Identification and Prediction of Flux Tower Latent Heat Data and Their Source Variables (Time Series Imputation)

This manuscript (permalink) was automatically generated from uiceds/project-team-wres@ee92e82 on November 15, 2024.

Authors

- **Jiaze Cao** ✉
  · JiazeCLeo
  Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign; WRES

- **Yuyao Huang**
  · Yuyao-Huang
  Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign; EWES

- **Yue Wan**
  · clarawan
  Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign; EWES

- **Hsing-Yu Huang** ✉
  · Hsing-Yu
  Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign; WRES

✉ — Correspondence possible via GitHub Issues or email to Jiaze Cao <jiazec2@illinois.edu>, Hsing-Yu Huang <hsingyu3@illinois.edu>.

Dataset Description

We propose to use Goose Creek Eddy Covariance Flux Tower Sensor Data(Kumar & Sargent, 2024). The data is collected from the Eddy Covariance Flux Tower in Goose Creek, Piatt County. The dataset consists of time series data spanning from Spring 2016 to Spring 2023 with 15 minutes time interval. Dataset involves 167 variables shown in Table 1 including latent heat, sensible heat, wind speed, temperature, and changes in the ecosystem with respect to water, carbon, and temperature. Figure 1 illustrates part of variables in 2022. The data collected by flux tower provides a foundation for further investigation into hydrological, meteorological, and environmental phenomena. The format of dataset is CSV file (generated from raw PICKLE file). The dataset can be found through link: https://www.hydroshare.org/resource/c276c71e8d1246e29d8502f5b2054668/

Fig 1: Variables Latent Heat, Sensible Heat, Air Pressure and Vapor Pressure in 2022

Table 1: A table with a variables.

Variable Name	Units	Description
TIMESTAMP	TS	
RECORD	RN	
Hs	W/m^2	sensible heat flux
tau	$kg/(m \cdot s^2)$	shear stress
u_star	m/s	friction velocity

Variable Name	Units	Description
Ts_stdev	$^{\circ}C$	instantaneous stdev of temperature
Ts_Ux_cov	$^{\circ}C \cdot m/s$	inst. cov(temp, Ux)
Ts_Uy_cov	$^{\circ}C \cdot m/s$	inst. cov(temp, Uy)
Ts_Uz_cov	$^{\circ}C \cdot m/s$	inst. cov(temp, Uv)
Ux_stdev	m/s	instantaneous stdev of Ux
Ux_Uy_cov	$(m/s)^2$	instantaneous cov of (Ux, Uy)
Ux_Uz_cov	$(m/s)^2$	instantaneous cov of (Ux, Uz)
Uy_stdev	m/s	instantaneous stdev of Uy
Uy_Uz_cov	$(m/s)^2$	instantaneous cov of (Uy, Uz)
Uz_stdev	m/s	instantaneous stdev of Uz
wnd_spd	m/s	wind speed (horizontal) - different from next?
rslt_wnd_spd	m/s	wind speed (horizontal)
wnd_dir_sonic	$degrees$	wind direction from CSAT3, deg from N?
std_wnd_dir	$degrees$	inst. stdev of wind direction
wnd_dir_compass	$degrees$	wind direction from compass (from N?)
Ux_Avg	m/s	average horiz windspeed x
Uy_Avg	m/s	average horiz windspeed y
Uz_Avg	m/s	average vertical windspeed z
Ts_Avg	$^{\circ}C$	air temperature at 25 m
sonic_azimuth	$degrees$	180 is direction is pointing - can change this value
sonic_samples_Tot	samples	10 Hz sampling rate (cycles per 15 mins = 9000)
Fc_li_wpl	$mg/(m^2 \cdot s)$	carbon flux upward (+ = upward) with Webb et al Term
LE_li_wpl	W/m^2	latent heat flux with Webb et al term
Hc_li	W/m^2	sensible heat flux
CO2_li_mean	mg/m^3	CO2 conc –> need to convert to ppm units
H2O_li_mean	g/m^3	water vapor conc at 25 m
amb_press_li_mean	kPa	air pressure at 25 m
Tc_li_mean	$^{\circ}C$	CSAT air temperature at 25 m
rho_a_li_mean	kg/m^3	density of air with water vapor
Fc_li_irga	$mg/(m^2 \cdot s)$	carbon flux without Webb et al. Term
LE_li_irga	W/m^2	latent heat flux without Webb et al. Term
irga_li_samples_Tot	$samples$	should be around 60 - quality indicator of LiCor

