

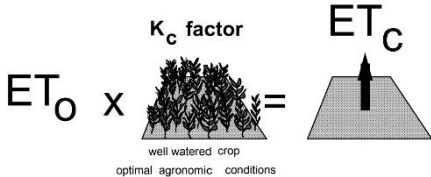
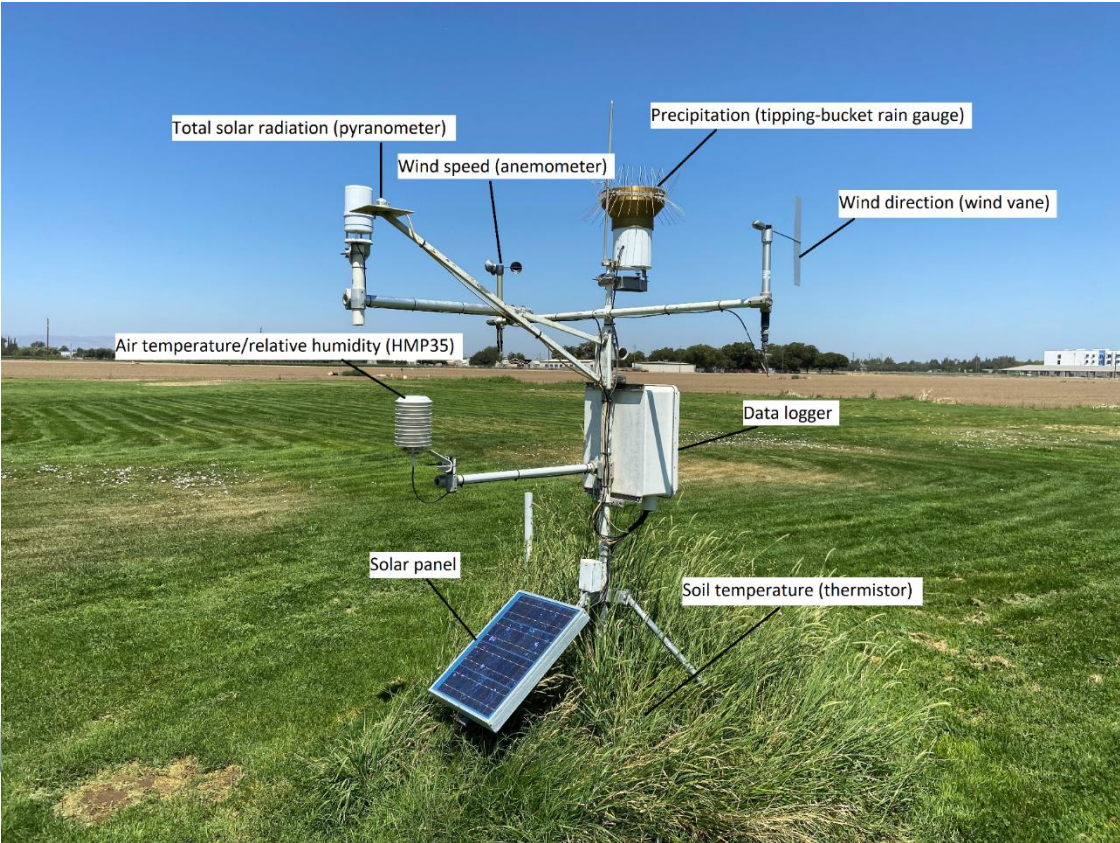
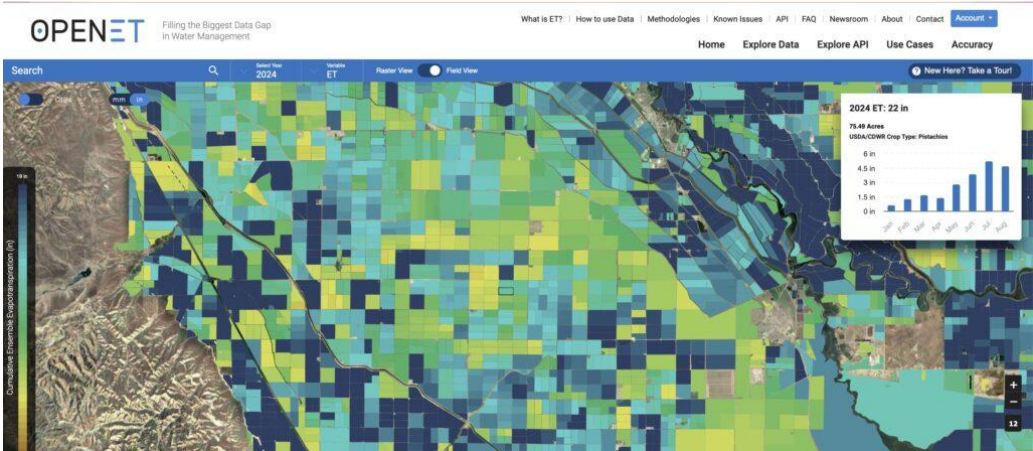
Estimation and Forecasting of Reference Evapotranspiration by Machine Learning

Andre Daccache
adaccache@ucdavis.edu

Why ET Is So Important?

- ✓ Important component of the hydrological cycle
 - ✓ About 2/3 of precipitation returns to the atmosphere through ET
- ✓ An important consumer of the atmosphere's energy
 - ✓ Regulates surface temperature and local weather patterns
 - ✓ Climate Change Indicator
- ✓ Ecosystem Health: Influences vegetation stress, soil moisture, and biodiversity
- ✓ Water Resource Management
 - ✓ Critical for reservoir, drought, and flood planning
- ✓ Agricultural Importance:
 - ✓ Drives irrigation scheduling and crop water demand

Reference (ET_o), Potential (ET_c), and Actual ET (ET_a)

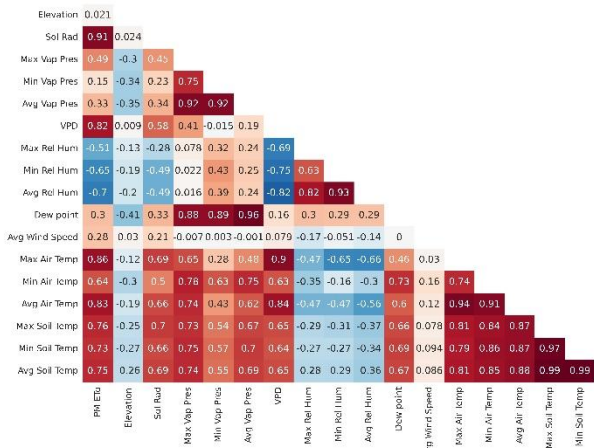
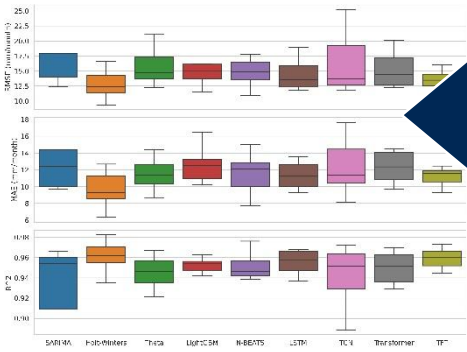
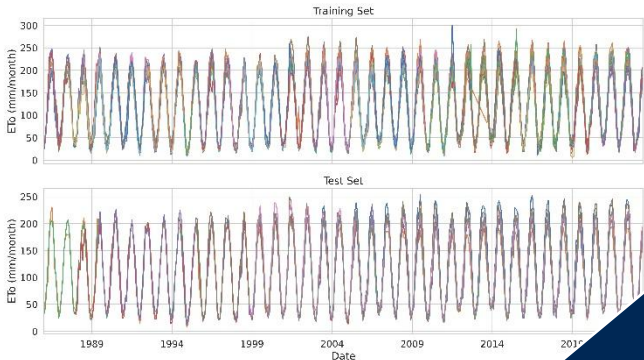


$$ET = \frac{\Delta \cdot (R_n - G) + \rho_a \cdot c_p \cdot \frac{(e_s - e_a)}{r_a}}{\lambda \cdot \left[\Delta + \gamma \cdot \left(1 + \frac{r_s}{r_a} \right) \right]}$$

$$ET_o = \frac{0.408 \Delta (R_n - G) + \gamma \frac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma (1 + 0.34 u_2)}$$

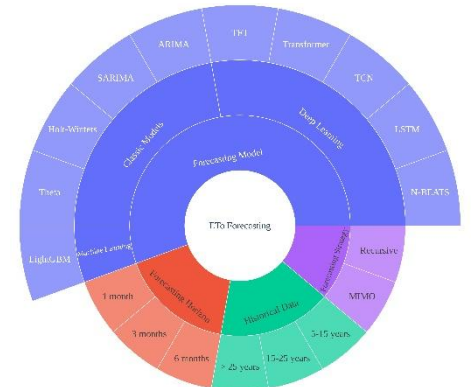
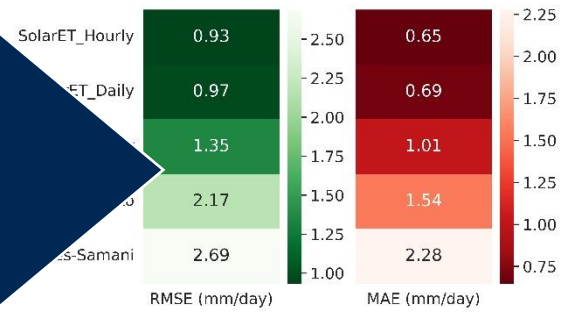
- Reference surface:
- Hypothetical grass reference crop
 - Crop height of 0.12 m
 - Fixed surface resistance of 70 s m⁻¹
 - Albedo of 0.23

Research Approach



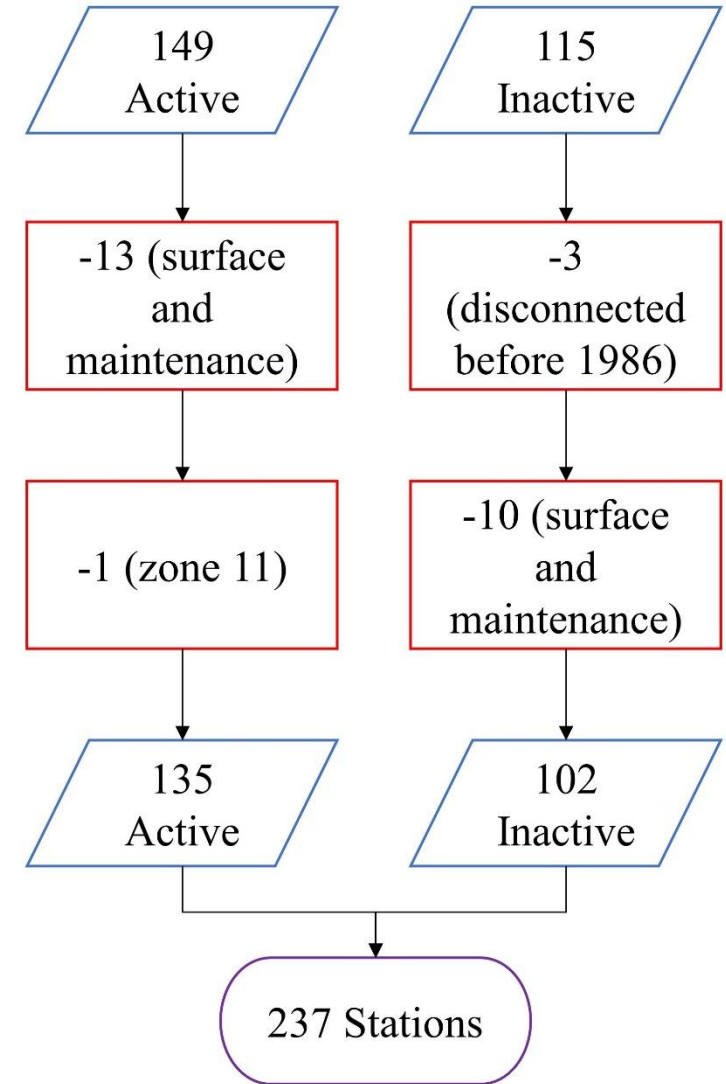
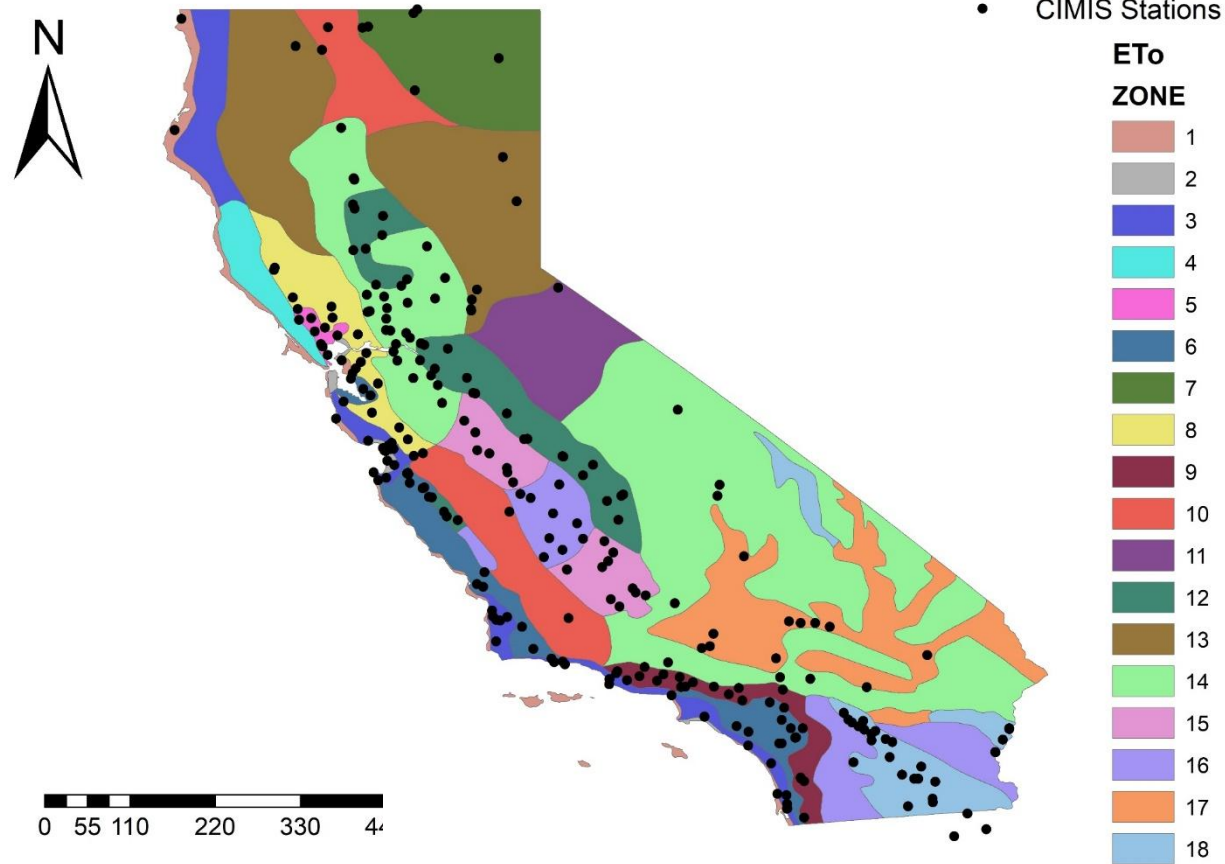
Exploratory Data Analysis

