water

MDPI

Article

# Gap-Filling of Surface Fluxes Using Machine Learning Algorithms in Various Ecosystems

**I-Hang Huang and Cheng-I Hsieh ***

Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 10617, Taiwan; scream831003@gmail.com
* Correspondence: hsieh@ntu.edu.tw; Tel.: +886-2-3366-3473

Received: 10 September 2020; Accepted: 1 December 2020; Published: 4 December 2020

check for updates

**Abstract:** Five machine learning (ML) algorithms were employed for gap-filling surface fluxes of $CO_2$, water vapor, and sensible heat above three different ecosystems: grassland, rice paddy field, and forest. The performance and limitations of these ML models, which are support vector machine, random forest, multi-layer perception, deep neural network, and long short-term memory, were investigated. Firstly, the accuracy of gap-filling to time and hysteresis input factors of ML algorithms for different ecosystems is discussed. Secondly, the optimal ML model selected in the first stage is compared with the classic method—the Penman–Monteith (P–M) equation for water vapor flux gap-filling. Thirdly, with different gap lengths (from one hour to one week), we explored the data length required for an ML model to perform the optimal gap-filling. Our results demonstrate the following: (1) for ecosystems with a strong hysteresis between surface fluxes and net radiation, adding proceeding meteorological data into the model inputs could improve the model performance; (2) the five ML models gave similar gap-filling performance; (3) for gap-filling water vapor flux, the ML model is better than the P–M equation; and (4) for a gap with length of half day, one day, or one week, an ML model with training data length greater than 1300 h would provide a better gap-filling accuracy.

**Keywords:** flux gap-filling; machine learning; Penman–Monteith equation; artificial neural network

## 1. Introduction

In order to study climate change, hydrological cycle, and atmosphere–surface interactions, information on fundamental factors such as carbon dioxide ($CO_2$), latent heat (LE), and sensible heat (H) fluxes are indispensable. To date, the most accurate and reliable method to obtain these surface flux data is the eddy-covariance method. However, this method relies on high frequency measurements of three-dimensional sonic anemometers and $CO_2$/$H_2O$ infrared gas analyzers, which are often forced to stop by rain or instrument maintenance, and complete flux time series data sometimes cannot be obtained. Furthermore, the process of data quality inspection will also result in missing flux data. Therefore, gap-filling missing flux data is a key process for calculating mass and energy budgets, such as ecological carbon budgets, evapotranspiration, and water resource balance. Commonly used gap-filling methods adopt low-frequency meteorological information such as temperature, humidity, wind speed, and net radiation to perform linear or non-linear regressions to make up loss data. These methods include the following: non-linear regression, interpolation, mean diurnal variation, look-up tables, multiple imputation, and marginal distribution sampling [1,2]. The advantages and disadvantages of these methods can be found in [1,3–11].

Aside from classic statistics-based and process-based gap-filling methods, gap-filling based on data-driven, that is, machine learning (ML) algorithms have also become popular in recent years. Van Wijk and Bouten [12] used artificial neural networks (ANNs), a type of classic ML algorithm, to simulate $CO_2$ and water vapor fluxes at six different forest stations in Europe. Their results showed that ANNs can effectively simulate $CO_2$ and water vapor fluxes without very detailed observed meteorological data and geographic environmental information. Carrara et al. [13] adopted average daily sunshine duration, net radiation, air temperature, soil temperature, rainfall, and relative humidity as input factors of ANNs to implement gap-filling of $CO_2$ flux for a mixed temperate forest. The correlation coefficient (CC) in their study was higher than 0.9, showing outstanding model performance. Schmidt et al. [14] used radial-based function neural network (RBFNN) to fill in missing $CO_2$ and water vapor fluxes above an urban area. The results indicated that the CC can reach 0.85. However, Kordowski and Kutterl [15] used the same method and input combination to estimate $CO_2$ flux over an urban park area, but the CC between the simulated and observed values was only 0.76, and there was a severe underestimation. The difference between these two studies reveals that the model performance could be different under different ecosystems.

Presently, for surface flux gap-filling, the ML algorithms were found to perform better than the other statistics- or process-based methods for both diurnal and nocturnal periods [2,3,16,17]. With the enhancement of computing power, more ML algorithms that require huge amounts of calculations have appeared one after another. Therefore, there are more follow-up studies focused on the use of data-preprocessing and novel algorithms for gap-filling. For example, Dengel et al. [18] attempted to construct an ANN model for gap-filling methane ($CH_4$) flux. Nguyen and Halem [19] adopted two deep learning (DL) algorithms, feed forward neural network (FFNN) and long short-term memory (LSTM), to predict $CO_2$ flux with the following input variables: $CO_2$ concentration, relative humidity, air pressure, air temperature, and wind speed. Their results showed that both FFNN and LSTM could predict $CO_2$ flux accurately ($R^2$ = 0.83 and 0.88, respectively; $R^2$: coefficient of determination). Kim et al. [20] selected three types of classic ML algorithms, ANN, support vector machine (SVM), and random forest (RF), to implement the gap-filling of $CH_4$ flux. Their results demonstrated that all the three ML models effectively filled in the missing values of $CH_4$ flux, of which the RF accuracy was the highest and the ANN was second. They also concluded that, if the gap of flux was less than monthly scale, two to three years of data (24–36 times of the gap length) can effectively construct a robust ML model. Kang et al. [21] used SVM with a long-term flux and meteorological data of 17 years to supplement the long missing data (i.e., gaps longer than 30 days) of water vapor and $CO_2$ fluxes. They concluded that "using a longer training dataset in the machine learning generally produced better model performance", although using long-term data might average out the effect of spatiotemporal variability.

In short, ML algorithms have been demonstrated to be able to effectively fill in gaps in flux data by using low-frequency meteorological data as the model training inputs. However, few studies have discussed the effects of various ecosystems on gap-filling. The purposes of this study are (1) to compare five ML methods for gap-filling of $CO_2$, water vapor, and sensible heat fluxes above three ecosystems: grassland, rice paddy field, and forest; (2) to compare the classic Penman–Monteith equation with the ML algorithm for gap-filling of water vapor flux; and (3) to investigate how much data are required for training when applying the ML model for gap-filling fluxes with various gap-length (one hour, half day, one day, and one week). The five ML algorithms employed in this study are three classic ones: SVM, RF, and multi-layer perceptron (MLP), and two relatively novel deep learning ones: deep neural network (DNN) and LSTM.

## 2. Study Sites and Data

Fluxes of $CO_2$, water vapor, and sensible heat collected from three ecosystems, grassland, rice paddy, and forest, were adopted for this study. Experiments at these sites are described below.

### *2.1. Grassland*

The flux measurements were conducted above a flat grassland (52°8′24″ N, 8°39′36″ W) in Cork, Ireland. Because of the temperate zone, the average annual precipitation and temperature of the experiment field were 1470 mm and 11.5 °C, respectively. Data used in this study were collected from 1 January 2002 until 31 December 2002. The grassland was predominantly pasture and meadow; the dominant plant species was the perennial ryegrass. The canopy height varied from 0.1 to 0.45 m.

The three-dimensional sonic anemometer (RM Young 81,000) and open-path $CO_2/H_2O$ infrared gas analyzer (LI7500, Li-Cor) were installed on the top of a 10 m high meteorological tower to measure $CO_2$, water vapor, and sensible heat fluxes. These flux measurements were sampled at 10 Hz and averaged every 30 min. A HMP45A sensor (Vaisala, Helsinki, Finland) was also installed at 3 m to measure air temperature and humidity, and a CNR1 net radiometer (Kipp & Zonen, Delft, The Netherlands) was used to measure net radiation. More details about this grassland and experiment can be found in Jaksic et al. [22] and Hsieh et al. [23].

### 2.2. Rice Paddy Field

The experiment was conducted over a rice paddy in Jiaosi, Yilan County (24°48′07″ N, 121°47′58″ E) in northeastern Taiwan. The experimental period was from 11 March 2010 to 31 August 2011. The site is flat and the fetch along the mean wind direction is about 1.5 km. The local climate is East Asian subtropical monsoon climate, hot and humid in summer, and rainy in winter; the average annual temperature and precipitation were about 22 °C and 3000 mm, respectively.

The rice harvest frequency in this area was once a year, the period of crop growth was from March to July, and the rice paddy field was flooded with water for 4 months from early January. The height of the crop gradually grew from 0.1 to 1.2 m, so the surface roughness changed with time. In this study, only data during the growing period were used and the leaf area index (LAI) varied from 0.82 to 5.68 ($m^2/m^2$) with an average of 3.35 in the growing season.

An eddy-covariance system was installed at 2.94 m above the ground to measure sensible heat, water vapor, and $CO_2$ fluxes. This system consisted of a three-dimensional sonic anemometer (R.M. Young 81,000) and an open-path gas analyzer (LI7500, Li-Cor) to measure wind velocity, sonic temperature, and concentrations of water vapor and $CO_2$. The sampling frequency and averaging period for the eddy-covariance system were 10 Hz and 30 min, respectively. In addition, a net radiometer and a temperature and humidity sensor were set at 2.49 and 2.80 m above the ground, respectively, to measure the mean net radiation, air temperature, and humidity every 30 min. A data-logger (CR23X, Campbell Scientific) was used to collect the signals of measurements and all the data were then transmitted to a lap-top computer for further analysis. For the flux calculation, the general FLUXNET standard process [24] was adopted and the Webb–Pearman–Leuning correction was applied to correct the fluctuation of air density.

### 2.3. Forest

This forest site was located in the Chi-lan Mountain (24°35′27″ N, 121°29′56″ W), Yilan County in northeastern Taiwan. The data adopted for this study spanned from 1 January 2007 to 31 December 2007. This forest is a yellow Cypress that covers an area of about 374 hectares, with an altitude ranging from 1650 to 2432 m with a uniform slope of 15 degrees. The average canopy height of the forest was about 10.3 m with an LAI of 6.3 ($m^2/m^2$), and the surface roughness was around 1.0 m. Because of the location and elevation of the study area, the climate was consistently mild and warm, with an average annual precipitation of 4000 mm and an average temperature of 13 °C. The frequent precipitation resulted in a relatively steady water content in the soil, about 0.3–0.4 ($m^3/m^3$) at 30 cm depth.

In this experiment, the high-frequency flux measuring instruments, consisting of a three-dimensional sonic anemometer (R. M. Young 81,000) and an LI7500 open path infrared $CO_2/H_2O$ analyzer, were installed on the top of a 24 m tall meteorological tower. Moreover, net radiation was measured at 22.5 m by a CNR1 radiometer (Kipp & Zonen, Delft, Netherlands); air temperature and humidity were measured at 23.5 m with a HMP45A sensor (Vaisala, Helsinki, Finland). More details about the site can be found in the work by Lin et al. [2] and Chu et al. [25].

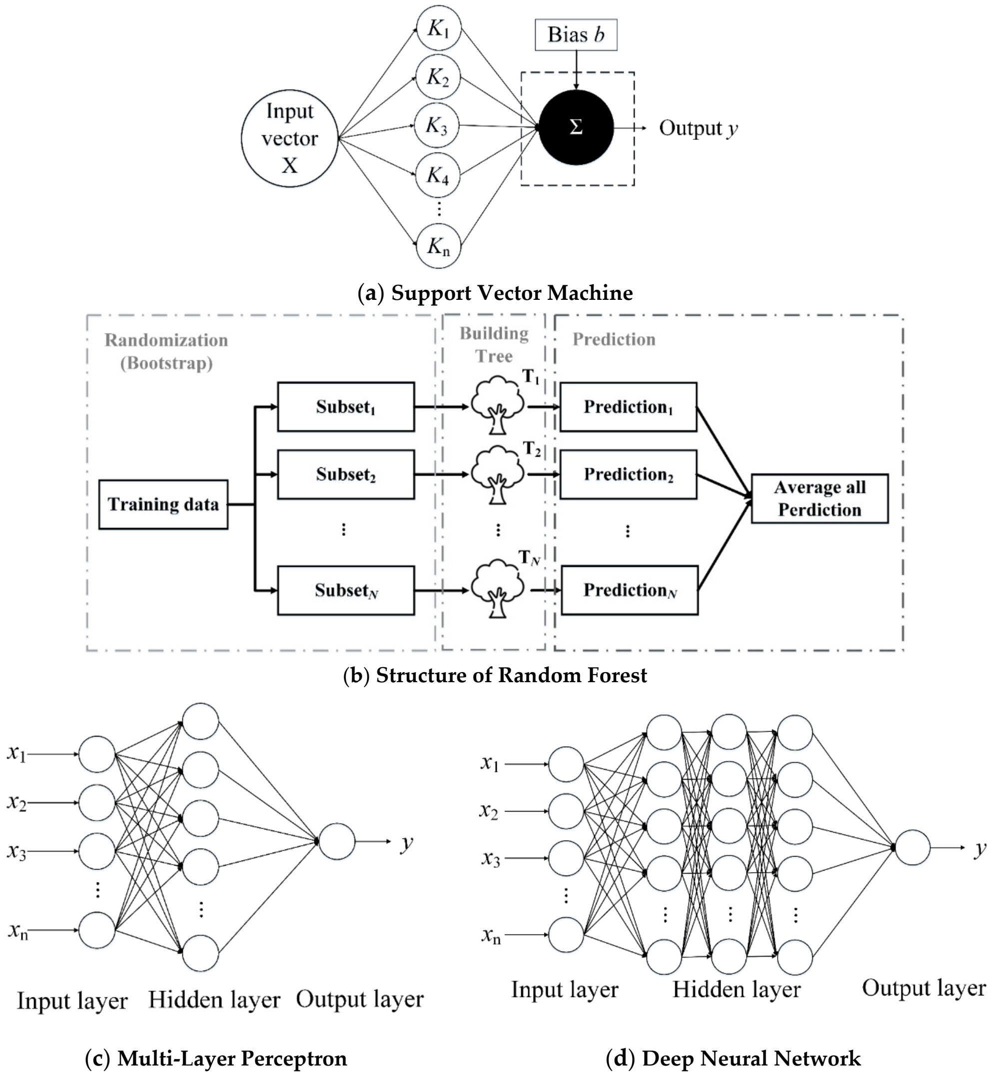
## 3. Methods

### *3.1. Machine Learning Algorithms*

The ML algorithm is a data-driven approach. The ML models analyze the internal correlation of input data to construct a model that applies regression or classification to future or missing data. Implementation of the five ML algorithms (SVM, RF, MLP, DNN, LSTM) adopted in this study is described below. We used the grid-search method [26] (see Appendix A for a brief introduction) to determine the hyper-parameter optimization in all the five models.