Variable Name	Units	Description
Precip_Tot	mm	rainfall
T_tmpr_rh_mean	$^{\circ}C$	air temperature at 25 m
e_tmpr_rh_mean	kPa	vapor pressure at 25 m
e_sat_tmpr_rh_mean	kPa	saturated vapor pressure at 25 m
H2O_tmpr_rh_mean	g/m^3	water vapor conc at 25 m
RH_tmpr_rh_mean		Relative Humidity at 25 m (e/e_sat)
rho_a_tmpr_rh_mean	kg/m^3	air density
slowsequence_1_Tot	$samples$	cycles per 15 mins - scanning every 10 secs
CS655_Wcr_Avg	m^3/m^3	soil water content
CS655_Ec_Avg	dS/m	soil conductivity
CS655_Tmpr_Avg	$^{\circ}C$	soil temperature
mean_wind_speed	m/s	wind speed at 10 m height...not average?
mean_wind_direction	$degrees$	wind direction
std_wind_dir	$degrees$	mean wind vector stdev of direction
NDVI_Avg		Normalized Difference Vegetation Index
NDVIUpRed_Avg	$W/m^2 \cdot nm$	NDVI is calculated from upward and canopy facing sensors that measure IR and NIR radiation
NDVIUpNIR_Avg	$W/m^2 \cdot nm$	NDVI is calculated from upward and canopy facing sensors that measure IR and NIR radiation
NDVIIndUp		NDVI is calculated from upward and canopy facing sensors that measure IR and NIR radiation
NDVIDownRed_Avg	$W/m^2 \cdot nm$	NDVI is calculated from upward and canopy facing sensors that measure IR and NIR radiation
NDVIDownNIR_Avg	$W/m^2 \cdot nm$	NDVI is calculated from upward and canopy facing sensors that measure IR and NIR radiation
NDVIIndDown		NDVI is calculated from upward and canopy facing sensors that measure IR and NIR radiation
PRI_Avg		Photochemical Reflectance Index
PRIUp531_Avg	$W/m^2 \cdot nm$	PRI calculated from updward and canopy facing sensors that measure 2 wavelengths of radiation
PRIUp570_Avg	$W/m^2 \cdot nm$	PRI calculated from updward and canopy facing sensors that measure 2 wavelengths of radiation

Variable Name	Units	Description
PRIIndUp		PRI calculated from updward and canopy facing sensors that measure 2 wavelengths of radiation
PRIDown531_Avg	$W/m^2 \cdot nm$	PRI calculated from updward and canopy facing sensors that measure 2 wavelengths of radiation
PRIDown570_Avg	$W/m^2 \cdot nm$	PRI calculated from updward and canopy facing sensors that measure 2 wavelengths of radiation
PRIIndDown		PRI calculated from updward and canopy facing sensors that measure 2 wavelengths of radiation
D5TE_VWC_5cm_Avg	m^3/m^3	volumetric water content
D5TE_P_5cm_Avg		bulk dielectric permittivity
D5TE_EC_5cm_Avg	dS/m	soil electrical conductivity
D5TE_T_5cm_Avg	$^{\circ}C$	soil temperature
D5TE_VWC_15cm_Avg	m^3/m^3	volumetric water content
D5TE_P_15cm_Avg		bulk dielectric permittivity
D5TE_EC_15cm_Avg	dS/m	soil conductivity
D5TE_T_15cm_Avg	$^{\circ}C$	soil temperature
D5TE_VWC_30cm_Avg	m^3/m^3	volumetric water content
D5TE_P_30cm_Avg		bulk dielectric permittivity
D5TE_EC_30cm_Avg	dS/m	soil conductivity
D5TE_T_30cm_Avg	$^{\circ}C$	soil temperature
D5TE_VWC_50cm_Avg	m^3/m^3	volumetric water content
D5TE_P_50cm_Avg		bulk dielectric permittivity
D5TE_EC_50cm_Avg	dS/m	soil conductivity
D5TE_T_50cm_Avg	$^{\circ}C$	soil temperature
D5TE_VWC_100cm_Avg	m^3/m^3	volumetric water content
D5TE_P_100cm_Avg		bulk dielectric permittivity