Data-driven Modeling



Data & Quality Control

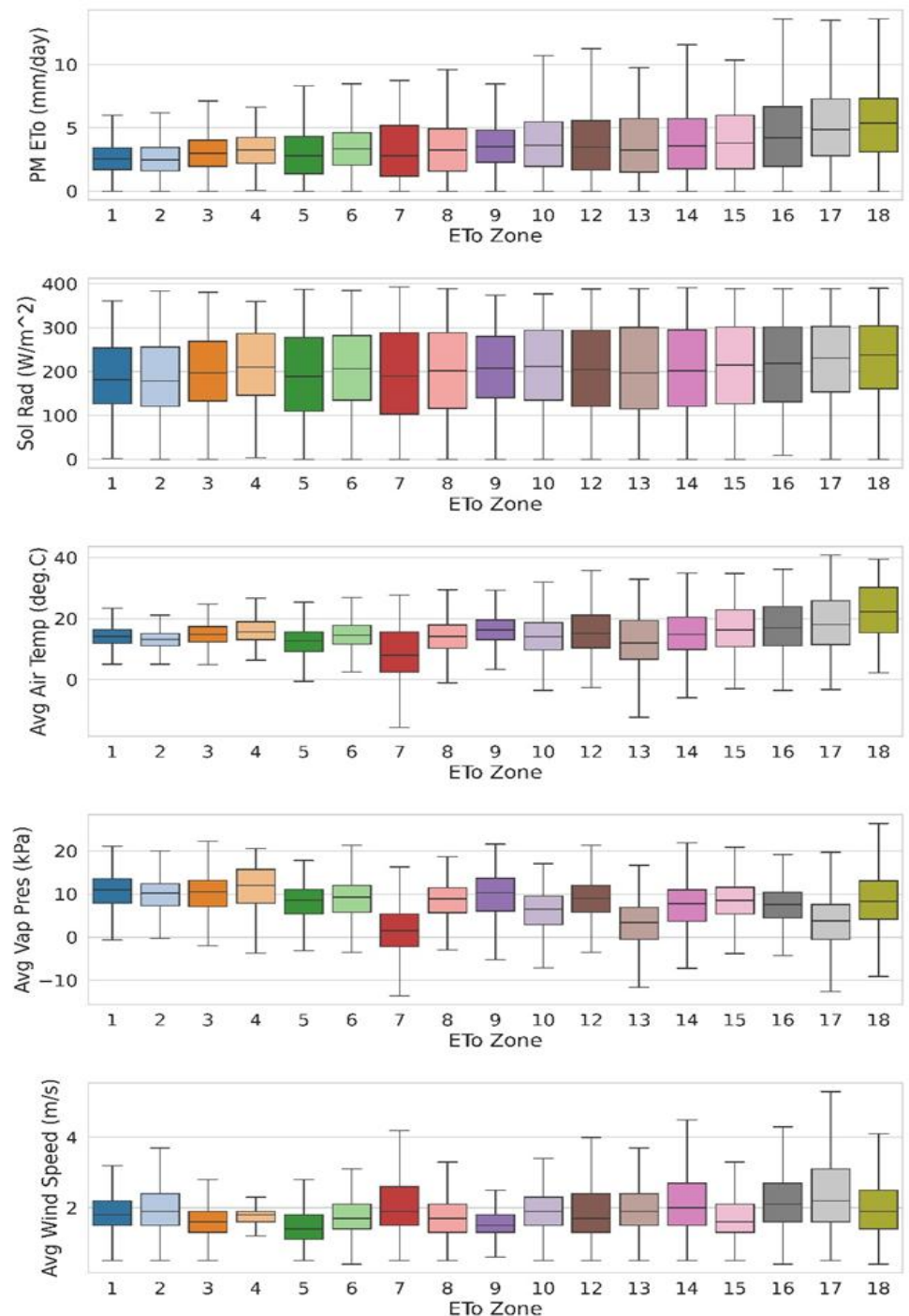
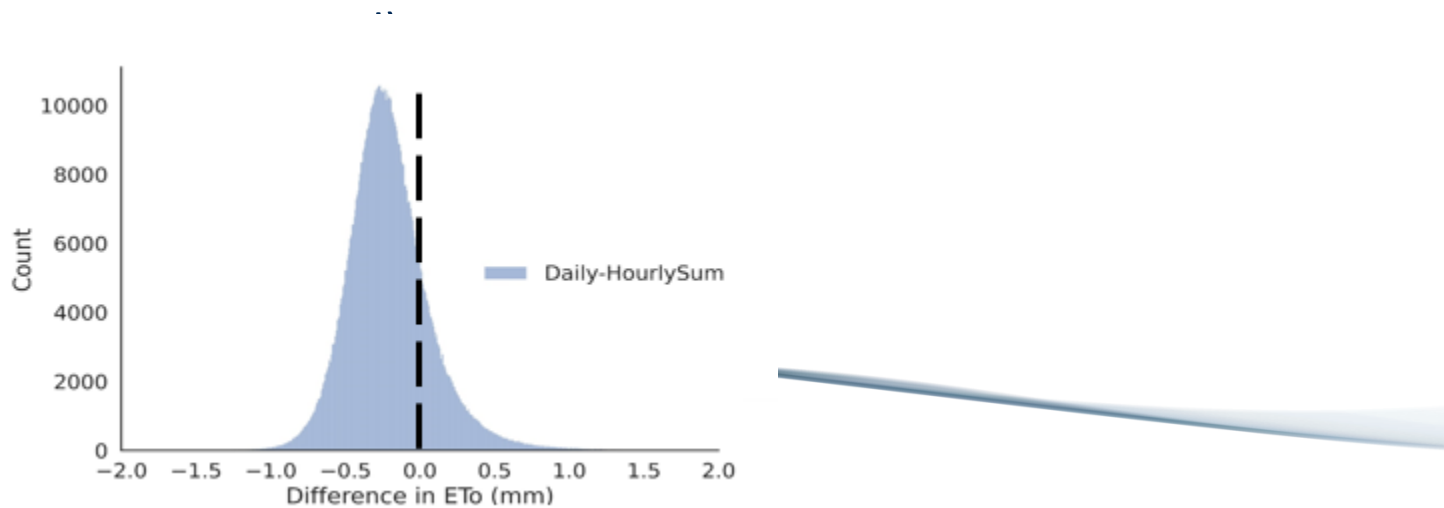
1,365,823 daily data samples from 237 CIMIS stations, from 1986 to 2021 (36 years of daily data)



Daily vs. Hourly Summation

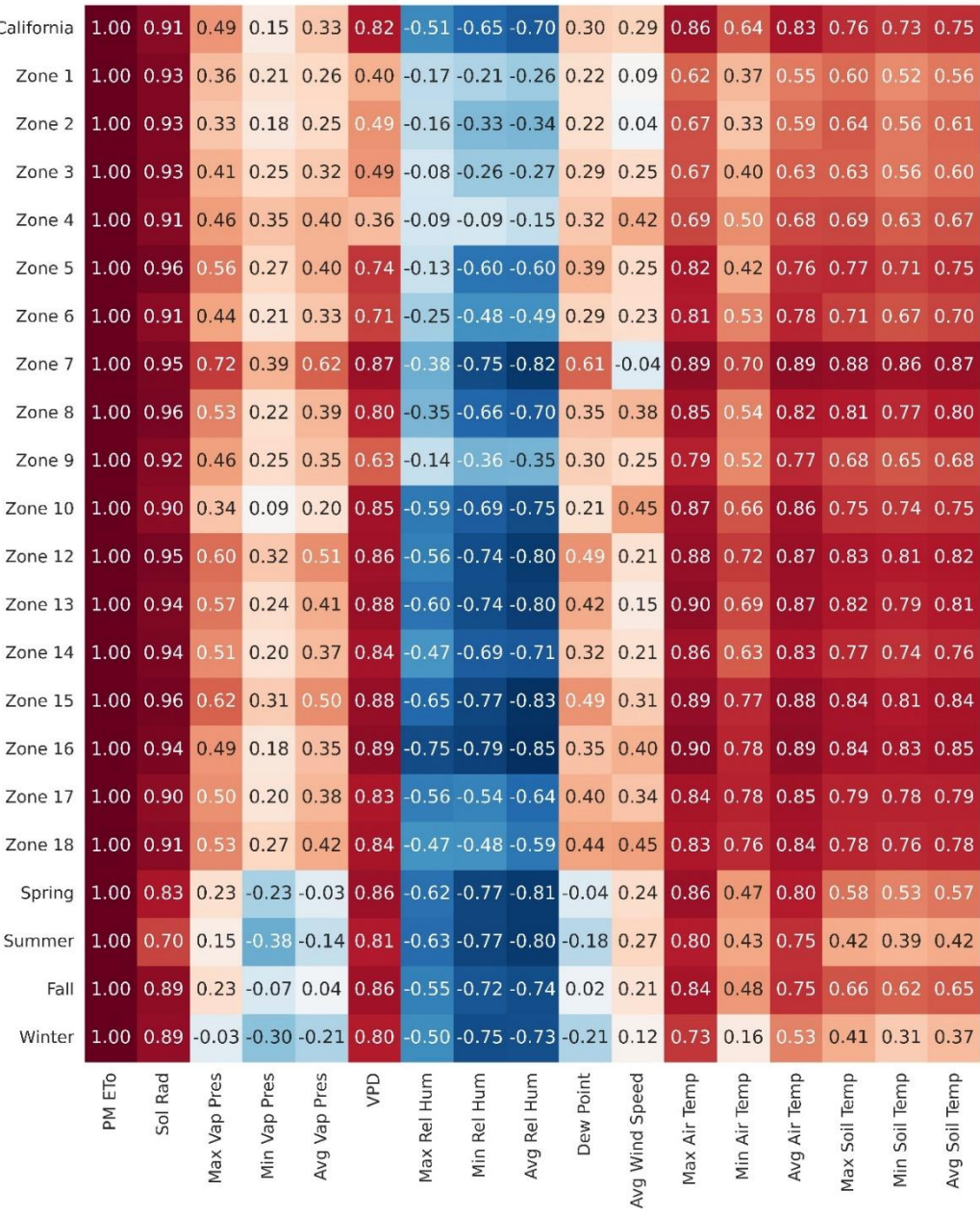
Exploratory Analysis

- **Feature importance:**
 - Pearson correlation (linear)
 - Mutual information (non-linear relationship)
 - Random forest feature importance (ML-based)
- **Trend analysis:**
 - Mann-Kendall test (monotonic upward or downward)