#### 3.1.1. Support Vector Machine (SVM)

The support vector machine in this study was mainly constructed by the Scikit-Learn package in Python. The grid-search method was applied to determine the optimal combination of hyper-parameters of SVM. The hyper-parameters (i.e., kernel functions) included radial basis function (RBF) and linear function and the penalty coefficient ranged from $2^{-5}$ to $2^{5}$. For RBF, the Gamma range was from $2^{-5}$ to $2^{5}$. The model structure of SVM is shown in Figure 1a; the X is the input vector. $K_n$ is the kernel function used to draw the hyperplane; b is the bias term; and y is the output. More details about the SVM can be found in Cortes and Vapnik [27].



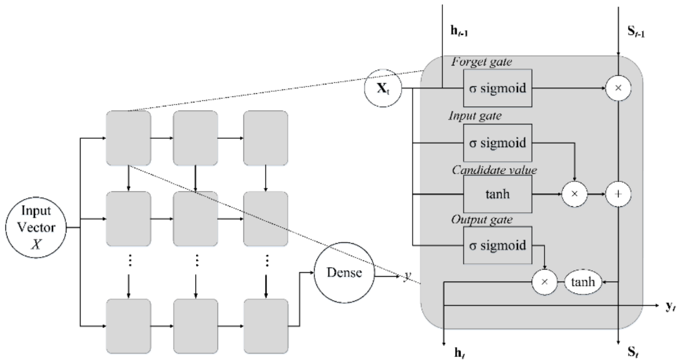

(**a**) **Support Vector Machine**

(**b**) **Structure of Random Forest**

(**c**) **Multi-Layer Perceptron**

(**d**) **Deep Neural Network**

**Figure 1.** *Cont.*




(**e**) **Long Short-Term Memory**

**Figure 1.** Model structure for (**a**) support vector machine (SVM), (**b**) random forest (RF), (**c**) multi-layer perceptron (MLP), (**d**) deep neural network (DNN), and (**e**) long short-term memory (LSTM).

### 3.1.2. Random Forest (RF)

The present study also used the Scikit-learn package in Python to build the random forest, in which the number of decision trees was gradually increased from 100 to 500. The maximum number of features is automatically selected by the Scikit-learn package. In order to avoid overfitting, the maximum depth of the decision tree in this study ranged from 1 to 10. The structure of the RF is presented in Figure 1b; a detailed description can be found in Breiman [28].

### 3.1.3. Multi-Layer Perceptron (MLP)

The package used to build MLP in this study was TensorFlow developed by Google. The model structure is shown in Figure 1c; in between the input and output layers, only one hidden layer is used to build the neural network. The number of neurons used in the hidden layer was from 3 to 512. The grid-search method was used to determine the hyper-parameters. In order to distinguish from the novel deep learning networks, the activation function used here is Sigmoid function. A detailed description can be found in Rumelhart et al. [29].

### 3.1.4. Deep Neural Network (DNN)

The main improvement of DNN is the introduction of ReLU as the activation function to avoid the gradient vanishing; therefore, most DL models can contain three or more hidden layers. In addition, the invention of the new optimizers such as Adam or Nadam [30] allows us to train a large number of hyper-parameters more quickly; the regularization and dropout [31] enable us to avoid overfitting. The DNN used in this study was the TensorFlow package published by Google. The model structure is shown in Figure 1d, the number of hidden layers is three, and the number of neurons used in each hidden layer ranges from 4 to 32 neurons. The activation function used is ReLU, and the optimizer is either Rmsprop, Adam, or Nadam (based on optimization). More details about DNN can be found in Hinton et al. [32].

### 3.1.5. Long Short-Term Memory (LSTM)

LSTM is a type of improved recurrent neural network (RNN)—details of LSTM are given in Hochreiter et al. [33]. As many studies have shown [34–36], LSTM can more effectively predict the trend of mid- and long-term events than other ML algorithms. The LSTM used in this study was the Keras package launched by Google. The structure of an LSTM memory cell is presented in Figure 1e. The memory cell includes forget, input, and output gates. The forget gate is used to filter whether the data memorized in the model is still valuable. The valuable data will be used to forecast in this round and will be retained in the next round. This process would be implemented through a Sigmoid function filter. The input gate and output gate are used to determine whether the new input data are useful, and as the output value of the neuron. Beside three gate units, there is a candidate value, which will be used to determine the magnitude of the updated neuron. In Figure 1e, $X_t$ is the input vector given to the LSTM model in this round of forecast; $h_{t-1}$ and $h_t$ are the forecasting results of the previous and current rounds ($h_{t-1}$ and $h_t$ also passed to the next forecasting round as inputs to deliver the model state); $S_{t-1}$ and $S_t$ are the cell state at time t − 1 and time t. tanh is the hyperbolic tangent function. In this study, the LSTM contained two to three hidden layers, and each layer contained 4 to 32 memory cells (neurons). The activation function used here for the layer connection was ReLU. The optimizer was selected from Rmsprop, Adam, or Nadam.

#### 3.1.6. Input Variables for Training the ML Models

In this study, the input variables for training the ML models included the following: (1) time factors: Julian day (JD) and day and night time (DN); the DN is a time index that converts the 24 h in a day into 0–1; (2) meteorological factors: air temperature ($T_a$), relative humidity (RH), net radiation ($R_n$), and wind speed (U); and (3) hysteresis factors.

The hysteresis phenomenon between surface fluxes and net radiation has been widely noted (e.g., Cui et al. [37]; Lin et al. [38]). However, none of the ML models for flux gap-filling have taken this into account. Other environmental factors (e.g., air temperature, wind speed, soil water content, relative humidity) also have hysteresis relationships with surface fluxes (Cui et al. [37]; Zhang et al. [39]). In addition to net radiation, we also select one of the environmental factors, air temperature, to explore its hysteresis relation with surface fluxes at these three sites. Hence, in this study, the third input variable group is the hysteresis factors consisting of $R_n$ (t − 2), $R_n$ (t − 1), $T_a$ (t − 2), $T_a$ (t − 1), and $T_{avg}$; $T_{avg}$ is the average of air temperature at time t − 2, t − 1, and t.; $R_n$ (t − 2) and $R_n$ (t − 1) are the net radiation at time t − 2 and t − 1, respectively; $T_a$ (t − 2) and $T_a$ (t − 1) are the air temperature at time t − 2 and t − 1, respectively. In this study, each time step is 30 min.

All the input variables from these three groups are listed in Table 1. In this study, we explore the best input combinations from these three groups for training the ML models for various sites and fluxes and compare the model performance of the five ML models for gap-filling the three fluxes at the three sites.

**Table 1.** List of input variables for machine learning (ML) model training.

| Input Factors | Abbreviation | Definition |
|---|---|---|
| Time factors | JD | Julian day |
| | DN | Day and night time index, which converts 24 h in a day to a continuous value from 0 to 1. |
| Meteorological factors | $T_a(t)$ | air temperature at time t (°C) |
| | RH(t) | relative humidity at time t (%) |
| | $R_n(t)$ | net radiation at time t (W/m$^2$) |
| | $U(t)$ | wind speed at time t (m/s) |
| Hysteresis factors | $R_n(t-1)$ | net radiation at time t − 1 (i.e., 30 min before time t) (W/m$^2$) |
| | $R_n(t-2)$ | net radiation at time t − 2 (i.e., one hour before time t) (W/m$^2$) |
| | $T_a(t-1)$ | air temperature at time t − 1 (°C) |
| | $T_a(t-2)$ | air temperature at time t − 2 (°C) |
| | $T_{avg}(t)$ | average of the air temperatures measured at time t, t − 1, and t − 2 (°C) |

### *3.2. Penman–Monteith Equation*

In this study, the gap-filling of latent heat flux (LE) was also performed by the classic physical-process based model, the Penman–Monteith (P–M) equation:

$$\mathrm{LE} = \frac{\Delta Q_n + \rho c_p D / r_{av}}{\Delta + \gamma(1 + r_{st}/r_{av})} \tag{1}$$

where $\Delta$ (kPa $K^{-1}$) represents the slope of the saturated vapor pressure-temperature curve at the air temperature $T_a$; $\gamma(=\frac{\rho c_p}{0.622L_v})$ is the psychrometric constant; $\rho$ (= 1.2 kg $m^{-3}$) and $c_p$ (= 1005 J $kg^{-1}K^{-1}$) represent the air density and specific heat for the air, respectively; $L_v$ is the latent heat of vaporization and is equal to $2.45 \times 10^6$ (J $kg^{-1}$); D (kPa) represents the vapor pressure deficit; $r_{av}$ and $r_{st}$ are the aerodynamic resistances of water vapor and stomatal resistance; $Q_n$ (= $R_n - G_s$) is the available energy; and $G_s$ is the soil heat flux. The $r_{av}$ can be expressed as [36]

$$r_{av} = \frac{ln\left[\frac{z-d}{z_{0m}}\right]ln\left[\frac{z-d}{z_{0v}}\right]}{k^2U} \tag{2}$$

where $k$ (=0.4) is the von Karman constant, $z$ is the measurement height, $d$ is the zero-plane displacement height (≈2/3 of the canopy height), $z_{0m}$ is the surface roughness for momentum, and $z_{0v}$ is the surface roughness for water vapor. With Equation (1), we can fill the gap of LE with measured low-frequency meteorological data and local $r_{st}$ derived from the measured latent heat flux (see Appendix B).

### 3.3. Flux Gap Scenario

Different gap lengths may require different data lengths to train the ML model. In this study, we consider four gap-length scenarios: one hour, half day, one day, and one week. The following steps were taken for each gap length to investigate the relation between gap length and data length of training. (1) The highest point of the measured flux was selected as the central point of the gap-length as this point was the most difficult point for gap-filling; (2) the closest neighbor points to the gap were selected as the training data; and (3) the training data length was started from 20 h and was increased gradually to a maximum of 1600 h with a time step of 20 h.

### 3.4. Research Process and Performance Metrics

#### 3.4.1. Research Process

The research process of this study includes three stages described below.

(1) Explore the optimal combinations of input variables for constructing the five ML models for gap-filling of surface fluxes at the three sites, and then compare the model performance. For constructing the ML model, the ratio of data sets for training, validation, and testing was 5:3:2, and each of the data sets was randomly selected with uniform distribution.

(2) In the second stage, the best ML model selected from the first stage was compared with the P–M equation to explore the water vapor flux gap-filling accuracy of both methods at the three ecosystems. The determination of $r_{st}$ for the P–M equation at the three sites is described in Appendix B.

(3) In the third stage, the relation between gap length (one hour, half day, one day, and one week) and training data length (20 to 1600 h) was investigated by the steps in Section 3.3. The ML model adopted here was the best model selected from the first stage.

#### 3.4.2. Performance Metrics

In order to objectively quantify the performance of each model, four commonly used performance metrics for linear or nonlinear data were selected and are described below.

(1) Root mean square error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum\nolimits_{t=1}^{n}(C_t - \hat{C}_t)^2} \tag{3}$$

where $C_t$ and $\hat{C}_t$ are the values from observation and model prediction, respectively, and n represents the total number of data points.

