

Received 18 January 2025, accepted 3 February 2025, date of publication 5 February 2025, date of current version 14 February 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3539581

RESEARCH ARTICLE

Estimating Near-Surface Air Temperature From Satellite-Derived Land Surface Temperature Using Temporal Deep Learning: A Comparative Analysis

JANGHO LEE 

Earth and Environmental Sciences, University of Illinois Chicago, Chicago, IL 60607, USA

e-mail: jholee@uic.edu


This work was supported by U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research's Urban Integrated Field Laboratories CROCUS Project Research Activity, under Award DE-SC0023226.

ABSTRACT This study develops and compares three deep learning methods—LSTM, TCN, and N-BEATS—for estimating near-surface air temperature (T2M) from satellite-derived land surface temperature (LST) and land cover metrics such as NDVI and NDBI. By incorporating temporal context through varying look-back windows, these models substantially outperform non-temporal baselines, reducing root-mean-square error (RMSE) from around 2.6–2.8°C to below 1.8°C, and underscoring the value of historical LST observations for capturing the evolving surface–air temperature relationship. Longer lags generally improve accuracy, although N-BEATS performance plateaus beyond a certain window, reflecting both diminishing returns and practical limitations linked to missing cloud-free satellite data. Seasonal and diurnal evaluations show higher errors in spring and midday hours, likely due to rapid vegetation changes and stronger physical and dynamical processes that make T2M less predictable. Spatially, stations with denser vegetation exhibit elevated errors, suggesting that transpiration and canopy effects complicate the LST–T2M linkage. For extreme-event detection, LSTM provides the fewest false alarms (highest precision), N-BEATS captures the most extremes (highest recall), and TCN offers the best overall balance in precision and recall (highest F1). While cloud-free satellite coverage remains a limitation, future work could explore adaptive lag strategies, additional data sources, and more advanced data-fusion techniques. These results highlight that satellite-based temperature monitoring, when combined with suitable deep learning architectures, can reliably estimate T2M based on LST, further addressing gaps in near-surface observations and facilitating the detection of critical T2M extremes. This framework has direct applications in heat-warning systems, resource management, precision agriculture, and urban climate adaptation, and stands to benefit further from ongoing advancements in satellite sensing technology.

INDEX TERMS Air temperature, atmosphere, climate informatics, deep learning, GOES-R, land-atmosphere interaction, land surface temperature, landsat, LSTM, machine learning, N-BEATS, near surface air temperature, remote sensing, statistical climatology, TCN.

I. INTRODUCTION

A major consequence of global warming is the increase in near-surface air temperature, or 2-m air temperature (T2M), often resulting in more frequent and intense temperature extremes. These extremes can negatively affect human

The associate editor coordinating the review of this manuscript and approving it for publication was Stefania Bonafoni .

health [1], [2], [3], energy consumption [4], [5], [6], agricultural productivity [7], [8], [9], and infrastructure [10], [11], [12]. The spatiotemporal variability of T2M is further complicated by the heterogeneity of land cover and other environmental factors—such as vegetation, urbanization, latitude, and time of day—that influence the land–atmosphere energy balance [13], [14], [15], [16]. Despite the importance of characterizing T2M across extensive spatial and temporal

scales, continuous and comprehensive measurements of T2M remain limited.

Several sources of T2M data exist, each with its own benefits and drawbacks. Station-based measurements, for example, offer local, high-quality records of temperature. However, their utility often suffers from sparse spatial coverage and issues related to station placement. In Illinois (USA), only 63 meteorological stations serve an area of approximately 150,000 km², providing insufficient data density to capture finer-scale temperature variations.

Reanalysis datasets can act as another major T2M source; they integrate diverse observational inputs into numerical weather prediction models and typically span large geographic regions (global) and extended periods [17], [18], [19]. However, reanalysis products can be prone to biases arising from model assimilation processes. Moreover, their relatively coarse resolution—on the order of 9 km for ERA-5—can hinder detailed local analyses [20], even though combining reanalysis data with regional climate or high-resolution numerical weather prediction models can partially improve spatial detail. These models, however, remain reliant on reanalysis-derived boundary conditions, which can cause underlying biases [21], [22], [23].

Finally, targeted observational campaigns—such as vehicle-based measurements—can collect high-resolution, localized T2M data. These campaigns are valuable for capturing microclimatic variations, especially in urban settings [24], [25], [26]. However, they tend to be limited in both temporal span and spatial coverage, and they can be logistically complex and resource-intensive—in both time and cost—to maintain on a large scale.

In contrast, satellite retrievals of land surface temperature (LST) are frequently available at finer resolutions and broader coverage. For instance, the Geostationary Operational Environmental Satellites (GOES-R) series (GOES-16, 17, and 18) provides hourly LST data over North America at roughly 2 km resolution [27], while instruments such as Moderate Resolution Imaging Spectroradiometer (MODIS) [28] or Ecosystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) [29] can collect near-daily or finer-scale measurements. These observations offer valuable insights into thermal conditions across diverse landscapes.

Thus, estimating T2M from satellite-derived LST appears attractive given the strong correlation between the two temperature variables. However, accurately retrieving T2M from LST remains challenging because LST and T2M exhibit distinct physical and dynamical characteristics—such as variations in the near-surface lapse rate [30], surface energy balance [31], and atmospheric mixing [32] within the boundary layer. Furthermore, vegetation strongly influences LST–T2M dynamics [33], [34], [35] by modulating energy exchanges through evapotranspiration, altering the partitioning of incoming radiation into latent and sensible heat fluxes, and exhibiting relatively low albedo [36], [37], [38]. These differences can introduce biases if not properly

accounted for, necessitating sophisticated modeling and validation techniques.

Nevertheless, leveraging the widespread availability of satellite-based LST data, multiple studies have proposed various approaches to estimate T2M. One common strategy is to develop physically based energy-balance models, in which net incoming radiation (including anthropogenic heat flux, where relevant) is partitioned into sensible and latent heat fluxes at the surface [39], [40], [41]. Although this approach offers a strong theoretical foundation, it is hindered by the limited availability of key input variables from satellite data alone (e.g., soil moisture, detailed vegetation parameters, or meteorological forcing). Furthermore, advanced land surface models can be computationally demanding when applied at large spatial or temporal scales, restricting their practicality for widespread operational use.

Another widely adopted approach involves statistical techniques, including advanced machine learning and deep learning methods. For example, Support Vector Machines (SVMs) are often chosen for their ability to handle high-dimensional data and effectively model nonlinear relationships [42], [43], [44]. Meanwhile, deep learning architectures—such as Recurrent Neural Networks (RNNs) [45], [46], [47] or Long Short-Term Memory (LSTM) networks [48], [49]—can incorporate temporal dependencies, thereby capturing the complex, time-varying dynamics that characterize T2M–LST interactions. Moreover, these models can be enhanced by integrating land cover information (e.g., NDVI or NDBI), which helps account for the influence of vegetation and urbanization on near-surface temperature patterns [50], [51], [52], [53], [54], [55], [56]. Overall, these data-driven methods offer strong predictive performance and flexibility, yet they often lack the physical interpretability of process-based models, which can limit their applicability in certain climate and environmental analyses.

This study seeks to advance T2M estimation from satellite-based remote sensing by introducing and rigorously evaluating state-of-the-art deep learning methods specifically designed for temporal analysis, with a focus on the Illinois region in the United States. Although many previous studies have employed ML and DL to estimate T2M from land surface temperature data [48], [49], [57], [58], this work differs in two ways. First, this study simultaneously compares three modern sequence-based architectures—LSTM, TCN, and N-BEATS—offering a broader exploration of how distinct temporal DL approaches perform under the same experimental setup. Second, I explicitly incorporate a suite of land cover metrics (NDVI and NDBI) at multiple spatial scales to capture vegetation and urbanization influences on surface–atmosphere dynamics, thus addressing a critical gap in existing studies that often overlook the impact of heterogeneous land properties. By examining multiple model architectures and systematically assessing their performance, I aim to highlight both the potential and the limitations of these methods for capturing variations in near-surface air temperature.

