

Article

Kilometre-Scale Climate Projections for Nicosia: Dynamical and Machine Learning Downscaling to Detect Future Intra-Urban Heat Hotspots

Konstantina Koutroumanou-Kontosi ^{1,2} , Panos Hadjinicolaou ¹ , Constantinos Cartalis ² and Jos Lelieveld ^{1,3,*} 

¹ Climate and Atmosphere Research Center, The Cyprus Institute, 20 Konstantinou Kavafi Street, Nicosia 2121, Cyprus; k.koutroumanou@cyi.ac.cy (K.K.-K.); p.hadjinicolaou@cyi.ac.cy (P.H.)

² Department of Physics, National and Kapodistrian University of Athens, University Campus, 15784 Athens, Greece; ckartali@phys.uoa.gr

³ Max Planck Institute for Chemistry, 55128 Mainz, Germany

* Correspondence: jos.lelieveld@mpic.de

Abstract

This study investigates the potential of machine learning-based statistical downscaling to generate high-resolution (1 km) urban climate projections over Nicosia, Cyprus. A dynamical and a hybrid dynamical–statistical framework was applied to CMIP6 data under the SSP2-4.5 scenario to assess mid-twenty-first-century summer thermal conditions, focusing on daily maximum (T2max) and minimum (T2min) 2 m air temperature. The WRF model was used to dynamically downscale ERA5 reanalysis data to produce reference datasets for training three statistical models of increasing complexity: Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), and Convolutional Neural Networks (CNN). All models successfully reproduced the spatial temperature patterns simulated by WRF, with the CNN achieving the best performance and improved spatial consistency. Evaluation against observations confirmed that the statistical models capture daily maximum and minimum temperature variability with accuracy comparable to the dynamical model. Both the WRF model and the trained statistical models were subsequently applied to CMIP6-driven predictors to produce high-resolution projections. Results indicate pronounced summer warming over Nicosia by mid-century (1.9 °C for T2max and 1.1 °C for T2min), with strong spatial heterogeneity and intensified heat hotspots in compact urban areas (LCZ 3). These findings highlight the value of integrating dynamical and machine learning approaches for efficient urban-scale climate assessments.

Keywords: dynamical downscaling; statistical downscaling; WRF model; convolutional neural networks; CMIP6 projections; urban climate; urban heat hotspots



Academic Editors: Wen Wu, Jing Wang, Andong Guo, Jin Rui and Enxu Wang

Received: 24 May 2026

Revised: 6 July 2026

Accepted: 14 July 2026

Published: 21 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

As global warming is breaching the Paris Agreement threshold of 1.5 °C above pre-industrial levels [1], the intensity of 10-year extreme heat events is projected to increase to approximately 1.9 °C relative to 1850–1900 conditions [2]. Such intensification of heat extremes poses substantial risks to human health, contributing to heat-related illnesses, increased morbidity, and mortality. Vulnerable population groups, including the elderly, women, and outdoor workers, are disproportionately affected [3]. Beyond public health, extreme heat events also threaten agricultural productivity and critical infrastructure, including healthcare systems and energy supply networks.

The frequency and intensity of heatwaves vary considerably across regions, with the eastern Mediterranean and Middle East region identified as a prominent climate change hotspot [4,5], and recent events such as the 2024 Hajj heatwave illustrate the growing severity of heat extremes [6]. As an eastern Mediterranean island, Cyprus is especially exposed to these trends: observed records show statistically significant temperature increases of 0.4–0.6 °C per decade since 1981, while future projections under high-emission scenarios suggest annual warming exceeding 4 °C and precipitation reductions of 20–30% by the end of the century [7].

Urban areas are particularly susceptible due to high population density and the amplification of heat through the Urban Heat Island (UHI) effect, with consistent upward UHI trends, particularly during nighttime, documented across cities of the region, including Nicosia [8,9]. Consequently, reliable high-resolution urban climate projections are essential to support targeted mitigation and adaptation strategies. Such projections can help identify intra-urban hotspots where future heat stress is expected to be most intense, thereby supporting targeted urban planning and heat mitigation measures.

Future climate conditions are typically simulated using General Circulation Models (GCMs) under different emission and socioeconomic scenarios [10]. The IPCC Sixth Assessment Report (AR6) defines five Shared Socioeconomic Pathways (SSPs): SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5, corresponding to radiative forcing levels of 1.9, 2.6, 4.5, 7.0, and 8.5 W/m² by 2100 [2]. Although advancements in computational capacity have improved GCM spatial resolution [11], typical resolutions remain on the order of ~100 km, which is insufficient for assessing climate change impacts at the urban scale [12].

To bridge the gap between global and local scales, downscaling techniques are widely applied [13–15]. These approaches are broadly categorized into Dynamical Downscaling (DD) and Statistical Downscaling (SD). Dynamical downscaling employs Regional Climate Models (RCMs) driven by GCM boundary and lateral conditions to generate higher-resolution climate information [16,17]. While physically consistent and capable of explicitly resolving atmospheric processes and, thereby, facilitating interpretation, DD is computationally demanding. Multi-year simulations at kilometer-scale resolution, particularly across multiple climate scenarios and/or models, require substantial computational resources, limiting scalability or compromising detailed urban treatment [18,19]. Statistical and machine learning approaches have emerged as computationally efficient alternatives.

SD establishes statistical relationships between large-scale variables (predictors) and local-scale variables (predictands) [20,21]. These relationships are then applied to large-scale model outputs to derive local-scale projections [12,22,23]. Statistical methods include Multiple Linear Regression (MLR) [24,25], k-nearest neighbors (KNN) [26], and Support Vector Machines (SVM) [27–29]. Machine Learning (ML) techniques have also been utilized for downscaling such as Artificial Neural Networks (ANN) [25,27,30,31]. Furthermore, with advancements in data science and computational methods, deep learning architectures, including Convolutional Neural Networks (CNN), have been increasingly adopted. CNN are particularly suitable for climate applications because they can exploit the spatial structure of gridded datasets and learn complex nonlinear relationships between large-scale predictors and local-scale predictands [32–34]. Recent studies have particularly highlighted the potential of CNN-based architectures, including U-Net-type models and related convolutional frameworks, because they can process gridded climate predictors directly and automatically extract spatially coherent features without extensive manual predictor selection [35].

Compared to traditional statistical approaches, deep learning methods have demonstrated enhanced capability in capturing complex spatial patterns and nonlinear predictor–predictand relationships, often achieving performance comparable to standard ESD meth-

ods and, in some cases, to dynamical downscaling benchmarks [36,37]. These characteristics make deep learning approaches attractive for producing high-resolution climate information and for extending downscaling applications to larger domains, multiple variables, and multi-model projection ensembles.

Compared to DD, SD approaches are computationally efficient and scalable [38,39]. When trained on sufficiently large datasets, well-designed neural network architectures can achieve performance comparable to numerical models while requiring substantially lower computational cost. Nevertheless, important challenges remain. Since statistical downscaling models are commonly trained on historical observations or reanalysis-based datasets, they rely on the assumption that predictor–predictand relationships remain valid under future climate conditions.

Consequently, both dynamical and statistical downscaling approaches present distinct advantages and limitations that should be considered depending on the objectives of the application. To summarize the main characteristics of the different downscaling approaches discussed above, Table 1 compares their underlying principles, representative methods, major strengths, and limitations.

Table 1. Comparison of the main characteristics of Dynamical Downscaling (DD), traditional Statistical Downscaling (SD), and deep learning-based SD approaches.

Aspect	DD	Traditional SD	Deep Learning SD
Underlying principle	Uses numerical models to simulate atmospheric processes at finer scales	Learns predefined statistical relationships between predictors and predictands	Learns hierarchical nonlinear relationships directly from data
Representative methods	Regional climate models, mesoscale models	MLR, KNN, SVM, ANN	CNN, U-Net, GAN, Graph Neural Networks
Strengths	Physically consistent and explicitly resolves regional processes	Computationally efficient and interpretable	Automatic feature extraction, captures complex spatial patterns, highly scalable
Limitations	High computational cost and sensitivity to model configuration	Expert-driven feature selection, limited representation of complex nonlinear processes, stationarity assumption	Higher data requirements, stationarity assumption

In this study, the Weather Research and Forecasting (WRF) model was employed as a RCM to dynamically downscale CMIP6 data to 1 km resolution over the urban area of Nicosia city under the SSP2-4.5 scenario. In addition, a WRF simulation driven by ERA5 reanalysis was performed to generate the high-resolution datasets required for training the statistical downscaling models. Three SD approaches of increasing complexity were developed: MLR, ANN, and CNN to learn the relationships between 12 km and 1 km resolution fields. These relationships were subsequently applied to CMIP6-driven simulations for both a historical and a future period to derive high-resolution projections of daily T2max and T2min and to compute the corresponding climate change signal (ΔT).

The present study investigates whether ML-based statistical downscaling can provide a computationally efficient yet reliable alternative to dynamical downscaling for urban-scale climate projections at 1 km resolution. Specifically, the objectives of this study are to:

1. Evaluate the performance of MLR, ANN, and CNN models in reproducing high-resolution spatial temperature fields relative to dynamical simulations.

2. Quantify projected changes in daily T2max and T2min over Nicosia city between the historical and future periods at high-resolution.
3. Compare projected summer warming signals derived from dynamical and statistical approaches.
4. Identify future intra-urban heat hotspots within the Nicosia urban area based on the ensemble mean of the projections.

Recent studies have dynamically downscaled CMIP6 data over the Mediterranean region using RCMs, however, most of these simulations remain at spatial resolutions of several kilometers. For example, ref. [40] dynamically downscaled CMIP6 projections using WRF to 5 km over the Italy, while [41] produced simulations at 6 km resolution over Portugal. Although these studies significantly improve the spatial detail of climate projections, such resolutions remain insufficient to resolve intra-urban temperature variability and urban climate processes.

Only a limited number of studies have dynamically downscaled CMIP6 projections to the urban scale. For instance, ref. [42] performed WRF simulations at 1 km resolution for a metropolitan region in China. However, similar high-resolution dynamical downscaling studies remain extremely limited for Mediterranean cities.

The present work addresses this gap by producing high-resolution (1 km) projections of daily maximum and minimum air temperature for Nicosia, the largest city in Cyprus, which experiences some of the highest summer temperatures on the island due to its inland location. This enables the analysis of intra-urban warming patterns and the identification of future urban heat hotspots. Furthermore, by integrating dynamical downscaling with ML-based statistical approaches, this study provides a computationally efficient framework for generating urban-scale climate projections while retaining the physical consistency of RCMs and benefiting from the robustness of multi-model ensemble projections.

2. Materials and Methods

2.1. Study Area

The study area focuses on Nicosia, the capital and largest city of Cyprus, located in the central part of the island. Cyprus is the third largest island in the Mediterranean Sea, covering an area of approximately 9251 km². The current population of the island is about 1.38 million inhabitants, of which approximately 61% reside in urban areas, while the urban agglomeration of Nicosia hosts roughly 350,000 inhabitants. Cyprus is located in the Eastern Mediterranean region and is characterized by a hot-summer Mediterranean climate (Csa) according to the Köppen–Geiger climate classification [43]. The climate is characterized by very hot and dry summers and relatively mild winters, with mean annual precipitation of approximately 470 mm [44]. During the summer season, Nicosia experiences particularly high temperatures, with average daily T2max and T2min of approximately 35.8 °C and 22.6 °C respectively.