Variable Name	Units	Description
D5TE_EC_100cm_Avg	dS/m	soil conductivity
D5TE_T_100cm_Avg	$^{\circ}C$	soil temperature
D5TE_VWC_200cm_Avg	m^3/m^3	volumetric water content
D5TE_P_200cm_Avg		bulk dielectric permittivity
D5TE_EC_200cm_Avg	dS/m	soil conductivity
D5TE_T_200cm_Avg	$^{\circ}C$	soil temperature
slowsequence_2_Tot	$samples$	cycles - 1 minute loops (number of times scanned)
SB121TempC_Avg	$^{\circ}C$	SB = sensor body, temp of body of sensor
Targ121TempC_Avg	$^{\circ}C$	surface temperature
Targ121mV_Avg	$^{\circ}C$	
SB1H1TempC_Avg	$^{\circ}C$	SB = sensor body, temp of body of sensor
Targ1H1TempC_Avg	$^{\circ}C$	surface temperature
Targ1H1mV_Avg	$^{\circ}C$	
short_up_Avg	W/m^2	Incoming shortwave radiation detected by the upward facing instrument
short_dn_Avg	W/m^2	Outgoing shortwave radiation detected by the downward facing instrument
long_up_Avg	W/m^2	incoming longwave radiation detected by upward facing instrument
long_dn_Avg	W/m^2	outgoing longwave radiation detected by downward facing instrument
cnr4_T_C_Avg	$^{\circ}C$	temperature of sensor
cnr4_T_K_Avg	K	temperature of sensor in Kelvin
long_up_corr_Avg	W/m^2	Incoming longwave radiation detected by the upward facing instrument, corrected
long_dn_corr_Avg	W/m^2	Outgoing longwave radiation detected by the downward facing instrument, corrected
Rs_net_Avg	W/m^2	Shortwave net radiation (Rshort_up - Rshort_down)
Rl_net_Avg	W/m^2	Longwave net radiation (Rlong_up - Rlong_down)
albedo_Avg	W/m^2	Albedo
Rn_Avg	W/m^2	Net radiation (Rs_net + Rl_net)
SQ_110_Avg	μmol photons $/(m^2 \cdot s)$	PAR (photosynthetically active radiation)

Variable Name	Units	Description
shf_Avg(1)	W/m^2	Ground heat flux
shf_Avg(2)	W/m^2	Ground heat flux
slowsequence _3_Tot	$samples$	number of times scanned in 15 mins (once per min)

Proposal

Background

Evapotranspiration (ET) is the process of water transferring from land to the atmosphere, accompanying the phase change of water from liquid to gas. This process plays a critical role in the ecohydrological system and profoundly affects the hydrological cycle. The processes of evapotranspiration and energy exchange are interdependent. Both latent heat (LE) and evapotranspiration (ET), from the perspective of energy and water flux, are key terms for anticipating weather conditions, simulating climate, and diagnosing climate change. However, the measurement of evapotranspiration is challenging because the process itself is invisible and complex.

