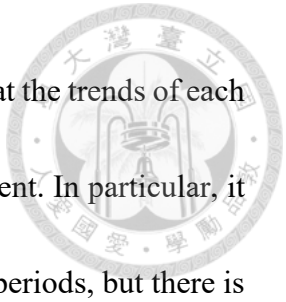


Figure 18. Input data of monthly scale

doi:10.6342/NTU202300497



However, from the half-hour data (**Figure 19**), it can be found that the trends of each parameter are quite similar, representing a relatively stable environment. In particular, it can be found that some data in PAR in 2008 are higher than in other periods, but there is no special increase in Rn. It can be as high as 1350 (W/m^2), I think it may be the relationship between the instrument, which also affects the monthly average value in May and June, which is higher than that in other years.

doi:10.6342/NTU202300497

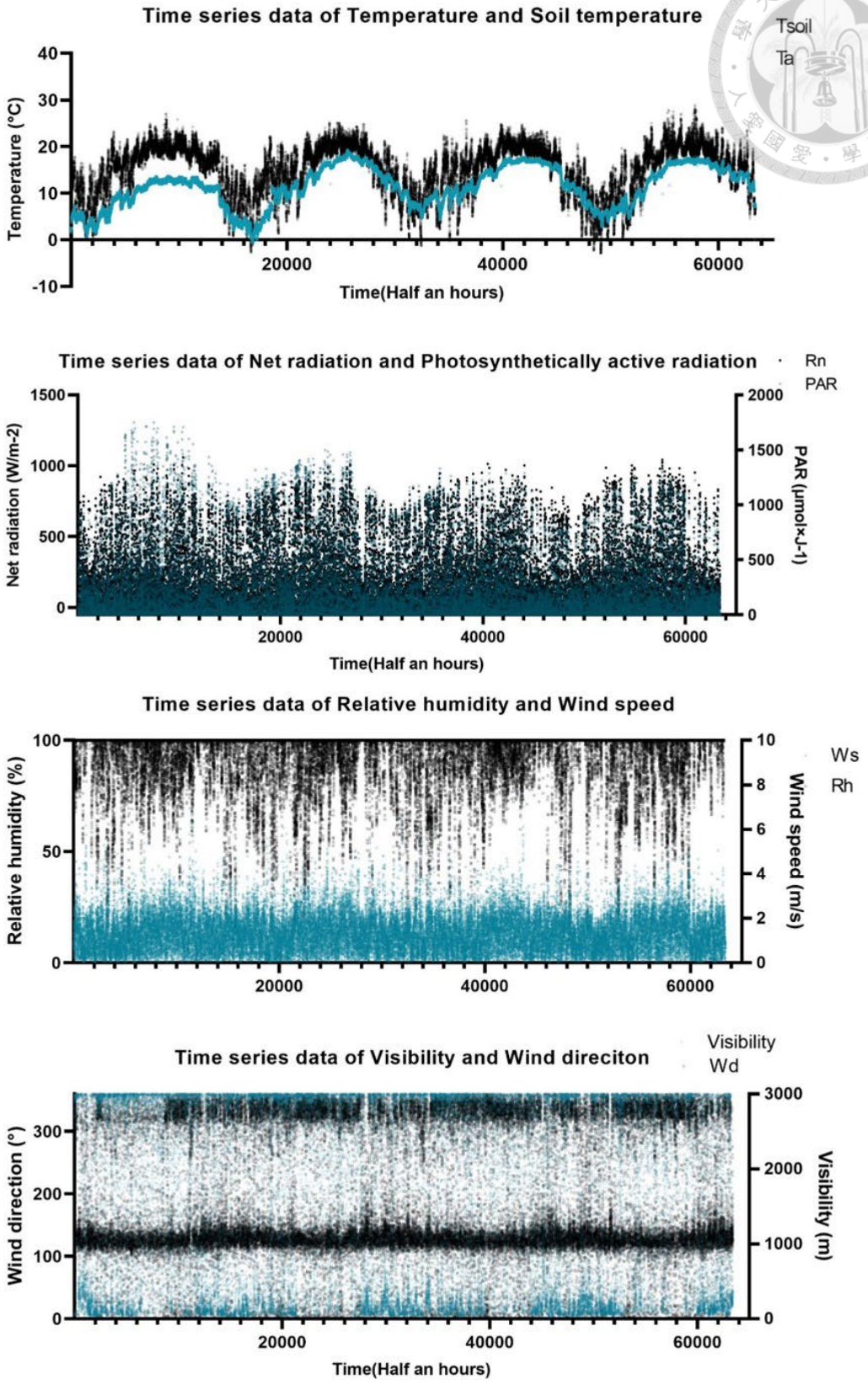


Figure 19. Meteorological figures

doi:10.6342/NTU202300497

3.5 Partial Derivation (PaD)

Derivate one of the variables in a multi-variable function through partial differentiation to obtain the constant of other variables. It is mostly used in vectors and widely used in machine learning because the slope of the parameters can be known, and then the relationship between the parameters can be understood. The relationship between and the degree of influence on the output.

A multivariable function that derivation one of the variables while keeping the other variable constant, if $(a, b) \in \text{Dom } f$, then the partial derivative of f at (a, b) for x is the definition that can be written as the following function:

$$\frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{f(a+h, b) - f(a, b)}{h} \quad (11)$$

the partial derivative for y is:

$$\frac{\partial f}{\partial y} = \lim_{h \to 0} \frac{f(a, b+h) - f(a, b)}{h} \quad (12)$$

Wherein, h and k are variable values. In this way, a poly variable function value can be calculated concerning one of the variables, while keeping the other variables constant.

doi:10.6342/NTU202300497

Chapter 4 Results and discussions

4.1 Correlation coefficients for input parameters

Before the training of machine learning, the main study was carried out to plot the correlation between individual parameters (**Figure 20**) which included 8 parameters, Ta, Tsoil, Rh, PAR, Visibility, Wd, Ws, CO_2, and LE respectively for the input parameters. The main reason is that there is a strong correlation between photosynthesis and Rn, so it is assumed that Rn and PAR also play an important role in the machine learning model. However, it can be seen that there is a high positive correlation between Rn and PAR, with values as high as 0.67 and 0.62. The formula shows that LE is also controlled by Rn, and PAR and Rn are already highly correlated, so these two values play important roles in the machine learning model, but other parameters including Wd and Ws are somewhat correlated with CO_2 and LE, which may have a greater impact on the model.

doi:10.6342/NTU202300497

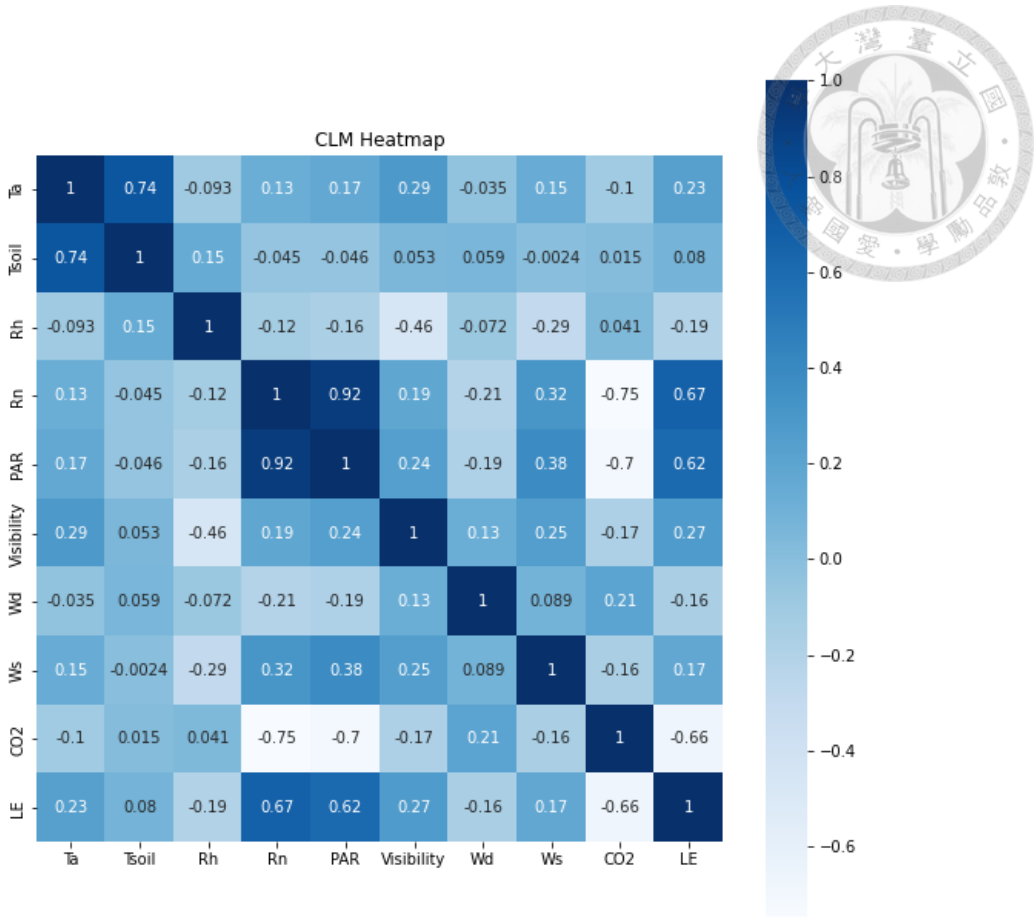


Figure 20. Heat map of correlations between parameters

Furthermore, the R^2 of all the inputs and outputs were plotted in **Figure 21** and **Figure 22**. It was found that most of the parameters had no linear relationship with CO_2, and most of the R^2 values were close to 0. It is probable that the period used in the ML data covers the whole period, including both daytime and nighttime. However, some CLM areas at nighttime are more susceptible to the influence of many factors, which may make the correlation between the parameters not high. If the model uses the parameters separately, the ML model may have good results. The analysis result shows a clear boundary in visibility at around 3000 (m), mainly because the visibility in the area can be

doi:10.6342/NTU202300497

as high as 3000 m for un-foggy conditions. Furthermore, there are obvious values in the middle section and 360 (degree) section of the wind direction, mainly because the CLM area is dominated by the mountain-valley wind system over different times of the day. Therefore, the number of values in this period will be more than the other wind directions, and the R^2 map can also be used as a reference point to characterize the difference between machine learning and the generally linear relationship.

However, in the LE section (**Figure 23**, **Figure 24**), the results are similar to those of CO_2, with the highest R^2 for input and output being Rn and PAR, with values of 0.5450 and 0.4796, respectively. It shows that LE is also influenced by solar radiation in CLM, with stronger evaporation at noon, which further affects the potential heat flux in the CLM. This indicates that LE is also affected by solar radiation in the CLM, with strong evaporation during midday, and further affects the exchange of potential heat flux in the ecosystem.

The relationship between each parameter and LE is shown in **Figure 23**, and **Figure 24**. It is found that some values of RH exceed 100 (%), which is a more unreasonable value. The temperature and soil temperature have a similar trend, mainly because the soil temperature near the surface is more influenced by the atmospheric temperature, but the data were filtered in the subsequent data to facilitate the model computation and remove the extreme values.