2.2. Data

2.2.1. ERA5 Reanalysis

The first WRF simulation was forced with initial and lateral boundary conditions derived from surface and pressure-level data of the fifth-generation reanalysis dataset (ERA5), provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5 provides hourly global reanalysis data from 1979 onwards at a horizontal resolution of 0.25°, assimilating numerical weather forecasts, meteorological station data and satellite remote sensing data [45]. Lateral boundary conditions were updated at 6 h intervals. The ERA5 data used in this study covered the period from 1 October 2007 to 1 January 2012. All data were

obtained from the Copernicus Climate Data Store (CDS, <https://cds.climate.copernicus.eu>, accessed on 20 May 2024).

2.2.2. CMIP6 Climate Projections

The second and third simulations were driven by outputs from the high-resolution version of the Max Planck Institute Earth System Model, MPI-ESM1.2-HR, obtained from the Coupled Model Intercomparison Project Phase 6 (CMIP6) archive. CMIP6 is an international framework coordinated by the World Climate Research Programme (WCRP) that brings together multiple GCMs to provide simulations of the past, present, and future climate conditions [46]. MPI-ESM1.2-HR is a fully coupled atmosphere–ocean–land global climate model, with an atmospheric horizontal resolution of approximately 100 km and an ocean resolution of approximately 40 km [47]. Six-hourly output data were used to provide the initial and lateral boundary conditions for the WRF simulations. The model was selected due to its relatively high spatial and temporal resolution within the CMIP6 ensemble and its demonstrated skill in providing reliable large-scale fields for regional climate modeling [48]. Furthermore, previous evaluations of CMIP6 models have shown that it performs particularly well in simulating climate variables over Europe and the Mediterranean region [49]. Also, the MPI-ESM1.2-HR exhibits an equilibrium climate sensitivity of approximately 3K [50], which is close to the CMIP6 multi-model median reported by the IPCC, making it suitable for climate impact assessment studies. Historical simulation was conducted for the period 2008–2012, while future projections covered 2048–2052 under the SSP2-4.5 scenario. The SSP2-4.5 scenario represents the “middle-of-the-road” socioeconomic pathway, characterized by moderate mitigation and adaptation challenges and it was selected as a plausible intermediate emissions trajectory for assessing mid-century regional climate changes. All CMIP6 data were obtained through the Earth System Grid Federation (ESGF) data portal, <https://metagrid.esgf-west.org/search>, accessed on 11 October 2025.

2.2.3. Meteorological Observations

To evaluate the performance of the models, in-situ meteorological observations provided by the Department of Meteorology of Cyprus were used. The analysis is based on data from an automated weather station located within the innermost modelling domain (d03) in central Nicosia (35.16° N, 33.36° E, 162 m above sea level). The station provides hourly measurements of 2 m air temperature and relative humidity, from which daily maximum and minimum temperatures were derived. Within the d03 domain, only two meteorological stations are available. However, only the Nicosia station is located within the urban core and is therefore representative of urban climatic conditions, while the second station (Athalassa) is situated in a suburban environment. Given the focus of this study on the urban thermal environment, the evaluation was performed using the Nicosia station to ensure consistency between the observations and the urban characteristics represented in the model. Records available from 1 January 2010 onwards. The station is situated in the urban core of Nicosia and is classified as urban. According to the Local Climate Zone (LCZ) classification, it belongs to LCZ 8 (Large low-rise), characterized by predominantly large, low-rise buildings with an average building height of approximately 8 m. For the purposes of this study, daily maximum and minimum temperatures were calculated from the hourly observations for the year 2012, considering only complete days (days with incomplete records were excluded from the analysis).

2.3. Dynamical Downscaling Framework

2.3.1. WRF Configuration

In this study the WRF model (version 4.5), was used to dynamically downscale the ERA5 reanalysis and CMIP6 climate model data. Figure 1 illustrates both the geographical

context of the study area and the WRF domain configuration. The left panel presents the Köppen–Geiger climate classification of the broader EMME region, showing that Cyprus is located within the hot-summer Mediterranean climate class (Csa), while the surrounding area encompasses mainly hot desert (BWh) and Csa climates. The Lambert conformal projection was applied with a reference latitude of 35.07° N and a reference longitude 33.49° E. The domain configuration (right panel of Figure 1) consists of one outermost domain covering the EMME region (d01, 13.75° E– 53.23° E, 25.0° N– 43.99° N) and two two-way nested domains over Cyprus (d02, 27.58° E– 39.40° E, 32.21° N– 37.8° N) and Nicosia (d03, 33.19° E– 33.48° E, 35.0° N– 35.21° N), with horizontal spatial resolutions of 12 km, 4 km, and 1 km, respectively. The number of grid points for d01, d02, and d03 was 271×167 , 262×153 , and 33×28 , respectively. In the vertical direction, the simulations comprised 35 terrain-following levels extending from the surface up to 50 hPa. Among the available urban canopy parameterization options, the Single-Layer Urban Canopy Model (SLUCM, [51]) was selected to better represent urban surface–atmosphere interactions. In addition, the Local Climate Zone (LCZ) classification scheme was used to provide a more detailed representation of the urban morphology of Nicosia. The innermost WRF domain (d03), which covers the Nicosia metropolitan area at 1 km resolution, is illustrated in Figure 2. Panel (a) shows the satellite imagery of the study area, while panel (b) presents the LCZ classification used in the simulations, displaying only the urban classes within the d03 domain. Lake surface temperature was prescribed using the `avg_tsfc.exe` utility provided within the WRF system.

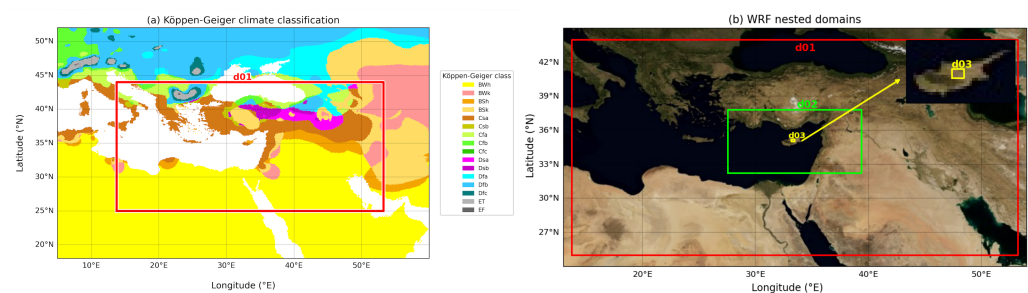


Figure 1. (a) Köppen–Geiger climate classification of the wider EMME region showing the location of the WRF outer domain (d01). (b) WRF domain configuration consisting of the outermost domain over the EMME region (d01), the nested domain over Cyprus (d02), and the innermost domain over Nicosia city (d03). Source: [52].

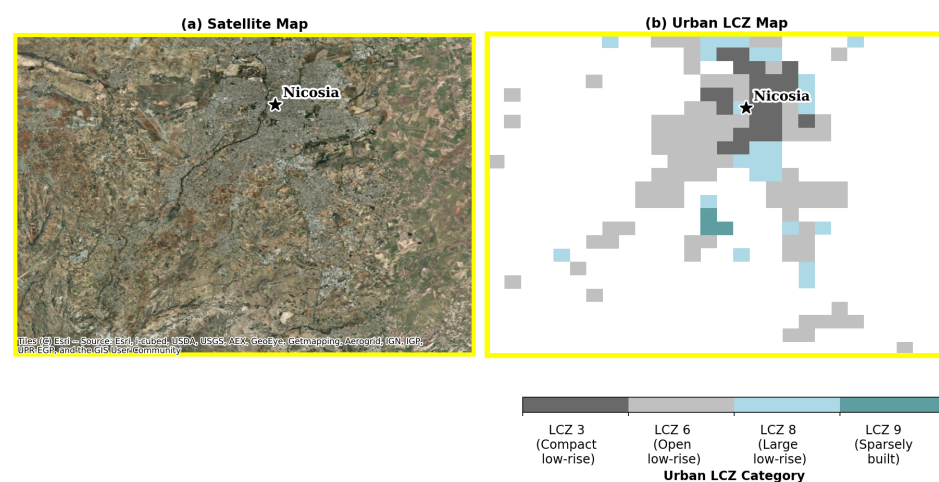


Figure 2. Study area of Nicosia showing (a) satellite imagery of the d03 domain and (b) the LCZ classification map including only the urban classes within d03. The location of the Nicosia meteorological station is indicated by the star.

A detailed description of the WRF configuration, including the physical parameterization schemes (Land Surface Model, Planetary Boundary Layer scheme, radiation and microphysics schemes) and the production of the LCZ classification map, is provided in [52], as the same model setup was adopted in the present study.

2.3.2. Experimental Design

The first simulation (WRF-ERA5) was forced with ERA5 reanalysis data and covered the period from 1 January 2008 to 31 December 2012. This five-year analysis period was selected as it was identified as representative of the 1991–2020 climatological reference period over Cyprus, while also coinciding with the availability of observational data at Nicosia station [52]. This simulation provided the reference high-resolution dataset required to construct the training databases for the statistical downscaling framework at 12 km and 1 km resolutions, where the machine learning models were developed. The second simulation (WRF-HIST) was driven by MPI-ESM1.2-HR initial and boundary conditions for the same historical period (2008–2012). This run served as the historical baseline for comparison with future projections. The third simulation (WRF-SSP2) was also driven by MPI-ESM1.2-HR for the future period 2048–2052 under the SSP2-4.5 scenario. This time slice was selected to represent mid-century climate conditions. The simulation enabled the assessment of projected changes over Nicosia at high spatial resolution and the computation of temperature differences between future and historical periods ($\Delta T = T_{\text{future}} - T_{\text{historical}}$).

All simulations were initialized on 1 October of the preceding year, with the first three months treated as spin-up and, therefore, the analysis was conducted from 1 January onward. The choice of spin-up duration depends on the modelling objectives and the total simulation length. For example, studies using the WRF model have shown that spin-up periods from weeks to several months may be required depending on the variables analyzed [53]. The WRF model was configured to produce hourly outputs, while the present analysis focused on daily maximum and minimum near-surface 2 m air temperature over the innermost domain (d03). The main characteristics of the WRF simulations performed in this study are summarized in Table 2.

Table 2. Summary of the WRF simulations performed in this study.

