IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE GEOSCIENCE AND REMOTE SENSING SOCIETY SECTION

Received 14 May 2026, accepted 30 May 2026, date of publication 3 June 2026, date of current version 9 June 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3700054

RESEARCH ARTICLE

Spatiotemporal Downscaling and Nowcasting of Urban Land Surface Temperatures With Deep Neural Networks

SOLOMIIA KURCHABA AND ANGELA MEYER
Department of Geoscience and Remote Sensing, Delft University of Technology, 2600 GA Delft, The Netherlands
School of Engineering and Computer Science, Bern University of Applied Sciences, 2501 Biel, Switzerland

Corresponding author: Solomiia Kurchaba (skurchaba@tudelft.nl)

This work was supported in part by the Horizon Europe Project UrbanAIR through the Swiss State Secretariat for Education, Research and Innovation (SERI).

ABSTRACT Land Surface Temperature (LST) is a key variable for various applications, such as urban climate and ecology studies. Yet, existing satellite-derived LST products provide either high spatial or high temporal resolution, resulting in a fundamental trade-off between the two. To address this trade-off, we combine observations from a geostationary and a polar orbiting satellite and provide LST fields at high spatial and high temporal resolution (1 km at 15-min intervals). We demonstrate their application for intraday forecasting of LSTs. To estimate LST fields at high spatiotemporal resolution, a U-Net model is trained to map LST fields from SEVIRI/MSG (3 km and 15 min resolution) along with corresponding solar zenith angles to LST fields from Terra/Aqua MODIS (1 km, 4 overpasses per day) that are collocated in space and time. The presented model has been trained on LSTs across large European cities with a population exceeding 1 million inhabitants, and achieves an RMSE = 1.92 °C and near-zero bias MBE = 0.01 °C on the hold-out test set. As a second step, we present an LST nowcasting model based on ConvLSTM architecture, trained across downscaled LST fields with forecast lead times of 15 to 75 minutes. The nowcasting model outperforms a persistence and a Climatological Rolling Median benchmarks, with RMSEs of 0.57 to 1.15 °C for the considered lead times and biases ranging from −0.1 to 0.14 °C. An additional validation conducted against independent MODIS overpasses confirms robust performance. Our LST forecast model at high spatiotemporal resolution is directly applicable to operational satellite-based LST monitoring.

INDEX TERMS Deep learning, downscaling, Europe, forecasting, land surface temperature, meteosat, MODIS, SEVIRI, urban environments.

I. INTRODUCTION

Land surface temperature (LST) is a critical parameter in various fields. It plays an important role in the monitoring of climate change, urban heat islands, droughts, and heatwaves, being also relevant in such fields as agriculture, hydrology, weather forecasting, ecosystem monitoring, and more generally in surface energy balance studies. As we are facing unprecedented challenges related to global warming, one of the most relevant applications of satellite-derived LSTs is the detection and characterization of surface urban heat islands. The impact of urban heat islands is particularly critical in rapidly growing cities worldwide, where urbanization, land use changes, and population density amplify thermal stress and energy demand [1]. Since LST can vary substantially over short time and small spatial scales, an accurate short-term forecasting of high-resolution LSTs becomes crucially important for understanding and management of urban heat islands [2], as well as a wide range of other applications such as ecosystem monitoring [3] and energy demand forecasting [4].

Despite their high potential, several challenges must be addressed to enable the effective use of satellite-derived LSTs in urban applications. A major limiting factor is the trade-off

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

between spatial and temporal resolution of the available satellite-derived LST field estimates. Geostationary sensors such as Meteosat Spinning Enhanced Visible and Infrared Imager (SEVIRI) usually provide observations with high temporal resolution of several observations per hour, while having lower spatial resolution of around 3 km at nadir in mid latitudes. Sensors in low Earth orbit offer a higher spatial resolution of 1 km or less, but typically at most two observations per day [5], which is insufficient for studying sub-diurnal LST processes [5], [6], [7].

Several studies have investigated the problem of downscaling LST from geostationary satellites across urban areas (e.g. [8], [9], [10]) using various statistical methods. For example, in [11] the authors used several support vector machine (SVM) models to downscale geostationary LST to a 1 km resolution for Athens. In [12], a least square support vector machine was used to downscale geostationary LST to 1 km resolution, for one day per season during 2012. In [13], the authors used a multi-layer perceptron model to downscale LST from approximately 4.5 km to 750 m for the city of Madrid. However, all the above-mentioned studies used classical machine learning models that use human-defined features and require spatial integration of the data, which usually results in the loss of information and sub-optimal model performance. In [14], the authors used a convolutional neural network (CNN) to downscale LST fields from 2 km to 70 meters for the city of Los Angeles. The study uses LST from the GOES-R geostationary sensor as input, and LST from ECOSTRESS as a target variable. However, an average revisiting time of an ECOSTRESS satellite is 3-4 days, which makes it difficult to collect enough data for the training of a deep learning model for multiple urban areas, especially when considering urban areas with a limited number of days suitable for satellite observations. Such a limitation makes the proposed approach difficult to scale.

Research on forecasting of LST fields at intraday timescales remains limited, particularly for products derived from Meteosat satellites or across European cities. In [15], several machine learning models were proposed for one-step-ahead forecasting of MODIS LST. While maintaining the spatial resolution of 1 km, such a model operates at the native time resolution of MODIS, and therefore cannot provide intraday forecasts. In [16], the authors introduced a deep learning architecture for nowcasting very high resolution (30 m) LST. This approach relies on Landsat-8 measurements to achieve the fine spatial detail, but as a result, the forecasts are tied to a single typical overpass time (approximately 10:30 am local time). Moreover, the 16-day revisit cycle of the satellite limits the method's ability to support intraday or even medium-range LST forecasting.

Our study introduces the first intraday satellite-based deep learning forecast model of high-resolution LST fields. We also demonstrate the first deep learning LST downscaling model that enables 1 km resolution LSTs updated every 15 minutes across Europe. We propose a combined LST downscaling and nowcasting approach. First, we develop a pan-European LST downscaling model trained using SEVIRI-derived LST fields as input and co-located MODIS-derived LST fields as high-resolution targets. This model is designed to generalize across major European urban areas (population > 1 million), enabling the generation of high-resolution (1 km), high-frequency (15-minute) LST fields across the continent.

Building on this generalized capability, we then focus on a city-scale application to demonstrate the practical utility of the downscaled data for short-term forecasting. We present an LST nowcast model that forecasts LST fields across European cities for up to 75 minutes ahead. We demonstrate our approach for three representative European cities: Bucharest, Antwerp, and Berlin. This two-step approach allows us to first establish a robust, transferable spatial downscaling framework and subsequently assess its value in a localized, operational nowcasting setting. Finally, we evaluate our proposed LST downscaling and nowcasting approach against actual MODIS observations.

The main contributions of this study are as follows:

- We present the first deep learning LST downscaling model that enables 1 km resolution LSTs updated every 15 minutes across Europe. Our model maps 15-min 3-km SEVIRI LST to 1-km MODIS LST with 2-4 overpasses per day.
- Our downscaling model is designed to generalize across many urban areas, not just one, including all major European cities (population>1 million) at a low error and a near-zero bias.
- We introduce the first intraday satellite-based deep learning nowcasting model for high-resolution LST fields and demonstrate its application at the city scale using Bucharest, Antwerp, and Berlin as representative case studies. Building on the downscaled 1 km, 15-minute LST fields generated by our pan-European model, this nowcasting framework enables short-term urban LST forecasting with lead times ranging from 15 to 75 minutes. This two-step design highlights the practical value of the downscaling approach by bridging continental-scale data generation with localized forecasting applications. Our proposed nowcasting model significantly outperforms the two data-driven benchmark models and shows strong agreement with MODIS LST observations.

This paper is organized as follows. In Section II, we describe the data used in this study, as well as the preprocessing workflow used to construct training and test datasets for the LST downscaling and nowcasting approaches. Section III presents the methodological framework, including the U-Net downscaling model, the ConvLSTM nowcasting method, and the benchmark predictors. In Section IV, we report the results of the performed experiments for LST downscaling and nowcasting, followed by a further validation against MODIS observations. Finally, in Section V,

we discuss the main findings, limitations, and directions for future work.

II. DATA AND PREPROCESSING

A. DATA SOURCES

In this study, we integrate several satellite and auxiliary datasets to construct a multi-sensor machine learning dataset for the downscaling and nowcasting of LST fields.

1) SEVIRI/MSG LST

We use the clear-sky Land Surface Temperature (MLST) product [17] provided by the Satellite Applications Facility on Land Surface Analysis (LSA SAF) of the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT). The LSTs were derived from spectral radiance measurements of the Spinning Enhanced Visible and Infrared Imager (SEVIRI) onboard the Meteosat Second Generation (MSG) [18] geostationary Earth monitoring satellite series. SEVIRI provides full-disk observations with a viewing zenith angle between 0° and 80°. It offers a temporal resolution of 15 minutes and a spatial resolution at nadir of approximately 3 km (0.05°).

2) MODIS LST

As a high-resolution LST target variable, we use daily Level-2 LST products [19] MOD21 (Terra) and MYD21 (Aqua), product version: 0.61. These products provide global coverage at 1 km spatial resolution at nadir. Equatorial overpass times are approximately 10:30 a.m./p.m. local solar time for Terra and 1:30 a.m./p.m. for Aqua.

3) AUXILIARY DATA

We include the Solar Zenith Angle (SZA), computed for each pixel at the satellite overpass time using the pvlib Python package (v0.13.0), as an auxiliary predictor. SZA directly controls the amount of incoming shortwave radiation reaching the surface and is therefore a primary driver of the diurnal and spatial variability of LST. Including SZA introduces physically meaningful information on solar illumination geometry, helping the model account for insolation-driven temperature variations and improving the physical consistency of the predicted LST fields.

We note, however, that SZA represents only a first-order control on radiative forcing. Other factors such as land cover, vegetation state (e.g., NDVI), urban morphology (e.g., sky view factor or building volume), and anthropogenic influences (e.g., population density or pollution) also affect the surface energy balance through mechanisms such as heat storage, radiative trapping, and anthropogenic heat release. These variables are not explicitly included in the current framework. Instead, we assume that part of their aggregated effect is implicitly encoded in the coarse-resolution LST signal used as input. This design choice allows us to develop a sensor-driven and spatially transferable approach that does not rely on auxiliary datasets, which are often heterogeneous, static, or unavailable at appropriate spatial and temporal resolutions.

B. DATA PREPROCESSING

We make use of 21 years of satellite observations (2004–2025). Because our primary application is the characterization of urban heat islands [20], we restrict the dataset to the climatologically warmest months of the year: 15 May–15 September. Training data are collected for European cities with populations exceeding one million inhabitants, ensuring diversity in climatic, morphological, and land-cover characteristics. As a hold-out test set for both downscaling and nowcasting models, we use the satellite measurements from the years 2007, 2013, 2019, and 2025. The rest of the years from the studied period are used for the training and validation of the models.

1) DOWNSCALING

To prepare the data for the training of the LST downscaling model, we perform the following data preprocessing steps. First, we acquire MODIS LST Level-2 granules and perform quality filtering based on quality flag values (only pixels with good or nominal quality are kept), LST accuracy (only pixels with good or excellent LST retrieval performance are kept), viewing angles ($< 50°$), and cloud flag (cloud, cloud shadow, or cirrus pixels are removed). In addition, we perform the removal of outliers: LST pixels with values below 0 °C and above 65 °C are removed.

Next, we spatially subset the data to patches covering each target city and regrid these patches onto a common, uniform grid of a resolution 0.01°. The regridding step is necessary because the original MODIS data are provided on a swath-based grid, where pixel locations and spacing vary with satellite viewing geometry. By ''uniform grid,'' we refer to a regular, fixed-resolution grid with consistent spacing in latitude and longitude, which ensures spatial alignment across samples. Only images with less than 50% missing or poor-quality pixels (i.e., pixels removed during filtering) are retained. These images serve as the ground-truth targets for training the LST downscaling model.

For each MODIS overpass included in the training set, we identify the temporally closest SEVIRI acquisition. Since SEVIRI observations are available every 15 minutes, the time difference between the paired MODIS and SEVIRI acquisitions is always less than 15 minutes. SEVIRI scenes undergo the same spatial subsetting and are additionally filtered to remove cloud- or water-contaminated pixels before being regridded to the same uniform grid. Due to the coarser spatial resolution of SEVIRI, each pixel represents a larger area and thus has a higher likelihood of being affected by clouds or other invalid conditions. To ensure sufficient data quality, we therefore impose a stricter requirement on data coverage compared to MODIS: only scenes with more than 80% valid pixels (i.e., pixels remaining after filtering) are retained.

