A Transformer-CNN Imputation Model for Turbulent Heat Fluxes in the Tibetan Plateau Grassland

Quanzhe Hou[1], Zhiqiu Gao[1,2,3], Zexia Duan[4], and Minghui Yu[1]

[1]School of Atmospheric Physics, Nanjing University of Information Science and Technology, Nanjing 210044, China
[2]State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China
[3] Geovis Environment Technology Co. Ltd., Nanjing, 211802, China
[4]School of Electrical Engineering, Nantong University, Nantong 226019, China

Correspondence to:Dr.Zhiqiu Gao (zgao@nuist.edu.cn)

Abstract. Based on the turbulent heat flux from the third scientific expedition to the Tibetan Plateau in 2012, imputation evaluations were conducted using algorithms like mean diurnal variation (MDV), nonlinear regression(NR), and look up tables(LUT), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Gradient Boosting (XGBoost), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and the Transformer model with deep self-attention mechanism. Results indicated that the Transformer model performed optimally. To further enhance imputation accuracy, a combined model of Transformer and Convolutional Neural Network (CNN), termed as Transformer_CNN, was proposed. Herein, while the Transformer primarily focused on global attention, the convolution operations in the CNN provided the model with local attention. Experimental outcomes revealed that the imputations from Transformer_CNN surpassed the traditional single artificial intelligence model approaches. The coefficient of determination (R^2) reached 0.949 in the sensible heat flux test set and 0.894 in the latent heat flux test set, thereby confirming the applicability of the Transformer_CNN model for data imputation of turbulent heat flux in the Tibetan Plateau. Ultimately, the turbulent heat flux observational database from 2007 to 2016 at the station was imputed using the Transformer_CNN model.

1 Introduction

Tibetan land surface processes play a significant role in influencing Asian weather and climate, primarily through the surface-atmosphere exchange of energy, momentum, and CO_2 across the atmospheric boundary layer (Zhang et al., 1996; Collatz et al., 2000; Bounoua et al., 2002; Defries et al., 2002; Chen et al., 2003, Gao et al., 2004, Jiao et al., 2023). Surface turbulent heat fluxes, including sensible and latent heat fluxes, are fundamental determinants of local microclimate formation and serve as crucial regulators for vegetation activity (Chapin et al., 2011). With global climate change, the ecosystems and water resources of the Tibetan Plateau have undergone significant impacts. Turbulent heat flux data provide key insights to assess these changes and devise countermeasures. Therefore, the long-term continuous observational data of land-atmosphere

turbulent heat flux on the Tibetan Plateau hold significant value for studying the region's weather and climate (Swinbank and W.C., 1951; Baotian et al., 1996; Zheng et al., 2000; Baldocchi, 2014; Yu Guirui et al., 2017).

As a direct observation technique for turbulent heat flux, the Eddy Covariance (EC) method stands as the primary observational means for the international flux network (FLUXNET) and a plethora of meteorological, ecological, and hydrological observation sites (Shaoying et al., 2020). Initially proposed by Swinbank (1951), EC directly measures the turbulent pulsations of various physical quantities based on micrometeorological principles. It calculates flux by evaluating the covariance produced by wind speed pulsations and physical quantity pulsations during atmospheric turbulent motion, thereby measuring heat, mass, and momentum exchanges between the land and the atmosphere. While this method does not rely on assumptions like the near-surface similarity theory, due to observational principles and instrument construction, a series of necessary corrections, quality controls, and quality assurances must be applied to raw data before obtaining final flux calculation results (Lee et al., 2004). Additionally, given the Tibetan Plateau's geographical location, high altitude, and harsh natural conditions, during continuous observations of material and energy exchanges between the land and atmosphere, data omissions account for 40% to 60% of the total data. This significant omission rate profoundly affects data integrity and accuracy (Falge et al., 2001; Lee et al., 2004), subsequently influencing its application in climate and weather models (Stull, 1988). Over the past two decades, domestic and international scientists have extensively researched the quality control and assurance of turbulent flux data, forming standardized processing procedures (Papale et al., 2006; Mauder et al., 2008; Wang Shaoying et al., 2009; Wutzler et al., 2018). However, discussions regarding the imputation of missing flux data remain necessary (Foltynova et al., 2020). Rational imputation methods can enhance the integrity of observational data series, facilitating a more accurate understanding of dynamic changing processes and laying the foundation for simulation experiments. In research on energy and material exchanges between terrestrial ecosystems and the atmosphere, the choice of imputation method is paramount. Seasonal changes in ecological processes and soil moisture can influence measurements of turbulent exchanges (Reichstein et al., 2005). Therefore, selecting imputation methods that capture such complexities is crucial.

Currently, researchers have developed dozens of imputation methods, which can be mainly categorized into the following four types: (1) imputation methods based on mean values; (2) non-linear regression methods driven by environmental factors; (3) imputation methods based on artificial neural networks; and (4) imputation methods based on machine learning algorithms (Falge et al., 2001; Hui et al., 2004; Ooba et al., 2006; Moffat et al., 2007; Soloway et al., 2017; Wang et al., 2020). A rational approach to imputing missing flux data serves as a crucial foundation for data integration among stations and flux observation networks and is a key factor in enhancing data comparability (Wang et al., 2009). However, the methods for imputing flux data across various flux observation networks have not been standardized. For instance, FLUXNET and the European flux network, CarboEurope, adopted the Marginal Distribution Sampling (MDS) from the mean value imputation methods and successfully applied it to the FLUXNET2015 Dataset (Papale et al., 2006; Soloway et al., 2017). Meanwhile, ChinaFLUX and the Japanese flux network opted for non-linear regression methods to impute the net ecosystem exchange, while the sensible and latent heat fluxes were imputed using the day-night average transition method and the lookup table method (Li et al., 2008).

The Australian National Ecosystem Research Network, OzFlux, employed artificial neural network algorithms for imputation (Beringer et al., 2017), while the U.S. flux network, AmeriFlux, selected both MDS and artificial neural networks to impute missing flux data (Agarwal et al., 2014). The aforementioned studies suggest that machine learning algorithms have gradually been incorporated into the domain of missing flux data imputation and have demonstrated promising performance (Moffat et al., 2007; Dengel et al., 2013; Knox et al., 2015; Beringer et al., 2017).

Machine learning technology, as a rapidly advancing super-computing domain (Ortega et al., 2023), has already demonstrated its potential value in data imputation across sectors such as transportation, healthcare, and sensor networks (Duan et al., 2014; Matusowsky et al., 2020; Gad et al., 2021). The variability in turbulent heat flux represents an extremely complex process. This suggests that simple linear models may not accurately capture the complex relationships between meteorological elements and turbulent heat flux. Consequently, in some instances, traditional statistical methods may fail to provide accurate predictions. Machine learning models can handle the complex nonlinear relationships between predictive variables, regardless of their interdependencies or correlations, and the expected outcomes. Compared to traditional machine learning techniques, the superiority of deep learning in heat flux data imputation lies not only in its capacity to integrate more environmental driving variables that affect flux exchanges but also in its more precise ability to handle non-linear data patterns(Fawaz et al., 2019). This is attributed to deep learning's ability to learn complex data features through multi-layer neural networks, thereby more precisely capturing intricate relationships inherent in the data. Over the past decade, deep learning has expanded from image and text processing domains to time series analysis. Specifically, Recurrent Neural Networks (RNN) and their variants, such as Long Short-Term Memory networks (LSTM) and Gated Recurrent Units (GRU), have been proven to excel in handling sequential data. They capture long-term dependencies and non-linear patterns in the data, optimizing the accuracy of time series predictions. Furthermore, the Transformer architecture, with its attention mechanism, has offered a novel approach to processing time series data, showcasing superiority in time series simulations (Vaswani et al., 2017). Convolutional Neural Networks (CNN), initially designed primarily for image recognition, have in recent years been successfully applied to the analysis and forecasting of time series data. Unlike traditional image processing, time series CNN models typically operate on one-dimensional data. CNNs capture local patterns and trends in time series data through local receptive fields and weight sharing. Local features and dependencies in time series, such as periodic patterns or breakpoints, can be efficiently captured by convolutional layers. These attributes allow CNNs to outperform traditional methods and other deep learning models in certain time series tasks, such as anomaly detection, pattern recognition, and forecasting. Concurrently, the multi-layered convolutional structure enables the model to automatically extract multi-scale features from the data. However, the practical application of deep learning in turbulent heat flux imputation remains nascent, especially in the Tibetan Plateau region, with related studies still being sparse.

Unlike most of the prior studies, the present work attempted to evaluate the capability of various artificial intelligence models in imputing the turbulent heat flux data for the QOMS site during the third Tibetan Plateau Experiment in 2012. The objectives of this study are to quantitatively compare the outcomes of different artificial intelligence models and to propose a novel

turbulent heat flux imputation method based on deep learning. The ultimate goal is to complete the imputation of turbulent heat flux for this site spanning from 2007 to 2016 and make this dataset publicly accessible.

2 Materials and Methods

2.1 Site

The data used in this study originates from the third Tibetan Plateau experiment at the QOMS station located in the bottom of the Rongbuk Valley, to the north of Mount Everest (28.36°N, 86.95°E, at an altitude of 4298m), as shown in Fig.1(adapted from Ma et al.,2020). The surface at the observation point is barren with relatively flat and open terrain, sparse and low vegetation. From the surface to the deeper soil layers, it mainly consists of sand and gravel. Not only is this observation station influenced by climate variations and weather processes, but it is also affected by local circulations of the Himalayan range, such as valley winds, making it an ideal location for monitoring surface processes on the Tibetan Plateau.