Simulation	Forcing Data	Period	Scenario	Purpose
WRF-ERA5	ERA5	2008–2012	Historical	Reference dataset for ML training
WRF-HIST	MPI-ESM1.2-HR	2008–2012	Historical	Historical GCM-driven simulation
WRF-SSP2	MPI-ESM1.2-HR	2048–2052	SSP2-4.5	Future projection (mid-century)

2.4. Statistical Downscaling Framework

2.4.1. Perfect Prognosis Approach

This study adopts the Perfect-Prognosis (PP) approach, in which statistical relationships between large-scale predictors and local-scale predictands are extracted using observational (or reanalysis) data and subsequently applied to GCM outputs to generate future local-scale projections [23,54–56]. The PP approach relies on predictor and predictand datasets with day-to-day correspondence, ensuring that the derived statistical relationships can be applied to any GCM outputs to generate future projections under different climate scenarios.

In the present study, both predictor and predictand datasets were derived from the dynamical downscaling of ERA5 reanalysis data using the WRF model (Simulation 1). The large-scale predictors were extracted from the outermost domain (d01) at 12 km spatial resolution, while the predictands were obtained from the innermost domain (d03) at 1 km resolution for the period 2008–2012.

The predictor set includes five near-surface thermodynamic variables: 2 m air temperature, 2 m relative humidity, 10 m zonal wind, 10 m meridional wind, and surface pressure. In addition, upper-air variables were considered, including temperature, relative humidity, zonal wind, meridional wind, and geopotential height at 500 hPa, as well as temperature at 850 hPa. These variables are commonly employed in temperature downscaling studies [57] as they capture large-scale atmospheric circulation and thermodynamic conditions.

From the hourly WRF outputs, daily maximum (T2max) and daily minimum (T2min) 2 m air temperatures were computed and used as predictands. Daily mean values of the predictor variables were calculated and used as inputs to the statistical downscaling models. A summary of the predictor variables is provided in Table 3.

Table 3. Summary of predictor variables used in this study.

Name	Code	Level	Time	Unit
Temperature	T	2 m, 500 hPa, 850 hPa	Daily mean	K
Relative Humidity	RH	2 m, 500 hPa	Daily mean	%
Zonal Wind	U	10 m, 500 hPa	Daily mean	m s ^{−1}
Meridional Wind	V	10 m, 500 hPa	Daily mean	m s ^{−1}
Surface Pressure	PSFC	Model Surface	Daily mean	Pa
Geopotential Height	GHT	500 hPa	Daily mean	m ² s ^{−2}

2.4.2. Machine Learning Models

1. MLR

Using the aforementioned predictor variables, MLR was implemented to establish linear relationships between large-scale atmospheric independent variables (predictors) and local-scale dependent variables (predictands, T2max and T2min). MLR models the predictand as a linear combination of multiple independent variables, aiming to explain the maximum variance of the dependent variable while minimizing model complexity.

The WRF-ERA5 dataset (2008–2012) was divided into training and testing subsets using a chronological split rather than a random partition. The first four years (2008–2011) were used for training, while the final year (2012) was reserved for testing. This approach preserves the temporal structure of the time series, avoids artificial mixing of seasons, and allows for evaluation over a complete independent year. Furthermore, 2012 coincides with the availability of observational data from the Nicosia station, enabling direct comparison between model outputs and measurements.

Prior to model training, predictor variables were standardized using z-score normalization ($\mu = 0$, $\sigma = 1$). The scaling parameters were computed exclusively from the training dataset and subsequently applied to both training and testing sets.

Given the high dimensionality of the predictor space, Principal Component Analysis (PCA) was applied to reduce multicollinearity among atmospheric predictors [58]. This is particularly important in linear regression frameworks to avoid overfitting and enhance numerical stability. PCA was fitted exclusively on the standardized training dataset, and the leading principal components explaining 95% of the total variance were retained. The resulting transformation was subsequently applied to both the training and testing datasets.

The MLR model was fitted using ordinary least squares regression on the transformed training dataset. Predictions were generated for the independent test period. Model outputs were compared against WRF-derived predictands and against Nicosia station observations.

2. ANNs

ANNs are models inspired by the structure and functioning of the human brain [59]. An ANN consists of interconnected processing units (neurons) organized in layers, where each connection is associated with a weight. These networks are particularly suitable for modeling complex and nonlinear relationships between input predictors and output variables, which has led to their increasing application in climate modeling and downscaling studies.

In a typical feedforward neural network architecture, neurons are arranged in an input layer, one or more hidden layers, and an output layer. Each layer consists of one or more neurons that are fully connected to neurons in the subsequent layer, while no connections exist between neurons within the same layer. Each neuron computes a weighted sum of its inputs, adds a bias term, and passes the result through a nonlinear activation (transfer) function, such as the sigmoid or Rectified Linear Unit (ReLU), to produce its output. During supervised training, the network parameters (weights and biases) are optimized by minimizing a predefined loss function that quantifies the difference between the predicted and the targeted values. This optimization is commonly performed using gradient-based algorithms through backpropagation, enabling the network to iteratively update its parameters and improve predictive performance.

A fully connected feedforward ANN was developed and was implemented using Keras within the TensorFlow framework. The WRF-ERA5 dataset (2008–2012) was split chronologically into training (2008–2010), validation (2011), and testing (2012) subsets. Predictor variables were standardized using z-score normalization, with scaling parameters estimated from the training set and applied consistently to the validation and test sets. The predictands were also standardized using a separate scaler fitted on the training targets to facilitate stable optimization and to ensure comparable weighting across grid cells.

The ANN architecture consisted of two fully connected hidden layers with 128 and 64 neurons, respectively, using the ReLU activation function. The output layer was a dense linear layer with dimensionality equal to the number of 1 km grid cells, corresponding to a multi-output regression framework. Model training was performed by minimizing the mean squared error (MSE) loss using the Adam optimizer. Training was conducted for up to 50 epochs with a batch size of 128, and early stopping was applied based on validation loss to reduce overfitting.

3. CNNs

A CNN was implemented to explicitly account for the spatial structure of predictors and predictand variables fields. Unlike fully connected ANN architectures, where neurons in successive layers are fully connected, CNNs rely on local receptive fields and shared convolutional kernels. In spatially structured data, localized connections are essential for capturing neighborhood-scale patterns that progressively contribute to larger-scale structures, whereas full connectivity becomes unnecessary. This architectural design allows CNNs to extract hierarchical spatial features while at the same time reducing the number of trainable parameters.

The dataset was split chronologically in the same manner as the ANN model (training: 2008–2010, validation: 2011, testing: 2012), and standardization was performed identically using training-set statistics to ensure methodological consistency. A key advantage of the CNN framework is its ability to operate directly on high-dimensional spatial tensors, eliminating the need for explicit feature extraction techniques such as PCA, which are typically required in linear models (e.g., MLR). Through its convolutional layers, the CNN internally learns the relative contribution and hierarchical

interactions of predictors, allowing feature extraction to be performed automatically during training.

The adopted architecture followed a U-Net-type fully convolutional encoder–decoder framework. The encoder consisted of three convolutional blocks with 32, 64, and 128 filters, respectively, each followed by max-pooling layers for spatial downsampling. A bottleneck layer with 256 filters captured high-level abstract features, followed by dropout for regularization. The decoder progressively restored spatial resolution through upsampling layers and skip connections that concatenated encoder feature maps to preserve fine-scale information. A final 1×1 convolutional layer produced the 1 km temperature field. L2 weight regularization and batch normalization were applied within convolutional blocks to avoid overfitting and improve the performance of the model to unseen data.

Model training was performed using backpropagation to minimize a composite loss function combining MAE and mean squared error (MSE), with a small total variation penalty to encourage spatial smoothness. The Adam optimizer [60] was employed with an adaptive learning rate [61], and training was conducted for up to 200 epochs using a batch size of 16. Early stopping and learning-rate reduction were applied to terminate the training process if only the performance of the training set is improved and not that of the validation set. The final architecture was selected after extensive experimentation with different network depths, filter sizes, and regularization configurations, with model performance systematically evaluated against the 1 km WRF reference dataset.

2.5. Identification of Future Urban Heat Hotspots

After evaluation, the statistical relationships learned by the MLR, ANN, and CNN models from the WRF-ERA5 dataset were applied, within the PP framework, to the large-scale (12 km) predictor variables of the WRF-HIST and WRF-SSP2 simulations. This process enabled the production of high-resolution (1 km) temperature fields for both historical and future climate periods using each statistical downscaling method. Consequently, both the dynamically (WRF) and statistically (MLR, ANN, CNN) downscaled projections were available on a common spatial resolution, facilitating direct comparison and assessment of future urban temperature patterns.

To identify urban heat hotspots within the study area, the high-resolution projections (1 km resolution) produced by all models (WRF, MLR, ANN, and CNN) were combined to form an ensemble estimate of near-surface air temperature. Seasonal averages were first computed from daily values for both maximum (T2max) and minimum (T2min) temperatures. The analysis focused on the summer season (JJA), when heat stress conditions in Nicosia are most pronounced.

For each predictand variable, the ensemble mean of the temperature field was calculated for both the historical and future periods. A threshold-based approach was then applied to identify the warmest grid cells. Specifically, the 95th percentile of the historical summer temperature distribution over urban grid cells was computed and used as a reference threshold. Grid cells exceeding this threshold were identified for both the historical and future projections.

To highlight areas consistently exposed to elevated thermal conditions, grid cells exceeding the historical 95th percentile during the future period for both T2max and T2min were considered as potential intra-urban heat hotspots. Considering both daytime and nighttime temperature extremes ensures that the identified hotspots represent areas exposed to persistent heat stress conditions. The intersection of these exceedance regions was used to identify locations experiencing both extreme daytime and nighttime temperatures.

These cells were ranked according to the combined magnitude of daytime and nighttime temperatures, defined as the mean of T2max and T2min values. The twenty highest-ranked grids were selected to represent the most intense intra-urban heat hotspots.

2.6. Performance Evaluation Metrics

Daily maximum and minimum 2 m air temperatures were derived from the hourly model outputs and compared with the corresponding observational data. WRF and ML models' performance was evaluated using several statistical metrics, including the Mean Bias Error (MBE, Equation (1)) where S_i and O_i are the simulated and observed values, respectively, and n is the number of data points, the Mean Absolute Error (MAE, Equation (2)), the Root Mean Square Error (RMSE, Equation (3)) and the Correlation Coefficient (r , Equation (4)) where $\bar{S}$ and $\bar{O}$ are the means of the simulated and observed values, respectively. The MBE quantifies the systematic bias of the model simulations relative to the observations. However, it may underestimate overall error due to the compensation of positive and negative deviations. The MAE represents the average magnitude of the errors, whereas the RMSE assigns greater weight to larger deviations and is therefore more sensitive to outliers, reflecting the variability of the error distribution [62].

$$\text{MBE} = \frac{\sum_{i=1}^n (S_i - O_i)}{n}, \quad (1)$$

$$\text{MAE} = \frac{\sum_{i=1}^n |S_i - O_i|}{n}, \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (S_i - O_i)^2}{n}}, \quad (3)$$

$$r = \frac{\sum_{i=1}^n (S_i - \bar{S})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^n (S_i - \bar{S})^2 \sum_{i=1}^n (O_i - \bar{O})^2}}, \quad (4)$$

3. Results

The capability of the WRF model to dynamically downscale near-surface air temperature (T2max and T2min) over Nicosia driven by ERA5 reanalysis data has been previously demonstrated in detail by [52]. Therefore, the present study focuses on evaluating the ability of the machine learning approaches to reproduce the high-resolution WRF outputs.

