https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

NeuralFAO56 v1.0: A Scalable Physics-Informed Deep Learning Framework for Real-Time Evapotranspiration Estimation Across CONUS

Adarsha Neupane[1], Vidya Samadi[1,2]
1. Department of Agricultural Sciences, Clemson University, Clemson, SC, USA
2. Artificial Intelligence Research Institute for Science and Engineering (AIRISE), School of Computing, Clemson University, Clemson, SC, USA
Corresponding author: samadi@clemson.edu

Abstract

Accurate estimation and forecasting of reference evapotranspiration (ETo) are essential for irrigation demand estimation. The FAO-56 Penman–Monteith formulation remains the physical standard for ETo computation, while recent advances in deep learning (DL) have demonstrated strong predictive skill for ETo forecasting. However, real-time ETo forecasting remains constrained by manual meteorological station identification, heterogeneous data acquisition, and labor-intensive preprocessing workflows. Existing software tools primarily support physics-based ETo estimation without real-time data integration or forecasting capability, whereas DL-based approaches often require manual data preparation, limiting automation and real-time applicability. This study introduces NeuralFAO56 Python package, a hybrid physics–data DL computational framework that embeds neural network–driven forecasting architectures for on-demand ETo estimation and forecasting across the continental United States (CONUS). NeuralFAO56 couples physics-based FAO-56 with DL sequence modeling within a unified pipeline that enables automated data acquisition, standardized preprocessing, and scalable deployment. The framework operates in dual modes: (i) physics-based FAO-56 ETo estimation using observed and forecasted meteorological inputs, and (ii) data-driven ETo forecasting using Long Short-Term Memory (LSTM) and Transformer architectures for multi-horizon (up to 7-day lead time) real-time forecast. The framework is evaluated across 845 stations at continental US spanning different climate regions. Results demonstrate strong short-term predictive skill, with performance degradation at longer lead times driven by reduced temporal predictability. Higher forecasting skill is observed in climatologically stable regions, while comparatively lower performance occurs in humid, convectively active regions. Overall, NeuralFAO56 provides a scalable, real-time framework that integrates physically based ETo modeling grounded in energy and mass conservation with DL forecasting and automated meteorological data pipelines to support short- to medium-range irrigation planning and management.

Keywords: Reference evapotranspiration; Irrigation demand estimation; FAO-56 Penman–Monteith; Deep learning; real-time forecasting; NeuralFAO56.

1 Introduction

Efficient irrigation management depends fundamentally on accurate quantification of atmospheric water demand, an imperative that is intensifying under accelerating global population growth and climate variability (De Oliveira E Lucas et al., 2020; Neupane and Sawada, 2023). This demand is commonly represented through evapotranspiration, a hydrologic process involving evaporation from land and plant surfaces and transpiration from plant tissue to the atmosphere (ASCE, 2025). Because evapotranspiration is difficult to measure directly at field scales (Jensen and Allen, 2016), reference evapotranspiration (ETo), defined as evapotranspiration from a standardized reference surface that is not short of water (Allen et al., 1998), is commonly estimated. ETo therefore provides a standardized and climate-consistent measure of atmospheric evaporative demand and underpins most crop water requirement and irrigation scheduling methodologies through crop coefficient-based approaches (Pereira et al., 2020).

Among the available formulations for ETo estimation, the FAO-56 Penman–Monteith method and its American Society of Civil Engineers (ASCE) standardized implementation (hereafter FAO-56 ETo) have emerged as the widely accepted reference standard due to their physically grounded structure, robustness across diverse hydroclimatic regimes, and widespread validation in both research and operational contexts

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

(4/30/2026 12:05:00 PMAllen et al., 1998; ASCE, 2005). Although simplified empirical approaches such as Hargreaves–Samani (Hargreaves and Samani, 1985) and Priestley–Taylor (Priestley and Taylor, 1972) remain useful in data-limited settings, the FAO-56 ETo remains the benchmark method where sufficient meteorological observations are available.

Concurrently, data-driven approaches, particularly deep learning (DL) models, have gained prominence for ETo estimation and forecasting. Various DL models, such as Long short-term memory (LSTM; Hochreiter and Schmidhuber, 1997) and Transformer (Vaswani et al., 2017), have demonstrated strong capacity to learn complex temporal structures, lagged dependencies, and multi-scale variability in ETo time series (Babaeian et al., 2022; Karbasi et al., 2022; Valipour et al., 2023; Umutoni and Samadi, 2024; Ali et al., 2025; Neupane and Samadi, 2025; Tayyaba et al., 2025). Comparative assessments indicate that these recurrent and attention-based models frequently outperform both empirical formulations and traditional data-driven models for short-term ETo forecasting tasks (Ahmadi et al., 2023; Qian et al., 2024). For example, Karbasi et al. (2022) demonstrated superior multi-step ETo forecasting using LSTM models under semiarid climatic conditions. Similarly, Tayyaba et al. (2025) evaluated multiple Transformer variants, including the vanilla Transformer, Informer, Autoformer, and FEDformer, and found that even the baseline Transformer architecture achieved competitive performance relative to more complex attention mechanisms.

Despite their demonstrated predictive capability, real-time implementation of both FAO-56–based and DL–based approaches remain nontrivial. Users must often identify appropriate meteorological data sources, retrieve heterogeneous datasets from multiple web servers, harmonize variable definitions and units, and perform temporal aggregation prior to model execution. These fragmented workflows introduce technical barriers, limit reproducibility, and constrain scalability across large geographic domains. Moreover, DL models require additional data pre-processing, feature engineering, and post-processing, which are among the most time-consuming components of model development (Bilal et al., 2022). These limitations highlight the need for an integrated framework that enables automated data acquisition, standardized preprocessing, and reproducible pipelines for both physical and data-driven modeling approaches.

Several tools have been developed to facilitate FAO-56–based ETo computation, including the FAO-56 calculator (Allen et al., 1998), “PyETo” (Richards, 2019), “pyfao56” (Thorp, 2022), and “PyET” (Vermec et al., 2024). However, these tools are primarily designed for retrospective, station-based analysis and typically lack forecasting capabilities, real-time data ingestion, and scalable deployment mechanisms. In parallel, a number of studies have leveraged DL models for ETo estimation and forecasting (Bellido-Jiménez et al., 2022; Sun et al., 2025; Ye et al., 2025). While these implementations demonstrate the effectiveness of data-driven approaches, they are typically provided as stand-alone packages that assume externally acquired datasets. This limits automation and applicability for real-time use. Consequently, a persistent gap remains in current irrigation demand calculation: the absence of automated, integrated, reproducible, and operationally deployable systems that combine physics–based FAO-56 computation with data-driven forecasting approaches within a modular architecture, while leveraging publicly available real-time, forecasted, and historical meteorological data streams.

Despite these advances, most existing studies remain limited to site- or region-specific experiments conducted under controlled, offline evaluation settings for ETo estimation and forecasting tasks. Consequently, large scale DL–based ETo forecasting remains comparatively sparse. Nevertheless, a limited number of studies have expanded spatial coverage within the United States using numerical weather prediction inputs or multi-station observational networks (Medina et al., 2018; Ahmadi et al., 2023; Valipour et al., 2023). Ahmadi et al., (2023), for example, conducted monthly multi-step forecasting across 107 California stations, while Medina et al. (2018) employed numerical weather prediction models to forecast ETo at 101 U.S. Climate Reference Network stations distributed across nine NOAA climate regions. Valipour et al., (2023) implemented hybrid DL frameworks for multi-horizon (1–10 day) ETo forecasting across 30 stations spanning the continental United States (CONUS). However, scalable station-scale DL frameworks evaluated uniformly across continental spatial extents remain limited.

To address these gaps, this study introduces NeuralFAO56, an open-source Python framework designed for on-demand ETo estimation and forecasting using DL approaches across CONUS. NeuralFAO56 is built as

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

a modular framework that integrates physics–based FAO-56 modeling with DL forecasting within a unified computational pipeline. The framework leverages real-time and forecasted meteorological data provided by the National Weather Service (NWS) and historical archives maintained by the National Centers for Environmental Information (NCEI) through their respective Application Programming Interfaces (APIs). First, we developed an end-to-end NeuralFAO56 pipeline that automates real-time and forecast-driven FAO-56 ETo computation using meteorological data services. Second, we implemented a hybrid workflow that flexibly integrates physics–based modeling with DL–based forecasting, enabling adaptive operation under varying data availability scenarios while maintaining physical interpretability. Finally, we evaluated the framework's capability through a scalable, continental-scale application across stations spanning all nine NOAA climate regions. This evaluation includes both retrospective validation and a real-time forecast-window experiment, comparing FAO-56 ETo derived from observed and forecasted meteorological inputs with data-driven LSTM and Transformer forecasts.

The remainder of this paper is structured as follows. Section 2 describes the system architecture, including key components, design, and operational workflow of NeuralFAO56. Section 3 presents CONUS application of NeuralFAO56, beginning with the data and experimental setup, followed by results including regional performance analyses, distributional error characteristics, station-level forecasts, real-time forecasting performance, and computational runtime assessment. Section 4 summarizes key conclusions and outlines future research directions.

2 Software Description

NeuralFAO56 is a modular, object-oriented framework that integrates FAO-56 with DL methods for on-demand ETo estimation and forecasting. The software enables end-to-end workflows encompassing data acquisition, preprocessing, standardization, model execution, and postprocessing. The workflow of NeuralFAO56 is described as follows:

2.1 Software Architecture

NeuralFAO56 integrates data acquisition and modeling tools into a unified and modular system, as illustrated in the Unified Modeling Language (UML) component diagram (Fig. 1). The major component of the framework includes:

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

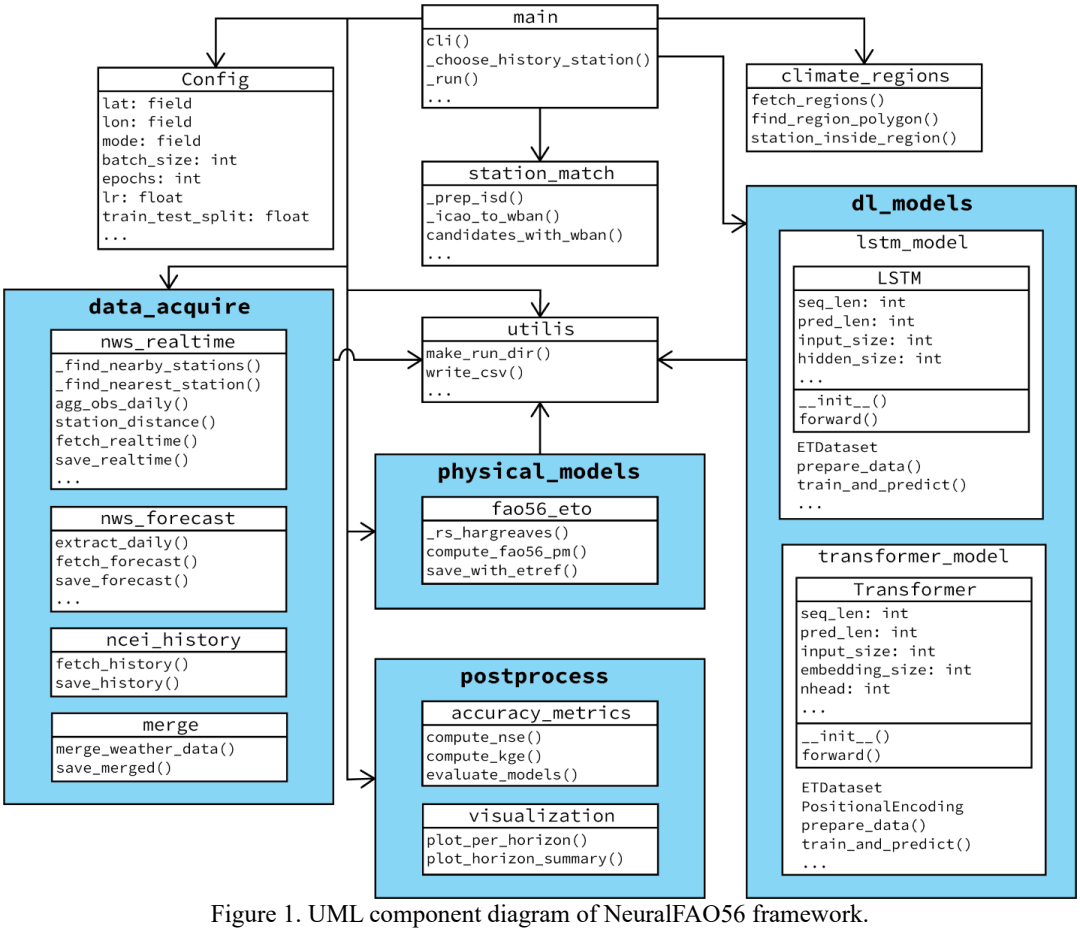


Figure 1. UML component diagram of NeuralFAO56 framework.