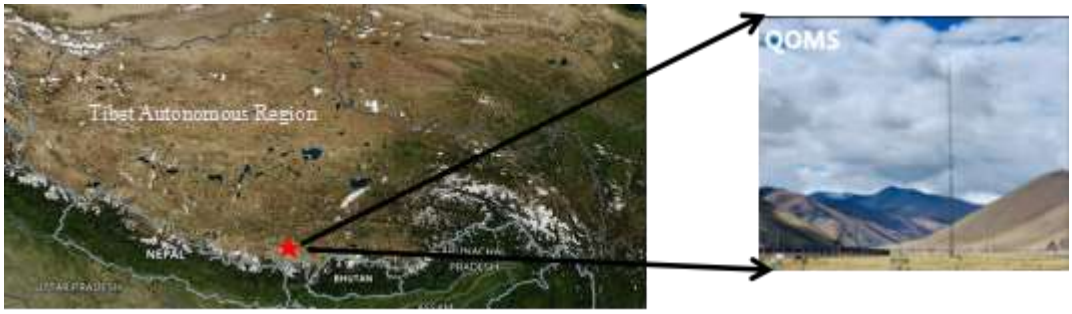

Figure 1：Geographical location and site images of the QOMS station (The map on the left is from Google Maps, and the site on the right is adapted from Ma et al.,2020).

Figure 2 showcases the daily average time series of the main meteorological elements recorded at this observation station, reflecting the unique meteorological characteristics of the alpine region. During the observation period, the average values for temperature, relative humidity, and annual precipitation were 4.16ºC, 43.47%, and 289 mm respectively. Windspeeds observed in the winter are generally high, reaching up to 16 m/s, while being relatively lower in the summer. During midday in the summer, surface temperatures can rise to 60 ºC, displaying a gradually increasing pattern throughout the year. Correspondingly, consistent with the annual summer rainfall pattern, surface humidity peaks in the summer.

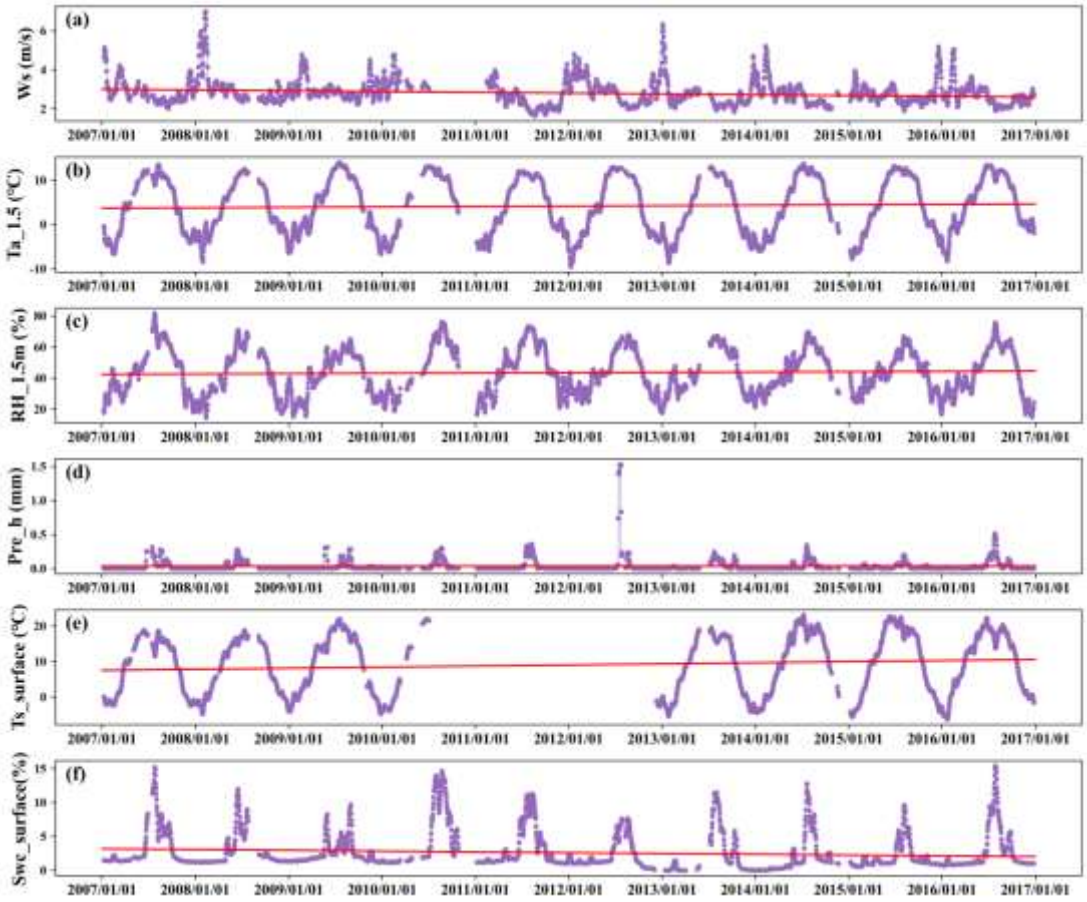

Figure 2:Time series of daily meteorological elements collected from January 1, 2007, to December 31, 2016:(a) Wind Speed (WS);(b)Air Temperature at 1.5m height (Ta_1.5m);(c)Relative Humidity at 1.5m (RH_1.5m);(d) Hourly Average Precipitation (Prec._h);(e)Surface Soil Temperature (Ts_surface);(f) Surface Soil Water Content (Swc_surface).

In recent years, many studies have focused on the climatic change characteristics of the Tibetan Plateau, especially the phenomena of increasing temperatures and decreasing wind speeds (Yang et al., 2014). As shown in Fig. 2, a linear fit of the basic meteorological variables indicates that the temperature and relative humidity at 1.5 m and surface temperature displayed an upward trend from 2007 to 2016, while the wind speed at 1.5 m exhibited a downward trend, consistent with previous findings (Yang et al., 2014).

The Mann-Kendall (MK) trend test, with a p-value set at 0.05, offers a reliable method for identifying and quantifying potential trends in time series data. This method has been widely employed in studies in the fields of hydrology, climate change, and environmental science and has proven to be an effective tool for assessing long-term trends (Hirsch et al., 1982; Yue et al., 2002). In this study, we employed the MK trend test with a p-value set at 0.05, which likewise corroborated the aforementioned

results. The outcomes of the MK trend test and the fitting equations are presented in Table 1, where the unit for X in the fitting equation is hours.

Table 1:MK Statistics and Fitting Equations

Indicator	MK	P-value	Fitting Equation	Trend
Wind speed	-0.036	$4.48e^{-51}$	$Y = -5.19\times10^{-8}X + 3.01$	Downward
Air temperature	0.023	$3.90e^{-23}$	$Y = 1.19\times10^{-5}X + 3.69$	Upward
Humidity	0.017	$7.19e^{-13}$	$Y = 2.26\times10^{-5}X + 42.43$	Upward
Precipitation	0.023	$4.60e^{-16}$	$Y = 1.75\times10^{-8}X + 0.03$	Upward
Soil temperature	0.017	3.14^{-10}	$Y = 4.99\times10^{-5}X + 7.49$	Upward
Soil water content	-0.19	0	$Y=-1.44\times10^{-5}X+3.18$	Downward

2.2 Site Data

This study analyzes the observational data from the QOMS observation site from January 1, 2007, to December 31, 2016. Specific variables include: sensible heat flux H, latent heat flux LE, soil heat flux SHF, air temperature at five levels Tair (1.5, 2, 4, 10, 20 m), relative humidity at five levels RH (1.5, 2, 4, 10, 20 m), wind speed at five levels WS (1.5, 2, 4, 10, 20 cm), wind direction at five levels WD(1.5, 2, 4, 10, 20 cm), downward shortwave radiation Rsd, upward shortwave radiation Rsu, downward longwave radiation Rld, upward longwave radiation Rlu, soil temperature at six levels Tsoil(0, 0.1, 0.2, 0.4, 0.8, 1.6 m), and soil volumetric water content at six levels SWC(0, 0.1, 0.2, 0.4, 0.8, 1.6 m). The instruments used at the site are shown in Table 2.

Table 2:Installation Heights and Burial Depths of the Observation Instruments

Variables	Sensor models	Manufacturers	Heights	Units
Air temperature	HMP45C-GM	Vaisala	1.5,2,4,10and20m	^{o}C
Wind speed and direction	034B	MetOne	1.5,2,4,10and20m	$ms^{-1}/^{o}$
Humidity	HMP45C-GM	Vaisala	1.5,2,4,10and20m	%
Pressure	PTB220A	Vaisala	-	hPa
Radiations	CNR1	Kipp&Zonen	-	Wm^{-2}
Precipitation	RG13H	Vaisala	-	mm
Soil temperature	Model107	Campbell	0,0.1,0.2,0.4,0.8and1.6m	◦C
Soil water content	CS616	Campbell	0,0.1,0.2,0.4,0.8and1.6m	v/v%

Soil heat flux	HFP01	Hukseflflux	0.05m	Wm^{-2}
H	CSAT3	Campbell	3.25m	Wm^{-2}
LE	LI-7500	Li-COR		

From 2007 to 2016, the missing rates for H and LE at the observation site (including missing and distorted data), denoted as gap_H and gap_LE, were 21.7 % and 21.4 % respectively (Table 3).

Table 3:Missing Rates for Sensible Heat Flux and Latent Heat Flux

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
gap_H	39.4%	10.3%	22.2%	9.8%	32.2%	29.6%	11.7%	10.4%	18.1%	33.3%
gap_LE	37.65%	8.28%	21.30%	8.48%	23.63%	28.52%	9.90%	8.84%	33.57%	33.34%

2.3 Data Preprocessing

To interpolate missing flux data using machine learning algorithms, it is essential to ensure the completeness of environmental driving variables (Wang Shaoying et al., 2009). Therefore, during the data preprocessing phase, this study employed the K-Nearest Neighbors (K-NN) interpolation method to address the missing data of environmental driving variables. The choice of setting the number of neighbors to 3 is based on the consideration that a smaller number of neighbors can reduce computational complexity and enhance the efficiency of the imputation process while maintaining accuracy. Additionally, "distance" was used as the weight calculation method to ensure that observations closer in distance receive higher weights (Friedman et al., 2009). The equation is given by Eq. (1).

$$\text{Gap\_filling value} = \frac{\sum_{i=1}^{3} \frac{y_i}{d_i}}{\sum_{i=1}^{3} \frac{1}{d_i}} \quad (1)$$

Where y_i is the observation of the ith nearest neighbor; d_i is the distance between the missing value and the ith nearest neighbor, In this study, by setting weights= "distance" parameter, KNN imputation (KNNImputer) is calculated using weighted Euclidenian distance formula; The numerator involves the weighted sum based on the observations and the reciprocal of the distances of the 3 nearest neighbors; The denominator is the sum of the reciprocals of the distances for these 3 nearest neighbors. Subsequently, the "fit_transform" method is utilized to fit and transform the chosen data, facilitating the imputation of missing values. Lastly, the imputed data is merged with the original sensible and latent heat fluxes to derive a comprehensive dataset of environmental driving variables.