doi:10.6342/NTU202300497

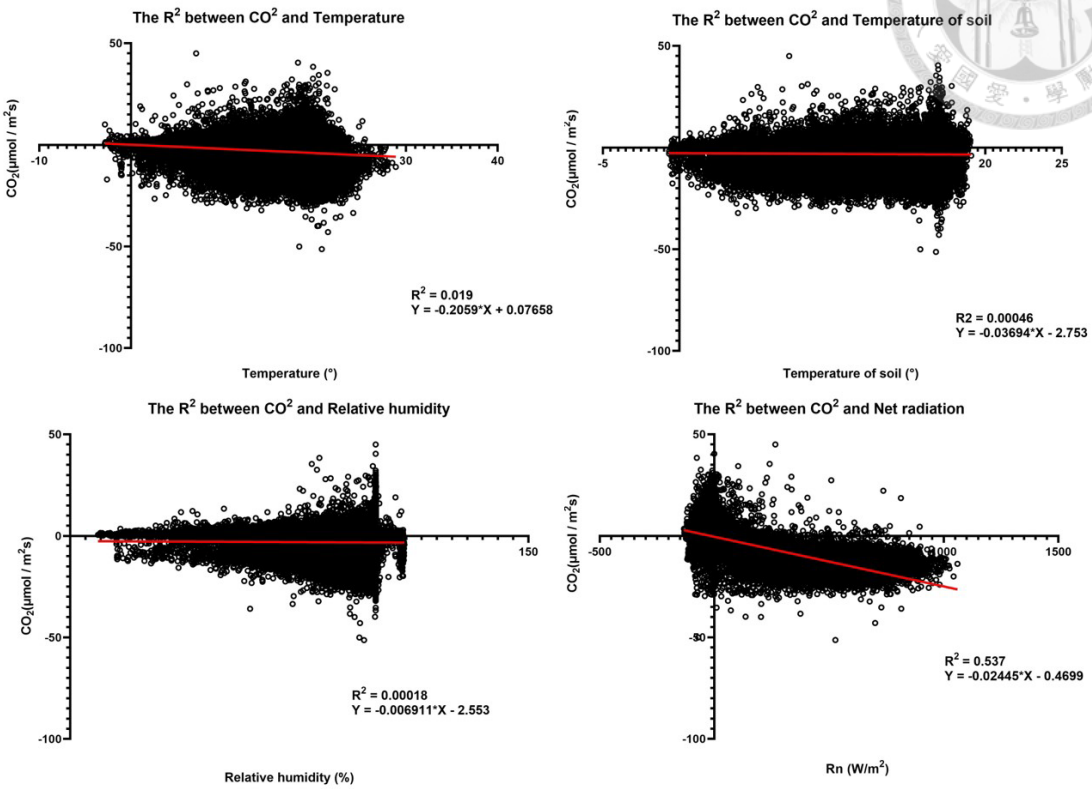


Figure 21. The R^2 between Ta、Tsoil、Rh、Rn and CO_2

doi:10.6342/NTU202300497

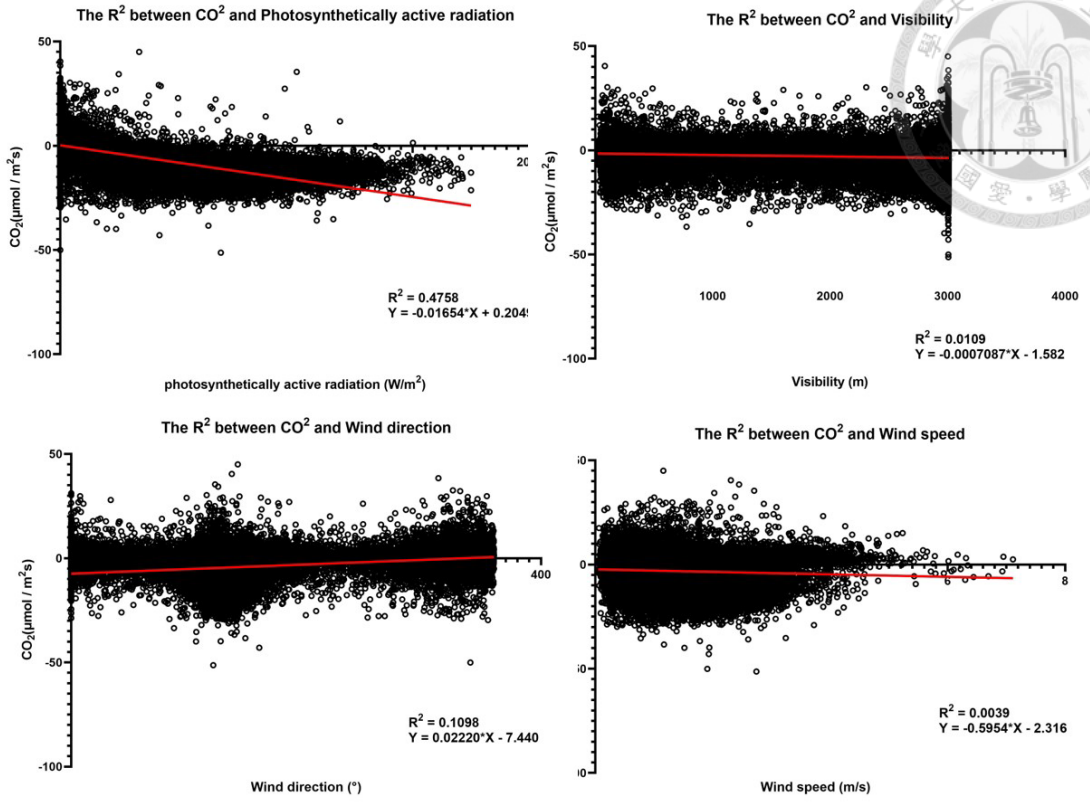


Figure 22. The R^2 between PAR、Visibility、Wd、Ws and CO_2

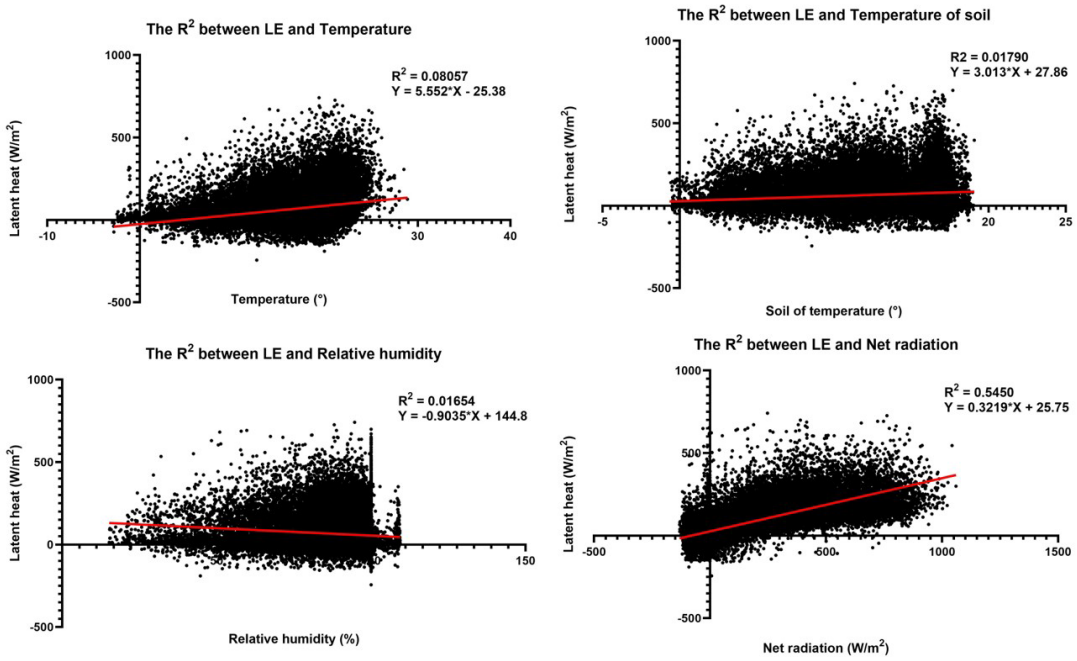


Figure 23. The R^2 between Ta、Tsoil、Rh、Rn and LE

doi:10.6342/NTU202300497

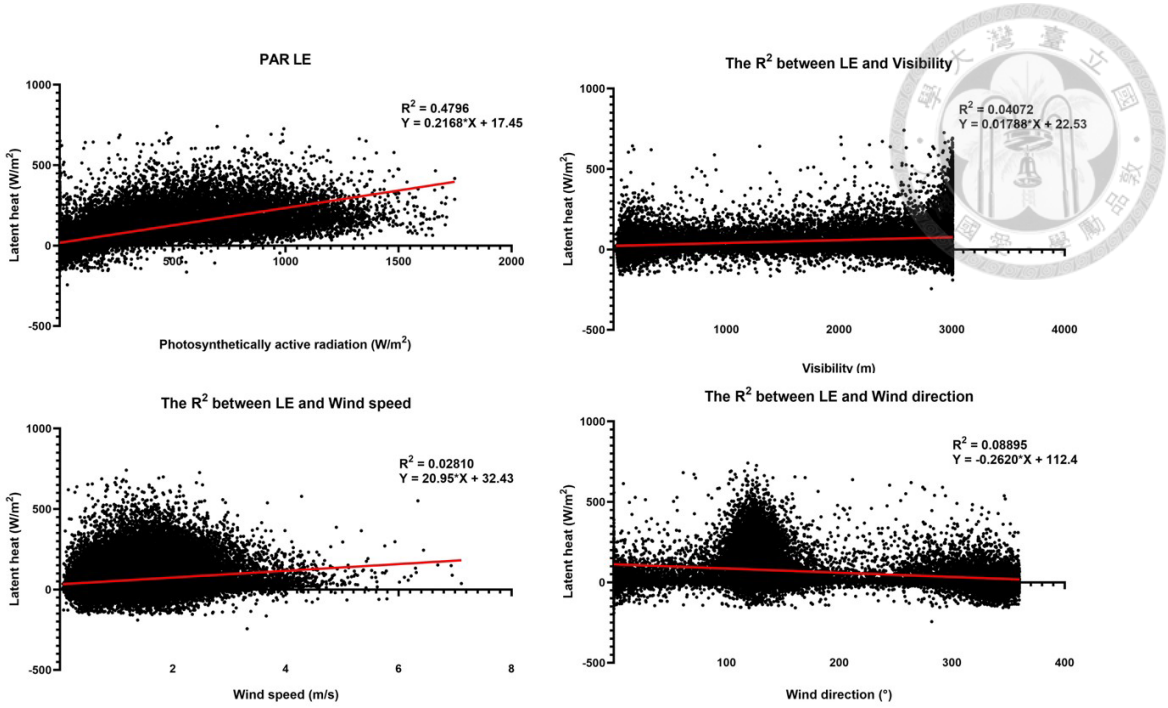


Figure 24. The R^2 between PAR、Visibility、Wd、Ws and LE

4.2 Machine learning results

Before the training, the data were filtered to remove the extreme values. If the value is larger than 97% or smaller than 3%, which is slightly higher than two standard deviations, in the training completion, the data are standardized. It is less likely that the data with too large a discrepancy will dominate the model.

4.2.1 Compare the R^2 relationship between different inputs

The following is the architecture of ANN (**Table 6**). ANN uses four layers, including an input layer, two hidden layers, and an output layer. The maximum shape used is 64. If too many nodes are used, the performance of the model may decline. However, After testing, the activation functions that can be successfully predicted in this research include

doi:10.6342/NTU202300497

linear and tanh. Other functions may not be able to be effectively predicted due to their limitations. The Batch size used is 100, which represents the data that machine learning has seen once. If the model uses a smaller Batch size, the training time will be longer, and Epochs are the number of times the model has to see the data completely. If the model uses a longer time for simulation, the model will run to the specified number of times, and it will stop. In this study, the data is divided into the training set and test set, and the proportions are 80% and 20% respectively. In the past literature, the proportion of this division is recommended, mainly because the training set needs to have enough data for training, and the test set is an appropriate amount of data needed to verify the correctness of the training results and ensure overfitting, which reduces the applicability of the data. In the ANN model, the data will be randomly assigned to random numbers to reduce the impact of the time series.

doi:10.6342/NTU202300497

Table 6. Neural network architecture

Layer	Output shape	Parameters	Activation function
Input layer 1	Units = 64	320	linear
Hidden layer 2	Units = 16	1040	tanh
Hidden layer 3	Units = 4	68	tanh
Output layer 4	Units = 1	5	linear

Other information

Total parameters: 1433

LOSS = MSE

Learning rate: 0.001

Batch size = 100

Epochs: 500

Validation split = 0.2

The model architecture of LSTM is shown in **Table 7**. It shows that the number of layers is the same as four layers, but there is a big difference in the activation function and other settings. The activation functions used include linear and relu. After testing, it shows that using relu is better than other functions, mainly because adding a little negative value to the time series will help to make good use of data that is nonlinear and contains signs. However, in this part, if the model uses a generally determined optimizer like Adam, it will be unpredictable, but you need to use the optimizer of SGD so that the training set will not be overfitted.

In addition, we found that training can not reduce the LOSS, mainly because LSTM is a time series data model, so for 80% and 20%, it will cause the distribution continuity

doi:10.6342/NTU202300497

of the data, so it is necessary to use the SGD method to reduce the learning rate of the model over time, and the model uses MSE to reduce the LOSS of the data just like ANN to obtain close values.

Table 7. LSTM architecture

Layer	Output shape	Activation function
Input layer 1	Units = 16	linear
Hidden layer 2	Units = 8	relu
Hidden layer 3	Units = 4	relu
Output layer 4	Units = 1	linear

Other information

Learning rate: SGD: 0.01, decay = 1e-6, momentum = 0.9

Loss = MSE

Batch size = 500

Epochs: 100

Validation split = 0.2

The comparison of the results for CO_2 prediction by ANN and LSTM is shown in **Figure 25**, and it shows similar results with the journal published in Dou and Yang (2018), the increase of parameters helps to improve the results. However, in the LSTM model, the R^2 is increased from 0.7095 to 0.7138, which is an increase of 0.0043. The upper and lower bounds are different for different inputs. For CO_2 with 4 inputs, the upper bound is about 2~3 ($\mu mol/m^2$), but for 6 and 8 inputs, the upper bound is between 3~4 ($\mu mol/m^2$). In general, it can be found that ANN outperforms LSTM in terms of R^2 and MSE regardless of the input. This may be because LSTM itself is not a random model but a

doi:10.6342/NTU202300497

time series model, when the data is split into 80% and 20%, it may cause the data to change seasonally in the middle of the season, instead of using the full-year to predict, so this study may conduct a more in-depth study on ANN.

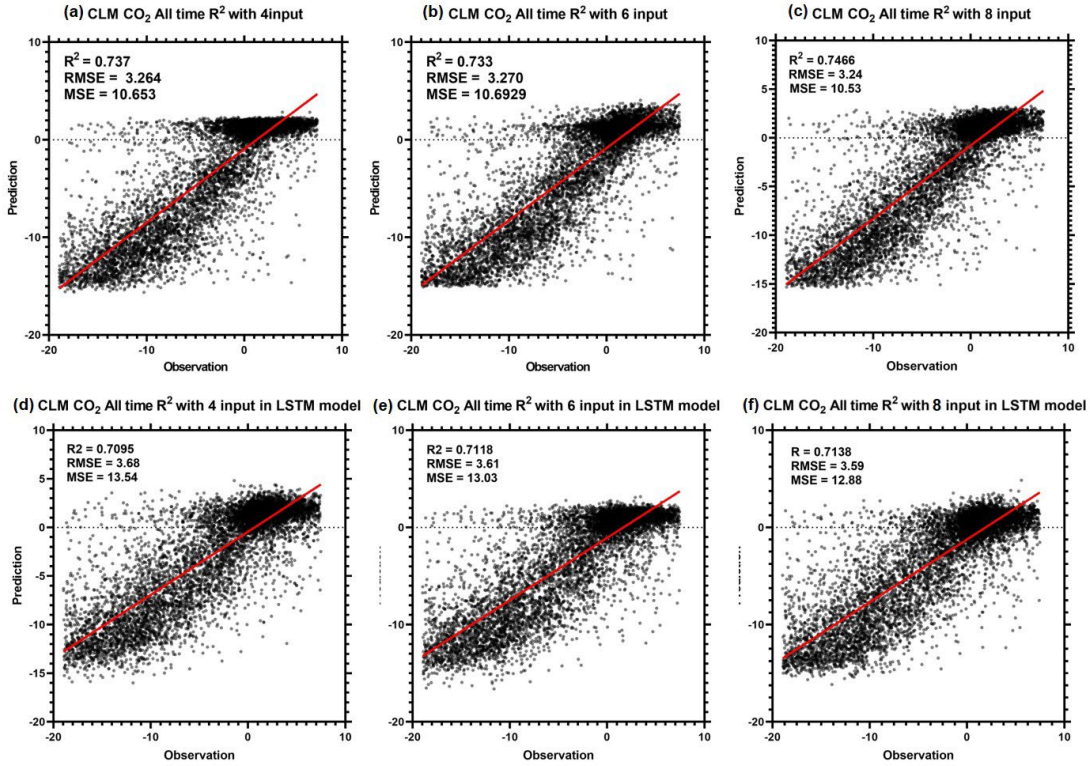


Figure 25. The R^2 of the different inputs in CO_2

(a) This figure is the ANN forecast of the CO_2 with 8 inputs. (b) This figure is the ANN forecast of the CO_2 with 6 inputs. (c) This figure is the ANN forecast of the CO_2 with 4 inputs. (d) This figure is the LSTM forecast for the CO_2 with 8 inputs. (e) This figure is the LSTM forecast of the CO_2 with 6 inputs. (f) This figure is the LSTM forecast of the CO_2 with 4 inputs.