In particular, this study explores how model performance varies across different station locations, addressing the persistent challenge of sparse in-situ observational networks. I also investigate the effects of incorporating various temporal “look-back” windows to determine how the volume and distribution of recent LST observations shape the models’ ability to capture T2M–LST dynamics. Through this comparative analysis, this study seeks to identify best practices for tailoring deep learning frameworks to diverse climates and land-use contexts, extending beyond the capabilities of prior studies that focused on either a single model architecture or a narrow range of temporal inputs.

Beyond its methodological contributions, this work offers practical insights that facilitate large-scale, rapid T2M estimation using satellite-derived LST. Moreover, by clarifying the conditions under which different lag windows boost T2M predictions, we provide actionable insights for optimizing data pipelines in applications such as urban heat island mapping, agricultural monitoring, and climate change assessments. By showcasing both the effectiveness and scalability of these approaches, this study paves the way for broader, more efficient T2M inference from satellite observations.

II. DATA

A. LOCATION AND STATION MEASUREMENTS

Station measurements for this study come from the National Centers for Environmental Information (NCEI) Land-Based Station product, operated by the National Oceanic and Atmospheric Administration (NOAA). In this study, I select stations within the state of Illinois (IL), a major midlatitude region in the United States. There are a total of 63 stations available, from which I extract near-surface air temperature (T2M) values. The frequency of the data varies, but for all stations, measurements are taken at intervals of less than 10 minutes. I focus on warm season (March to October) T2M data from 2019 to 2023 (5 years). The locations of the NCEI stations used in this study, as well as general 10m-land cover information [59], [60] for IL, are represented in Fig. 1. Further details, including the location and dominant land cover information for each station, can be found in Appendix.

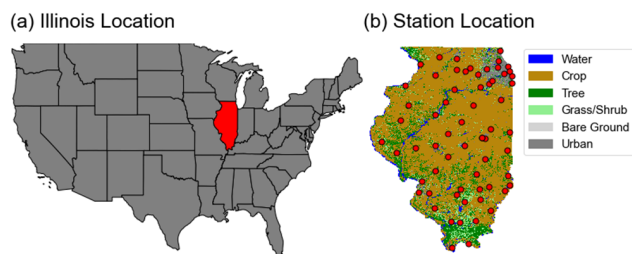


FIGURE 1. (a) Location of the study area (Illinois). (b) Major land cover in Illinois and the location of each NCEI stations used in this study. The red dots denote the location of the NCEI stations.

B. GOES-R LST ESTIMATES

In this study, I utilize land surface temperature (LST) measurements from GOES-16, 17, and 18 satellites, collectively referred to as GOES. These measurements have a spatial resolution of 2 km and provide hourly data over the North American region. For this analysis, I focus on the data for IL during the warm season from 2019 to 2023, consistent with the NCEI station measurements. One inherent limitation of satellite-based products is their inability to measure surface properties under cloudy conditions. Therefore, this study only use cloud-free GOES pixels to ensure the reliability of the LST product. As a result, the LST data in this study should be considered cloud-free. Prior research has shown the error margin of GOES LST to be under 2 K [27], [29].

C. LANDSAT-DERIVED LAND COVER METRICS

I also utilize high-resolution land cover metrics derived from Landsat-8 Collection 2, Tier 2, Level 2 (hereafter, Landsat) data. Landsat is equipped with 15m panchromatic and 30m shortwave-infrared sensors, which enable the calculation of various land cover metrics, namely the Normalized Difference Vegetation Index (NDVI), and the Normalized Difference Built-up Index (NDBI) at 30m resolution. These metrics are selected for their relevance in analyzing vegetation (NDVI) and built-up areas (NDBI). The specific equations for calculating each metric are provided below:

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{1}$$

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \tag{2}$$

In Equations (1) and (2), NIR represents near-infrared band (0.85–0.88 μ m wavelength), Red represents red band (0.64–0.67 μ m), and SWIR represents shortwave-infrared band (1.57–1.65 μ m). I considered other metrics, such as the Enhanced Vegetation Index (EVI) and the Normalized Difference Water Index (NDWI), but since they were strongly correlated with NDVI and NDBI, respectively, they were not included in the analysis.

Given that Landsat has a 16-day revisit interval, each metric (NDVI and NDBI) is initially calculated for every 16-day period at a 30 m spatial resolution, covering the entire Illinois region. To obtain daily values, spline smoothing is then applied, a method widely used in previous studies [61], [62], [63], [64] to generate fine temporal resolutions from multi-day imagery. As a validation, I compared the spline-interpolated daily dataset against daily MODIS NDVI and NDBI at 500m resolution for 100 grid points in Illinois (March–October 2023). The mean absolute error was 0.08 for NDVI and 0.07 for NDBI, suggesting that this interpolation approach reasonably preserves day-to-day variability (figure not shown).

D. DATA INTEGRATION

As a final step, I co-located NCEI station data with the GOES LST by performing a nearest-neighbor search in

geographical space. Specifically, I calculated the great-circle distance between each station (based on its latitude and longitude) and every GOES grid cell, then assigned the station to the grid cell whose center was closest. After this spatial matching, we temporally matched each hourly GOES LST measurement to station T2M observations within a 30-minute window. For example, if GOES LST was measured at 15:02, I averaged all T2M readings from the corresponding station between 14:47 and 15:17.

To understand the interaction between GOES-based LST, land cover metrics, and station-specific metrics, I created land property metrics for each station using Landsat data. Specifically, for each NCEI station, I calculated the average values NDVI and NDBI, within radii of 50m, 100m, and 200m to characterize the immediate surrounding environment. Additionally, I calculated the same metrics for the broader 2 km × 2 km area around each GOES grid point, representing larger-scale land cover influences, and to quantify the effect of land cover related to GOES derived LST. These station-specific land cover metrics have daily resolution, which are smoothed from the 16-day raw NDVI and NDBI product.

This multi-scale approach allows for the examination of how land cover characteristics at different spatial scales interact with surface temperatures observed by GOES LST and those measured at station-specific locations. By comparing station-specific metrics within smaller radii to broader metrics around the GOES grid, the study considers the extent to which local land cover features influence temperature dynamics at both fine and coarse spatial resolutions [24], [51], [65].

Fig. 2a illustrates the relationship between LST and T2M across all stations during the warm season from 2019 to 2023, capturing the relationship for all sampled data points. Fig. 2b presents the correlation metrics between Landsat-derived indices (NDVI and NDBI) at various spatial scales. The results demonstrate that NDVI and NDBI at different scales are positively correlated, while NDVI and NDBI exhibit an inverse relationship.

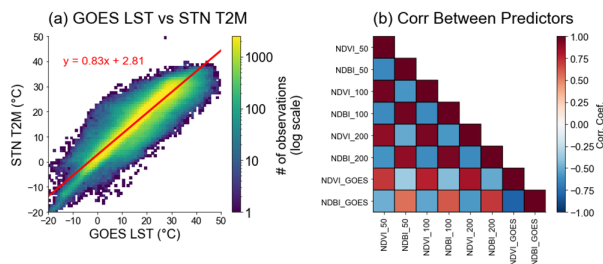


FIGURE 2. a) Relationship between GOES LST and NCEI station T2M during the warm season from 2019 to 2023, aggregated over all NCEI stations. b) Correlation matrix of Landsat metrics (NDVI and NDBI) at different spatial scales.

III. METHOD OF ANALYSIS

A. GENERAL MODEL SETUP

The objective of the deep learning methods employed in this study is to predict T2M at a given hour, t . To achieve

this, the model uses a set of predictor variables, including GOES-based LST at a 2 km resolution surrounding the weather stations, as well as NDVI and NDBI indices calculated within radii of 50 m, 100 m, and 200 m around the weather stations. Additionally, the average NDVI and NDBI values over the 2 km GOES grid are included. To incorporate diurnal and seasonal patterns, the model also utilizes the month and hour of the day corresponding to hour t .