This K-NN-based imputation method has been demonstrated to be effective for datasets exhibiting similar patterns or local consistency (Little and Rubin, 2002). In other words, the KNN approach is applicable when the correlation length scale significantly exceeds the distance between missing and available data points. By considering the distances between observation

points over time, this method can accurately estimate missing values, thereby enhancing the utilityof the data for subsequent analyses.

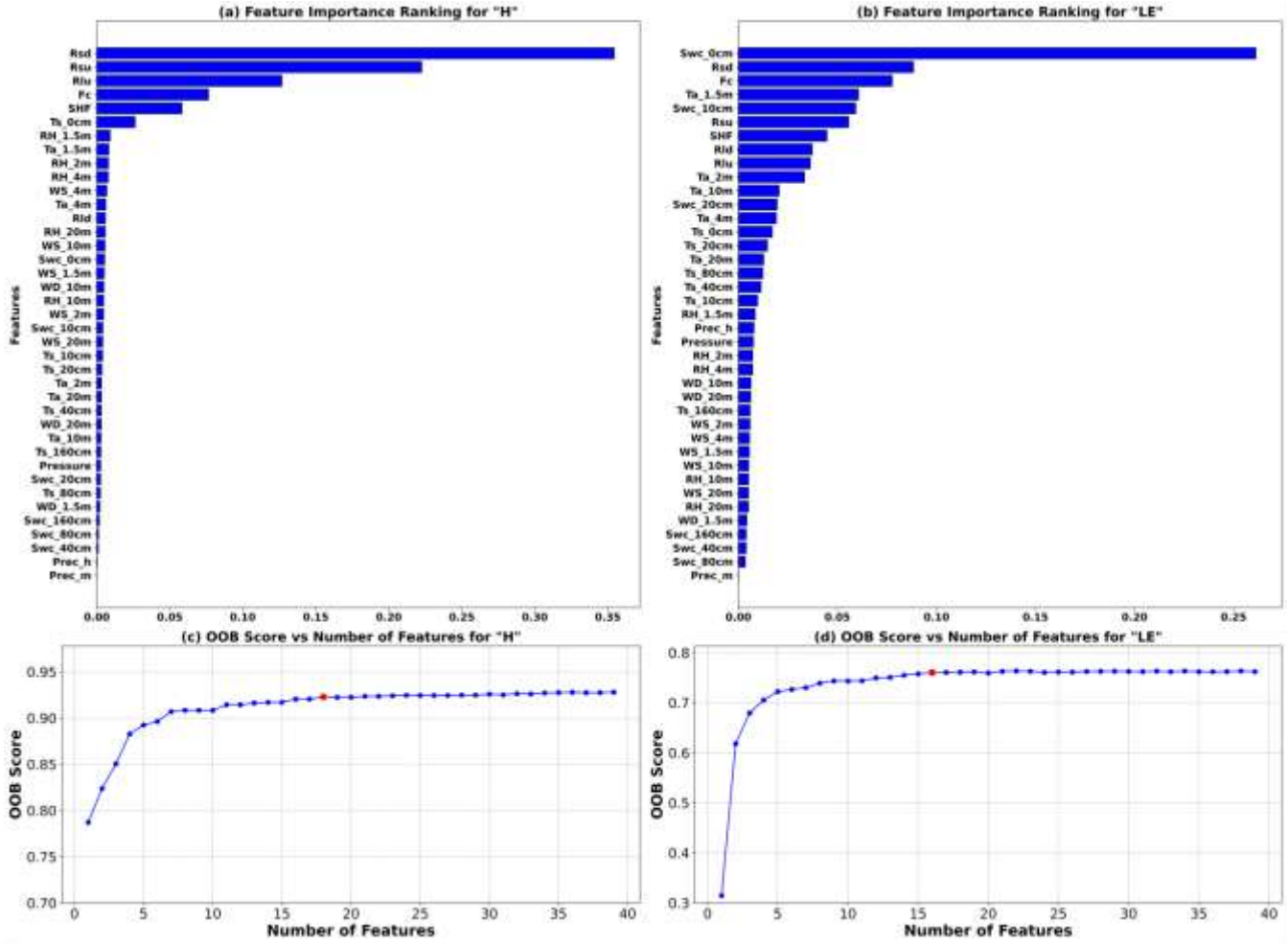

Figure 3:(a) Importance index for sensible heat flux (H), (b) Importance index for latent heat flux (LE), (c) OOB scores for different feature combinations of sensible heat flux based on Random Forest (the red dot indicates the maximum value), (d) OOB scores for different feature combinations of latent heat flux based on Random Forest (the red dot indicates the maximum value).

Random forests assess the contribution of each variable to model predictive performance through importance ranking, with variables that contribute significantly to predictive performance receiving higher rankings. By sorting features based on their importance, random forests select the optimal feature combination, not only effectively reducing the dimensionality of input features but also aiding in the selection of variables within machine learning models. In this study, the number of trees for the random forest model was set to 159 based on ten-fold cross-validation and grid search algorithms, meaning the model consists of 159 decision trees. Since bootstrapping (sampling with replacement) is used to generate random decision trees, not all

samples participate in the tree generation process. The unused samples are referred to as out-of-bag (OOB) samples, which can be used to evaluate the accuracy of the trees. OOB scores provide an unbiased estimate of the model's generalization ability by effectively assessing the model's capability to predict unknown data. The higher the OOB score, the stronger the model's generalization capability (Wang et al., 2023).

Random Forest analysis at the QOMS site on the Tibetan Plateau reveals that downward shortwave radiation has the greatest impact on sensible heat flux, while soil water content is most influential for latent heat flux, aligning closely with their respective roles in physical processes. The site is characterized by arid, barren, flat, and open terrain with sparse, low vegetation, primarily composed of sand and gravel from the surface to deeper soil layers. Such environmental conditions mean that less solar radiation is absorbed by plants for photosynthesis, allowing more shortwave radiation energy to reach the ground directly, thus increasing surface heating. Moreover, the flat and open terrain, combined with sandy and gravelly textures, enables efficient absorption and re-radiation of solar energy, significantly affecting the formation of sensible heat flux. In this arid and barren environment, soil water content plays a crucial role in regulating the surface energy balance, affecting predictions of latent heat flux significantly, where even minimal variations can have substantial impacts under conditions of water scarcity. In conclusion, in subsequent analyses of turbulent heat flux, characteristics such as radiation, air temperature, soil temperature, and soil water content exhibit high importance. However, whether to exclude some relatively less important features still requires further consideration and analysis.

The OOB scores under different input feature dimensions have been calculated, with variables input in order of importance, as shown in Fig. 3c. For the sensible heat flux, the OOB score is highest when considering the top 18 features ranked by importance; as shown in Fig. 3d, for the latent heat flux, the OOB score peaks when considering the top 16 features. Additional features added afterward no longer impact the results; in other words, the optimal combination for sensible heat comprises the top 18 features, and for latent heat, it's the top 16 features.

3 Experiments

3.1 Experimental design

We present here the experimental design and the different statistical and learning methods used in this study.The aim of this study is to use a decade's worth of observational data to fit missing values for sensible and latent heat fluxes. We explored the turbulence flux changes at the QOMS site from 2007 to 2016 and treated the missing parts of the turbulence flux in the dataset as quantitative prediction variables. The objective is to use other meteorological elements as environmental drivers to impute these missing data, forming a complete heat flux dataset.

In the research application of the model, it is crucial to correctly divide the training, validation, and test sets (Bishop and M, 2006; Friedman et al., 2009). The training set is used to train the model's parameters so that the model can learn and capture the underlying patterns and structures from the given data (Goodfellow et al., 2016). The primary purpose of the validation set

is for model selection and hyperparameter tuning, enhancing the model's generalization capability (Cawley and Talbot, 2010). The test set offers a completely independent evaluation method to more accurately assess the model's performance on unseen data (Arlot and Celisse, 2010). This dataset has never been used in the training or validation process, so it can serve as an unbiased estimate of the model's performance in practical applications (Kohavi, 1995). This study utilizes ten years of data and employs a rolling forecasting approach for training and testing the model. Specifically, each year is sequentially selected as the test set, with the remaining nine years used for training. For instance, the data from 2007 is initially used as the test set, while data from other years serves as the training set. This process is then repeated with 2008 as the test year, and so on. Notably, due to significant missing turbulence heat flux data in 2012, this paper will primarily present the data imputation for that year, while the data handling for other years is detailed in supplementary files. According to the research objectives and the length of the interpolated dataset, all samples are divided into three groups: a training set (2007-2011, 2013-2016), with 10% randomly extracted as the validation set, and the test set (2012). In total, there are 87,673 samples, with 80% used for training, 10% for validation, and the remaining 10% for testing. Traditional statistical methods do not involve the division into training, validation, and testing sets. However, for ease of comparison in this paper, the data imputation results from 2012 using traditional statistical methods are used as the test set, while the remaining imputation results serve as the training set This design plan fully considers the complexity and diversity of time series analysis while ensuring the rigor of model validation and testing. In this way, it provides an accurate and reliable means to predict soil turbulence heat flux.