Among the models of LE **(Figure 26)**, it shows that the ANN model improves from 0.6783 to 0.7122 for 4 parameters, which is about 0.0339, while the LSTM improved from 0.6610 to 0.6746 with an improvement of 0.0136. It can be found that the improvement in the improvement of ANN is larger than LSTM, on the one hand, it may

doi:10.6342/NTU202300497

be because of the data with a large range gap. However, it is found that because the range of LE is larger, the RMSE is also larger than that of CO_2, ranging from 41 to 49. Although ANN performs better in R^2, it shows that LSTM has a better ability to predict the off-average value in either CO_2 or LE. The better ability of the LSTM to predict values far from the mean may be due to the time series, as it is probably because the data from 10 days ago are used to predict the data of the next day. Therefore it has a better ability to combine with the changes of the previous period. In summary, both ANN and LSTM increase the prediction ability with the increase of parameters. Furthermore, ANN performs better in predicting, but LSTM is easier to predict for the more extreme values.

doi:10.6342/NTU202300497

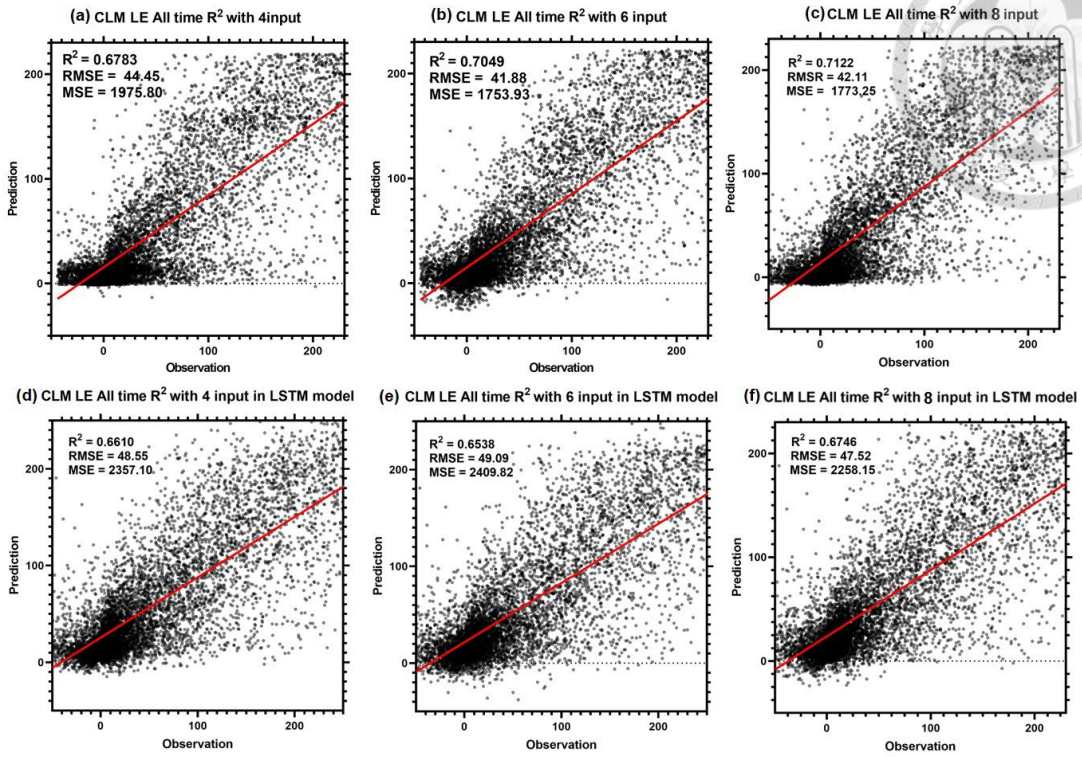


Figure 26, The R^2 of the different inputs in LE

(a) This figure is the ANN forecast for the LE with 8 inputs. (b) This figure is the ANN forecast of the CO_2 with 6 inputs. (c) This figure is the ANN forecast for the LE with 4 inputs. (d) This figure is the LSTM forecast for the LE with 8 inputs. (e) This figure is the LSTM forecast for the LE with 6 inputs. (f) This figure is the LSTM forecast for the LE with 4 inputs.

4.2.2 The R^2 between the different periods

After the comparison in the previous sections, it shows that 8 parameters can get the best results, so in this section, the Eight parameters will be used to predict day and night. However, the day and night boundary in the model for net radiation is 0 W m^{-2}. If it is greater than 0, it means daytime, if it is less than 0, it means nighttime. It shows (**Figure 27**) that there is a big difference between daytime and nighttime after separation, ANN

doi:10.6342/NTU202300497

calculates the daytime results (**Figure 27 b**), the value ranges from -16 to 3 (μ mol/m^2) during the day, while R^2 is about 0.67 and RMSE is about 3.43, which is an acceptable value, but in At night (**Figure 27 c**) cannot be effectively predicted, and there is an obvious boundary. The main reason is that there is no correlation between CO_2 and atmospheric values at night, but is affected by more environmental factors and vegetation so it cannot be effectively predicted at night, its R^2 is only 0.1276, but the RMSE is lower, mainly because the value at night is more concentrated around 2 (μ mol/m^2), so the value obtained is closer to the average value, so Lower RMSE can be obtained.

However, it is slightly different in LSTM (**Figure 27 e**). LSTM cannot be effectively predicted during the day. Its R^2 dropped from 0.7138 to 0.5059 during the whole day, a drop of about 0.2, which may be due to the Rn during the day. Most of the values are not continuous with those of the previous period. Furthermore, it indicates that some values in the observed values are greater than 0 (μ mol/m^2), but most of the predicted values do not exceed 0 (μ mol/m^2), so it may also be related to the activation function. After testing, the parameter adjustment of LSTM during the day has not been effective, so only the result with the highest test value can be used for analysis. However, at night (**Figure 27 f**). LSTM is also unable to effectively predict, and it is estimated that CO_2 at night has of few time-series relationship.

doi:10.6342/NTU202300497

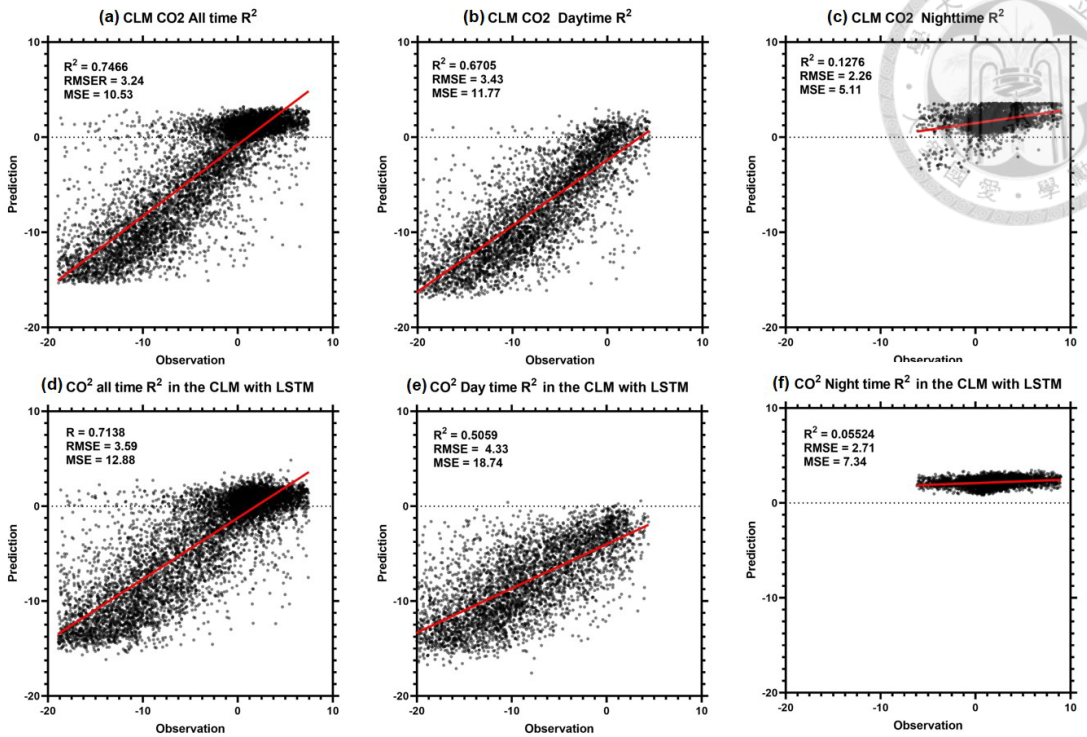


Figure 27, The R^2 of the different inputs in CO_2

(a) This figure is the ANN forecast for the CO_2 in full-time. (b) This figure is the ANN forecast for the CO_2 in the daytime. (c) This figure is the ANN forecast of the CO_2 at nighttime. (d) This figure is the LSTM forecast for the CO_2 at all times. (e) This figure is the LSTM forecast for the CO_2 in the daytime. (f) This figure is the LSTM forecast for the CO_2 at nighttime.

LE (**Figure 28**) behaves similarly to CO2 on both day and night, but most of the values fall in the range of -50 to 250 (W/m^2), while the nighttime values are concentrated around 0 (W/m^2). The daytime values are positive, mainly due to the influence of solar radiation and temperature during the daytime, which leads to more intense evaporation. It is found that most of the values are still concentrated between 0~50(W/m^2), but near noon, the potential heat flux will gradually increase with time because of the stronger radiation, while the daytime R^2 is 0.6429, and the nighttime (**Figure 28 c**) cannot be predicted effectively like CO_2, and there is a boundary, which is presumably related to

doi:10.6342/NTU202300497

the model used. Although linearity is used for the final output layer, the boundary often appears.

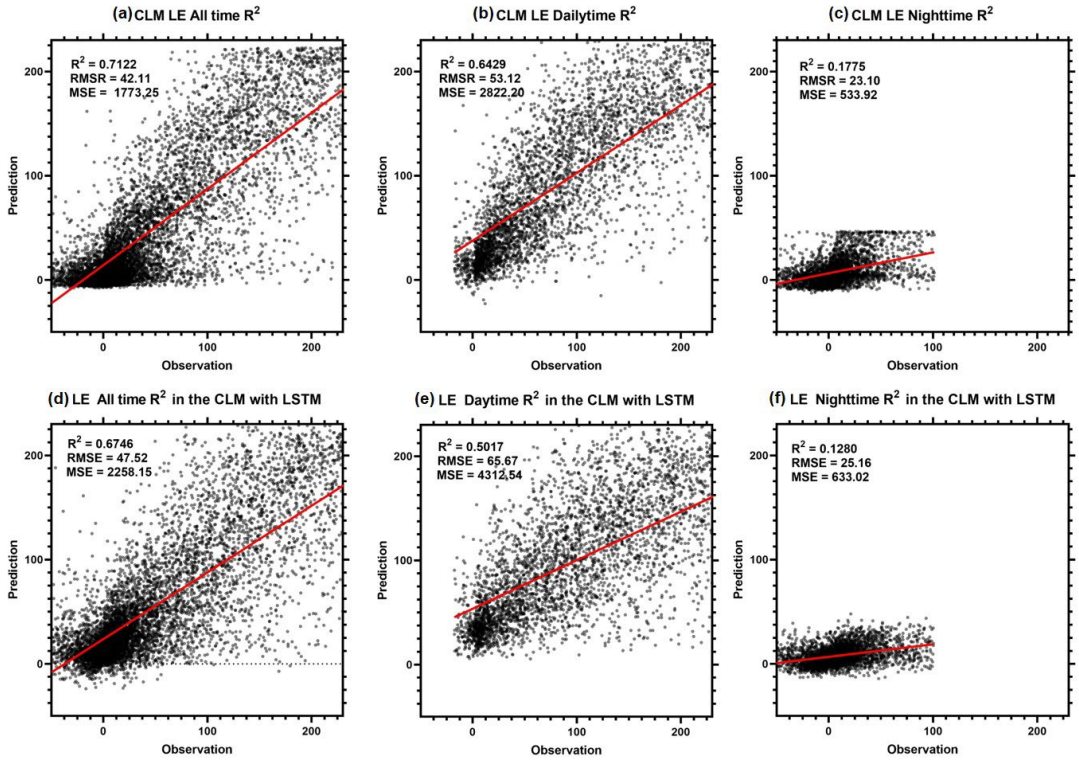


Figure 28. The R^2 of the different inputs in LE.

(a) This figure is the ANN forecast for the LE full-time. (b) This figure is the ANN forecast of the LE in the daytime. (c) This figure is the ANN forecast for the LE at nighttime. (d) This figure is the LSTM forecast for the LE at all times. (e) This figure is the LSTM forecast of the LE in the daytime. (f) This figure is the LSTM forecast for the LE at nighttime.

However, in the LSTM (**Figure 28 e**), the study can find that the prediction ability also decreases in the daytime part, from 0.6746 to 0.5017 with a decrease of 0.1729. It is presumably related to the discontinuity of Rn and CO_2 data. However, in the nighttime

doi:10.6342/NTU202300497

(**Figure 28 f**), there is no clear boundary as in ANN, and the prediction ability is better than that of CO_2.

In the CO_2 (**Figure 29**) section, all the performance evaluations of ANN are better than those of LSTM, and it shows that the R^2 (blue-green) has higher values of about 0.75 for 8 parameters in the full-time period, and the same for 4 (purple) or 6 (yellow) inputs. However, the ANN decreases in the daytime, with the LSTM decreasing more, and in the nighttime, neither model can predict effectively, so although the RMSE are both lower, it is impossible to build an effective model.