(2) Mean absolute error (MAE)

$$\mathrm{MAE} = \frac{1}{n}\sum\nolimits_{1}^{n}\left|C_t - \hat{C}_t\right| \quad (4)$$

It is worth noting that the RMSE amplifies extreme errors more and ignores the small errors, while the MAE considers the average of all errors.

(3) Coefficient of determination ($R^2$)

$$R^2 = \left(\frac{\sum_{t=1}^{n}\left(C_t - \overline{C}\right)\left(C_t - \overline{\hat{C}}\right)}{\sqrt{\sum_{t=1}^{n}\left(C_t - \overline{C}\right)^2 \sum_{t=1}^{n}\left(\hat{C}_t - \overline{\hat{C}}\right)^2}}\right)^2 \quad (5)$$

where $\overline{C}$ is the mean of the observed values and $\overline{\hat{C}}$ represents the mean of the estimated values.

(4) Coefficient of efficiency (CE)

$$\mathrm{CE} = 1 - \frac{\sum_{t=1}^{n}\left(C_t - \hat{C}_t\right)^2}{\sum_{t=1}^{n}\left(C_t - \overline{C}\right)^2} \quad (6)$$

The CE evaluates whether the estimated value generated by the model is better than the estimated value using a direct average. The closer the value is to 1, the more ideal it is.

## 4. Results and Discussion

### 4.1. Optimal Input Combinations for Training ML Models

The optimal input combinations for $CO_2$, latent heat, and sensible heat fluxes at the three sites determined by the model performance metrics are listed in Tables 2–4, respectively. It is noticed that, for the grassland site, the hysteresis factors played no roles in all three fluxes, and the time factors have some influences on all fluxes at the three sites. To further examine the influences of time factors and hysteresis factors on the three fluxes at the three sites, the averaged model performances with and without these two factors are summarized in Tables 5–7 for the grassland, rice paddy field, and forest, respectively. The individual model performance with and without time and hysteresis factors is listed in Appendix C. Tables 5–7 demonstrate the following:

(1) For the grassland, the improvements by including time factors in the input combination are less than 5% for all three fluxes. For the rice paddy field, the improvements for RMSE for the three fluxes range from 8.6 to 19.7%, showing that time factors' influence is larger at this site. For the forest, this influence on $CO_2$ flux is small (2.9%), but it is larger for water vapor and sensible heat fluxes (7.26–7.9%, respectively).

(2) Concerning the hysteresis factors, at the grassland site, these factors have no influence on all three fluxes. For the rice paddy, the influence on $CO_2$ flux is less important, but important for water vapor and sensible heat fluxes (RMSE improved by 8.72–9.50%). For the forest site, the influence on $CO_2$ flux is important (RMSE improved by 8.10%), but the influence on both LE and H is small (improvement rates of RMSE both less than 2%). Cui et al. [37] found that the magnitude of hysteresis between LE and net radiation is large on water surfaces and small on land surfaces. Our results for LE reveal that the hysteresis factors are stronger for the rice paddy field (flooded with water during growing season), but small for forest and grassland sites. This is consistent with the finding of Cui et al. [37].

**Table 2.** The optimal input combinations for $CO_2$ flux gap-filling at the three ecosystems using different machine learning models. SVM, support vector machine; RF, random forest; MLP, multi-layer perceptron; DNN, deep neural network; LSTM, long short-term memory.

| Model | Study Site | Optimal Input Combinations |
|---|---|---|
| SVM | Grassland | JD, DN, $T_a(t)$,RH, $R_n(t)$, $U(t)$ |
| | Rice paddy field | JD, DN, $T_a(t)$, $R_n(t)$, $U(t)$ |
| | Forest | JD, DN, $T_a(t)$, RH, $R_n(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$ |
| RF | Grassland | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$ |
| | Rice paddy field | DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |
| | Forest | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |
| MLP | Grassland | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$ |
| | Rice paddy field | JD, DN, $T_a(t)$, $R_n(t)$, $U(t)$, $R_n$ $(t-1)$, $T_{avg}(t)$ |
| | Forest | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-1)$, $T_{avg}(t)$ |
| DNN | Grassland | JD, $T_a(t)$, RH, $R_n(t)$, $U(t)$ |
| | Rice paddy field | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |
| | Forest | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |
| LSTM | Grassland | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$ |
| | Rice paddy field | JD, DN, $T_a(t)$, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |
| | Forest | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, |

**Table 3.** Same as Table 2, but for latent heat flux.

| Model | Study Site | Optimal Input Combinations |
|---|---|---|
| SVM | Grassland | $T_a(t)$, RH, $R_n(t)$, $U(t)$ |
| | Rice paddy field | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |
| | Forest | JD, DN, $T_a(t)$, RH, $R_n(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |
| RF | Grassland | DN, $T_a(t)$, $R_n(t)$, $U(t)$ |
| | Rice paddy field | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |
| | Forest | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-1)$, $T_{avg}(t)$ |
| MLP | Grassland | $T_a(t)$, RH, $R_n(t)$, $U(t)$ |
| | Rice paddy field | JD, $T_a(t)$, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |
| | Forest | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-1)$, $T_{avg}(t)$ |
| DNN | Grassland | DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$ |
| | Rice paddy field | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |
| | Forest | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |
| LSTM | Grassland | $T_a(t)$, RH, $R_n(t)$, $U(t)$ |
| | Rice paddy field | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |
| | Forest | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |

**Table 4.** Same as Table 2, but for sensible heat flux.

| Model | Study Site | Optimal Input Combinations |
|---|---|---|
| SVM | Grassland | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$ |
| | Rice paddy field | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |
| | Forest | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-1)$, $T_a(t-1)$, $T_{avg}(t)$ |
| RF | Grassland | JD, $T_a(t)$, RH, $R_n(t)$, $U(t)$ |
| | Rice paddy field | JD, DN, $R_n(t)$, $U(t)$, $T_{avg}(t)$ |
| | Forest | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-1)$, $T_{avg}(t)$ |
| MLP | Grassland | JD, $T_a(t)$, $R_n(t)$, $U(t)$ |
| | Rice paddy field | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |
| | Forest | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-1)$, $T_a(t-1)$, $T_{avg}(t)$ |
| DNN | Grassland | JD, $T_a(t)$, $R_n(t)$, $U(t)$ |
| | Rice paddy field | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |
| | Forest | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $T_{avg}(t)$ |
| LSTM | Grassland | JD, $T_a(t)$, RH, $R_n(t)$, $U(t)$ |
| | Rice paddy field | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $R_n(t-2)$, $R_n(t-1)$, $T_a(t-2)$, $T_a(t-1)$, $T_{avg}(t)$ |
| | Forest | JD, DN, $T_a(t)$, RH, $R_n(t)$, $U(t)$, $T_{avg}(t)$ |

**Table 5.** Summary of averaged model performance with and without time factors and hysteresis factors as inputs for machine learning (ML) model training at the grassland. The model's average was taken from the five ML algorithms with the optimal input combinations listed in Tables 2–4. RMSE, root mean square error; MAE, mean absolute error; CE, coefficient of efficiency.

| Flux | Type | RMSE | MAE | $R^2$ | CE |
|---|---|---|---|---|---|
| $CO_2$ flux | Model's average | 4.94 | 3.09 | 0.59 | 0.59 |
| | without time factors | 5.10 | 3.25 | 0.57 | 0.56 |
| | without hysteresis factors | 4.94 | 3.09 | 0.59 | 0.59 |
| | Improvement rate with time factors (%) | 3.10 | 4.74 | 4.58 | 4.96 |
| | Improvement rate with hysteresis factors (%) | 0.00 | 0.00 | 0.00 | 0.00 |
| Latent heat flux | Model's average | 23.34 | 13.47 | 0.85 | 0.85 |
| | without time factors | 23.44 | 13.58 | 0.85 | 0.85 |
| | without hysteresis factors | 23.34 | 13.47 | 0.85 | 0.85 |
| | Improvement rate with time factors (%) | 0.43 | 0.80 | 0.24 | 0.00 |
| | Improvement rate with hysteresis factors (%) | 0.00 | 0.00 | 0.00 | 0.00 |
| Sensible heat flux | Model's average | 16.73 | 10.77 | 0.83 | 0.83 |
| | without time factors | 17.00 | 10.96 | 0.83 | 0.83 |
| | without hysteresis factors | 16.73 | 10.77 | 0.83 | 0.83 |
| | Improvement rate with time factors (%) | 1.60 | 1.77 | 0.48 | 0.48 |
| | Improvement rate with hysteresis factors (%) | 0.00 | 0.00 | 0.00 | 0.00 |

**Table 6.** Same as Table 5, but for the rice paddy field.

| Flux | Type | RMSE | MAE | $R^2$ | CE |
|---|---|---|---|---|---|
| $CO_2$ flux | Model's average | 2.42 | 1.80 | 0.89 | 0.87 |
| | without time factors | 3.01 | 2.30 | 0.87 | 0.79 |
| | without hysteresis factors | 2.46 | 1.86 | 0.89 | 0.86 |
| | Improvement rate with time factors (%) | 19.73 | 21.70 | 3.23 | 10.10 |
| | Improvement rate with hysteresis factors (%) | 1.95 | 2.80 | 0.22 | 0.93 |
| Latent heat flux | Model's average | 18.51 | 12.81 | 0.89 | 0.89 |
| | without time factors | 20.25 | 13.52 | 0.87 | 0.87 |
| | without hysteresis factors | 20.45 | 14.30 | 0.86 | 0.86 |
| | Improvement rate with time factors (%) | 8.59 | 5.27 | 2.30 | 2.31 |
| | Improvement rate with hysteresis factors (%) | 9.50 | 10.39 | 2.78 | 2.55 |
| Sensible heat flux | Model's average | 10.41 | 6.53 | 0.87 | 0.87 |
| | without time factors | 11.98 | 7.47 | 0.83 | 0.83 |
| | without hysteresis factors | 11.40 | 7.05 | 0.85 | 0.85 |
| | Improvement rate with time factors (%) | 13.09 | 12.58 | 4.80 | 5.06 |
| | Improvement rate with hysteresis factors (%) | 8.72 | 7.29 | 2.82 | 2.83 |

**Table 7.** Same as Table 5, but for the forest site.

| Flux | Type | RMSE | MAE | $R^2$ | CE |
|---|---|---|---|---|---|
| $CO_2$ flux | Model's average | 3.43 | 2.32 | 0.81 | 0.81 |
| | without time factors | 3.53 | 2.42 | 0.80 | 0.80 |
| | without hysteresis factors | 3.73 | 2.56 | 0.78 | 0.77 |
| | Improvement rate with time factors (%) | 2.95 | 4.30 | 1.00 | 1.00 |
| | Improvement rate with hysteresis factors (%) | 8.10 | 9.60 | 3.87 | 4.40 |
| Latent heat flux | Model's average | 52.52 | 36.71 | 0.71 | 0.71 |
| | without time factors | 57.01 | 40.11 | 0.66 | 0.66 |
| | without hysteresis factors | 53.43 | 37.28 | 0.70 | 0.70 |
| | Improvement rate with time factors (%) | 7.88 | 8.49 | 7.25 | 8.23 |
| | Improvement rate with hysteresis factors (%) | 1.70 | 1.53 | 1.14 | 1.43 |
| Sensible heat flux | Model's average | 61.77 | 41.31 | 0.84 | 0.84 |
| | without time factors | 66.54 | 44.09 | 0.82 | 0.82 |
| | without hysteresis factors | 62.85 | 41.44 | 0.84 | 0.84 |
| | Improvement rate with time factors (%) | 7.17 | 6.31 | 3.43 | 3.43 |
| | Improvement rate with hysteresis factors (%) | 1.72 | 0.32 | 0.96 | 0.96 |

In addition to the time and hysteresis factors, at the rice paddy field, the LAI measurements were also carried out and the values varied with time; hence, LAI's influence on model performance was also considered. The results are presented in Appendix D. The ML model adopted in Appendix D was SVM. From Appendix D, it is noticed that the influences of LAI on gap-filling of $CO_2$ flux, LE, and H at the rice paddy field are all small.

### 4.2. Comparisons of Gap-Filling by ML Models

In this section, we compare the gap-filling results generated from the five ML models for $CO_2$, water vapor, and sensible heat fluxes at the three ecosystems.

#### 4.2.1. Carbon Dioxide Flux

Figure 2 shows a typical time series for $CO_2$ flux obtained from the experimental measurements and the five ML models at the three sites. As Figure 2a shows, in the grassland, the results of the five ML models were similar and could underestimate the peak values of $CO_2$ fluxes around noon. For the rice paddy field and forest ecosystem, all five models produced similar predictions and could accurately reproduce the flux peak values.