3.2 Traditional statistical methods

Current techniques used for imputing missing data in turbulent heat flux include linear interpolation (Alavi et al., 2006), variable relationships(Soloway et al., 2017), mean diurnal variation (MDV) (Falge et al., 2001), nonlinear regression (NR) (Chen et al., 2012), and look up tables (LUT) (Falge et al., 2001). Linear interpolation is suitable only for small gaps (1-3 consecutive missing data points), but in the Tibetan Plateau, turbulent heat flux often exhibits prolonged periods of missing observations, rendering linear interpolation unreliable. Variable relationships utilize linear relationships between meteorological variables for mutual interpolation; however, due to the strong nonlinear relationships between turbulent heat flux and environmental driving factors, the mere method of variable relationships struggles to accurately capture the changes in turbulent heat flux. Additionally, the variable relationships vary across different sites on the vast and geographically diverse Tibetan Plateau. The daily variation method involves establishing a time window, typically between 4-15 days, with 7 and 14 days being the most common selections. Within this window, averages of observations at the same time are computed to obtain a set of daily variation data, and missing data within these averages are imputed using linear interpolation, filled with corresponding daily variation data for the respective times. Nonlinear regression is based on an understanding of the main factors controlling the flux, thereby effectively capturing the impact of major environmental element changes on the flux, allowing for more accurate data imputation. The lookup table method is based on creating a data retrieval table from valid data, searching for valid data under similar environmental conditions according to major environmental factors, and averaging the

found data to impute missing data. Given the harsh geographical conditions of the Tibetan Plateau, we employ the daily variation method, nonlinear regression, and look up tables at the QOMS site to impute data, exploring the gap between these methods and machine learning approaches.

3.3 Traditional machine learning methods

The Support Vector Machine (SVM) is versatile and can be applied not only as a linear classifier but also for non-linear classification through the use of kernel functions. Moreover, beyond its capability for classification, SVM can be adapted for regression tasks—known as Support Vector Regression (SVR). This approach aims to find an optimal hyperplane in a high-dimensional kernel space that best fits the data points, thereby ensuring optimal regression performance. This versatility allows SVM to address both classification and regression problems effectively (Cortes and Vapnik, 1995). XGBoost is a decision tree based gradient boosting algorithm that enhances the model's performance by progressively adding new trees and adjusting the errors of previous trees. It has been proven to perform excellently in various competitions and practical applications (Chen et al., 2016). The K-Nearest Neighbors (KNN) algorithm is an instance-based learning method. It classifies or predicts by calculating the distance between the input data point and the data points in the training dataset, selecting the nearest K points, and voting based on their labels (Cover and Hart, 1967). Each algorithm has its unique principle, offering multiple choices for addressing the problem of turbulent heat flux imputation.

3.4 Recurrent neural network

Recurrent Neural Networks (RNNs) are a class of deep learning models designed for processing sequential data (Goodfellow et al., 2016). The core idea is to share weights between the hidden layers of the network to capture temporal dependencies within sequences. However, standard RNNs suffer from issues of vanishing and exploding gradients, which limit their ability to capture long-term dependencies. Long Short-Term Memory networks (LSTM) address this problem by introducing special units with three gate structures, allowing the network to learn and remember long-term dependencies within sequences (Hochreiter and Schmidhuber, 1997). Gated Recurrent Units (GRUs) are a variant of LSTM that improve computational efficiency by simplifying the gate structure and reducing the number of parameters, while retaining the ability to capture long-term dependencies (Cho et al., 2014). These recurrent neural network architectures have achieved significant success in many sequence modeling and prediction tasks.

3.5 Transformer

The Transformer model is a deep learning architecture widely used in natural language processing and other sequence-to-sequence tasks (Vaswani et al., 2017). It mainly consists of two parts: an encoder and a decoder. Transformers capture long-distance dependencies in sequences through the self-attention mechanism. Self-attention allows the model to consider other positions in the input sequence simultaneously at all positions, which, unlike traditional RNNs and LSTMs, eliminates the

need for sequential computation, thereby greatly enhancing parallel computation capabilities. Following each self-attention layer is a feed-forward neural network, accompanied by layer normalization, which contributes to training stability and convergence. The Transformer exhibits outstanding performance in data fitting and prediction (LI et al., 2019). Its ability for parallel computation allows it to process large datasets more quickly. The self-attention mechanism ensures that the model can capture complex dependencies, surpassing previous methods in many tasks (Wu et al., 2020).

3.6 Transformer_CNN

To address the high complexity of data from the Tibetan Plateau, a deep neural network model based on the PyTorch framework was adopted in this study, as illustrated in Fig. 4. At the initialization of the model (Feed-Forward), a layer normalization component was introduced with the aim of normalizing the input along the embedding dimension, thereby enhancing the stability and convergence rate of network training. Subsequently, a feed-forward neural network comprising three fully connected layers was defined, incorporating a ReLU activation function to capture non-linear features (please refer to the supplementary information). This paper considers employing kernels of sizes 3, 5, and 7 to capture multi-scale features. In the simulation of turbulent heat flux, turbulence phenomena display distinct characteristics at various spatial scales. Smaller kernels (such as 3) can capture more localized features, while larger kernels (such as 5 and 7) are able to cover a wider area, capturing more global features. This combination enables the model to concurrently learn features across different scales, thereby enhancing the model's understanding and predictive capacity regarding changes in turbulent heat flux. Following this, another one-dimensional convolutional layer was defined to integrate the outputs from the previous three convolutional layers, forming a comprehensive feature representation.

The multi-head self-attention mechanism was realized through the Multi-Head Attention component, which boasts four attention heads capable of capturing long-distance dependencies within the input sequence. The Decoder section of the model is responsible for mapping the encoded features to the target space. The initialization of weights and biases is designed to ensure the stability and efficiency of the model training process. Specifically, we adopted a variant of the He initialization method (He et al., 2015), a scientific approach to weight initialization frequently employed in deep learning models to improve convergence speed and stability during training. The core idea of He initialization is to adjust the initial standard deviation of weights based on the number of nodes in the previous layer (i.e., fan_in). This initialization method is particularly crucial for the training of deep neural networks as it helps prevent issues of vanishing or exploding gradients, which often occur when traditional random weight initialization methods are used.

The forward propagation process was defined by the feed-forward function, and ultimately, when the model is invoked, two data views are generated: F1 (primary view) and F2 (contrast view). Despite the presence of dropout, the two views might still differ. The loss function employed is the Smooth L1 Loss, comprised of three parts: the loss between F1 and the true value, the loss between F2 and the true value, and the distance between F1 and F2, which serves as a regularization term multiplied by 0.1. During model inference, the final prediction is derived from the average of F1 and F2.

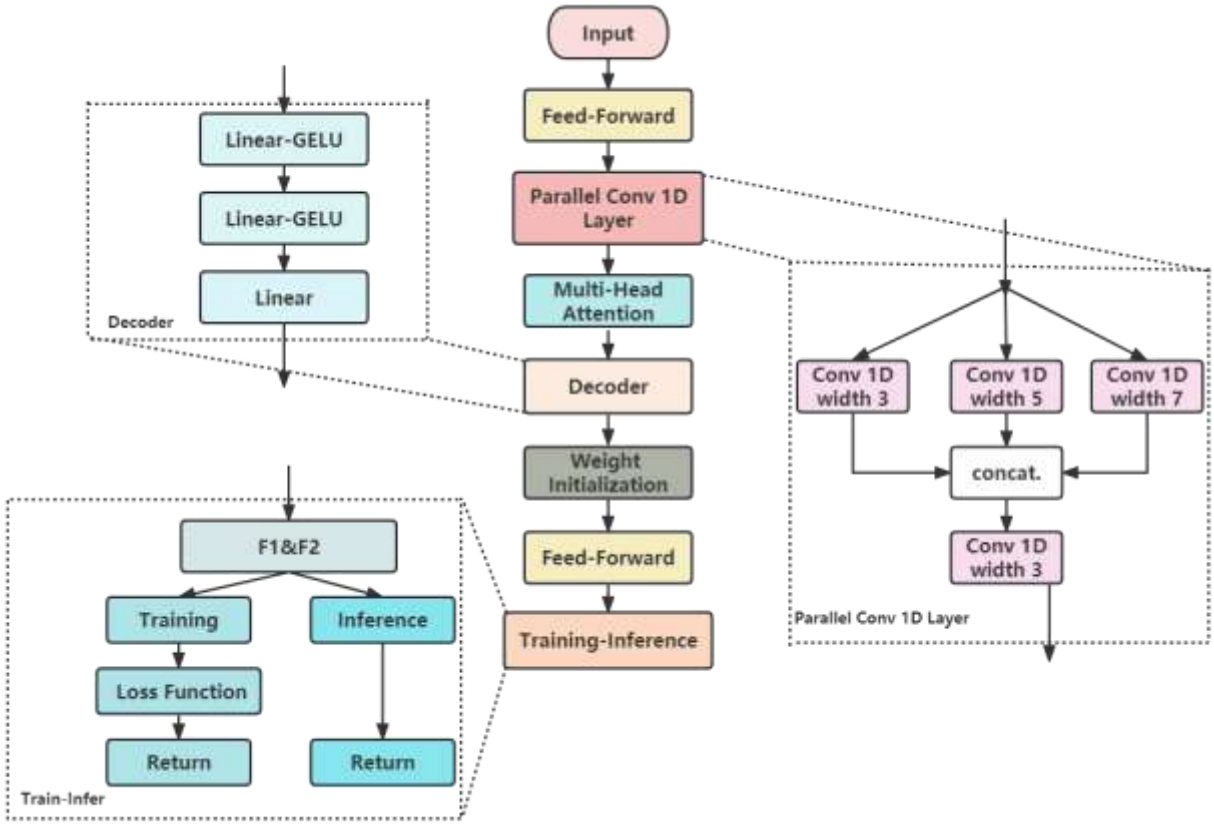

Figure 4:The model framework and network structure of Transformer_CNN.

The Transformer_CNN model is a novel deep learning framework specifically designed to address the complex physical phenomena of turbulent heat flux or similar challenges. It integrates the features of Convolutional Neural Networks (CNN) and Transformer, aiming to capture the intricate relationships temporal dimensions. In this model, the CNN, through its convolutional layers, manages to capture the evident seasonal and cyclical variations in turbulent heat fluxes. By identifying these local patterns, it discerns the daily, monthly, and seasonal variations in turbulent fluxes, thereby holding an edge in capturing intricate patterns within the data (Krizhevsky et al., 2012). Secondly, predicting turbulent heat fluxes encompasses intricate physical processes and multi-scale interactions. The prowess of the Transformer model in capturing long-distance dependencies (Vaswani et al., 2017), combined with the CNN's local feature extraction capability, facilitates a superior grasp of the interactions among wind speed, temperature, and radiation with respect to turbulent heat fluxes, thereby enhancing the model's versatility and diversity. Moreover, the efficiency of convolutional operations might contribute to elevating the speed and efficiency of model training (Goodfellow et al., 2016). By employing a hybrid model of CNN and Transformer, both local

and global features can be concurrently captured, thereby manifesting an adaptive advantage in data fitting for turbulent heat fluxes (Bello et al., 2019).

