# 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@38bac40 on October 28, 2024.

#### **Authors**

- Jiaze Cao <sup>™</sup>
  - · 🕝 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
  - · 🕠 <u>clarawan</u>

Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign; EWES

- Hsing-Yu Huang <sup>™</sup>
  - · G Hsing-Yu

Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign; WRES

■ — Correspondence possible via <u>GitHub Issues</u> 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: <a href="https://www.hydroshare.org/resource/c276c71e8d1246e29d8502f5b2054668/">https://www.hydroshare.org/resource/c276c71e8d1246e29d8502f5b2054668/</a>

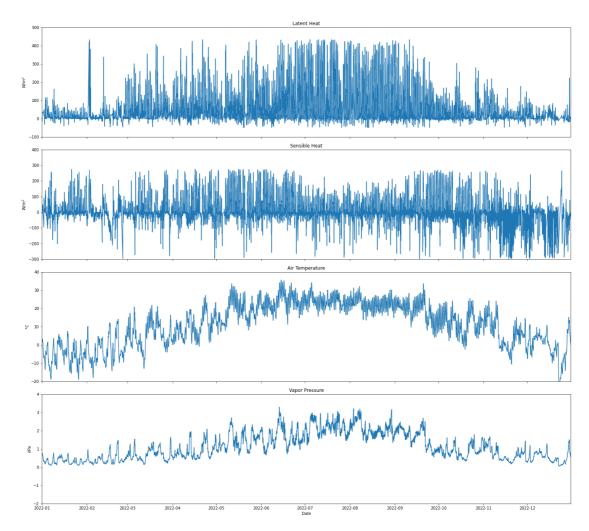


Fig 1: Variables Latent Heat, Sensible Heat, Air Pressure and Vapor Pressure in 2022

**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_comp ass	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_sample s_Tot	samples	should be around 60 - quality indicator of LiCor				

Variable Name	Units	Description
Precip_Tot	mm	rainfall
T_tmpr_rh_me an	$^{\circ}C$	air temperature at 25 m
e_tmpr_rh_me an	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_m ean		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_Av	$m^3/m^3$	soil water content
CS655_Ec_Avg	dS/m	soil conductivity
CS655_Tmpr_A vg	$^{\circ}C$	soil temperature
mean_wind_sp eed	m/s	wind speed at 10 m heightnot average?
mean_wind_di rection	degrees	wind direction
std_wind_dir	degrees	mean wind vector stdev of direction
NDVI_Avg		Normalized Difference Vegetation Index
NDVIUpRed_A vg	$W/m^2 \cdot nm$	NDVI is calculated from upward and canopy facing sensors that measure IR and NIR radiation
NDVIUpNIR_Av	$W/m^2 \cdot nm$	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_5c m_Avg	$m^3/m^3$	volumetric water content
D5TE_P_5cm_A vg		bulk dielectric permittivity
D5TE_EC_5cm_ Avg	dS/m	soil electrical conductivity
D5TE_T_5cm_A vg	$^{\circ}C$	soil temperature
D5TE_VWC_15 cm_Avg	$m^3/m^3$	volumetric water content
D5TE_P_15cm_ Avg		bulk dielectric permittivity
D5TE_EC_15c m_Avg	dS/m	soil conductivity
D5TE_T_15cm_ Avg	$^{\circ}C$	soil temperature
D5TE_VWC_30 cm_Avg	$m^3/m^3$	volumetric water content
D5TE_P_30cm_ Avg		bulk dielectric permittivity
D5TE_EC_30c m_Avg	dS/m	soil conductivity
D5TE_T_30cm_ Avg	$^{\circ}C$	soil temperature
D5TE_VWC_50 cm_Avg	$m^3/m^3$	volumetric water content
D5TE_P_50cm_ Avg		bulk dielectric permittivity
D5TE_EC_50c m_Avg	dS/m	soil conductivity
D5TE_T_50cm_ Avg	$^{\circ}C$	soil temperature
D5TE_VWC_10 0cm_Avg	$m^3/m^3$	volumetric water content
D5TE_P_100c m_Avg		bulk dielectric permittivity

Variable Name	Units	Description
D5TE_EC_100c m_Avg	dS/m	soil conductivity
D5TE_T_100cm _Avg	$^{\circ}C$	soil temperature
D5TE_VWC_20 0cm_Avg	$m^3/m^3$	volumetric water content
D5TE_P_200c m_Avg		bulk dielectric permittivity
D5TE_EC_200c m_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
Targ121Temp C_Avg	$^{\circ}C$	surface temperature
Targ121mV_Av	$^{\circ}C$	
SB1H1TempC_ Avg	$^{\circ}C$	SB = sensor body, temp of body of sensor
Targ1H1Temp C_Avg	$^{\circ}C$	surface temperature
Targ1H1mV_A vg	$^{\circ}C$	
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	$\mu  ext{mol photons} / (m^2 \cdot s)$	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 (<a href="https://etdata.org/">https://etdata.org/</a>). 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.

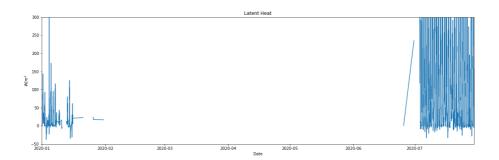


Fig 2: Data Gap in 2020

#### **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\_0bs (Latent Heat Flux), due to its environmental relevance. The model plan includes:

- 1. **Problem Definition**: Use time-series forecasting to predict LE\_0bs based on the past sequences of environmental variables. Predicting LE\_0bs 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 timeseries 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.