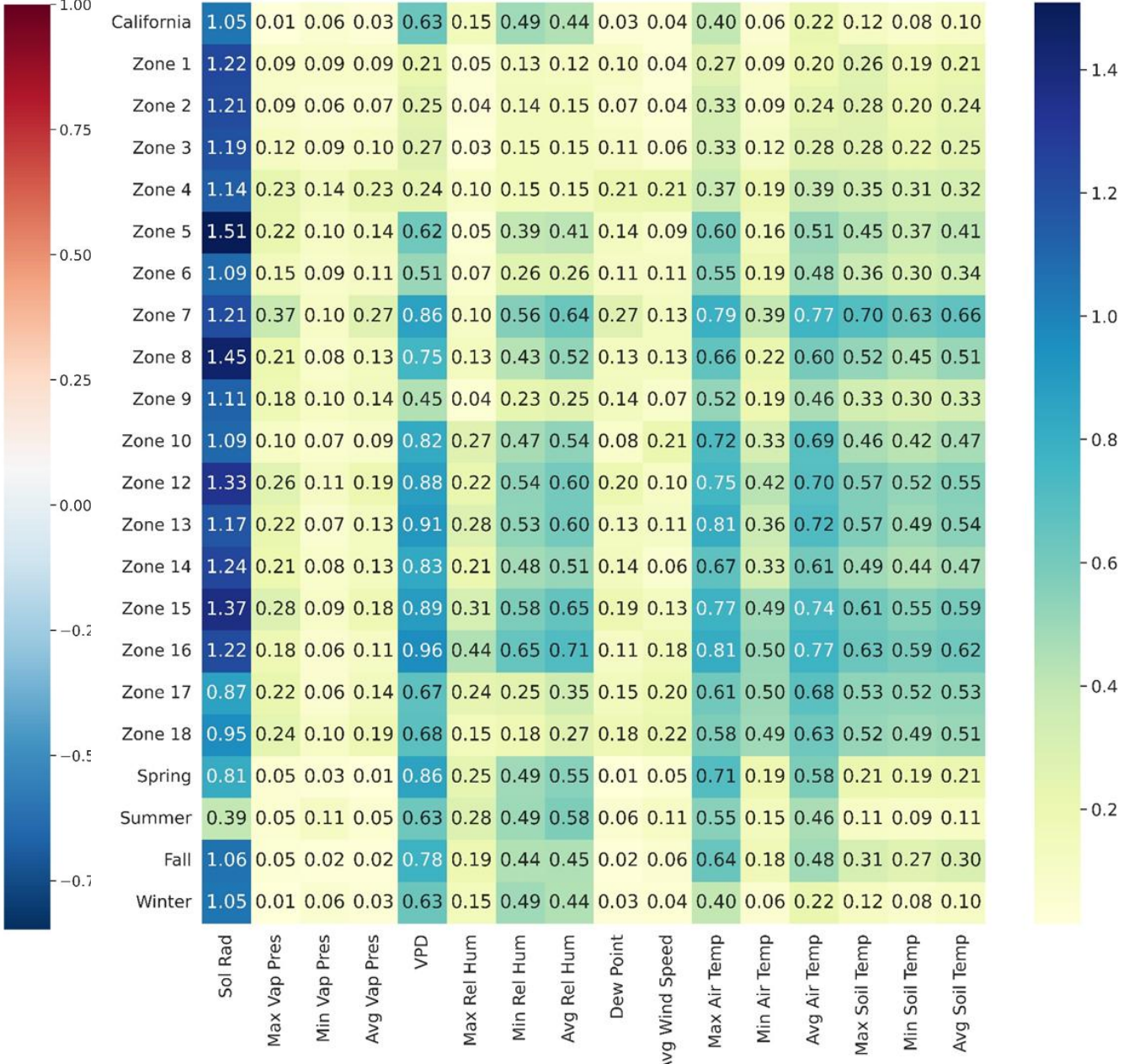


ETo & Driving factors

Pearson Correlation



Mutual information



ETo & Driving factors

Random Forest

California	0.78	0.00	0.00	0.00	0.18	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.01	0.00	0.00	0.00
Zone 1	0.87	0.00	0.00	0.00	0.08	0.00	0.00	0.01	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
Zone 2	0.86	0.00	0.00	0.00	0.07	0.00	0.00	0.01	0.00	0.01	0.04	0.00	0.01	0.00	0.00	0.00
Zone 3	0.86	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.00
Zone 4	0.85	0.00	0.00	0.00	0.09	0.00	0.00	0.01	0.00	0.00	0.03	0.00	0.01	0.00	0.00	0.00
Zone 5	0.92	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.00	0.00	0.00
Zone 6	0.79	0.00	0.00	0.00	0.16	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.01	0.00	0.00	0.00
Zone 7	0.86	0.00	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.10	0.00	0.00	0.00
Zone 8	0.91	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00
Zone 9	0.82	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.02	0.07	0.00	0.01	0.00	0.00	0.00
Zone 10	0.76	0.00	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.02	0.00	0.00	0.00
Zone 12	0.89	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00
Zone 13	0.84	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00
Zone 14	0.87	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.00
Zone 15	0.91	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00
Zone 16	0.85	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.00
Zone 17	0.72	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.18	0.00	0.00	0.00
Zone 18	0.81	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.02	0.00	0.00	0.00
Spring	0.24	0.00	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00
Summer	0.28	0.00	0.00	0.00	0.62	0.00	0.00	0.01	0.00	0.07	0.01	0.00	0.01	0.00	0.00	0.00
Fall	0.74	0.00	0.00	0.00	0.18	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.01	0.00	0.00	0.00
Winter	0.74	0.00	0.00	0.00	0.19	0.00	0.00	0.01	0.00	0.05	0.00	0.00	0.01	0.00	0.00	0.00
	Sol Rad	Max Vap Pres	Min Vap Pres	Avg Vap Pres	VPD	Max Rel Hum	Min Rel Hum	Avg Rel Hum	Dew Point	Avg Wind Speed	Max Air Temp	Min Air Temp	Avg Air Temp	Max Soil Temp	Min Soil Temp	Avg Soil Temp



January	0.47	0.00	0.00	0.00	0.43	0.00	0.00	0.01	0.00	0.07	0.00	0.00	0.01	0.00	0.00	0.00
February	0.66	0.00	0.00	0.00	0.26	0.00	0.00	0.01	0.00	0.05	0.00	0.00	0.01	0.00	0.00	0.00
March	0.64	0.00	0.00	0.00	0.29	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.01	0.00	0.00	0.00
April	0.26	0.00	0.00	0.00	0.68	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00
May	0.22	0.00	0.00	0.00	0.72	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00
June	0.17	0.00	0.00	0.00	0.74	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00
July	0.17	0.00	0.00	0.00	0.16	0.00	0.00	0.55	0.00	0.08	0.03	0.00	0.00	0.00	0.00	0.00
August	0.18	0.00	0.00	0.00	0.70	0.00	0.00	0.01	0.00	0.08	0.01	0.00	0.01	0.00	0.00	0.00
September	0.23	0.00	0.00	0.00	0.64	0.00	0.00	0.01	0.00	0.10	0.00	0.00	0.01	0.00	0.00	0.00
October	0.31	0.00	0.00	0.00	0.55	0.00	0.00	0.01	0.00	0.11	0.00	0.00	0.01	0.00	0.00	0.00
November	0.28	0.00	0.00	0.00	0.59	0.00	0.00	0.01	0.00	0.10	0.00	0.00	0.01	0.00	0.00	0.00
December	0.25	0.00	0.00	0.00	0.58	0.00	0.01	0.05	0.00	0.08	0.00	0.00	0.02	0.00	0.00	0.00
	Sol Rad	Max Vap Pres	Min Vap Pres	Avg Vap Pres	VPD	Max Rel Hum	Min Rel Hum	Avg Rel Hum	Dew Point	Avg Wind Speed	Max Air Temp	Min Air Temp	Avg Air Temp	Max Soil Temp	Min Soil Temp	Avg Soil Temp

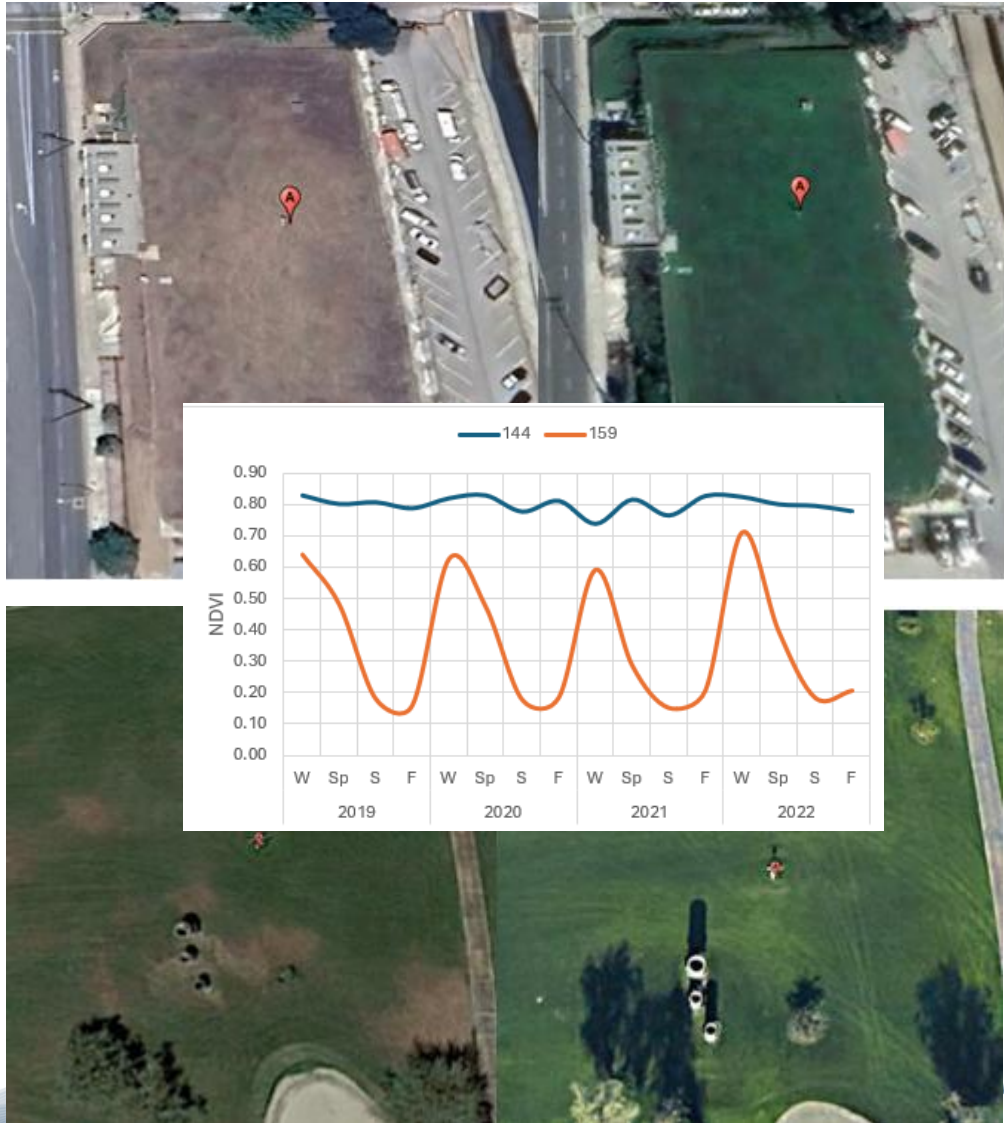


Trend Analysis (Mann-Kendall Test)

		California	Zone 2	Zone 3	Zone 6	Zone 7	Zone 8	Zone 9	Zone 10	Zone 12	Zone 13	Zone 14	Zone 15	Zone 16	Zone 18
Summer	PM ETo	-0.42	-1.42	0.42	0.45	1.57	-2.50	-1.99	3.07	2.41	-0.45	0.59	0.83	5.00	-2.31
	Max Air Temp	0.45	-0.60	0.67	1.21	2.90	-1.11	-1.51	2.93	2.03	4.20	2.33	2.87	1.13	3.02
	Avg Air Temp	1.16	0.06	-0.34	1.54	2.76	-1.22	0.24	3.53	1.35	2.73	3.15	3.12	-0.37	3.94
	Avg Vap Pres	-1.02	0.85	-1.48	-0.75	-0.18	2.16	1.60	-3.14	-1.21	-0.71	0.75	-1.18	-3.28	1.84
	Sol Rad	-2.36	-0.14	0.39	-0.04	-1.43	-1.22	0.18	-1.86	-2.60	-0.43	-2.63	-0.99	-1.92	-4.03
	Wind Speed	-0.99	-3.98	0.83	-4.13	-0.12	-0.34	-3.94	1.57	1.02	-3.72	1.10	-1.46	4.37	-3.44

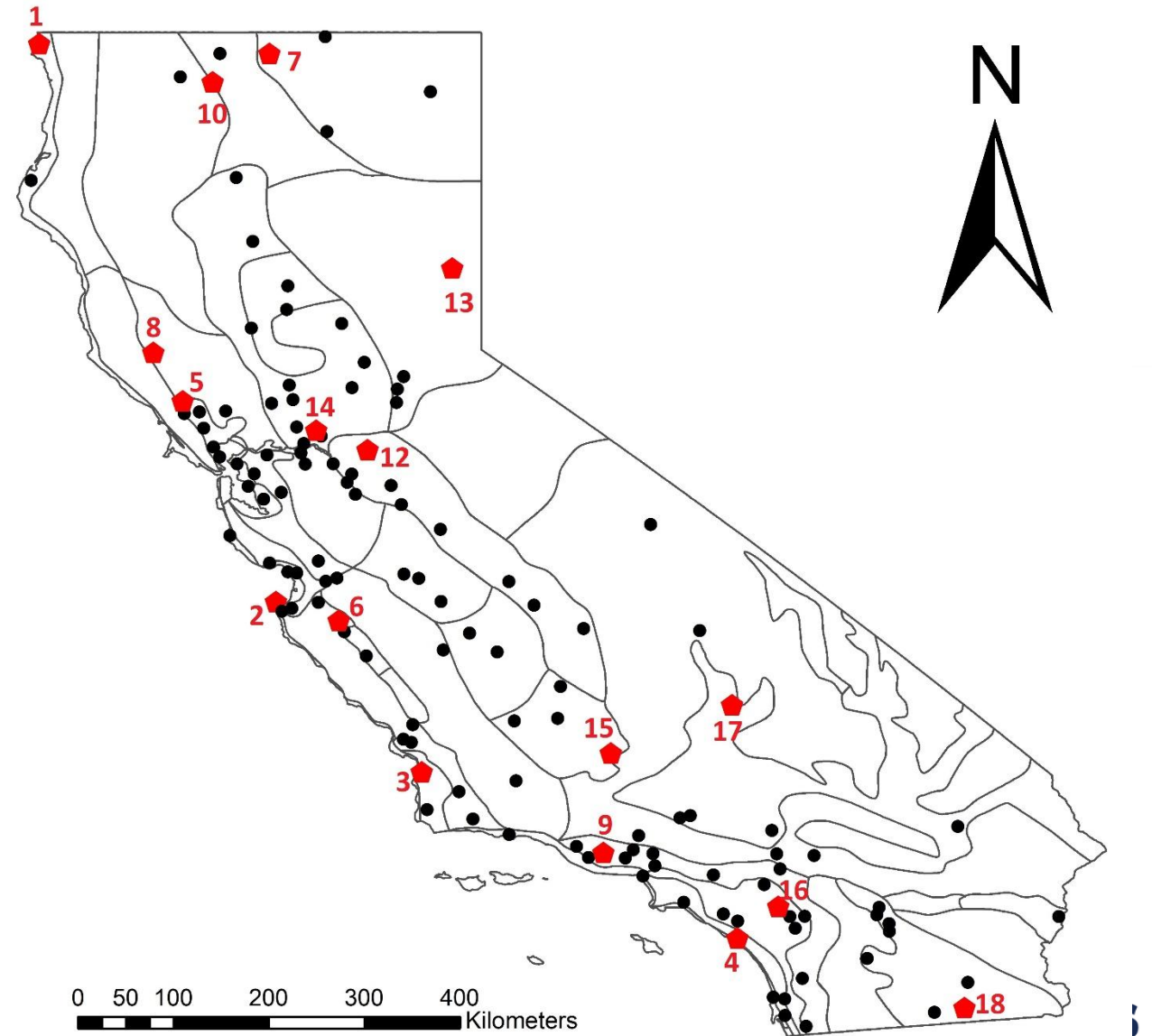
Research Objective

To develop a data-driven approach to estimate ETo without reference grass surface