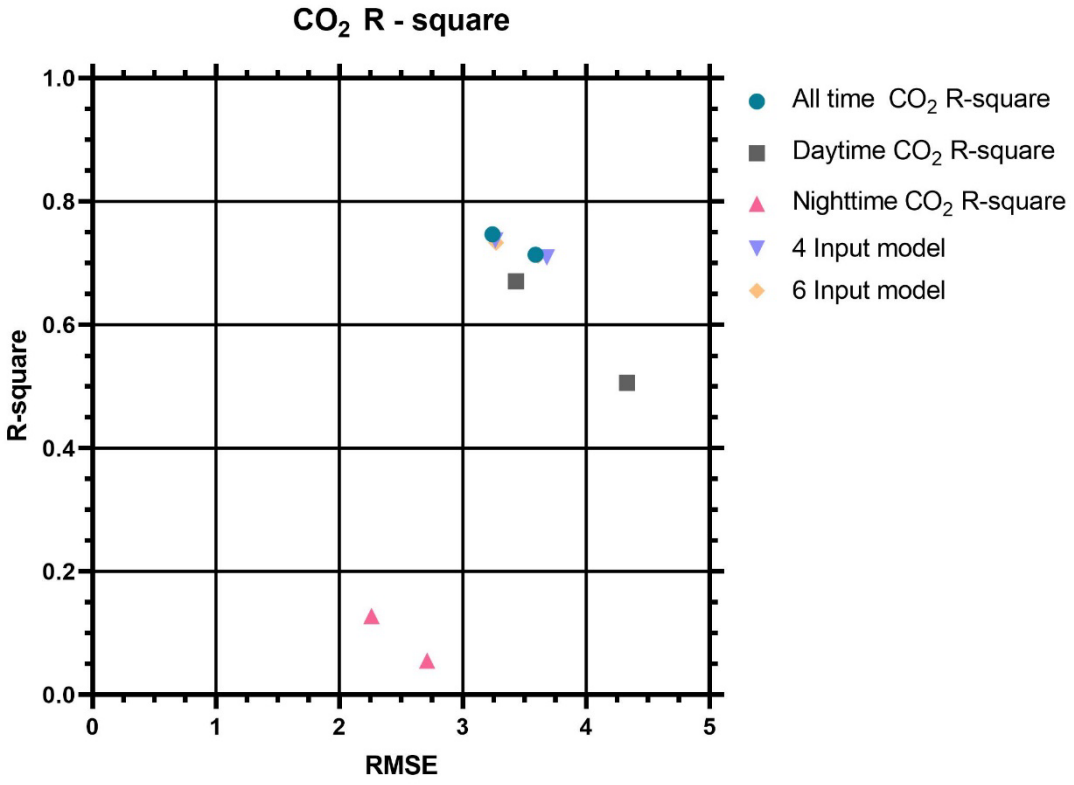


Figure 29. Comparison diagram of R^2 for each method

doi:10.6342/NTU202300497

In the LE section (**Figure 30**), it shows that the prediction ability of ANN is higher than that of LSTM, and the highest value of R^2 is 0.7122 for the data of 8 parameters in the full-time period. In the daytime data, the predictive ability of the data decreases, but in the nighttime data, it is not possible to build a valid model. Therefore, the model can use the nighttime data for more research in the future study.

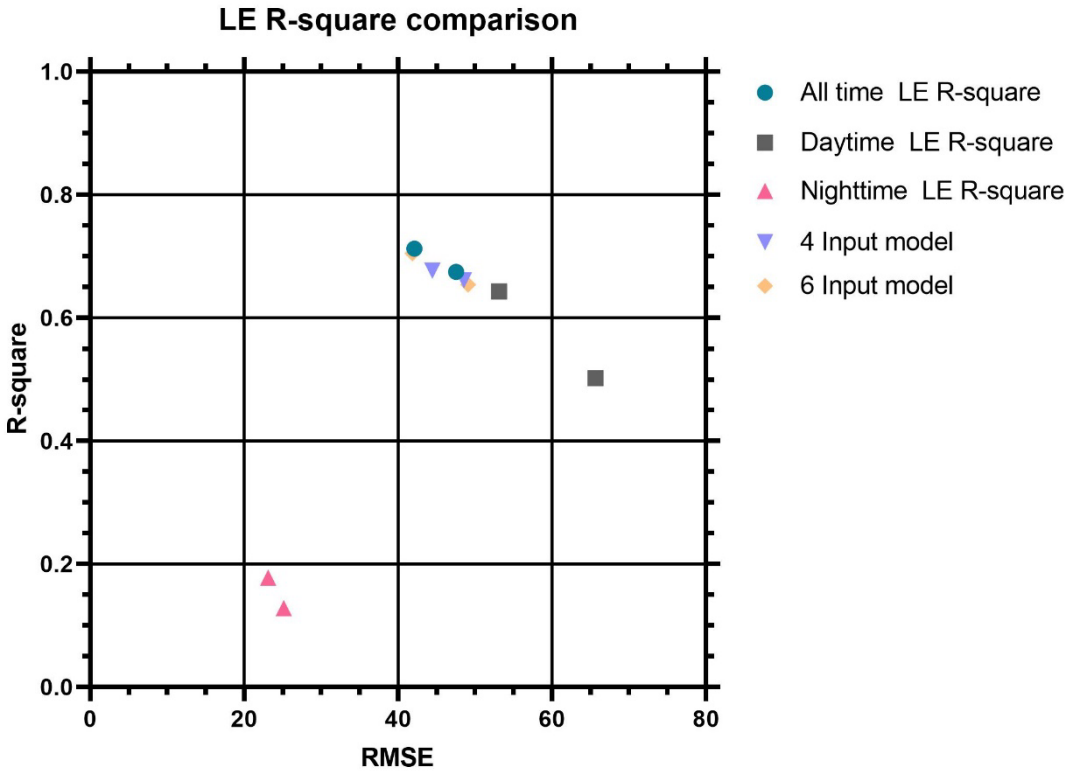


Figure 30. Comparison diagram of R^2 for each method

4.2.3 ANN model plus LSTM model

After the above comparison, it shows that LSTM may be affected in time series due to the quality of the data and the discontinuity of the preceding and following time, and

doi:10.6342/NTU202300497

the results indicate that the missing value is up to 40% in CLM data, which may have a great impact on the data.

Therefore, this section first uses the full-time 8-parameter ANN model to predict CO_2 and LE and then puts the predicted CO_2 and LE into the LSTM model to get better results. The result (**Figure 31**) shows that the value of prediction after ANN predicted can get a very high R^2, the value is as high as 0.8660, and LE (**Figure 32**) can also get a value of 0.8939, which can show that the data is relatively complete and continuous in some cases, LSTM can still get good results. On the one hand, it is because the increase of the overall value can improve R^2; To make the data more concentrated, the CO_2 is between -15~-10 (μ mol/m^2) and -4~4 (μ mol/m^2), the LE value between 0~100 (W/m^2) and 170~230 (W/m^2) has an increasing trend, and the overall number of entries has increased from 39,000 to 63,000. The increased data is helpful for the prediction of seasonal changes, but care must be taken when using it because the values predicted by ANN are not completely observed, so there are still some limitations in research.

doi:10.6342/NTU202300497

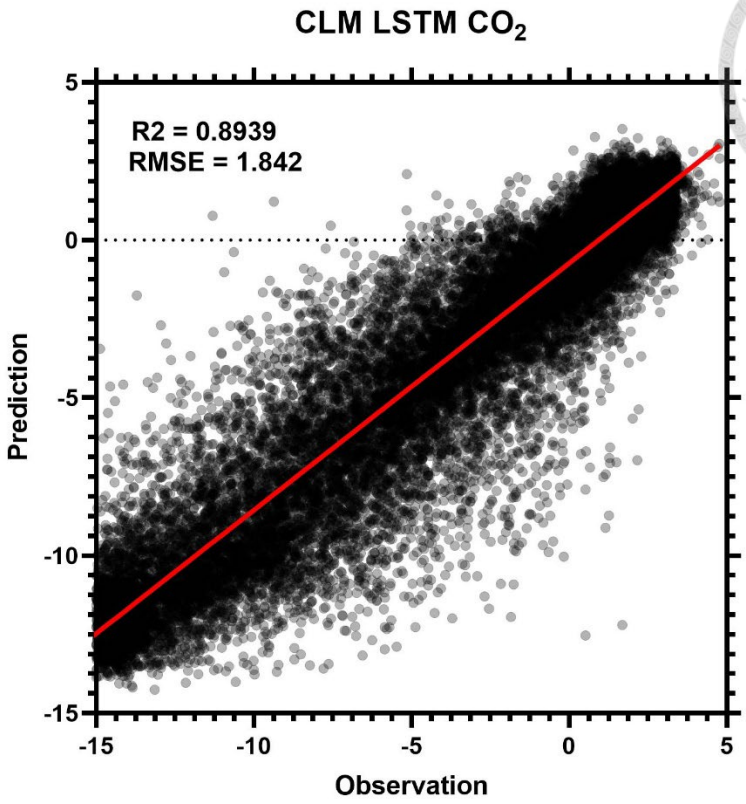


Figure 31. Results of ANN prediction of CO_2

This result is that ANN predicts all the data and then uses LSTM to make predictions

doi:10.6342/NTU202300497

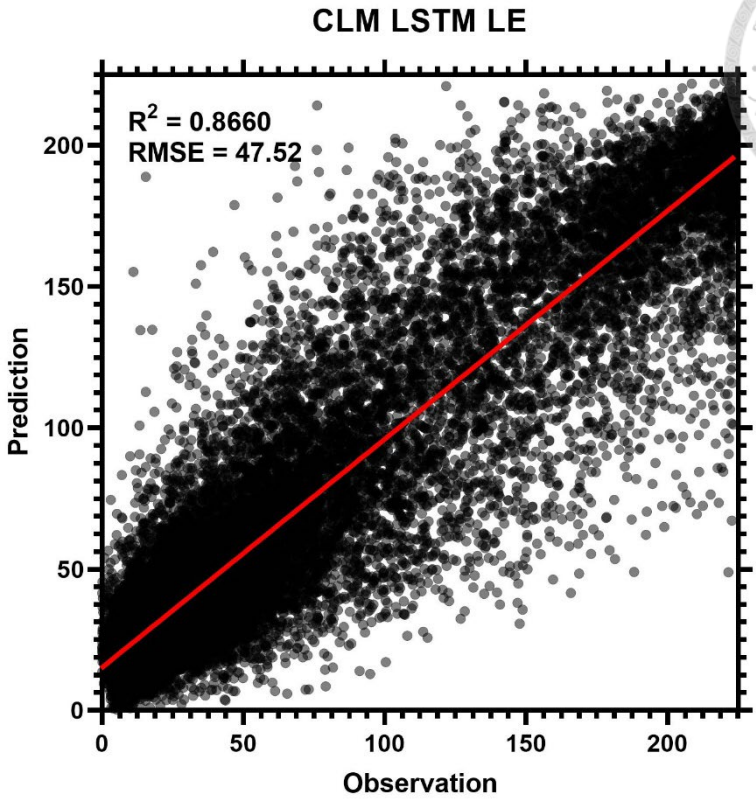


Figure 32. Results of ANN prediction of LE
This result is that ANN predicts all the data and then uses LSTM to make predictions

4.3 Results of sensitivity analysis

4.3.1 Formulation of sensitivity analysis

To quantify the sensitivity of the CO_2 and LE values to the input and explore the most dominant parameters under different climate conditions, the weight and Partial derivation (PaD) methods were used in this study. The connection weights are divided to determine the relative importance of various inputs. The method involves dividing the hidden output connection weights of each hidden neuron into components associated with

doi:10.6342/NTU202300497

each neuron. The sensitivity of an artificial neural network model can be expressed as the first-order partial derivative between the output variable and the input parameter and relevant mathematical concepts will be presented below.

The ANN architecture (**Figure 33**) used one layer of the hidden layer and one layer of the output layer. Previous sections pointed out that this architecture can effectively predict fluxes. Because it is smaller in the selection of Units, plus layers the number is small, so the efficiency of the operation is relatively well. The selected activation functions include Tanh and linear. The data processing must first be standardized. The Min-Max Standard method is selected in this model, the method will change the data from 0 to 1, and it is used in the input and output. Because the relationship between positive and negative values can be reflected later, after experiments, it indicates that the architecture has good results, and the parameter selection includes the following, the learning rate is 0.001, the batch size is 1000, and the epochs is 1000, good results can be obtained.

doi:10.6342/NTU202300497

Table 8. Neural network architecture

Layer	Output shape	Parameters	Activation function
Hidden layer 1	Units = 4	25	Tanh
Output layer 2	Units 1	1	linear

Other information

Total parameters: 25

Learning rate: 0.001

Batch size = 1000

Epochs: 1000

The following is the figure for neural network workflow:

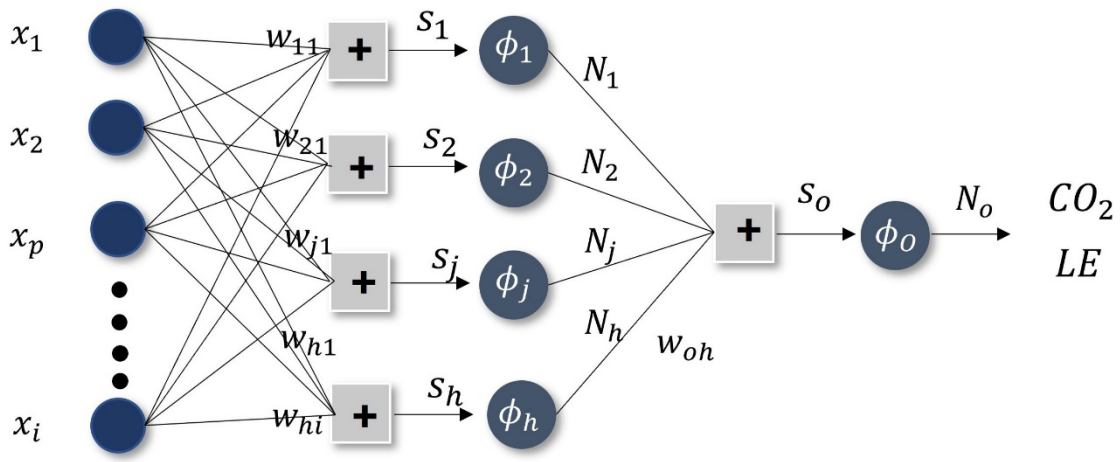


Figure 33. ANN structure schematic diagram (Redrawn from (Nourani & Fard, 2012))