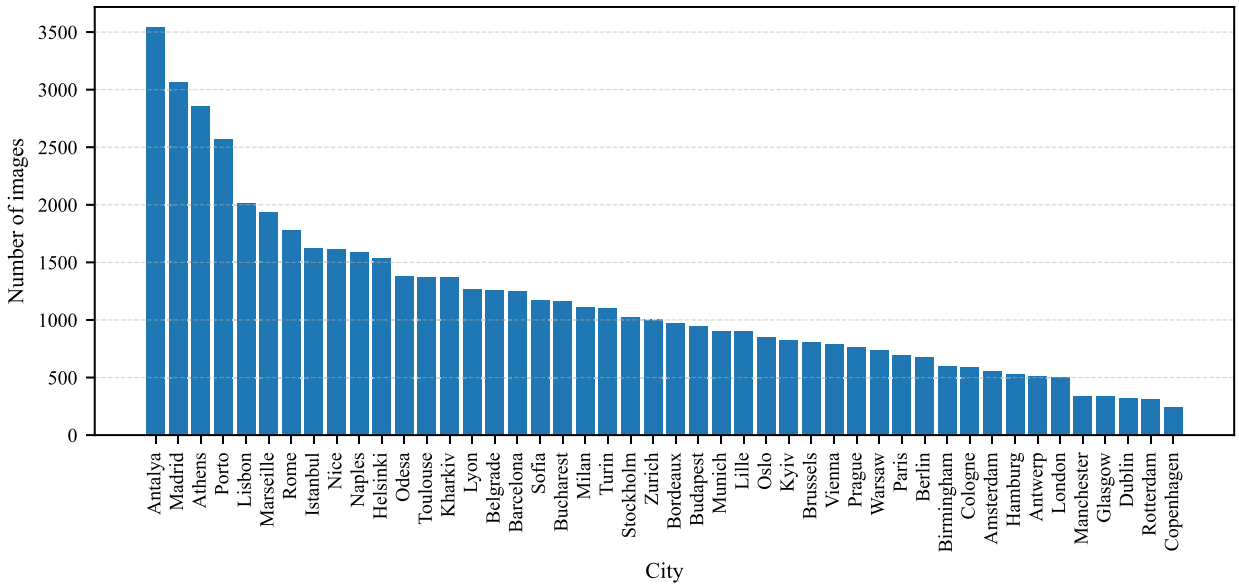

FIGURE 1. Distribution of image patches per city in the dataset before training-test split and dataset balancing. For model training, a maximum of 1500 samples per city is used to ensure a more uniform representation across regions.

We only retain MODIS–SEVIRI pairs when at least 75% of their valid-pixel areas spatially overlap. All thresholds were derived from an empirical examination of MODIS–SEVIRI pairs to balance the size of the resulting dataset and the image quality. The distribution of image patches per city is shown in Fig. 1. To ensure a balanced representation across cities and to prevent over-representation of specific climatic regions, we limit the number of training samples to a maximum of 1500 images per city. This downsampling strategy mitigates biases toward cities with higher data availability and promotes more uniform learning across different geographic and climatic conditions. Ultimately, our dataset contains 52938 datapoints.

2) NOWCASTING

We include all available SEVIRI acquisitions for the studied period of 2004-2025 in the training of the LST nowcasting model. We spatially subset the data around cities of interest, and regrid the resulting image patches to the grid used for the LST downscaling model. We apply the pre-saved downscaling model to estimate the prepared SEVIRI image patches at higher spatial resolution. The resulting patches are then arranged into time-series sequences to be used as an input to the nowcasting model. Ultimately, for Bucharest, Antwerp, and Berlin, the training set is composed of respectively 98330, 54031, and 51237 datapoints. The test set contains 24152, 12042, and 11834 datapoints respectively.

III. METHOD

This Section presents the methodological framework of the proposed downscaling–nowcasting pipeline. We first describe the formulation of the LST downscaling task and the U-Net architecture used to reconstruct high-resolution MODIS-like LST fields from lower-resolution SEVIRI inputs. We then introduce the nowcasting setup, including the temporal prediction problem, the construction of training samples, and the ConvLSTM-based model used to forecast future high-resolution LST fields. Finally, we define the benchmark predictors used for evaluation.

The proposed framework follows a two-stage design, where spatial downscaling and temporal prediction are learned separately. This choice is motivated by both methodological and practical considerations. First, the spatial and temporal components address distinct learning problems: the U-Net focuses on reconstructing spatial detail from coarse-resolution observations, while the ConvLSTM captures temporal dynamics of already downscaled fields. Decoupling these tasks simplifies training and allows each model to specialize in a well-defined objective. Moreover, the availability of supervision differs between tasks. High-resolution MODIS LST provides a direct target for training the spatial downscaling model, whereas temporally consistent high-resolution sequences are not available without applying a downscaling model. The sequential design enables the use of all available high-resolution snapshots for spatial learning, while leveraging temporally dense SEVIRI observations for forecasting.

A. DOWNSCALING

Formally, the LST downscaling task is defined as a map

$$Y = f(X) \tag{1}$$

where X is the low-resolution SEVIRI LST field (augmented as in our case by SZA or potentially other auxiliary predictors), Y is the corresponding high-resolution MODIS LST field, and f is a regression function approximated via a

machine learning model. In this work, we deliberately restrict the predictor space to LST fields and solar illumination geometry (SZA), in order to investigate how much of the kilometer-scale spatial variability can be recovered from satellite-derived thermal signals alone. Thus, the proposed framework should be interpreted as a data-minimal and sensor-driven approach, rather than a fully physically explicit urban climate model.

We implement a convolutional encoder–decoder U-Net architecture [21] (architecture details can be found in Appendix A). A U-Net consists of a contracting path that captures contextual information through successive convolution and pooling operations, and an expansive path that progressively reconstructs high-resolution predictions. Skip connections between encoder and decoder layers help preserve fine-scale spatial patterns, making the U-Net particularly effective for tasks requiring both global context and local detail. In remote sensing, U-Net architectures have been successfully applied to cloud segmentation [22], land-cover and land-use classification [23], and image super-resolution [24], demonstrating robustness under varying spatial resolutions and sensor characteristics.

Because the cities in our dataset vary in spatial extent, input images are padded to a standardized dimension of 128×128 pixels, which is computationally efficient and compatible with the downsampling scheme of the U-Net. To fill missing values in the input, we use k-nearest neighbors (KNN) imputation with k = 5, following common practice. We note that, in urban environments, LST fields can exhibit strong spatial heterogeneity with sharp thermal gradients (e.g., between built-up areas, vegetation, and water bodies), which may not be fully captured by the local smoothness assumption underlying KNN. As a result, this approach can introduce localized smoothing or interpolation artifacts, particularly when the fraction of missing pixels is high. However, in our dataset, missing values typically affect a limited portion of each image, and KNN imputation provided a practical trade-off between simplicity and performance. We did not perform explicit hyperparameter optimization for k, as the focus of this study is on the downscaling model rather than the imputation method. In addition, such experiments would significantly increase computational costs. During preliminary experiments, this method yielded better downstream performance than image-wise mean and median imputation, as reflected by lower RMSE (and consistent trends in other regression metrics) on a validation set. Moreover, any potential smoothing effects introduced during imputation are partially mitigated by the learning-based downscaling model, which is trained to recover high-resolution spatial variability from the input data.

B. NOWCASTING

We refer to intraday forecasts with lead times up to 75 minutes as nowcasting and use these terms interchangeably here.

TABLE 1. Performance of the LST downscaling model on the hold-out test set.

R^2	RMSE [°C]	MAE [°C]	MBE [°C]
0.97	1.92	1.44	0.01

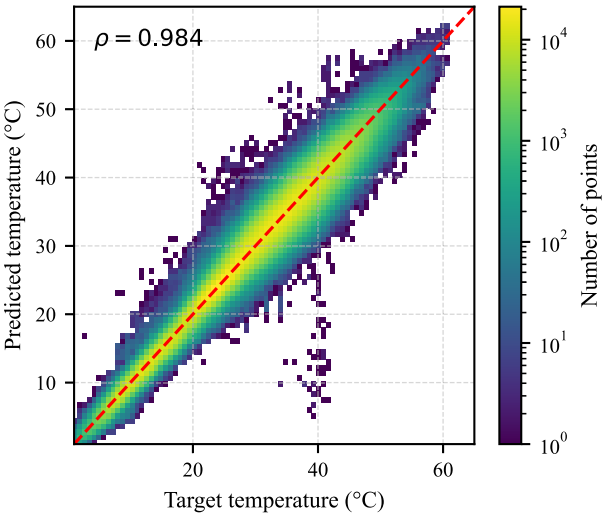

FIGURE 2. Predictions versus target values of the LST downscaling model. The dashed red line corresponds to 45-degree line. ρ corresponds to the Pearson correlation coefficient.

We define a nowcasting task as follows:

$$\hat{X}_{t+\Delta t} = g\left(Xt - 2\delta t, X_{t-\delta t}, X_t\right) \tag{2}$$

where X_t denotes the downscaled SEVIRI LST field at time t, $\delta t = 15$ min is the native SEVIRI sampling interval, $\Delta t \in 15, 30, 45, 60, 75$ min is the prediction lead time, and g is a non-linear mapping learned by the model. We use the three most recent time steps as input to capture short-term temporal dynamics while keeping the input dimensionality manageable. We did not perform an explicit hyperparameter search over the number of input frames, as preliminary experiments indicated that three time steps provide a reasonable trade-off between performance and computational cost.

To construct the training samples, we use trailing temporal windows of three consecutive downscaled SEVIRI-derived LST fields as predictors, while the target is the future downscaled LST field at lead time Δt at MODIS resolution. A separate model is trained for each lead time using a corresponding dataset.

For the task of nowcasting, we use a Convolutional LSTM Network, which combines recurrent temporal memory with convolutional operators [25], [26], [27]. For architecture specification, see Appendix B. Such a model processes spatiotemporal data by stacking ConvLSTM layers to capture spatial features and temporal dependencies simultaneously. We use an encoder-decoder structure as it is shown to be effective for time-dependent images or video predictions [26], [27]. This architecture is suitable for LST nowcasting because it preserves spatial coherence (e.g., intra-urban thermal gradients) while learning temporal dependencies in