3.7 Statistical Analysis

In this study, traditional statistical analysis indices (RMSE, MAE, and R^2) were used to evaluate the accuracy of various models. The comparison statistics were calculated as follows:

Root Mean Square Error (RMSE): RMSE represents the square root of the mean of the squared errors, which is the average of the differences between the simulated and observed values. The lower the RMSE value, the better the model's fit. The relationship between them can be expressed as Eq. (2):

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \quad (2)$$

Mean Absolute Error (MAE): MAE calculates the average of the absolute differences between the observed and predicted values. A lower MAE value indicates a better fit of the model. The relationship between them can be expressed as Eq. (3):

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i| \quad (3)$$

Coefficient of Determination (R^2): R^2 measures the proportion of the variance in the dependent variable that is predictable from the independent variables. This metric indicates how close the data are to the fitted regression line. The closer R^2 is to 1, the more effectively the model explains the data's variability. The relationship between them can be expressed as Eq. (4):

$$R^2=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2} \quad (4)$$

Where y_i represents the model-simulated value, $\hat{y}$ denotes the observed value, $\bar{y}$ signifies the mean of the observed values. The subscript i represents the serial number of samples, and N represents the total number of samples.

3.8 Hyperparameter Optimization

In this study, we employed a standardized hyperparameter optimization method. The batch size was set to 32 to control the number of samples used for each parameter update. The learning rate and weight decay parameters were gradually adjusted during training to optimize model performance. The initial learning rate was set at 0.0005 and halved every 6 epochs, with the weight decay parameter set at 0.01. The training was conducted over 100 epochs, with multiple iterations for model training and validation. During each iteration, we recorded the training loss and validation loss, using Mean Squared Error Loss (MSELoss) and Smooth L1 Loss (SmoothL1Loss) to calculate the loss. Gradient clipping (with a maximum norm of 10) was applied to ensure training stability. Specifically, we updated model parameters using the training set and evaluated model performance on the validation set. The primary evaluation metric was the R-squared (R^2) score. The hyperparameter optimization followed these steps:

A) Initial Training: The model was trained with the initial hyperparameter settings and evaluated on the validation set, recording the initial R^2 score on the validation set.

B) Learning Rate Adjustment: After every 6 training epochs, the R^2 score on the validation set was checked. If the score improved, the current model parameter configuration was saved. Otherwise, the learning rate was halved, and training continued.

C) Weight Decay and Batch Size Optimization: Throughout the training process, the weight decay and batch size were kept constant to ensure training stability and convergence.

D) Saving the Best Model: At the end of each training epoch, the R^2 score on the validation set was evaluated. If the current score was better than the previous best score, the current model parameter configuration was saved.

4 Results

The performance of the nine models (MDV, NR, LUP, SVM, KNN, XGBoost, LSTM, GRU and Transformer) was assessed by calculating the RMSE and MAE values between the predicted and actual values for both the sensible and latent heat fluxes. These evaluations are presented in Table 4, encompassing the training, validation, and test sets, with the best results highlighted in bold. The results are clear; whether for H or LE, traditional statistical methods are comprehensively outperformed by machine learning algorithms. Among the three traditional machine learning methods, the SVM model demonstrated superior performance in simulating the sensible heat flux, whereas XGBoost excelled in simulating the latent heat flux. Surprisingly, both RNN models exhibited subpar performance in the task of simulating turbulent fluxes. Two predominant factors might account for these observations. One pertains to the challenges of gradient vanishing and explosion. Although LSTM and GRU alleviate the issues of gradient vanishing and explosion through gating mechanisms, these problems can still affect model performance when dealing with particularly long sequences. The second factor concerns the nuances of hyperparameter optimization in RNN models. Choosing the right set of hyperparameters, which are particularly numerous in RNNs, is crucial to achieving optimal model performance. Fortunately, the Transformer model showcased exceptional prowess in the task of simulating turbulent fluxes. In almost all simulations, the Transformer model achieved the best performance, boasting the smallest RMSE and MAE on the test set. As a result, the Transformer model architecture was integrated into the neural network framework, and by further incorporating convolutional layers and multi-head attention mechanisms, a Transformer_CNN model was proposed, which was found to be superior in simulating turbulent fluxes.

Table 4:Model performance evaluation (RMSE and MAE) for MDV, NR, LTT, SVM, KNN, XGBoost, LSTM, GTU, and Transformer. Bold values highlight the best performance.

	H	LE

Sets	Training		Validation		Test		Training		Validation		Test	
	RMSE	MAE	RMSE	MAE	RMES	MAE	RMSE	MAE	RMSE	MAE	RMES	MAE
MDV	42.34	6.32	/	/	40.26	5.91	97.67	17.02	/	/	95.17	17.19
NR	40.59	6.54	/	/	38.44	6.90	115.86	19.37	/	/	107.54	18.76
LUT	51.46	8.89	/	/	52.64	9.86	132.24	20.59	/	/	142.18	21.43
SVM	19.38	2.856	19.88	2.965	25.89	3.182	25.12	3.093	21.12	3.174	19.79	3.124
KNN	20.83	2.946	25.18	3.016	31.41	3.519	**14.61**	**2.692**	19.99	3.095	20.05	3.271
XGBoost	25.58	3.124	25.07	3.549	29.34	4.178	16.21	2.977	19.11	3.066	16.88	3.034
LSTM	25.13	3.029	24.59	3.481	28.68	4.159	18.86	2.859	20.72	3.017	19.47	3.033
GRU	22.14	3.004	21.74	3.257	26.99	4.036	17.59	2.818	21.76	3.157	20.96	3.198
Transformer	**16.65**	**2.531**	**18.07**	**2.814**	**24.04**	**3.029**	18.15	2.883	**19.10**	**3.079**	**14.57**	**2.830**

Table 5 juxtaposes the results of the Transformer_CNN with various artificial intelligence models, illustrating the predictive outcomes in terms of the coefficient of determination (R^2). Evidently, the Transformer_CNN possesses a distinct advantage in long-term predictions. The performance of the Transformer_CNN model in data fitting surpassed that of the conventional Transformer model. The incorporation of the Convolutional Neural Network (CNN) enhanced the model's capability to extract local features (Lecun et al., 1998). In summary, the Transformer_CNN model, by amalgamating the Transformer's global dependency capture capability with the CNN's local feature extraction prowess, offers a richer and more flexible model representation, thereby exhibiting superior performance in data fitting.

Table 5:Comparison of the Coefficient of determination (R^2) predicted by multiple models. Bold values highlight the best performance.

	H			LE		
Sets	Training	Validation	Test	Training	Validation	Test
MDV	0.65	/	0.68	0.50	/	0.49
NR	0.62	/	0.72	0.48	/	0.53
LUT	0.59	/	0.54	0.45	/	0.40
SVM	0.92	0.91	0.90	0.78	0.80	0.82
KNN	0.91	0.90	0.85	0.86	0.78	0.79
XGBoost	0.90	0.90	0.87	0.83	0.78	0.85
LSTM	0.93	0.93	0.88	0.82	0.80	0.83
GRU	0.94	0.94	0.89	0.84	0.78	0.81
Transformer	0.93	0.93	0.91	0.85	0.81	0.87

Transformer_CNN	**0.94**	**0.94**	**0.95**	**0.90**	**0.81**	**0.90**

To more comprehensively and intuitively describe the Transformer_CNN, Fig. 5 displays a scatter plot of predicted values against actual values. The distance between the data points and the diagonal line indicates prediction errors. The results suggest that the majority of the turbulent heat flux values are centered around 0. Owing to the large volume of low-value data, the model is adept at capturing the characteristics of environmental driving forces when observed values are near 0, thereby achieving more accurate predictions. As observed values increase, the prediction error of the model gradually amplifies. Furthermore, it was observed that when the turbulent heat flux is substantial, predicted values typically fall below the observed values. This phenomenon is more pronounced in the fitting of LE.

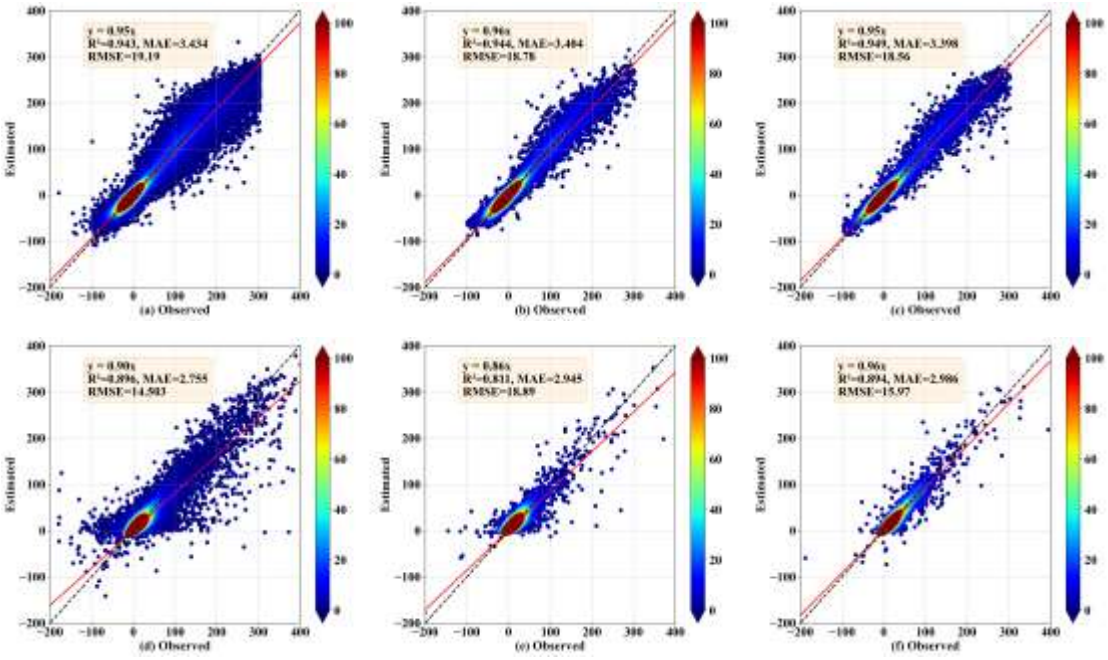

Figure 5:Scatter density plots of observed values and Transformer_CNN estimated values, where a) and d) are for the training dataset, b) and e) are for the validation dataset, and c) and f) are for the test dataset. a), b), and c) are estimates for H, while d), e), and f) are estimates for LE.