In this study, a four-layer network is used; the sensitivity of the artificial can be expressed as the output variables and the first-order derivation of the input parameters, which are input without processing as follows:

$$N_i = x_i \quad (13)$$

doi:10.6342/NTU202300497

The formula for the hidden layer is as follows:

$$S_h = N_p W_{hp} + \sum_{i \neq p} N_i W_{hi} \quad (14)$$

$$N_h = \phi(S_h) \quad (15)$$

Also, the study has:

$$\frac{\partial S_h}{\partial N_p} = W_{hp} \quad (16)$$

$$\frac{\partial S_o}{\partial N_h} = W_{oh} \quad (17)$$

The first-order derivative of the output concerning the input parameter x_p is:

$$\frac{\partial N_o}{\partial X_p} = \frac{\partial N_o}{\partial N_p} = \frac{\partial N_o}{\partial N_h}\frac{\partial N_h}{\partial N_p} = \left(\frac{dN_o}{dS_o}\frac{\partial S_o}{\partial N_h}\right)\left(\frac{\partial N_o}{\partial S_h}\frac{\partial S_h}{\partial N_p}\right) \quad (18)$$

Which can be transformed into the following equation:

doi:10.6342/NTU202300497

$$\frac{dN_o}{dX_o} = \phi'(S_o) \quad (19)$$

$$\frac{dN_h}{dS_h} = \phi'(S_h) \quad (20)$$

Then (30) can be brought into the following equation:

$$\frac{\partial N_o}{\partial X_p} = \phi_o{}'(S_o) W_{oh} \phi_h{}'(S_h) W_{hp} \quad (21)$$

Finally, the values of each point are added together to obtain the contribution of the input value to the output value:

$$\frac{\partial N_o}{\partial X_p} = \sum_{h=1}^{nh} \phi_o{}'(S_o) W_{oh} \phi_h{}'(S_o)\, W_{hp} \quad (22)$$

In this study, the formulation used the two activations function. The first one is linear, and the second one is the tanh function. The two functions can express the linear ($\phi_o\ (S_o) = S_o$), and tanh function ($\phi_h\ (S_h) = \frac{(Exp(2S)-1)}{Exp(2S)+1}$, and the $\phi_h{}'(S_h) = 1 - N^2$), According to (29), we used the following formula:

doi:10.6342/NTU202300497

$$\frac{\partial N_o}{\partial X_p} = \sum_{h=1}^{nh} W_{oh}\,(1 - N^2) W_{hp} \qquad (23)$$

The derivative is the input sensitivity, which can quantify the expected variation of N_o, where is the variation of X_o on its input, while the other factors are relatively constant. However, because each parameter has its range, the output cannot be used to determine its influence. Because this study uses normalization, it can be used to express the relative sensitivity. The relative contribution of the ANN output to the data set concerning the input can be calculated by summing the squared derivatives of the input variables:

$$SSD_p = \sum_{d=1}^{n} \left(\frac{\partial N_o}{\partial X_p} \right)_d^2 \qquad (24)$$

The input variable with the highest SSD value has the maximum impact on the output variable, and its sensitivity is affected by the following factors, (A) The number of layers and the number of hidden nodes in the ANN, (B) The impact of the ANN data set (C) The weight value of the ANN, and (D) The output value of the nodes.

doi:10.6342/NTU202300497

After the model is trained, factors one to three are fixed, and the influence of the input parameters on the output is entirely influenced by the (D) factor, which can be plotted to understand the weight value of each factor.

4.3.2 Results of the sensitivity analysis

This section is based on the previous framework for sensitivity analysis because in the previous chapter, it was found that the R^2 differences in the 4, 6, and 8 parameters are not significant, and most of the other stations only have the basic meteorological parameters, so this analysis refers to the analysis of the ANN model with 4 parameters. The results of this sensitivity analysis (**Figure 34**) are mainly calculated by partial differentiation. It is found that the value of the X-axis falls between 0 and 1 after the Min Max Standard, while the Y-axis is the result of differentiation, which is the small change of each point for CO_2. The other parameters do not have a fixed direction of change, but the study can still find that they are concentrated in the part of Ta, where most of the values are concentrated in the partial differentiation of 0.08, which is not large for other values. The values of Tsoil are concentrated at 4 and Rh are around -13, which has a large degree of variation for CO2 values.

doi:10.6342/NTU202300497

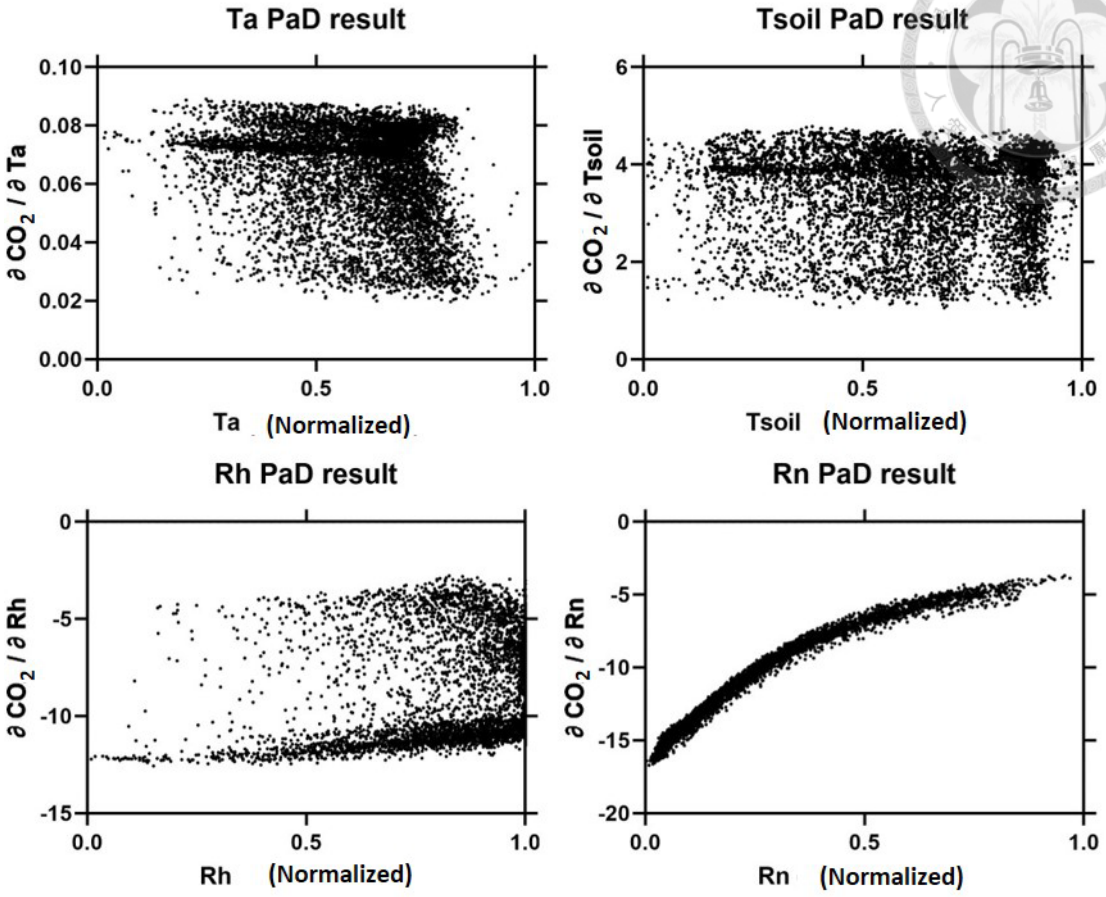


Figure 34. CO_2 Partial derivation results

However, for LE (**Figure 35**) section, the results show that the value of Rn also shows a long bar, and its value is also negative, which means that it has a negative correlation for LE. Furthermore, it decreases with the increase of its value, but the empirical formula shows a positive correlation between them. It is because the model is highly variable and will be affected by the function and learning rate, Epochs, and Batch

doi:10.6342/NTU202300497

size, but the study can know that the change in the value is very large for the partial differential,

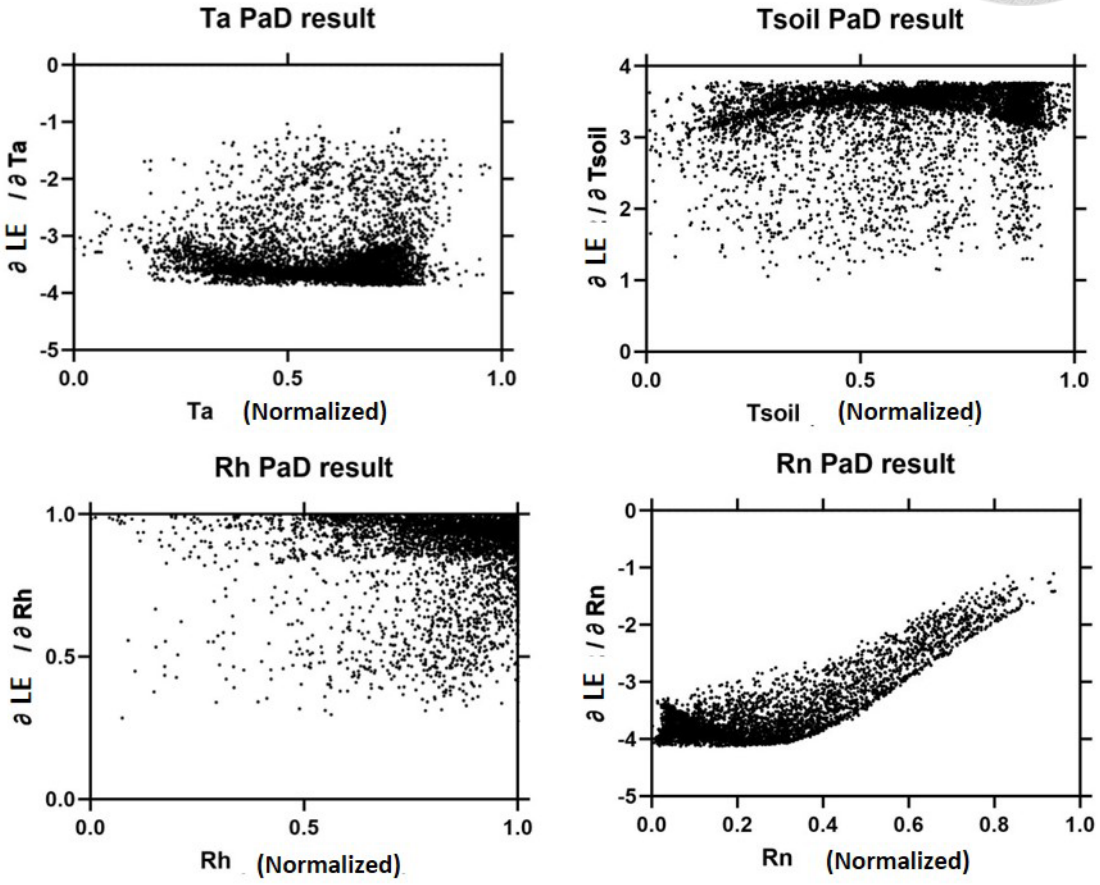


Figure 35. LE Partial derivation results

partial differential, while the other values are more concentrated and the rate of change between values is lower. The results show that Ta is concentrated between -3~-4, Tsoil is concentrated between 3~4, and Rh is between 0.8~1, representing the change in the value of the partial differential. From this analysis, the study can conclude that the change in Rn has a very important contribution to CO_2 and LE.

doi:10.6342/NTU202300497

Finally, it is important to calculate the contribution of all parameters to the output. The following results can be obtained through the previous SSD formula. For CO_2 (**Figure 36**), the results indicate that the contribution of Rn and Rh is the highest, which can reach 60% of the whole and 38%. The reason for the high contribution of the Rh value is that Rh provides a stable value for the model to make predictions during the nighttime period, so the contribution is relatively high. In contrast, Ta and Tsoil are almost the same in this model. There is no contribution, which means that its value is closer to 0.

On the other hand, for LE (**Figure 36**), it is found that although Rn is the value with the largest value change. Its contribution is not particularly high. Its value is about 35%, and the contribution values of Ta and Tsoil are also relatively high, which means that their values have a greater impact on LE for the model. Contrary to CO_2, Rh contributes less to LE, which may be due to the long-term dataset. The value at 100 has no trend to predict in LE with a large positive range. All in all, Rn is the most important value for the model.

doi:10.6342/NTU202300497

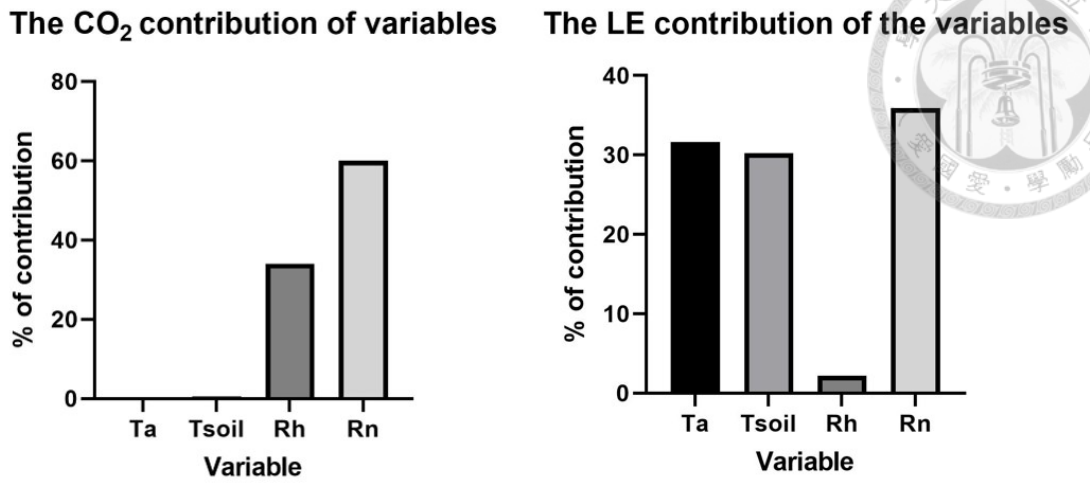


Figure 36. SSD results

4.4 Analysis of the gap-filling data

4.4.1 Monthly data before and after filling

This section mainly explores the difference between the value after the model filling and the original observed value, before filling in the data (**Table 9**). Before filling in the data, the model must first fill in the meteorological data (**Table 10**) in the prediction of data. In this study, the data with missing values of more than 14 half hours of meteorological data were filled by using the average of the time in other years, while the data with less than 24 half an hour of meteorological data were filled by selecting the previous time to facilitate the continuity of the time series. After filling in the data, the results present that the highest value is -2.07 ($\mu mol/m^2$) in 2008 and the lowest is -3.17($\mu mol/m^2$) in 2011. It shows that most of the values do not show many trends after

doi:10.6342/NTU202300497

filling in the data and then averaging, but there may be a big difference in the seasonality. Therefore, the study will make a graph for discussion later.