3.1. Performance of the ML Models to Reproduce WRF Fields

3.1.1. Annual

The statistical relationships derived for each machine learning model (MLR, ANN, and CNN) between the low-resolution (12 km) and high-resolution (1 km) datasets based on WRF-ERA5 simulation were applied to the independent test period (2012). Specifically, the low-resolution predictor variables were provided as inputs to the trained statistical models in order to generate the high-resolution temperature fields.

Firstly, the ability of the ML models to reproduce the WRF simulations was evaluated by calculating the annual MBE and MAE for both T2max and T2min. Figure 3 summarizes the spatial distributions of these performance metrics for the three machine learning approaches.

As shown in the first row of Figure 3, all methods reproduce the WRF results satisfactorily, with relatively small biases across the urban domain. Overall, all models exhibit a slight overestimation of T2max. The ANN model presents the largest positive bias (approximately 0.13 °C), while the CNN shows the smallest deviations from the WRF simulations (around 0.08 °C). The MLR model exhibits exclusively positive biases across the domain (around 0.1 °C), whereas the ANN and CNN models display a limited number of grid cells with small negative biases. Furthermore, a comparison of the three models reveals clear

differences in the spatial distribution of the biases. The MLR exhibits relatively smooth and spatially homogeneous biases, whereas the ANN and CNN show greater spatial variability between adjacent grid cells. This variability is particularly pronounced for the ANN, which contains several isolated grid cells with comparatively larger positive biases. In contrast, the CNN exhibits a more homogeneous spatial pattern than the ANN, although it remains less spatially uniform than the MLR.

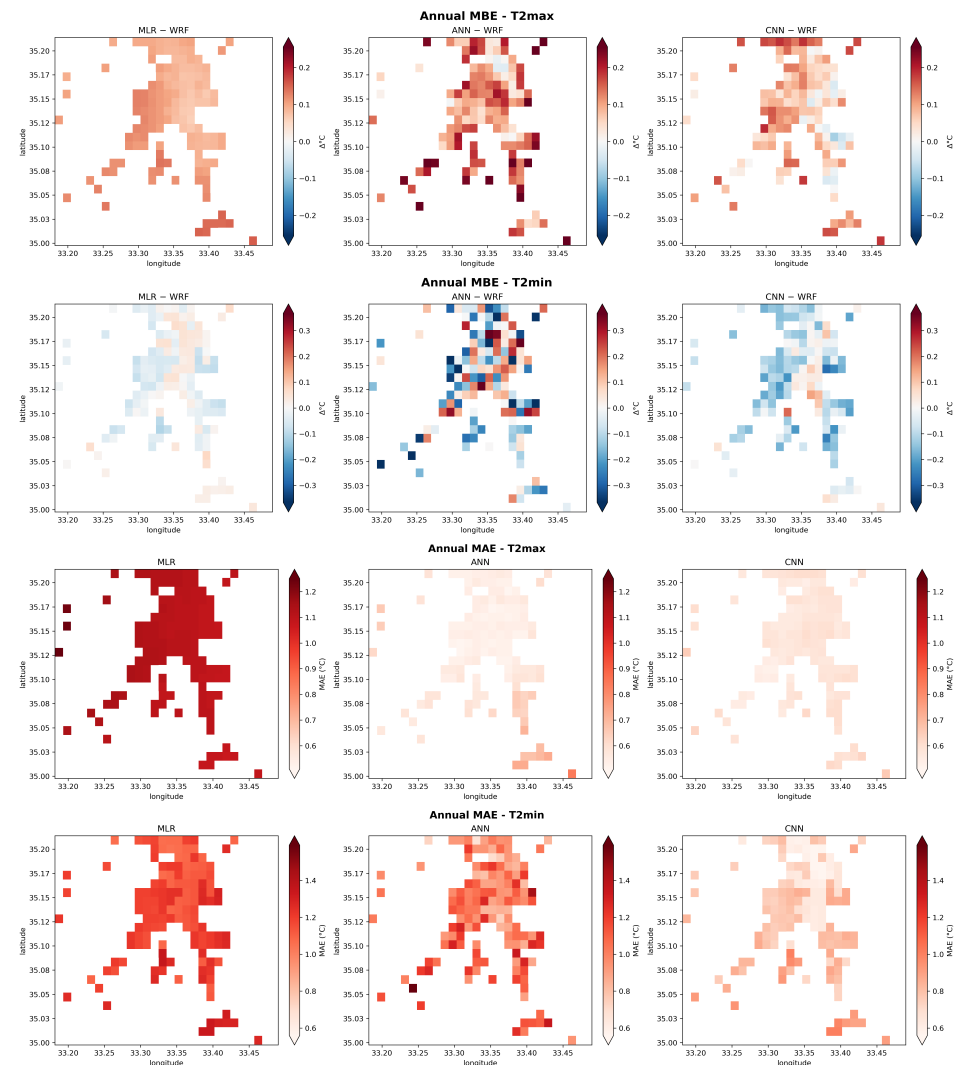


Figure 3. Annual mean bias error (MBE, 1st and 2nd rows) and mean absolute error (MAE, 3rd and 4th rows) of T2max and T2min between each machine learning model (MLR, ANN and CNN) and the WRF simulations for the test period (2012).

For T2min (2nd row of Figure 3), the spatial patterns differ slightly. Both the MLR and CNN models tend to underestimate T2min across most of the domain, while small overestimations are observed mainly in the eastern part of the urban area. In contrast, the ANN model does not exhibit a clear spatial bias pattern, with alternating positive and negative biases occurring between adjacent grid cells ranging from approximately $-0.3\text{ }^{\circ}\text{C}$ to $+0.3\text{ }^{\circ}\text{C}$. Overall, the MLR approach demonstrates the best performance for T2min, with biases generally close to zero and deviations typically within approximately $0.1\text{ }^{\circ}\text{C}$, whereas the CNNs exhibit biases around $-0.2\text{ }^{\circ}\text{C}$ across much of the domain. The spatial heterogeneity of the errors follows a pattern similar to that observed for T2max, with the MLR exhibiting the smoothest and most homogeneous biases and the ANNs displaying the greatest spatial variability.

While the MBE provides information on the systematic overestimation or underestimation of the models, it does not fully capture the overall model performance because positive and negative errors may cancel each other out. Therefore, the spatial distribution of the MAE was also calculated to quantify the overall magnitude of the deviations between the ML models and the WRF simulations for both T2max and T2min (3rd–4th rows of Figure 3).

For T2max (3rd row of Figure 3), the ANN and CNN models exhibit substantially better performance compared to the MLR model. Both ANN and CNN present average MAE values of approximately 0.6 °C, whereas the MLR model shows considerably larger errors, with an average MAE close to 1.1 °C. The spatial distribution of the MAE for all methods remains relatively uniform across the domain, indicating that the models reproduce the spatial structure of the WRF T2max field consistently across the study area.

For T2min (4th row of Figure 3), the CNN model provides the best overall performance with an average MAE of approximately 0.8 °C. The ANN model follows with an average MAE of around 1 °C, while the MLR model presents the largest deviations, reaching approximately 1.2 °C. In contrast to T2max, slightly smaller MAE values are observed in the urban core for T2min, suggesting that the models are able to reproduce nighttime temperature patterns more accurately within the central urban area. In particular, the MAE over the central urban area is approximately 1.0 °C for the MLR, 0.8 °C for the ANN, and 0.6 °C for the CNN.

3.1.2. Summer

The annual evaluation provides an overall assessment of the model performance. However, since the primary objective of this study is to focus on the capability of the ML models to reproduce temperature patterns during periods of enhanced heat stress, the analysis was further extended to the summer season (JJA). The summer period is of particular importance for the Eastern Mediterranean region, as it is associated with pronounced urban heat stress conditions. Therefore, summer MBE and MAE maps for both T2max T2min variables were also examined for the test period in order to evaluate the ability of the ML models to reproduce the spatial structure of the WRF simulations under these thermally demanding conditions (Figure 4).

In contrast to the annual results, the summer biases exhibit more pronounced differences among the models. As shown in the 1st row of Figure 4, the CNN shows a relatively uniform overestimation of approximately 0.2 °C across most of the domain. The ANN also tends to overestimate T2max, however, its spatial bias pattern is considerably more heterogeneous, with alternating positive and negative values between neighboring grid cells ranging from −0.3 °C to 0.3 °C. In the case of the MLR model, positive biases of about +0.1 °C are mainly observed in the southern part of the domain reaching up to +0.2 °C in the southernmost grid cells, while slight underestimations (around −0.1 °C) occur in the northern areas. In contrast to the ANN, where neighboring grid cells often exhibit alternating positive and negative biases, the MLR displays a smooth north-to-south transition from slight underestimation to slight overestimation, indicating a more spatially coherent bias pattern.

For T2min variable (2nd row of Figure 4), all models generally underestimate the nighttime temperatures. The CNN demonstrates the best performance, with biases around −0.2 °C, whereas the ANN presents the largest underestimation, averaging approximately −0.5 °C, while the MLR lies in between with close to −0.3 °C. The CNN also exhibits a small number of positive biases (around 0.25 °C) in the central–eastern part of the domain, whereas the corresponding area in the MLR is characterized by nearly zero biases. Compared with the MLR and CNN, the ANN displays a much more heterogeneous spatial pattern, with several localized grid cells where the underestimation approaches −1 °C.

Overall, the magnitude of the biases is larger for T2min compared to T2max during the summer season, indicating that reproducing nighttime temperature conditions remains more challenging for the ML models.