To better display the predictions of the Transformer_CNN model, Fig. 6a and Fig. 6b respectively show the monthly average diurnal variation curves for H and LE in the test set. The red line represents observed values, while the blue and green lines represent the predictions of the Transformer and Transformer_CNN models respectively.

In the prediction of sensible heat flux, both the Transformer and Transformer_CNN models perform excellently for the hours of 0-9 and 19-23, where their predicted values closely align with the observed values. However, between 9-19 hours, as solar radiation intensity increases and the sensible heat flux rapidly grows, the Transformer model struggles to capture this escalating trend, resulting in a notable underestimation. The Transformer_CNN model, having incorporated convolutional layers, is better equipped to recognize periodic data changes and the impacts of environmental drivers on sensible heat, substantially rectifying the underestimation issues observed in high values.

In the latent heat flux prediction, the performance superiority of the Transformer_CNN model is even more pronounced. While the Transformer model exhibits significant over estimations in low-value periods and struggles to capture high values, the Transformer_CNN model's predictions largely coincide with observed values, significantly reducing the prediction errors exhibited by the Transformer model. Not only does it excel during the low-value periods of LE in January-March and October-December, but it also accurately predicts the pronounced increase of LE in July-September. The experiments demonstrate that the Transformer_CNN model is well-suited to serve as an artificial intelligence model for imputing turbulent heat fluxes.

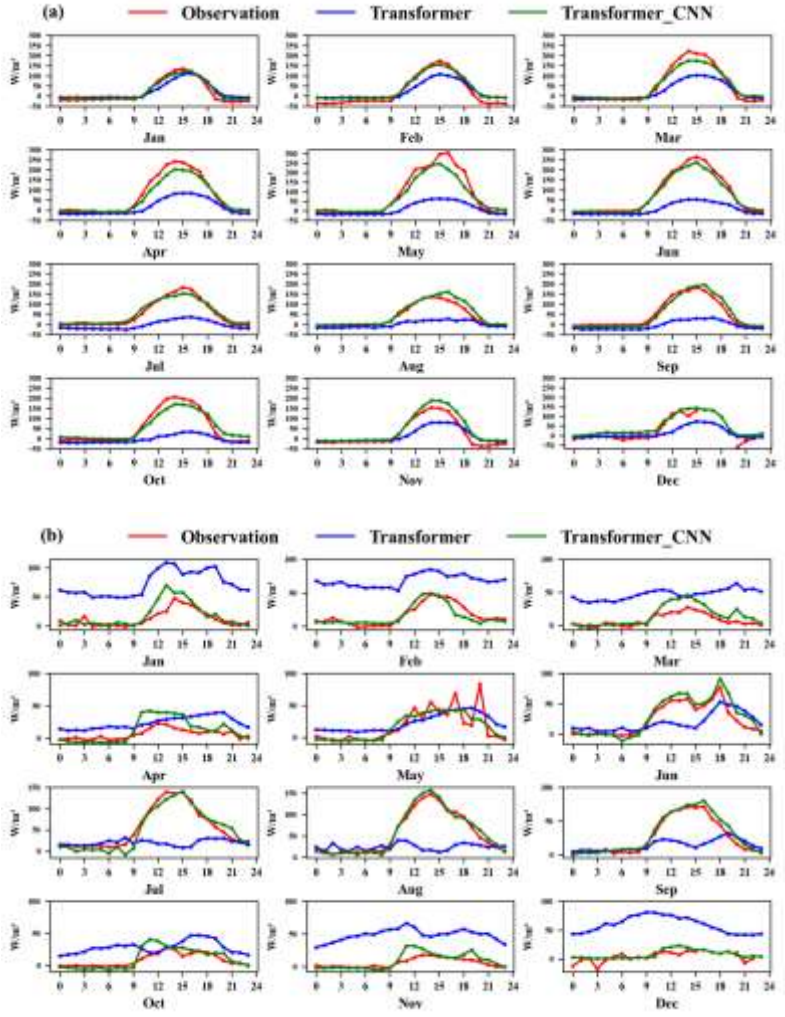

Figure 6:Observed values, Transformer estimated values, and Transformer_CNN estimated values for the monthly average diurnal variation curves in the test set (2012) are shown (x-axis is in hours). Specifically, (a) represents the sensible heat flux (H), and (b) represents the latent heat flux (LE).

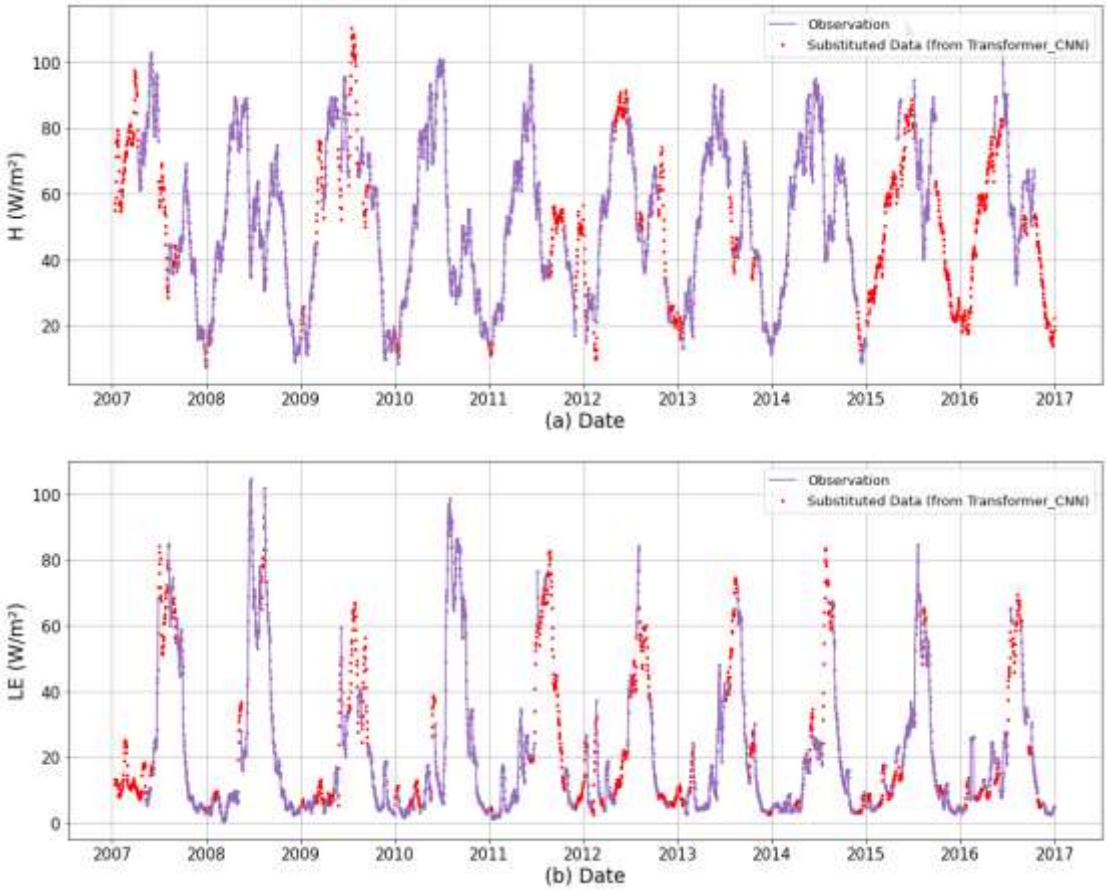

Figure 7:Observed values and Transformer_CNN estimated values for the variation curves from 2007 to 2016 are presented. Specifically, (a) depicts the sensible heat flux (H), and (b) illustrates the latent heat flux (LE).

Based on the research presented earlier, it was determined that the Transformer_CNN model can serve as an artificial intelligence model for imputing turbulent heat fluxes. To delve deeper into the variation of turbulent heat fluxes at the QOMS site, the model was employed to impute data from 2007 to 2016 for the QOMS site, with the results shown in Fig. 7. The variation in sensible heat flux, as depicted in Fig. 7a, indicates that prior to the monsoon season, the sensible heat flux is the primary consumer of the available energy at the Earth's surface. With the onset of the summer monsoon, the diurnal variation of sensible heat flux significantly decreases, equating to the latent heat flux. In other words, during the pre-monsoon period, the exchange of sensible heat flux dominates. Influenced by the interaction of mid-latitude westerlies and the summer monsoon, the summer sensible heat flux is significantly lower than that of the spring. In contrast to the bi-modal seasonal variation of sensible heat flux, the seasonal variation of latent heat flux exhibits a single peak pattern. That is, during the pre-monsoon period, the latent heat flux is small, but with the outbreak of the monsoon, it rapidly increases due to frequent precipitation and

the moistening of the surface soil. Subsequently, the latent heat flux gradually increases, equating to the sensible heat flux during the summer monsoon period. A comparison of the seasonal variations of the sensible heat flux (Fig. 7a) and the latent heat flux (Fig. 7b) suggests that during the Asian summer monsoon season, the impacts of latent and sensible heat fluxes at the QOMS site are comparable; during the non-monsoon season in Asia, the site's sensible heat flux has a greater impact.