Table 9. Before filling in the data

Fluxes	CO_2 (μmol /m^2/month)				LE (W/m^2/month)			
Year/Month	2008	2009	2010	2011	2008	2009	2010	2011
Jan	-1.99	-1.55	-1.85	-2.19	34.38	27.40	38.60	33.94.
Feb	-1.65	-2.45	-2.31	-2.75	26.30	48.75	43.59	44.30
Mar	-2.65	-3.21	-3.10	-2.51	43.01	52.21	49.03	32.76
Apr	-2.67	-2.58	-3.11	-3.17	49.26	39.56	48.09	51.89
May	-2.27	-2.82	-2.71	-2.42	50.05	63.78	54.24	56.87
Jun	-2.34	-2.29	-1.83	-2.95	57.19	63.85	57.60	69.01
Jul	-1.84	-2.75	-2.66	-3.02	55.92	69.32	69.98	69.29
Aug	-1.86	-2.89	-2.64	-3.03	62.44	68.61	68.20	79.39
Sep	-1.52	-2.61	-2.04	-2.88	49.16	58.36	60.65	63.70
Oct	-2.30	-1.19	-1.77	-1.25	41.22	38.36	39.01	33.42
Nov	-1.39	-1.67	-2.20	-1.83	35.88	38.39	32.37	38.77
Dec	-2.05	-1.67	-2.17	-2.51	38.56	35.76	41.69	35.95
Annual	-2.07	-2.31	-2.37	-3.17	52.52	50.38	50.32	50.69

In the part of LE, it can be found that its value varies greatly from year to year, from about 60(W/m^2) before filling to about 50(W/m^2) in 2009, 2010, and 2011. It can be inferred from the table that the value is higher in June, July, and August before filling, but after filling, its value decreases a lot.

doi:10.6342/NTU202300497

Table 10. After filling in the data

Fluxes	CO_2 ($\mu mol\ /m^2/month$)				LE ($W/m^2/month$)			
Year/Month	2008	2009	2010	2011	2008	2009	2010	2011
Jan	-1.86	-1.88	-1.31	-3.83	32.16	29.00	39.01	58.64
Feb	-1.51	-3.03	-2.10	-2.79	29.09	52.01	36.30	46.97
Mar	-2.82	-3.77	-3.26	-2.88	47.37	60.16	48.09	32.19
Apr	-2.82	-2.85	-3.09	-3.97	52.48	46.99	56.66	56.48
May	-2.45	-3.13	-2.61	-2.59	54.25	66.23	63.45	66.40
Jun	-2.20	-2.33	-1.83	-3.62	63.47	67.70	62.24	84.93
Jul	-1.83	-3.10	-2.98	-3.35	74.22	79.15	84.95	81.30
Aug	-2.05	-3.74	-2.80	-3.48	65.02	88.43	83.96	100.48
Sep	-1.35	-3.54	-3.18	-2.92	56.02	78.13	98.48	70.07
Oct	-1.62	-2.18	-4.02	-2.89	55.96	60.58	66.28	38.44
Nov	-1.51	-2.45	-3.97	-3.01	40.17	59.64	48.02	48.98
Dec	-1.97	-1.68	-2.86	-2.44	36.31	55.47	44.16	21.49
Annual	-2.00	-2.86	-2.81	-3.17	52.52	62.34	60.78	60.19

However, during the one year year year (**Figure 37**), it usually declines first in spring, rises and then falls in summer, rises in autumn, and falls in winter with the seasons, and its monthly median value is concentrated between 0 and -5. It can be seen that the CO_2 of the forest ecosystem is considered to be in a stable state during this period. The blue value is the value after filling, while the pink value is the value after filling. It is found that if the data is not filled in, there will be many extreme values in certain periods. It is because there are more missing data this month, so it is easy to cause the quintile values to be more abnormal. In 2008 (**Figure 37**), the results indicate that the value has more extreme values in July and October, which often affects the average. However, it is found that the

doi:10.6342/NTU202300497

value is slightly lower than the median, which may be because of the night time during this period. In 2009 (**Figure 37**), it is found that the quintile values difference in the winter period is very large. The results shown in **Figure 37**, shows that this season has a large number of missing values, and they are concentrated in the evening, which is the reason for the large extreme value in this period. However, after the value is filled, the value is approximately between 0~-5 ($\mu mol/m^2$). In 2010 (**Figure 37**), it can be found that extreme values are also prone to appear in winter, and its missing values (**Figure 41**) are also mostly concentrated in winter, while they are relatively stable in other seasons. In 2011 (**Figure 39**) the annual change is small, but there is a problem with the data in spring (**Figure 42**), and the data shows negative values at night, so the discussion of the previous seasons will not be discussed first. In winter, there are the same problems as before, but except for spring, other trends are roughly the same as in other years.

doi:10.6342/NTU202300497

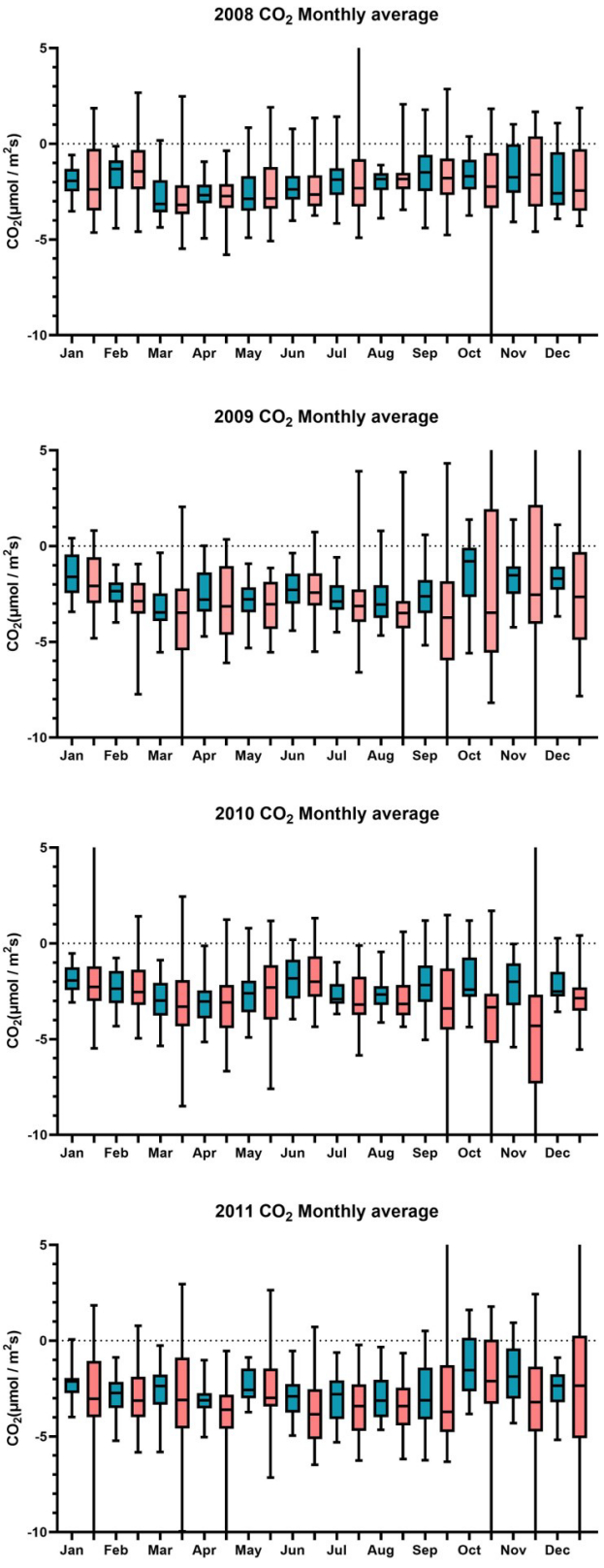


Figure 37. CO_2 annual average

doi:10.6342/NTU202300497

In the part of LE (**Figure 38**), it is found that the trend is generally higher at the end of summer, and in 2008 (**Figure 38**), it shows that the quartiles after filling are smaller than the value before filling, but the difference in the median above July, September, October, and November is larger by (**Figure 40**). It indicates that the color is lighter in the fall and winter seasons, which can reflect the narrower quartiles in that period, while in the early spring, it can be found that there is a yellow color between 33 and 55, which shows a lower value. In 2009 (**Figure 38**), the results show that the latter difference is larger, (**Figure 40**). We can see that there are more missing values in this period, which affects the distribution of quartiles, and after filling in the data, the study can show more values with a day-night variation. We can find that there are more missing values in the winter period (**Figure 41**), mostly after 4:00 p.m., and in the late winter and early spring before 2011 (**Figure 42**), the data is not discussed because of anomalies, but the trend is the same with higher LE in summer and lower in winter. However, because the distribution of the missing values in (**Figure 42**) is mostly sporadic, the phenomenon of a larger gap between the previous winter quartiles does not appear.

On the whole, it shows that the highest median values of LE quartiles are located in July, August, and September, which are the seasons with high convection in summer, about 75(W/m^2) in 2008, and about 100(W/m^2) in 2009, 2010, and 2011. This is mainly because the CLM tends to have missing values in winter.

doi:10.6342/NTU202300497

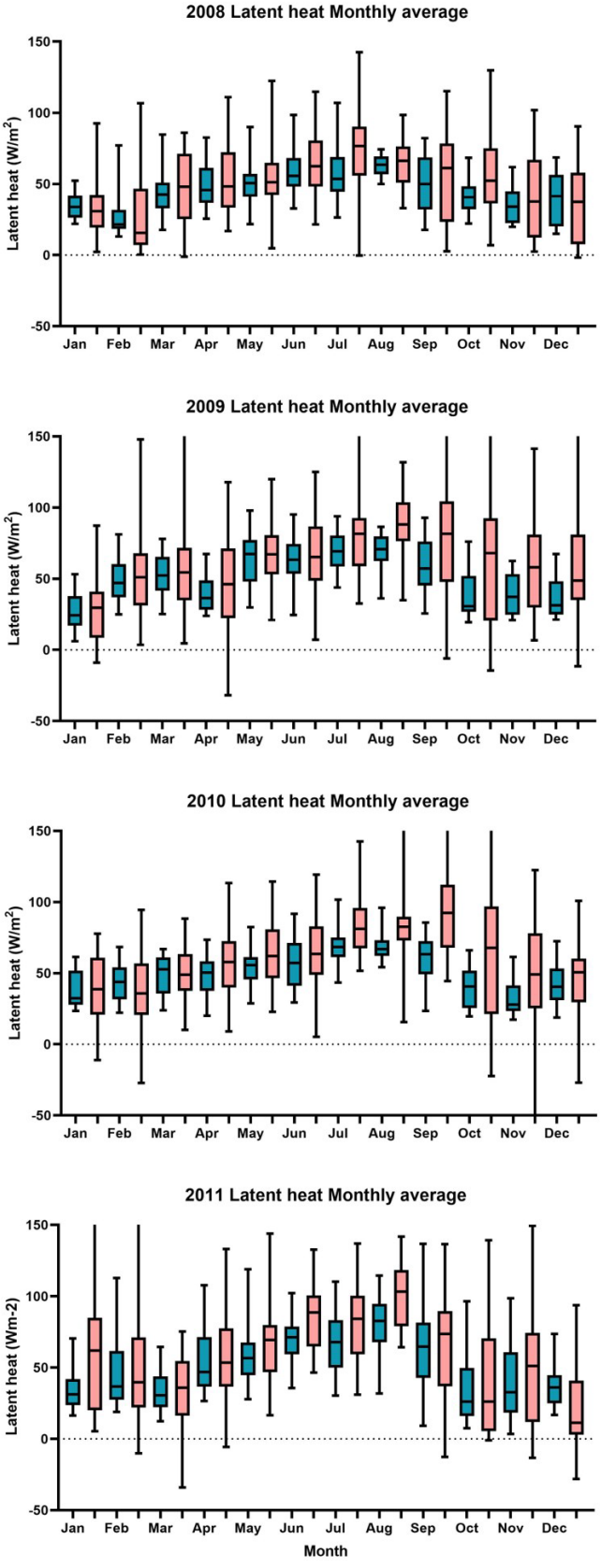


Figure 38. Annual average of LE

doi:10.6342/NTU202300497

4.4.2 Heat map before and after filling

(**Figure 39**) displays the filled data of CO_2 flux, and this result finds that in 2008, the CO_2 values mostly ranged from -20 to 10, with some larger values reaching about 20 at night, and mostly in the summer.