To leverage the temporal sequence capabilities of the models, historical data is included in the input. Specifically, for a given look-back window (lag) l , the model uses LST values at $t, t-1, \dots, t-l+1$, enabling it to learn from patterns and dependencies in prior observations. Different lags are tested in this study, ranging from 2 to 18 hours with 2-hour intervals (e.g., 2, 4, 6, ..., 18). This approach ensures that the model captures both the immediate and extended temporal relationships between the predictor variable (LST) and the target variable, T2M.

Since land cover information is included in the model, I do not account for individual stations—meaning the models are not developed separately for each station but are instead designed as a universal framework capable of capturing the T2M–LST relationship across varying land cover types and temporal dynamics.

Using a 5-year dataset (2019–2023) covering the months from March to October, I designate 2021 as the test set and use 2019, 2020, 2022, and 2023 for training. This study specifically chose 2021 because it falls in the middle of the time span, creating an approximate 80%/20% split between training and testing. This arrangement ensures that the test set is temporally independent, providing a robust evaluation of the model’s generalization capabilities. In total, the training dataset contains 1,471,239 samples, while the test set comprises 365,778 samples, offering sufficient data to assess how well the model captures spatial and temporal variability in T2M under diverse conditions.

B. LONG SHORT-TERM MEMORY (LSTM)

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) designed to address the vanishing gradient problem, which often limits the effectiveness of traditional RNNs in learning temporal dependencies within sequential data [66], [67], [68]. LSTM networks incorporate memory cells and gating mechanisms—input, forget, and output gates—that allow the network to retain, discard, or update information as needed. These mechanisms make LSTMs particularly effective in capturing both short-term and long-term patterns, enabling them to excel in various time series forecasting tasks.

In this study, I design a two-layer LSTM network that receives (lag × 11) input sequences (GOES_LST, NDVI_50, NDBI_50, NDVI_100, NDBI_100, NDVI_200, NDBI_200, NDVI_GOES, and NDBI_GOES, plus two columns for Month and Hour). The first LSTM layer takes these 11 features per time step and has 32 hidden units, using the

sigmoid gates and tanh state updates. I apply a dropout rate of 0.2 to the output of this first LSTM. Next, the second LSTM layer, also with dropout 0.2, reduces the hidden dimension from 32 down to 16 hidden units. After processing the entire sequence, I extract only the final time step's output (dimension 16) and feed it into a fully connected layer (linear) of size 1, yielding a single scalar prediction (T2M). No explicit activation function is used after the final linear layer, allowing the output to remain in the real-valued range for regression.

C. TEMPORAL CONVOLUTIONAL NETWORKS (TCN)

Temporal Convolutional Networks (TCNs) are a convolutional architecture designed for sequential data, offering an alternative to recurrent models such as LSTMs [69], [70], [71]. TCNs use causal convolutions, ensuring that predictions for a given time step are influenced only by past inputs, and employ dilated convolutions to efficiently capture long-range dependencies. Residual connections within TCN layers improve gradient flow and model stability during training. These features enable TCNs to model temporal patterns effectively across varying time scales.

In this study, I design a two-block TCN to estimate T2M with a same input setup as LSTM. The TCN itself consists of two sequential blocks, each comprising two causal one-dimensional convolution layers with padding and a chomp operation to ensure that the output at any time step depends only on past inputs. Within each block, ReLU activations introduce nonlinearity, dropout (0.2) helps prevent overfitting, and residual connections support deeper stacks without vanishing gradients by merging the original input with the convolutions' output. The first TCN block receives the raw input features and increases the channel dimension to 32, while the second block processes these 32 channels at a higher dilation rate, thus expanding the network's receptive field to capture broader temporal dependencies. Finally, a linear layer then maps the output into a single scalar value corresponding to the predicted T2M. This design, featuring stacked causal convolutions, progressive dilations, and residual pathways, allows the model to efficiently learn multiscale temporal dependencies in the time series data.

D. NEURAL BASIS EXPANSION ANALYSIS FOR TIME SERIES (N-BEATS)

Neural Basis Expansion Analysis for Time Series (N-BEATS) is a state-of-the-art deep learning architecture designed specifically for time series forecasting [72], [73], [74]. N-BEATS uses a fully connected neural network framework without reliance on domain-specific feature engineering, making it highly flexible and adaptable to diverse forecasting problems. The architecture is based on a stack of blocks, each of which contains a set of fully connected layers designed to decompose the input time series into interpretable components, such as trend and seasonality. These blocks operate in a forward and backward direction,

allowing the model to capture both historical context and future patterns in the data.

In this study, I design a two-block N-BEATS model to forecast T2M with identical input variables as LSTM and TCN. Each time window, input data, is flattened into a single vector before entering the first block. This block comprises four fully connected layers (each followed by a ReLU activation) that produce two outputs: a backcast, aiming to reconstruct the input window, and a forecast, representing a single-step prediction. The residual is then computed by subtracting the backcast from the input vector and passed on to the second block, which has the same four-layer structure. The second block similarly outputs a backcast and a forecast; the final model prediction is the sum of the two forecasts. In this way, the two-block design captures a multi-stage residual learning process in which each block refines the forecast based on the partial reconstruction of prior inputs. A RMSE loss drives the training, and dropout (0.2) is used to mitigate overfitting. Once trained, the model produces a single T2M estimate per sample by summing the forecasts from both blocks.

E. EVALUATION METRIC

To compare the performance of the different methods, this study uses several metrics—including root mean squared error (RMSE), mean absolute error (MAE), R-squared (R^2), and slope—to assess overall accuracy. Additionally, to evaluate how effectively the models capture extreme T2M events, this study employs standard classification metrics (precision, recall, and F1 score). Specifically, I first determine the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) for each model. This study then compute precision, recall, and F1 based on the following formulas:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

These classification metrics enable a direct comparison of each model's ability to detect extreme cases.

IV. RESULTS

A. OVERALL PERFORMANCE

To evaluate the overall performance of each model (LSTM, TCN, and N-BEATS) across different look-back periods (lags, ranging from 2 to 18 hours at 2-hour intervals), I analyze metrics including RMSE, MAE, R^2 , and slope ($\Delta\text{Model}/\Delta\text{Station}$), as shown in Fig. 3. These metrics are calculated using the 2021 data, which were not included in the training of the models and were specifically reserved as a test set. Overall, all three models perform well. Notably, when compared to non-temporal models such as Support Vector Machine (SVM), Random Forest (RF) and RNN, which are

applied without temporal context (i.e., comparing LST and T2M only at time t while considering land cover), incorporating temporal information leads to a significant reduction in RMSE. Specifically, the RMSE for non-temporal SVM, RF and RNN models was 2.84, 2.56 and 2.37°C, respectively (Figure not shown). In contrast, temporal models such as TCN, N-BEATS, and LSTM demonstrate substantially lower RMSE (below 1.8°C, Fig. 3a), highlighting the importance of leveraging temporal dependencies.

Among the temporal models, TCN and N-BEATS consistently outperform LSTM across most metrics, with TCN and N-BEATS showing comparable performance depending on the metric being evaluated. This is probably because TCN can effectively capture multi-scale temporal dependencies via dilated, causal convolutions and residual connections, which often leads to more stable gradient propagation than standard LSTM. Meanwhile, N-BEATS employs a basis expansion strategy that better decomposes time series into trend-like components, thereby enhancing its representation of T2M patterns.

Model performance generally improves as the lag increases, reflecting the benefit of incorporating more temporal information. However, an exception is observed for N-BEATS beyond a 10-hours lag, where the performance starts to plateau or worsen slightly. This suggests that for N-BEATS, incorporating additional temporal inputs beyond this threshold may not yield further gains and could even introduce noise or redundancy into the model.

Looking at the slope, one might notice a disturbance of trend with changing lags. A likely explanation is that the slope metric is particularly sensitive to small changes in which data points are included at each lag, especially when missing LST observations alter the valid sample. Even subtle shifts in data composition can make the slope fluctuate, even if RMSE and MAE remain relatively stable.