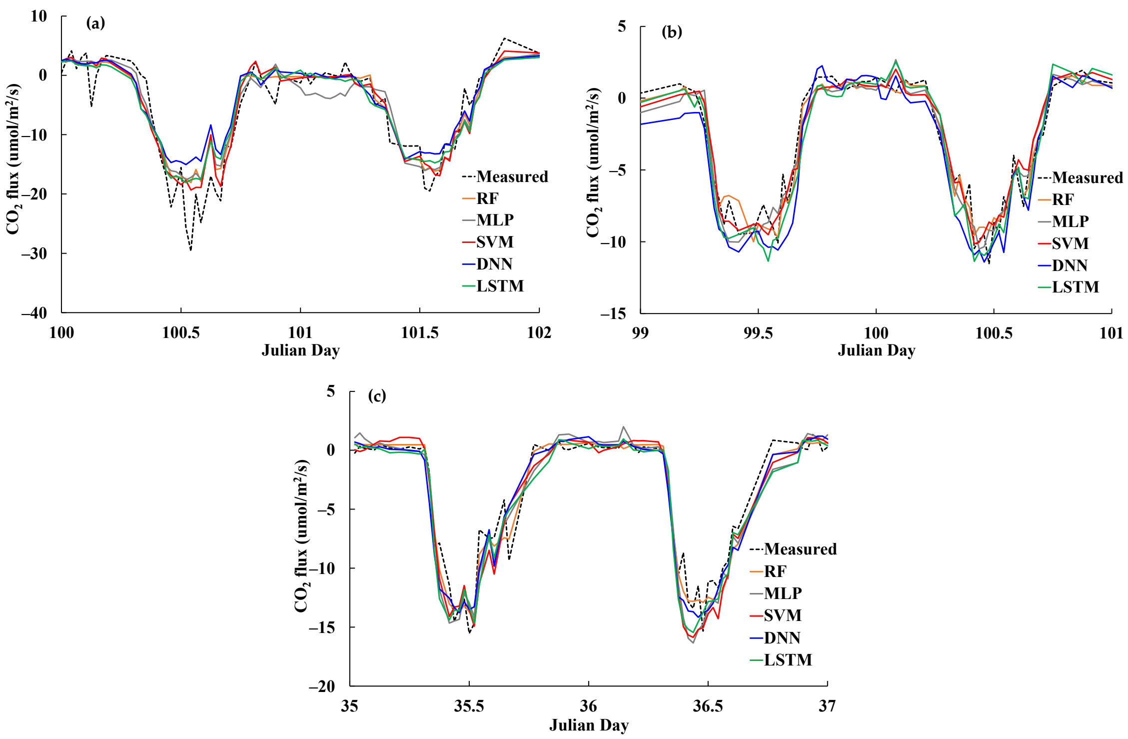



**Figure 2.** Typical time series of $CO_2$ flux obtained from the measurements and the five machine learning (ML) algorithms: (**a**) grassland, (**b**) rice paddy field, and (**c**) forest.

The linear regression analyses between measured and model predicted fluxes at the three sites are summarized in Table 8. From Table 8 and Appendix C, notice that RF had the best performance in the grassland ecosystem, while the SVM performed best in both the rice paddy field and forest ecosystem. Moreover, from Tables 5–7, for gap-filling of $CO_2$ flux, the ML models performed best in the rice paddy field, followed by the forest, and lastly the grassland ($R^2$ values were around 0.9, 0.8, and 0.6, respectively).

**Table 8.** Summary of regression analysis between measured and ML model predicted fluxes at the three sites. Y = aX + b; Y is measured flux; X is predicted flux.

| | | $CO_2$ Flux | | | Latent Heat Flux | | | Sensible Heat Flux | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Site** | **Model** | **a** | **b** | **$R^2$** | **a** | **b** | **$R^2$** | **a** | **b** | **$R^2$** |
| Grassland | SVM | 1.14 | 0.14 | 0.60 | 0.99 | −0.26 | 0.85 | 1.02 | 0.16 | 0.84 |
| | RF | 1.10 | −0.10 | 0.61 | 1.02 | −1.15 | 0.83 | 1.03 | −0.10 | 0.83 |
| | MLP | 1.00 | 0.05 | 0.59 | 1.01 | −0.33 | 0.85 | 1.01 | 0.12 | 0.82 |
| | DNN | 1.02 | −0.03 | 0.57 | 0.98 | −0.63 | 0.86 | 1.03 | 5.34 | 0.83 |
| | LSTM | 1.07 | 0.41 | 0.60 | 1.02 | −1.86 | 0.86 | 1.05 | 1.33 | 0.83 |
| Rice paddy field | SVM | 1.06 | −0.16 | 0.91 | 0.98 | 1.33 | 0.90 | 1.00 | 0.10 | 0.89 |
| | RF | 1.06 | 0.21 | 0.89 | 0.93 | 3.41 | 0.87 | 1.03 | −0.32 | 0.89 |
| | MLP | 1.03 | −0.02 | 0.90 | 0.91 | 4.43 | 0.89 | 1.00 | 0.02 | 0.87 |
| | DNN | 0.99 | 0.57 | 0.89 | 0.97 | 3.23 | 0.89 | 1.04 | 0.08 | 0.83 |
| | LSTM | 0.94 | 0.08 | 0.89 | 0.92 | 4.11 | 0.89 | 0.99 | 0.61 | 0.89 |
| Forest | SVM | 1.00 | 0.07 | 0.81 | 1.03 | 1.34 | 0.76 | 1.01 | 0.49 | 0.85 |
| | RF | 1.02 | −1.27 | 0.81 | 1.08 | −8.68 | 0.70 | 1.02 | −3.74 | 0.85 |
| | MLP | 0.98 | −0.02 | 0.80 | 1.00 | −0.99 | 0.71 | 0.99 | −0.03 | 0.85 |
| | DNN | 1.00 | 0.01 | 0.80 | 1.07 | 2.50 | 0.69 | 1.02 | −7.29 | 0.83 |
| | LSTM | 0.98 | −0.03 | 0.81 | 1.00 | 1.86 | 0.72 | 1.11 | −0.34 | 0.84 |

### 4.2.2. Latent Heat Flux

Figure 3 presents a typical time series for latent heat flux obtained from the experimental measurements and predictions by the five ML models. As Figure 3a,c show, the five models produced similar results for flux gap-filling at the grassland and forest ecosystems. For the rice paddy field (Figure 3b), the DNN could underestimate the day time peak values, while the other four models resulted in similar predictions.

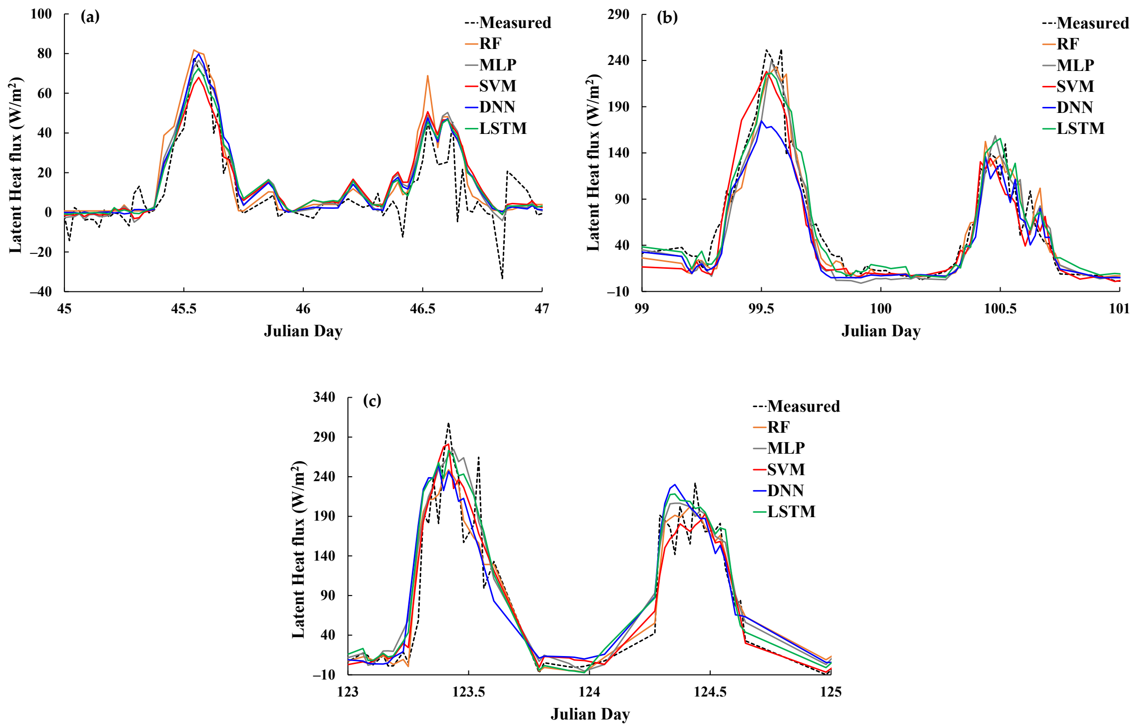



**Figure 3.** Typical time series of latent heat flux obtained from the measurements and the five ML algorithms: (**a**) grassland, (**b**) rice paddy field, and (**c**) forest.

The regression analyses between simulated and measured latent heat fluxes are listed in Table 8. From Table 8 and Appendix C, though the differences between the five models' performances were marginal, for both the rice paddy field and forest ecosystem, SVM was the best model; for the grassland, the model with the best accuracy was LSTM. Moreover, from Tables 5–7, for gap-filling of LE, the ML models performed well over the grassland, rice paddy field, and forest, where $R^2$ values were 0.85, 0.89, and 0.71, respectively.

#### 4.2.3. Sensible Heat Flux

A typical time series of sensible heat flux gap-filling from the five ML models is shown in Figure 4. It is noticed that the five models produced similar results for all three sites and no model was found to systematically over- or underestimate the flux peak values.

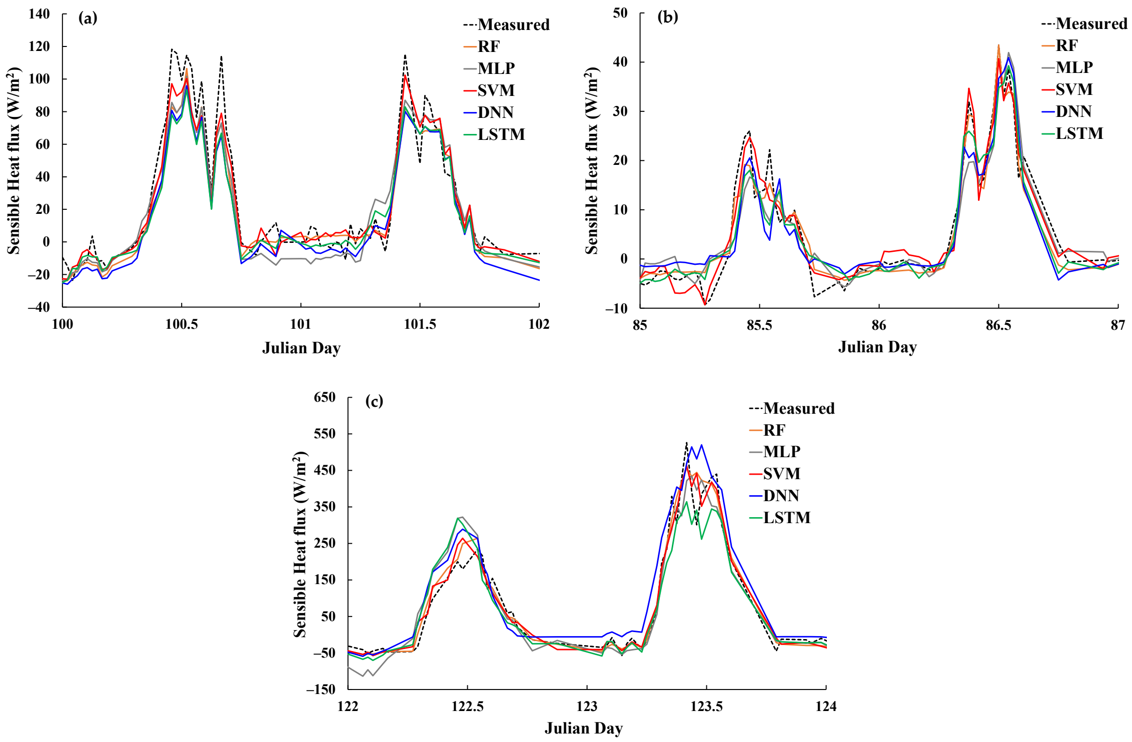



**Figure 4.** Typical time series of sensible heat flux obtained from the measurements and the five ML algorithms: (**a**) grassland, (**b**) rice paddy field, and (**c**) forest.

The regression analyses between measured and ML model predicted sensible heat fluxes and are also summarized in Table 8. Based on Table 8 and Appendix C, the performance of SVM was the best for all the three sites, though the differences between the five models were quite small. Moreover, on the basis of Tables 5–7, for gap-filling of H, the ML models performed well over the grassland, rice paddy field, and forest, where $R^2$ values were 0.83, 0.87, and 0.84, respectively. To summarize, we list the best ML models for gap-filling $CO_2$ flux, LE, and H at the three sites in Table 9.