Data

- 19,088,736 hourly data records from 131 active CIMIS stations
- Training on 114 stations
- Testing on 17 stations, representing California climatic zones



Data Preprocessing and Feature Selection

- Input-output schemes
 - Hourly Rs data as input, daily ETo as output [24,1]
 - Daily Rs data and cyclic Julian date ($J_c = \sin(\frac{2\pi}{365}J)$) as input, daily ETo as output [2,1]
- Data preprocessing
 - Eliminating records with null values
 - Eliminating records with ETo values higher than 15 mm/day
 - Using robust scaler to mitigate the influence of the outliers in the input

variable: $x_{scaled} = \frac{x_i - x_{median}}{Q_3 - Q_1}$

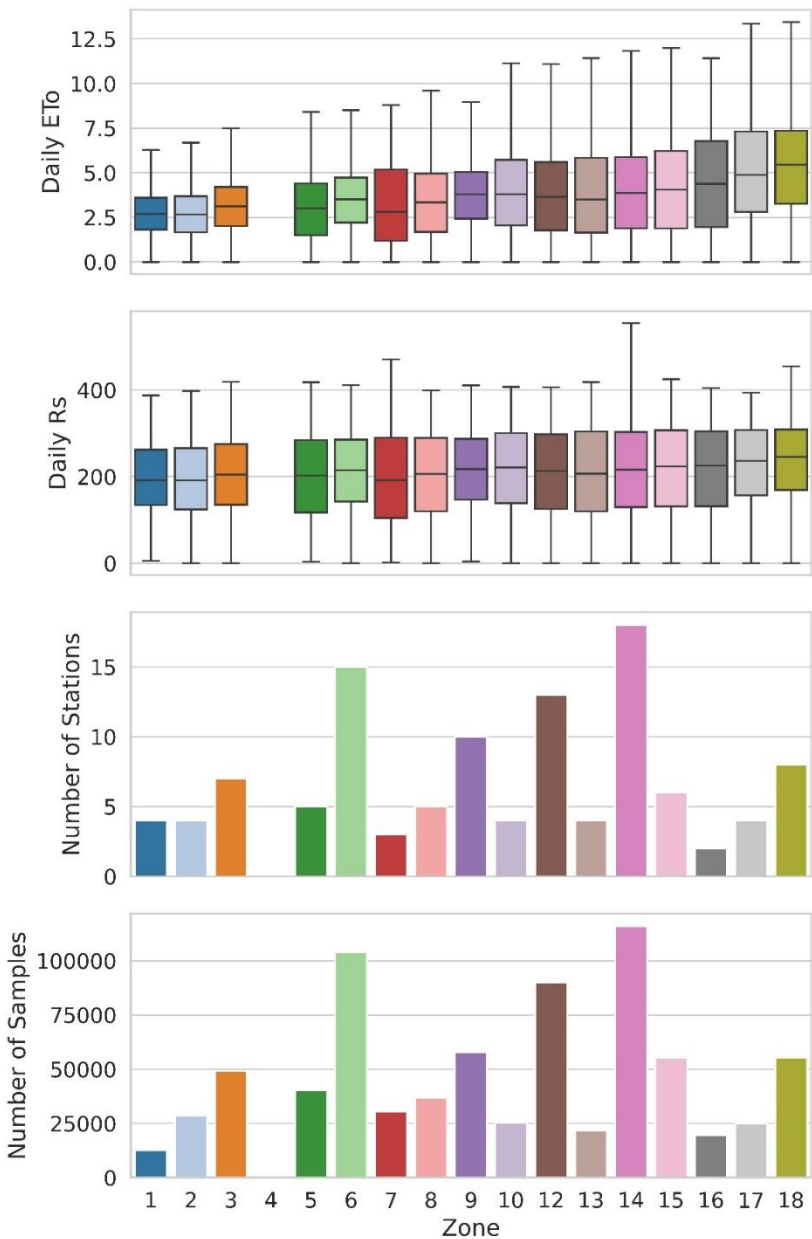
Feature Selection

- Mean substitution
 - Substituting zero Rs (nighttime) with daily mean Rs
[24,1]
- Feature selection
 - Pearson correlation
 - Spearman correlation
 - Mutual information
 - CatBoost feature importance

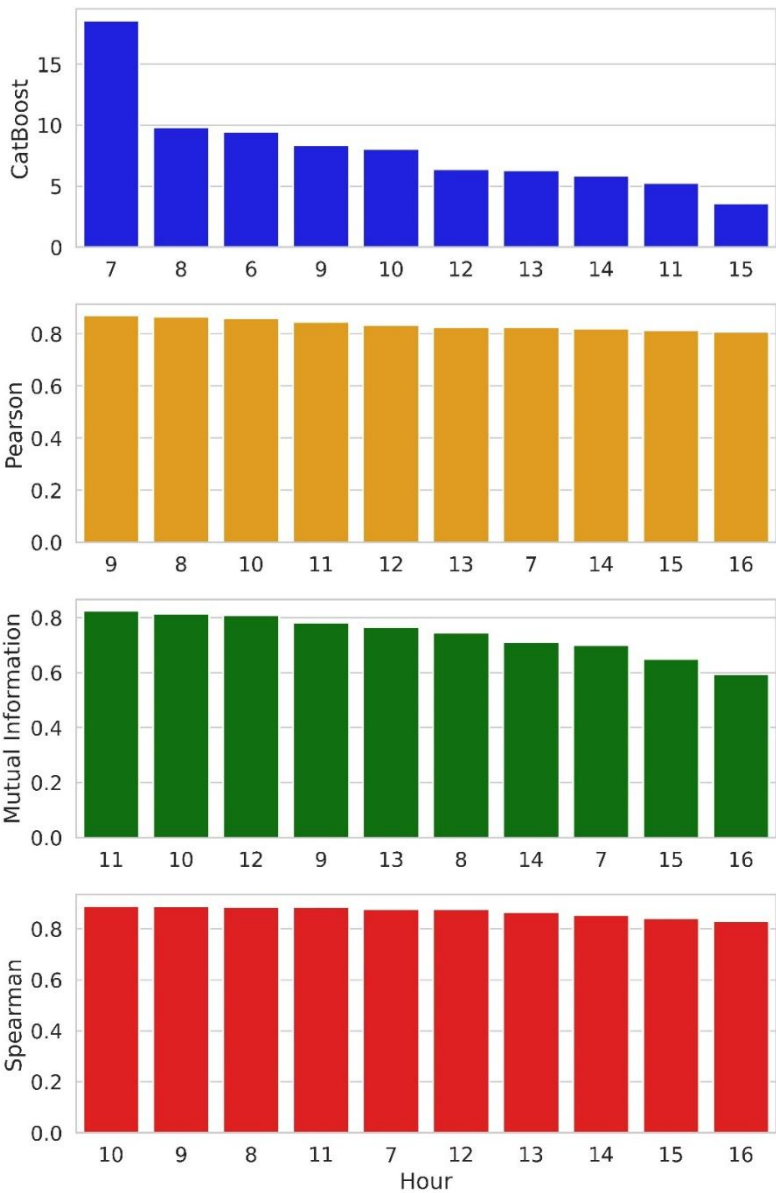
Machine learning Models

- Decision tree-based algorithms
 - CatBoost
 - LightGBM
 - Random Forest
 - XGBoost
- Neural networks
 - Deep neural network

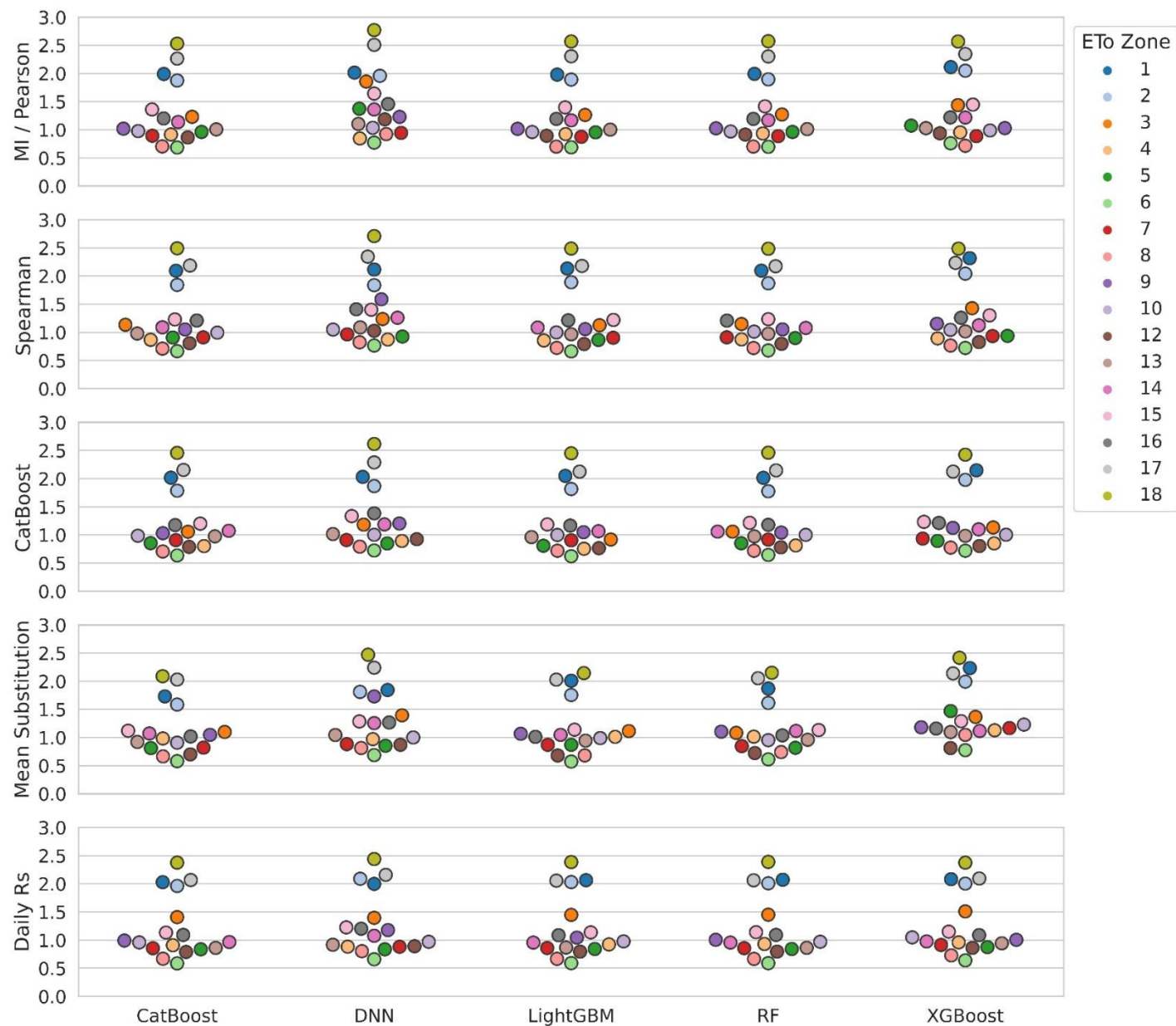
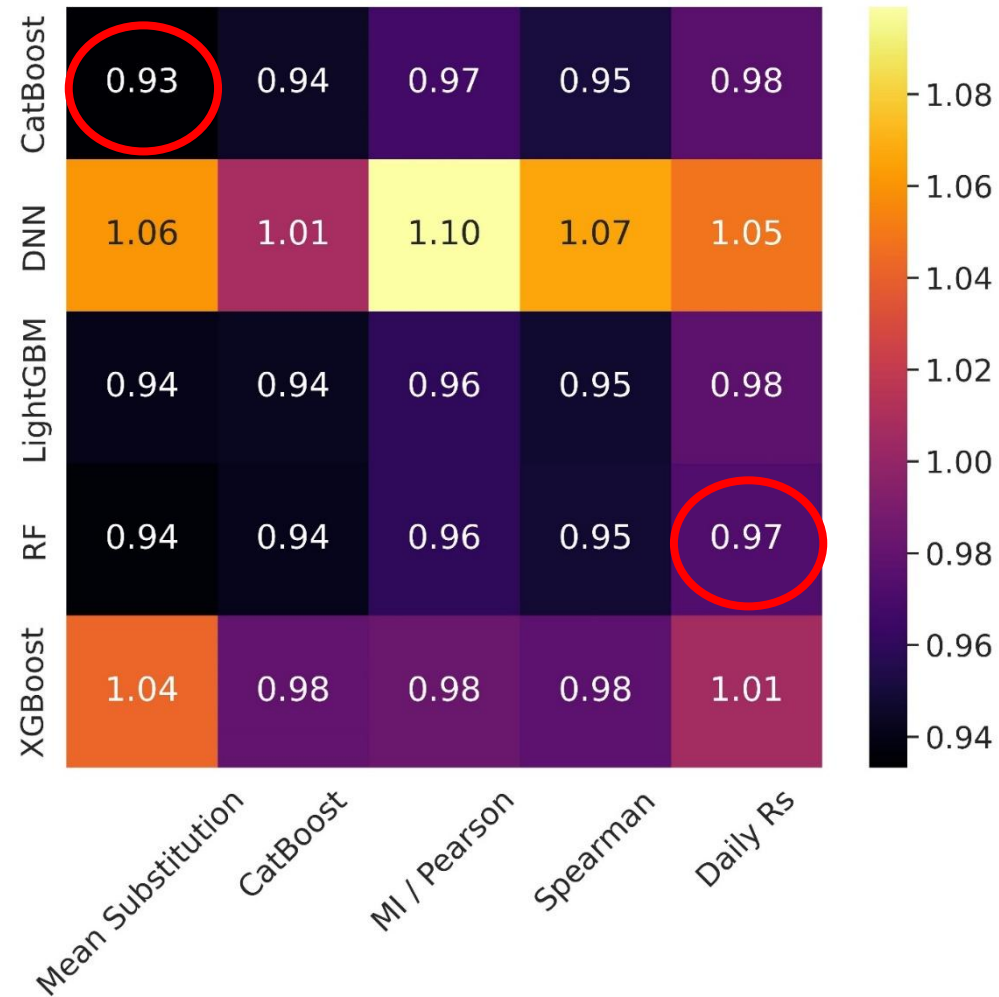
Training data investigation