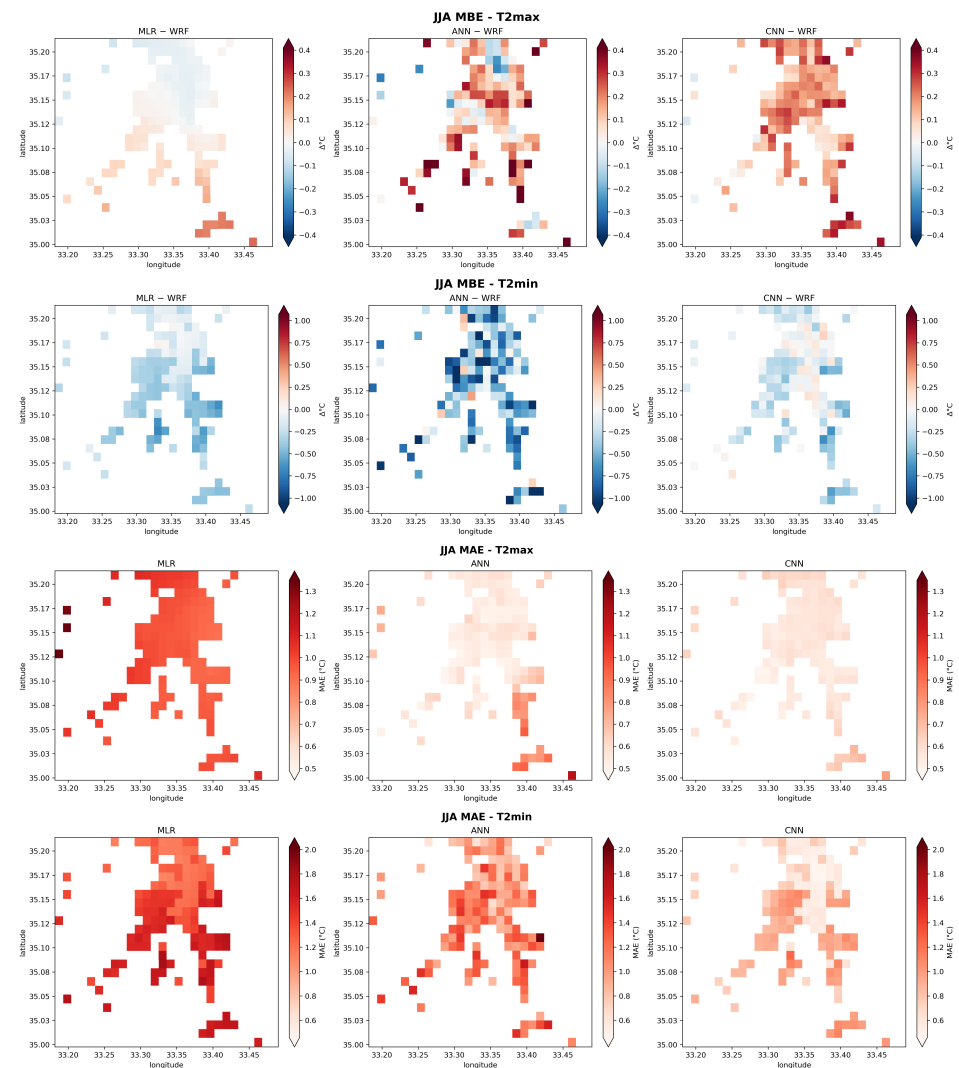


Figure 4. Summer mean bias error (MBE, 1st and 2nd rows) and mean absolute error (MAE, 3rd and 4th rows) of T2max and T2min between each machine learning model (MLR, ANN and CNN) and the WRF simulations for the test period (2012).

The 3rd and 4th rows of Figure 4 present the MAE distributions. The CNN demonstrates the best performance among the examined ML models, with average MAE values of approximately 0.6 °C for T2max and 0.8 °C for T2min. In contrast, the MLR model exhibits the largest errors, with MAE values of about 1 °C for T2max and 1.4 °C for T2min. The ANN follows with average MAE values of approximately 0.6 °C and 1.1 °C for T2max and T2min, respectively.

The spatial patterns of the MAE values also highlight differences between the models. The CNN presents the most homogeneous error distribution across the domain for T2max, indicating a more consistent reproduction of the WRF spatial patterns across the domain. In contrast, both the MLR and ANN models exhibit larger spatial variability, with localized areas of increased errors. Specifically, the ANN presents larger MAE values over the southeastern part of the domain, locally approaching 1.0 °C, whereas the MLR reaches values up to approximately 1.3 °C over the northwestern part. For the T2min variable, the CNN again exhibits the lowest MAE values across most of the domain. Both the CNN

and MLR show slightly lower errors over the central urban area (around 1.2 °C and 0.6 °C respectively), whereas the ANN displays a heterogeneous spatial distribution with several localized areas of higher errors.

As also observed in the bias analysis, the MAE values are generally larger for T2min compared to T2max, suggesting that reproducing nighttime temperature conditions during the summer period is more challenging for the ML models.

To provide a quantitative summary of the results discussed above, the domain-averaged MBE and MAE values between each ML model and the WRF simulations are presented in Tables 4 and 5, respectively. The metrics are reported for both T2max and T2min variables during the annual and summer periods.

Table 4. Domain-averaged mean bias error (MBE) (°C) between the machine learning models and WRF simulations for T2max and T2min during the annual and summer periods.

Period	MLR		ANN		CNN	
	T2max	T2min	T2max	T2min	T2max	T2min
Annual	0.10	−0.01	0.13	−0.04	0.08	−0.08
Summer	0.02	−0.29	0.13	−0.52	0.17	−0.17

Table 5. Domain-averaged mean absolute error (MAE) (°C) between the machine learning models and WRF simulations for T2max and T2min during the annual and summer periods.

Period	MLR		ANN		CNN	
	T2max	T2min	T2max	T2min	T2max	T2min
Annual	1.11	1.18	0.57	1.03	0.58	0.76
Summer	0.98	1.43	0.60	1.11	0.58	0.79

3.2. Evaluation of WRF and ML Models Against Observations

Subsequently, the performance of the WRF and ML models was evaluated against the observational data from the Nicosia meteorological station. In contrast to the previous analysis, where biases were computed for each grid cell across the domain, this evaluation focuses on the specific grid cell corresponding to the station location. The model outputs at this grid cell were compared with the observed measurements during the test period.

Figure 5 shows the monthly biases between the simulated and observed values for each model (WRF, MLR, ANN, CNN) for the T2max and T2min variables.

For T2max (1st and 2nd rows of Figure 5), all models tend to overestimate the observed temperatures during the winter and early spring months, with median biases generally ranging between 1 and 2 °C. During the warm period (May–September), the biases become negative, indicating a tendency of the models to underestimate the maximum temperatures by approximately 2 °C. The strongest underestimation is generally observed in September, when the median biases approach about −3 °C. Despite these seasonal differences, all models consistently reproduce the seasonal evolution of the biases, with a gradual transition from positive winter biases to negative summer biases and a return towards near-zero values during autumn. Among the examined models, the CNN and ANN show slightly smaller spreads in their distributions compared to the MLR and WRF, suggesting a more stable reproduction of the observed variability. In contrast, the MLR generally exhibits larger interquartile ranges during several months, indicating greater variability in the simulated biases.

In contrast, the T2min biases (3rd and 4th rows of Figure 5) reveal a systematic overestimation of nighttime temperatures throughout most of the year for all models. The largest biases are observed during winter and spring, with median values frequently

exceeding 4 °C. The biases decrease markedly during the summer months, with all models approaching near-zero values, particularly in July and August, before increasing again during autumn. This pronounced seasonal cycle is consistently reproduced by both the WRF and the ML models, indicating that the statistical approaches successfully preserve the temporal evolution of the nighttime temperature biases. Overall, the ML models reproduce the seasonal behavior of the observed temperatures similarly to the WRF simulations, with the CNN frequently presenting a slightly reduced spread compared to the other approaches.

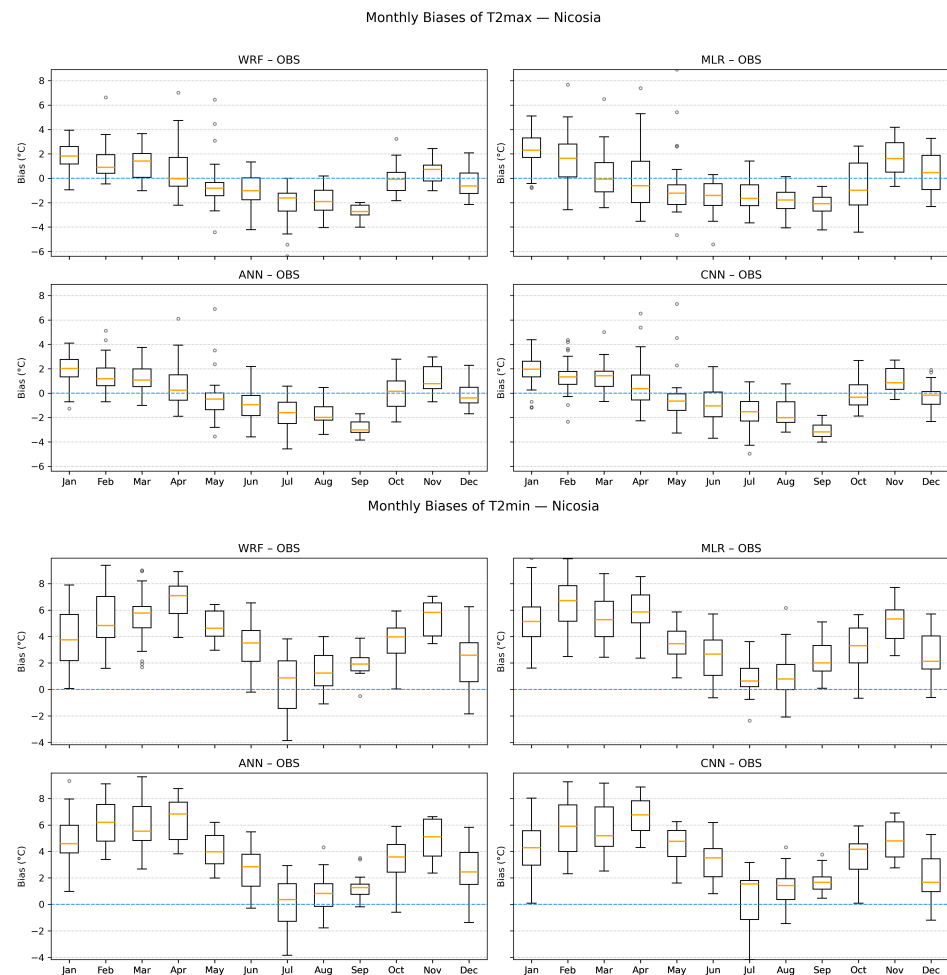


Figure 5. Monthly biases between simulated and observed T2max (1st and 2nd rows) and T2min (3rd and 4th rows) at the Nicosia station during the test period (2012).

To quantify the agreement between the simulated and observed values, the statistical metrics of MBE, MAE, RMSE and were calculated. Although the metrics were computed for all seasons during the test period (2012), only the annual and summer results are presented here (Tables 6 and 7) since the primary focus of the current study is on summer heat conditions. A detailed seasonal evaluation of the WRF model performance has already been documented in [52].

On an annual basis, the ANN and CNN models show comparable MAE and RMSE magnitudes with WRF for both T2max and T2min variables whereas the MLR performs worse for T2max but slightly better than the other three models for T2min. During the summer period the ML models exhibit smaller MAE values than WRF for T2max, indicating an improved representation of daytime temperature conditions. For T2min, the MLR and ANN models also present lower MAE values compared to WRF, while the CNN performance is very similar to that of the dynamical model. It is also worth noting that

the MAE values for T2max remain similar between the annual and summer evaluations (approximately 1.7 °C). In contrast, the annual MAE values for T2min are considerably larger (around 3.9 °C) than those of the summer period (about 2 °C), mainly because the annual statistics are influenced by the winter and spring seasons, during which larger deviations between the simulations and observations occur. These results suggest that the statistical downscaling methods are capable of effectively reproducing the observed temperature and, in some cases, better than the WRF simulations.