**Table 9.** Summary of the best ML model for gap-filling of $CO_2$, water vapor, and sensible heat fluxes at the grassland, rice paddy field, and forest.

| Site | $CO_2$ Flux | Latent Heat Flux | Sensible Heat Flux |
|---|---|---|---|
| Grassland | RF | LSTM | SVM |
| Rice paddy field | SVM | SVM | SVM |
| Forest | SVM | SVM | SVM |

### 4.3. Comparison between Machine Learning Model and Penman–Monteith Equation

The comparisons between gap-filling results from the ML model and physical-process based model (P–M equation) over the three ecosystems are presented in this section. According to Section 4.2, the ML algorithm used in the rice paddy field and forest ecosystem was SVM; for the grassland ecosystem, LSTM was adopted.

Figure 5 shows the comparison between measured and simulated latent heat flux by LSTM and the Penman–Monteith equation in the grassland ecosystem. The regression analyses for Figure 5 and model performance metrics are also summarized in Table 10. From Figure 5 and Table 10, we noticed that both the data-driven model (LSTM) and process-based model (P–M equation) performed well (both $R^2$ values greater than 0.82) for gap-filling LE at the grassland site. Moreover, it is noticed that, for the peak values of LE, the LSTM predictions were closer to the measurements than the P–M equation. As shown through the RMSE and MAE, the error of peak value and the average error obtained by LSTM were both smaller than those by the P–M equation.

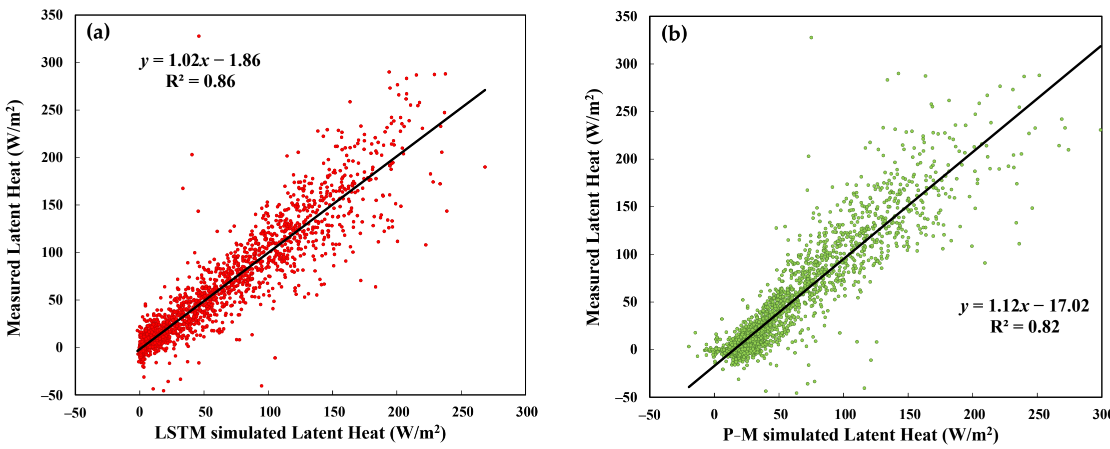



**Figure 5.** Comparison between measured and simulated latent heat flux by (**a**) LSTM and (**b**) the Penman–Monteith equation in the grassland ecosystem.

**Table 10.** Summary of model performance for predicting latent heat flux at the three ecosystems by the machine learning model and Penman–Monteith equation. Y = aX + b; Y is measured flux; X is predicted flux.

| Site | Model | RMSE ($W/m^2$) | MAE ($W/m^2$) | $R^2$ | CE | a (Slope) | b (Intercept) |
|---|---|---|---|---|---|---|---|
| Grassland | LSTM | 22.76 | 12.99 | 0.86 | 0.86 | 1.02 | −1.23 |
| | P–M equation | 28.19 | 20.33 | 0.82 | 0.78 | 1.12 | −17.02 |
| Rice paddy field | SVM | 17.41 | 11.92 | 0.90 | 0.90 | 0.98 | 1.33 |
| | P–M equation | 13.32 | 9.50 | 0.95 | 0.94 | 1.07 | −8.51 |
| Forest | SVM | 50.90 | 35.09 | 0.76 | 0.73 | 1.03 | 1.34 |
| | P–M equation | 67.73 | 46.27 | 0.59 | 0.54 | 0.86 | 28.95 |

The comparison between measured and simulated latent heat flux by SVM and the Penman–Monteith equation at the rice paddy field is plotted in Figure 6. The regression analyses for Figure 6 and model performance metrics are also listed in Table 10. According to Figure 6 and Table 10, it is noticed that both SVM and the P–M equation could accurately reproduce LE (both $R^2$ values greater than 0.9) at the rice paddy field. Because of the lack of measurements of soil heat flux and water heat storage at the rice paddy field (the field was flooded with water), the available energy ($Q_n$) was calculated by H+LE with the assumption of energy balance, and applied to Equation (A1) for calculating $r_{st}$. The outstanding performance of the P–M equation benefits from this forced energy balance. This result implies that, if the surface energy balance of the experimental site is satisfied (the assumption of the P–M equation), then the P–M equation can effectively gap-fill the LE flux. In addition to the energy conservation assumption, the hysteresis between $R_n$ and LE is another factor to influence the accuracy of the P–M equation (but this hysteresis factor is not considered in the P–M equation). Though this hysteresis in magnitude is stronger in the rice paddy field than the forest, the outstanding performance of the P–M equation in the rice paddy field implies that the energy conservation assumption might be a key factor in this field.

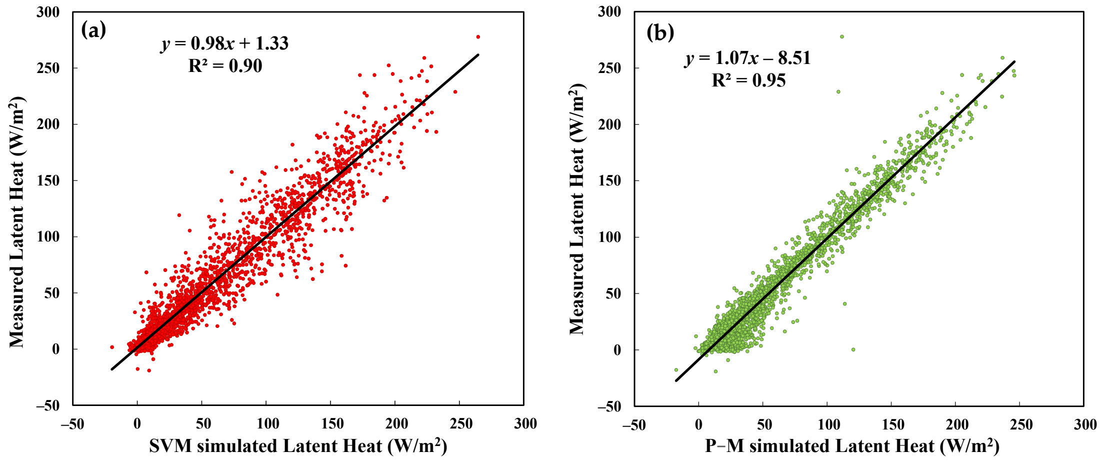



**Figure 6.** Comparison between measured and simulated latent heat flux by (**a**) SVM and (**b**) the Penman–Monteith equation in the rice paddy field.

Figure 7 shows the comparison between measured and predicted LE by SVM and the P–M equation at the forest. The regression analyses for Figure 7 and model performance metrics are also summarized in Table 10. It is obvious that, at the forest site, the ML model was better than the process-based model. The reason that the P–M equation did not reproduce LE well might be because the half-hourly $r_{st}$ varied strongly with the environmental factors (e.g., $R_n$, air temperature) and could not be predicted by the process in Appendix B. For the forest ecosystem, the two methods were both less accurate than the previous two ecosystems. Especially for the P–M equation, the $R^2$ and CE were 0.59 and 0.54, showing the simulated values were only moderately correlated with the measured values. (RMSE and MAE were 67.73 W/m$^2$ and 46.27 W/m$^2$, respectively). Concerning SVM, the $R^2$ and CE were 0.76 and 0.73, respectively. The RMSE and MAE were 50.90 W/m$^2$ and 35.09 W/m$^2$, respectively, which are significantly better than those by the P–M equation.

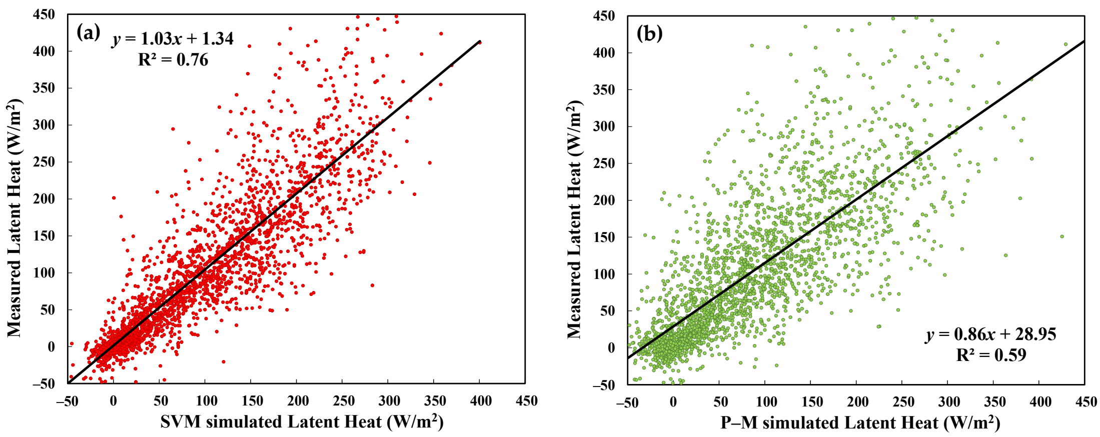



**Figure 7.** Comparison between measured and simulated latent heat flux by (**a**) SVM and (**b**) the Penman–Monteith equation in the forest ecosystem.

Many factors (e.g., energy conservation ratio, hysteresis, $r_{st}$) can affect the accuracy of the P–M equation. The accuracy of the process-based model (P–M equation) at the three ecosystems clearly reflects the environmental and physiological uncertainties at the site. On the other hand, the ML models do not suffer from these environmental and physiological factors. However, when using ML models to extract features from historical data, it might not perform well when the missing data exceed the range of the training data.

### 4.4. Effect of Data Length on Flux Gap-Filling

In this section, we present the relation between gap length and data length of training for the three fluxes at the three sites. The ML model adopted here is SVM, with the optimal input combinations listed in Tables 2–4. Figure 8a shows the RMSE of the model predicted LE as a function of training data length for various gap lengths (one hour, half day, one day, and one week) at the grassland site. The curve for one-hour gap length (blue line) reveals that RMSE was reduced from the highest value (≈27.4 W/m$^2$) to a local minimum (≈17 W/m$^2$) as the training data length increased from 20 h to around 280 h. Once the data length was more than 280 h, the RMSE oscillated and increased to around 19.5 W/m$^2$ at 900 h data length. After 900 h, the RMSE was reduced again and reached a relative stable value (≈17 W/m$^2$) when the data length was longer than 1300 h. This result shows that a longer training data length does not promise better performance for the one-hour gap length.

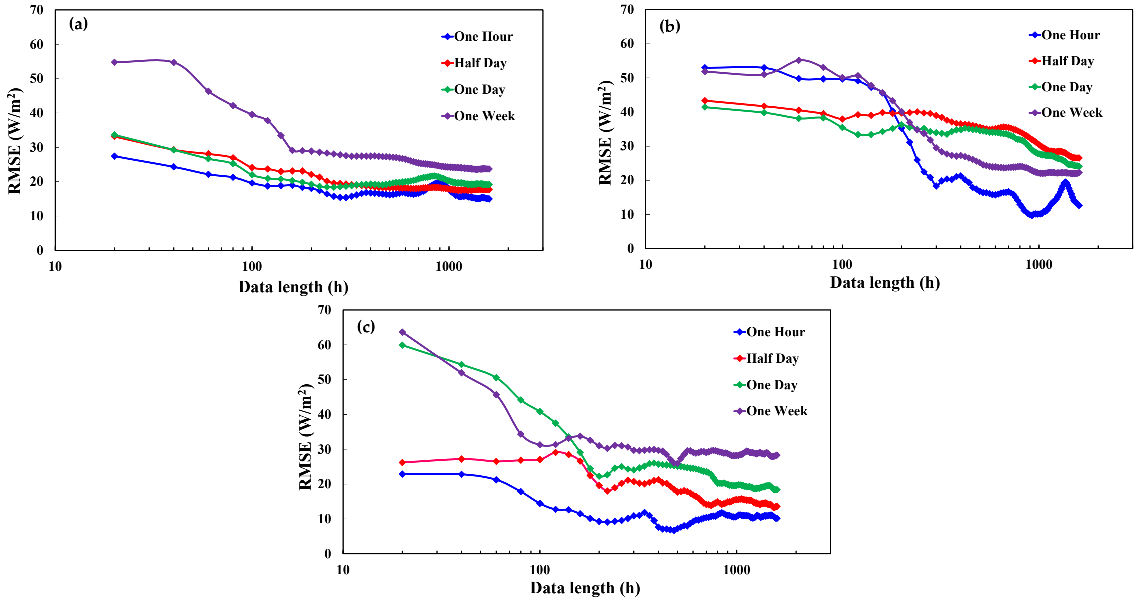



**Figure 8.** The root mean square error (RMSE) of simulated latent heat flux from SVM under different lengths of training data for (**a**) grassland, (**b**) rice paddy field, and (**c**) forest ecosystem.