2.1.1 Data Acquisition and Preprocessing

Data acquisition componenet (`data_acquire` argument in the code) implements the automated retrieval and preprocessing of real-time, forecasted, and historical meteorological data through dedicated modules, including `nws_realtime, nws_forecast, and ncei_history`. These modules leverage publicly available API services provided by NWS and NCEI. The `nws_realtime` and `nws_forecast` modules retrieve real-time observations and short-term forecasts from the NWS API using the functions `fetch_realtime()` and `fetch_forecast()`, respectively. In contrast, the `ncei_history` module uses `fetch_history()` function to provide long-term historical data from the NCEI Daily Summaries archive.

The retrieved variables across both data sources include air temperature, relative humidity, dew point temperature, wind speed, and precipitation. NWS provides meteorological data at sub-daily temporal resolution, whereas NCEI archives provide data on a daily scale. Following data acquisition, these variables are preprocessed and transformed into a standardized set of daily inputs for the FAO-56 ETo computation. Stations lacking the minimum set of required variables, particularly daily maximum and minimum air temperature, are excluded from further processing.

Following preprocessing, data from these multiple sources are integrated using the `merge_weather_data()` function within the `merge` module. The resulting multivariate daily time series serves as inputs for both the FAO-56 and DL models. These processed datasets are subsequently stored

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

using dedicated I/O routines (`save_realtime()`, `save_forecast()`, `save_history()`, `save_merged()`).

2.1.2 Physics-Based Modeling

The physics-based modeling (`physical_models` argument in the code) implements physics-based ETo estimation through the `fao56_eto` module. The `fao56_eto` module computes ETo using the FAO-56 Penman–Monteith formulation as subsequently standardized by ASCE for short and tall reference crops (Allen et al., 1998; ASCE, 2005). Daily ETo is calculated using the ASCE standardized equation for a short reference crop:

$$ETo = \frac{0.408\,\Delta\,(R_n - G) + \frac{\gamma C_n}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + C_d u_2)} \quad (1)$$

where R_n is net radiation at the crop surface (MJ m^{-2} day^{-1}), G is soil heat flux density (MJ m^{-2} day^{-1}), T is mean air temperature (°C), u_2 is the mean wind speed at 2 m height (m s^{-1}), e_s and e_a are saturation and actual vapor pressure (kPa), Δ is the slope of the saturation vapor pressure curve (k Pa °C^{-1}), γ is the psychrometric constant (k Pa °C^{-1}), and C_n and C_d are empirical coefficients defined by the reference crop and time step. Detailed formulations of intermediate values follow ASCE (2005).

This computation is implemented in the `compute_fao56_pm()` function using the pyfao56 package (Thorp, 2022), an open-access Python implementation of the ASCE (2005) methodology. Daily ETo computation requires, at a minimum, elevation, latitude, minimum and maximum air temperature, along with incoming solar radiation. While station metadata from NWS and NCEI archives provide elevation and latitude, neither data source provides direct measurements of incoming solar radiation. Consequently, solar radiation is estimated within the `_rs_hargreaves()` function using the Hargreaves–Samani empirical formulation (Eq. 2), as described in Appendix E of ASCE (2005):

$$R_s = k_{RS}\sqrt{(T_{max} - T_{min})}\, R_a \quad (2)$$

where R_a is the extraterrestrial radiation (MJ m^{-2} d^{-1}), T_{max} and T_{min} are daily maximum and minimum air temperature (°C), and k_{RS} is an empirical adjustment coefficient. k_{RS} was fixed at 0.16 to minimize user-specified parameters.

To maintain a streamlined user interface and avoid excessive configuration requirements, NeuralFAO56 adopts a fixed short reference crop (crop height = 0.12 m). The resulting ETo estimates are stored as structured Comma-Separated Values (CSV) outputs in dedicated directories for subsequent analysis and integration within the framework.

2.1.3 Deep Learning Models

In NeuralFAO56, two established DL architectures (`dl_models` argument in the code), LSTM (Hochreiter and Schmidhuber, 1997) and Transformer (Vaswani et al., 2017), are implemented within the `dl_models` component via `lstm_model` and `transformer_model` modules for ETo forecasting. Both modules follow a consistent workflow encompassing data preprocessing, standardization, model training, and multi-horizon forecast generation.

Both modules utilize a core function, `prepare_data()`, in which the datasets are first screened for missing values. The proportion of missing observations is evaluated independently for each input variable as well as for the target ETo series. In hydrologic time-series analyses, missing-data proportions up to 10% are commonly considered acceptable, and interpolation within this range has been shown to yield reliable estimates (Zhang and Post, 2018). Based on this, a threshold of 5% missingness is adopted. Input features exceeding this threshold are excluded prior to model training, and only the remaining features are used as model inputs. Stations for which the target ETo series exceeds this threshold, or for which all candidate input features exceed it, are discarded entirely.

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

Following feature and station-level filtering, remaining missing values in both the retained input features and the target ETo series are imputed using forward and backward filling to produce temporally continuous sequences required for DL training. The processed input variables and target values are standardized using Min–Max Scaling technique (Han et al., 2022) defined as:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

where x denotes the original variable value, and x_{min} and x_{max} represent the minimum and maximum values computed from the training subset. Each variable is linearly rescaled to the [0,1] interval. This normalization prevents variables with larger numerical ranges from disproportionately influencing model optimization and facilitates stable convergence during gradient-based learning (Ferreira and da Cunha, 2020; Umar et al., 2025).

Finally, for model development, the datasets are partitioned using a year-based split, with 60% of the available record (rounded to the nearest complete year) used for training and the remainder for testing. Both modules employ the `train_and_predict()` method for model training and inference. Mean Squared Error (MSE) is used as loss function for parameter optimization during training and is defined as:

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

where y_i denotes observed ETo, $\hat{y}_i$ represents predicted ETo, and n is the number of evaluation samples.

Moreover, to balance predictive capability with computational efficiency for real-time deployment, a lightweight configuration is adopted as the default across both DL models, with implementation details provided in the following subsections. These default hyperparameters are defined within the `Config` class and are used unless modified by the user. To limit computational overhead and facilitate near-real-time execution within the NeuralFAO56 framework, increases in architectural depth and model complexity are intentionally avoided. This design provides sufficient expressive capacity to capture dominant temporal patterns in ETo dynamics while remaining suitable for on-demand forecasting applications. These trained models generate predictions across the specified forecast horizon which are stored as structured output files for subsequent analysis.

2.1.3.1 LSTM

The package implements an LSTM architecture using a dedicated `LSTM` class (`lstm_model` module in the package) that inherits from PyTorch's `torch.nn.Module`, with standard `__init__()` and `forward()` methods defining the model structure and forward propagation. LSTM networks are gated recurrent neural networks designed to capture long-range temporal dependencies in sequential data. Owing to their ability to model nonlinear temporal dynamics and mitigate vanishing gradient issues, LSTM architectures have been widely adopted in hydrological and environmental time-series forecasting applications (Zhang et al., 2023; Li et al., 2024; Saberian et al., 2025; Umutoni et al., 2025). Figure 2 illustrates the internal structure of an LSTM unit, including forgetting, input, and output gates that regulate information flow through the cell state. As illustrated, at each time step, the model processes the current input vector together with the hidden state from the previous time step, enabling the network to retain relevant historical information while discarding less informative signals through its gating mechanism (Gers et al., 2000; Ferreira and da Cunha, 2020). The resulting hidden representations are passed through a fully connected layer to produce multi-horizon ETo forecasts. In NeuralFAO56, the implemented default LSTM configuration consists of a single recurrent layer with 64 hidden units, followed by a fully connected output layer to generate ETo forecasts over the specified prediction horizon.

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

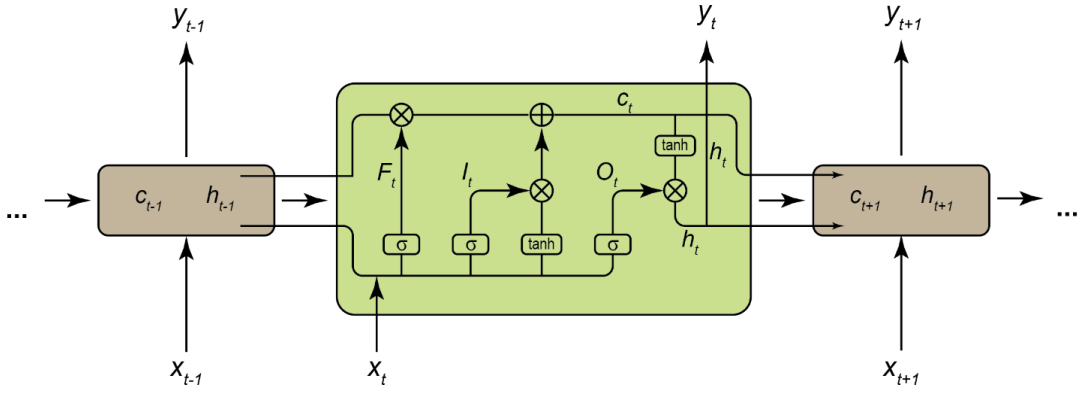


Figure 2. Schematic representation of LSTM unit used in NeuralFAO56. At each time step t, the LSTM cell updates its internal memory state (c_t) and hidden state (h_t) based on the previous states (c_{t-1}, h_{t-1}) and the current input vector (x_t). The forget gate (F_t), input gate (I_t), and output gate (O_t) regulate information flow through sigmoid activations (σ), while candidate cell updates are modulated using a hyperbolic tangent activation (*tanh*). Element-wise operations control memory retention, update, and output generation.

2.1.3.2 Transformer

The Transformer (`transformer_model` module in the code) contains a `PositionalEncoding` class and a `Transformer` class, both derived from `torch.nn.Module`. The `PositionalEncoding` class encodes positional information in the input sequence, while the `Transformer` class implements an encoder-based Transformer architecture for temporal sequence modeling. Transformer models, originally introduced by Vaswani et al., (2017), have gained widespread adoption in time-series forecasting due to their ability to model long-range temporal dependencies using self-attention mechanisms. Unlike recurrent architectures, Transformer encoders process entire sequences in parallel, enabling efficient learning of temporal relationships over long input windows (Katrompas et al., 2022).

As illustrated in Fig. 3, input meteorological time series are first mapped to a higher-dimensional latent space through a linear embedding layer. Positional encodings are then added to the embedded inputs to preserve temporal ordering, as the Transformer architecture does not inherently encode sequence position (Dufter et al., 2022; Li, 2025). These embedded representations are passed through stacked Transformer encoder blocks consisting of multi-head self-attention and position-wise feed-forward layers. The self-attention mechanism enables each time step to attend selectively to all other time steps within the input window, facilitating the learning of long-range dependencies. Multi-head attention further enables the model to capture multiple dependency patterns simultaneously (Vaswani et al., 2017). Residual connections and layer normalization are applied after each sub-layer to promote stable optimization and mitigate gradient degradation during training (Ba et al., 2016; Wang et al., 2019). The resulting encoded representations are then projected through a fully connected output layer to generate multi-horizon ETo forecasts. The Transformer model in NeuralFAO56 package employs a single encoder layer with an embedding dimension of 128, four attention heads, and a position-wise feed-forward network dimension of 256.