Table 6. Annual performance metrics between the simulated and observed T2max and T2min during the test period (2012).

Model	MBE (°C)		MAE (°C)		RMSE (°C)		<i>r</i>	
	T2max	T2min	T2max	T2min	T2max	T2min	T2max	T2min
WRF	−0.12	3.77	1.57	3.97	1.99	4.57	0.98	0.94
MLR	−0.13	3.75	1.88	3.85	2.31	4.52	0.97	0.96
ANN	0.00	3.74	1.56	3.94	1.94	4.61	0.98	0.95
CNN	−0.02	3.74	1.56	3.94	1.99	4.55	0.98	0.94

Table 7. Summer performance metrics between the simulated and observed T2max and T2min during the test period (2012).

Model	MBE (°C)		MAE (°C)		RMSE (°C)		<i>r</i>	
	T2max	T2min	T2max	T2min	T2max	T2min	T2max	T2min
WRF	−1.66	1.61	1.80	2.20	2.17	2.64	0.86	0.54
MLR	−1.59	1.47	1.65	1.80	1.99	2.32	0.89	0.66
ANN	−1.52	1.13	1.71	1.83	2.04	2.25	0.87	0.59
CNN	−1.42	1.55	1.69	2.21	2.02	2.58	0.86	0.54

3.3. Future Temperature Projections

The CMIP6 predictor variables were dynamically downscaled to 1 km resolution using the WRF model for both the historical (2008–2012) and future (2048–2052) periods (Simulations 2 and 3). The evaluation results presented in the previous sections demonstrated that the ML models successfully reproduce the main spatial temperature patterns simulated by the WRF model, providing confidence in their application for future climate projections. The predictor datasets produced by the WRF simulations driven by CMIP6 data at 12 km were subsequently used as inputs to the ML models in order to generate high-resolution (1 km) temperature projections using the statistical downscaling framework. To quantify the projected climate change signal, the differences between the future and historical periods were calculated for both T2max and T2min variables with the focus given on the summer season.

The projected summer changes in T2max are shown in Figure 6. All models consistently project a pronounced warming signal across the study area. The WRF simulation indicates an average increase of approximately 2 °C between the historical and future periods. The ANN produces a slightly stronger warming signal, reaching about 2.1 °C, while the CNN and MLR project a more moderate increase close to 1.8 °C. The ANN reproduces the WRF spatial pattern most closely, while the CNN and MLR exhibit systematically lower warming magnitudes. Despite these differences in magnitude, all models display a relatively uniform spatial warming pattern, indicating that the projected increase in daytime temperatures is broadly consistent across the urban domain.

Figure 7 presents the projected changes in T2min. The nighttime warming signal is generally smaller compared to that of T2max. The WRF simulation indicates an average

increase of approximately 1 °C, while the ANN and CNN models produce comparable warming patterns of similar magnitude. The MLR model exhibits a slightly stronger signal, with warming values reaching around 1.3 °C. In contrast to T2max, the spatial distribution of T2min warming displays greater variability across the domain. The MLR model projects a more spatially homogeneous warming field than the other approaches, whereas the WRF, ANN, and CNN simulations exhibit stronger heterogeneity. Furthermore, slightly lower warming is projected in the central urban area by the WRF, CNN, and MLR simulations, while higher increases appear mainly in the south-western part of the domain. This spatial heterogeneity suggests that local-scale characteristics may influence nighttime temperature changes more strongly than daytime temperatures.

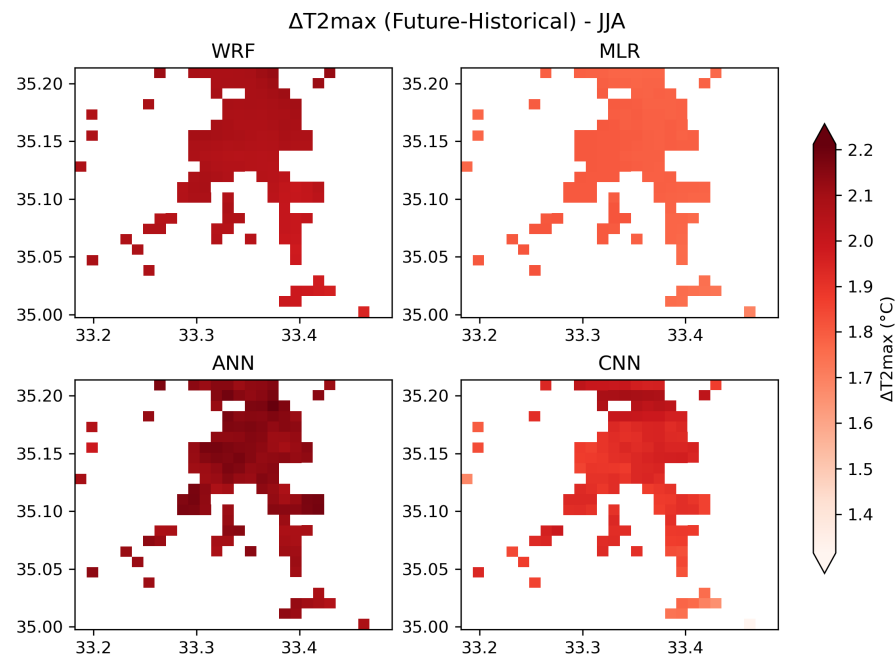


Figure 6. Projected summer temperature changes (Future–Historical) simulated by each model for T2max.

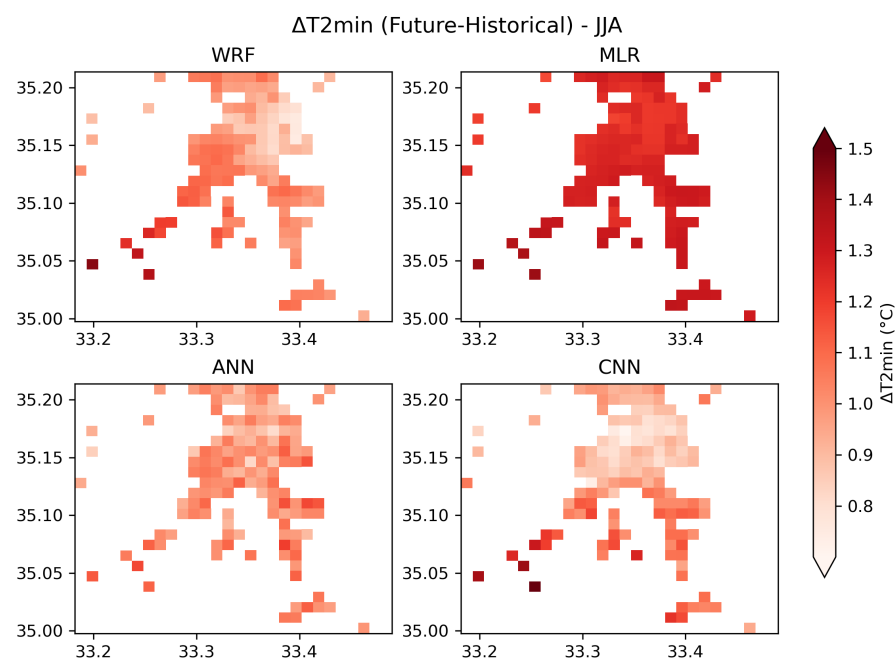


Figure 7. Projected summer temperature changes (Future–Historical) simulated by each model for T2min.

Overall, the results indicate that all models consistently project substantial summertime warming over Nicosia, with stronger increases in daytime maximum temperatures compared to nighttime minima. The close agreement between the dynamical and statistical downscaling approaches further supports the robustness of the projected climate change signal.

To further reduce projection uncertainty and improve the robustness of the downscaling results, a multi-model ensemble approach was applied. Multi-model ensembles have been shown to generally outperform individual model projections by reducing the influence of model-specific biases and uncertainties [63,64]. The ensemble mean was calculated by assigning equal weights to each model (WRF, MLR, ANN, and CNN). Specifically, the ensemble was obtained by averaging the outputs of the four models for each grid cell.

The projected ensemble mean warming for both variables during the summer season is presented in Figure 8. The ensemble results indicate a pronounced warming signal between the historical and future periods across the study area, with an average increase of approximately 1.9 °C for T2max and about 1.1 °C for T2min. The spatial pattern of the ensemble projections shows that the increase in daytime temperatures (T2max) is relatively uniform across the domain. In contrast, the projected warming for T2min exhibits greater spatial variability, with relatively lower increases in the urban core. These spatial patterns are consistent with those obtained from the individual model projections, indicating a high degree of agreement among the different modelling approaches regarding the location of the strongest and weakest warming signals. Overall, the ensemble projections confirm the strong summertime warming signal identified by the individual models, while providing a more robust estimate of the expected temperature increase over Nicosia. The average warming signals projected by each model, as well as the ensemble mean, are summarized in Table 8 for both T2max and T2min variables.

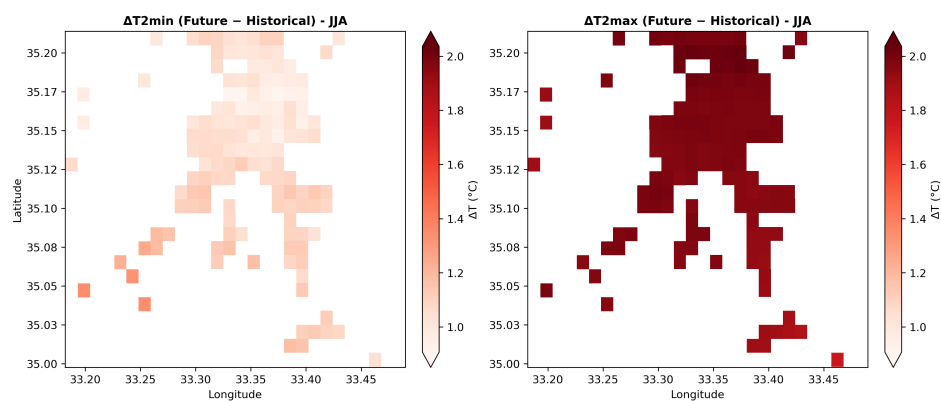


Figure 8. Ensemble mean of the summer temperature changes (Future–Historical) for T2min (**left**) and T2max (**right**), calculated as the equal-weighted average of the WRF, MLR, ANN, and CNN model projections.

Table 8. Projected summer temperature changes (Future–Historical) for T2max and T2min simulated by each model and the ensemble mean. Uncertainty is expressed as the inter-model standard deviation.

	WRF	MLR	ANN	CNN	Ensemble Mean
ΔT_{2max} (°C)	2.06	1.79	2.13	1.92	1.97 ± 0.15
ΔT_{2min} (°C)	1.00	1.27	1.00	0.97	1.06 ± 0.14

To further assess the robustness of the projected warming signal, a Welch two-sample *t*-test [65] was applied to compare the summer mean temperatures between the historical and future periods for both T2max and T2min across all models (WRF, MLR, ANN, and CNN). The test was performed using daily summer domain-averaged temperatures

(92 summer days per year over 5 years, resulting in 460 values for each period). For all models and both variables, the projected future warming was found to be statistically significant ($p < 0.001$), confirming that the increases in both daytime and nighttime temperatures are robust and unlikely to be due to random variability alone.