Figure 2 shows the latent heat data gap in 2020 due to covid-19 and overhaul of equipment. Our project goal is to fill in these missing data. The ground truth data is collected from satelite sensors (https://etdata.org/). Despite the existence of numerous classical evapotranspiration simulation models, such as Bowen Ratio, Priestley-Taylor and Penman-Monteith models, the predictive accuracy of these models is inferior to that of deep learning models. Therefore, we plan to use RNN and LSTM deep learning models to predict latent heat and fill the gap.

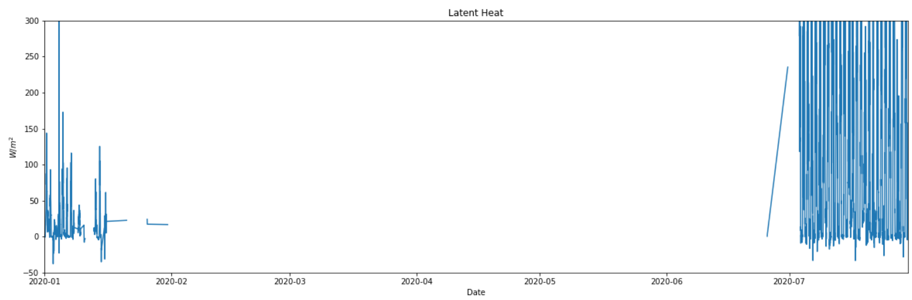


Fig 2: Data Gap in 2020

Step 1: Exploratory Data Analysis

We have 167 variables in the dataset. Although we can filter some ET related variables based on empirical models, these variables may not accurate and AI models tend to obtain adequate information. Therefore, we propose to conduct regression analysis to find out variables highly correlated to latent heat. These variables will be input variables in deep learning model. To simplify the process of figuring out how these factors interact and vary across different times of the day and seasons, we focus on the first five data of main variables first. The main variables include Timestamp, Observed latent heat flux (LE_Obs), Net radiation (Rn), Temperature in degrees Celsius (T), Air_Pressure and etc. Table 2 shows the first 5 data of each variable as follows:

Step 2: Predictive Modeling (Deep Learning Time Series Forecast, Time Series Imputation)

Once we confirm the input variables, we plan to use RNN or LSTM forecast models to predict latent heat in 2020. All the input are divided into training datasets and the validation datasets. After the RNN model is trained, the validation datasets are used to verify the model. At last, the missing data are generated by the model.For Deliverable 3, the goal is to create a predictive model to estimate missing values or forecast future measurements, specifically targeting `LE_Obs` (Latent Heat Flux), due to its environmental relevance. The model plan includes:

1. **Problem Definition**: Use time-series forecasting to predict `LE_Obs` based on the past sequences of environmental variables. Predicting `LE_Obs` could help estimate latent heat flux in unmeasured periods, contributing to more complete energy balance data.
2. **Feature Selection**: Include variables closely related to `LE_Obs`, such as `Rn`, `T`, `Relative Humidity`, and `Sensible_H`. These will serve as predictive features as they capture the interactions of surface energy components.
3. **Model Choice**: Implement an LSTM (Long Short-Term Memory) model due to its suitability for time-series data, capturing sequential dependencies and trends. Using an LSTM allows the model to learn from past data sequences, making it appropriate for forecasting latent heat flux.
4. **Evaluation Metric**: Use Mean Squared Error (MSE) to evaluate the model's performance, as it penalizes larger errors and is commonly used in regression tasks.
5. **Training Plan**: Divide the dataset into training, validation, and test sets, ensuring that the test set consists of the most recent data. Train the model on scaled data to improve performance and use the validation set for hyperparameter tuning.