The trade-off of using longer lags needs careful consideration. While longer lags tend to improve model accuracy, as evidenced by the decreasing RMSE and MAE and increasing R^2 , they also require more extensive historical data. For satellite-derived LST, such as from the GOES-R series, obtaining continuous data for extended time periods (e.g., 18 hours) can be challenging due to occasional disruptions caused by cloud cover. Missing data in these scenarios could adversely affect model performance, particularly for methods like TCN and N-BEATS that rely on uninterrupted temporal sequences. In that context, the N-BEATS model with 10 hours of lag seems to be the most reasonable choice in overall performance and consideration, by balancing between the model performance and the lookback window. But that could change based on data availability and one's priority.

B. SEASONAL AND DIURNAL CHARACTERISTICS

As a next step, this study examines how the model performance changes with seasonality and the diurnal cycle. In terms of seasonal RMSE, all three models display a

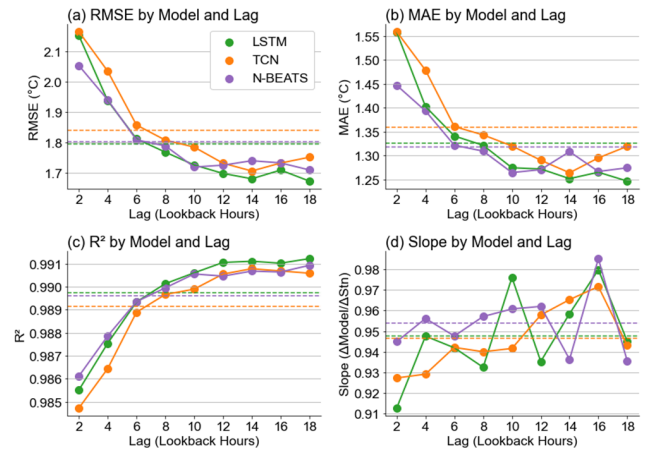


FIGURE 3. (a) RMSE calculated for each lag (x-axis) for different models used in this study (LSTM, TCM, and N-BEATS). (b-d) same as (a), but for MAE, R^2 , and slope values.

similar evolution: a peak in April and May, followed by a notable minimum in July and August (Fig. 4a-c). This indicates that the models generally become more accurate during mid-summer, most likely due to a stable boundary layer and a more consistent coupling between LST and T2M at higher temperatures [75], [76], [77]. The presence of higher NDVI in mid-summer, reflecting the growth of denser vegetation, can further increase evapotranspiration and moderate local temperature fluctuations, thus strengthening the LST–T2M relationship. Another significant result is that the best-performing lag is shorter in spring but becomes longer in summer. In spring, abrupt temperature changes and rapid vegetation transitions often make recent data more critical for capturing T2M variability, whereas in summer, vegetation has reached a more mature state, land surface conditions shift more slowly, and built surfaces can accumulate heat over longer times [78], [79], [80]. As a result, the model benefits from drawing on a longer memory window to capture these gradual changes in both vegetated and built-up areas (as captured by NDVI and NDBI).

Examining the diurnal cycle (Fig. 4d-f), the RMSE is highest at noon, suggesting that the midday surge of heat fluxes, stronger convection, and more vigorous boundary-layer mixing reduce the accuracy of T2M predictions. During these midday conditions, NDVI-related transpiration and the urban heat storage reflected by NDBI can either dampen or magnify the thermal contrast between surface and air, making the LST–T2M linkage more volatile. The optimal lag also becomes smaller during noon to afternoon, implying that short-term data are especially critical when temperature ramps up rapidly under intense solar heating. By contrast, in the morning and overnight hours, temperature evolution proceeds more steadily, allowing longer lags to remain informative.

There appears to be a discrepancy regarding model accuracy at higher temperatures: seasonally, performance tends to improve as temperatures rise, whereas diurnally,

error increases during peak heat. On a seasonal scale, mature vegetation and relatively stable boundary-layer conditions in midsummer reinforce the coupling between LST and T2M, which explains why higher temperatures can yield better performance overall. In contrast, midday hours often involve rapid heat fluxes, intense solar radiation, and enhanced convection, leading to pronounced short-term variability in T2M and thus higher RMSE at the daily temperature peak. Essentially, strong coupling at broader seasonal scales coexists with short-term diurnal fluctuations that reduce accuracy during midday extremes.

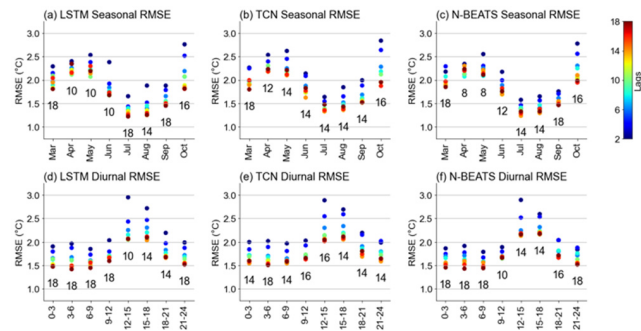


FIGURE 4. (a) RMSE calculated for each month (x-axis) for different lags (colored dots) for LSTM model. The color bar for the model lag can be found in the right side of the plot. Blue dots represent the shorter lags and the red represent the longer lags used in the model. The optimal lag (lag with minimum RMSE) is shown on the bottom of each column of scatterplot (b, c) same as (a), but for TCN and N-BEATS model. (d-f) same as (a-c), but for the diurnal cycle (x-axis), sharing the same color bar.

C. SPATIAL CHARACTERISTICS

The model performance can vary due to not only the temporal factors, but the spatial factors. In that context, this study also examines how spatial characteristics—particularly vegetation and built environments—affect model accuracy. I focus on the N-BEATS model with a 10-hours lag window because it achieves a good balance between predictive performance and lookback length (see Fig. 3), and all models share broadly similar spatial patterns of error. I plot the 10-hours lag N-BEATS RMSE against station-level NDVI and NDBI at the 2 km scale in Fig. 5. Here, the RMSE is computed for each NCEI station by comparing predicted and observed T2M, and the station’s NDVI and NDBI values are taken as warm season averages from 2021.

As shown, model performance deteriorates as NDVI increases, and there is also a weaker negative trend with NDBI. The stronger slope in NDVI implies that denser vegetation tends to amplify the discrepancy between surface (LST) and near-surface (T2M) thermal behavior, leading the model to incur higher errors. One explanation is that vegetation can enhance evaporative cooling and modulate boundary-layer exchanges in a way that does not map consistently onto LST changes. A canopy’s shading and transpiration can decouple surface warming from the ambient air temperature, making it more difficult for the model to track T2M solely from LST-based features.

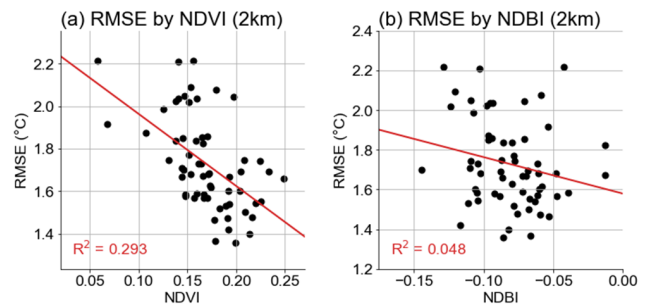


FIGURE 5. (a) Station-specific RMSE of the model performance versus the NDVI (black dots), as well as its regression fit (red line). (b) Same as (a), but for RMSE versus the NDBI.

D. EXTREME CASES

Another important consideration for T2M estimation is how well the models can capture extreme temperatures. In this analysis, I define an extreme case as any T2M value exceeding the 95th percentile of all T2M measurements recorded throughout the study period for all NCEI stations, which corresponds to 29.4°C.

This study then compare Precision, Recall, and F1 score with 14-hour lagged models, because all three models perform best during midday to early afternoon (Fig. 4d–f), when T2M reaches its daily peak. Furthermore, changing the lag window does not significantly alter the relative trends in extreme detection.