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

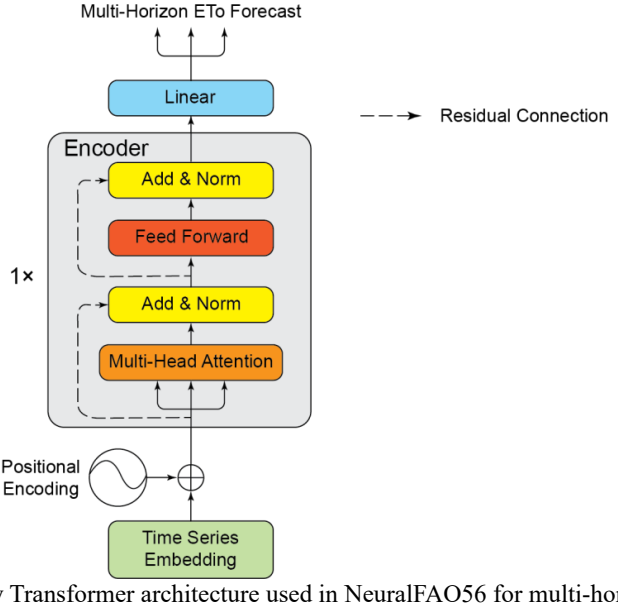


Figure 3. Encoder-only Transformer architecture used in NeuralFAO56 for multi-horizon ETo forecasting. Input multivariate meteorological time series are first mapped to a fixed-dimensional embedding and augmented with sinusoidal positional encodings. The embedded sequence is processed through a Transformer encoder block consisting of multi-head self-attention and position-wise feed-forward layers with residual connections and layer normalization. The resulting representations are projected through a linear output layer to generate multi-step ETo forecasts.

2.1.4 Postprocessing and Model Evaluation

The postprocessing component (`postprocess` module in the code) provides mathematical functions for evaluating model performance and analyzing outputs generated by the DL-based models. These functions support the computation of accuracy metrics and the visualization of forecast results. In addition the evaluation metrics (`accuracy_metrics` module in the code) assesses the performance of DL-based ETo forecasts using three widely adopted metrics in hydrological and environmental modeling: Nash–Sutcliffe Efficiency (NSE; Nash & Sutcliffe, 1970; Eq. 5), Kling–Gupta Efficiency (KGE; Gupta et al., 2009; Eq. 6), and Mean Absolute Error (MAE; Eq. 7). Together, these metrics provide complementary perspectives on model performance by quantifying temporal agreement, statistical consistency, and the magnitude of prediction errors between predicted and observed ETo values. NSE and KGE are computed using custom implementations (`compute_nse()` and `compute_kge()`), while MAE is computed using the `mean_absolute_error()` function from scikit-learn. The computed metrics are stored as structured CSV files in a dedicated directory.

NSE assesses the ability of a model to reproduce observed temporal variability by comparing the residual variance of predictions to the variance of the observations. NSE values range from $-\infty$ to 1, with values closer to 1 indicating better predictive performance. An NSE value of 0 corresponds to performance equivalent to using the mean of the observed series. NSE is computed as:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (5)$$

where y_i denotes observed ETo, $\hat{y}_i$ represents predicted ETo, $\bar{y}_i$ is the mean of observed ETo, and n is the number of evaluation samples.

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

To address limitations of NSE related to bias and variability sensitivity, KGE is also employed. KGE decomposes model performance into correlation (r), relative variability (α), and relative bias (β), combining these components into a single diagnostic measure:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (6)$$

where $\alpha=\sigma_p/\sigma_o$ and $\beta=\mu_p/\mu_o$, with σ and μ denoting standard deviation and mean of predicted (p) and observed (o) ETo series, respectively.

Because both NSE and KGE are influenced by squared deviations and may be sensitive to extreme errors, MAE is additionally computed to characterize typical magnitude of prediction errors in physical units. MAE is defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (7)$$

2.2 NeuralFAO56 Workflow Design

NeuralFAO56 is designed as a dual-mode operational framework for on-demand ETo estimation and forecasting. The system integrates physics-based FAO-56 modeling with DL forecasting, enabling flexible operation under varying data availability scenarios. Two execution modes are supported: Mode 1 performs physics-based FAO-56 ETo computation using real-time and forecasted meteorological data, whereas Mode 2 extends this capability by incorporating DL-based ETo forecasting using historical weather records. The complete operational workflow of the NeuralFAO56 framework is illustrated in Fig. 4.

The framework can be executed through a command-line interface (CLI) after installation by specifying geographic coordinates and execution mode as:

```
neuralfao56 run --lat 35.0 --lon -101.0 --mode 1
```

where Mode 1 corresponds to FAO-56–based ETo computation. Similarly, Mode 2 enables additional DL-based forecasting. An interactive execution mode is also available:

```
neuralfao56 run --interactive
```

in which the user is prompted to provide geographic coordinates and execution mode.

Following user input of geographic coordinates and execution mode, the system first identifies the corresponding NOAA climate region using the `climate_regions` module. This classification constrains subsequent station selection to ensure climatological consistency and reduce cross-regime bias. Candidate observation stations are then retrieved from NWS observation network and filtered to retain only those within the same climate region. Stations are ranked by geodesic distance, and the nearest suitable station is selected.

In Mode 1, real-time meteorological observations are retrieved from the selected NWS stations, while short-term forecasted meteorological variables are obtained from gridded NWS forecast products corresponding to the user-specified location. These datasets, available at sub-daily temporal resolution, are aggregated to daily values prior to FAO-56 ETo computation. This mode enables real-time ETo estimation and forecasting with fine spatial resolution across CONUS.

In Mode 2, the framework enables DL-based ETo forecasting using historical meteorological records. Because NWS and NCEI employ different station identifier conventions, the selected NWS station is matched to a corresponding NCEI station using metadata from the Integrated Surface Database maintained by NCEI (NOAA NCEI, 2026) within the `station_match` module. If no suitable match is found, or if the matched station yields an empty historical record, the system iteratively evaluates the next nearest NWS station until a valid historical dataset is identified or no candidates remain, as illustrated in Fig. 4. Historical

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

meteorological data are retrieved from the NCEI Daily Summaries archive using a nominal start date of 2006. If no valid historical dataset can be identified after exhausting all candidate stations, the system automatically reverts to Mode 1.

Historical NCEI records are merged with available real-time NWS observations to form station-specific multivariate daily time series using the `merge` module within the `data_acquire` component. These datasets undergo automated data validation and quality control. Stations are excluded from data-driven processing if the required variables for FAO-56 ETo computation or the target ETo series exceeds 5% missingness threshold, or if no valid predictor variables remain after feature screening. In such case, the station-matching procedure advances to the next candidate. For retained stations, remaining gaps in both input variables and the target series are filled using forward and backward filling. The processed datasets are subsequently normalized using Min–Max scaling.

The `visualization` module in the NeuralFAO56 package provides functions for plotting simulation results and forecasts generated by both FAO-56 and DL-based models. Two primary functions are implemented within this module: `plot_per_horizon` and `plot_horizon_summary`. The `plot_per_horizon` function generates time-series plots for the testing period, comparing DL-based model outputs with FAO-56 ETo estimates. These plots also incorporate evaluation metrics computed by the `accuracy_metrics` module to provide additional context on model performance. To enhance interpretability, visualization is typically limited to the most recent years of the testing period (up to five years). The `plot_horizon_summary` function generates forecast-window visualizations, comparing multi-horizon DL-based forecasts with FAO-56 ETo derived from forecasted meteorological inputs. These plots facilitate the assessment of model performance across different forecast lead times.

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

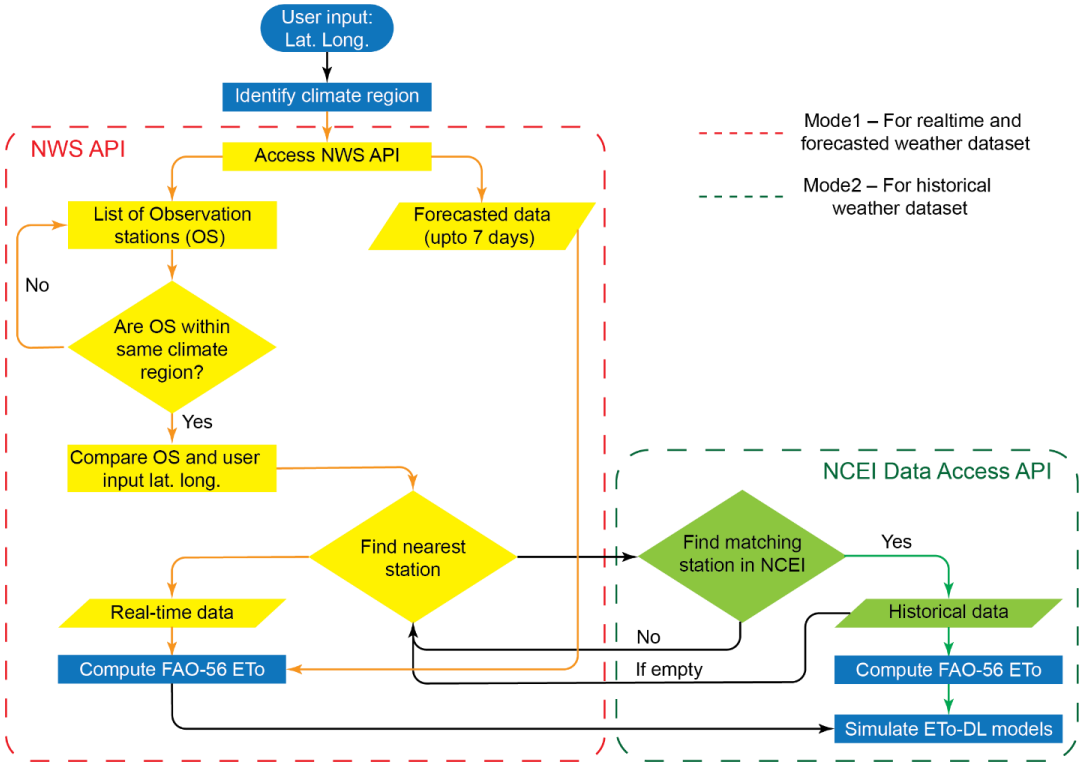


Figure 4. Operational workflow of the NeuralFAO56 framework. User-specified geographic coordinates are first mapped to NOAA climate regions, after which nearby NWS observation stations are identified and filtered. In Mode 1, real-time observations and short-term forecasted weather data are retrieved from NWS for FAO-56 ETo computation. In Mode 2, the selected NWS station is matched to a corresponding station in the NCEI archive; where sufficient historical records are available, historical data are retrieved to support DL-based ETo forecasting.

2.3 Study Area and Data

NeuralFAO56 was applied across all stations for which consistent NWS–NCEI station matches were identified, yielding a total of 845 surface observation stations distributed across CONUS. These stations span all nine NOAA climate regions (see Fig. 5), enabling regional stratification of results under diverse hydroclimatic conditions. Real-time observations and forecasted meteorological data were retrieved from the NWS API, while historical meteorological records were obtained from the NCEI API within the framework. Instead of demonstrating the framework at a single station, the implementation was conducted across a network of stations spanning CONUS to assess its performance and scalability for DL-based ETo forecasting at continental scale.