Feature selection (6 most important features/hours used)



RMSE (mm/day)



ET_o Prediction

- Priestley-Taylor (radiation-based)

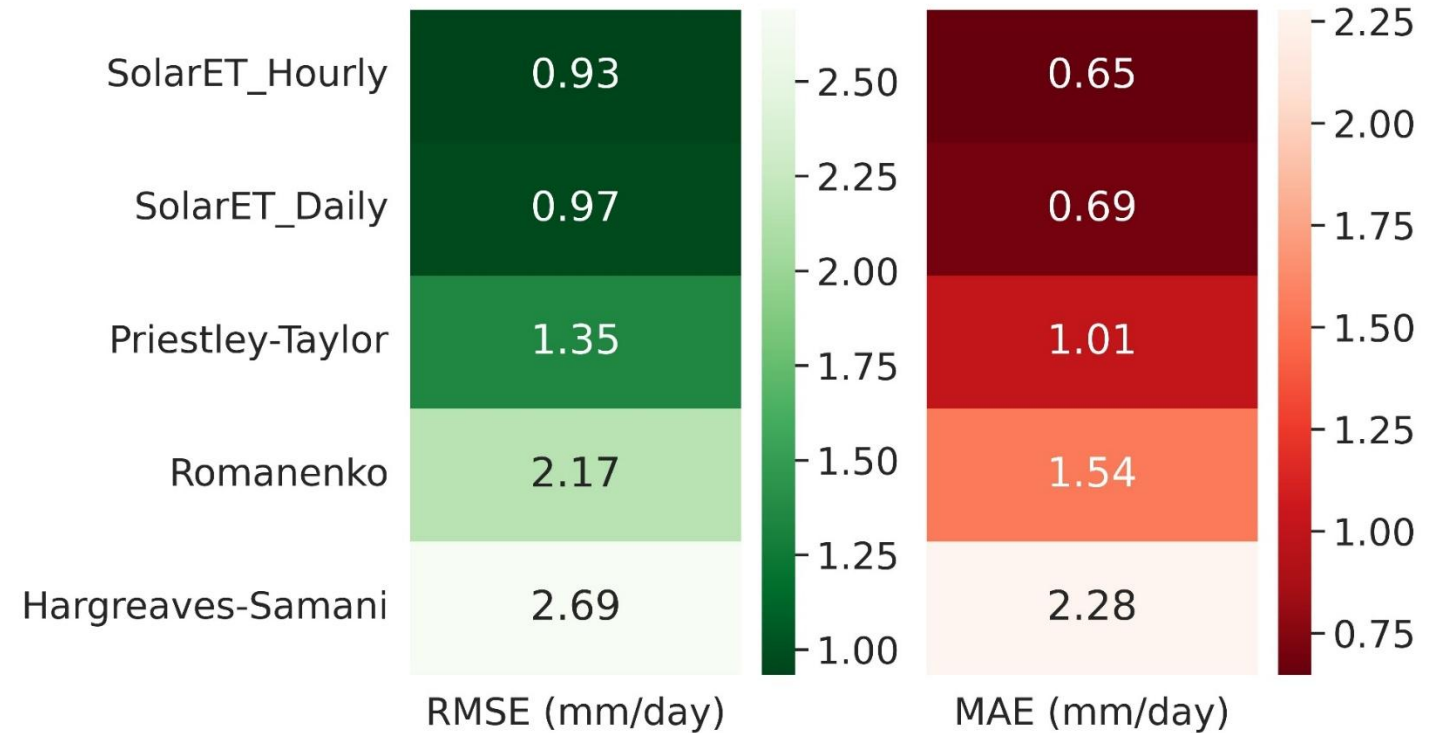
- $ET_o = 1.26 \frac{\Delta}{\Delta + \gamma} \frac{R_n - G}{\lambda}$

- Hargreaves-Samani (temperature-based)

- $ET_o = 0.0023 R_a (T_{avg} + 17.8)(T_{max} - T_{min})^{0.5}$

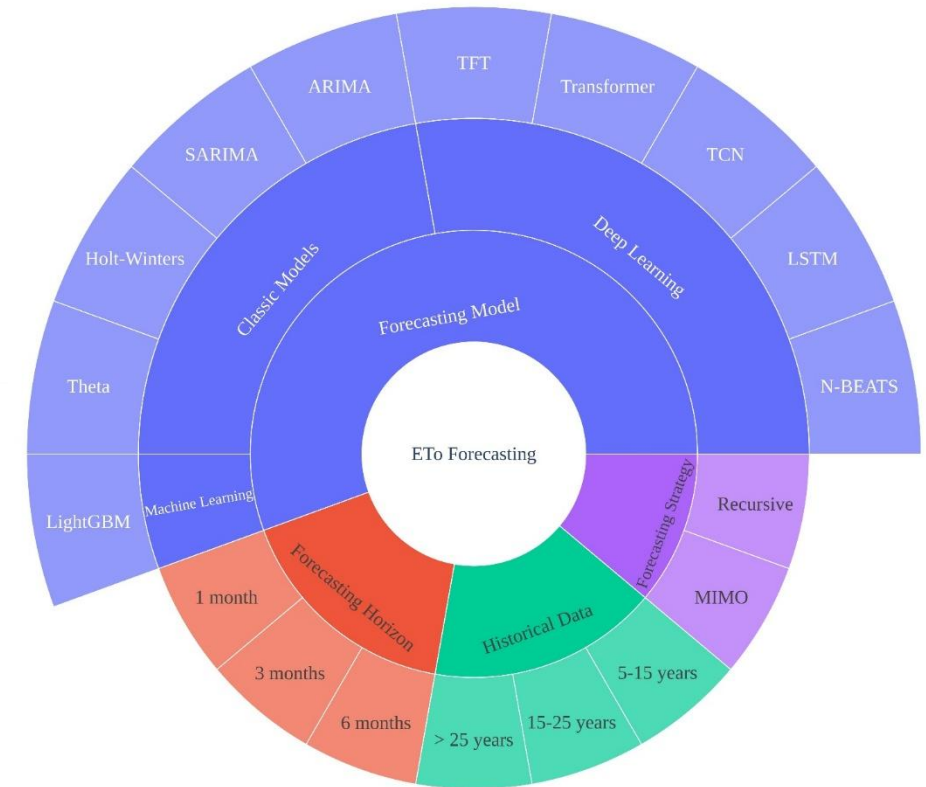
- Romanenko (mass transfer-based)

- $ET_o = 0.00006(25 + T_{avg})^2(100 - RH)$



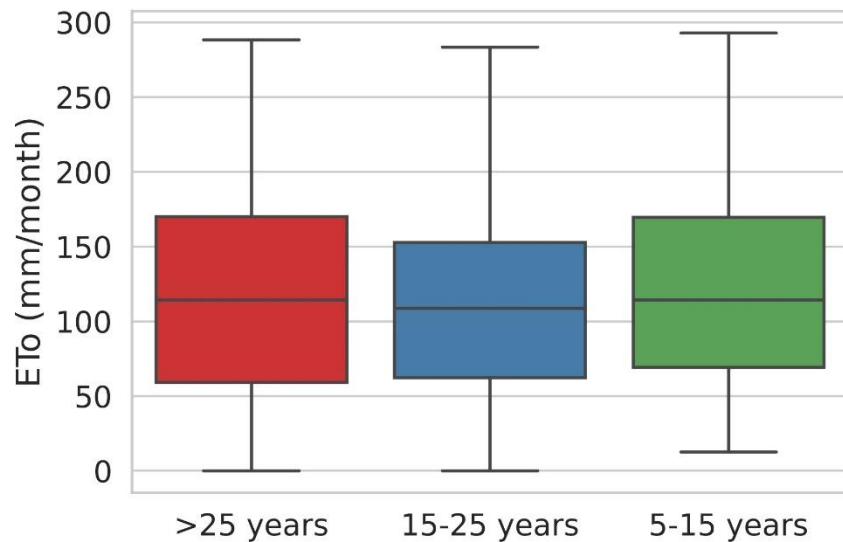
Research Objectives (ETo Forecasting)

- I. To analyze the accuracy, complexity, and data efficiency of statistical and DL models for monthly ETo forecasting

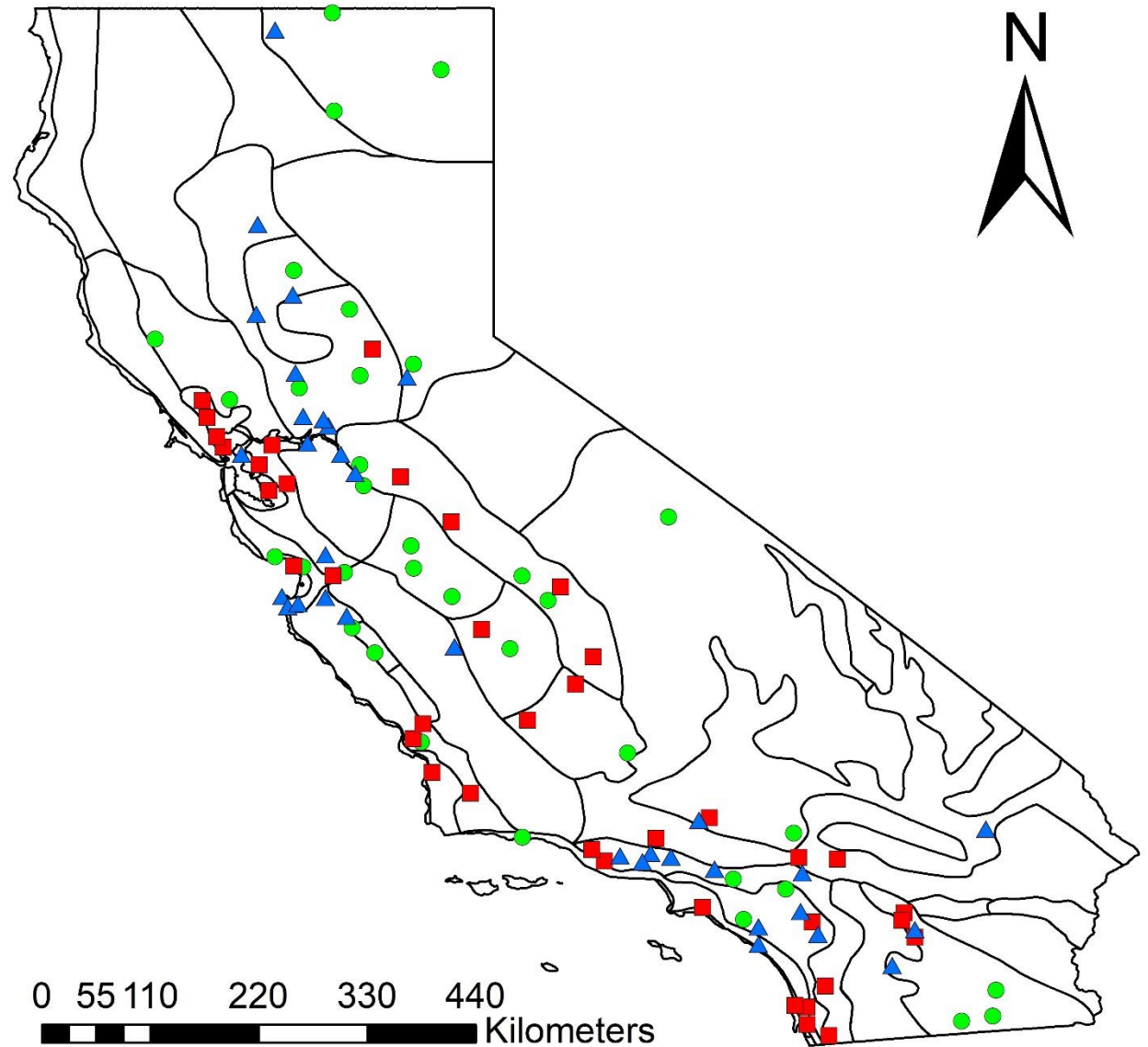


- II. To develop a globally-learned DL model to forecast monthly ETo in the Central Valley