As shown in Fig. 6, the highest precision is achieved by the LSTM, while recall is highest with the N-BEATS, and F1 is highest in the TCN. This indicates that LSTM is most conservative (fewer false alarms), N-BEATS catches the largest proportion of actual extremes, and TCN strikes a balance between the two, leading to the strongest F1 performance.

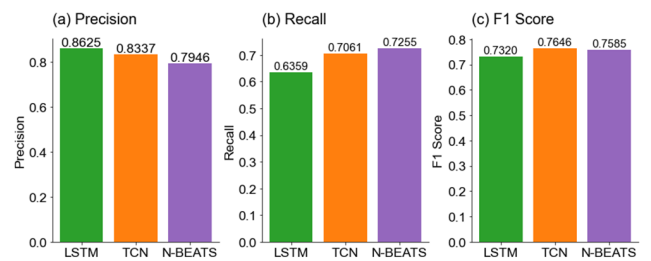


FIGURE 6. (a) Precision of LSTM, TCN, and N-BEATS in extreme case detection. (b, c) Same as (a), but for Recall and F1 Score.

V. SUMMARY AND DISCUSSION

A. SUMMARY

In this study, I utilized and compared three temporal deep learning methods—LSTM, TCN, and N-BEATS—for estimating T2M from satellite-derived LST and land cover metrics such as NDVI and NDBI. By incorporating temporal context through various look-back windows (lags), these models substantially outperformed non-temporal approaches like support vector machine, random forests and simpler

neural networks. This underscores the critical role that historical LST play in capturing the evolving relationship between surface and near-surface temperatures.

Evaluating the three methods revealed several key insights. First, the models generally benefited from longer lags, as more historical data improved accuracy. However, N-BEATS performance plateaued at around 10 hours of lag, suggesting diminishing returns beyond this threshold. Since these models rely on cloud-free satellite data, pushing to 18-hour lags or more can encounter practical issues where missing LST observations break the temporal continuity. Consequently, while a longer memory window can theoretically capture extended dynamics, N-BEATS with a 10-hour lag offers a balanced solution that is more robust against gaps in satellite measurements.

Seasonally and diurnally, all models showed higher RMSE in spring (April–May) and lower RMSE in mid-summer (July–August). The abrupt vegetation transitions in spring likely reduce predictability, whereas mature vegetation and relatively stable conditions in summer support stronger T2M–LST coupling. A similar pattern emerged in the diurnal cycle, where the highest midday (12–18 hours) temperatures introduced more error in T2M, suggesting that heat fluxes, convection, and boundary-layer mixing reduce the accuracy of T2M predictions. Interestingly, a shorter optimal lag tends to align with higher RMSE, implying that these rapid changes are inherently harder to predict.

Spatially, all models exhibited a trend of increasing RMSE with higher NDVI values at station locations, highlighting how denser vegetation modifies surface–air temperature relationships (for instance, through enhanced evapotranspiration). In detecting extreme T2M events, LSTM achieved the highest precision (fewest false alarms), N-BEATS achieved the highest recall (capturing the truest extremes), and TCN provided the best balance overall (highest F1). This variability in extreme detection suggests that model choice could be tailored to particular operational needs: for example, N-BEATS with a 10-hour lag for general T2M forecasting, but TCN for more balanced extreme-event monitoring.

B. DISCUSSIONS

These findings carry significant implications for near-surface air temperature (T2M) monitoring, particularly in applications that require frequent and fine-scale temperature forecasts. By leveraging satellite-derived LST alongside multi-scale NDVI and NDBI, the deep learning models evaluated here—LSTM, TCN, and N-BEATS—demonstrate a marked improvement in capturing diurnal and seasonal T2M variability. This level of accuracy supports more targeted resource allocation in sectors such as public health, where early heat-wave detection can mitigate risks [1], [2], [3], and agriculture, where timely temperature information can inform irrigation and crop management decisions [7], [9].

Compared with similar studies that commonly employed either traditional machine learning algorithms or

single-architecture deep learning approaches under shorter or fixed lag times [46], [56], this work offers a side-by-side examination of three advanced temporal methods applied over varying lag windows. Such a comparative framework highlights both the shared and distinct mechanisms by which these models learn dependencies between historical LST observations and current T2M, providing novel insight into how different architectures cope with missing data and complex land cover influences. In particular, the multi-scale design for NDVI and NDBI used here moves beyond the narrower focus on a single vegetation or urban index, shedding light on localized variations in vegetation structure and built surfaces that more strongly modulate T2M than previously recognized [42].

An immediate advantage of these temporal deep learning architectures is their flexibility in incorporating additional features—such as lagged meteorological variables or other satellite products—without major restructuring of the model. This opens the door to operational applications, where new data streams can be seamlessly integrated to refine forecasts. Nonetheless, disadvantages do surface in practice. Models like TCN and N-BEATS often require extensive training data, and maintaining continuity in LST observations can be challenging when cloud cover disrupts measurements over extended windows. The computational cost may also rise rapidly as the input sequence lengthens or the spatial coverage expands, potentially limiting real-time deployment in resource-constrained settings.

Certain limitations further constrain how well these approaches capture local land–atmosphere interactions. While NDVI and NDBI provide valuable information on vegetation and built-up areas, they do not encompass other critical factors such as soil moisture, detailed canopy structures, or fine-scale topography. If any of these missing variables heavily influence T2M in specific regions, model performance may suffer. Moreover, relying on sparse station-based observations as ground-truth can introduce biases where local microclimates deviate from regional conditions. Lastly, incorporating additional variables and more extensive ground observations could yield a more comprehensive framework for T2M estimation.

Looking ahead, future directions include integrating physical model outputs or reanalysis datasets to fill in gaps during cloudy conditions, employing adaptive lag strategies that shorten or lengthen historical windows based on data continuity or known meteorological regimes, and expanding the land-surface descriptors to encompass soil moisture, fine-scale urban morphology, or high-resolution canopy structure. These enhancements could help the models better disentangle rapidly changing processes—such as evapotranspiration in newly leafed vegetation or the heat storage dynamics in densely built areas—and thus more reliably detect extreme events. As satellite sensing technology moves toward higher spatial and temporal resolution, and as ground-based sensor networks expand, the combined use of multi-source data with robust temporal learning frameworks stands to further

advance climate monitoring, urban heat adaptation efforts, and sustainable resource management at multiple scales.

APPENDIX STATION INFORMATIONS

TABLE 1. Location and land cover information of the NCEI stations.