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

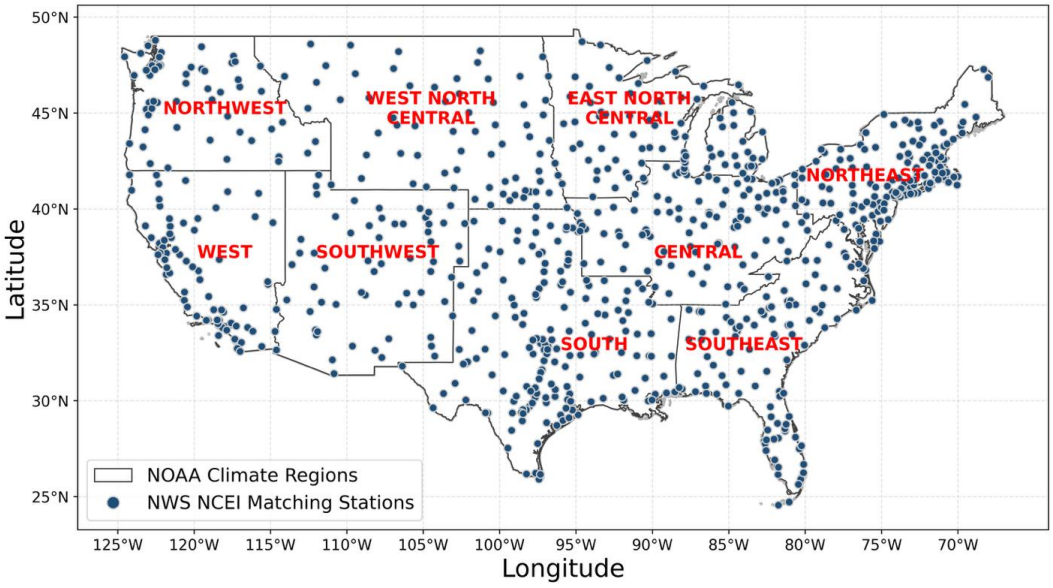


Figure 5. Spatial distribution of 845 matched NWS-NCEI stations across CONUS used for DL-based ETo forecasting. Polygons indicate NOAA climate regions.

2.4 Results and Discussion

2.4.1 NeuralFAO56 Implementation

We implemented NeuralFAO56 in Mode 2 across all 845 matched NWS-NCEI stations. Under this configuration, both physics-based FAO-56 ETo computation and DL-based ETo forecasting were executed at each station. The FAO-56 ETo was computed using observed meteorological variables, providing a baseline representation of ETo dynamics. In parallel, DL-based ETo forecasts were generated using two architectures, LSTM and Transformer, trained on station-specific historical datasets following the data preprocessing pipeline and default model configurations defined within the framework.

DL forecasting is performed using default parameters defined in the `Config` class. To support real-time deployment, model configurations are intentionally constrained to limit computational overhead. Training is conducted with a fixed random seed to ensure reproducibility, using a batch size of 64, a learning rate of 0.001, and a maximum of 100 epochs with early stopping applied using a patience of 10 epochs. Preliminary sensitivity experiments were conducted with multiple look-back windows (e.g., 3, 7, 10, and 14 days). Performance improvements dropped beyond 7 days, while computational costs increased. Therefore, a 7-day window was adopted as a parsimonious configuration for look-back window and used as model hyperparameter. These parameters can be modified by the user through the `Config` class.

The DL-based forecasts are then merged with FAO56 estimates and used for computing performance evaluation and visualization through the `accuracy_metrics` and `visualizations` modules within the `postprocess` component respectively. All outputs generated by the framework are organized within a structured directory (`runs/`), where a subdirectory named `{lat}_{lon}` is created during execution. Through this adaptive, data-aware workflow, the proposed NeuralFAO56 is able to balance spatial proximity, climatological consistency, and data availability. Physics-based ETo estimation is available

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

across a broad spatial domain using real-time and forecasted data, while DL-based forecasting is selectively enabled where historical records permit.

Model evaluation was conducted at regional scale by aggregating station-level metrics within NOAA climate regions. This enabled the characterization of spatial patterns in forecasting performance and the assessment of regional variability in model behavior across heterogeneous climatic regimes. Moreover, station-level forecasts were analyzed for four representative stations (KBGM, Binghamton, NY; KLLJ, Challis, ID; KSLN, Salina, KS; and KRQE, Window Rock, AZ), selected across distinct NOAA climate regions. Station identifiers correspond to International Civil Aviation Organization (ICAO) codes. To assess forecast performance under real-time conditions, FAO-56 ETo estimates derived from short-term forecasted meteorological inputs were additionally generated for each station. DL-based ETo forecasts were produced over the same forecast horizon. Validation was performed once corresponding observed meteorological data became available, enabling both forecast-driven FAO-56 ETo and DL-based ETo forecasts to be evaluated against observed FAO-56 ETo. This comparison was conducted exclusively over the seven-day forecasting period beyond the final date of testing dataset (8 – 14 November 2025), ensuring consistency across all forecasting approaches. Finally, we assessed the computational performance of NeuralFAO56 to evaluate the feasibility of real-time deployment. This analysis compared the training overhead associated with DL-based forecasting against the runtime cost of physics-based FAO-56 computation.

2.4.2 Daily ETo Simulations

Tables 1 and 2 summarize station-level simulation performance of NeuralFAO56 averaged over each NOAA climate region for two DL architectures, LSTM and Transformer, at short-range (1-day-ahead) and medium-range (7-day-ahead) forecast horizons. These results are further illustrated spatially in Fig. 6, which shows the corresponding spatial distribution of NSE, KGE, and MAE across stations for both models and forecast horizons. Overall, predictive performance degrades with increasing lead time, while no significant differences are observed between the two DL architectures.

The superior performance at the 1-day forecast horizon compared to the 7-day horizon is primarily attributed to the decay of temporal autocorrelation and the inherent difficulty of multi-step sequence prediction. At short lead times, ETo exhibits strong persistence, allowing both DL models to leverage recent observations for accurate predictions. In contrast, at longer lead times, dependence on recent states weakens, requiring the models to infer future conditions without direct input information, leading to reduced predictive skill.

At the 1-day forecast horizon (Table 1; Fig. 6a), both models achieved strong predictive performance across most regions. The northwest and southwest regions exhibit the highest performance with NSE and KGE values of at least 0.84 and 0.87 respectively. This behavior reflects strong seasonal forcing and relatively stable atmospheric conditions, where evapotranspiration is primarily governed by smooth variations in radiation and temperature values. Similarly, the northeast and east north central regions demonstrate robust performance, with mean NSE values approaching 0.80 and MAE values ranging from 0.58 to 0.62 mm day^{-1}, indicating reliable short-range forecasting under humid continental climate. In contrast, lower predictive skill is observed in the south and southeast regions, where higher humidity variability, convective weather systems, and stronger synoptic-scale disturbances introduce increased short-term atmospheric variability. This resulted in reduced NSE values (~0.72) in these regions.

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

Table 1. Station-level simulation performance of NeuralFAO56 at 1-day lead time, summarized as mean values within each NOAA climate region. The best-performing model for each region is highlighted in bold.

Region	NSE		KGE		MAE (mm day^{-1})	
	LSTM	Transformer	LSTM	Transformer	LSTM	Transformer
Central	**0.76**	**0.76**	0.81	**0.82**	**0.64**	**0.64**
East North Central	0.79	**0.80**	**0.84**	**0.84**	**0.62**	**0.62**
Northeast	**0.78**	**0.78**	0.83	**0.84**	**0.58**	**0.58**
Northwest	**0.87**	**0.87**	**0.90**	**0.90**	**0.52**	**0.52**
South	**0.72**	**0.72**	**0.78**	**0.78**	**0.79**	**0.79**
Southeast	0.71	**0.72**	**0.77**	**0.77**	**0.56**	**0.56**
Southwest	0.84	**0.85**	0.87	**0.88**	0.74	**0.73**
West	**0.80**	**0.80**	**0.85**	**0.85**	0.61	**0.60**
West North Central	**0.78**	**0.78**	**0.83**	**0.83**	**0.84**	**0.84**

On the other hand, a decline in predictive skill is observed across all regions at the 7-day forecast horizon (see Table 2 and Fig. 6b), reflecting increased uncertainty at longer lead times. The northwest region continues to demonstrate the strongest performance with consistently high NSE (0.81) and KGE (≥0.85) values across both architectures. This indicates the persistence of predictable atmospheric patterns in this region. Similarly, the northeast and east north central regions also maintain relatively stable performance with NSE values of 0.74 and moderate increases in MAE. On the contrary, greater degradation is observed in the south, west, and west north central regions, where NSE decreases by approximately 0.10 and MAE increases by up to 0.19 mm day^{-1} relative to the 1-day-ahead forecasts. This degradation reflects reduced temporal persistence and increased variability in evapotranspiration dynamics, which limit the ability of DL models to accurately predict longer forecast horizons.

Table 2. Station-level simulation performance of NeuralFAO56 at 7-day lead time, summarized as mean values within each NOAA climate region. The best-performing model for each region is highlighted in bold.

Region	NSE		KGE		MAE	
	LSTM	Transformer	LSTM	Transformer	LSTM	Transformer
Central	**0.70**	**0.70**	**0.77**	**0.77**	**0.74**	**0.74**
East North Central	**0.74**	**0.74**	**0.80**	**0.80**	**0.71**	**0.71**
Northeast	**0.74**	**0.74**	**0.80**	**0.80**	**0.63**	0.64

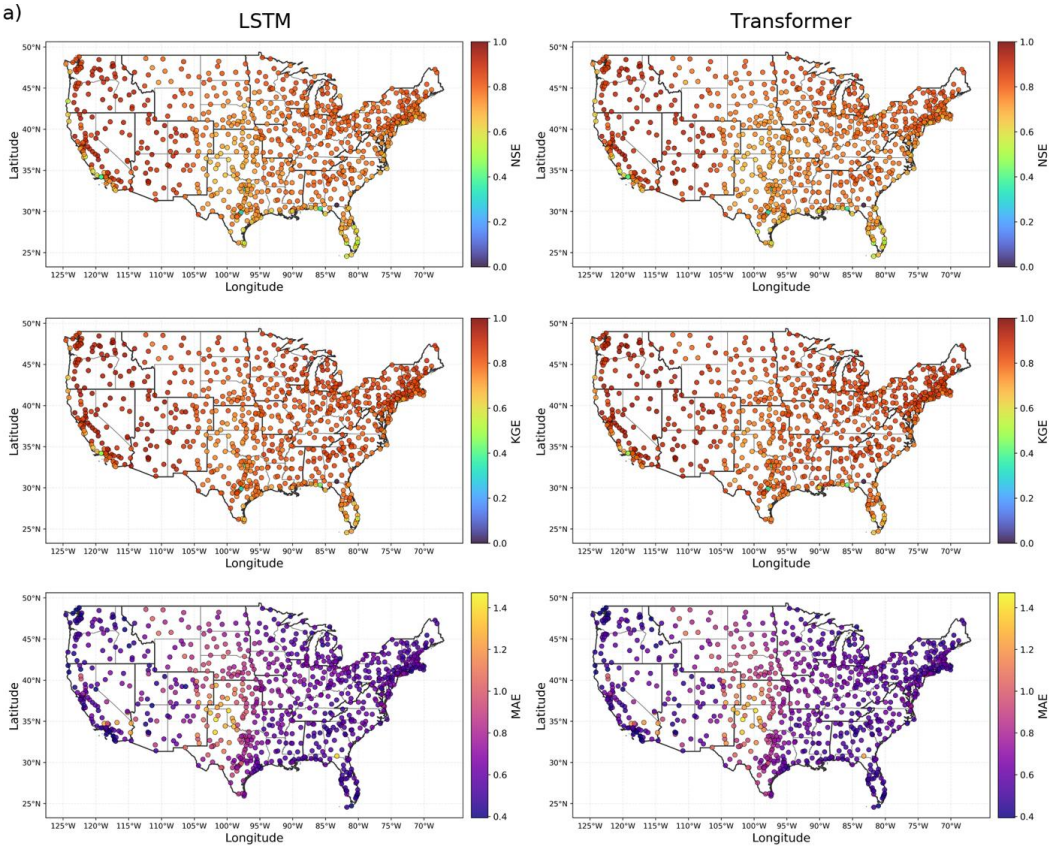
https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

Northwest	**0.81**	**0.81**	**0.86**	0.85	**0.65**	**0.65**
South	**0.62**	**0.62**	0.70	**0.71**	**0.94**	0.95
Southeast	**0.66**	0.65	**0.73**	**0.73**	**0.63**	**0.63**
Southwest	**0.77**	0.76	**0.81**	**0.81**	**0.93**	**0.93**
West	**0.71**	0.70	**0.76**	**0.76**	**0.76**	0.77
West North Central	**0.68**	**0.68**	**0.75**	**0.75**	**1.03**	**1.03**

Across both lead times, predictive skill is generally higher in northern regions relative to southern humid zones. This spatial pattern aligns with prior studies indicating reduced performance of data-driven models in warm and humid climates, where atmospheric variability and moisture dynamics introduce increased modeling complexity (Fan et al., 2018; Zhou et al., 2020).