In Figure 8a, the RMSE curve for one day (green line) had the same trend as the one hour curve did, but with slightly higher RMSE values. Concerning the curve for the half day gap length (red line), the RMSE decreased with increasing training data length from 20 h to the maximum point (1600 h), where after 400 h, the decrease in RMSE was relatively small. For the one week gap length case, the RMSE curve (purple line) decreased rapidly with the increase of data length and reached a stable RMSE value after 1000 h.

From these four curves, it is noticed that, when using the ML algorithm to gap-fill the missing LE at the grassland ecosystem, the error was larger when the length of training data was short. As the data length increased, the accuracy also increased, and then gradually converged to the model limitation; that is, the RMSE no longer decreased with an increase of data length. For the case herein, it is found that, for gaps less than one day, using about 400 h training data could result in a relatively small RMSE (less than 25 W/m$^2$).

Figure 8b is the same as Figure 8a, but for the rice paddy field. The major trend in Figure 8b is the same as in Figure 8a, that is, the RMSE curves (for all four gap lengths) had the highest values at the beginning and then dropped as the data length increased. For the one hour gap length case, the RMSE had its minimum ($\approx$10 W/m$^2$) at 900 h and then oscillated up and down with data length till the end at 1600 h. For cases of half day and one day, the minimum values ($\approx$25 W/m$^2$) of RMSE happened at 1600 h, but the RMSEs only changed a bit after 1300 h. For the one week case, the RMSE had its lowest value ($\approx$22 W/m$^2$) at 1000 h and remained stable afterwards.

Figure 8c is the same as Figure 8a, but for the forest ecosystem. For the one hour gap length case, the RMSE had its minimum ($\approx$8 W/m$^2$) at 500 h and then increased to 11 W/m$^2$ with the increase of data length till 800 h, and then remained stable till the end at 1600 h, with an RMSE around 11 W/m$^2$. Same as that in Figure 8b, for the one week gap length case, the RMSE also had its lowest value ($\approx$31 W/m$^2$) at 1000 h and remained stable afterwards. For the cases of half day and one day, the minimum values of RMSE occurred at 1300 h and the RMSE remained stable afterwards.

In summary, Figure 8 reveals the following characteristics:

(1) The gap-filling accuracy increased with the increase of data length and reached the model limitation when the data length is longer than 1300 h, except for the one hour gap length case.
(2) For cases of one hour, the best model performance happened at different data lengths for different ecosystems (around 300, 900, and 500 h for grassland, rice paddy, and forest, respectively).

Figure 9 plots the RMSE of gap-filling for $CO_2$ flux by the SVM at the three ecosystems as a function of data length for various gap lengths. In Figure 9, the following issues are noticed.

(1) For all three sites, the RMSE curve of half day and one day gap lengths had the highest value at the beginning and then dropped as the data length increased; after a local RMSE minimum was reached, the RMSE oscillated up and down with the increase of data length and then reached its minimum at around 1300 h and remained stable till the end at 1600 h. The one week gap length case at the grassland also followed this trend.
(2) For the one week gap length case at both the rice paddy field and forest, the RMSE decreased with the increase of data length and had the minimum after 1300 h.
(3) For the one hour gap length case, the RMSE curve trend differed from each ecosystem. The minimum RMSE occurred at around 1000, 200, and 300 h for the grassland, rice paddy field, and forest, respectively.
(4) Figure 9c shows that too much training data could result in a decrease in gap-filling accuracy for the short gap length cases (one hour, half day, and one day). This is because too much training data might average out the necessary features (e.g., peak values) of a short period.

Figure 10 is the same as Figure 9, but for sensible heat flux. Figure 10b shows that the variations of all four RMSE curves at the rice paddy field were relative stable with the increase of data length. This is because the sensible heat flux measurements were small at this site. In Figure 10a,b, for both the one day and one week cases at both the grassland and forest ecosystems, the RMSE decreased with the increase of data length and had the minimum after 1300 h. However, the RMSE curves for the one hour and half day cases at these two sites oscillated a bit after the local minimum occurred.

In summary, (1) for the one week gap length case, the RMSE value decreased with the increase of the data length, and had the minimum value around 1300 h; (2) for both half day and one day cases, the RMSE value sometimes oscillated with the increase of the data length, but it would converge and reach the lowest value around 1300 h; (3) in the case of one hour, using the longest data length does not guarantee the best predicted value, and sometimes, the lowest RMSE value occurs in a shorter data length. Here, the shortest data length required to obtain an RMSE less than 1.05 times the lowest RMSE is summarized in Table 11.

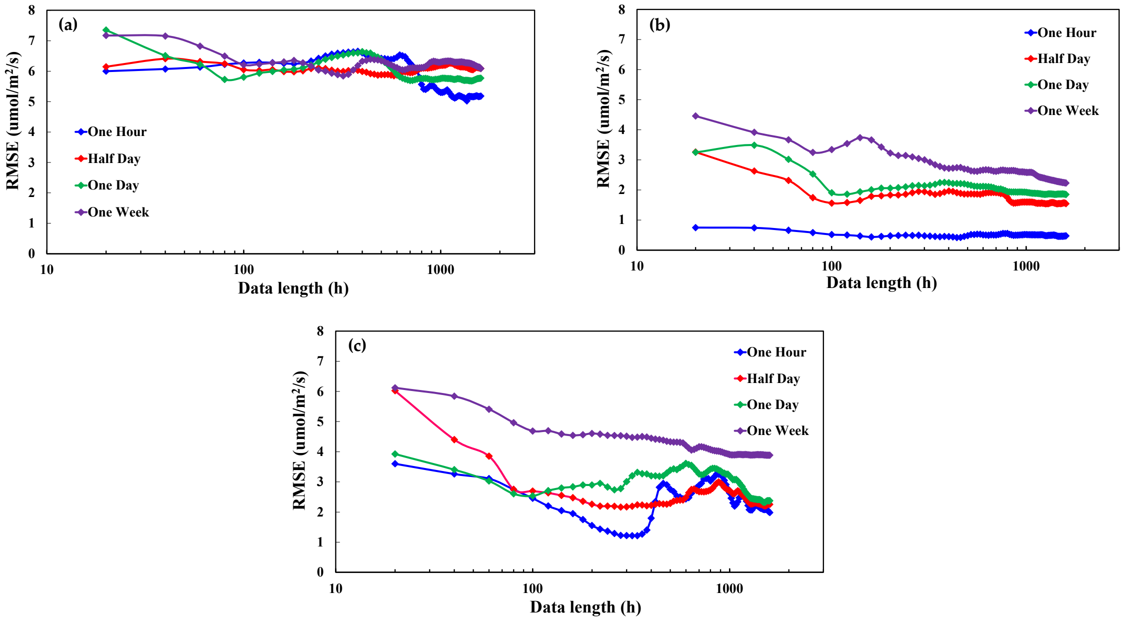



**Figure 9.** The RMSE of simulated $CO_2$ flux from SVM under different lengths of training data for the (**a**) grassland, (**b**) rice paddy field, and (**c**) forest ecosystem.

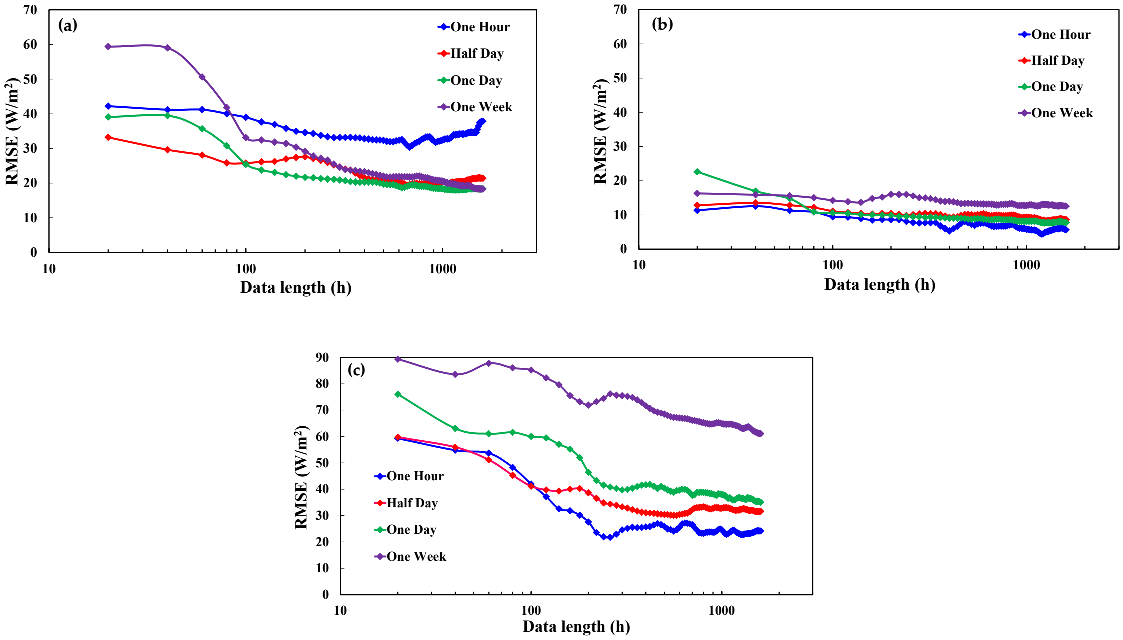



**Figure 10.** The RMSE of simulated sensible heat flux from SVM under different lengths of training data for the (**a**) grassland, (**b**) rice paddy field, and (**c**) forest ecosystem.

**Table 11.** The shortest data length required to obtain an RMSE less than 1.05 times the lowest RMSE for various gap lengths of $CO_2$, latent heat, and sensible heat fluxes at the three sites.

| | LE | | | | $CO_2$ | | | | H | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Site | One Hour | Half Day | One Day | One Week | One Hour | Half Day | One Day | One Week | One hour | Half Day | One Day | One Week |
| Grassland | 280 | 440 | 200 | 860 | 1120 | 100 | 80 | 240 | 640 | 620 | 620 | 1280 |
| Rice paddy field | 880 | 1380 | 1380 | 920 | 160 | 100 | 100 | 1340 | 1180 | 1180 | 1180 | 580 |
| Forest | 460 | 740 | 1120 | 460 | 280 | 200 | 1340 | 640 | 240 | 380 | 1080 | 1200 |

The above analyses are for individual gap cases; it is also interesting to know the relation between total gap length and data length of training where the total length of the gaps is the same (but the individual gap length is different). The analysis is presented in Appendix E, and it shows that, when the total gap length was the same, the RMSE curves of the four gaps were quite similar (regardless of the length of each gap) and all of them decreased with the increase of data length and converged to the lowest values after 1300 h.

## 5. Conclusions

In this study, five ML algorithms were adopted, including three conventional algorithms (SVM, RF, and MLP) and two deep learning algorithms, (DNN and LSTM) to gap-fill the missing fluxes of $CO_2$, water vapor, and sensible heat in three ecosystems (grassland, rice paddy field, and forest). We conclude the following.

(1) In addition to the mean meteorological parameters, including the time factors (i.e., Julian day and decimal time) is important for all fluxes of $CO_2$, water vapor, and sensible heat at the rice paddy field. However, the influences of time factors on these three fluxes are small (less than 5%) at the grassland. For the forest, this influence on $CO_2$ flux is small, but it is larger for water vapor and sensible heat fluxes.
(2) The hysteresis factors have no influence on all three fluxes at the grassland site. For the rice paddy, this influence on $CO_2$ flux is not important, but it is important for water vapor and sensible heat fluxes. For the forest site, the hysteresis influence is important on $CO_2$ flux, but it is small on both water vapor and sensible heat fluxes.
(3) For all three ecosystems, the five ML models produced similar results for gap-filling of $CO_2$, water vapor, and sensible heat fluxes. A list of the best ML model for flux gap-filling at the three sites is provided in Table 9. All in all, the SVM model is the most recommended model.
(4) In terms of water vapor flux gap-filling, the ML model was better than the P–M equation, especially for forests; however, historical data are required a priori for training ML models.
(5) The following general rule for the relation between gap length and data length of training can be made: if the gap length is less than one week, the training data length for achieving the best model performance is around 1300 h (i.e., 7.7 times the gap length).
(6) For a particular gap that we are concerned about (especially where the flux peak values occurred), if training data length longer than 1300 h are not available when doing gap-filling, the data length listed in Table 11 is recommended.