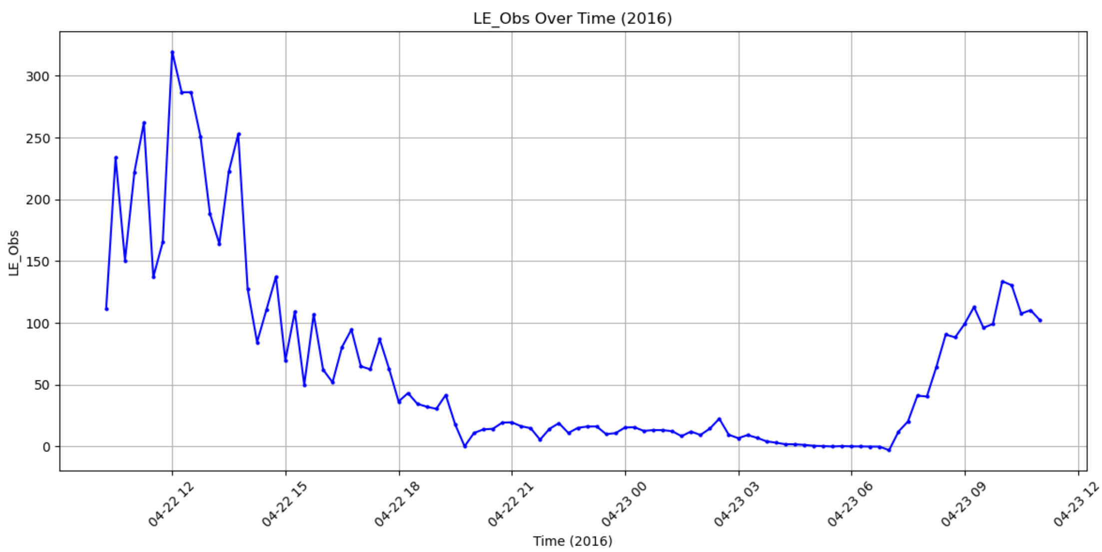


Fig 3: LE_Obs over time for the year 2016

Table 2: Dataset

Timestamp	LE_Obs	Rn	T	Air_Pressure	G	VWC_30cm	VWC30cm_diff	Sensible_H	Relative_Humidity	Vapor_pressure	Saturated_Vapor_Pressure
2016/4/22 10:15	111.1646	550.7244	15.39692	98.65031	-3.305334	0.307	NaN	66.62987	75.43023	1.318720	1.748264

Timestamp	LE_Obs	Rn	T	Air_Pressure	G	VWC_30cm	VWC30cm_diff	Sensible_H	Relative_Humidity	Vapor_pressure	Saturated_Vapor_Pressure
2016/4/22 10:30	233.7630	566.5699	15.75987	98.67207	-1.070949	0.307	0.0	97.43565	74.24434	1.328549	1.789427
2016/4/22 10:45	150.6655	568.7559	15.84722	98.69296	1.211160	0.307	0.0	74.73856	74.90156	1.347835	1.799476
2016/4/22 11:00	222.2112	631.6904	16.14020	98.71219	3.689381	0.307	0.0	89.94750	77.31715	1.417601	1.833488
2016/4/22 11:15	261.9899	704.8135	16.37151	98.72832	6.272957	0.307	0.0	99.07247	78.72035	1.464757	1.860709