- Monthly ETo data from 107 active CIMIS stations
- Stations categorized based on their historical data availability:
 - Long: more than 25 years (34 stations)
 - Medium-length: 15 to 25 years (38 stations)
 - Short: 5 to 15 years (35 stations)



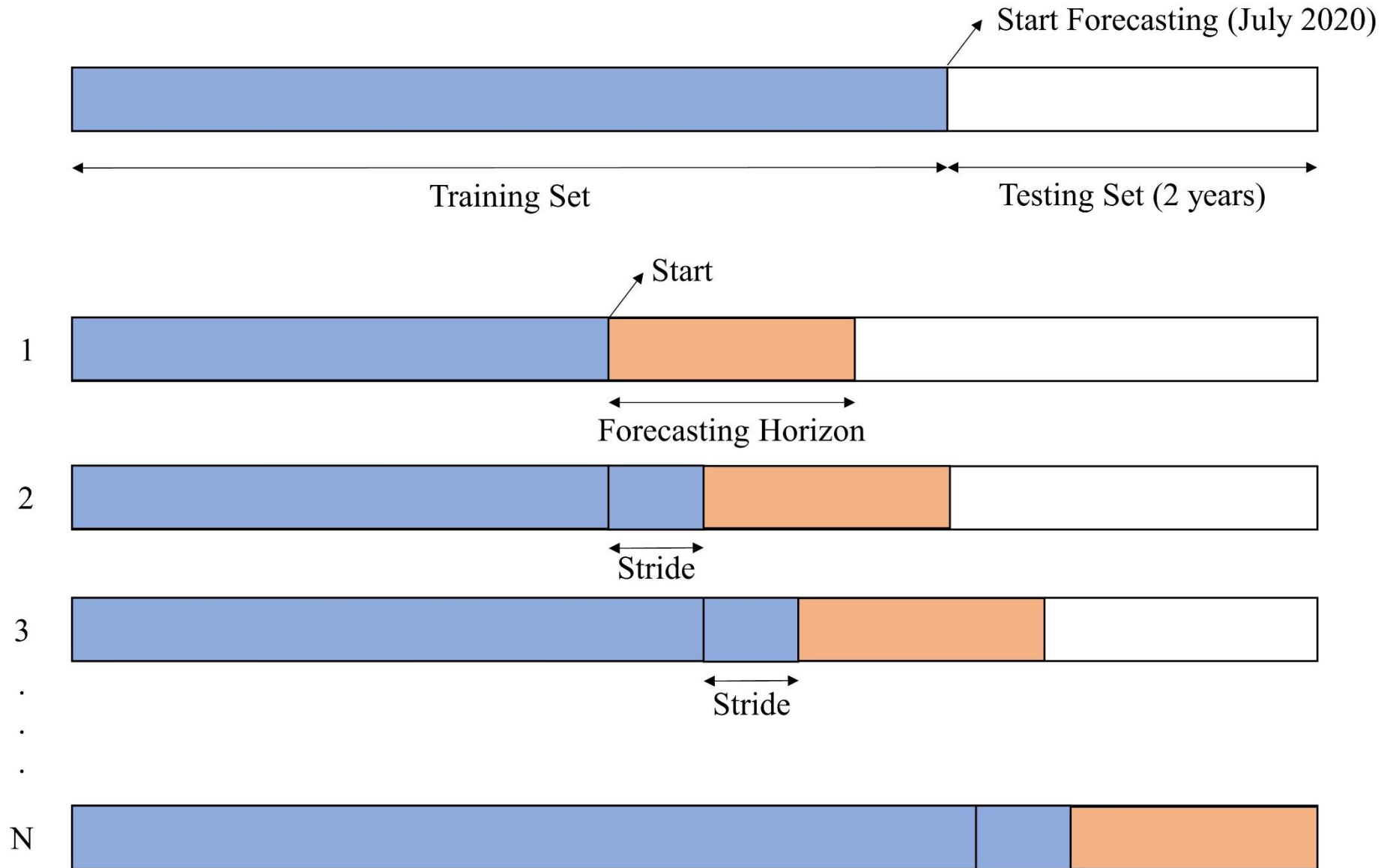
Data



Forecasting Strategy and Horizon

- ❖ Univariate time series forecasting (no exogenous variables)
- ❖ Last two years of data (July 2020 to June 2022) as the test set
- ❖ Forecasting horizons:
 - One month ahead
 - Three months ahead
 - Six months ahead
- ❖ Multi-step ahead forecasting strategies:
 - Recursive: $y_{t+1} = f(y_t, \dots, y_{t-k+1})$
 - Multi-input multi-output (MIMO): $[y_{t+H}, \dots, y_{t+1}] = F(y_t, \dots, y_{t-k+1})$

Data Splitting and Forecasting Time Steps



Models

❖ Statistical Forecasting Models:

- (Seasonal) Autoregressive integrated moving average (ARIMA and SARIMA)
- Holt-Winters' exponential smoothing
- Theta method

❖ Machine Learning Model:

- Light gradient-boosting machine (LightGBM)

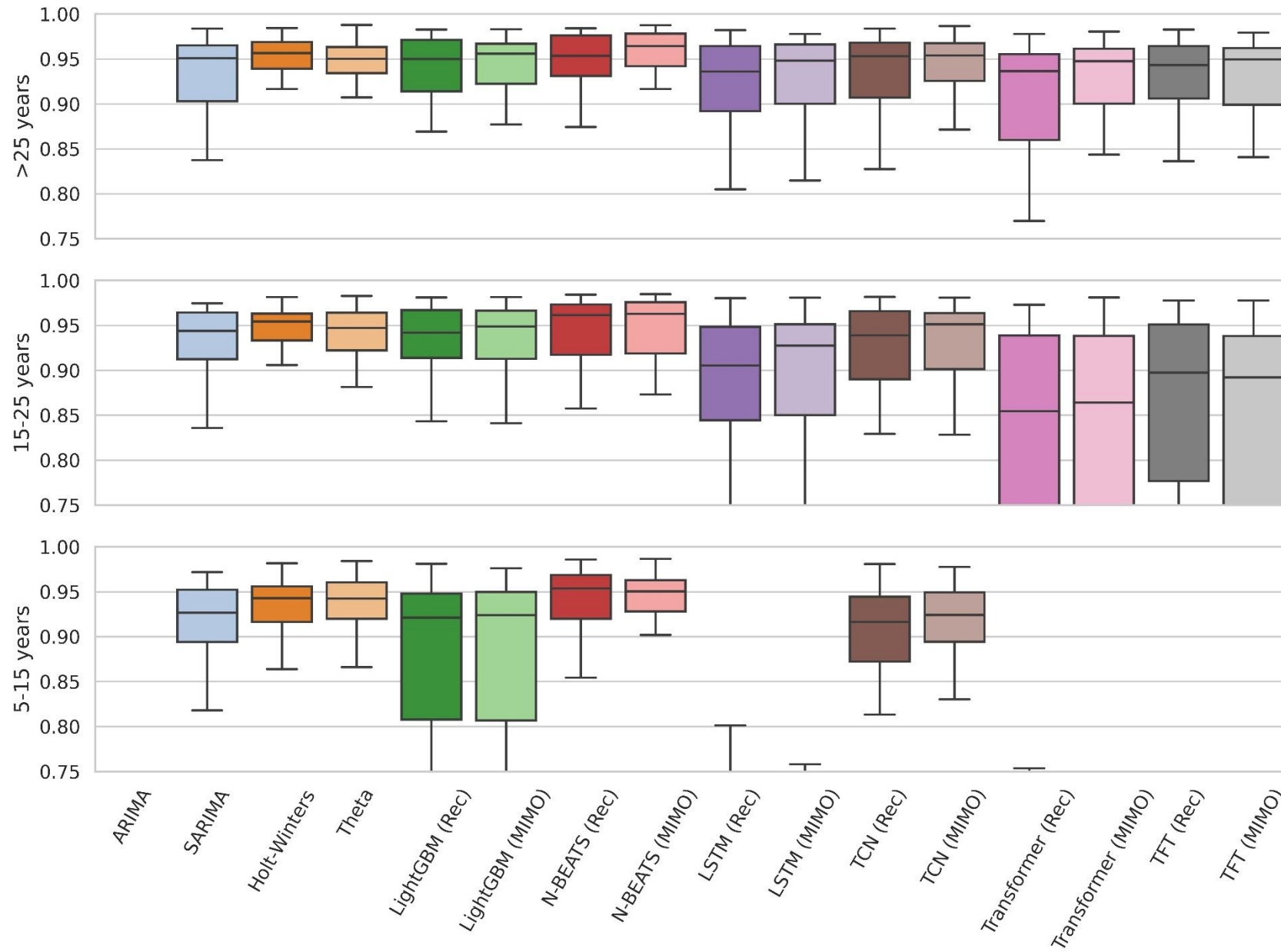
❖ Deep Learning Models:

- Neural basis expansion analysis for interpretable time series forecasting (N-BEATS)
- Long short-term memory (LSTM)
- Temporal convolutional network (TCN)
- Transformer model
- Temporal fusion transformer (TFT)

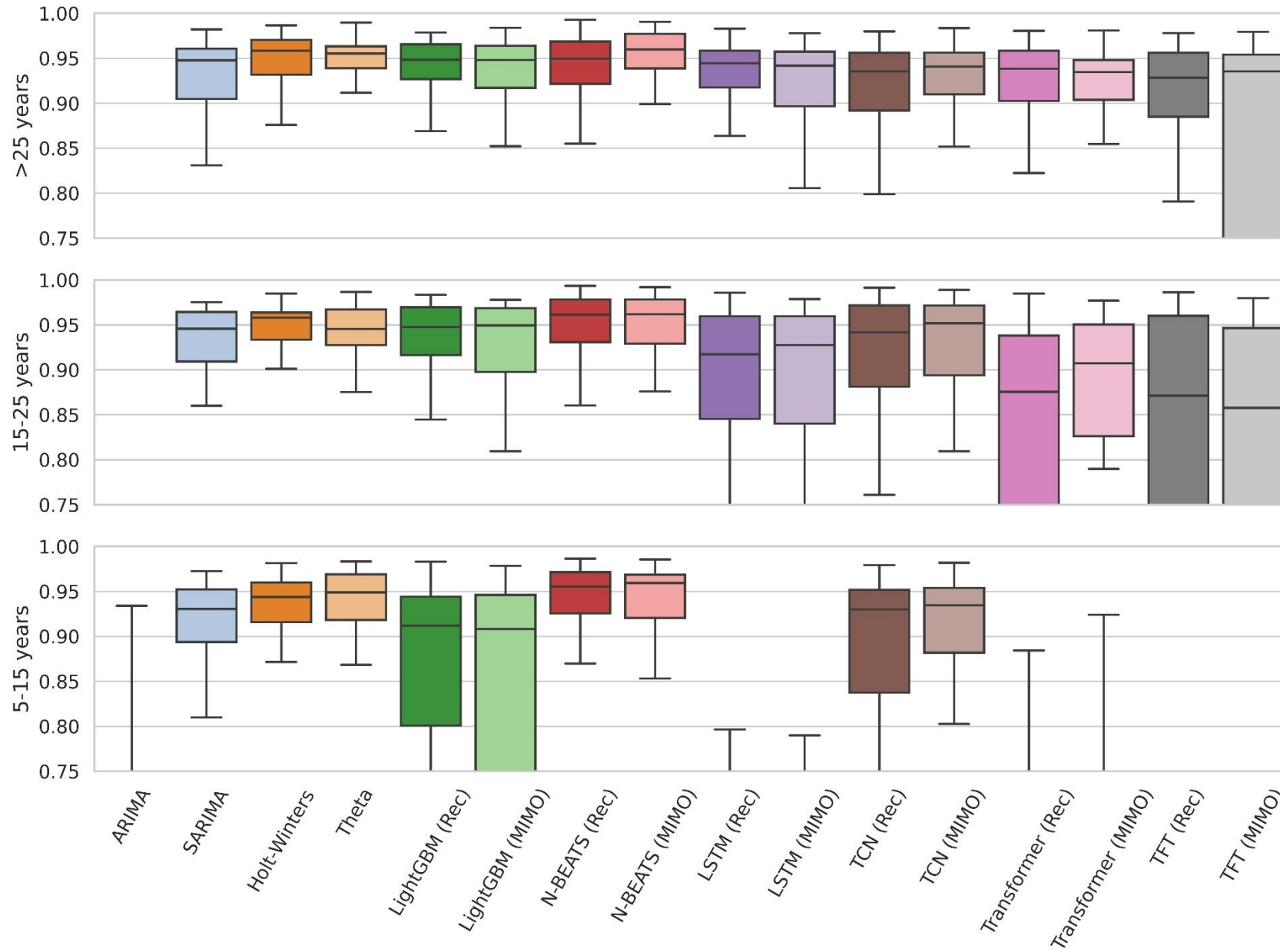
Model Complexity

Model	Runtime (seconds)	Number of trainable parameters of deep learning models
ARIMA	168 (152 + 16)	-
SARIMA	266 (152 + 114)	-
Holt-Winters	4	-
Theta	1	-
LightGBM	2	-
N-BEATS	59	~ 20,700
LSTM	645	733
TCN	253	~ 4,300
Transformer	367	~ 12,100
TFT	473	~ 15,400

Forecasting Accuracy (R^2 for three months ahead)