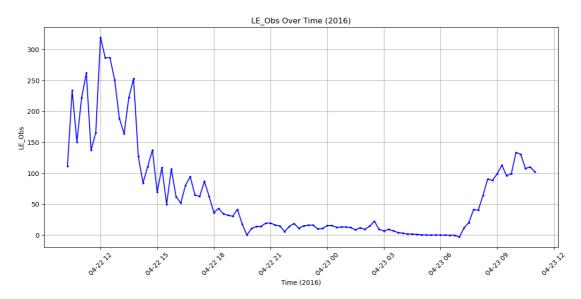


Fig 3: LE\_Obs over time for the year 2016

Table 2: Dataset

Timestam p	LE_ Obs	Rn	Т	Air_Pre ssure	G	VW C_3 Oc m	VWC30 cm_diff	Sensi ble_H	Relative_H umidity	Vapor_pr essure	Saturated_Vapo r_Pressure
2016/4/22 10:15	111. 164 6	550. 724 4	15.3 969 2	98.6503 1	-3.30 5334	0.30 7	NaN	66.629 87	75.43023	1.318720	1.748264

Timestam p	LE_ Obs	Rn	Т	Air_Pre ssure	G	VW C_3 0c m	VWC30 cm_diff	Sensi ble_H	Relative_H umidity	Vapor_pr essure	Saturated_Vapo r_Pressure
2016/4/22 10:30	233. 763 0	566. 569 9	15.7 598 7	98.6720 7	-1.07 0949	0.30 7	0.0	97.435 65	74.24434	1.328549	1.789427
2016/4/22 10:45	150. 665 5	568. 755 9	15.8 472 2	98.6929 6	1.211 160	0.30 7	0.0	74.738 56	74.90156	1.347835	1.799476
2016/4/22 11:00	222. 211 2	631. 690 4	16.1 402 0	98.7121 9	3.689 381	0.30 7	0.0	89.947 50	77.31715	1.417601	1.833488
2016/4/22 11:15	261. 989 9	704. 813 5	16.3 715 1	98.7283 2	6.272 957	0.30 7	0.0	99.072 47	78.72035	1.464757	1.860709

**Table 3:** Correlation matrix

	LE_ Obs	Rn	Т	Air_Pre ssure	G	VW C_3 0cm	VWC30c m_diff	Sensi ble_ H	Relative_ Humidity	Vapor_pr essure	Saturated_Vap or_Pressure
LE_Obs	1.00 000 0	0.74 000 1	0.36 908 0	-0.0667 91	0.41 924 4	-0.0 235 82	-0.0375 40	0.390 064	-0.288203	0.278373	0.425352
Rn	0.74 000 1	1.00 000 0	0.35 436 1	0.01825 1	0.41 241 5	-0.0 405 60	-0.0594 80	0.642 862	-0.416080	0.198780	0.407578
т	0.36 908 0	0.35 436 1	1.00 000 0	-0.3650 36	0.64 293 0	-0.1 662 46	-0.0010 06	0.171 602	-0.262734	0.866856	0.959008
Air_Pressure	-0.0 667 91	0.01 825 1	-0.3 650 36	1.00000	-0.3 137 25	-0.0 710 28	-0.0208 30	0.103 935	-0.167686	-0.34851 4	-0.289969
G	0.41 924 4	0.41 241 5	0.64 293 0	-0.3137 25	1.00 000 0	-0.0 248 29	0.05816 9	0.247 713	-0.382420	0.473200	0.689871
VWC_30cm	-0.0 235 82	-0.0 405 60	-0.1 662 46	-0.0710 28	-0.0 248 29	1.00 000 0	-0.0430 53	-0.07 4545	0.144828	-0.11026 7	-0.177017
VWC30cm_di ff	-0.0 375 40	-0.0 594 80	-0.0 010 06	-0.0208 30	0.05 816 9	-0.0 430 53	1.00000	-0.04 9210	-0.000121	-0.00924 2	-0.003133
Sensible_H	0.39 006 4	0.64 286 2	0.17 160 2	0.10393 5	0.24 771 3	-0.0 745 45	-0.0492 10	1.000 000	-0.395557	0.011293	0.210560
Relative_Hum idity	-0.2 882 03	-0.4 160 80	-0.2 627 34	-0.1676 86	-0.3 824 20	0.14 482 8	-0.0001 21	-0.39 5557	1.000000	0.149367	-0.311806
Vapor_pressu re	0.27 837 3	0.19 878 0	0.86 685 6	-0.3485 14	0.47 320 0	-0.1 102 67	-0.0092 42	0.011 293	0.149367	1.000000	0.863301

	LE_ Obs	Rn	Т	Air_Pre ssure	G	VW C_3 0cm	VWC30c m_diff	Sensi ble_ H	Relative_ Humidity	Vapor_pr essure	Saturated_Vap or_Pressure
Saturated_Va por_Pressure	0.42 535 2	0.40 757 8	0.95 900 8	-0.2899 69	0.68 987 1	-0.1 770 17	-0.0031 33	0.210 560	-0.311806	0.863301	1.000000

**Table 4:** Statistics of variables in the dataset

	LE_Ob	Rn	Т	Air_Pr essur e	G	VWC _30c m	VWC3 0cm_d iff	Sensi ble_ H	Relative_ Humidity	Vapor_ pressur e	Saturated_Vap or_Pressure
mean	53.976 825	109.99 9297	12.524 988	98.813 399	-0.159 915	0.308 942	-0.000 015	13.96 2563	71.872526	1.21845 2	1.783960
std	94.877 870	227.81 9402	11.105 505	0.6708 04	16.345 332	0.033 217	0.0002 57	72.76 8378	17.570532	0.74356 4	1.113461

#### References

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