Station ID	Lat	Lon	NDVI (100m)	NDBI (100m)	Land Cover (100m)
1	41.43	-88.42	0.19	-0.08	Crop
2	41.54	-87.53	0.19	-0.05	Urban
3	41.07	-87.85	0.21	-0.04	Crop
4	41.35	-89.15	0.23	-0.08	Crop
5	40.52	-90.65	0.20	-0.12	Crop
6	40.92	-88.62	0.21	-0.10	Crop
7	39.70	-87.67	0.19	-0.08	Crop
8	41.89	-89.08	0.20	-0.09	Crop
9	40.29	-88.14	0.15	-0.01	Crop
10	42.05	-90.11	0.17	-0.04	Crop
11	39.16	-89.67	0.23	-0.11	Crop
12	37.19	-88.75	0.25	-0.06	Grass
13	38.65	-88.97	0.23	-0.08	Crop
14	38.32	-88.86	0.22	-0.05	Grass
15	37.78	-89.25	0.22	-0.02	Crop
16	38.55	-89.85	0.26	-0.10	Crop
17	37.75	-89.00	0.29	-0.12	Grass
18	39.84	-89.68	0.22	-0.07	Crop
19	38.88	-90.05	0.24	-0.08	Crop
20	40.48	-88.95	0.07	-0.05	Urban
21	39.94	-91.19	0.19	-0.05	Crop
22	37.06	-89.22	0.23	-0.05	Crop
23	39.02	-87.65	0.27	-0.11	Crop
24	41.96	-87.93	0.16	-0.09	Urban
25	38.57	-90.16	0.21	-0.07	Crop
26	40.04	-88.28	0.28	-0.11	Crop
27	41.91	-88.25	0.19	-0.09	Urban
28	39.83	-88.87	0.26	-0.11	Crop
29	39.48	-88.28	0.19	-0.06	Crop
30	40.67	-89.68	0.23	-0.09	Crop
31	41.74	-89.68	0.22	-0.10	Crop
32	41.79	-87.75	0.17	-0.06	Urban
33	38.76	-87.61	0.13	0.05	Crop
34	38.61	-87.73	0.22	-0.07	Crop
35	41.50	-88.17	0.17	-0.10	Crop
36	42.42	-87.87	0.12	-0.06	Urban
37	41.60	-88.08	0.20	-0.11	Crop
38	42.19	-89.09	0.15	-0.07	Crop
39	41.47	-90.52	0.24	-0.17	Crop
40	37.81	-88.55	0.27	-0.07	Crop
41	38.15	-89.70	0.27	-0.10	Crop
42	41.77	-88.48	0.18	-0.08	Crop
43	38.38	-88.41	0.24	-0.06	Crop
44	38.51	-89.09	0.24	-0.04	Crop
45	38.09	-88.12	0.24	-0.09	Crop
46	38.66	-88.45	0.24	-0.06	Crop
47	38.72	-88.18	0.26	-0.11	Crop
48	39.53	-89.33	0.23	-0.09	Crop
49	39.64	-90.78	0.25	-0.14	Crop
50	42.12	-87.90	0.18	-0.09	Urban
51	39.78	-90.24	0.23	-0.14	Crop
52	40.16	-89.33	0.25	-0.10	Crop
53	41.73	-87.54	0.16	-0.05	Urban
54	41.86	-87.61	0.16	-0.05	Urban
55	40.06	-88.37	0.19	-0.02	Crop
56	41.93	-88.71	0.18	-0.07	Crop
57	40.05	-88.37	0.27	-0.11	Crop
58	41.84	-88.85	0.24	-0.11	Crop
59	39.07	-88.53	0.21	-0.06	Crop
60	41.02	-89.39	0.18	-0.04	Crop
61	40.20	-87.60	0.17	-0.08	Crop
62	42.25	-89.58	0.22	-0.11	Crop
63	40.93	-90.43	0.24	-0.18	Crop

REFERENCES

[1] J. Lee and A. E. Dessler, "Future temperature-related deaths in the U.S.: The impact of climate change, demographics, and adaptation," *GeoHealth*, vol. 7, no. 8, 2023, Art. no. e2023GH000799.

[2] K. L. Ebi, A. Capon, P. Berry, C. Broderick, R. de Dear, G. Havenith, Y. Honda, R. S. Kovats, W. Ma, A. Malik, N. B. Morris, L. Nybo, S. I. Seneviratne, J. Vanos, and O. Jay, "Hot weather and heat extremes: Health risks," *Lancet*, vol. 398, no. 10301, pp. 698–708, Aug. 2021.

[3] J. E. Bell, C. L. Brown, K. Conlon, S. Herring, K. E. Kunkel, J. Lawrimore, G. Luber, C. Schreck, A. Smith, and C. Uejio, "Changes in extreme events and the potential impacts on human health," *J. Air Waste Manage. Assoc.*, vol. 68, no. 4, pp. 265–287, Apr. 2018.

[4] J. Lee and A. E. Dessler, "The impact of neglecting climate change and variability on ERCOT's forecasts of electricity demand in Texas," *Weather, Climate, Soc.*, vol. 14, no. 2, pp. 499–505, 2022.

[5] T. E. Morakinyo, C. Ren, Y. Shi, K. K.-L. Lau, H.-W. Tong, C.-W. Choy, and E. Ng, "Estimates of the impact of extreme heat events on cooling energy demand in Hong Kong," *Renew. Energy*, vol. 142, pp. 73–84, Nov. 2019.

[6] M. A. D. Larsen, S. Petrović, A. M. Radoszynski, R. McKenna, and O. Balyk, "Climate change impacts on trends and extremes in future heating and cooling demands over Europe," *Energy Buildings*, vol. 226, Nov. 2020, Art. no. 110397.

[7] F. M. Aragón, F. Oteiza, and J. P. Rud, "Climate change and agriculture: Subsistence farmers' response to extreme heat," *Amer. Econ. J., Econ. Policy*, vol. 13, no. 1, pp. 1–35, Feb. 2021.

[8] L. E. Parker, A. J. McElrone, S. M. Ostoja, and E. J. Forrester, "Extreme heat effects on perennial crops and strategies for sustaining future production," *Plant Sci.*, vol. 295, Jun. 2020, Art. no. 110397.

[9] I. Haqiqi, D. S. Grogan, T. W. Hertel, and W. Schlenker, "Quantifying the impacts of compound extremes on agriculture," *Hydrol. Earth Syst. Sci.*, vol. 25, no. 2, pp. 551–564, Feb. 2021.

[10] S. S. Clark, M. V. Chester, T. P. Seager, and D. A. Eisenberg, "The vulnerability of interdependent urban infrastructure systems to climate change: Could Phoenix experience a Katrina of extreme heat?" *Sustain. Resilient Infrastruct.*, vol. 4, no. 1, pp. 21–35, Jan. 2019.

[11] D. Burillo, M. V. Chester, S. Pincetl, and E. Fournier, "Electricity infrastructure vulnerabilities due to long-term growth and extreme heat from climate change in Los Angeles county," *Energy Policy*, vol. 128, pp. 943–953, May 2019.

[12] B. Stone, E. Mullen, M. Rajput, C. J. Gronlund, A. M. Broadbent, E. S. Krayenhoff, G. Augenbroe, M. S. O'Neill, and M. Georgescu, "Compound climate and infrastructure events: How electrical grid failure alters heat wave risk," *Environ. Sci. Technol.*, vol. 55, no. 10, pp. 6957–6964, May 2021.

[13] T. Yokobori and S. Ohta, "Effect of land cover on air temperatures involved in the development of an intra-urban heat island," *Climate Res.*, vol. 39, pp. 61–73, Jun. 2009.

[14] V. Cermak, L. Bodri, M. Kresl, P. Dedecek, and J. Safanda, "Eleven years of ground-air temperature tracking over different land cover types," *Int. J. Climatol.*, vol. 37, no. 2, pp. 1084–1099, Feb. 2017.

[15] L. Yu and G. Leng, "Global effects of different types of land use and land cover changes on near-surface air temperature," *Agricult. Forest Meteorol.*, vol. 327, Dec. 2022, Art. no. 109232.

[16] D. X. Tran, F. Pla, P. Latorre-Carmona, S. W. Myint, M. Caetano, and H. V. Kieu, "Characterizing the relationship between land use land cover change and Land Surface Temperature," *ISPRS J. Photogramm. Remote Sens.*, vol. 124, pp. 119–132, Feb. 2017.

[17] J. Muñoz-Sabater, E. Dutra, A. Agustí-Panareda, C. Albergel, G. Arduini, G. Balsamo, S. Boussetta, M. Choulga, S. Harrigan, H. Hersbach, B. Martens, D. G. Miralles, M. Piles, N. J. Rodríguez-Fernández, E. Zsoter, C. Buontempo, and J.-N. Thépaut, "ERA5-land: A state-of-the-art global reanalysis dataset for land applications," *Earth Syst. Sci. Data*, vol. 13, no. 9, pp. 4349–4383, Sep. 2021.

[18] S. Saha et al., "The NCEP climate forecast system reanalysis," *Bull. Amer. Meteorol. Soc.*, vol. 91, no. 8, pp. 1015–1058, Apr. 2010.

[19] E. Kalnay et al., "The NCEP/NCAR 40-year reanalysis project," in *Renewable Energy*. Evanston, IL, USA: Routledge, 2018, pp. 146–194.