The values of LE (**Figure 39**) are significantly lower in spring and winter. The value of LE is higher from 6:00 am after sunrise to 3:00 pm, mainly due to the solar radiation,

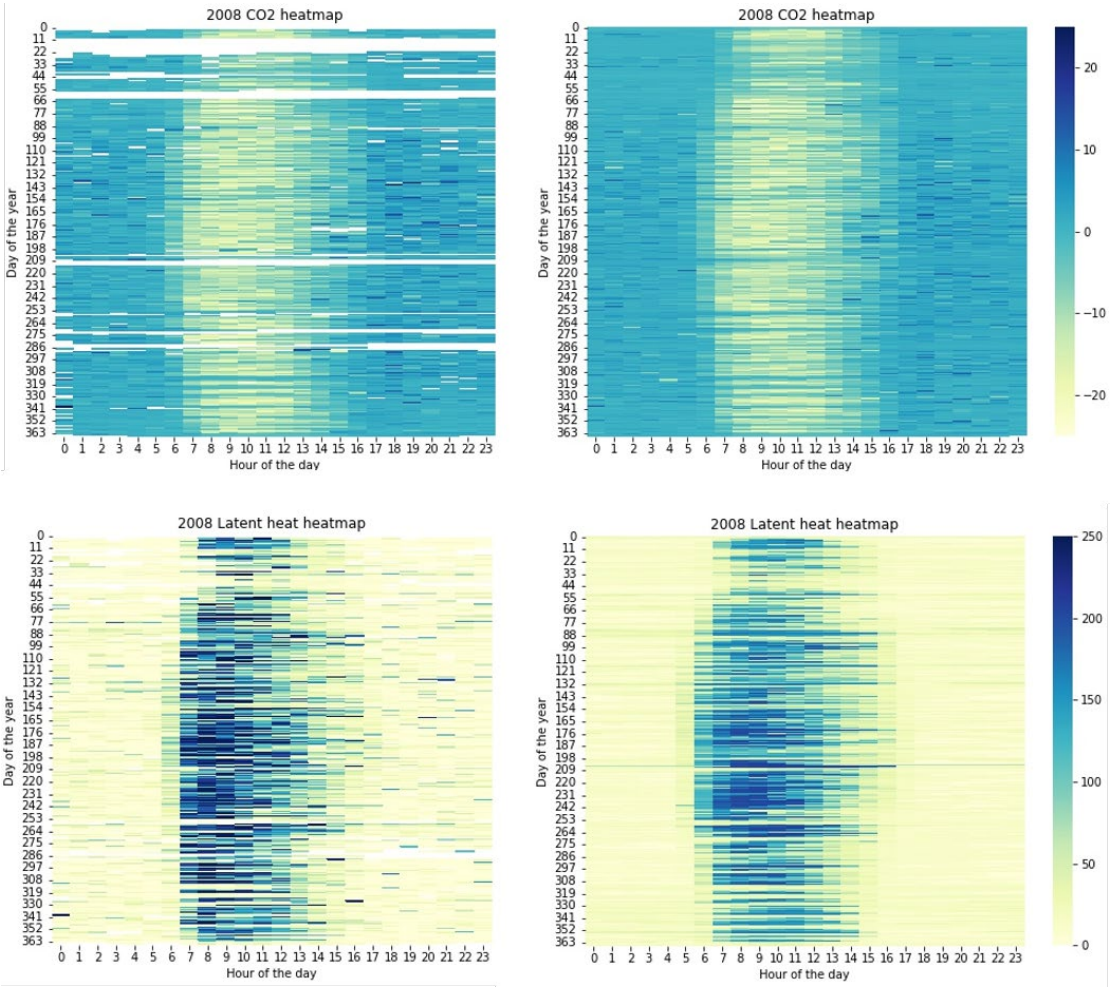


Figure 39. Filling gap results in 2008.

doi:10.6342/NTU202300497

while the value is mostly 0 in winter when the night time is longer, (**Figure 40**) is the 2009 CO_2 heatmap. The value mostly ranged from -20 to 20 ($\mu mol/m^2$), in the middle of the day, the values are mostly negative, but there are some differences in the days of 276~286, the values are lower than the surrounding. Among the data not yet filled, there is a long gap, but **(Figure 40)** 2009 CO_2 Filling gap results in the meteorological getting the lower values, which causes no pattern in the middle of the days.

In 2010 (**Figure 41**), it can be found that most of the missing data are concentrated in winter, and there are similar missing data on the CO_2 and LE figures. It may be mainly

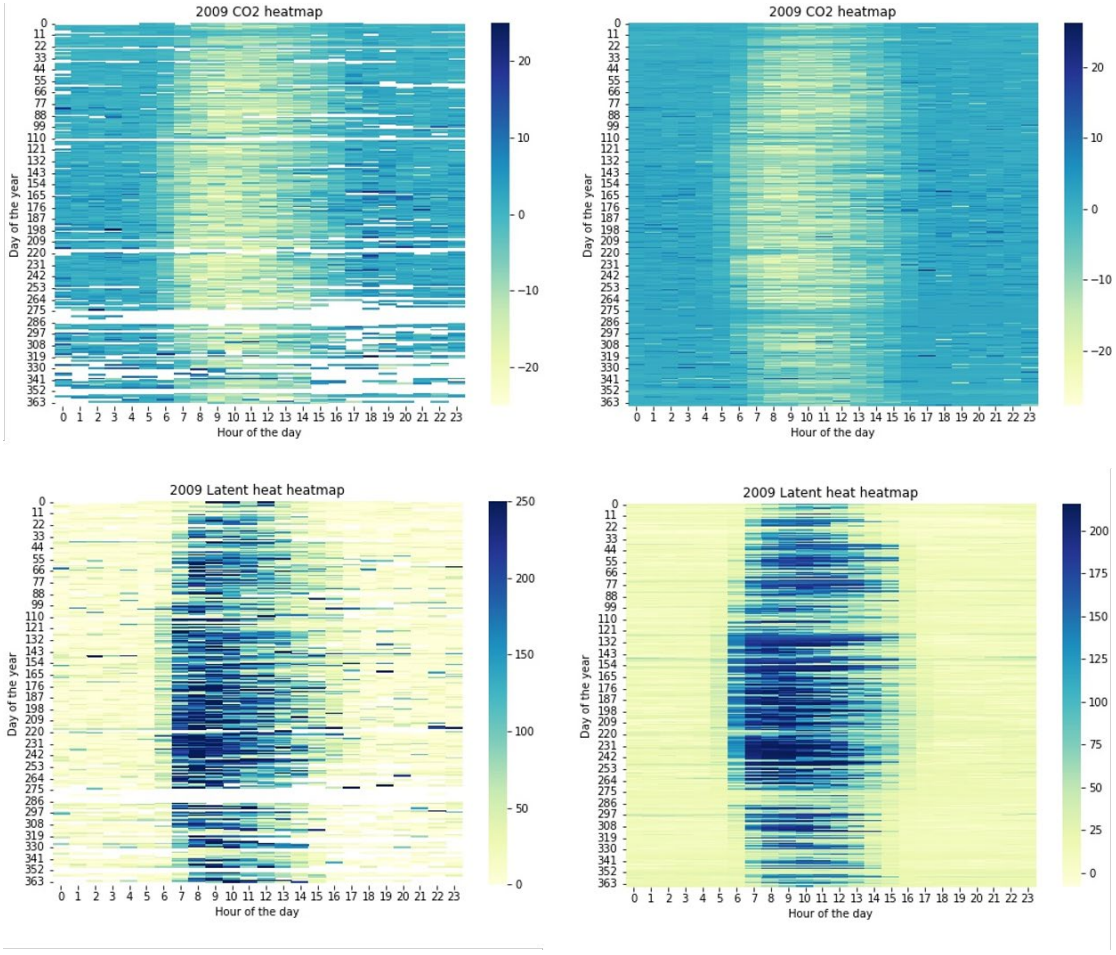


Figure 40. Filling gap results in 2009.

doi:10.6342/NTU202300497

because the wind is weak during this period, which leads to filtering out due to the QA and QC, but after filling, the period between 286 and 308 may be due to the low net radiation flux and lower temperature in the winter, resulting in lower simulate the data, which indeed reflects that CO_2 and LE may be affected by solar radiation in winter.

In days 176 to 187, the simulated value is deepest in this section, mainly because there is strong solar radiation in June and July, and at noon there are high temperatures and net radiation, it can be found that the 5:00 am to the 3:00 pm in this period are deepest yellow, and the LE is mostly dark blue, that was inflected by the solar radiation.

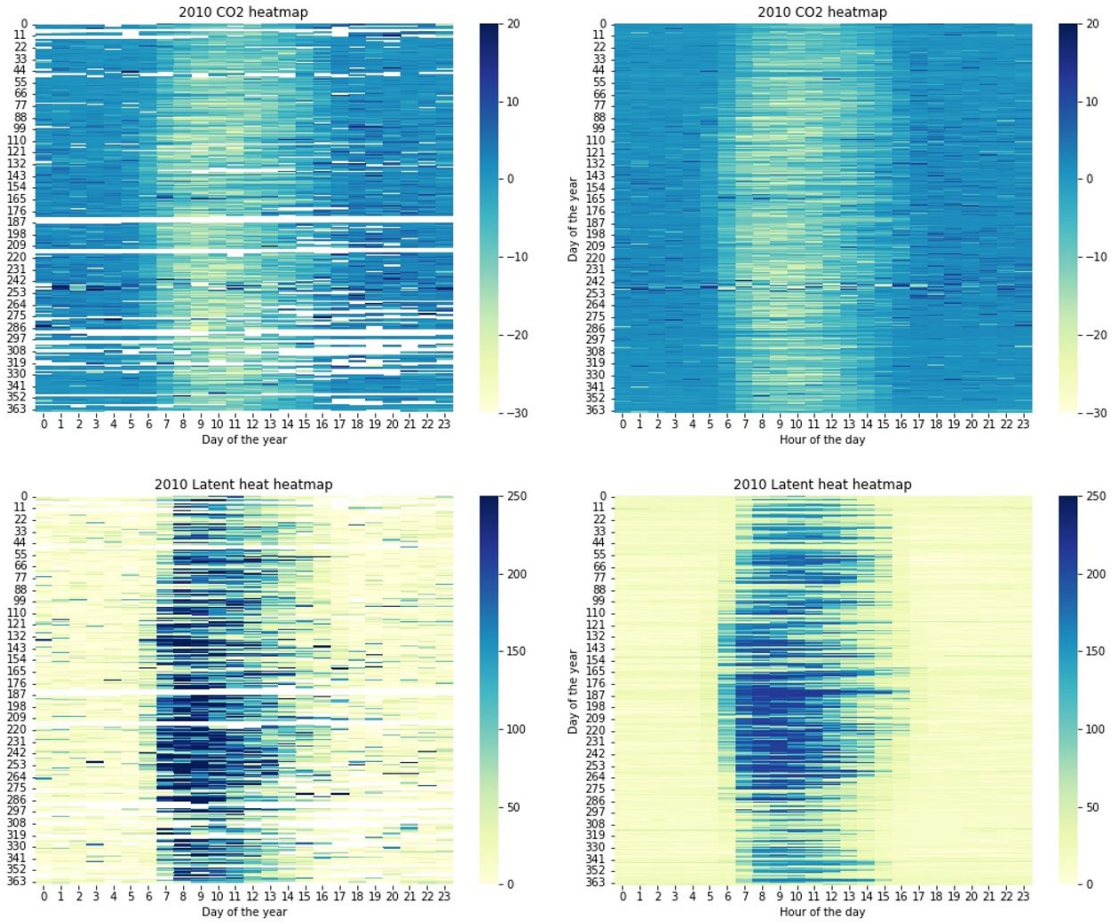


Figure 41. Filling gap results in 2010.

doi:10.6342/NTU202300497

From the data in 2011, it is found that the trend is the same as before, and the lack of value is concentrated in winter. In the part on CO_2 (**Figure 42**), it shows that there are relatively strange lines between days 242 and 253, which may be because the instrument was affected by the climate during that time. After filling in the LE (**Figure 42**), it indicates that the value is more continuous and can truly show a deeper value between days 176 and 187. The result shows that machine learning can also predict the more extreme value of the year.

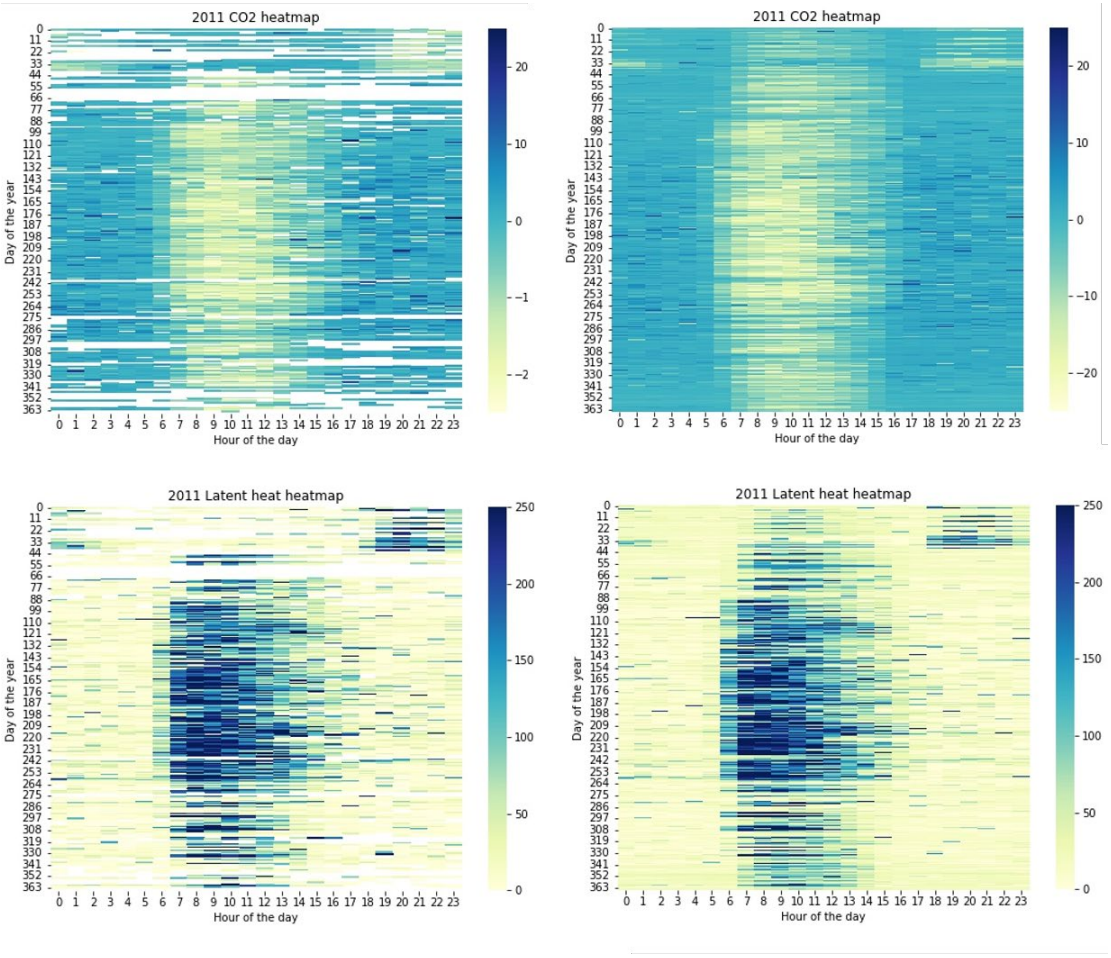


Figure 42. Filling gap results in 2011.

doi:10.6342/NTU202300497

4.4.3 Cumulative chart

This part (**Figure 43**) shows the difference between the cumulative fluxes of the model and the missing value of the annual cumulative amount. However, the CO_2 flux value is converted into kilograms, which is easier to understand the meaning. It can be found that most of the values in the CO_2 part are negative, so the value after accumulation is negative. Before filling, the accumulated value is about 6~7 (tons ha^{-1} $year^{-1}$), but after filling, the value is close to 7~8 (tons ha^{-1} $year^{-1}$), which shows that the difference between years can be as high as 1 (tons ha^{-1} $year^{-1}$). However, it will have a great impact on the overall annual estimate, and it is because most of the missing values are concentrated in winter. In the second half of each year, there is usually a relatively large increase trend. As for cumulative LE, it is converted to annual evapotranspiration (mm). The part of evapotranspiration increases greatly in the winter period, because the value of evapotranspiration itself is relatively large, and its impact on the annual total is greater than that of CO_2, so when the missing value is relatively large in winter, there will be a significant increase from the original 550 (mm $year^{-1}$) to more than 600 (mm $year^{-1}$), and the increase is as high as 50-100 (mm $year^{-1}$). Through the cumulative graph, we know

doi:10.6342/NTU202300497

which season has more missing values in the four years, and after the model is simulated, the cumulative change after imputation can be known.