Table 3: Correlation matrix

	LE_Obs	Rn	T	Air_Pressure	G	VWC_30cm	VWC30cm_diff	Sensible_H	Relative_Humidity	Vapor_pressure	Saturated_Vapor_Pressure
LE_Obs	1.000000	0.740001	0.369080	-0.066791	0.419244	-0.023582	-0.037540	0.390064	-0.288203	0.278373	0.425352
Rn	0.740001	1.000000	0.354361	0.018251	0.412415	-0.040560	-0.059480	0.642862	-0.416080	0.198780	0.407578
T	0.369080	0.354361	1.000000	-0.365036	0.642930	-0.166246	-0.001006	0.171602	-0.262734	0.866856	0.959008
Air_Pressure	-0.066791	0.018251	-0.365036	1.000000	-0.313725	-0.071028	-0.020830	0.103935	-0.167686	-0.348514	-0.289969
G	0.419244	0.412415	0.642930	-0.313725	1.000000	-0.024829	0.058169	0.247713	-0.382420	0.473200	0.689871
VWC_30cm	-0.023582	-0.040560	-0.166246	-0.071028	-0.024829	1.000000	-0.043053	-0.074545	0.144828	-0.110267	-0.177017
VWC30cm_diff	-0.037540	-0.059480	-0.001006	-0.020830	0.058169	-0.043053	1.000000	-0.049210	-0.000121	-0.009242	-0.003133
Sensible_H	0.390064	0.642862	0.171602	0.103935	0.247713	-0.074545	-0.049210	1.000000	-0.395557	0.011293	0.210560
Relative_Humidity	-0.288203	-0.416080	-0.262734	-0.167686	-0.382420	0.144828	-0.000121	-0.395557	1.000000	0.149367	-0.311806
Vapor_pressure	0.278373	0.198780	0.866856	-0.348514	0.473200	-0.110267	-0.009242	0.011293	0.149367	1.000000	0.863301

	LE_Obs	Rn	T	Air_Pressure	G	VWC_30cm	VWC30cm_diff	Sensible_H	Relative_Humidity	Vapor_pressure	Saturated_Vapor_Pressure
Saturated_Vapor_Pressure	0.425352	0.407578	0.959008	-0.289969	0.689871	-0.177017	-0.003133	0.210560	-0.311806	0.863301	1.000000

Table 4: Statistics of variables in the dataset

	LE_Obs	Rn	T	Air_Pressure	G	VWC_30cm	VWC30cm_diff	Sensible_H	Relative_Humidity	Vapor_pressure	Saturated_Vapor_Pressure
mean	53.976825	109.999297	12.524988	98.813399	-0.159915	0.308942	-0.000015	13.962563	71.872526	1.218452	1.783960
std	94.877870	227.819402	11.105505	0.670804	16.345332	0.033217	0.000257	72.768378	17.570532	0.743564	1.113461