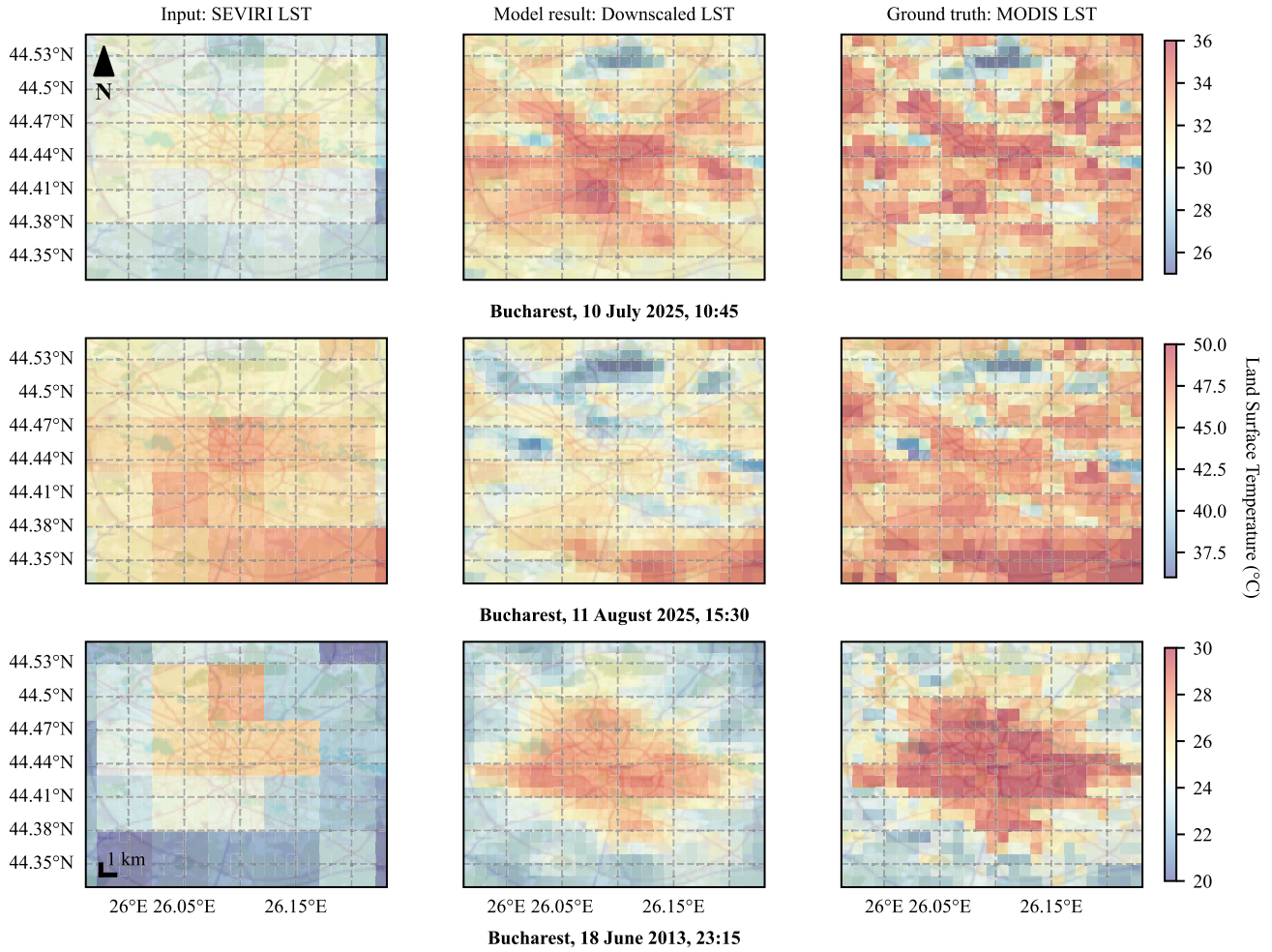

FIGURE 3. Visual examples of the performance of the downscaling model on the example city of Bucharest. All times are in local summer time of Bucharest (UTC+3). Background maps of Bucharest are generated using python package contextily v.1.6.2 with OpenStreetMap Humanitarian map provider.

image sequences, required for sub-hourly urban temperature nowcasting.

1) BENCHMARKS

To evaluate the ConvLSTM-based LST forecast model, we compare it against two complementary data-driven baseline predictors: the Persistence and Climatological Rolling Median benchmarks.

The Persistence benchmark assumes the LST field does not change as time progresses, so the Persistence benchmark propagates the latest available field to all forecast lead times. Formally, for each $\Delta t \in \{15, 30, 45, 60, 75\}$ min,

$$\hat{X}^{\text{persistence}}_{t+\Delta t} = X_t. \tag{3}$$

Persistence is a reasonable baseline in high-frequency geophysical nowcasting because LST fields are usually strongly autocorrelated in time at intraday forecast horizons.

The Climatological Rolling Median benchmark accounts for the typical diurnal behavior at each time slot while remaining insensitive to short-term fluctuations. For a given timestamp t and lead time Δt, the prediction is defined pixel-wise as

$$\hat{X}^{\text{climatology}}_{t+\Delta t} = \text{median}\left\{X_{t+\Delta t-k\cdot 24\text{h}}\right\}_{k=1}^{5}, \tag{4}$$

that is, the median over the previous five days at the same local time of day and location.

These two benchmarks are relevant and complementary for LST nowcasting: Persistence tests whether the model improves over a strong short-memory predictor, while Climatological Rolling Median tests whether the model captures event-specific departures from the expected diurnal cycle. Demonstrating gains over both baselines indicates that the model learns both immediate temporal evolution and non-stationary dynamics.

IV. RESULTS

In this Section, we present the results of our experiments. We start with the evaluation of the LST downscaling model, which will then be used for improving the spatial resolution of the inputs of the LST nowcasting model.

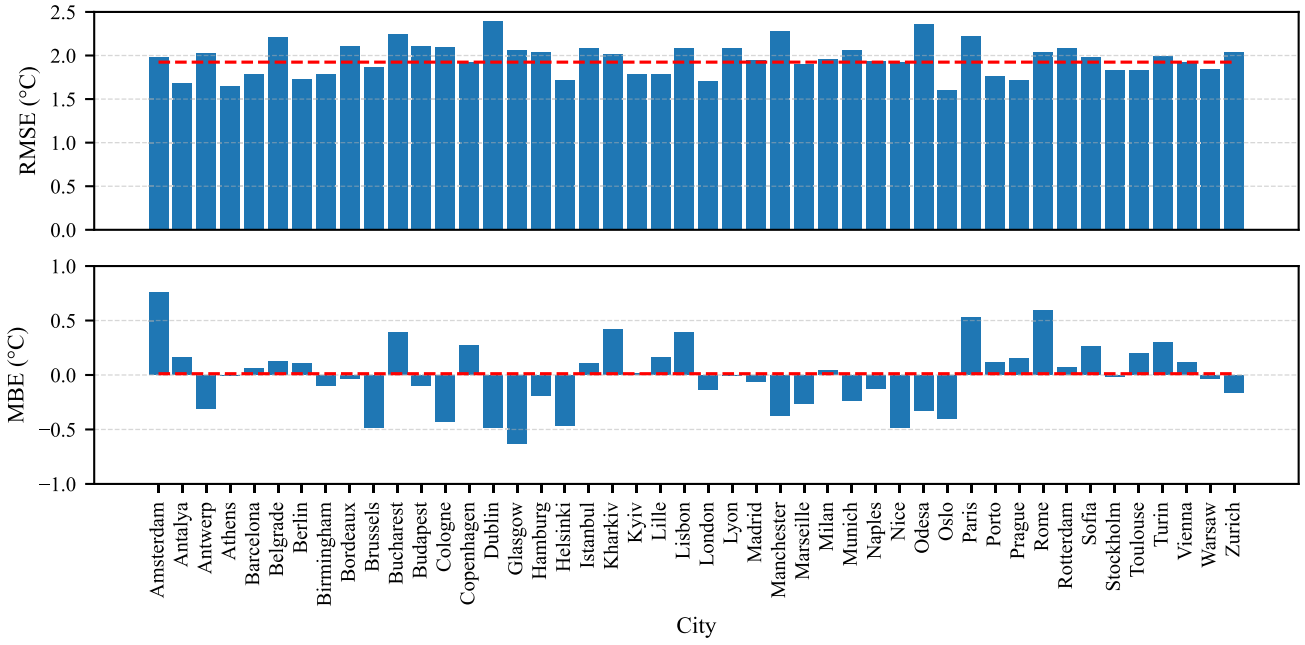

FIGURE 4. Downscaling performance per city. Top panel: Root Mean Square Error (RMSE) per studied city. Bottom panel: Mean Bias Error (MBE) per studied city. Red dashed lines: RMSE/MBE over the whole dataset.

TABLE 2. Performance of the LST nowcasting models on the hold-out test set for different lead times.

City	Lead time [min]	R^2	RMSE [°C]	MAE [°C]	MBE [°C]
Bucharest	15	0.99	0.59	0.42	0.04
	30	0.99	0.76	0.55	-0.09
	45	0.99	0.87	0.62	-0.10
	60	0.99	0.99	0.70	-0.04
	75	0.98	1.15	0.82	-0.02
Antwerp	15	0.99	0.57	0.40	0.06
	30	0.99	0.67	0.47	-0.08
	45	0.99	0.77	0.55	0.01
	60	0.99	0.85	0.60	0.10
	75	0.98	0.97	0.68	0.04
Berlin	15	0.99	0.58	0.41	0.13
	30	0.99	0.67	0.48	0.13
	45	0.99	0.75	0.54	0.08
	60	0.99	0.86	0.62	-0.01
	75	0.99	0.97	0.70	0.04

Next, we present an evaluation of the performance of the nowcasting model. Finally, we present an evaluation of the whole downscaling-nowcasting pipeline against the actual high-resolution MODIS-derived LST fields at the available timestamps.

A. LST DOWNSCALING MODEL

In this Subsection, we present the results of the LST downscaling model. The model is trained to map low-resolution SEVIRI LST fields to high-resolution MODIS (Terra/Aqua) LST fields using the U-Net encoder–decoder architecture described in Section III. Evaluation on the hold-out test set shows a strong agreement between predictions and targets (Fig. 2) with an overall R^2 of 0.97, Root Mean Square Error (RMSE) of 1.92 °C, Mean Absolute Error (MAE) of 1.44 °C, and negligible Mean Bias Error (MBE) of 0.01 °C (see Table 1). The near-zero bias indicates that the model does not systematically over- or underestimate LST across the test set. In Fig. 3, we provide visual examples of the performance of the downscaling model on the example city of Bucharest over different observation times.

As a next step, we evaluate the robustness of the model across different geographic and temporal conditions by assessing its error metrics by city, month, and hour of day. Fig. 4 summarizes the RMSE and MBE per city included in this study. The spread in RMSE shows that performance is not uniform across all urban areas, reflecting differences in local climate and land-cover heterogeneity. The city-wise MBE remains close to zero for most cities, suggesting that the model does not introduce strong location-specific biases.

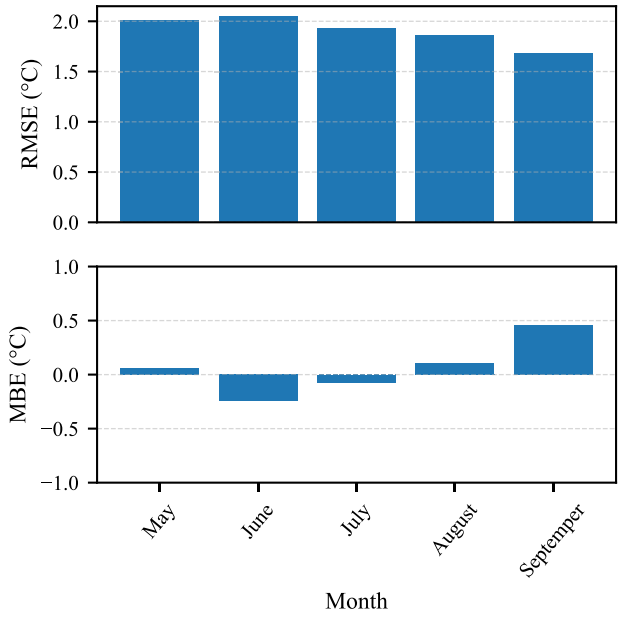

FIGURE 5. Downscaling performance per month. Top panel: Root Mean Square Error (RMSE) per analyzed month. Bottom panel: Mean Bias Error (MBE) per analyzed month.

Seasonal variations are shown in Fig. 5. Errors change across the warm-season months only slightly. The observed changes may reflect changing surface conditions (e.g., vegetation state and soil moisture) and the varying availability of clear-sky observations used to form the training data pairs.

Finally, Fig. 6 shows the LST downscaling model performance as a function of local hour. Results cluster around the nominal overpass times of the MODIS Terra and MODIS Aqua instruments, with a spread driven by longitudinal differences across cities, the use of civil time (including daylight saving time) rather than local solar time, and smaller timing offsets introduced by the wide MODIS swath. The hourly patterns indicate that the RMSE of the LST downscaling model depends on the diurnal cycle, just like the LSTs themselves. We observe a noticeably better performance of the model during the night when LSTs are lower. We also notice a deterioration of the model performance for hours that are further away from the typical overpass time of the satellite - this is likely caused by lower training and test data availability for those time stamps. Despite this variability, the MBE remains small across hours, confirming stable bias behavior throughout the day.

Overall, the results demonstrate that the proposed U-Net-based model can reliably enhance the spatial resolution of 15-min LST fields derived from geostationary satellite while maintaining low bias, which is a key prerequisite for the subsequent nowcasting experiments based on the downscaled LST field sequences.

B. LST NOWCASTING MODEL

In this Subsection, we present the results of the nowcasting models. For the nowcasting part of the experiment, we demonstrate the LST nowcasting model for three exemplary cities: Bucharest, Antwerp, and Berlin, selected to represent different urban and climatic conditions across Europe. Bucharest is characterized by a humid subtropical climate (Cfa in the Köppen classification [28]), Antwerp by a temperate oceanic climate (Cfb), and Berlin by a transitional climate between oceanic and continental regimes (Cfb/Dfb).