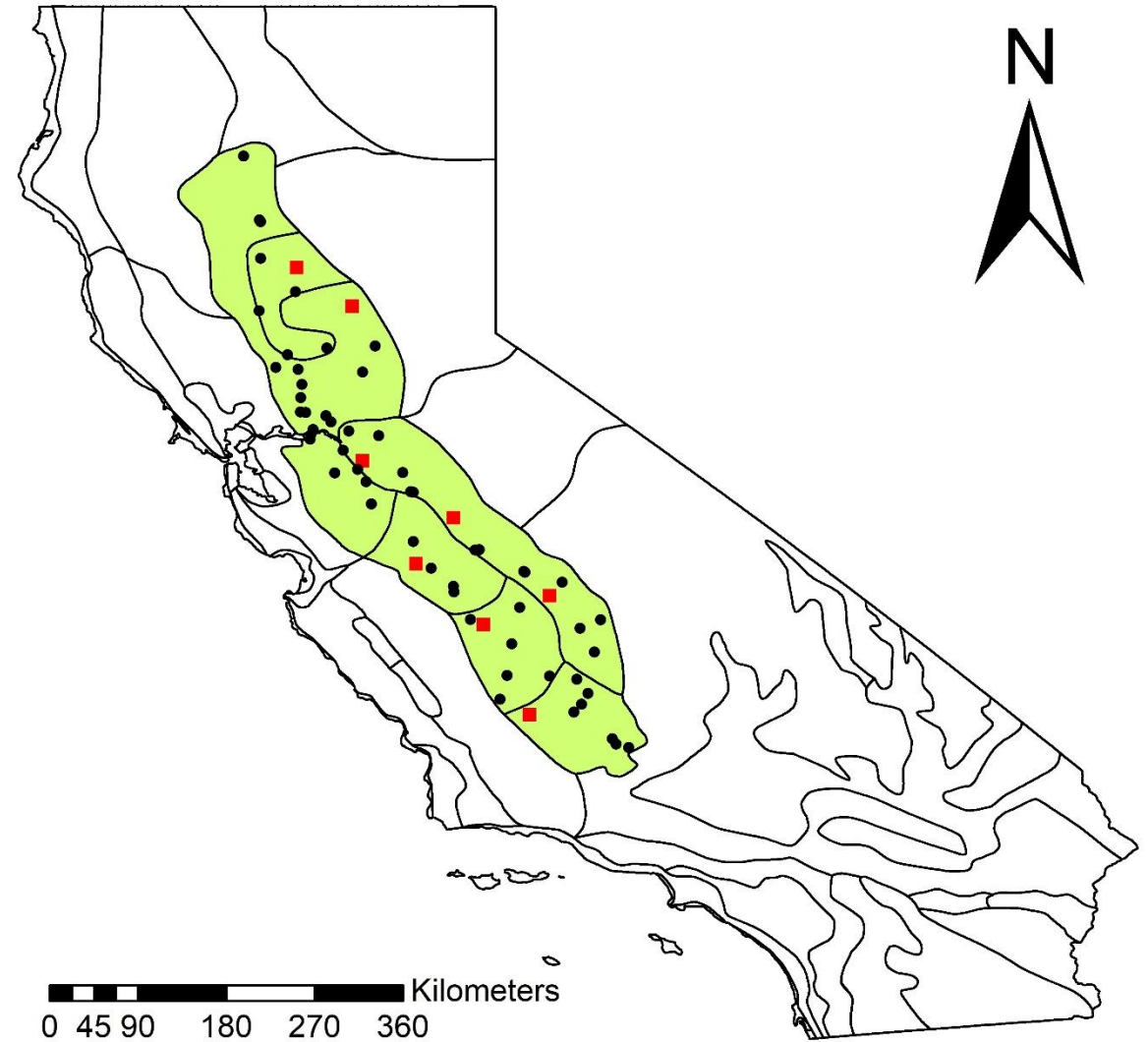


Forecasting Accuracy (R^2 for six months ahead)

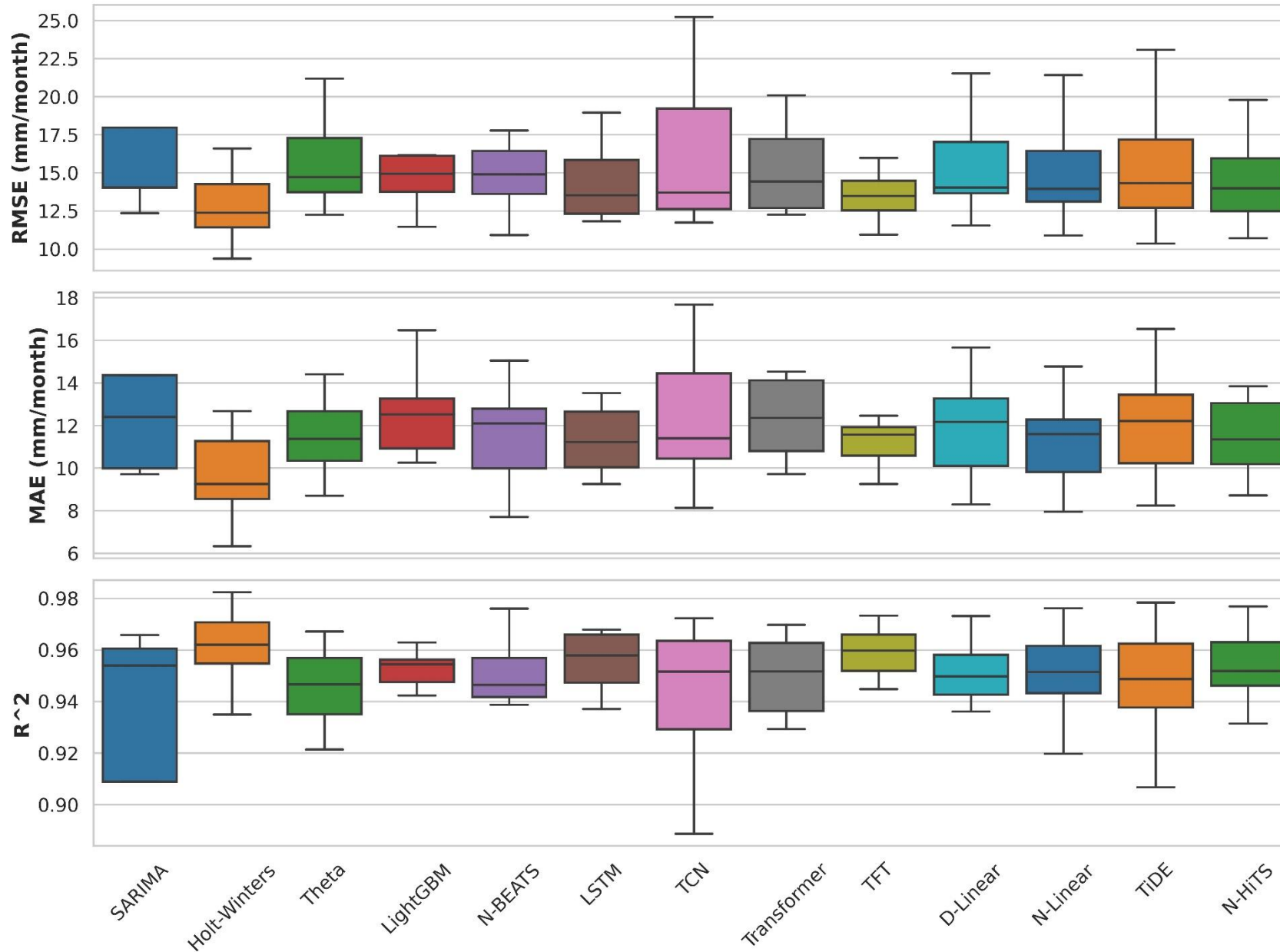


Global Learning

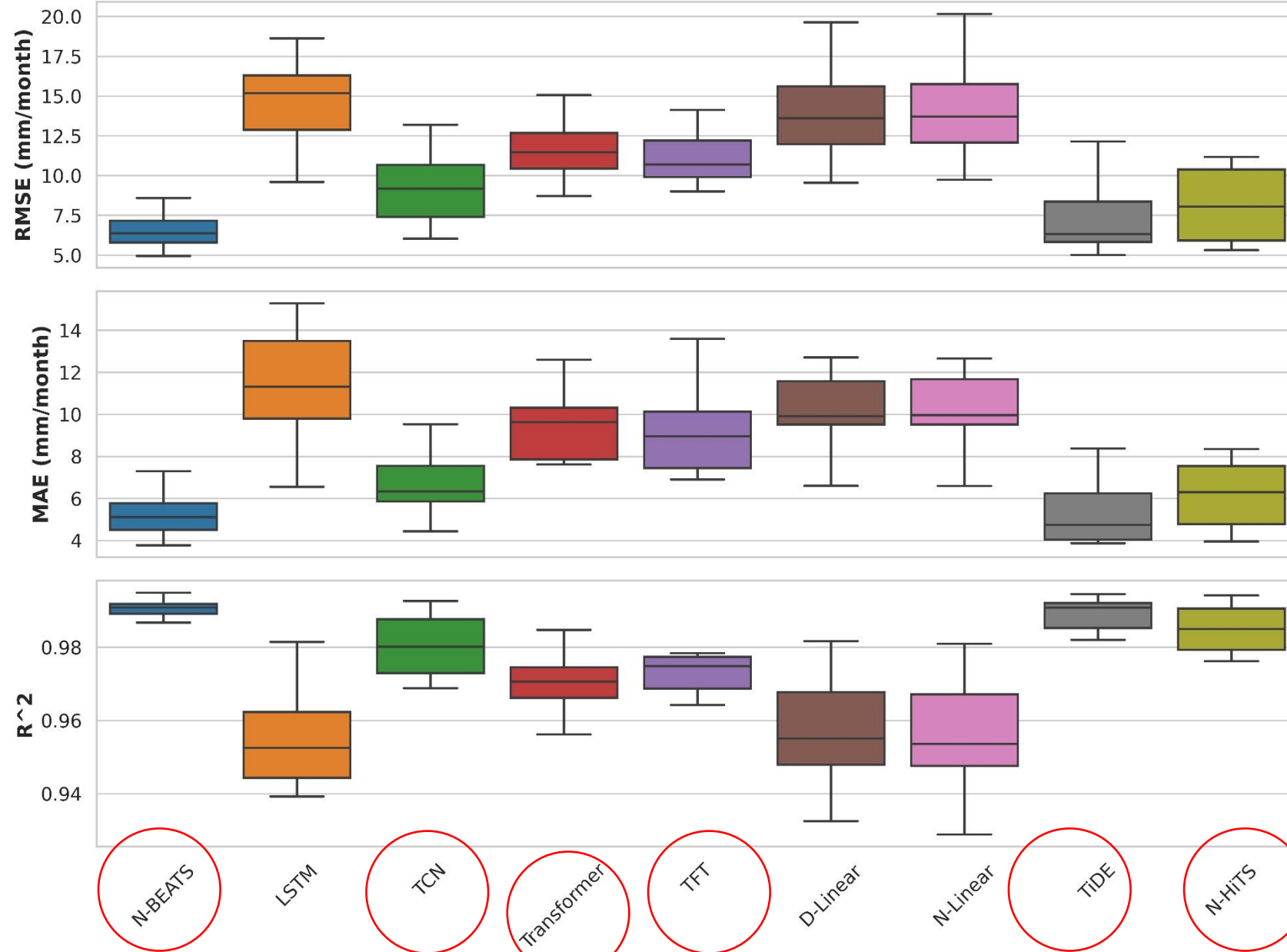
- ❖ Global Learning: training a deep forecasting model over multiple time series instead of one (model learns everything)
- ❖ Testing the performance of the model over unseen locations (Generalizability)
- ❖ 47 stations training
- ❖ 8 stations for testing