Table 6:The imputation results for some elements at the QOMS and SETORS sites

Sets	QOMS						SETORS					
	H			LE			H			LE		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
Transformer (selected elements)	34.76	4.36	0.74	37.58	4.77	0.69	39.96	5.28	0.70	42.48	4.79	0.67
Transformer_CNN (selected elements)	29.34	3.44	0.83	30.25	3.93	0.78	31.66	4.83	0.80	34.22	4.61	0.79
Transformer_CNN	18.56	3.40	0.95	15.97	2.98	0.89	22.46	3.97	0.91	19.26	3.18	0.86

We all know that many stations on the Tibetan Plateau cannot fully measure the 39 variables listed in the importance ranking of Figure 3. To better validate the model's applicability on the Tibetan Plateau, Table 6 present below the imputation results using Transformer_CNN at the QOMS and SETORS sites (29.77° N, 94.73° E, at an altitude of 3327m), with the year 2012 as the test set. The model employs basic meteorological elements, including single-layer air temperature, pressure, single-layer air humidity, single-layer wind speed, single-layer wind direction, site hourly average precipitation, ground net radiation, single-layer soil temperature, and single-layer soil moisture content. As shown in Table 6, the imputation performance using basic meteorological elements is significantly lower than that using all variables, but it still presents a generally good result. When employing basic meteorological elements for data imputation, the Transformer_CNN model consistently outperforms the single Transformer model. This superiority is evident at both the QOMS and SETORS sites. Particularly at the SETORS site, the better imputation performance further validates the high applicability of the Transformer_CNN model on the Tibetan Plateau.

5 Conclusions

During the period from 2007 to 2016, deep learning methods were employed to impute turbulent heat flux observational data for the QOMS site. To optimize predictive performance and simplify model complexity, we first utilized the random forest algorithm to extract features from basic meteorological, turbulent, radiation, and soil data, eliminating redundant data. Subsequently, three traditional statistical methods (MDV, NR, LUP), three machine learning methods (SVM, XGBoost, KNN), and two recurrent neural networks (LSTM, GRU) were employed, along with the deep learning model Transformer introduced in 2017, for model evaluation and comparison. The results indicated that the Transformer exhibited superior performance in imputing turbulent heat fluxes.

To further optimize predictive performance, CNN was introduced and combined with the Transformer, forming a new model named Transformer_CNN. The CNN was designed to capture the periodic change features of turbulent heat fluxes across different time scales, while the Transformer effectively captured long-distance dependencies in time-series data, aiding in revealing intricate temperature variation patterns under environmental driving variables more precisely. Upon evaluation, Transformer_CNN significantly outperformed other traditional artificial intelligence models in terms of predictive performance.

More specifically, Transformer_CNN excelled in predicting H, with a determination coefficient (R^2) for its test set reaching 0.949. It could predict not only low values accurately but also achieved precise predictions as the magnitude of observed values increased, addressing the shortcomings of the traditional Transformer model in predicting higher values. In terms of predicting LE, its test set determination coefficient reached 0.894, effectively resolving the issues of overestimation of low values and underestimation of high values by the Transformer model. In summary, the experimental results thoroughly validated that the Transformer_CNN model provides a novel and efficient solution for imputing turbulent heat fluxes.

Lastly, the Transformer_CNN model was utilized to impute turbulent heat flux data from 2007 to 2016 for the QOMS site. It was found that during non-monsoon periods, H dominated. However, during the summer monsoon season, influenced by the interactions of mid-latitude westerlies and the monsoon, the H decreased and became similar to the latent heat flux. Overall, during the summer, the impacts of H and LE fluxes at the QOMS site were comparable, while the influence of H was more pronounced during non-monsoon periods.

Author contributions

GZ and HQ designed the experiments and carried them out. GZ, HQ and DZ performed data processing, organization, and figure generation. GZ, HQ and DZ wrote the manuscript, and all authors participated in the revision of the paper.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Code and data availability

The dataset and code are both available at https://doi.org/10.5281/zenodo.10005741 (Hou et al., 2023). The local time (UTC+8) was used at the site.

Acknowledgements

We sincerely thank all the scientists, engineers, and students who participated in the field surveys, maintained the measurement instruments, and processed the observational data. We are very grateful to anonymous reviewers for their careful review and valuable comments, which led to substantial improvement of this manuscript.

Financial support

This work was funded by the Second Tibetan Plateau Scientific Expedition and Research Program (Grant 2019QZKK0102) and the Fengyun-3 (03) Batch Meteorological Satellite Project (FY-3 (03) -AS-11.10-ZT, FY-3 (03) -AS-11.12-ZT)..

References

Agarwal, D., Pastorello, G., Poindexter, C., Papale, D., Trotta, C., Ribeca, A., Canfora, E., Faybishenko, B., and Samak, T.: The data post-processing pipeline for AmeriFlux data products, Agu Fall Meeting, ui.adsabs.harvard.edu/abs/2014AGUFM.B53A0159A, 2014.

Alavi, N., Warland, J. S., and Berg, A. A.: Filling gaps in evapotranspiration measurements for water budget studies: Evaluation of a Kalman filtering approach, Agricultural and Forest Meteorology, 141, 57–66, https://doi.org/10.1016/j.agrformet.2006.09.011, 2006.

Arlot, S. and Celisse, A.: A survey of cross-validation procedures for model selection, Statistics Surveys, 4, 40-79, 10.1214/09-SS054, 2010.

Baldocchi, D.: Measuring fluxes of trace gases and energy between ecosystems and the atmosphere - the state and future of the eddy covariance method, Glob. Change Biol., 20, 3600-3609, 10.1111/gcb.12649, 2014.

Baotian, P., Jijun, L., and Fahu, C.: Qinghai-Tibetan Plateau: a Driver and Amplifier of Global Climatic Changes──I Basic Characteristics of Climatic Changes in Cenozoic Era, Journal of Lanzhou University(Natural Sciences), 32, 108-115, 10.13885/j.issn.0455-2059.1995.03.024, 1996.

Bello, I., Zoph, B., Vaswani, A., Shlens, J., Le, Q. V., and Ieee: Attention Augmented Convolutional Networks, IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, SOUTH KOREA, Oct 27-Nov 02, WOS:000531438103044, 3285-3294, 10.1109/iccv.2019.00338, 2019.

Beringer, J., McHugh, I., Hutley, L. B., Isaac, P., and Kljun, N.: Technical note: Dynamic INtegrated Gap-filling and partitioning for OzFlux (DINGO), Biogeosciences, 14, 1457-1460, 10.5194/bg-14-1457-2017, 2017.

Bishop and M, C.: Pattern Recognition and Machine Learning, Springer-Verlag, New York, 10.1007/978-3-030-57077-4_11, 2006.

Cawley, G. C. and Talbot, N. L. C.: On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation, J. Mach. Learn. Res., 11, 2079-2107, 10.5555/1756006.1859921, 2010.

Chapin, F. S., Matson, P. A., and Mooney, H. A.: Principles of Terrestrial Ecosystem Ecology, Springer Verlag, Seconded, New York, USA, 10.1007/978-1-4419-9504-9, 2011.

Chen, B., Chao, W. C., Liu, X. J. C. d. O., theoretical, and system, c. r. o. t. c.: Enhanced climatic warming in the Tibetan Plateau due to doubling CO2: a model study, 20, 10.1007/s00382-003-0308-6, 2003.

Chen, T. Q., Guestrin, C., and Assoc Comp, M.: XGBoost: A Scalable Tree Boosting System, 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, Aug 13-17, WOS:000485529800092, 785-794, 10.1145/2939672.2939785, 2016.

Chen, Y.-Y., Chu, C.-R., and Li, M.-H.: A gap-filling model for eddy covariance latent heat flux: Estimating evapotranspiration of a subtropical seasonal evergreen broad-leaved forest as an example, Journal of Hydrology, 468–469, 101–110, https://doi.org/10.1016/j.jhydrol.2012.08.026, 2012.

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y.: Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), EMNLP 2014, Doha, Qatar, 1724–1734, https://doi.org/10.3115/v1/D14-1179, 2014.

Collatz, G. J., Bounoua, L., Los, S. O., Randall, D. A., Fung, I. Y., and Sellers, P. J.: A mechanism for the influence of vegetation on the response of the diurnal temperature range to changing climate, 27, 3381-3384, 10.1029/1999GL010947, 2000.

Cortes, C. and Vapnik, V.: Support-vector networks, Machine Learning, 20, 273-297, 10.1007/BF00994018, 1995.

Cover, T. and Hart, P.: Nearest neighbor pattern classification, IEEE Transactions on Information Theory, 13, 21-27, 10.1109/TIT.1967.1053964, 1967.

Defries, R. S., Bounoua, L., and Collatz, G. J.: Human modification of the landscape and surface climate in the next fifty years, 8, 438-458, 10.1046/j.1365-2486.2002.00483.x, 2002.

Dengel, S., Zona, D., Sachs, T., Aurela, M., Jammet, M., Parmentier, F. J. W., Oechel, W., and Vesala, T.: Testing the applicability of neural networks as a gap-filling method using CH4 flux data from high latitude wetlands, Biogeosciences, 10, 8185-8200, 10.5194/bg-10-8185-2013, 2013.

Duan, Y. J., Lv, Y. S., Kang, W. W., and Zhao, Y. F.: A Deep Learning Based Approach for Traffic Data Imputation, IEEE 17th International Conference on Intelligent Transportation Systems (ITSC), Qingdao, PEOPLES R CHINA, Oct 08-11, WOS:000357868700163, 912-917, 10.1109/ITSC.2014.6957805, 2014.

Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grunwald, T., Hollinger, D., Jensen, N. O., Katul, G., Keronen, P., Kowalski, A., Lai, C. T., Law, B. E., Meyers, T., Moncrieff, H., Moors, E., Munger, J. W., Pilegaard, K., Rannik, U., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: Gap filling strategies for defensible annual sums of net ecosystem exchange, Agric. For. Meteorol., 107, 43-69, 10.1016/s0168-1923(00)00225-2, 2001.

Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., and Muller, P. A.: Deep learning for Time Series Classification: a review, Data Mining and Knowledge Discovery, 33, 917–963, https://doi.org/10.1007/s10618-019-00619-1, 2019.