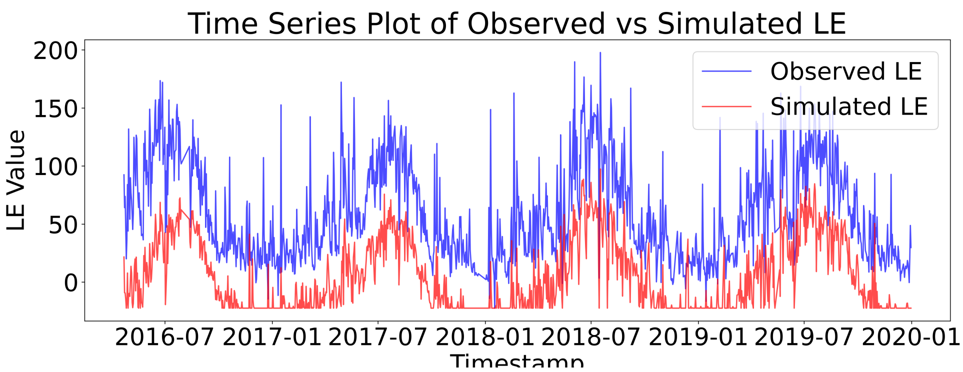
Predictive Modeling

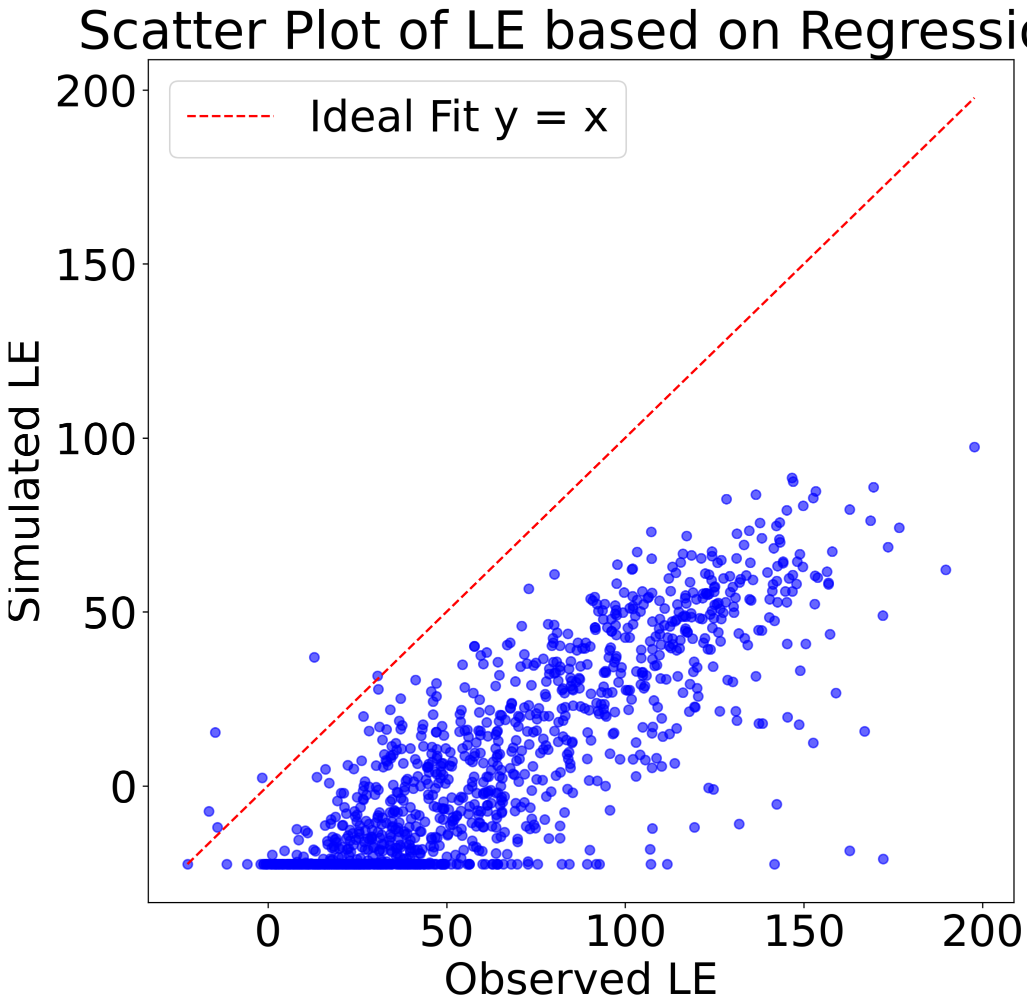
In this section, we propose to use Machine Learning methods to simulate target variable latent heat. Source variables are shown in the above table. Specifically, our methods include regression, single layer neural network, multi layer neural network, CNN network and LSTM. Figures and results are shown in following content including the scatter, time series plot, R^2, MSE value between observed and simulated latent heat.

Regression

Based on gradient descent, we look for a vector $beta$ which can minimize the difference between simulated and observed value. Since our system has 9 source variables, we want to obtain a vector with 9 values which are the coefficient for each source variable. The equation is,
$LE = \beta_1 Rn + \beta_2 T + \beta_3 Pa + \beta_4 G + \beta_5 VWC + \beta_6 VWC \text{ diff} + \beta_7 H + \beta_8 RH + \beta_9 VP$

At beginning, we didn't apply any processing on our data and we input the original data directly.





References

Kumar, P., & Sargent, S. (2024). Goose Creek Eddy Covariance Flux Tower Sensor Data - Sep 2020-ongoing. HydroShare. http://www.hydroshare.org/resource/c276c71e8d1246e29d8502f5b2054668