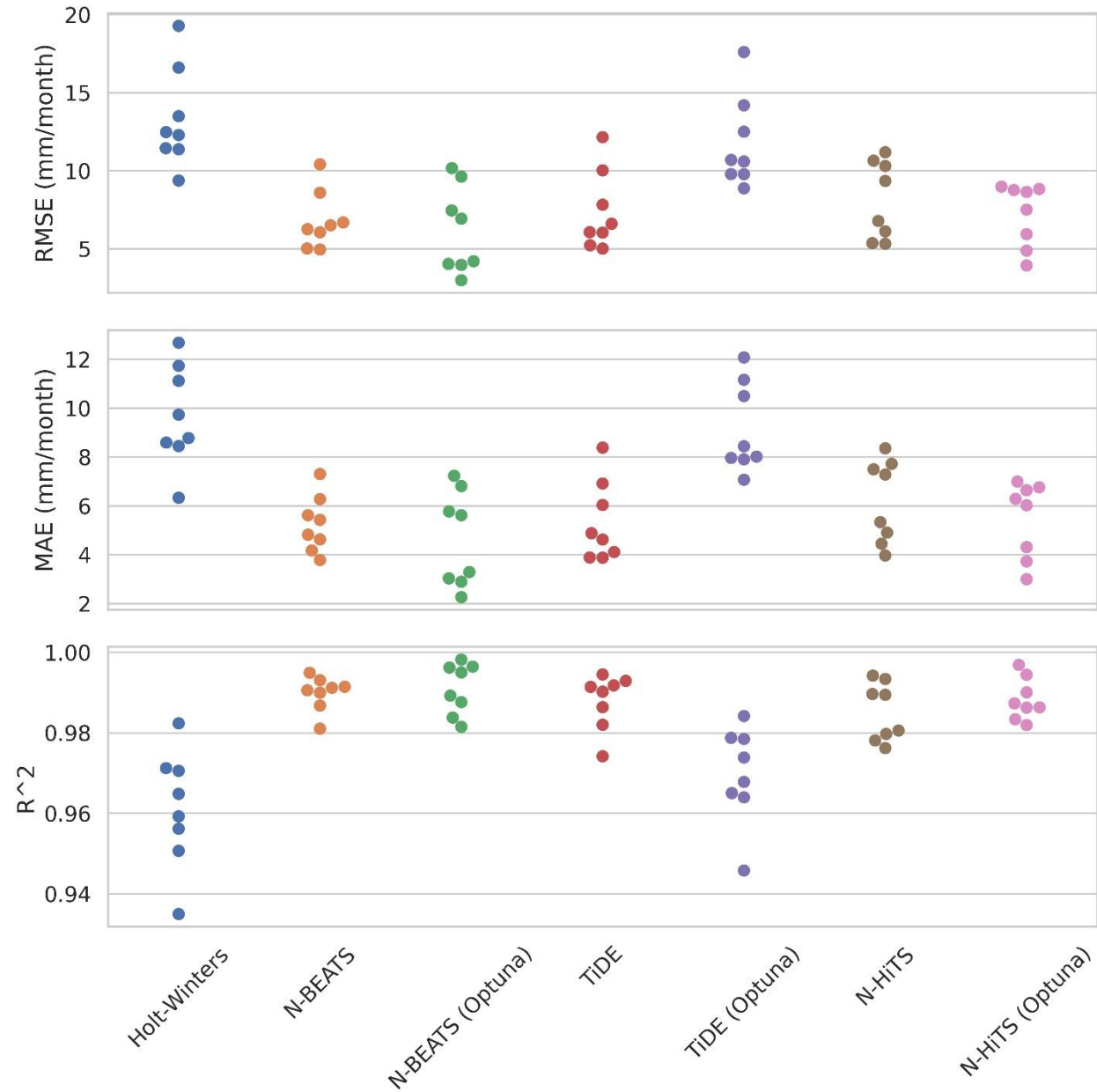
Local Learning with All Thirteen Models



Global Learning

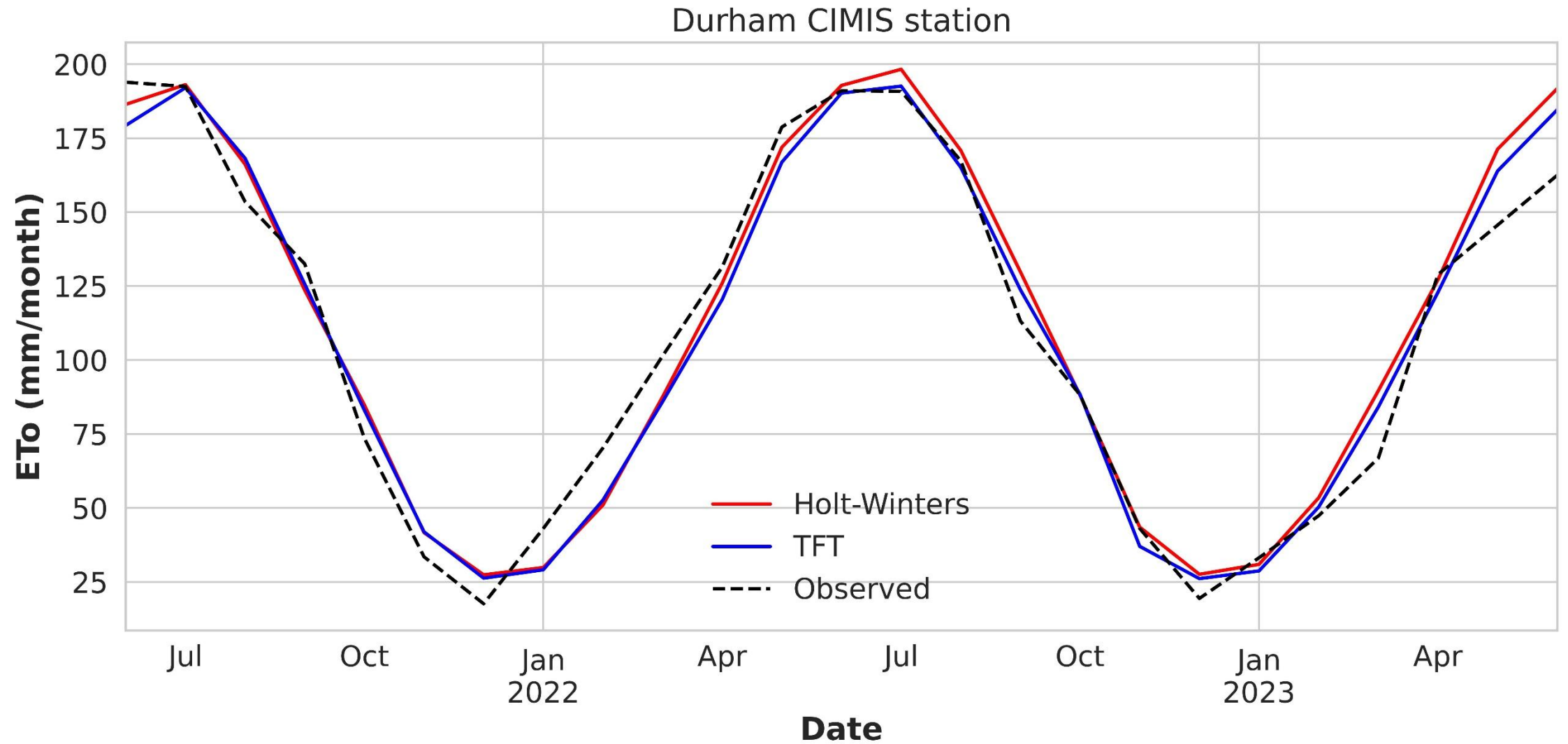


Global Learning with Optuna



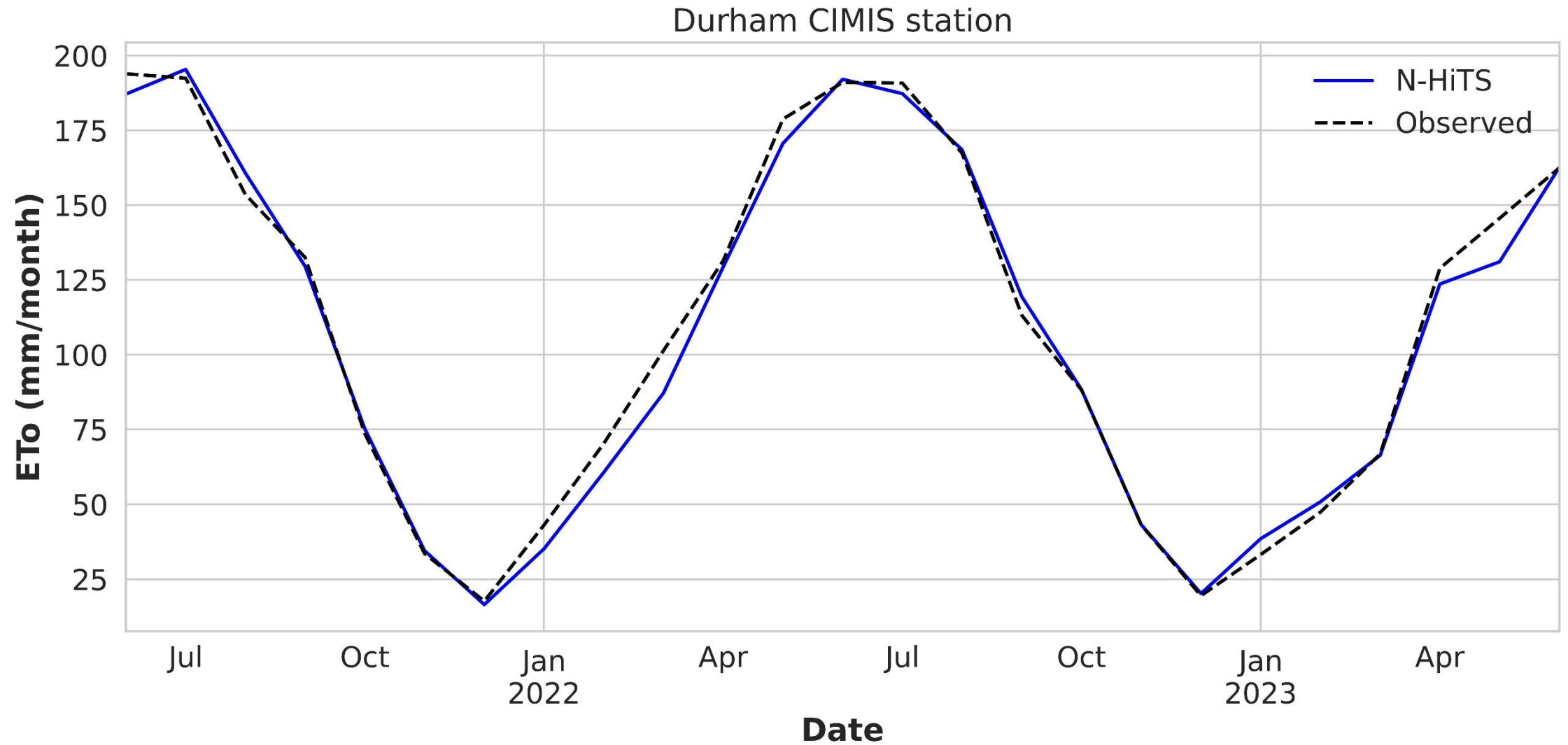


Local Learning (Durham CIMIS Station)





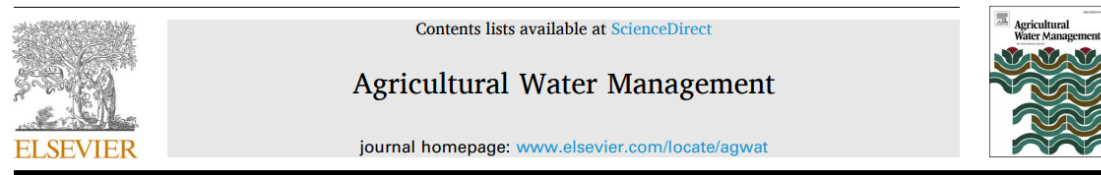
Global Learning + Optuna (Durham CIMIS Station)



Conclusion

- Global learning significantly enhances the accuracy of DL models for ETo forecasting
- GL enhances the generalizability of deep forecasting models & less prone to overfitting
- Potential to generalize findings to ungauged regions and even newly established weather stations
- Con
 - Relies on availability of sufficient and diverse training data
 - Requires substantial computational resources

Acknowledgement



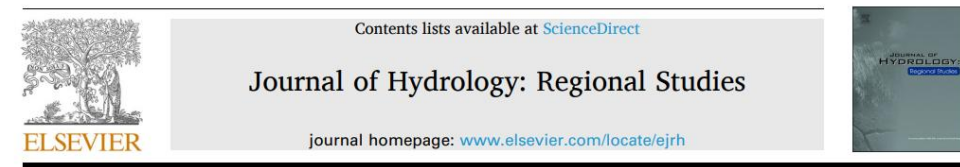
SolarET: A generalizable machine learning approach to estimate reference evapotranspiration from solar radiation

Arman Ahmadi^a, Mohammad Hossein Kazemi^b, Andre Daccache^{a,*}, Richard L. Snyder^c

^a Department of Biological and Agricultural Engineering, University of California, Davis, CA 95616, USA

^b Department of Civil Engineering, Sharif University of Technology, Tehran, Iran

^c Department of Land, Air and Water Resources, University of California, Davis, CA 95616, USA



Enhancing the accuracy and generalizability of reference evapotranspiration forecasting in California using deep global learning

Arman Ahmadi^{a,*}, Andre Daccache^b, Minxue He^c, Peyman Namadi^c, Alireza Ghaderi Bafti^d, Prabhjot Sandhu^e, Zhaojun Bai^e, Richard L. Snyder^f, Tariq Kadir^c

^a Department of Environmental Science, Policy and Management, University of California, Berkeley, Berkeley, CA, USA

^b Department of Biological and Agricultural Engineering, University of California, Davis, Davis, CA, USA

^c California Department of Water Resources, Sacramento, CA, USA

^d Department of Ocean Engineering, University of Rhode Island, Kingston, RI, USA

^e Department of Computer Science, University of California, Davis, Davis, CA, USA

^f Department of Land, Air and Water Resources, University of California, Davis, CA 95616, USA

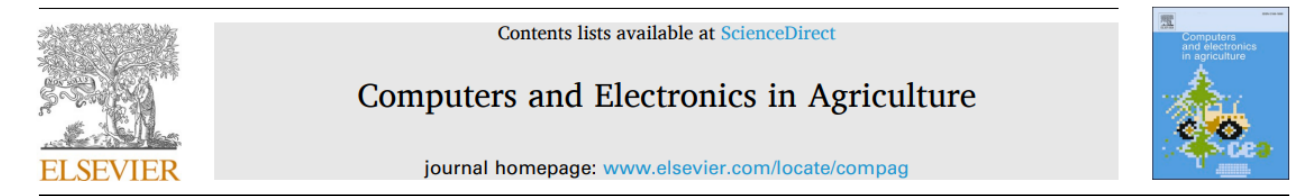


Meteorological driving forces of reference evapotranspiration and their trends in California

Arman Ahmadi^a, Andre Daccache^{a,*}, Richard L. Snyder^b, Kosana Suvočarev^b

^a Department of Biological and Agricultural Engineering, University of California, Davis, CA 95616, USA

^b Department of Land, Air and Water Resources, University of California, Davis, CA 95616, USA



Statistical and deep learning models for reference evapotranspiration time series forecasting: A comparison of accuracy, complexity, and data efficiency

Arman Ahmadi^a, Andre Daccache^{a,*}, Mojtaba Sadegh^b, Richard L. Snyder^c

^a Department of Biological and Agricultural Engineering, University of California, Davis, CA 95616, USA

^b Department of Civil Engineering, Boise State University, Boise, ID 83706, USA

^c Department of Land, Air and Water Resources, University of California, Davis, CA 95616, USA

UC DAVIS