The performance of LSTM and Transformer models remains nearly identical across both horizons, indicating that ETo forecasting is primarily constrained by data characteristics and inherent predictability rather than model architecture. Both models effectively capture the dominant seasonal and short-term dependencies in the data. This convergence in performance is further influenced by the use of lightweight model configurations designed for real-time applicability. In addition, the absence of station-specific hyperparameter tuning limits model-specific optimization, which may otherwise improve performance but would increase computational overhead and reduce operational efficiency.

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026



508

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

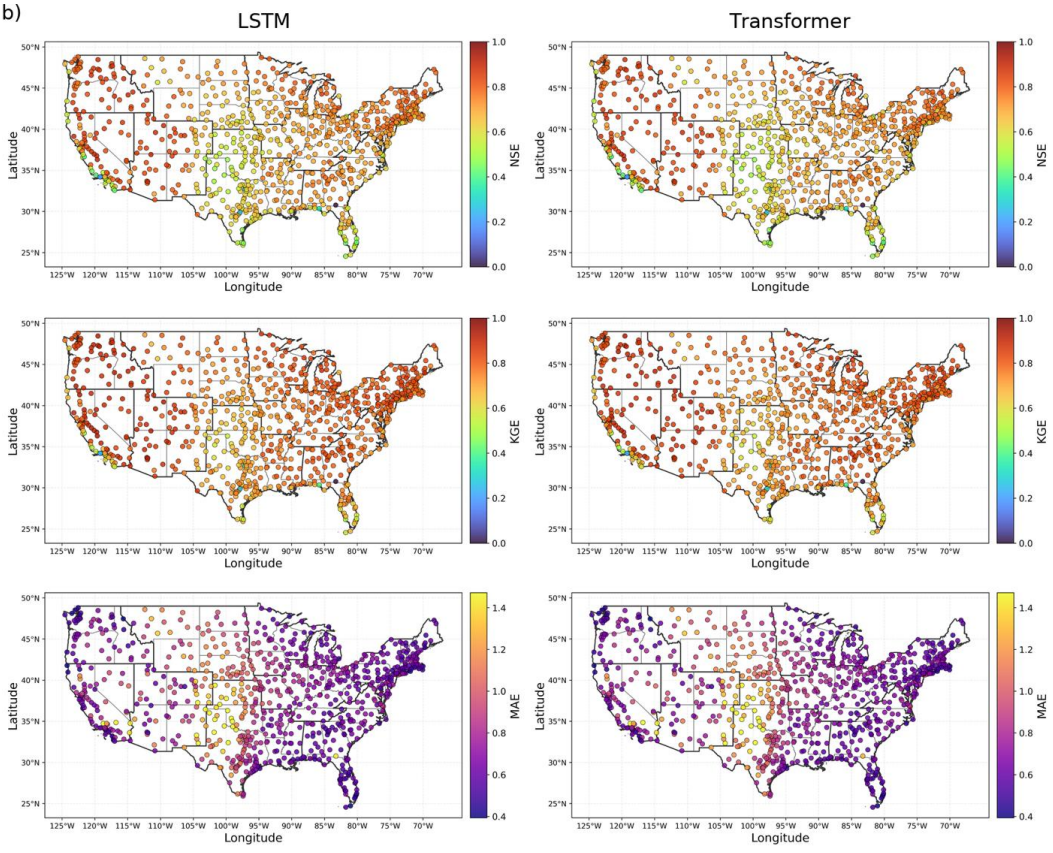


Figure 6. Spatial distribution of NeuralFAO56 forecasting performance across the nine NOAA climate regions over CONUS for (a) 1-day-ahead and (b) 7-day-ahead forecast horizons. Maps show station-level NSE, KGE, and MAE for LSTM (left column) and Transformer (right column) models.

2.4.3 Cumulative Distribution Functions (CDFs)

Figures 7 and 8 present the CDFs of station-level NSE and MAE for both DL architectures across all nine NOAA climate regions. These CDFs complement the regional mean statistics by providing a population-level perspective on forecast skill, allowing direct assessment of performance dominance, tail behavior, and robustness across heterogeneous hydroclimatic conditions. Across both metrics and forecast horizons, no systematic differences between model architectures are observed, with CDFs for LSTM and Transformer largely overlapping across most regions. This indicates that predictive performance is primarily constrained by data characteristics rather than model structure.

Across all regions, the NSE CDFs demonstrate a clear separation between short-range (1-day) and medium-range (7-day) forecasts (Fig. 7). At the 1-day horizon, both LSTM and Transformer models exhibit steep CDF slopes concentrated in the high-skill regime, indicating that a large fraction of stations achieve strong predictive performance. In the east north central, northeast, northwest, southwest, and west regions, most stations exceed NSE values of 0.80, reflecting highly consistent short-term forecasting capability across diverse climatic regimes. At the 7-day horizon, the NSE distributions shift leftward and become more gradual, consistent with increasing forecast uncertainty at longer lead times. Despite this shift, median NSE

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

values remain above ~0.70 in most regions, except for the south and west north central regions, where skill degradation is more pronounced.

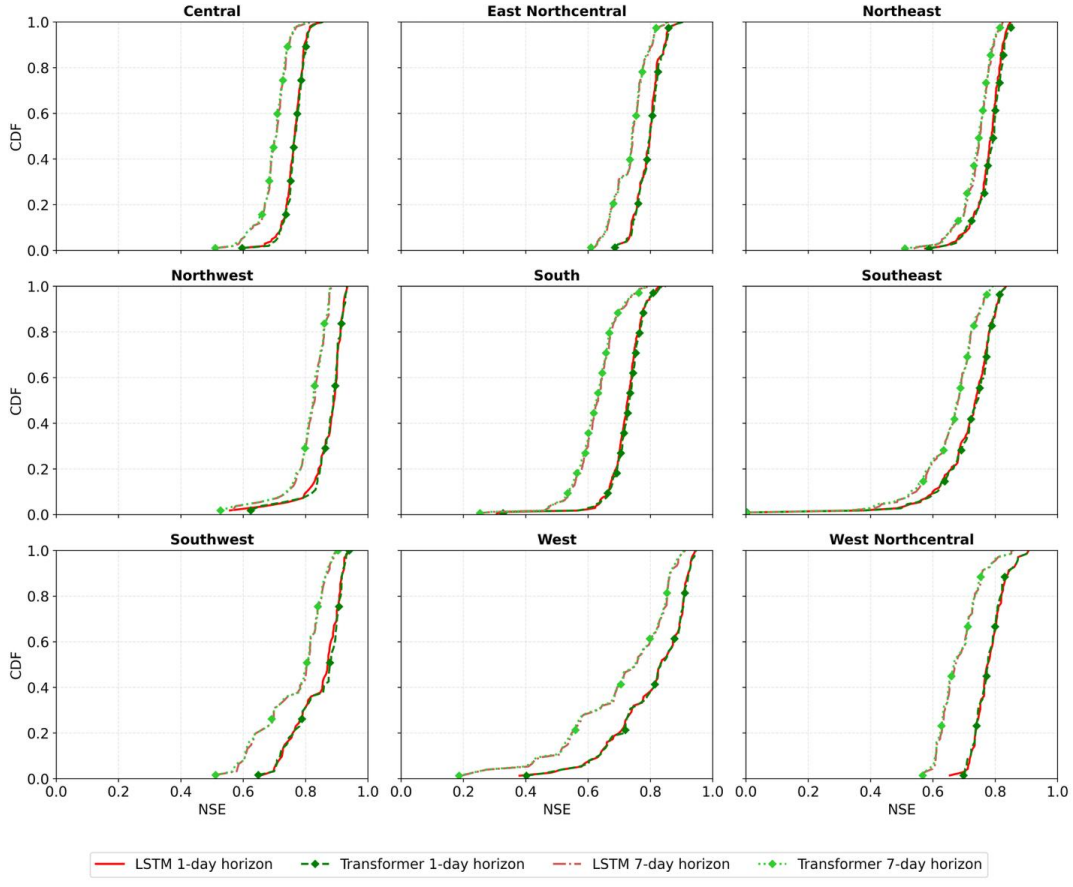


Figure 7. Regional CDFs of station-level NSE for NeuralFAO56 forecasts across nine NOAA climate regions. Results are shown for LSTM and Transformer models at 1-day and 7-day forecast horizons.

Complementary patterns are observed in the MAE CDFs (Fig. 8). For 1-day forecasts, both models produce steep CDFs concentrated at low error values, indicating that most stations achieve small absolute prediction errors. This behavior is particularly pronounced in the northwest and northeast regions, where a substantial proportion of stations maintain MAE values below approximately 0.60 mm day^{-1}. These results reflect the relatively consistent and predictable evapotranspiration dynamics in these regions. At the 7-day horizon, the MAE distributions shift rightward and develop heavier upper tails across all regions which suggest increasing uncertainty and cumulative propagation of meteorological forecast errors. This effect is strongest in the south and southwest regions, where convective activity, high evaporative demand, and greater atmospheric variability amplify sensitivity to forecast inaccuracies. In contrast, the northeast and southeast regions exhibit comparatively smaller rightward shifts, likely due to more moderate temperature variability and stronger climatological constraints on evapotranspiration processes.