For the nowcasting task, a separate ConvLSTM model was trained for each city, for each lead time using downscaled sequences of SEVIRI-derived LST fields. The performance of the models is evaluated on a hold-out test set. The aggregated metrics for all three cities are summarized in Table 2. Across all cities, the nowcasting models demonstrate consistently high predictive skill. The R^2 values remain very high (0.98–0.99) for all lead times, indicating that the models effectively capture the short-term spatio-temporal evolution of LST. A gradual degradation of performance with increasing lead time is observed, as expected for forecasting tasks. For instance, RMSE increases from approximately 0.57–0.59 °C at 15 minutes to 0.97–1.15 °C at 75 minutes, while MAE increases from about 0.40–0.42 °C to 0.68–0.82 °C. Biases remain small for all cities and lead times, with MBE values close to zero (typically within ±0.1 °C). No systematic over- or underestimation is observed across the forecasting horizon. While small positive or negative biases appear at specific lead times, their magnitude remains negligible compared to the total prediction error. To provide a visual context, Fig. 7 shows prediction-versus-target scatter plots for Bucharest across different lead times, illustrating a close alignment between predictions and observations. Equivalent plots for Antwerp and Berlin are provided in Appendix C.

Despite the common trend, some inter-city differences emerge. From Table 2, we can see that Antwerp and Berlin generally exhibit slightly lower errors than Bucharest, particularly at longer lead times, whereas Bucharest exhibits the strongest increase in error with lead time.

We now further break down the results for the city of Bucharest. First, Fig. 8 and Fig. 9 show the distribution of RMSE and MBE per month. Similarly to the downscaling model, we do not see strong variations in the performances of the models across the studied months. Importantly, the relative ranking across lead times is preserved month by month (Fig. 8): shorter lead times consistently produce lower errors than longer lead times. From Fig. 9, we also see that even though the MBE exhibits slight seasonal dependency, it remains close to zero for all studied months. We also observe similar patterns for Antwerp and Berlin. Corresponding plots are provided in Appendix C.

Finally, with the diurnal analysis in Fig. 10 (respectively Fig. 15, and Fig. 16 in Appendix C), we compare our ConvLSTM approach with two benchmark methods: Persistence and Climatological Rolling Median. Our ConvLSTM-based LST nowcasting models achieve lower RMSEs across most hours of the day and for all lead times, indicating that our LST nowcasting models learn meaningful temporal

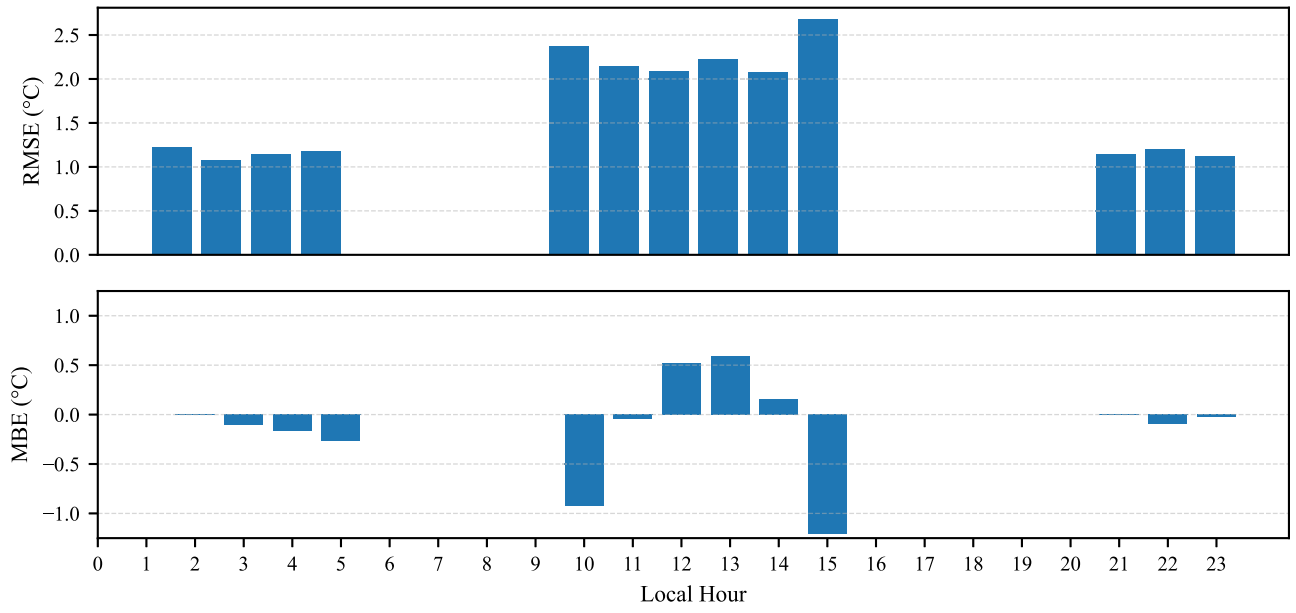

FIGURE 6. Downscaling performance per hour of day. Top panel: Root Mean Square Error (RMSE) per hour of the day. Bottom panel: Mean Bias Error (MBE) per hour of the day.

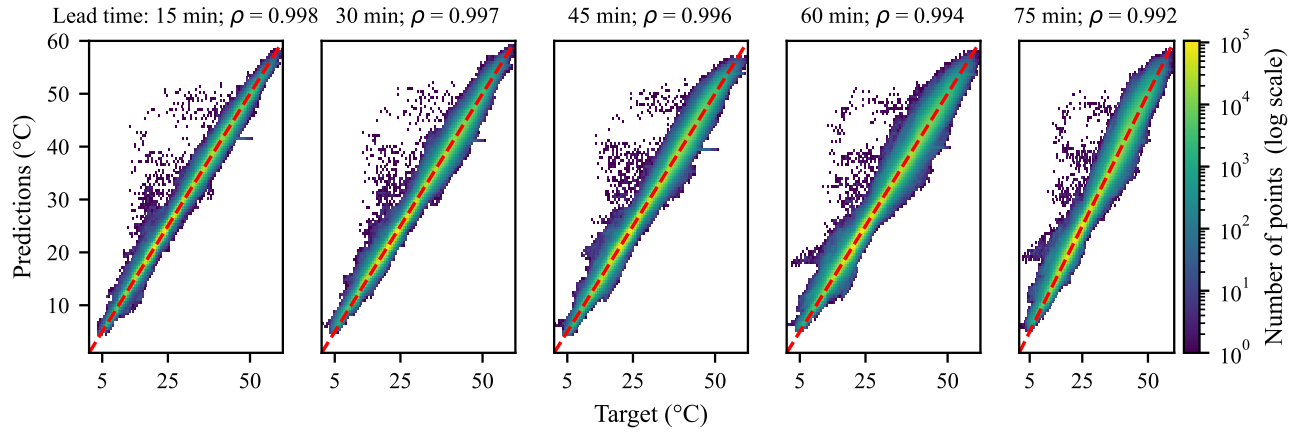

FIGURE 7. Predictions versus target values of the LST nowcasting model for different lead times for the city of Bucharest. The red line corresponds to 45-degree line. ρ corresponds to the Pearson correlation coefficient.

dynamics beyond simple extrapolation. At the same time, the curves reveal two time windows in which outperforming Persistence becomes more challenging, specifically near the transition points of the diurnal cycle where the temperature tendency changes sign. This is visible around approximately 14:00 local time, which is typically the time when LST reaches its daytime maximum and starts to decrease, and between approximately 03:00 and 05:00, when LSTs stabilize following nighttime cooling before increasing again after sunrise. Around these points, Persistence can be more difficult to surpass because the recent trajectory is less predictive of the immediate future than during monotonic warming or cooling periods. Taken together, these results show that our proposed LST nowcasting framework provides accurate and stable sub-hourly LST forecasts at 1 km resolution, making it suitable for urban heat monitoring applications.

C. VALIDATION AGAINST MODIS-DERIVED LST

To provide further validation, in this Subsection, we compare the LST forecasts of our model against spatiotemporally collocated MODIS-derived LSTs. In contrast to the previous Subsection, which evaluates the model against downscaled SEVIRI-derived LST fields, here we provide a direct comparison with independent MODIS-derived observations taken from the hold-out test set used for the evaluation

TABLE 3. Forecast LSTs compared to MODIS-derived LSTs, by lead time. We report results for daytime/nighttime overpasses of Terra and Aqua.

City	Lead time [min]	R^2	RMSE [°C]	MAE [°C]	MBE [°C]
Bucharest	15	0.78/0.92	2.47/1.10	1.91/0.86	0.66/0.27
	30	0.79/0.91	2.46/1.12	1.89/0.88	0.57/0.00
	45	0.81/0.88	2.32/1.19	1.75/0.93	0.55/0.04
	60	0.80/0.86	2.36/1.28	1.81/0.99	0.65/0.17
	75	0.78/0.82	2.45/1.44	1.89/1.05	0.44/-0.15
Antwerp	15	0.62/0.97	2.32/1.09	1.75/0.84	-0.19/-0.09
	30	0.61/0.97	2.33/1.09	1.77/0.86	-0.66/-0.07
	45	0.61/0.97	2.37/1.01	1.82/0.81	-0.40/-0.15
	60	0.58/0.97	2.45/1.00	1.90/0.79	-0.50/0.09
	75	0.44/0.96	2.70/1.22	2.10/0.99	-0.76/0.08
Berlin	15	0.89/0.97	1.96/0.98	1.54/0.74	-0.20/0.01
	30	0.89/0.97	1.97/0.93	1.55/0.71	-0.23/0.09
	45	0.88/0.96	2.11/0.89	1.68/0.69	-0.54/0.06
	60	0.87/0.96	2.15/0.88	1.67/0.68	-0.31/0.06
	75	0.86/0.96	2.25/0.93	1.73/0.72	-0.07/-0.09

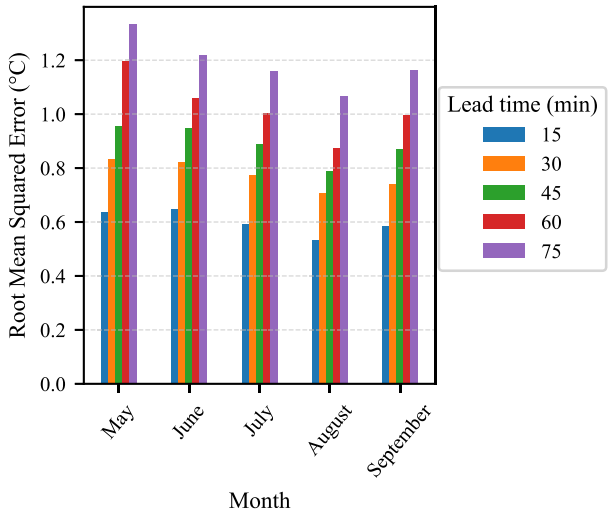

FIGURE 8. Nowcasting performance per studied month for the city of Bucharest: RMSE per month per lead time.

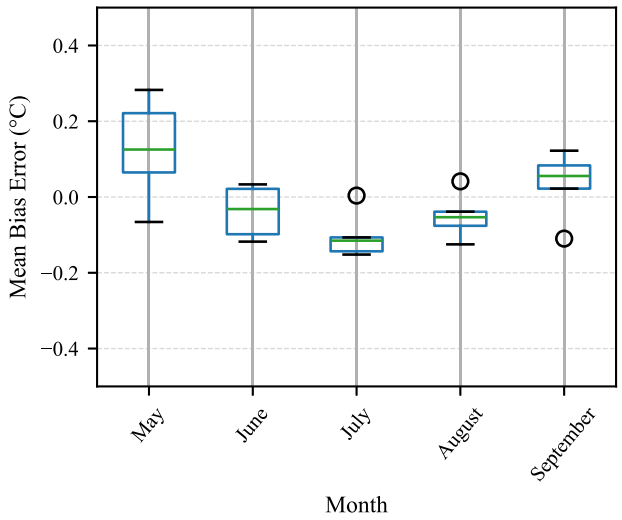

FIGURE 9. Nowcasting performance per studied month for the city of Bucharest: MBE distribution over lead times per month.

of the downscaling model. As in the previous subsection, we consider three cities for the analysis: Bucharest, Antwerp, and Berlin.