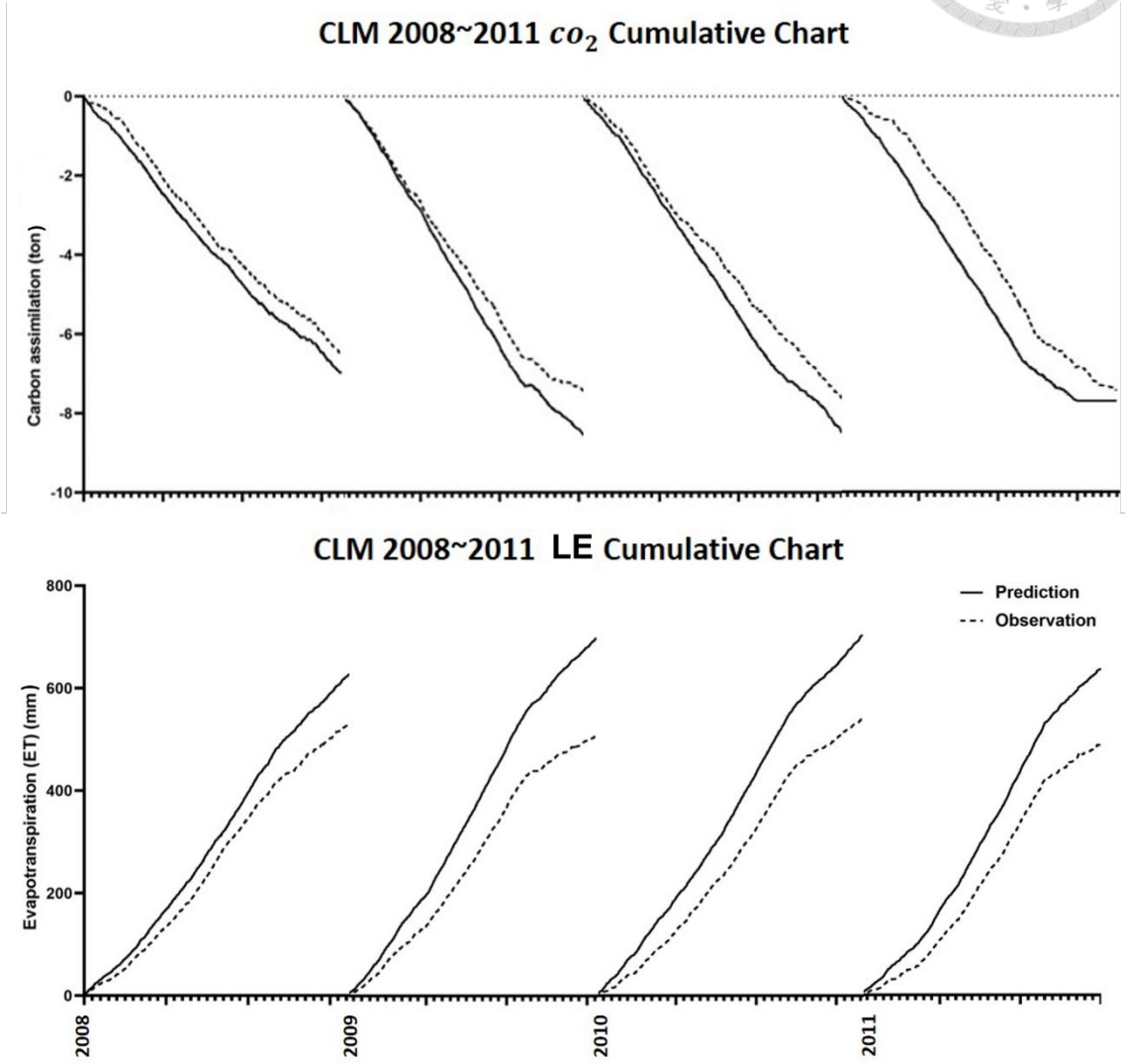


Figure 43. The cumulative chart

doi:10.6342/NTU202300497

Chapter 5 Conclusions

Based on the results of this study, the conclusions and suggestions are shown as:

1. The fluxes of CO_2 and LE in the CLM are most affected by Rn, and it can be seen from the monthly average data that each input parameter has a similar trend in four years, except for the wind direction in 2008.
2. A good prediction model can be obtained through machine learning, which can be used to fill in the missing data of CO_2 and LE.
3. After experiments, it is found that the neural network model of ANN has a better prediction results than LSTM in all scenarios, but if there is continuous time series data, LSTM has a better prediction ability.
4. The results of the day and night forecasts show that the highest R^2 can be found in the whole period, followed by the daytime, and the predictive ability at nighttime is very limited.
5. In the prediction of each parameter combination, it shows that the use of four parameters Ta, Tsoil, Rh, and Rn can achieve good results. If more parameters are added, the prediction ability will be improved with better performance.

doi:10.6342/NTU202300497

6. From the sensitivity analysis, it is known that both CO_2 and LE have the highest contribution to Rn, but the contribution of LE to temperature is similar to that of Rn.
7. Through the thermal power, quartile map, and cumulative map, it is found that the lack of value is concentrated in winter, and it will cause the winter value to be more inaccurate. The quartile value has the problem of overdispersion. The impact is huge.

The research suggestions are as follows: machine learning methods have gradually matured in recent years. However, because machine learning is mostly a black-box job, due to the rise of explainable AI, it is possible to roughly understand some of the internal situations of the model, so it can be used in it. Find some way to understand the relationship between parameters for the model.

Finally, there are still many limitations in the filling of the flux. For example, the selected station and geographical location, vegetation type, latitude, etc. will affect the accuracy of the model. This study has previously experimented with the model adjustment of the tea garden in the Pinglin area. The preliminary results found that no matter how the model adjusts it, the ML can't get good results. After exploring, this study found that because the tea garden is divided into the harvesting period and growing period, there is no certain rule in the measurement of CO_2 and LE. However, in the forest, because

doi:10.6342/NTU202300497

there is no direct influence of human interference, this method is suitable for filling compared to other stations. In the end, it is also found that the results after filling are also very different, mainly because it will have a great impact on the annual carbon exchanges of the ecosystem. It will affect the estimation of carbon uptake and carbon release. Therefore, if the annual data can be fully estimated, it will help to know whether the total carbon is in balance in the region and whether latent heat will affect the growth of the ecosystem. It has a great correlation with temperature, so if the annual data can be filled in, it will also help to understand the seasonal changes of evapotranspiration in this area.

doi:10.6342/NTU202300497

Reference

Alkama, R., & Cescatti, A. (2016). Biophysical climate impacts of recent changes in global forest cover. *Science, 351*(6273), 600-604.

Aubinet, M., Vesala, T., & Papale, D. (2012). *Eddy covariance: a practical guide to measurement and data analysis*: Springer Science & Business Media.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., et al. (2001). FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society, 82*(11), 2415-2434.

Braswell, B. H., Sacks, W. J., Linder, E., & Schimel, D. S. (2005). Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Global change biology, 11*(2), 335-355.

Desai, A. R., Richardson, A. D., Moffat, A. M., Kattge, J., Hollinger, D. Y., Barr, A., et al. (2008). Cross-site evaluation of eddy covariance GPP and RE decomposition techniques. *Agricultural and Forest Meteorology, 148*(6-7), 821-838.

Dou, X., & Yang, Y. (2018). Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems. *Computers and Electronics in Agriculture, 148*, 95-106.

Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., et al. (2001). Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agricultural and forest meteorology, 107*(1), 43-69.

Fares, S., Conte, A., & Chabbi, A. (2018). Ozone flux in plant ecosystems: new opportunities for long-term monitoring networks to deliver ozone-risk assessments. *Environmental Science and Pollution Research, 25*. doi:10.1007/s11356-017-0352-0

Gleick, P. H. (1989). The implications of global climatic changes for international security. *Climatic Change, 15*(1), 309-325.

Gove, J., & Hollinger, D. (2006). Application of a dual unscented Kalman filter for simultaneous state and parameter estimation in problems of surface-atmosphere exchange. *Journal of Geophysical Research: Atmospheres, 111*(D8).

Granata, F. (2019). Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. *Agricultural Water Management, 217*, 303-315.

Gu, R. Y., Lo, M. H., Liao, C. Y., Jang, Y. S., Juang, J. Y., Huang, C. Y., et al. (2021). Early Peak of Latent Heat Fluxes Regulates Diurnal Temperature Range in

doi:10.6342/NTU202300497

Montane Cloud Forests. *Journal of Hydrometeorology, 22*(9), 2475-2487. doi:10.1175/jhm-d-21-0005.1

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*(8), 1735-1780.

Huang, I., & Hsieh, C.-I. (2020). Gap-Filling of Surface Fluxes Using Machine Learning Algorithms in Various Ecosystems. *Water, 12*(12), 3415.

Hung, L.-S., & Li, M.-H. (2020). Extreme weather events and health responses in Taiwan. In *Extreme Weather Events and Human Health* (pp. 197-207): Springer.

Janssen, P., & Heuberger, P. (1995). Calibration of process-oriented models. *Ecological Modelling, 83*(1-2), 55-66.

Jensen, M. E., Burman, R. D., & Allen, R. G. (1990). *Evapotranspiration and irrigation water requirements*.

Khan, M. S., Jeon, S. B., & Jeong, M. H. (2021). Gap-Filling Eddy Covariance Latent Heat Flux: Inter-Comparison of Four Machine Learning Model Predictions and Uncertainties in Forest Ecosystem. *Remote Sensing, 13*(24). doi:10.3390/rs13244976

Kim, S., Shiri, J., Kisi, O., & Singh, V. P. (2013). Estimating daily pan evaporation using different data-driven methods and lag-time patterns. *Water resources management, 27*(7), 2267-2286.

Kim, Y., Johnson, M. S., Knox, S. H., Black, T. A., Dalmagro, H. J., Kang, M., et al. (2020). Gap-filling approaches for eddy covariance methane fluxes: A comparison of three machine learning algorithms and a traditional method with principal component analysis. *Global change biology, 26*(3), 1499-1518.

Menzer, O., Moffat, A. M., Meiring, W., Lasslop, G., Schukat-Talamazzini, E. G., & Reichstein, M. (2013). Random errors in carbon and water vapor fluxes assessed with Gaussian Processes. *Agricultural and Forest Meteorology, 178*, 161-172.

Moffat, A. M., Papale, D., Reichstein, M., Hollinger, D. Y., Richardson, A. D., Barr, A. G., et al. (2007). Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agricultural and Forest Meteorology, 147*(3-4), 209-232.

Moureaux, C., Debacq, A., Bodson, B., Heinesch, B., & Aubinet, M. (2006). Annual net ecosystem carbon exchange by a sugar beet crop. *Agricultural and Forest Meteorology, 139*(1), 25-39. doi:https://doi.org/10.1016/j.agrformet.2006.05.009

Nourani, V., & Fard, M. S. (2012). Sensitivity analysis of the artificial neural network outputs in simulation of the evaporation process at different climatologic regimes. *Advances in Engineering Software, 47*(1), 127-146.

doi:10.6342/NTU202300497

Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378*.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., et al. (2006). Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. *Biogeosciences, 3*(4), 571-583.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., et al. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific data, 7*(1), 1-27.

Peng, D., Zhang, B., & Liu, L. (2012). Comparing spatiotemporal patterns in Eurasian FPAR derived from two NDVI-based methods. *International Journal of Digital Earth, 5*(4), 283-298.

Ramachandran, P., Zoph, B., & Le, Q. V. (2018). Searching for Activation Functions. *ArXiv, abs/1710.05941*.

Schmidt, A., Wrzesinsky, T., & Klemm, O. (2008). Gap filling and quality assessment of CO 2 and water vapour fluxes above an urban area with radial basis function neural networks. *Boundary-Layer Meteorology, 126*(3), 389-413.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena, 404*, 132306.

Stauch, V. J., & Jarvis, A. J. (2006). A semi-parametric gap-filling model for eddy covariance CO2 flux time series data. *Global change biology, 12*(9), 1707-1716.

Ueyama, M., Ichii, K., Iwata, H., Euskirchen, E. S., Zona, D., Rocha, A. V., et al. (2013). Upscaling terrestrial carbon dioxide fluxes in Alaska with satellite remote sensing and support vector regression. *Journal of Geophysical Research: Biogeosciences, 118*(3), 1266-1281.

Wang, H. J., Riley, W. J., & Collins, W. D. (2015). Statistical uncertainty of eddy covariance CO2 fluxes inferred using a residual bootstrap approach. *Agricultural and Forest Meteorology, 206*, 163-171. doi:10.1016/j.agrformet.2015.03.011

Wang, S.-C. (2003). *Interdisciplinary computing in Java programming* (Vol. 743): Springer Science & Business Media.

Wang, T., Brender, P., Ciais, P., Piao, S., Mahecha, M. D., Chevallier, F., et al. (2012). State-dependent errors in a land surface model across biomes inferred from eddy covariance observations on multiple timescales. *Ecological Modelling, 246*, 11-25.

doi:10.6342/NTU202300497

Webster, P. J., Holland, G. J., Curry, J. A., & Chang, H. R. (2005). Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science, 309*(5742), 1844-1846. doi:10.1126/science.1116448

Yu, G.-R., Wen, X.-F., Sun, X.-M., Tanner, B. D., Lee, X., & Chen, J.-Y. (2006). Overview of ChinaFLUX and evaluation of its eddy covariance measurement. *Agricultural and Forest Meteorology, 137*(3-4), 125-137.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting, 14*(1), 35-62.

Zhao, L., LI, Y., XU, S., ZHOU, H., GU, S., YU, G., et al. (2006). Diurnal, seasonal and annual variation in net ecosystem CO2 exchange of an alpine shrubland on Qinghai-Tibetan plateau. *Global Change Biology, 12*(10), 1940-1953. doi:https://doi.org/10.1111/j.1365-2486.2006.01197.x

doi:10.6342/NTU202300497