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

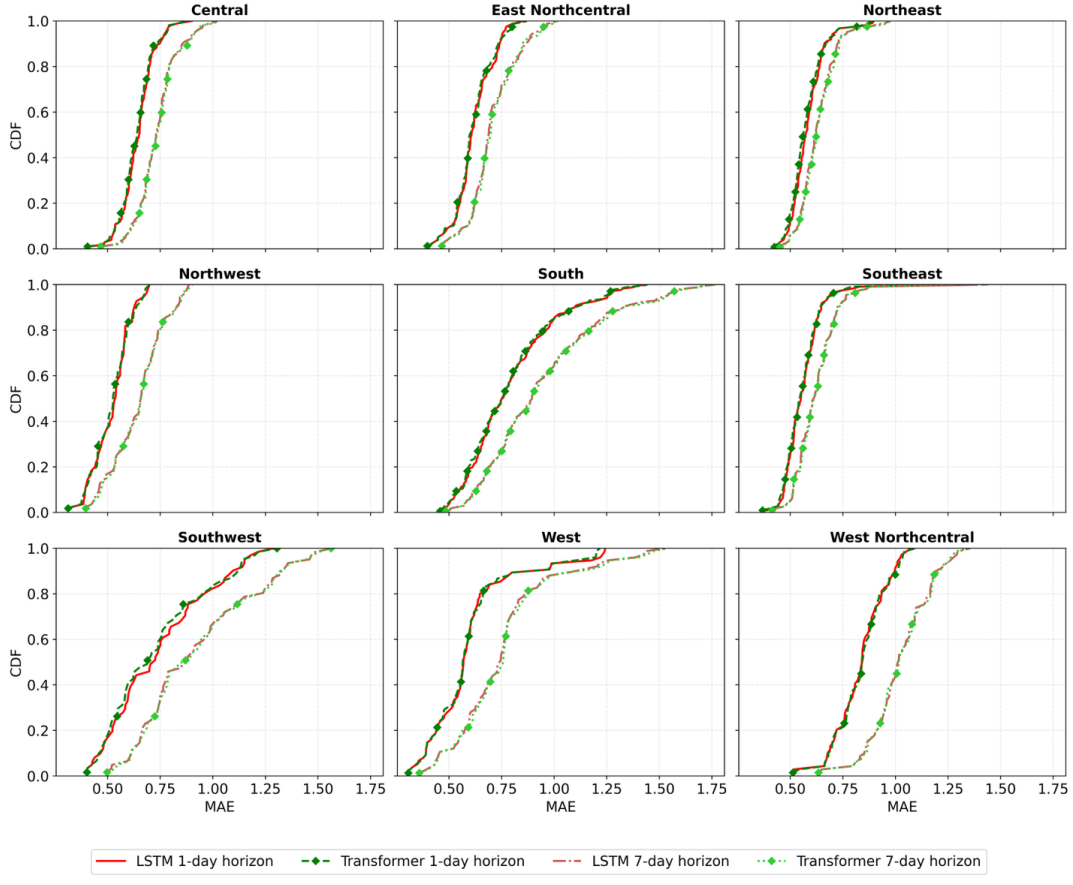


Figure 8. Regional CDFs of station-level MAE for NeuralFAO56 forecasts across nine NOAA climate regions. Distributions are shown for LSTM and Transformer models at 1-day and 7-day forecast horizons.

2.4.4 ETo Forecasts Across Gauging Stations

To complement the aggregate and distributional performance analysis, station-level forecasts were examined to evaluate the fidelity of NeuralFAO56 in reproducing daily ETo dynamics under contrasting hydroclimatic regimes. Representative stations were selected from four NOAA climate regions–northeast (KBGM), northwest (KLLJ), south (KSLN), and southwest (KRQE)–to capture a wide range of atmospheric variability, evaporative demand regimes, and seasonal forcing characteristics. For each station, both short-range (1-day-ahead) and medium-range (7-day-ahead) forecasts produced by the LSTM and Transformer models were compared against FAO-56 ETo (Fig. 9). Across all stations, the 1-day-ahead simulations demonstrate strong agreement with observed FAO-56 ETo, with both LSTM and Transformer accurately reproducing the dominant seasonal cycle, intra-seasonal variability, and short-term fluctuations.

In the northeast station (KBGM), characterized by moderate evaporative demand and relatively smooth seasonal transitions, both models closely track FAO-56 ETo with minimal phase lag and generally low peak bias. Peak summer demand and winter minima are captured with high fidelity, indicating robust short-term predictive capability under humid continental climate conditions. Both models exhibit almost similar performances, with NSE ~ 0.79 and MAE = 0.57 mm day^{-1}.

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

The northwest station (KLLJ) exhibits highly regular seasonal patterns and comparatively stable atmospheric forcing. Under these conditions, both models demonstrate particularly strong agreement with FAO-56 ETo, with nearly overlapping trajectories throughout the evaluation period. The smooth seasonal structure is well preserved, and daily variability is accurately represented, consistent with high NSE values (≥ 0.92) and low MAE values (0.48 mm day^{-1}). This performance reflects the enhanced predictability associated with persistent atmospheric patterns in this region.

In contrast, the south station (KSLN) presents a more challenging forecasting environment due to higher atmospheric variability, frequent convective activity, and elevated evaporative demand. While both models capture the overall seasonal structure, increased short-term variability and occasional underestimation of peak ETo values are evident. This behavior is reflected in reduced short-range performance skills (NSE ~ 0.72) and higher absolute errors (MAE ~ 1.00 mm day^{-1}). Despite these challenges, no systematic drift is observed, indicating stable temporal behavior even under dynamically complex atmospheric conditions.

The southwest station (KRQE), representing arid and semi-arid climatic conditions with radiation-driven evaporative demand, further highlights the models' ability to generalize across distinct regimes. Both architectures successfully track large seasonal amplitudes and elevated summer ETo values while preserving the timing of peak demand periods. Performance metrics exceed NSE = 0.91 for both models, with MAE values near 0.50 mm day^{-1}, demonstrating strong predictive skill under high-demand conditions. On the 7-day horizon, both models exhibit increased smoothing and reduced sensitivity to short-term fluctuations relative to the 1-day forecasts for all stations. While fine-scale variability is attenuated, the dominant seasonal cycle and longer-term trends remain well preserved. This controlled degradation is consistent with horizon-dependent performance patterns observed in the aggregate and distributional analyses.

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

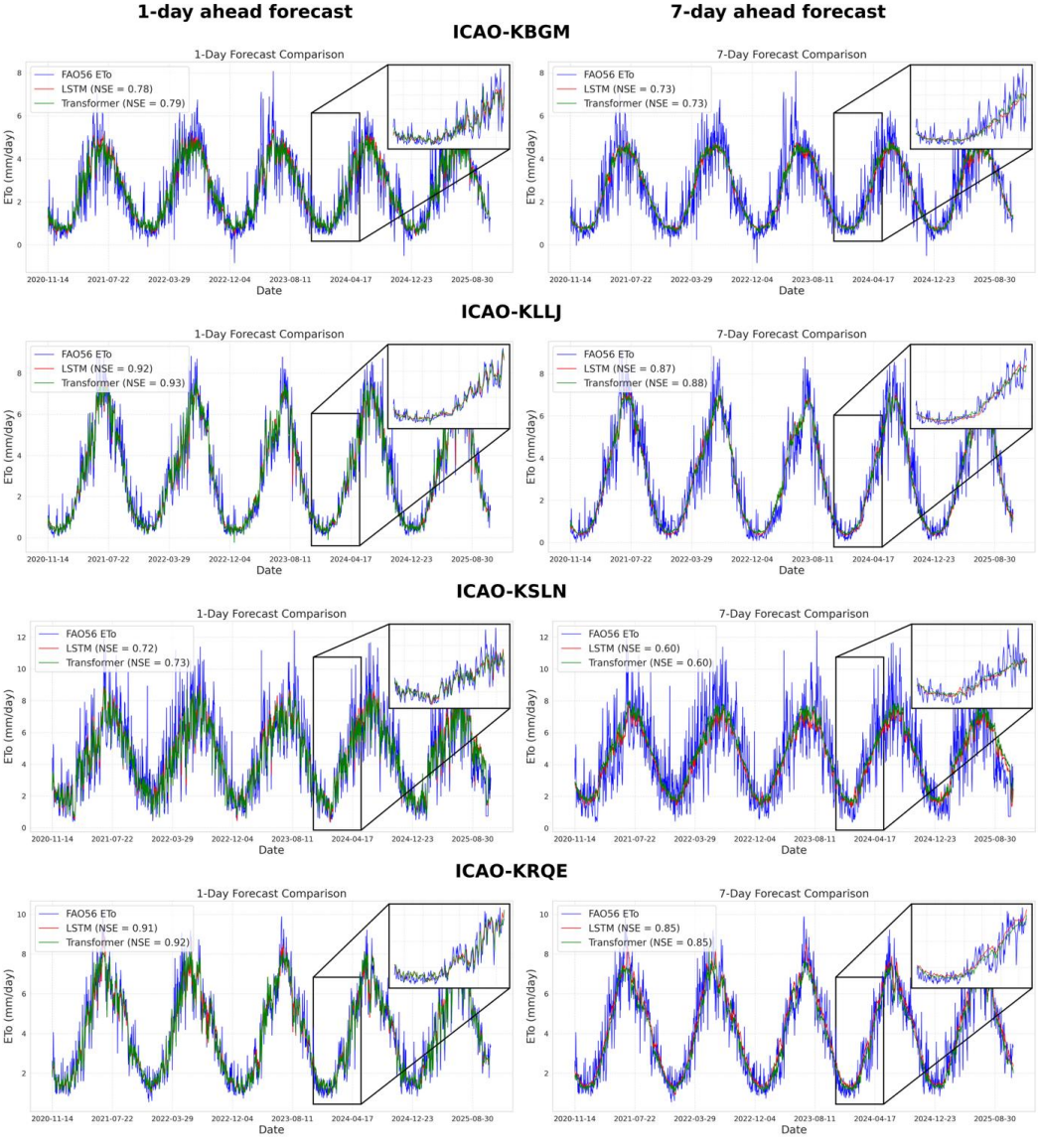


Figure 9. Time-series comparison of daily ETo forecasts using NeuralFAO56 for representative stations across four NOAA climate regions. Panels show observed FAO-56 ETo and corresponding predictions from LSTM and Transformer models for 1-day-ahead and 7-day-ahead forecast horizons. Stations include KBGM (northeast), KLLJ (northwest), KSLN (south), and KRQE (southwest), representing diverse hydroclimatic regimes. Subplots provide magnified views of selected time windows to illustrate the ability of the DL models to capture short-term ETo variability and peak dynamics.

2.4.5 Real-time ETo Forecasting

To evaluate NeuralFAO56 under true real-time forecasting conditions, model forecasts were assessed during the post-testing deployment window spanning 8-14 November 2025, corresponding to the 7-day

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

meteorological forecast horizon provided by NWS. Both approaches, FAO-56 ETo computed from forecasted meteorological data and DL-based forecasts, were subsequently evaluated against FAO-56 ETo computed from observed meteorological data once they were available, providing a consistent reference for real-time accuracy assessment.

Figure 10a presents the distribution of forecast errors (MAE) across 845 CONUS stations for lead times from 1 to 7 days. Figure 10b complements this analysis by illustrating horizon-wise ETo forecasts at four representative stations (KBGM, KLLJ, KSLN, and KRQE), consistent with those analyzed in Section 2.4.4. Overall, DL models provide smoother and more stable forecasts while preserving dominant temporal patterns, whereas physics-based forecasts remain more responsive to short-term variability but are increasingly affected by uncertainty in meteorological inputs at longer lead times.

The box-whisker plot (Fig. 10 a) shows that the forecast-driven FAO-56 approach exhibits a monotonic increase in MAE from day 1 to day 7. This trend reflects the accumulation of uncertainty in forecasted atmospheric variables, such as temperature, radiation, humidity, and wind speed, that directly propagate into physically based ETo calculations. In contrast, DL-based forecasts exhibit different error pattern, with an increase from day 1 to day 2, relatively stable errors through days 3–4, reduced errors at days 5–6, and increased dispersion again at day 7.

Across all horizons, LSTM and Transformer models show nearly overlapping MAE distributions, indicating comparable forecasting performance between these two architectures. Notably, at the intermediate lead times (5-6 days), both models achieve MAE values comparable to or lower than those of the forecast-driven FAO-56 estimates. This behavior indicates that both models maintain stable predictive skill by leveraging learned temporal structure in ETo dynamics allowing them to remain robust as temporal predictability decreases with increasing forecast horizon. The station-level analysis (Fig. 10b) further illustrates these patterns. Across all stations, both models consistently reproduce the overall trend of ETo forecasts across horizons, capturing gradual changes in evaporative demand. In contrast, observed FAO-56 ETo exhibits higher day-to-day variability, while forecast-driven FAO-56 partially captures these fluctuations but shows increasing deviation at longer lead times.