We split all available MODIS observations into those corresponding to the daytime and nighttime overpasses of Terra (equatorial overpass times are approximately 10:30 a.m./p.m.) and Aqua (equatorial overpass times are approximately 11:30 a.m./p.m.). The results are presented in Table 3. The results show a consistent contrast between nighttime and daytime observations. Nighttime performance remains high, with R^2 values typically ranging from 0.82 to 0.97 and RMSE between about 0.88 and 1.44 °C across lead times. In contrast, daytime performance is systematically lower, with R^2 values ranging more broadly (0.44–0.89) and RMSE values around 1.96–2.7 °C. This day–night contrast is coherent with earlier results from the downscaling subsection, where lower errors were also found at night, likely due to the higher amplitudes of LST variability during the daytime. Similar conclusions can be drawn from Fig. 11 (also Fig. 17 and Fig. 18 in Appendix C).

Some inter-city differences can also be identified. Berlin shows the most consistent performance across both daytime and nighttime conditions, maintaining relatively high R^2 values and comparatively low errors. Antwerp exhibits very strong nighttime agreement (RMSE below 1.22 °C) but lower daytime performance, particularly at longer lead times, where R^2 decreases more substantially. Bucharest shows intermediate behavior, with stable nighttime performance but slightly higher daytime errors compared to Berlin.

Biases remain relatively small overall but also exhibit a certain structure. Nighttime MBE values are generally close to zero for all cities, indicating the absence of systematic offsets. During the daytime, however, city-dependent biases emerge. Bucharest tends to show a positive bias (slight overestimation of MODIS LST), while Antwerp and Berlin exhibit negative biases. Moreover, the MBE of Berlin tends to be closer to 0 than the MBE of Antwerp and Bucharest. Those trends show similar patterns to what we observe in Figure 4 for the downscaling model. Nevertheless, all biases typically remain below 1 °C in magnitude and are small relative to the overall prediction error. Summing up, despite the slight daytime offset, this MODIS-based validation confirms that the proposed pipeline enables strong predictive skill and

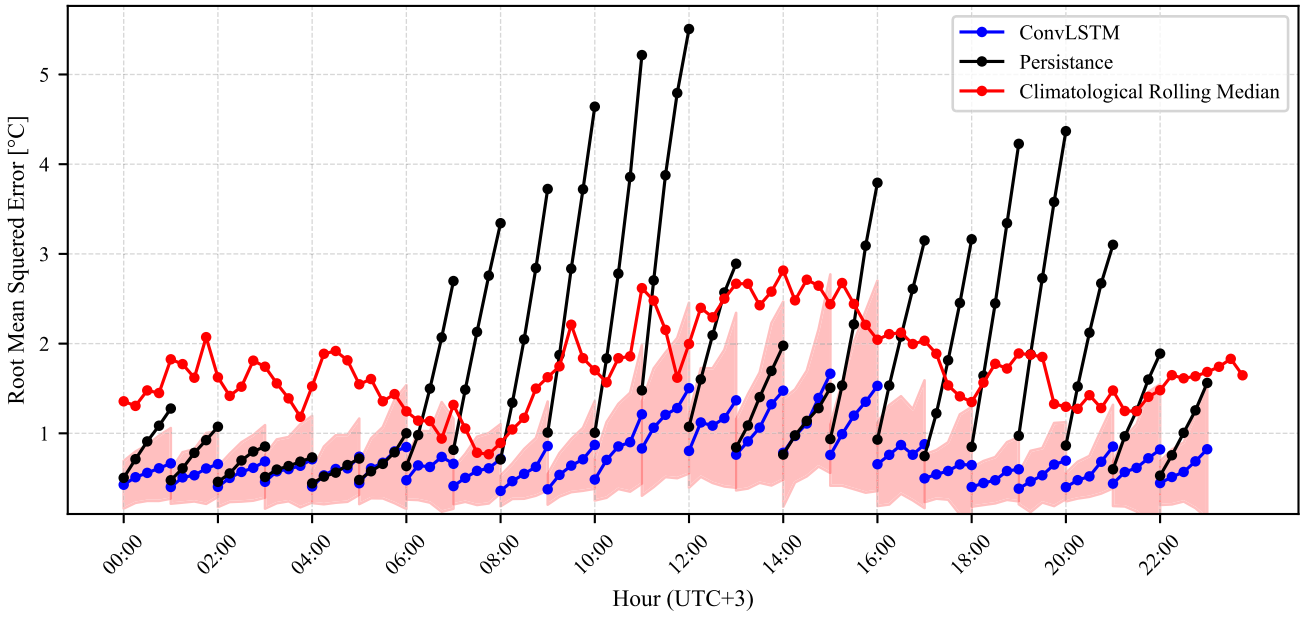

FIGURE 10. Diurnal evolution of the RMSE per lead time: ConvLSTM-based LST nowcasting model, Persistence benchmark and Climatological Rolling Median benchmark. All times are in local summer time of Bucharest (UTC+3).

remains robust for all studied lead times for high-resolution urban LST nowcasting.

V. DISCUSSION

In this study, we developed a comprehensive pipeline that combines downscaling and short-term nowcasting of land surface temperatures in urban areas. First, we presented a downscaling model trained to increase the spatial resolution of SEVIRI-derived LST fields using MODIS-derived LST as a target. The downscaling model was developed for large European cities and achieved an overall RMSE of 1.96 °C. A closer analysis of the error distribution revealed differences in downscaling quality across cities and months. We also observed that downscaling is generally more challenging for daytime observations than for nighttime observations. Nevertheless, MBE remained close to zero in most cases, suggesting that the model generalizes without introducing strong structural biases. Generalization across different climatic regimes is often a concern in data-driven LST modeling. However, based on the per-city evaluation, we do not observe a systematic dependency of model performance on geographic or climatic conditions within the studied domain. In particular, several northern cities exhibit among the lowest downscaling errors, while higher errors are observed in some warmer regions, suggesting that performance is not trivially linked to climate regime. This indicates that the proposed model is capable of capturing spatial variability across a range of environmental conditions during the warm season. Overall, these experiments indicated that the developed LST downscaling model is well-suited for preparing inputs for the follow-up LST nowcasting application.

As a second step, we developed LST nowcasting models trained on SEVIRI LST fields that were first downscaled by the downscaling model. For nowcasting, we focused on three cities of interest, Bucharest, Antwerp, and Berlin. We trained five separate LST nowcasting models per city for lead times ranging from 15 to 75 minutes. The general RMSE ranged from 0.57-0.59 to 0.97-1.15 °C, gradually increasing with the lead time. For all studied cities, the developed nowcasting models outperformed the two baselines, Persistence and Climatological Rolling Median, indicating that the proposed LST nowcasting models capture both immediate temporal evolution and non-stationary LST dynamics.

Finally, we compared the obtained nowcasting results with actual MODIS-derived LSTs from daytime and nighttime overpasses of MODIS Terra/Aqua. The obtained overall RMSE ranged between 1.96–2.7 °C for daytime predictions and between 0.88–1.44 °C for nighttime predictions. We also observed that nighttime predictions exhibited negligible bias centered around zero, while for daytime predictions, some low city-dependent biases emerged. The observed bias can be partially correlated with the city-specific bias of the downscaling model. The observed differences between daytime and nighttime performance can be interpreted in the context of urban surface energy balance processes. During daytime, LST variability is strongly driven by solar radiation, shadowing effects, and heterogeneous surface properties (e.g., vegetation, building materials), resulting in sharper spatial gradients that are more difficult to reconstruct and predict. In contrast, nighttime LST fields are governed by heat release from urban materials and reduced radiative forcing, leading to smoother spatial patterns and improved predictability. This behavior is consistent with the characteristic dynamics of the

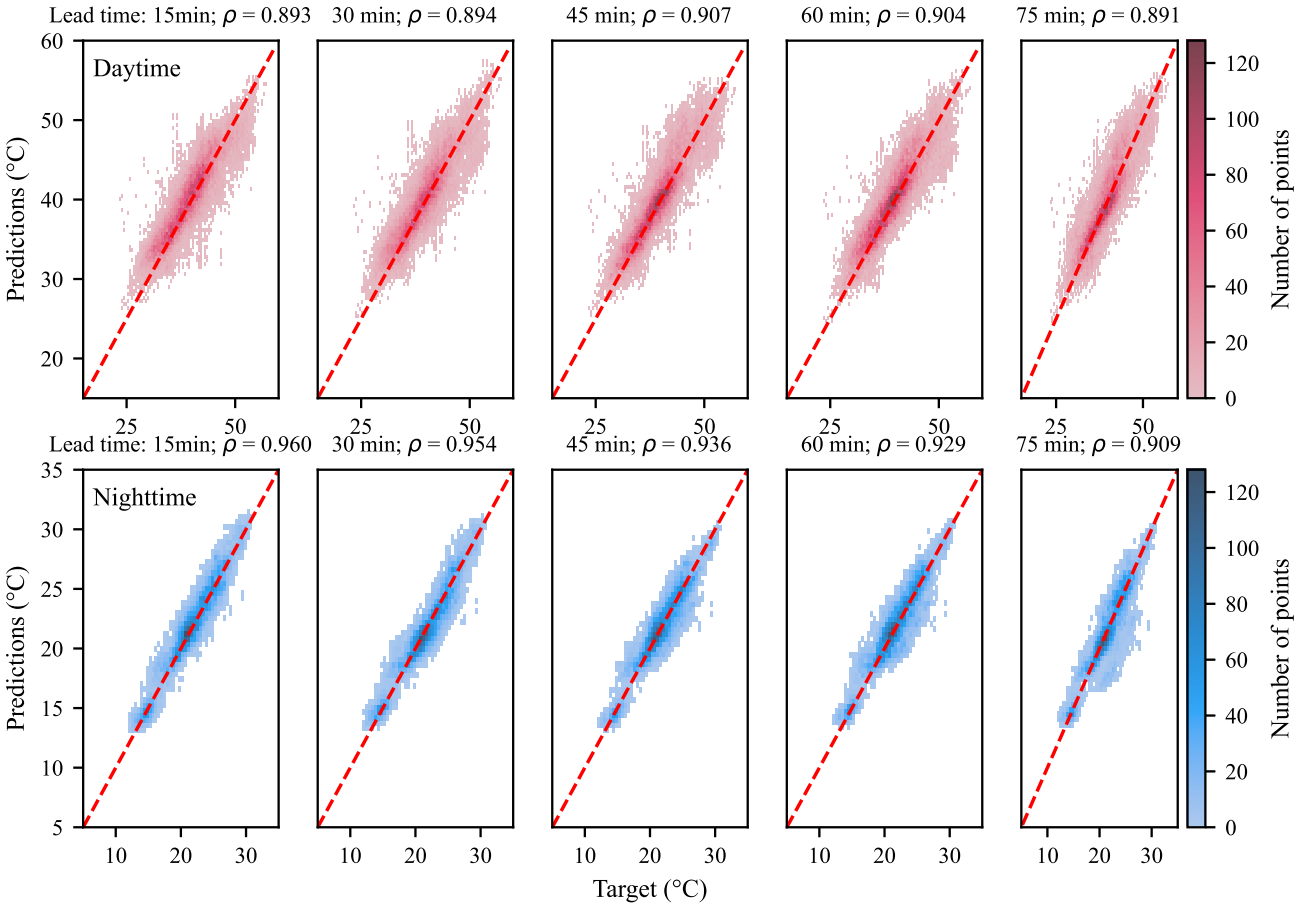

FIGURE 11. Nowcasting results vs MODIS measurement per lead time for the city of Bucharest. Top panel: LST estimates vs corresponding MODIS-derived LSTs from daytime overpasses of Terra and Aqua. Bottom panel: LST estimates vs corresponding MODIS-derived LSTs from nighttime overpasses of Terra and Aqua. ρ corresponds to the Pearson correlation coefficient.

surface urban heat island, which is typically more spatially coherent at night.

It is important to note, however, that the observed error values reflect the combined effect of both the LST downscaling and nowcasting models. Due to the sequential design of the framework, any spatial inaccuracies or biases introduced by the downscaling model are propagated into the temporal prediction stage. As a result, the reported performance does not exclusively quantify the predictive skill of the nowcasting model, but rather the joint behavior of the coupled spatio-temporal pipeline. Moreover, the evaluation is limited by the temporal sampling of MODIS observations, which are concentrated around the overpass times of Terra and Aqua. Consequently, the assessment does not fully capture model performance throughout the diurnal cycle, particularly during periods without high-resolution reference data.