3.4. Identification of Future Urban Heat Hotspots

To identify areas expected to experience the strongest heat stress in the future climate, the ensemble mean projections were used to detect areas that are projected to experience consistently high thermal conditions under the SSP2-4.5 scenario. Grid cells exceeding the historical 95th percentile thresholds of both T2max and T2min were first identified for the future period. Grids simultaneously exceeding both thresholds were considered potential hotspots, representing locations exposed to elevated daytime and nighttime temperatures.

Figure 9 presents the spatial distribution of the ensemble mean future summer temperatures for T2min (left) and T2max (right). A total of 61 grid cells exceed the historical 95th percentile thresholds for both variables. Among these cells, the twenty with the highest combined temperature magnitude, defined as the mean of T2max and T2min, were identified as the most intense intra-urban heat hotspots (yellow stars). The different outline colors indicate the LCZ classes: black for LCZ 3 (compact low-rise), orange for LCZ 6 (open low-rise), yellow for LCZ 8 (large low-rise), and green for LCZ 9 (sparsely built). The identified hotspots are predominantly concentrated in the central and north-eastern parts of the urban area, which are mainly characterized by compact low-rise urban fabric (LCZ 3). These areas correspond to the most densely built parts of the domain, where reduced ventilation and increased heat storage contribute to elevated air temperatures.

To further investigate the distribution of the identified hotspot cells and their relationship with the different LCZ classes, a quantitative analysis was performed (Table 9). The table summarizes the total number of grid cells within each LCZ together with the number and percentage of cells identified as hotspots and top 20 hotspots. The results show that all LCZ 3 cells (100%) and almost all LCZ 8 cells (96.2%) were classified as hotspots, whereas only 13.5% of LCZ 6 cells and none of the LCZ 9 cells exceeded the hotspot threshold. Similarly, 62.5% of LCZ 3 cells and 19.2% of LCZ 8 cells were classified among the top 20 most intense hotspots. Considering the complete hotspot dataset, 39.3% and 41.0% of the 61 identified hotspot cells belong to LCZ 3 and LCZ 8, respectively, while only 19.7% occur within LCZ 6. Furthermore, the top 20 most intense hotspots are exclusively located within LCZ 3 (75%) and LCZ 8 (25%), highlighting the strong association between compact urban morphology and the occurrence of the most intense future heat hotspots.

Table 9. For each LCZ, the table shows the total number of grid cells, together with the absolute number of hotspot cells, their percentage relative to the corresponding LCZ and to the total number of hotspot cells, and the absolute number of the top 20 most intense hotspot cells together with their percentage relative to the corresponding LCZ and to the total number of top 20 hotspot cells.

LCZ	Total Cells	Hotspot Cells	Hotspots (% of LCZ)	Hotspots (% of Total)	Top 20 Hotspots	Top 20 (% of LCZ)	Top 20 (% of Total)
LCZ 3	24	24	100.0	39.3	15	62.5	75.0
LCZ 6	89	12	13.5	19.7	0	0.0	0.0
LCZ 8	26	25	96.2	41.0	5	19.2	25.0
LCZ 9	3	0	0.0	0.0	0	0.0	0.0

Overall, the hotspot analysis highlights the specific urban locations and the corresponding LCZs that are expected to experience the strongest combined daytime and nighttime warming in Nicosia city, providing valuable information for the development of targeted urban heat mitigation strategies.

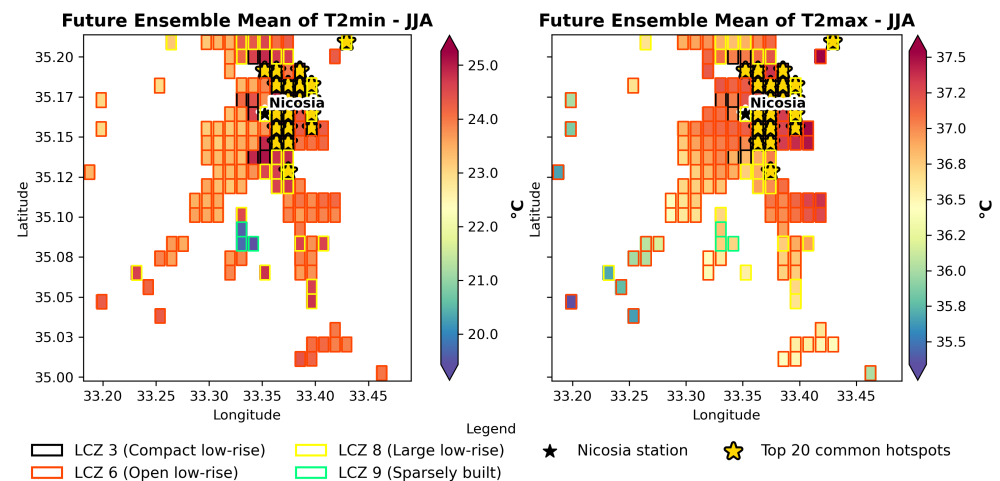


Figure 9. Spatial distribution of the ensemble mean future summer T2min (left) and T2max (right). The different outline colors indicate different LCZ classes. Yellow stars indicate the top 20 common urban heat hotspots based on the historical 95th percentile thresholds of both variables. The black star represents the location of the Nicosia meteorological station.

4. Discussion

4.1. Performance of Statistical Downscaling Methods

The performance of the statistical downscaling approaches was first assessed by evaluating their ability to reproduce the spatial temperature patterns generated by the WRF simulations for the T2max and T2min variables. After training, the derived statistical relationships were applied to the independent test period using the low-resolution predictor variables from the WRF–ERA5 simulation.

Overall, the results demonstrate that all examined ML models were capable of reproducing the spatial structure of the WRF temperature fields with relatively small deviations. Domain-averaged biases remain generally low across the study area for both variables, indicating that the ML models were able to capture the main large-scale to local-scale relationships governing the downscaling process. In terms of MAE, the simplest approach (MLR) performed reasonably well despite its linear formulation, while the ANN and CNN models achieved slightly better overall accuracy. The differences among the three methods remain relatively small, particularly for the T2min variable.

This behavior can be explained by the fact that near-surface temperature can be often adequately represented by linear statistical relationships with large-scale atmospheric conditions. Nevertheless, more complex ML architectures offer additional advantages when handling spatially structured predictors. In particular, CNN models are able to automatically extract spatial features and perform implicit dimensionality reduction through convolutional kernels, without the need for external feature selection procedures [66]. This capability becomes especially important when dealing with high-dimensional climate datasets.

Comparable performances of neural-network-based statistical downscaling methods have been reported in recent studies. For example, ref. [67] applied ANN, CNN and Random Forest models to downscale ERA5-Land daily mean temperature data from 9 km to 1 km resolution. Their results showed relatively small summer biases (approximately 0.2 °C) and MAE values close to 1 °C for both ANN and CNN models. These error magnitudes are comparable to those obtained in the present study, particularly for the CNN model. Furthermore, ref. [67] reported that the CNN architecture achieved the lowest MAE among the examined models, which is consistent with the findings of the present study.

Beyond the quantitative error metrics, the CNN model also exhibited a more spatially coherent representation of the temperature fields compared to the ANN model. In the ANN results, localized deviations between neighboring grid cells were occasionally observed, particularly for the T2min variable. This difference highlights the advantage of convolutional architectures, which explicitly account for spatial dependencies through localized connections and shared convolutional kernels. Such properties make CNN models particularly suitable for spatially structured datasets, such as climate variables.

Overall, the results indicate that while relatively simple statistical approaches (MLR) can already reproduce the main temperature patterns produced by the WRF model, deep learning methods (particularly CNN) provide improved spatial consistency and slightly better accuracy. These findings suggest that CNN-based statistical downscaling can serve as a reliable and computationally efficient alternative for generating high-resolution temperature fields derived from dynamical simulations. In terms of computational cost, the training and inference of the ML models require only a few minutes on a standard computing system, whereas the corresponding high-resolution WRF simulations (1 km resolution for a multi-year period) required several months of computational time on a high-performance computing system. This substantial difference highlights the potential of ML-based approaches for efficiently generating large ensembles of high-resolution climate projections.

An important characteristic of the proposed framework is that the statistical downscaling models are trained using high-resolution temperature fields generated by the WRF model rather than direct observations. This strategy follows the Perfect Prognosis approach, in which physically consistent large-scale atmospheric conditions from ERA5-driven WRF simulations are used to establish statistical relationships between regional- and urban-scale variables. This approach was adopted because the available observational network is insufficient to represent the spatial variability of air temperature across the study domain and does not provide the spatially continuous atmospheric predictor fields required for statistical downscaling. Consequently, the ML models are designed to emulate the behaviour of the reference dynamical model rather than directly learn from atmospheric observations and therefore inherit both its strengths and limitations. Therefore, confidence in the statistical downscaling framework ultimately depends on the ability of the reference WRF simulation to reproduce the observed temperature variability, which is evaluated in the following section.

4.2. Evaluation Against Observations

Subsequently, the performance of the WRF model and the ML models was evaluated based on their ability to reproduce the daily T2max and T2min at Nicosia station. The error magnitudes obtained in the present study are consistent with those reported in previous urban climate modelling studies. For example, ref. [68] evaluated the WRF model coupled with the SLUCM over a 50-day summer period in Athens at 1 km resolution and reported MAE values of approximately 1.7 °C and 1.5 °C for T2max and T2min, respectively. These values are in close agreement with the present study, where summer MAE values of approximately 1.8 °C for T2max and 2.2 °C for T2min were obtained. Likewise, ref. [69] applied WRF–SLUCM over Hangzhou city with 1 km resolution during a 14-day summer simulation and reported RMSE values of about 2.5 °C for T2max and 1.2 °C for T2min. These ranges are comparable to those obtained in the present study, particularly for T2max.

The relatively small MAE and RMSE values obtained during the summer period indicate that the statistical downscaling approaches are able to effectively capture the relationships between large-scale atmospheric predictors and local-scale temperature variability represented in the WRF simulations. The larger errors observed for T2min compared to T2max can be attributed to the greater complexity of nighttime processes in urban environ-

ments. During nighttime conditions, heat storage in urban material and surfaces together with anthropogenic heat release can significantly influence near-surface temperatures, making them more difficult to simulate accurately.

Overall, these results indicate that the ML models reproduced the observed temperature with a level of accuracy similar to that of the dynamical simulations. Consequently, the ML-based statistical downscaling methods can provide reliable representations of urban temperatures while maintaining a substantially lower computational cost compared to dynamical simulations. However, a limitation of the present evaluation is that it relies on a single urban meteorological station, reflecting the limited availability of urban observations within the study area. Consequently, while the temporal performance of both the WRF and ML models can be assessed at the station location, the spatial accuracy of the simulated 1 km temperature fields across the entire urban domain cannot be directly verified. Nevertheless, the physically based WRF simulations provide physically consistent spatial temperature fields that constitute the reference dataset for the statistical downscaling framework.