[20] J. Lee and A. E. Dessler, "Improved surface urban heat impact assessment using GOES satellite data: A comparative study with ERA-5," *Geophys. Res. Lett.*, vol. 51, no. 1, Jan. 2024, Art. no. e2023GL107364.

- [21] S.-Q. Zhang, G.-Y. Ren, Y.-Y. Ren, Y.-X. Zhang, and X.-Y. Xue, "Comprehensive evaluation of surface air temperature reanalysis over China against urbanization-bias-adjusted observations," *Adv. Climate Change Res.*, vol. 12, no. 6, pp. 783–794, Dec. 2021.
- [22] B. McNicholl, Y. H. Lee, A. G. Campbell, and S. Dev, "Evaluating the reliability of air temperature from ERA5 reanalysis data," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [23] T. Zhao, W. Guo, and C. Fu, "Calibrating and evaluating reanalysis surface temperature error by topographic correction," *J. Climate*, vol. 21, no. 6, pp. 1440–1446, Mar. 2008.
- [24] J. Lee and M. Berkelhammer, "Observational constraints on the spatial effect of greenness and canopy cover on urban heat in a major midlatitude city," *Geophys. Res. Lett.*, vol. 51, no. 21, Nov. 2024, Art. no. e2024GL110847.
- [25] M. Alonzo, M. E. Baker, Y. Gao, and V. Shandas, "Spatial configuration and time of day impact the magnitude of urban tree canopy cooling," *Environ. Res. Lett.*, vol. 16, no. 8, Aug. 2021, Art. no. 084028.
- [26] C. D. Ziter, E. J. Pedersen, C. J. Kucharik, and M. G. Turner, "Scale-dependent interactions between tree canopy cover and impervious surfaces reduce daytime urban heat during summer," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 15, pp. 7575–7580, Apr. 2019.
- [27] Y. Yu and P. Yu, "Land Surface Temperature product from the GOES-R series," in *The GOES-R Series*. Amsterdam, The Netherlands: Elsevier, 2020, pp. 133–144.
- [28] Z. Wan, "New refinements and validation of the MODIS land-surface temperature/emissivity products," *Remote Sens. Environ.*, vol. 112, no. 1, pp. 59–74, Jan. 2008.
- [29] G. C. Hulley, F. M. Göttsche, G. Rivera, S. J. Hook, R. J. Freepartner, M. A. Martin, K. Cawse-Nicholson, and W. R. Johnson, "Validation and quality assessment of the ECOSTRESS level-2 Land Surface Temperature and emissivity product," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5000523.
- [30] K. Zakšek and M. Schroedter-Homscheidt, "Parameterization of air temperature in high temporal and spatial resolution from a combination of the SEVIRI and MODIS instruments," *ISPRS J. Photogramm. Remote Sens.*, vol. 64, no. 4, pp. 414–421, Jul. 2009.
- [31] C. Vancutsem, P. Ceccato, T. Dinku, and S. J. Connor, "Evaluation of MODIS Land Surface Temperature data to estimate air temperature in different ecosystems over Africa," *Remote Sens. Environ.*, vol. 114, no. 2, pp. 449–465, Feb. 2010.
- [32] M. Z. Jacobson, *Fundamentals of Atmospheric Modeling*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
- [33] D. J. Mildrexler, M. Zhao, and S. W. Running, "A global comparison between station air temperatures and MODIS land surface temperatures reveals the cooling role of forests," *J. Geophys. Res., Biogeosci.*, vol. 116, no. 3, pp. 1–12, 2011.
- [34] Y. Qiu, J. Feng, J. Wang, Y. Xue, and Z. Xu, "Memory of land surface and subsurface temperature (LST/SUBT) initial anomalies over Tibetan Plateau in different land models," *Climate Dyn.*, vol. 62, no. 4, pp. 2703–2718, Apr. 2024.
- [35] P. Ji, X. Yuan, X. Liang, Y. Jiao, Y. Zhou, and Z. Liu, "High-resolution land surface modeling of the effect of long-term urbanization on hydrothermal changes over Beijing metropolitan area," *J. Geophys. Res., Atmos.*, vol. 126, no. 18, Sep. 2021, Art. no. e2021JD034787.
- [36] K. Jin, F. Wang, Q. Zong, P. Qin, and C. Liu, "Impact of variations in vegetation on surface air temperature change over the Chinese loess Plateau," *Sci. Total Environ.*, vol. 716, May 2020, Art. no. 136967.
- [37] Y. Cao, W. Guo, J. Ge, Y. Liu, C. Chen, X. Luo, and L. Yang, "Greening vegetation cools mean and extreme near-surface air temperature in China," *Environ. Res. Lett.*, vol. 19, no. 1, Jan. 2024, Art. no. 014040.
- [38] J. Ge, Y. Wang, H. Akbari, D. Zhou, Z. Gu, and X. Meng, "Cooling energy saving by vegetation planting in high-density districts: Evaluation using the coupled simulation," *Building Environ.*, vol. 232, Mar. 2023, Art. no. 110054.
- [39] Y.-J. Sun, J.-F. Wang, R.-H. Zhang, R. R. Gillies, Y. Xue, and Y.-C. Bo, "Air temperature retrieval from remote sensing data based on thermodynamics," *Theor. Appl. Climatol.*, vol. 80, no. 1, pp. 37–48, Feb. 2005.
- [40] S. Pareeth and P. Karimi, "Evapotranspiration estimation using surface energy balance model and medium resolution satellite data: An operational approach for continuous monitoring," *Sci. Rep.*, vol. 13, no. 1, p. 12026, Jul. 2023.
- [41] M. Mauder, T. Foken, and J. Cuxart, "Surface-energy-balance closure over land: A review," *Boundary-Layer Meteorol.*, vol. 177, nos. 2–3, pp. 395–426, Dec. 2020.
- [42] J. Cifuentes, G. Marulanda, A. Bello, and J. Reneses, "Air temperature forecasting using machine learning techniques: A review," *Energies*, vol. 13, no. 16, p. 4215, Aug. 2020.
- [43] S. M. Karimi, O. Kisi, M. Porrajabali, F. Rouhani-Nia, and J. Shiri, "Evaluation of the support vector machine, random forest and geostatistical methodologies for predicting long-term air temperature," *ISH J. Hydraulic Eng.*, vol. 26, no. 4, pp. 376–386, Oct. 2020.
- [44] G. Moser, M. De Martino, and S. B. Serpico, "Estimation of air surface temperature from remote sensing images and pixelwise modeling of the estimation uncertainty through support vector machines," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 1, pp. 332–349, Jan. 2015.
- [45] E. A. Nketiah, L. Chenlong, J. Yingchuan, and S. A. Aram, "Recurrent neural network modeling of multivariate time series and its application in temperature forecasting," *PLoS ONE*, vol. 18, no. 5, May 2023, Art. no. e0285713.
- [46] M. Bilgili, A. Ozbek, A. Yildirim, and E. Simsek, "Artificial neural network approach for monthly air temperature estimations and maps," *J. Atmos. Solar-Terrestrial Phys.*, vol. 242, Jan. 2023, Art. no. 106000.
- [47] E. Haque, S. Tabassum, and E. Hossain, "A comparative analysis of deep neural networks for hourly temperature forecasting," *IEEE Access*, vol. 9, pp. 160646–160660, 2021.
- [48] A. Sekertekin, M. Bilgili, N. Arslan, A. Yildirim, K. Celebi, and A. Ozbek, "Short-term air temperature prediction by adaptive neuro-fuzzy inference system (ANFIS) and long short-term memory (LSTM) network," *Meteorol. Atmos. Phys.*, vol. 133, no. 3, pp. 943–959, Jun. 2021.
- [49] A. Ozbek, A. Sekertekin, M. Bilgili, and N. Arslan, "Prediction of 10-min, hourly, and daily atmospheric air temperature: Comparison of LSTM, ANFIS-FCM, and ARMA," *Arabian J. Geosci.*, vol. 14, no. 7, pp. 1–16, Apr. 2021.
- [50] W. Runke, Y. Xiaoni, S. Yaya, W. Chengyong, and L. Baokang, "Study on air temperature estimation and its influencing factors in a complex mountainous area," *PLoS ONE*, vol. 17, no. 8, Aug. 2022, Art. no. e0272946.
- [51] F. Marzban, S. Sodoudi, and R. Preusker, "The influence of land-cover type on the relationship between NDVI–LST and LST-tair," *Int. J. Remote Sens.*, vol. 39, no. 5, pp. 1377–1398, Mar. 2018.
- [52] Y. Guo, T. Gál, G. Tian, and J. Unger, "Model development for the estimation of urban air temperature based on surface temperature and NDVI—A case study in Szeged," *Acta Climatologica*, vol. 54, no. 3, pp. 29–40, 2020.
- [53] W. Zhu, A. Lú, and S. Jia, "Estimation of daily maximum and minimum air temperature using MODIS Land Surface Temperature products," *Remote Sens. Environ.*, vol. 130, pp. 62–73, Mar. 2013.
- [54] J. Cristóbal, M. Ninyerola, and X. Pons, "Modeling air temperature through a combination of remote sensing and GIS data," *J. Geophys. Res., Atmos.*, vol. 113, no. D13, pp. 1–13, Jul. 2008.
- [55] K. Wang, Z. Li, and M. Cribb, "Estimation of evaporative fraction from a combination of day and night land surface temperatures and NDVI: A new method to determine the Priestley–Taylor parameter," *Remote Sens. Environ.*, vol. 102, nos. 3–4, pp. 293–305, Jun. 2006.
- [56] E. Bustos and F. J. Meza, "A method to estimate maximum and minimum air temperature using MODIS surface temperature and vegetation data: Application to the maipo basin, Chile," *Theor. Appl. Climatol.*, vol. 120, nos. 1–2, pp. 211–226, Apr. 2015.
- [57] J. Che, M. Ding, Q. Zhang, Y. Wang, W. Sun, Y. Wang, L. Wang, and B. Huai, "Reconstruction of near-surface air temperature over the Greenland ice sheet based on MODIS data and machine learning approaches," *Remote Sens.*, vol. 14, no. 22, p. 5775, Nov. 2022.
- [58] X. Zhu, Q. Zhang, C.-Y. Xu, P. Sun, and P. Hu, "Reconstruction of high spatial resolution surface air temperature data across China: A new geointelligent multisource data-based machine learning technique," *Sci. Total Environ.*, vol. 665, pp. 300–313, May 2019.
- [59] D. Phiri, M. Simwanda, S. Salekin, V. Nyirenda, Y. Murayama, and M. Ranagalage, "Sentinel-2 data for land cover/use mapping: A review," *Remote Sens.*, vol. 12, no. 14, p. 2291, Jul. 2020.
- [60] K. Karra, C. Kontgis, Z. Statman-Weil, J. C. Mazzariello, M. Mathis, and S. P. Brumby, "Global land use/land cover with Sentinel 2 and deep learning," in *Proc. 2021 IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2021, pp. 4704–4707.