The training and evaluation of the proposed framework are restricted to the period from mid-May to mid-September, corresponding to the climatologically warmest months of the year. This choice is motivated by our primary application, namely the characterization of surface urban heat islands, which are most pronounced during this period. Nevertheless, this temporal restriction limits the applicability of the model for year-round LST monitoring. In particular, the physical drivers of LST variability during colder seasons differ substantially from those in summer, with increased influence of factors such as reduced solar forcing, altered surface energy balance, and season-dependent land–atmosphere interactions. As a result, the relationships learned by the model during warm months may not generalize to winter conditions [29]. Extending the framework to a full annual cycle would likely require either season-specific models or the inclusion of additional predictors capturing landscape and environmental controls that become more dominant outside the summer period.

To further improve the presented approach, we propose to investigate the influence of the auxiliary data on the obtained results. Urban LST is known to be influenced by such factors as land cover composition, vegetation activity (e.g., NDVI), urban geometry (e.g., sky view factor), and anthropogenic factors (e.g. building density, population activity, and pollution levels). These variables affect surface

TABLE 4. The downscaling model U-Net architecture used. Convolutional layers are defined by kernel size, stride (*s*), and padding (*p*).

Stage	Operation	Channels	Spatial	Role
		Encoder		
1	Conv Block	3 → 64	128×128	Feature extraction
	MaxPool (2×2)	64 → 64	128→64	Downsampling
2	Conv Block	64 → 128	64×64	
	MaxPool (2×2)	128 → 128	64→32	Downsampling
3	Conv Block	128 → 256	32×32	
	MaxPool (2×2)	256 → 256	32→16	Downsampling
4	Conv Block	256 → 512	16×16	
	MaxPool (2×2)	512 → 512	16→8	Downsampling
		Bottleneck		
5	Conv Block	512 → 1024	8×8	Deep representation
		Decoder		
6	Transposed Conv (2×2, s=2)	1024 → 512	8→16	Upsampling
	Concatenate (skip)	512 + 512 → 1024	16×16	Skip connection
	Conv Block	1024 → 512	16×16	Feature fusion
7	Transposed Conv (2×2, s=2)	512 → 256	16→32	Upsampling
	Concatenate (skip)	256 + 256 → 512	32×32	Skip connection
	Conv Block	512 → 256	32×32	Feature fusion
8	Transposed Conv (2×2, s=2)	256 → 128	32→64	Upsampling
	Concatenate (skip)	128 + 128 → 256	64×64	Skip connection
	Conv Block	256 → 128	64×64	Feature fusion
9	Transposed Conv (2×2, s=2)	128 → 64	64→128	Upsampling
	Concatenate (skip)	64 + 64 → 128	128×128	Skip connection
	Conv Block	128 → 64	128×128	Feature fusion
10	Conv (1×1)	64 → 1	128×128	Output layer

TABLE 5. Architecture of the ConvLSTM-based nowcasting model. Convolutional layers are defined by kernel size, stride (*s*), and padding (*p*).

Stage	Operation	Channels	Spatial	Role
		Encoder		
Stage 1	Conv (3×3, s=1, p=1) + LeakyReLU	1 → 32	128×128	Feature extraction
	ConvLSTM (5×5)	32 → 64	128×128	Temporal modeling
Stage 2	Conv (3×3, s=2, p=1) + LeakyReLU	64 → 64	128→64	Downsampling
	ConvLSTM (5×5)	64 → 96	64×64	Temporal modeling
Stage 3	Conv (3×3, s=2, p=1) + LeakyReLU	96 → 96	64→32	Downsampling
	ConvLSTM (5×5)	96 → 128	32×32	Temporal modeling
		Decoder		
Stage 3	ConvLSTM (5×5)	128 → 128	32×32	Temporal modeling
	Transposed Conv (4×4, s=2, p=1) + LeakyReLU	128 → 128	32→64	Upsampling
Stage 2	ConvLSTM (5×5)	128 → 96	64×64	Temporal modeling
	Transposed Conv (4×4, s=2, p=1) + LeakyReLU	96 → 96	64→128	Upsampling
Stage 1	ConvLSTM (5×5)	96 → 64	128×128	Temporal modeling
	Conv (3×3, s=1, p=1) + LeakyReLU	64 → 32	128×128	Feature refinement
	Conv (1×1, s=1, p=0)	32 → 1	128×128	Output layer

energy balance components, including heat storage, radiative trapping, and anthropogenic heat release, and therefore contribute to the spatial variability of LST. By not explicitly incorporating these variables, we assume that their effects are encoded in the coarse-resolution LST signal and can be statistically inferred during downscaling. While this assumption enables a simplified and broadly applicable framework, it may limit the physical interpretability of the model and its ability to generalize across heterogeneous urban environments or extreme conditions. While potentially improving spatial detail, such variables may also introduce city-specific dependencies or diurnal biases (e.g., the NDVI–LST relationship is typically stronger during daytime than nighttime [30], [31]), which could reduce the robustness and transferability of the model across regions and times. For the downscaling model, the addition of the variables can be done as additional channels or in a multi-modal fashion [32]. For the nowcasting model, additional features will have to be split into stationary and dynamic, which, given the used ConvLSTM architecture, could add additional implementation complexity.

In conclusion, the proposed framework demonstrates that combining deep-learning-based downscaling with short-term nowcasting is a feasible and effective strategy for the diurnal monitoring of urban LST. The results indicate strong predictive performance, low systematic bias of the proposed models, and clear added value over the benchmark predictors. The proposed pipeline can support a broad range of applications, such as near-real-time urban heat island monitoring and forecasting, early warning for short-term heat stress episodes, improved city-scale heat-risk mapping, and targeted cooling interventions. It can also provide

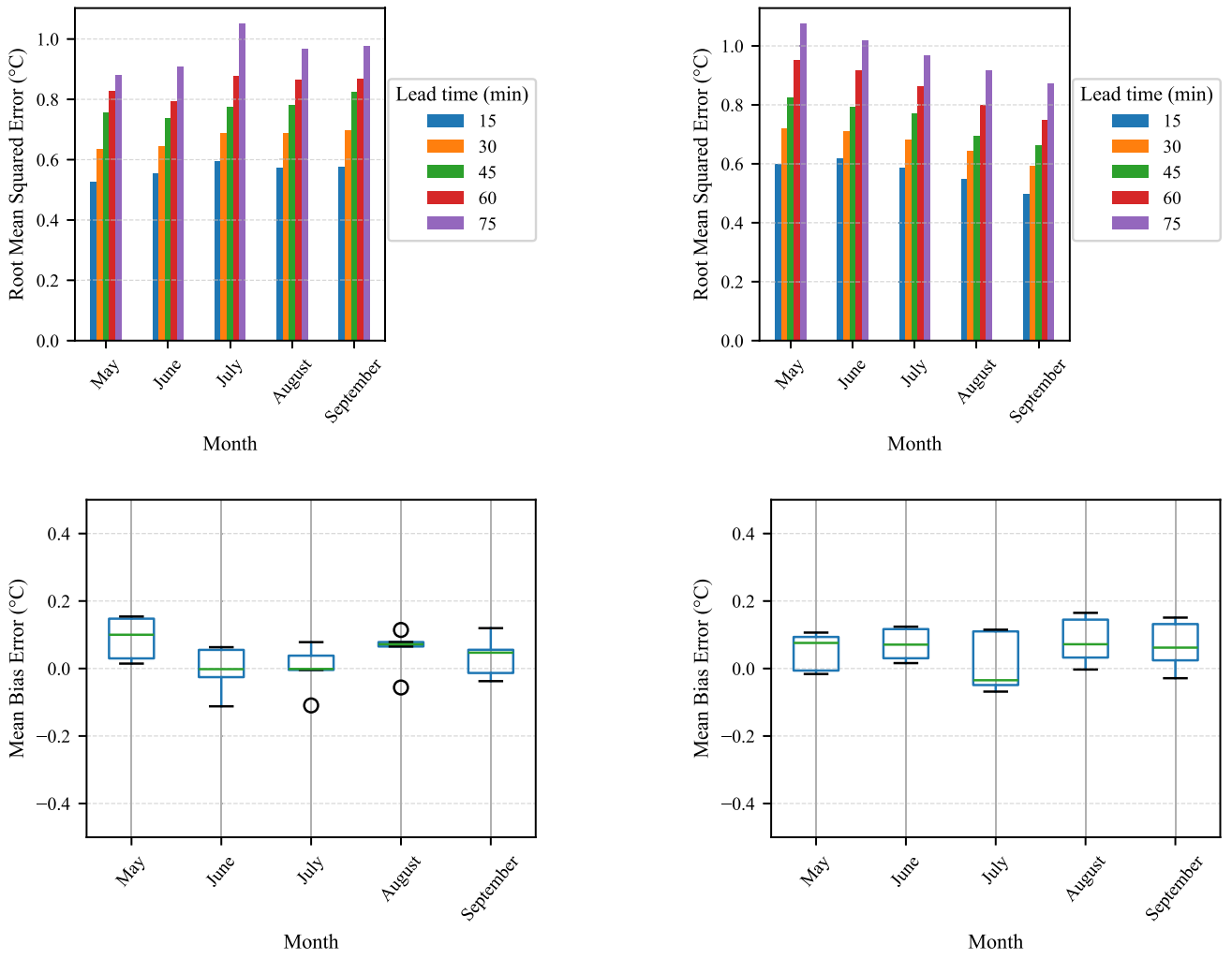

FIGURE 12. Nowcasting performance per studied month for Antwerp (left) and Berlin (right). Top row: RMSE per month per lead time. Bottom row: MBE distribution over lead times.

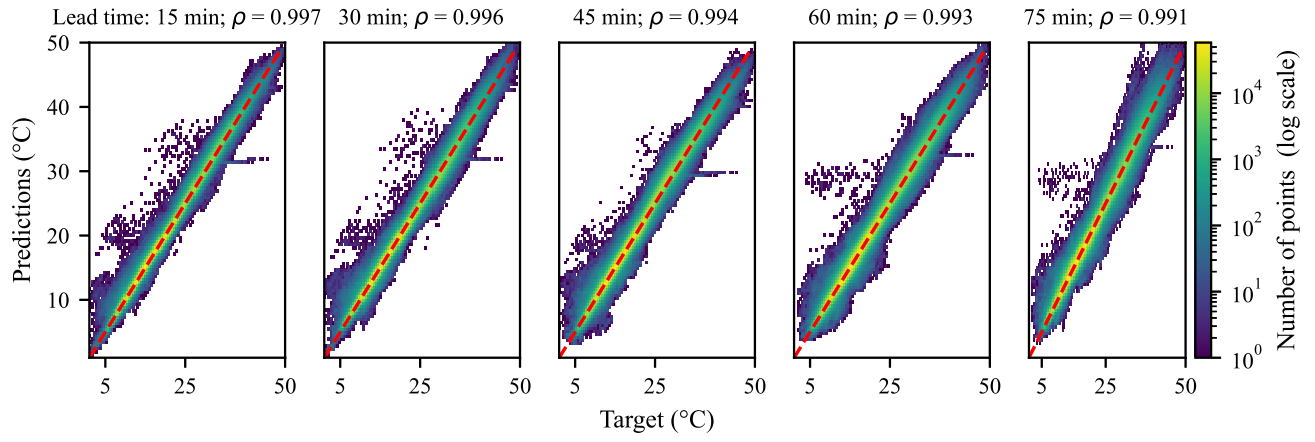

FIGURE 13. Predictions versus target values of the LST nowcasting model for different lead times for the city of Antwerp. The red line corresponds to 45-degree line. ρ corresponds to the Pearson correlation coefficient.

high-frequency thermal inputs for downstream models of energy demand or public-health risk assessment.