For KBGM (northeast), KLLJ (northwest), and KRQE (southwest), both DL models closely track observed ETo dynamics across most horizons, consistent with their strong performance reported in Section 3.3.3. Importantly, the forecast driven FAO-56 shows greater deviation in KBGM and KRQE compared to observed FAO-56 ETo. This is more pronounced for KRQE, where the ETo values computed using forecast driven FAO-56 exhibit higher error magnitude (ranging from 0.3 to 1.5 mm day^{-1}) as lead time increases. The DL-based forecasts, on the other hand, maintain lower error magnitudes (typically $\leq$ 0.4 mm day^{-1}), indicating more stable performance across horizons. These results suggest that stations exhibiting higher predictive skills during historical evaluation also tend to retain strong forecasting performance under real-time conditions.

In contrast, for KSLN from south region, which exhibits higher atmospheric variability, the forecast driven FAO-56 approach more closely follows short-term fluctuations in observed FAO-56 ETo. both models, however, primarily capture the broader trend and show larger deviations at shorter lead times (2–3 days), before improving at longer horizons. This behavior highlights a trade-off between capturing high-frequency variability and maintaining stable multi-step forecasts.

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

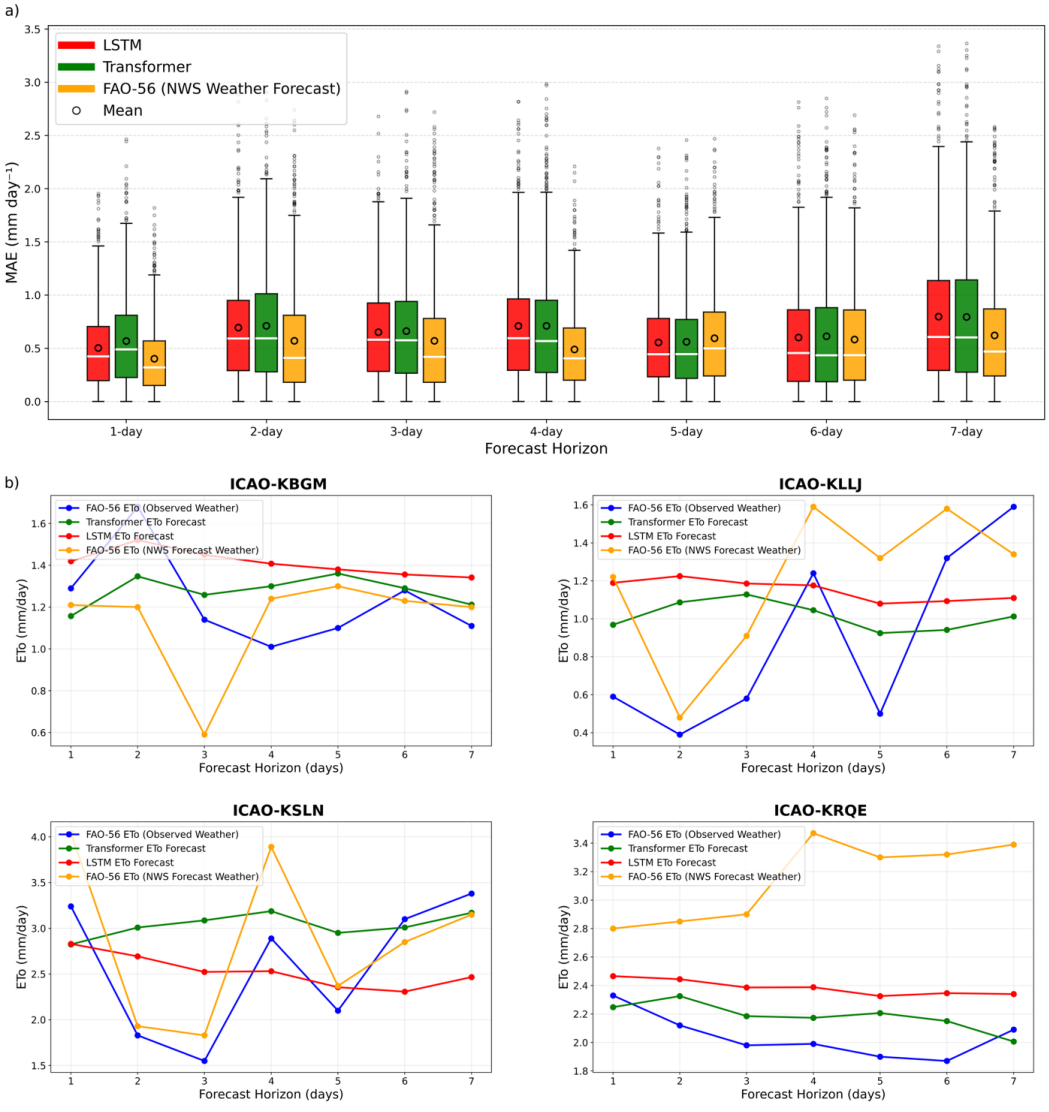


Figure 10. (a) Distribution of MAE for real-time ETo forecasts across 845 CONUS stations at 1–7 day lead times. Box-whisker plots summarize station-level MAE for LSTM and Transformer-based forecasts, along with FAO-56 ETo computed using NWS forecasted meteorological inputs, all evaluated against FAO-56 ETo derived from observed weather. White horizontal lines indicate medians; open circles denote mean MAE values; (b) Horizon-wise ETo forecasts for the same four representative stations analyzed in Fig. 9.

2.4.6 Computational Performance and Runtime Analysis

The computational feasibility of NeuralFAO56 for real-time deployment was assessed by evaluating the training overhead associated with DL-based forecasting (Mode 2) as well as the run time cost of FAO-56 ETo execution pipeline (Mode 1). Both models were executed on a high-performance computing cluster using an NVIDIA V100 GPU with 16 GB memory, while Mode 1 runtime was measured on CPU-based

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

execution. Table 3 summarizes the average training cost per station for the LSTM and Transformer models. The LSTM architecture converged in approximately 43 epochs on average, corresponding to a mean training time of 128.49 s per station. In contrast, the Transformer model required fewer convergence epochs (mean of 24.77), resulting in a substantially lower total training time of 78.73 seconds per station. This indicates that the reduced total training time of the Transformer model primarily results from faster convergence behavior. Despite this difference in overall training duration, both architectures exhibited comparable per-epoch computation costs, with average values of 3.02 s for LSTM and 3.23 s for Transformer.

Table 3. Average training computational cost per station for NeuralFAO56 DL models.

Model	Convergence Epochs	Training Time per Station (s)	Training Time per Epoch (s)
LSTM	43.31	128.49	3.02
Transformer	24.77	78.73	3.23

The relatively small per-epoch training time for both models reflects the lightweight architectural design adopted in NeuralFAO56, including the use of a single LSTM layer and a single Transformer encoder block with constrained embedding and feedforward dimensions. This design choice was motivated by the operational objective of enabling on-demand ETo forecasting while minimizing computational overhead. The observed training times demonstrate that station-specific model training can be completed within a few minutes, making large-scale deployment across hundreds of stations computationally feasible in batch-processing environments.

In addition to model training cost, the execution time of the physics-based FAO-56 pipeline (Mode 1) was evaluated. Across all stations, Mode 1 execution required an average of approximately 2.96 second per station, including automated station selection, data retrieval, preprocessing, and FAO-56 ETo computation. This low execution latency highlights the efficiency of the API-integrated workflow and supports near-real-time application of NeuralFAO56 for on-demand ETo estimation and short-term forecasting.

3 Conclusion

This study introduced NeuralFAO56, an integrated hybrid physics–based data-driven computational framework designed to automate real-time ETo estimation and forecasting across CONUS within a workflow to support irrigation demand estimation. By coupling FAO-56 Penman–Monteith computation with DL-based forecasting, the framework provides an end-to-end pipeline that enables automated data acquisition and standardized preprocessing for on-demand ETo prediction. This addresses a critical gap between methodological advances in evapotranspiration modeling and their practical deployment within irrigation decision support systems. The system was developed to operate in dual modes: (i) physics-based ETo estimation and forecast execution driven by real-time and forecasted meteorological forcings obtained through the NWS API, and (ii) data-driven ETo forecasting using sequential learning architectures trained on historical atmospheric observations retrieved from NCEI archives.

We demonstrated the applicability of NeuralFAO56 by evaluating the framework at continental scale across 845 NWS–NCEI matched stations. This large sample evaluation demonstrated that NeuralFAO56 achieves strong and spatially consistent predictive skill across heterogeneous hydroclimatic regimes. These assessments further reveal systematic spatial variability aligned with differences in regional atmospheric dynamics. Regions characterized by stable synoptic forcing and pronounced seasonal structure, such as the northwest and southwest, exhibited the highest forecasting skill. On the other hand, humid and convectively active regions, particularly the south and southeast, showed comparatively degraded performance, reflecting increased meteorological variability and nonlinear atmospheric interactions. Moreover, LSTM and Transformer models exhibited nearly identical performance across all NOAA regions and forecast

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

horizons. This indicates that ETo predictability is primarily constrained by data characteristics and temporal structure rather than model architecture, with the use of lightweight configurations potentially limiting further model differentiation.

Gauging station level experiments provided additional insight into model realism beyond statistical metrics. Across 845 stations selected to represent four distinct NOAA climate regions, both LSTM and Transformer architectures reproduced physically consistent ETo dynamics, including seasonal amplitudes, peak evaporative demand periods, and intra-seasonal variability. Even under extended forecast horizons, model trajectories retained phase coherence and avoided systematic drift, indicating stable temporal generalization and physically plausible forecast evolution.
Real-time ETo forecast further extended the analysis beyond retrospective testing, providing additional insight into real-world deployment performance. During the 7-day forecast horizon, physics-based forecasts exhibited systematic error growth with lead time, reflecting the accumulation of uncertainty in meteorological forcing variables. In contrast, DL-based forecasts displayed comparatively stable error distributions, with improved performance at intermediate lead times (5-6 days) compared to forecast-driven FAO-56 estimates. Station-level analysis for four representative stations indicated that LSTM and Transformer models consistently reproduced the overall trend of ETo forecasts, capturing gradual changes in evaporative demand. In comparison, forecast-driven FAO-56 estimates captured the short-term variability but exhibited increasing deviation at longer lead times. Importantly, in both models, stations exhibiting higher predictive skill during historical evaluation generally retained strong performance under real-time conditions, indicating stable generalization across deployment scenarios. The complementary strengths observed between physics-based and data-driven forecasts highlight the operational value of hybrid system design, wherein physically interpretable modeling and data-driven temporal learning jointly enhance evapotranspiration prediction under real-world forecasting constraints.

Computational runtime analysis further confirmed the operational feasibility of the framework. Physics-based FAO-56 execution demonstrated low per-station runtime, supporting near-real-time on-demand deployment. DL training costs, when executed on GPU infrastructure, remained computationally manageable, with convergence achieved within moderate epoch counts and stable per-epoch runtime, demonstrating that large-scale continental deployment is computationally tractable.

From an operational perspective, the forecasting skill demonstrated by NeuralFAO56 supports short- to medium-range irrigation planning, including irrigation timing optimization, deficit irrigation scheduling, and real-time water allocation decision making. The ability to deliver consistent predictive performance across geographically distributed stations and heterogeneous climatic regimes underscores the scalability of the framework. This spatial generalizability, combined with automated data ingestion and near-real-time execution capability, reinforces its reliability, scalability, and operational readiness of environmental prediction systems.

Code and Data Availability

The source code for NeuralFAO56 (v 1.0) is archived on Zenodo: https://doi.org/10.5281/zenodo.19696546 (Neupane and Samadi, 2026). The input datasets, model outputs, and scripts used to reproduce all figures and results presented in this study are archived on Zenodo: https://doi.org/10.5281/zenodo.19858713 (Neupane and Samadi, 2026). Meteorological data used in this study were obtained from publicly accessible sources provided by the National Oceanic and Atmospheric Administration (NOAA). Real-time and forecast meteorological data were retrieved from the National Weather Service (NWS) API (https://api.weather.gov), while historical observations were obtained from the National Centers for Environmental Information (NCEI) archives

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

(https://www.ncei.noaa.gov/access/services/data/v1). All data are openly available and can be accessed programmatically via the respective APIs.