**Author Contributions:** C.-IH. conceived the research idea; I-H.H. performed the simulations of ML models; C.-IH. and I-H.H. performed the simulations of the P–M equation; C.-IH. conducted the paddy rice field experiment. I-H.H. and C.-IH. took part in the discussion, analysis, and interpretation of the data and model predictions; I-H.H. and C.-IH. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** The authors are grateful to Yu-Li Wang for preparing the data for this work. This work was supported by the Core Research Project, National Taiwan University (project numbers: NTU-CC-109L892803; NTU-CC-109L203313).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest.

## Appendix A. Brief Introduction of the Grid-Search Method

The grid-search method [26] is one of the most detailed and comprehensive algorithms used to find the best combination of the hyper-parameters. The principle of it is to evaluate the performance metrics of models built by all parameter combinations in the feasible solution space and find the optimal one. The process of the grid-search method can be divided into the following three steps: (1) Set reasonable upper and lower boundaries of each parameter; these boundaries will determine the feasible solution space. (2) Set the grid size according to the user needs of accuracy and hardware efficiency; the upper and lower boundaries and the grid size will determine the number of parameter combinations to be calibrated. (3) Establish models using all the parameter combinations separately to obtain the performance metrics, and select the best one as the optimal parameter combination. More details about this method can be found in Lerman [26].

## Appendix B. Calculation of the Stomatal Resistance (rst)

When applying the P–M equation to do latent heat flux (LE) gap-filling, in addition to the general meteorological parameters, $R_n$, $G_s$, $T_a$, RH, and U, the stomatal resistance $r_{st}$ should also be given. In this study, when there is a missing point at time t, the $r_{st}$ for this point is the average of the $r_{st}$ at t − 1 and t + 1 (each time period is 30 min). By the measured LE and rearranging Equation (1), the $r_{st}$ was calculated as

$$r_{st} = \left(\frac{\Delta}{\gamma}\frac{Q_n - \mathrm{LE}}{\mathrm{LE}} - 1\right) r_{av} + \left(\frac{\rho c_p}{\gamma}\frac{D}{\mathrm{LE}}\right) \tag{A1}$$

If there was an outlier in the measured $r_{st}$ at t − 1 and t + 1, the following rules were applied to replace the outlier: firstly, an $r_{st}$ greater than 800 was replaced by 800; secondly, an $r_{st}$ less than 0 was replaced by 0. If the previous or next $r_{st}$ was missing, or if the data were the first or the last point of the data set, the long-term average of $r_{st}$ was used as a replacement. That is, the diurnal (07:00~17:00) missing $r_{st}$ was supplemented by the average of all $r_{st}$ from 10:00 to 14:00. On the other hand, the nocturnal (0:00~07:00 and 17:00~00:00) missing $r_{st}$ would be supplemented by the average of all $r_{st}$ during night time.

For the grassland, the measured data were divided into two seasons, the hot period (from May to October) and cold period (from November to April). In the hot season, the average $r_{st}$ for diurnal and nocturnal periods were 358.86 and 503.01 (s/m), respectively. In the cold period, the average $r_{st}$ for diurnal and nocturnal periods were 273.81 and 422.43 (s/m), respectively

For the rice paddy field, the average $r_{st}$ for diurnal and nocturnal periods in the growing season were 321.35 and 361.41 (s/m), respectively.

For the forest ecosystem, similar to the grassland site, the whole dataset was divided into hot and cold seasons. The average $r_{st}$ for diurnal period in the hot and cold seasons were 196.63 and 339.33 (s/m), respectively. As the forest canopy was always wet during night time [38], the missing $r_{st}$ at night for both hot and cold seasons were filled in by 0 (s/m).

## Appendix C. Summary of Individual Model Performance with and without Time Factors and Hysteresis Factors

**Table A1.** Summary of individual model performance with and without time factors and hysteresis factors as inputs for flux gap-filling at the grassland. The number in the first parenthesis is the result without time factors; the number in the second parenthesis is the result without hysteresis factors.

| Flux | Model | RMSE | MAE | $R^2$ | CE |
|---|---|---|---|---|---|
| $CO_2$ | SVM | 4.90 (5.04) (4.90) | 3.02 (3.16) (3.02) | 0.60 (0.58) (0.60) | 0.60 (0.58) (0.60) |
| | RF | 4.86 (5.06) (4.86) | 3.03 (3.22) (3.03) | 0.61 (0.57) (0.61) | 0.61 (0.57) (0.61) |
| | MLP | 4.96 (5.14) (4.96) | 3.19 (3.31) (3.19) | 0.59 (0.56) (0.59) | 0.59 (0.56) (0.59) |
| | DNN | 5.06 (5.16) (5.06) | 3.16 (3.31) (3.16) | 0.57 (0.56) (0.57) | 0.57 (0.55) (0.57) |
| | LSTM | 4.93 (5.10) (4.93) | 3.06 (3.23) (3.06) | 0.60 (0.57) (0.60) | 0.59 (0.56) (0.59) |
| LE | SVM | 23.01 (23.01) (23.01) | 13.23 (13.23) (13.23) | 0.85 (0.85) (0.85) | 0.85 (0.85) (0.85) |
| | RF | 24.88 (25.07) (24.88) | 14.72 (14.88) (14.72) | 0.83 (0.83) (0.83) | 0.83 (0.83) (0.83) |
| | MLP | 23.13 (23.13) (23.13) | 13.24 (13.24) (13.24) | 0.85 (0.85) (0.85) | 0.85 (0.85) (0.85) |
| | DNN | 22.92 (23.23) (22.92) | 13.19 (13.57) (13.19) | 0.86 (0.85) (0.86) | 0.85 (0.85) (0.85) |
| | LSTM | 22.76 (22.76) (22.76) | 12.99 (12.99) (12.99) | 0.86 (0.86) (0.86) | 0.86 (0.86) (0.86) |
| H | SVM | 16.16 (16.28) (16.16) | 10.33 (10.38) (10.33) | 0.84 (0.84) (0.84) | 0.84 (0.84) (0.84) |
| | RF | 16.66 (17.16) (16.66) | 10.54 (10.89) (10.54) | 0.83 (0.82) (0.83) | 0.83 (0.82) (0.83) |
| | MLP | 17.33 (17.35) (17.33) | 11.37 (11.28) (11.37) | 0.82 (0.82) (0.82) | 0.82 (0.82) (0.82) |
| | DNN | 16.87 (17.37) (16.87) | 10.79 (11.35) (10.79) | 0.83 (0.82) (0.83) | 0.83 (0.82) (0.83) |
| | LSTM | 16.63 (16.85) (16.63) | 10.80 (10.90) (10.80) | 0.83 (0.83) (0.83) | 0.83 (0.83) (0.83) |

**Table A2.** Same as Table A1, but for the rice paddy field.

| Flux | Model | RMSE | MAE | $R^2$ | CE |
|---|---|---|---|---|---|
| $CO_2$ | SVM | 2.27 (2.63) (2.27) | 1.70 (1.97) (1.70) | 0.90 (0.87) (0.90) | 0.89 (0.85) (0.89) |
| | RF | 2.55 (2.66) (2.66) | 1.86 (1.99) (1.98) | 0.89 (0.87) (0.87) | 0.86 (0.84) (0.84) |
| | MLP | 2.30 (2.83) (2.30) | 1.71 (2.16) (1.74) | 0.90 (0.86) (0.90) | 0.88 (0.82) (0.88) |
| | DNN | 2.56 (3.99) (2.59) | 1.95 (3.08) (1.99) | 0.89 (0.87) (0.89) | 0.86 (0.64) (0.85) |
| | LSTM | 2.40 (2.94) (2.50) | 1.80 (2.32) (1.87) | 0.89 (0.86) (0.90) | 0.87 (0.81) (0.86) |
| LE | SVM | 17.41 (18.64) (20.69) | 11.92 (12.26) (14.30) | 0.90 (0.89) (0.86) | 0.90 (0.89) (0.86) |
| | RF | 19.85 (21.14) (20.79) | 13.32 (13.90) (14.21) | 0.87 (0.86) (0.86) | 0.87 (0.86) (0.86) |
| | MLP | 18.29 (20.88) (19.95) | 12.74 (14.23) (14.10) | 0.89 (0.86) (0.87) | 0.89 (0.86) (0.87) |
| | DNN | 18.79 (20.88) (20.30) | 13.28 (14.23) (14.37) | 0.89 (0.86) (0.87) | 0.88 (0.86) (0.87) |
| | LSTM | 18.20 (19.70) (20.52) | 12.79 (12.99) (14.50) | 0.89 (0.87) (0.86) | 0.89 (0.87) (0.86) |
| H | SVM | 9.71 (10.92) (11.47) | 6.07 (6.74) (6.80) | 0.89 (0.86) (0.85) | 0.89 (0.86) (0.85) |
| | RF | 9.64 (11.79) (9.70) | 6.06 (7.19) (6.08) | 0.89 (0.84) (0.89) | 0.89 (0.84) (0.89) |
| | MLP | 10.64 (11.93) (11.48) | 6.75 (7.58) (7.21) | 0.87 (0.84) (0.85) | 0.87 (0.83) (0.85) |
| | DNN | 12.05 (13.92) (13.58) | 7.35 (8.84) (8.40) | 0.83 (0.78) (0.79) | 0.83 (0.77) (0.79) |
| | LSTM | 10.00 (11.32) (10.78) | 6.43 (7.01) (6.74) | 0.89 (0.85) (0.87) | 0.88 (0.85) (0.86) |

**Table A3.** Same as Table A1, but for the forest.

| Flux | Model | RMSE | MAE | $R^2$ | CE |
|---|---|---|---|---|---|
| $CO_2$ | SVM | 3.40 (3.52) (3.68) | 2.30 (2.37) (2.52) | 0.81 (0.80) (0.78) | 0.81 (0.80) (0.78) |
| | RF | 3.37 (3.45) (3.71) | 2.25 (2.33) (2.48) | 0.81 (0.81) (0.78) | 0.81 (0.81) (0.78) |
| | MLP | 3.46 (3.61) (3.80) | 2.40 (2.52) (2.66) | 0.80 (0.79) (0.77) | 0.80 (0.79) (0.76) |
| | DNN | 3.52 (3.59) (3.79) | 2.38 (2.50) (2.61) | 0.80 (0.79) (0.77) | 0.80 (0.79) (0.76) |
| | LSTM | 3.38 (3.48) (3.66) | 2.25 (2.38) (2.54) | 0.81 (0.80) (0.78) | 0.81 (0.80) (0.78) |
| LE | SVM | 50.90 (55.73) (52.76) | 35.09 (38.50) (36.07) | 0.73 (0.68) (0.72) | 0.73 (0.67) (0.71) |
| | RF | 53.05 (58.21) (53.21) | 36.42 (41.47) (36.62) | 0.70 (0.64) (0.70) | 0.70 (0.64) (0.70) |
| | MLP | 52.55 (57.32) (52.75) | 37.47 (40.44) (37.79) | 0.71 (0.66) (0.71) | 0.71 (0.65) (0.71) |
| | DNN | 54.06 (57.87) (54.26) | 37.89 (41.21) (38.94) | 0.69 (0.65) (0.69 | 0.69 (0.65) (0.69 |
| | LSTM | 52.04 (55.92) (54.16) | 36.67 (38.95) (36.97) | 0.72 (0.68) (0.69) | 0.72 (0.67) (0.69) |
| H | SVM | 60.42 (65.57) (60.74) | 39.83 (42.54) (39.36) | 0.85 (0.82) (0.85) | 0.85 (0.82) (0.85) |
| | RF | 60.96 (65.59) (61.48) | 40.61 (43.61) (40.71) | 0.85 (0.82) (0.85) | 0.85 (0.82) (0.85) |
| | MLP | 61.03 (64.28) (61.47) | 41.34 (43.67) (41.82) | 0.85 (0.83) (0.84) | 0.85 (0.83) (0.84) |
| | DNN | 64.76 (68.97) (67.82) | 44.82 (46.18) (44.32) | 0.83 (0.80) (0.81) | 0.83 (0.80) (0.81) |
| | LSTM | 61.67 (68.28) (62.72) | 39.95 (44.45) (41.01) | 0.84 (0.81) (0.84) | 0.84 (0.81) (0.84) |

## Appendix D. Model Performance with and without Leaf Area Index (LAI)

In this appendix, we present the model performance with and without leaf area index (LAI) as an input for flux gap-filling at the rice paddy field (Table A4). The ML model adopted here is the support vector machine (SVM). It is noticed that the influences of LAI on gap-filling of the three surface fluxes at the rice paddy field are all small.