The source code for the experiments described in this article can be found in the github repository:

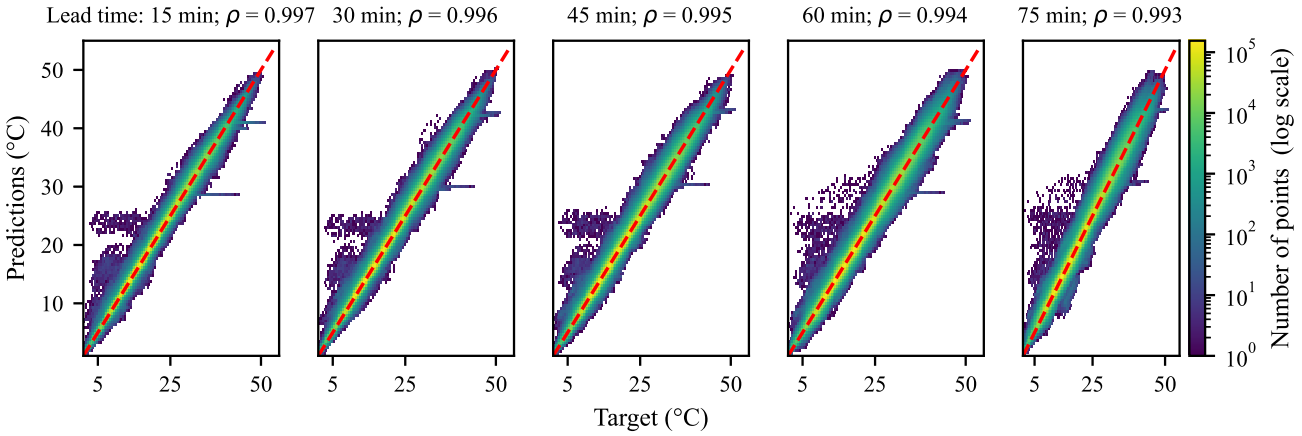

FIGURE 14. Predictions versus target values of the LST nowcasting model for different lead times for the city of Berlin. The red line corresponds to 45-degree line. ρ corresponds to the Pearson correlation coefficient.

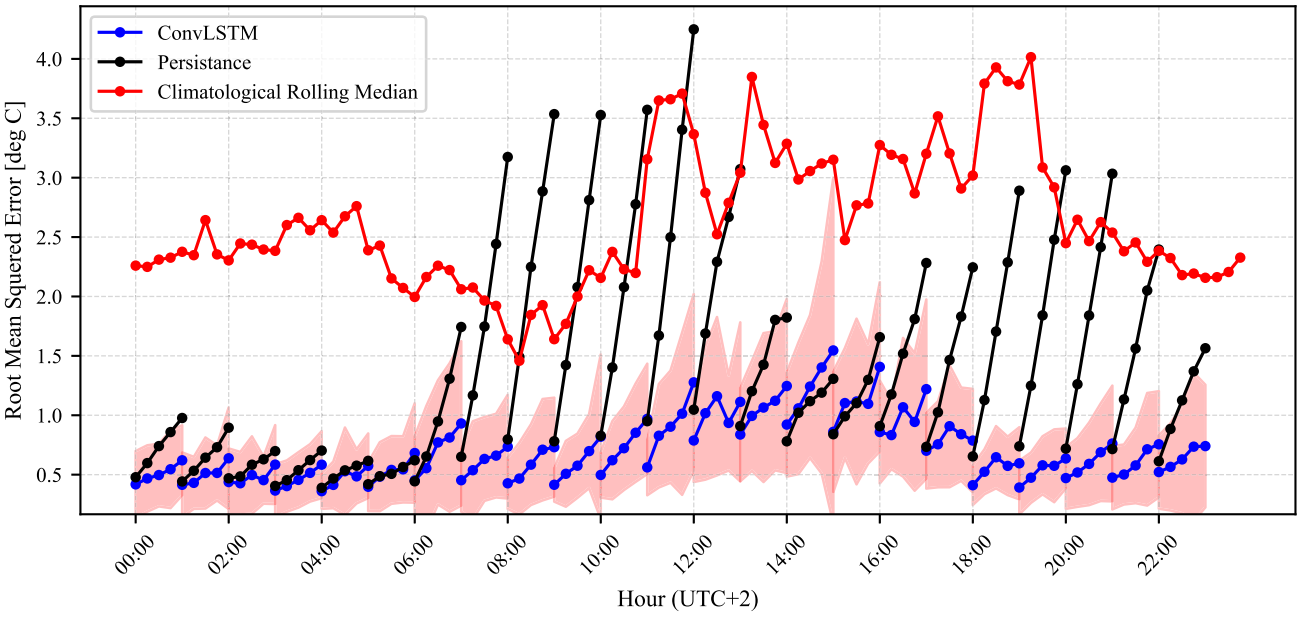

FIGURE 15. Diurnal evolution of the RMSE per lead time for the city of Antwerp: ConvLSTM-based LST nowcasting model, Persistence benchmark and Climatological Rolling Median benchmark (computed over two consecutive days due to the low data coverage). All times are in local summer time of Antwerp (UTC+2).

https://github.com/EnergyWeatherAI/LST_downscaling_and_nowcasting. We share the preprocessed machine learning-ready datasets used in this study via Zenodo repository [33].

APPENDIX A
DOWNSCALING MODEL SPECIFICATIONS

For the downscaling, we use a typical U-Net [21], composed of four downsampling (encoder) and upsampling (decoder) stages and skip connections between them. The details of the architecture used can be found in Table 4. For training, we use the Adam optimizer with a maximum of 1000 epochs and early stopping set at 10 epochs for the validation RMSE. The learning rate used is 2×10^{-5}, and the batch size is 32. The reproducibility seed is set at 0. Training time for a model run is approximately 3 hours on average using one NVIDIA A100 GPU node.

APPENDIX B
NOWCASTING MODEL SPECIFICATIONS

The nowcasting model architecture is a typical implementation of Convolutional LSTM Network, which is composed of an encoder and a decoder part [25], [26], [27], since the

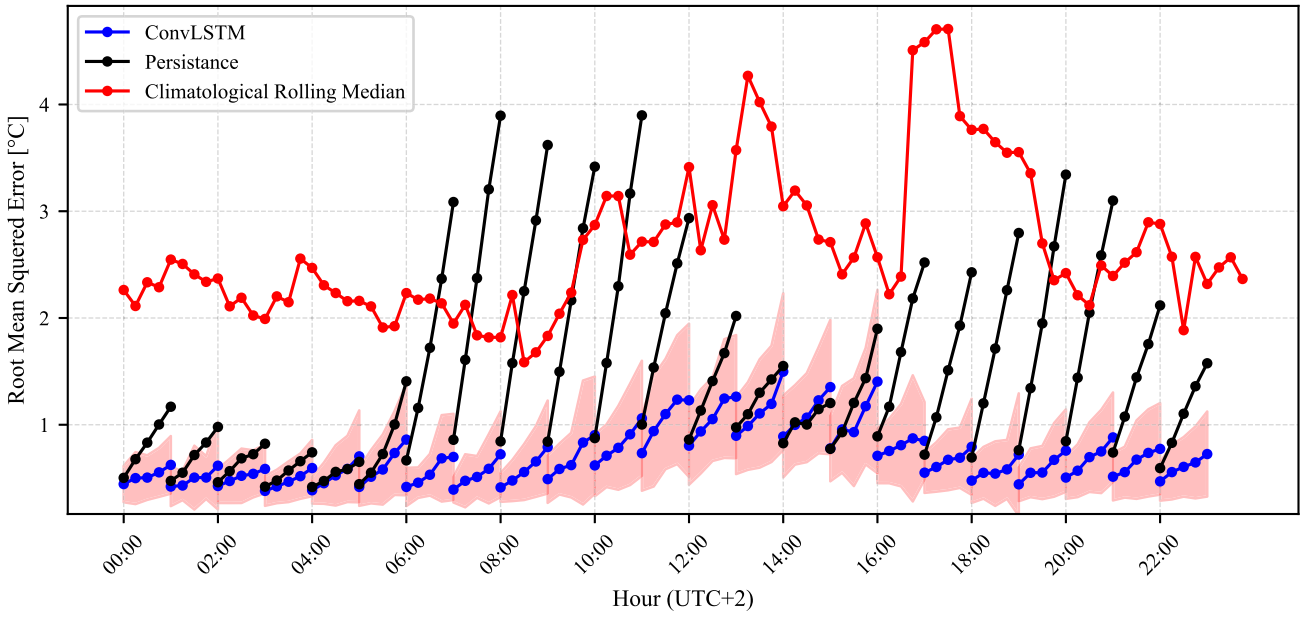

FIGURE 16. Diurnal evolution of the RMSE per lead time for the city of Berlin: ConvLSTM-based LST nowcasting model, Persistence benchmark and Climatological Rolling Median benchmark (computed over two consecutive days due to the low data coverage). All times are in local summer time of Berlin (UTC+2).

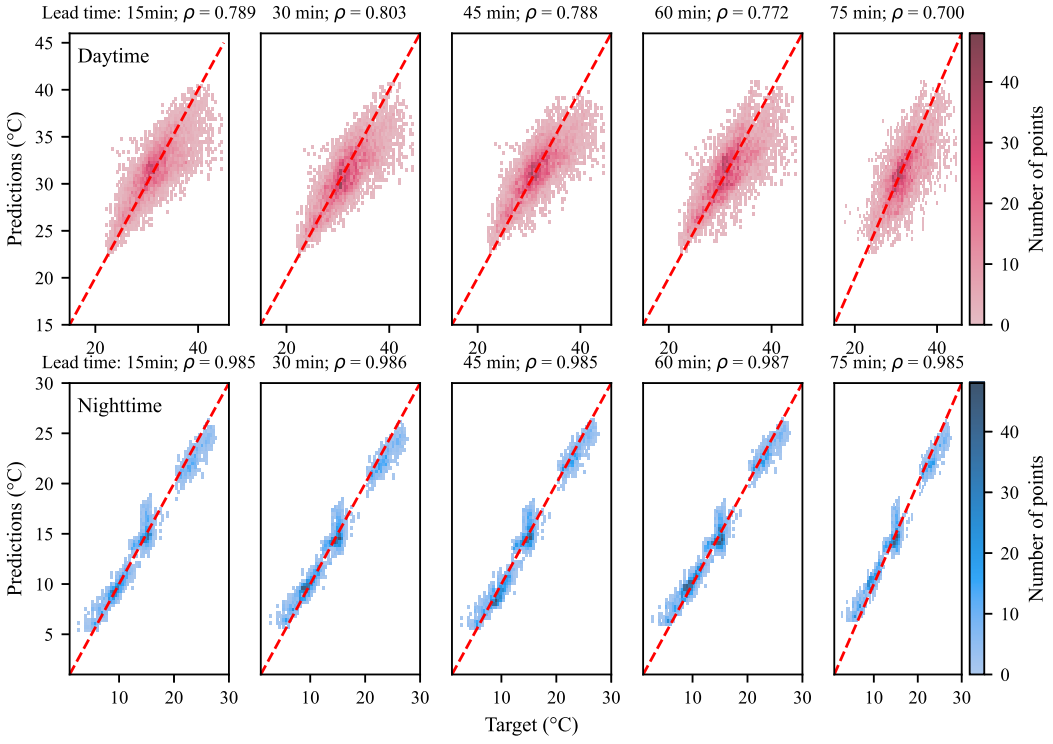

FIGURE 17. Nowcasting results vs MODIS measurement per lead time for the city of Antwerp. Top panel: LST estimates vs corresponding MODIS-derived LSTs from daytime overpasses of Terra and Aqua. Bottom panel: LST estimates vs corresponding MODIS-derived LSTs from nighttime overpasses of Terra and Aqua. ρ corresponds to the Pearson correlation coefficient.

desired output is an image. The encoder consists of three hierarchical stages, each combining:

- A convolutional block (for feature extraction and downsampling). The kernel size used is 3×3.