Author contribution

Adarsha Neupane: Writing – original draft, Data curation, Formal analysis, Conceptualization, Methodology, Visualization. Vidya Samadi: Writing – review and editing, Supervision, Conceptualization, Methodology, Funding acquisition, Project administration.

Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Acknowledgements

The authors appreciate the funding support from USDA-NIFA (grant # 2023000603). Any opinions, findings, and discussions expressed in this study are those of the authors and do not necessarily reflect the views of USDA-NIFA. We gratefully acknowledge NWS and NCEI for providing open-access meteorological data through their public APIs, which facilitated retrieval of sub-daily and daily climatic datasets used in this research. Clemson University is acknowledged for generously allocating computing time on the Palmetto cluster.

References

Ahmadi, A., Daccache, A., Sadegh, M., & Snyder, R. L. (2023). Statistical and deep learning models for reference evapotranspiration time series forecasting: A comparison of accuracy, complexity, and data efficiency. Computers and Electronics in Agriculture, 215, 108424. https://doi.org/10.1016/j.compag.2023.108424

Ali, M., Nayahi, J. V., Abdi, E., Ghorbani, M. A., Mohajeri, F., Farooque, A. A., & Alamery, S. (2025). Improving daily reference evapotranspiration forecasts: Designing AI-enabled recurrent neural networks based long short-term memory. Ecological Informatics, 85, 102995. https://doi.org/10.1016/j.ecoinf.2025.102995

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, Rome, 300(9), D05109.

ASCE Environmental and Water Resources Institute, Evapotranspiration in Irrigation and Hydrology Committee. (2025). Evapotranspiration terminology and definitions. Journal of Irrigation and Drainage Engineering, 151(5), 06025003

Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). Layer normalization. arXiv preprint arXiv:1607.06450.

Babaeian, E., Paheding, S., Siddique, N., Devabhaktuni, V. K., & Tuller, M. (2022). Short- and mid-term forecasts of actual evapotranspiration with deep learning. Journal of Hydrology, 612, 128078. https://doi.org/10.1016/j.jhydrol.2022.128078

Bellido-Jiménez, J. A., Estevez, J., Vanschoren, J., & García-Marín, A. P. (2022). AgroML: an open-source repository to forecast reference evapotranspiration in different geo-climatic conditions using machine learning and transformer-based models. Agronomy, 12(3), 656

Bilal, M., Ali, G., Iqbal, M. W., Anwar, M., Malik, M. S. A., & Kadir, R. A. (2022). Auto-prep: efficient and automated data preprocessing pipeline. Ieee Access, 10, 107764-107784

De Oliveira E Lucas, P., Alves, M. A., De Lima E Silva, P. C., & Guimarães, F. G. (2020). Reference evapotranspiration time series forecasting with ensemble of convolutional neural networks. Computers and Electronics in Agriculture, 177, 105700. https://doi.org/10.1016/j.compag.2020.105700

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

Dufter, P., Schmitt, M., & Schütze, H. (2022). Position information in transformers: An overview. Computational Linguistics, 48(3), 733-763

Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., Lu, X., & Xiang, Y. (2018). Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. Agricultural and Forest Meteorology, 263, 225–241. https://doi.org/10.1016/j.agrformet.2018.08.019

Ferreira, L. B., & da Cunha, F. F. (2020). Multi-step ahead forecasting of daily reference evapotranspiration using deep learning. Computers and electronics in agriculture, 178, 105728

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. Neural computation, 12(10), 2451-2471

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Han, J., Pei, J., & Tong, H. (2022). Data mining: concepts and techniques. Morgan kaufmann

Hargreaves, G., & Samani, Z. (1985). Reference Crop Evapotranspiration From Temperature. Applied Engineering in Agriculture, 1. https://doi.org/10.13031/2013.26773

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. Neural Computation. https://doi.org/10.1162/neco.1997.9.8.1735

Jensen, M. E., & Allen, R. G. (2016). Evaporation, evapotranspiration, and irrigation water requirements: Task Committee on Revision of Manual 70

Karbasi, M., Jamei, M., Ali, M., Malik, A., & Yaseen, Z. M. (2022). Forecasting weekly reference evapotranspiration using Auto Encoder Decoder Bidirectional LSTM model hybridized with a Boruta-CatBoost input optimizer. Computers and Electronics in Agriculture, 198, 107121. https://doi.org/10.1016/j.compag.2022.107121

Katrompas, A., Ntakouris, T., & Metsis, V. (2022, June). Recurrence and self-attention vs the transformer for time-series classification: A comparative study. In International Conference on Artificial Intelligence in Medicine (pp. 99-109). Cham: Springer International Publishing

Li, M., Zhou, Q., Han, X., & Lv, P. (2024). Prediction of reference crop evapotranspiration based on improved convolutional neural network (CNN) and long short-term memory network (LSTM) models in Northeast China. Journal of Hydrology, 645, 132223. https://doi.org/10.1016/j.jhydrol.2024.132223

Li, Y. (2025). Theoretical Analysis of Positional Encodings in Transformer Models: Impact on Expressiveness and Generalization (arXiv:2506.06398). arXiv. https://doi.org/10.48550/arXiv.2506.06398

Medina, H., Tian, D., Srivastava, P., Pelosi, A., & Chirico, G. B. (2018). Medium-range reference evapotranspiration forecasts for the contiguous United States based on multi-model numerical weather predictions. Journal of Hydrology, 562, 502–517. https://doi.org/10.1016/j.jhydrol.2018.05.029

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology, 10(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Neupane, A., & Samadi, V. (2025). “QuantumIrrigation” – a new quantum computing python package for irrigation demand assessment. Smart Agricultural Technology, 12, 101523. https://doi.org/10.1016/j.atech.2025.101523

Neupane, A., & SAMADI, V. (2026). NeuralFAO56 v1.0. Zenodo. https://doi.org/10.5281/zenodo.19696546

Neupane, A., & SAMADI, V. (2026). CONUS-scale implementation of NeuralFAO56 v1.0: data, model outputs, and supporting materials for GMD submission [Data set]. Zenodo. https://doi.org/10.5281/zenodo.19858713

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

Neupane, A., & Sawada, Y. (2023). Performance Assessment of Irrigation Projects in Nepal by Integrating Landsat Images and Local Data. Remote Sensing, 15(18), 4633. https://doi.org/10.3390/rs15184633

NOAA National Centers for Environmental Information. 2026. Integrated Surface Database (ISD) station history file (isd-history.txt). National Oceanic and Atmospheric Administration. Available at: https://www.ncei.noaa.gov/pub/data/noaa/isd-history.txt (accessed February 9, 2026).

Pereira, L. S., Paredes, P., & Jovanovic, N. (2020). Soil water balance models for determining crop water and irrigation requirements and irrigation scheduling focusing on the FAO56 method and the dual Kc approach. Agricultural Water Management, 241, 106357. https://doi.org/10.1016/j.agwat.2020.106357

Priestley, C. H. B., & Taylor, R. J. (1972). On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. Monthly Weather Review, 100(2), 81–92. https://doi.org/10.1175/1520-0493(1972)100%253C0081:OTAOSH%253E2.3.CO;2

Qian, H., Wang, W., & Chen, G. (2024). Assessing forecast performance of daily reference evapotranspiration: A comparison of equations, machine and deep learning using weather forecasts. Journal of Hydrology, 644, 132101. https://doi.org/10.1016/j.jhydrol.2024.132101

Richards, M. (2019). PyETo, https://github.com/woodcrafty/PyETo (last access: 20 December 2025)

Saberian, M., Zafarmomen, N., Neupane, A., Panthi, K., & Samadi, V. (2025). HydroQuantum: A new quantum-driven Python package for hydrological simulation. Environmental Modelling & Software, 106736

Sun, Zhenyuan, et al. "Estimation of daily reference crop evapotranspiration in China based on Time-Space-LSTM model." Journal of Hydrology 660 (2025): 133430

Tayyaba, B., Khan, M. U. G., Waheed, T., Al-Otaibi, S., & Saba, T. (2025). TMRE: Novel Algorithm for Computing Daily Reference Evapotranspiration Using Transformer-Based Models. Computers, Materials & Continua, 83(2), 2851–2864. https://doi.org/10.32604/cmc.2025.060365

Technical Committee on Standardization of Reference Evapotranspiration. (2005). The ASCE Standardized Reference Evapotranspiration Equation (R. G. Allen, I. A. Walter, R. L. Elliott, T. A. Howell, D. Itenfisu, M. E. Jensen, & R. L. Snyder, Eds.). American Society of Civil Engineers. https://doi.org/10.1061/9780784408056

Thorp, K. R. (2022). pyfao56: FAO-56 evapotranspiration in Python. SoftwareX, 19, 101208. https://doi.org/10.1016/j.softx.2022.101208

Umar, M. A., Chen, Z., Shuaib, K., & Liu, Y. (2025). Effects of feature selection and normalization on network intrusion detection. Data Science and Management, 8(1), 23-39

Umutoni, L., & Samadi, V. (2024). Application of machine learning approaches in supporting irrigation decision making: A review. Agricultural Water Management, 294, 108710. https://doi.org/10.1016/j.agwat.2024.108710

Umutoni, L., Samadi, V., Vellidis, G., Privette Iii, C., Payero, J., & Koc, B. (2025). Decoding time: Unraveling the power of N-BEATS and N-HiTS vs. LSTM for accurate soil moisture prediction. Computers and Electronics in Agriculture, 237, 110614. https://doi.org/10.1016/j.compag.2025.110614

Valipour, M., Khoshkam, H., Bateni, S. M., Jun, C., & Band, S. S. (2023). Hybrid machine learning and deep learning models for multi-step-ahead daily reference evapotranspiration forecasting in different climate regions across the contiguous United States. Agricultural Water Management, 283, 108311. https://doi.org/10.1016/j.agwat.2023.108311

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need (arXiv:1706.03762). arXiv. https://doi.org/10.48550/arXiv.1706.03762

Vremec, M., Collenteur, R. A., & Birk, S. (2024). PyEt v1. 3.1: a Python package for the estimation of potential evapotranspiration. Geoscientific Model Development, 17(18), 7083-7103

https://doi.org/10.5194/egusphere-2026-2300
Preprint. Discussion started: 3 July 2026

902 Wang, Q., Li, B., Xiao, T., Zhu, J., Li, C., Wong, D. F., & Chao, L. S. (2019, July). Learning deep
903 transformer models for machine translation. In Proceedings of the 57th annual meeting of the
904 association for computational linguistics (pp. 1810-1822)
905 Ye, Y., González-Vidal, A., Zamora-Izquierdo, M. A., & Skarmeta, A. F. (2025). Transfer and deep
906 learning models for daily reference evapotranspiration estimation and forecasting in Spain from
907 local to national scale. Smart Agricultural Technology, 11, 100886.
908 Zhang, L., Zhao, X., Zhu, G., He, J., Chen, J., Chen, Z., Traore, S., Liu, J., & Singh, V. P. (2023). Short-
909 term daily reference evapotranspiration forecasting using temperature-based deep learning
910 models in different climate zones in China. Agricultural Water Management, 289, 108498.
911 https://doi.org/10.1016/j.agwat.2023.108498
912 Zhang, Y., & Post, D. (2018). How good are hydrological models for gap-filling streamflow data?.
913 Hydrology and Earth System Sciences, 22(8), 4593-4604.
914 Zhou, Z., Zhao, L., Lin, A., Qin, W., Lu, Y., Li, J., Zhong, Y., & He, L. (2020). Exploring the potential of
915 deep factorization machine and various gradient boosting models in modeling daily reference
916 evapotranspiration in China. Arabian Journal of Geosciences, 13(24), 1287.
917 https://doi.org/10.1007/s12517-020-06293-8
918