**Table A4.** Summary of model performance with and without leaf area index (LAI) as an input for flux gap-filling at the paddy rice field. The model adopted here is the support vector machine (SVM) with the optimal input combination listed in Tables 2–4.

| Flux | Model | RMSE | MAE | $R^2$ | CE |
|---|---|---|---|---|---|
| $CO_2$ Flux | SVM | 2.27 | 1.70 | 0.90 | 0.89 |
| (umol/m$^2$/s) | SVM with LAI | 2.54 | 1.87 | 0.90 | 0.86 |
| Latent Heat Flux | SVM | 17.41 | 11.92 | 0.90 | 0.90 |
| (W/m$^2$) | SVM with LAI | 16.29 | 10.78 | 0.91 | 0.91 |
| Sensible Heat Flux | SVM | 9.71 | 6.07 | 0.89 | 0.89 |
| (W/m$^2$) | SVM with LAI | 9.75 | 6.10 | 0.89 | 0.89 |

**Appendix E. Gap Scenario with Equal Total Gap Length**

Here, we present the relation between total gap length and data length of training where the total length of the gaps is the same (but the individual gap length is different). The gap scenario is as follows. (1) The total gap length was one week (=168 h). (2) Four individual gap lengths were considered: one hour, half day, one day, and one week. That is, if the gap-length of each gap is a half day (=12 h), there will be 14 gaps in this case. (3) The highest point of the measured flux was selected as the central point of the first gap because this point was the most difficult point for gap-filling. (4) The remaining gaps were randomly selected from the rest of the data with a uniform distribution. (5) The closest neighbor points to the first gap were selected as the training data. (6) The training data length started from 20 h and increased gradually to a maximum of 1600 h.

Figure A1 plots the RMSE of the model predicted $CO_2$ flux as a function of training data length for four individual gap lengths (one hour, half day, one day, and one week) at the forest site. In Figure A1, the total gap length for the four individual gap lengths was the same (one week) and the $CO_2$ flux was predicted using the SVM model. Figure A1 shows that, when the total gap length was the same, the RMSE curves of the four different gaps were quite similar, and all of them decreased with the increase of data length and converged to the lowest values after 1300 h. In other words, if the total gap length is one week, the length of training data for achieving better model performance is around 1300 h, regardless of the length of each gap.

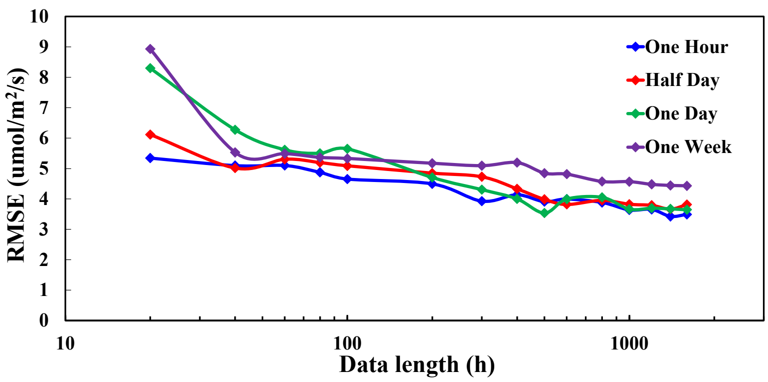



**Figure A1.** The RMSE of predicted $CO_2$ flux at the forest site as a function of training data length where the total gap length is one week for the four individual gap lengths: one hour, half day, one day, and one week.

## References

1. Falge, E.; Baldocchi, D.; Olson, R.J.; Anthoni, P.; Aubinet, M.; Bernhofer, C.; Burba, G.; Ceulemans, R.; Clement, R.; Dolman, H.; et al. Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agric. For. Meteorol.* **2001**, *107*, 43–69. [CrossRef]
2. Moffat, A.M.; Papale, D.; Reichstein, M.; Hollinger, D.Y.; Richardson, A.D.; Barr, A.G.; Beckstein, C.; Braswell, B.H.; Churkina, G.; Desai, A.R.; et al. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agric. For. Meteorol.* **2007**, *147*, 209–232. [CrossRef]
3. Barr, A.G.; Black, T.A.; Hogg, E.H.; Kljun, N.; Morgenstern, K.; Nesic, Z. Inter-annual variability in the leaf area index of a boreal aspen-hazelnut forest in relation to net ecosystem production. *Agric. For. Meteorol.* **2004**, *126*, 237–255. [CrossRef]
4. Desai, A.R.; Bolstad, P.; Cook, B.D.; Davis, K.J.; Carey, E.V. Comparing net ecosystem exchange of carbon dioxide between an old-growth and mature forest in the upper Midwest, USA. *Agric. For. Meteorol.* **2005**, *128*, 33–55. [CrossRef]
5. Hollinger, D.Y.; Aber, J.; Dail, B.; Davidson, E.A.; Goltz, S.M.; Hughes, H.; Leclerc, M.; Lee, J.T.; Richardson, A.D.; Rodrigues, C.; et al. Spatial and temporal variability in forest-atmosphere $CO_2$ exchange. *Glob. Chang. Biol.* **2004**, *10*, 1689–1706. [CrossRef]

6. Noormets, A.; Chen, J.; Crow, T.R. Age-dependent changes in ecosystem carbon fluxes in managed forests in northern Wisconsin, USA. *Ecosystems* **2007**, *10*, 187–203. [CrossRef]
7. Richardson, A.D.; Braswell, B.H.; Hollinger, D.Y.; Burman, P.; Davidson, E.A.; Evans, R.S.; Flanagan, L.B.; Munger, J.W.; Savage, K.; Urbanski, S.P.; et al. Comparing simple respiration models for eddy flux and dynamic chamber data. *Agric. For. Meteorol.* **2006**, *141*, 219–234. [CrossRef]
8. Richardson, A.D.; Hollinger, D.Y. Statistical modeling of ecosystem respiration using eddy covariance data: Maximum likelihood parameter estimation, and Monte Carlo simulation of model and parameter uncertainty, applied to three simple models. *Agric. For. Meteorol.* **2005**, *131*, 191–208. [CrossRef]
9. Stauch, V.J.; Jarvis, A.J. A semi-parametric model for eddy covariance $CO_2$ flux time series data. *Glob. Chang. Biol.* **2006**, *12*, 1707–1716. [CrossRef]
10. Hui, D.; Wan, S.; Su, B.; Katul, G.; Monson, R.; Luo, Y. Gap-filling missing data in eddy covariance measurements using multiple imputation (MI) for annual estimations. *Agric. For. Meteorol.* **2004**, *121*, 93–111. [CrossRef]
11. Du, Q.; Liu, H.; Feng, J.; Wang, L. Effects of different gap filling methods and land surface energy balance closure on annual net ecosystem exchange in a semiarid area of China. *Sci. China Earth Sci.* **2014**, *57*, 1340–1351. [CrossRef]
12. Van Wijk, M.T.; Bouten, W. Water and carbon fluxes above European coniferous forests modelled with artificial neural networks. *Ecol. Model.* **1999**, *120*, 181–197. [CrossRef]
13. Carrara, A.; Kowalski, A.S.; Neirynck, J.; Janssens, I.A.; Yuste, J.C.; Ceulemans, R. Net ecosystem $CO_2$ exchange of mixed forest in Belgium over 5 years. *Agric. For. Meteorol.* **2003**, *119*, 209–227. [CrossRef]
14. Schmidt, A.; Wrzesinsky, T.; Klemm, O. Gap Filling and Quality Assessment of $CO_2$ and Water Vapour Fluxes above an Urban Area with Radial Basis Function Neural Networks. *Bound.-Layer Meteorol.* **2008**, *126*, 389–413. [CrossRef]
15. Kordowski, K.; Kuttler, W. Carbon dioxide fluxes over an urban park area. *Atmos. Environ.* **2010**, *44*, 2722–2730. [CrossRef]
16. Papale, D.; Valentini, R. A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. *Glob. Chang. Biol.* **2003**, *9*, 525–535. [CrossRef]
17. Van Wijk, M.T.; Bouten, W.; Verstraten, J.M. Comparison of different modelling strategies for simulating gas exchange of a Douglas-fir forest. *Ecol. Model.* **2002**, *158*, 63–81. [CrossRef]
18. Dengel, S.; Zona, D.; Sachs, T.; Aurela, M.; Jammet, M.; Parmentier, F.J.W.; Oechel, W.; Vesala, T. Testing the applicability of neural networks as a gap-filling method using $CH_4$ flux data from high latitude wetlands. *Biogeosciences* **2013**, *10*, 8185–8200. [CrossRef]
19. Nguyen, P.; Halem, M. *Deep Learning Models for Predicting $CO_2$ Flux Employing Multivariate Time Series*; Mile TS: Anchorage, AK, USA, 2019.
20. Kim, Y.; Johnson, M.S.; Knox, S.H.; Black, T.A.; Dalmagro, H.J.; Kang, M.; Kim, J.; Baldocchi, D. Gap-filling approaches for eddy covariance methane fluxes: A comparison of three machine learning algorithms and a traditional method with principal component analysis. *Glob. Chang. Biol.* **2020**, *26*, 1499–1518. [CrossRef]
21. Kang, M.; Ichii, K.; Kim, J.; Indrawati, Y.M.; Park, J.; Moon, M.; Lim, J.H.; Chun, J.H. New Gap-Filling Strategies for Long-Period Flux Data Gaps Using a Data-Driven Approach. *Atmosphere* **2019**, *10*, 568. [CrossRef]
22. Jaksic, V.; Kiely, G.; Albertson, J.; Oren, R.; Katul, G.; Leahy, P.; Byrne, K.A. Net ecosystem exchange of grassland in contrasting wet and dry years. *Agric. For. Meteorol.* **2006**, *139*, 323–334. [CrossRef]
23. Hsieh, C.I.; Kiely, G.; Birkby, A.; Katul, G. Photosynthetic responses of a humid grassland ecosystem to future climate perturbations. *Adv. Water Resour.* **2005**, *28*, 910–916. [CrossRef]
24. Aubinet, M.; Grelle, A.; Ibrom, A.; Rannik, Ü.; Moncrieff, J.; Foken, T.; Kowalski, A.S.; Marin, P.H.; Berbigier, P.; Bernhofer, C.; et al. Estimates of the annual net carbon and water exchange of forests: The EUROFLUX methodology. *Adv. Ecol. Res.* **2000**, *30*, 113–175. [CrossRef]
25. Chu, H.S.; Chang, S.C.; Klemm, O.; Lai, C.W.; Lin, Y.Z.; Wu, C.C.; Lin, J.Y.; Jiang, J.Y.; Chen, J.; Gottgens, J.F.; et al. Does canopy wetness matter? Evapotranspiration from a subtropical montane cloud forest in Taiwan. *Hydrol. Process.* **2012**, *28*. [CrossRef]
26. Lerman, P.M. Fitting segmented regression models by grid search. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **1980**, *29*, 77–84. [CrossRef]
27. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]

28. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
29. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
30. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
31. Baldi, P.; Sadowski, P.J. Understanding dropout. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2013; pp. 2814–2822.
32. Hinton, G.E.; Osindero, S.; Teh, Y.W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [CrossRef]
33. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
34. Parascandolo, G.; Huttunen, H.; Virtanen, T. Recurrent neural networks for polyphonic sound event detection in real life recordings. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 6440–6444. [CrossRef]
35. Yang, Y.; Dong, J.; Sun, X.; Lima, E.; Mu, Q.; Wang, X. A CFCC-LSTM model for sea surface temperature prediction. *IEEE Geosci. Remote Sens. Lett.* **2017**, *15*, 207–211. [CrossRef]
36. Irmak, S.; Mutiibwa, D. On the dynamics of canopy resistance: Generalized linear estimation and relationships with primary micrometeorological variables. *Water Resour. Res.* **2010**, *46*. [CrossRef]
37. Cui, Y.; Liu, Y.; Gan, G.; Wang, R. Hysteresis behavior of surface water fluxes in a hydrologic transition of an ephemeral Lake. *J. Geophys. Res. Atmos.* **2020**, *125*, e2019JD032364. [CrossRef]
38. Lin, B.S.; Lei, H.; Hu, M.C.; Visessri, S.; Hsieh, C.I. Canopy Resistance and Estimation of Evapotranspiration above a Humid Cypress Forest. *Adv. Meteorol.* **2020**, *2020*, 4232138. [CrossRef]
39. Zhang, Q.; Manzoni, S.; Katul, G.; Porporato, A.; Yang, D. The hysteretic evapotranspiration—vapor pressure deficit relation. *J. Geophys. Res. Biogeosciences* **2014**, *119*, 125–140. [CrossRef]