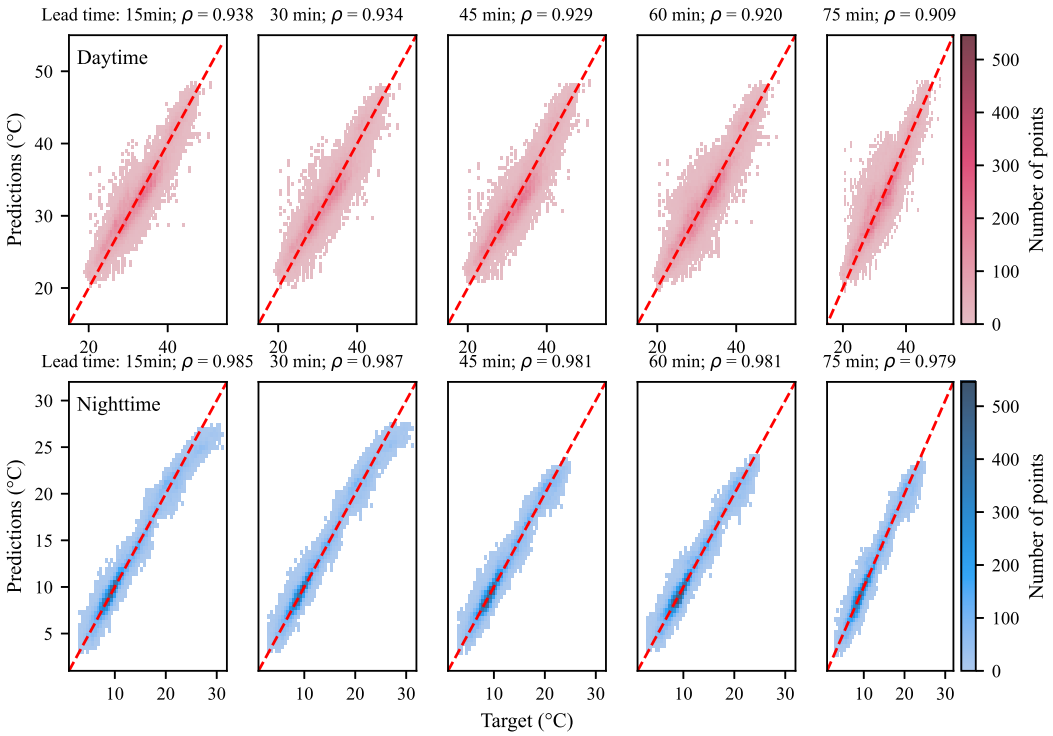

FIGURE 18. Nowcasting results vs MODIS measurement per lead time for the city of Berlin. Top panel: LST estimates vs corresponding MODIS-derived LSTs from daytime overpasses of Terra and Aqua. Bottom panel: LST estimates vs corresponding MODIS-derived LSTs from nighttime overpasses of Terra and Aqua. ρ corresponds to the Pearson correlation coefficient.

- A ConvLSTM layer (temporal modeling). The kernel size used is 5×5.

We use the LeakyReLU activation function with a slope set at the value 0.2.

The decoder mirrors the encoder in reverse order, combining:

- ConvLSTM layers (initialized with encoder outputs),
- Transposed convolutions for upsampling with a stride of 2, kernel size 4×4, and padding 1,
- Two final convolutional layers. The first convolutional layer uses a 3×3 kernel and the second a 1×1 to generate the final 128×128 prediction.

We also use the LeakyReLU activation function with a slope set at the value 0.2. The architecture is shown in Table 5. For training, we use the Adam optimizer with a learning rate of $1e-4$, a batch size of 128, a maximum of 1000 epochs, and early stopping set at 10 epochs for the validation RMSE. Training time for a model run is approximately 7 hours on average for Bucharest using one NVIDIA A100 GPU node. For Berlin and Antwerp, this is reduced to approximately 4 hours due to less data being available for those cities than for Bucharest.

APPENDIX C
RESULTS FOR ANTWERP AND BERLIN

To avoid the excessive number of figures in the main text, we present in this appendix the nowcasting results for Antwerp and Berlin. The conclusions drawn from the results of Bucharest presented in the main text also apply to the figures presented in this section.

REFERENCES

[1] E. Yeboah, I. Sarfo, C. Kwang, P. K. Alimo, M. A. Djan, M. S. Afenya, A. Okrah, and S. O. Y. Amankwah, "Influence of land use patterns on urban heat island dynamics in an emerging megacity: A case study of zhengzhou, henan, China," *J. Chin. Archit. Urbanism*, vol. 8, no. 1, p. 8412, May 2025.

[2] J. A. Voogt and T. R. Oke, "Thermal remote sensing of urban climates," *Remote Sens. Environ.*, vol. 86, no. 3, pp. 370–384, Aug. 2003.

[3] A. Karnieli, N. Agam, R. T. Pinker, M. Anderson, M. L. Imhoff, G. G. Gutman, N. Panov, and A. Goldberg, "Use of NDVI and Land Surface Temperature for drought assessment: Merits and limitations," *J. Climate*, vol. 23, no. 3, pp. 618–633, Feb. 2010.

[4] S. Mirasgedis, Y. Sarafidis, E. Georgopoulou, D. Lalas, M. Moschovits, F. Karagiannis, and D. Papakonstantinou, "Models for mid-term electricity demand forecasting incorporating weather influences," *Energy*, vol. 31, nos. 2–3, pp. 208–227, Feb. 2006.

[5] H. Shi, G. Xian, R. Auch, K. Gallo, and Q. Zhou, "Urban heat island and its regional impacts using remotely sensed thermal data—A review of recent developments and methodology," *Land*, vol. 10, no. 8, p. 867, Aug. 2021.

[6] I. D. Stewart, E. S. Krayenhoff, J. A. Voogt, J. A. Lachapelle, M. A. Allen, and A. M. Broadbent, "Time evolution of the surface urban heat island," *Earth's Future*, vol. 9, no. 10, Oct. 2021, Art. no. e2021EF002178.

[7] E. Parlow, "Regarding some pitfalls in urban heat island studies using remote sensing technology," *Remote Sens.*, vol. 13, no. 18, p. 3598, Sep. 2021.

[8] K. Zakšek and K. Oštir, "Downscaling Land Surface Temperature for urban heat island diurnal cycle analysis," *Remote Sens. Environ.*, vol. 117, pp. 114–124, Feb. 2012.

[9] B. Bechtel, K. Zakšek, and G. Hoshyaripour, "Downscaling Land Surface Temperature in an urban area: A case study for hamburg, Germany," *Remote Sens.*, vol. 4, no. 10, pp. 3184–3200, Oct. 2012.

[10] A. R. Bah, H. Norouzi, S. Prakash, R. Blake, R. Khanbilvardi, and C. Rosenzweig, "Spatial downscaling of GOES-R Land Surface Temperature over urban regions: A case study for New York city," *Atmosphere*, vol. 13, no. 2, p. 332, Feb. 2022.

[11] I. Keramitsoglou, C. T. Kiranoudis, and Q. Weng, "Downscaling geostationary Land Surface Temperature imagery for urban analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1253–1257, Sep. 2013.

[12] Q. Weng and P. Fu, "Modeling diurnal land temperature cycles over Los Angeles using downscaled GOES imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 97, pp. 78–88, Nov. 2014.

[13] A. Hurduc, S. L. Ermida, and C. C. DaCamara, "A multi-layer perceptron approach to downscaling geostationary Land Surface Temperature in urban areas," *Remote Sens.*, vol. 17, no. 1, p. 45, Dec. 2024.

[14] Y. Chang, Y. Cao, and Q. Weng, "Generating hourly 70-M Land Surface Temperature from GOES-R observations: A comparison of statistical downscaling and deep learning methods," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2024, pp. 918–920.

[15] S. Kartal and A. Sekertekin, "Prediction of MODIS Land Surface Temperature using new hybrid models based on spatial interpolation techniques and deep learning models," *Environ. Sci. Pollut. Res.*, vol. 29, no. 44, pp. 67115–67134, Sep. 2022.

[16] R. Brown, A. Domiter, and P. Vahabi, "Toward global multimodal high-resolution Land Surface Temperature forecasting and nowcasting," in *Proc. IEEE Int. Conf. Big Data (BigData)*, Dec. 2024, pp. 4292–4301.

[17] I. Trigo, S. Freitas, J. Bioucas-Dias, C. Barroso, I. Monteiro, and P. Viterbo, "Algorithm theoretical basis document for Land Surface Temperature (LST)," Instituto Portugues do Mar e da Atmosfera (IPMA), Lisbon, Portugal, Tech. Rep. 1.0, 2009.

[18] J. Schmetz, P. Pili, S. Tjemkes, D. Just, J. Kerkmann, S. Rota, and A. Ratier, "An introduction to Meteosat Second Generation (MSG)," *Bull. Amer. Meteorological Soc.*, vol. 83, no. 7, pp. 977–992, 2002.

[19] G. Hulley and S. Hook, "Modis/terra Land Surface Temperature/3-band emissivity 5-min l2 1km v061," NASA EOSDIS Land Processes Distributed Active Archive Center (DAAC), Sioux Falls, SD, USA, Tech. Rep. v061, 2021.

[20] L. Yang, F. Qian, D.-X. Song, and K.-J. Zheng, "Research on urban heat-island effect," *Proc. Eng.*, vol. 169, pp. 11–18, Dec. 2016.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[22] L. Jiao, L. Huo, C. Hu, and P. Tang, "Refined UNet: UNet-based refinement network for cloud and shadow precise segmentation," *Remote Sens.*, vol. 12, no. 12, p. 2001, Jun. 2020.

[23] P. Zhang, Y. Ke, Z. Zhang, M. Wang, P. Li, and S. Zhang, "Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery," *Sensors*, vol. 18, no. 11, p. 3717, Nov. 2018.

[24] W. Xu, G. Xu, Y. Wang, X. Sun, D. Lin, and Y. Wu, "High quality remote sensing image super-resolution using deep memory connected network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 8889–8892.

[25] X. Shi, Z. Chen, H. Wang, D. Yeung, W. K. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

[26] J. Kim, H. Choo, and J. Jeong, "Self-attention (SA)-ConvLSTM encoder–decoder structure-based video prediction for dynamic motion estimation," *Appl. Sci.*, vol. 14, no. 23, p. 11315, Dec. 2024.

[27] V. Vukotić, S.-L. Pintea, C. Raymond, G. Gravier, and J. V. Gemert, "One-step time-dependent future video frame prediction with a convolutional encoder–decoder neural network," in *Proc. Int. Conf. image Anal. Process.*, 2017, pp. 140–151.

[28] W. Köppen, "Die wärmezonen der erde, nach der dauer der heissen, gemässigten und kalten zeit und nach der wirkung der wärme auf die organische welt betrachtet," *Meteorologische Zeitschrift*, vol. 1, no. 21, pp. 5–226, 1884.

[29] Q. An, Y. Dong, W. Dong, and S. Xiao, "Spatiotemporal dynamics and nonlinear landscape-driven mechanisms of urban heat islands in a winter city: A case study of harbin, China," *Sustain. Cities Soc.*, vol. 133, Oct. 2025, Art. no. 106842.

[30] A. Ayanlade, "Seasonality in the daytime and night-time intensity of Land Surface Temperature in a tropical city area," *Sci. Total Environ.*, vols. 557–558, pp. 415–424, Jul. 2016.

[31] 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.

[32] D. Hong, L. Gao, N. Yokoya, J. Yao, J. Chanussot, Q. Du, and B. Zhang, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.

[33] S. Kurchaba and A. Meyer, "Spatiotemporal downscaling and nowcasting of urban land surface temperatures with deep neural networks," Tech. Rep., 2026, doi: 10.5281/zenodo.20168139.

SOLOMIIA KURCHABA received the B.Sc. degree in econophysics and the M.Sc. degree in theoretical physics from the University of Silesia in Katowice, Poland, and the Ph.D. degree in computer science from Leiden University, The Netherlands, where she worked on machine-learning-based solutions for NO_2 estimation from seagoing ships using TROPOMI/S5P data.

She was a Data Scientist with StorkJet, where she developed machine-learning-based solutions for aircraft performance optimization. She subsequently completed a Postdoctoral position with The Netherlands Organization for Space Research (SRON), focusing on the development of machine-learning-based methods for monitoring methane emission from super-emitters. She is currently a Postdoctoral Researcher with TU Delft and Bern University of Applied Sciences. Her research focuses on the development of machine-learning methods for monitoring land surface temperatures in urban areas. She is a member of the UrbanAIR Project Consortium.

ANGELA MEYER received the B.S. degree in physics from Imperial College London, the M.S. degree in mathematics from the University of Cambridge, U.K., and the Ph.D. degree in atmospheric physics from ETH Zürich, Switzerland. After Postdoctoral Researcher with the EU Project DNICast on short-term solar irradiance forecasting, she was a Research Scientist and a Project Leader in the fields of applied machine learning and data-driven condition monitoring with the Research and Development Centers, Hexagon AB and Siemens Smart Infrastructures. She is currently an Assistant Professor with BFH and TU Delft. Her research aims at developing intelligent decision support systems to increase the resilience and sustainability of industrial and energy systems with sensor-driven and machine learning approaches.