- [61] S.-Z. Liang, W.-D. Ma, X.-Y. Sui, H.-M. Yao, H.-Z. Li, T. Liu, X.-H. Hou, and M. Wang, "Extracting the spatiotemporal pattern of cropping systems from NDVI time series using a combination of the spline and HANTS algorithms: A case study for Shandong province," *Can. J. Remote Sens.*, vol. 43, no. 1, pp. 1–15, Jan. 2017.
- [62] N. Vorobiova and A. Chernov, "Curve fitting of MODIS NDVI time series in the task of early crops identification by satellite images," *Proc. Eng.*, vol. 201, pp. 184–195, May 2017.
- [63] Z. Cai, P. Jönsson, H. Jin, and L. Eklundh, "Performance of smoothing methods for reconstructing NDVI time-series and estimating vegetation phenology from MODIS data," *Remote Sens.*, vol. 9, no. 12, p. 1271, Dec. 2017.
- [64] X. Li, W. Zhu, Z. Xie, P. Zhan, X. Huang, L. Sun, and Z. Duan, "Assessing the effects of time interpolation of NDVI composites on phenology trend estimation," *Remote Sens.*, vol. 13, no. 24, p. 5018, Dec. 2021.
- [65] I. Kloog, A. Chudnovsky, P. Koutrakis, and J. Schwartz, "Temporal and spatial assessments of minimum air temperature using satellite surface temperature measurements in Massachusetts, USA," *Sci. Total Environ.*, vol. 432, pp. 85–92, Aug. 2012.
- [66] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019.
- [67] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Phys. D, Nonlinear Phenomena*, vol. 404, Mar. 2020, Art. no. 132306.
- [68] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space Odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
- [69] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 156–165.
- [70] C. Lea, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks: A unified approach to action segmentation," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Cham, Switzerland: Springer, 2016, pp. 47–54.
- [71] R. Wan, S. Mei, J. Wang, M. Liu, and F. Yang, "Multivariate temporal convolutional network: A deep neural networks approach for multivariate time series forecasting," *Electronics*, vol. 8, no. 8, p. 876, Aug. 2019.
- [72] A. Sbrana, A. L. D. Rossi, and M. C. Naldi, "N-BEATS-RNN: Deep learning for time series forecasting," in *Proc. 19th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2020, pp. 765–768.
- [73] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting," 2019, *arXiv:1905.10437*.
- [74] C. E. Hachimi, S. Belaqqiz, S. Khabba, and A. Chehbouni, "Evaluation of statistical and deep learning methods for short-term weather forecasting in semi-arid regions," in *Proc. Int. Conf. Medit. Geosci. Union*, Cham, Switzerland: Springer, Jan. 2022, pp. 203–206.
- [75] R. E. Dickinson, "Land-atmosphere interaction," *Rev. Geophysics*, vol. 33, no. S2, pp. 917–922, Jul. 1995.
- [76] P. Gentile, "Land-atmosphere interactions in the tropics—a review," *Hydrol. Earth Syst. Sci.*, vol. 23, no. 10, pp. 4171–4197, 2019.
- [77] L. Jach, T. Schwitalla, O. Branch, K. Warrach-Sagi, and V. Wulfmeyer, "Sensitivity of land-atmosphere coupling strength to changing atmospheric temperature and moisture over Europe," *Earth Syst. Dyn.*, vol. 13, no. 1, pp. 109–132, Jan. 2022.
- [78] M. Moon, D. Li, W. Liao, A. J. Rigden, and M. A. Friedl, "Modification of surface energy balance during springtime: The relative importance of biophysical and meteorological changes," *Agricult. Forest Meteorol.*, vol. 284, Apr. 2020, Art. no. 107905.
- [79] D. Zhou, J. Xiao, S. Bonafoni, C. Berger, K. Deilami, Y. Zhou, S. Frohking, R. Yao, Z. Qiao, and J. A. Sobrino, "Satellite remote sensing of surface urban heat islands: Progress, challenges, and perspectives," *Remote Sens.*, vol. 11, no. 1, p. 48, Dec. 2018.
- [80] J. Lee, "Assessment of U.S. urban surface temperature using GOES-16 and GOES-17 data: Urban heat island and temperature inequality," *Weather, Climate, Soc.*, vol. 16, no. 2, pp. 315–329, Apr. 2024.



JANGHO LEE received the B.S. degree in Earth and environmental sciences from Seoul National University, Seoul, South Korea, in 2018, and the Ph.D. degree in atmospheric sciences from Texas A&M University, College Station, TX, USA, in 2023.

Since 2023, he has been a Postdoctoral Researcher with the University of Illinois Chicago, as a part of the Community Research on Climate and Urban Science (CROCUS) Project. His research interests include climate informatics, statistical climatology, deep learning, remote sensing, climate impacts, and land-atmospheric modeling.

Dr. Lee has been a member of American Geophysical Union (AGU) and American Meteorological Society (AMS) since 2018.

...