Foltynova, L., Fischer, M., and McGloin, R. P.: Recommendations for gap-filling eddy covariance latent heat flux measurements using marginal distribution sampling, Theor. Appl. Climatol., 139, 677-688, 10.1007/s00704-019-02975-w, 2020.

Friedman, J., Hastie, J., and Tibshirani, R.: The elements of statistical learning, Springer, New York, NY, 10.1007/978-0-387-84858-7, 2009.

Gad, I., Hosahalli, D., Manjunatha, B. R., and Ghoneim, O. A.: A robust deep learning model for missing value imputation in big NCDC dataset, Iran Journal of Computer Science, 4, 67-84, 10.1007/s42044-020-00065-z, 2021.

Gao, Z., Chae, N., Kim, J., Hong, J., Choi, T., and Lee, H.: Modeling of surface energy partitioning, surface temperature, and soil wetness in the Tibetan prairie using the Simple Biosphere Model 2 (SiB2), 109, 10.1029/2003JD004089, 2004.

Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning, MIT Press, 2016.

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y.: Deep Learning (Vol. 1), The MIT Press, Cambridge, 10.1038/nature14539, 2016.

He, K., Zhang, X., Ren, S., and Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification, Proceedings of the IEEE International Conference on Computer Vision, 1026–1034, 2015.

Hirsch, R. M., Slack, J. R., and Smith, R. A.: Techniques of trend analysis for monthly water quality data, Water Resour. Res., 18, 107-121, 10.1029/WR018i001p00107, 1982.

Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Comput., 9, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

Hui, D. F., Wan, S. Q., Su, B., Katul, G., Monson, R., and Luo, Y. Q.: Gap-filling missing data in eddy covariance measurements using multiple imputation (MI) for annual estimations, Agric. For. Meteorol., 121, 93-111, 10.1016/s0168-1923(03)00158-8, 2004.

Idrees, S. M., Alam, M. A., Agarwal, P., and Ansari, L.: Effective predictive analytics and modeling based on historical data, in: Advances in computing and data sciences, Singapore, Citation Key: 10.1007/978-981-13-9942-8_52, 552–564, 2019.

Jiao, B., Su, Y., Li, Q., Manara, V., and Wild, M.: An integrated and homogenized global surface solar radiation dataset and its reconstruction based on a convolutional neural network approach, Earth Syst. Sci. Data, 15, 4519-4535, 10.5194/essd-15-4519-2023, 2023.

Knox, S. H., Sturtevant, C., Matthes, J. H., Koteen, L., Verfaillie, J., and Baldocchi, D.: Agricultural peatland restoration: effects of land-use change on greenhouse gas (CO2 and CH4) fluxes in the Sacramento-San Joaquin Delta, Glob. Change Biol., 21, 750-765, 10.1111/gcb.12745, 2015.

Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection, Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 1137-1143, 10.5555/1643031.1643047, 1995.

Krizhevsky, A., Sutskever, I., and Hinton, G.: ImageNet Classification with Deep Convolutional Neural Networks, Advances in neural information processing systems, 25, 1097-1105, 10.1145/3065386, 2012.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P.: Gradient-based learning applied to document recognition, Proceedings of the Ieee, 86, 2278-2324, 10.1109/5.726791, 1998.

Lee, X., Massman, W. J., and Law, B. E.: Handbook of micrometeorology: a guide for surface flux measurement and analysis, 2004.

Li, C., He, H., Liu, M., Su, W., Fu, Y., Zhang, L., Wen, X., and Yu, G.: ChinaFLUX CO2 flux data processing system and its application, Journal of Geo-Information Science, 10, 557-565, 10.3969/j.issn.1560-8999.2008.05.002, 2008.

Little, R. J. A. and Rubin, D. B.: Statistical Analysis with Missing Data, Second Edition, John Wiley & Sons, Inc., Hoboken, New Jersey., 10.1002/9781119013563, 2002.

LI, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X., and Yan, X.: Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting, ArXiv, 2019.

Ma, Y., Hu, Z., Xie, Z., Ma, W., Wang, B., Chen, X., Li, M., Zhong, L., Sun, F., Gu, L., Han, C., Zhang, L., Liu, X., Ding, Z., Sun, G., Wang, S., Wang, Y., and Wang, Z.: A long-term (2005–2016) dataset of hourly integrated land-atmosphere interaction observations on the Tibetan Plateau, Earth Syst. Sci. Data, 12, 2937-2957, 10.5194/essd-12-2937-2020, 2020.

Matusowsky, M., Ramotsoela, D. T., and Abu-Mahfouz, A. M.: Data Imputation in Wireless Sensor Networks Using a Machine Learning-Based Virtual Sensor, J. Sens. Actuar. Netw., 9, 25, 10.3390/jsan9020025, 2020.

Mauder, M., Foken, T., Clement, R., Elbers, J. A., Eugster, W., Grunwald, T., Heusinkveld, B., and Kolle, O.: Quality control of CarboEurope flux data - Part 2: Inter-comparison of eddy-covariance software, Biogeosciences, 5, 451-462, 10.5194/bg-5-451-2008, 2008.

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., Falge, E., Gove, J. H., Heimann, M., Hui, D. F., Jarvis, A. J., Kattge, J., Noormets, A., and Stauch, V. J.: Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes, Agric. For. Meteorol., 147, 209-232, 10.1016/j.agrformet.2007.08.011, 2007.

Ooba, M., Hirano, T., Mogami, J. I., Hirata, R., and Fujinuma, Y.: Comparisons of gap-filling methods for carbon flux dataset: A combination of a genetic algorithm and an artificial neural network, Ecol. Model., 198, 473-486, 10.1016/j.ecolmodel.2006.06.006, 2006.

Ortega, L. C., Otero, L. D., Solomon, M., Otero, C. E., and Fabregas, A.: Deep learning models for visibility forecasting using climatological data, International Journal of Forecasting, 39, 992-1004, 10.1016/j.ijforecast.2022.03.009, 2023.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., and Yakir, D.: Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation, Biogeosciences, 3, 571-583, 10.5194/bg-3-571-2006, 2006.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grunwald, T., Havrankova, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J. M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, Glob. Change Biol., 11, 1424-1439, 10.1111/j.1365-2486.2005.001002.x, 2005.

Shaoying, W., Yu, Z., Xianhong, M., Minhong, S., Lunyu, S., Youqi, S. U., and Zhaoguo, L. I.: Fill the Gaps of Eddy Covariance Fluxes Using Machine Learning Algorithms, Plateau Meteorology, 39, 1348-1360, 10.7522/j.issn.1000-0534.2019.00142, 2020.

Soloway, A. D., Amiro, B. D., Dunn, A. L., and Wofsy, S. C.: Carbon neutral or a sink? Uncertainty caused by gap-filling long-term flux measurements for an old-growth boreal black spruce forest, Agric. For. Meteorol., 233, 110-121, 10.1016/j.agrformet.2016.11.005, 2017.

Stull, R. B.: An Introduction to Boundary Layer Meteorology, Kluwer Academic Publishers Dordrecht., The Netherlands., 10.1007/978-94-009-3027-8, 1988.

Swinbank and W.C.: The measurement of vertical transfer of heat and water vapor by eddies in the lower atmosphere, Journal of Meteorology, 8, 135-145, 10.1175/1520-0469(1951)008<0135:TMOVTO>2.0.CO;2, 1951.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I.: Attention Is All You Need, 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, Dec 04-09, WOS:000452649406008, 5998-6008, 10.5555/3295222.3295349, 2017.

Wang, L., Wan, B., Zhou, S., Sun, H., and Gao, Z.: Forecasting tropical cyclone tracks in the northwestern Pacific based on a deep-learning model, Geosci. Model Dev., 16, 2167–2179, https://doi.org/10.5194/gmd-16-2167-2023, 2023.

Wang, S., Zhang, Y., Lv, S., Ao, Y., Li, S., and Chen, S.: Quality control research of turbulent data in Jinta oasis, Plateau Meteorology, 28, 1260-1273,ir.casnw.net/handle/362004/21874, 2009.

Wang, S., Zhang, Y., Meng, X., Song, M., Shang, L., Su, Y., and LI, Z.: Fill the Gaps of Eddy Covariance Fluxes Using Machine Learning Algorithms , Plateau Meteorology, 39, 1348-1360, 10.7522/j.issn.1000-0534.2019.00142, 2020.

Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Sigut, L., Menzer, O., and Reichstein, M.: Basic and extensible post-processing of eddy covariance flux data with REddyProc, Biogeosciences, 15, 5015-5030, 10.5194/bg-15-5015-2018, 2018.

Wu, N., Green, B., Ben, X., and O'Banion, S.: Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case, https://doi.org/10.48550/arXiv.2001.08317, 22 January 2020.

Yang, K., Wu, H., Qin, J., Lin, C. G., Tang, W. J., and Chen, Y. Y.: Recent climate changes over the Tibetan Plateau and their impacts on energy and water cycle: A review, Glob. Planet. Change, 112, 79-91, 10.1016/j.gloplacha.2013.12.001, 2014.

Yu, G., Chen, Z., Zhang, L., Peng, C., Chen, J., Pu, S., Zhang, Y., Niu, S., Wang, Q., Luo, Y., Ciais, P., and Baldocchi, D. D.: Re-recognizing the scientific mission of flux observation - Laying a solid data foundation for solving global sustainable development ecological issues, Journal of Resources and Ecology (English version), 8, 115-120, 10.5814/j.issn.1674-764x.2017.02.001, 2017.

Yue, S., Pilon, P., Phinney, B., and Cavadias, G.: The influence of autocorrelation on the ability to detect trend in hydrological series, Hydrological Processes, 16, 1807-1829, 10.1002/hyp.1095, 2002.

Zhang, C., Dazlich, D. A., Randall, D. A., Sellers, P. J., and Denning, A. S.: Calculation of the global land surface energy, water and CO2 fluxes with an off-line version of SiB2, 101, 19061-19075, 10.1029/96JD01449, 1996.

Zheng, D., Zhang, Q. s., and Wu, S.: Mountain Geoecology and Sustainable Development of the Tibetan Plateau, Kluwer Academic, Dordrecht, the Netherlands, 10.1007/978-94-010-0965-2, 2000.