4.3. Projected Warming in Nicosia

Following the extraction of the statistical relationships during the training period and their satisfactory performance during the test period, these relationships were applied to the low-resolution outputs of the WRF-HIST and WRF-SSP2 simulations in order to generate 1 km resolution T2max and T2min fields for the historical and future periods, respectively. The projected high-resolution summer climate change signal was then calculated as the difference between the future and historical periods for each of the models.

The strong agreement between the CNN downscaled outputs and the WRF simulations observed in this study is consistent with previous work demonstrating that CNN-based statistical downscaling methods are capable of reproducing temperature fields and climate change signals comparable to those produced by RCMs [70].

The projections indicated a pronounced warming signal over Nicosia for the future period under the SSP2-4.5 scenario. In particular, the increase in daytime T2max is stronger than that of nighttime T2min. A similar behaviour has been reported in previous studies. For instance, Gumus et al. [71] analysed future temperature projections using data from 15 CMIP6 GCMs under the SSP2-4.5 scenario and found that projected increases in T2max exceed those of T2min in Morocco, which is located in the Mediterranean region. This pattern is consistent with the results of the present study and highlights the tendency for daytime temperatures to respond more strongly to future climate forcing in Mediterranean environments.

High-resolution regional climate simulations incorporating detailed urban representations have also been used to investigate future changes in near-surface air temperature. For example, Ni et al. [42] dynamically downscaled CMIP6 climate data using the WRF model to 1 km resolution over a metropolitan region in China under the SSP2-4.5 scenario. Their results showed an increase of approximately 1.2 °C in near-surface air temperature between the present climate and mid-century conditions, which is comparable to the warming signal obtained in the present study.

Although the projected warming signal is of a similar magnitude to the absolute errors obtained during the evaluation against observations, these quantities represent fundamentally different aspects of model performance and should not be interpreted directly. The evaluation metrics quantify the ability of the models to reproduce absolute temperatures during the historical period, whereas the projected climate change signal is calculated as the difference between dedicated historical (WRF-HIST) and future (WRF-SSP2) simulations performed with the same modelling framework. Consequently, systematic model biases are expected to largely cancel out when computing the climate change signal, making the

projected warming less sensitive to the absolute errors of the individual simulations. Furthermore, the use of an equal-weighted ensemble combining the dynamical WRF simulation and the three statistical downscaling models helps reduce the influence of model-specific biases and provides a more robust estimate of the projected warming. Overall, the close agreement between the warming signals produced by the ML models and those obtained from the dynamical WRF simulations, together with the consistency of the projected warming with previous studies, increases confidence in the robustness of the projected climate change signal over Nicosia and supports the suitability of ML-based statistical downscaling methods for generating high-resolution future climate projections.

4.4. Urban Heat Hotspots and Implications for Urban Climate Adaptation

The availability of kilometre-scale projections allowed the investigation of the spatial heterogeneity of the intra-urban thermal environment in Nicosia and the identification of areas that may experience the most intense heat exposure under future climate conditions. To identify these areas, the ensemble mean of both the dynamical and statistical downscaling models was used to detect grid cells exceeding the historical 95th percentile thresholds for both T2max and T2min during the future period. Considering both daytime and nighttime temperature extremes enables the identification of areas exposed to persistent heat stress conditions, providing a more comprehensive assessment of the potential impacts of future climate warming on the urban thermal environment.

The identified hotspots are predominantly located in the central and north-eastern parts of the city and correspond mainly to compact low-rise urban areas (LCZ 3). These urban forms are characterized by high building density and limited ventilation, which favour increased heat storage and reduced nighttime cooling. Such morphological characteristics have been widely recognized as key contributors to the development of the urban heat island effect and elevated urban air temperatures. Previous LCZ-based urban climate studies have shown that compact urban zones generally exhibit higher air temperatures compared to sparsely built or open areas due to the combined effects of increased thermal storage in urban materials and lower vegetation coverage [72,73].

The spatial clustering of hotspots within the most densely built parts of the city highlights the important role of urban morphology in modulating future temperature patterns. These findings reveal the areas where projected heat stress is expected to be most amplified. Consequently, the identification of these hotspots provides valuable information for the development of targeted urban heat mitigation strategies, such as increasing urban greenery, creating pocket parks, or applying reflective materials to urban surfaces. Overall, the integration of high-resolution dynamical and statistical downscaling with urban morphology analysis provides a useful framework for identifying priority areas for climate adaptation planning in Mediterranean cities facing increasing heat stress under climate change, enabling the detection of intra-urban thermal patterns that cannot be resolved by typical regional climate simulations.

5. Summary and Conclusions

This study investigated the potential of statistical downscaling techniques to generate high-resolution urban climate projections for Nicosia under the SSP2-4.5 scenario. The WRF model was used to dynamically downscale ERA5 data in order to produce the reference datasets required for the statistical downscaling of daily T2max and T2min at both regional (12 km) and local (1 km) scales. Three machine learning models of increasing complexity (Multiple Linear Regression, MLR, Artificial Neural Networks, ANN, and Convolutional Neural Networks, CNN) were developed to reproduce the high-resolution temperature fields.

Regarding the first objective of this study, the results showed that all three ML models were capable of reproducing the high-resolution temperature fields simulated by the WRF model with relatively small MBE and MAE values. Among the examined approaches, the CNN demonstrated the best overall performance, providing improved spatial coherence and slightly lower errors compared to the ANN and MLR models. These findings support the use of CNN-based statistical downscaling as a computationally efficient alternative to dynamical downscaling for producing high-resolution urban climate simulations. The comparison with observations from Nicosia meteorological station further indicated that the ML models reproduced the observed temperature variability with an accuracy comparable to that of the dynamical simulations. These findings demonstrate that ML-based statistical downscaling can accurately reproduce kilometer-scale dynamical simulations while substantially reducing computational cost.

Regarding the second objective, both the DD and SD approaches projected a substantial summertime warming between the historical and future periods over Nicosia by mid-century. This signal was estimated to be approximately 1.9 °C for T2max and 1.1 °C for T2min based on the multi-model ensemble of both dynamical and statistical approaches.

Regarding the third objective, the comparison between the dynamical and statistical downscaling approaches showed that the ML models reproduced projected warming signals of similar magnitude to those obtained from the WRF simulations with the ANN showing the closest agreement and the MLR exhibiting the largest deviations from the WRF projections for both T2max and T2min. Therefore, once trained using dynamical simulations, ML-based statistical downscaling represents a reliable efficient alternative for generating urban climate projections.

Regarding the fourth objective, the high-resolution projections enabled the identification of intra-urban thermal variability and the detection of future urban heat hotspots. The most intense hotspots were mainly located in compact low-rise urban areas (LCZ 3), highlighting the important role of urban morphology in amplifying future heat stress.

Overall, these findings demonstrate that integrating high-resolution dynamical and machine learning-based statistical downscaling with urban morphology analysis provides an efficient and robust framework for urban-scale climate assessments by combining the physical consistency of dynamical models with the computational efficiency of machine learning approaches, while enabling the identification of priority areas for climate adaptation planning in Mediterranean cities facing increasing heat stress.

Future work could extend this approach by applying the trained ML models to a large number of GCM-RCM pairs, following the approach of Banõ-Medina et al. [70], to generate probabilistic 1 km projections spanning multiple GCMs and SSP scenarios at negligible additional computational cost, thereby sampling the full phase space of projection uncertainty. In addition, future research could explore the interpretability of the ML models using explainable artificial intelligence techniques, such as SHAP values, to identify the relative importance of the predictor variables and improve the physical understanding of the relationships between large-scale atmospheric conditions and local temperature patterns. The framework could further be leveraged to evaluate targeted urban heat mitigation strategies, such as increasing urban greenery or applying cool materials, under projected climate conditions.

Author Contributions: Conceptualization and methodology, K.K.-K., C.C., P.H. and J.L.; software and formal analysis, K.K.-K.; investigation and data curation, K.K.-K.; writing—original draft preparation, K.K.-K.; writing—review and editing, K.K.-K., C.C., P.H. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the EMME-CARE project, funded by the European Union's Horizon 2020 Research and Innovation Programme, under grant agreement no. 856612, and co-funded by the Government of the Republic of Cyprus.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors acknowledge the use of High Performance Computing (HPC) resources provided by The Cyprus Institute.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Network
CMIP6	Coupled Model Intercomparison Project Phase 6
CNN	Convolutional Neural Network
DD	Dynamical Downscaling
ERA5	ECMWF Reanalysis v5
GCM	General Circulation Model
JJA	June–July–August
LCZ	Local Climate Zone
MAE	Mean Absolute Error
MBE	Mean Bias Error
ML	Machine Learning
MLR	Multiple Linear Regression
RCM	Regional Climate Model
RMSE	Root Mean Square Error
SD	Statistical Downscaling
SLUCM	Single-Layer Urban Canopy Model
SSP	Shared Socioeconomic Pathway
T2	2 m air temperature
T2max	Daily maximum 2 m air temperature
T2min	Daily minimum 2 m air temperature
UHI	Urban Heat Island
WRF	Weather Research and Forecasting model

References

1. Cannon, A.J. Twelve months at 1.5 °C signals earlier than expected breach of Paris Agreement threshold. *Nat. Clim. Change* **2025**, *15*, 266–269. [[CrossRef](#)]
2. Masson-Delmotte, V.; Zhai, P.; Pirani, A.; Connors, S.L.; Péan, C.; Berger, S.; Caud, N.; Chen, Y.; Goldfarb, L.; Gomis, M.I.; et al. (Eds.) *Climate Change 2021: The Physical Science Basis*; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2021. [[CrossRef](#)]
3. Ebi, K.L.; Capon, A.; Berry, P.; Broderick, C.; de Dear, R.; Havenith, G.; Honda, Y.; Kovats, R.S.; Ma, W.; Malik, A.; et al. Hot weather and heat extremes: Health risks. *Lancet* **2021**, *398*, 698–708. [[CrossRef](#)] [[PubMed](#)]
4. Zittis, G.; Almazroui, M.; Alpert, P.; Ciais, P.; Cramer, W.; Dahdal, Y.; Fnais, M.; Francis, D.; Hadjinicolaou, P.; Howari, F.; et al. Climate Change and Weather Extremes in the Eastern Mediterranean and Middle East. *Rev. Geophys.* **2022**, *60*, e2021RG000762. [[CrossRef](#)]
5. Ntounos, A.; Hadjinicolaou, P.; Zittis, G.; Constantinidou, K.; Vautard, R.; Lelieveld, J. Escalating heat stress in Middle East–North Africa under global warming. *Crit. Insights Clim. Change* **2026**, *2*, 2610922. [[CrossRef](#)]
6. Zittis, G.; Alberti, T.; Almazroui, M.; Driouech, F.; Faranda, D.; Francis, D.; Hadjinicolaou, P.; Kurnaz, M.L.; Lazoglou, G.; Nikulin, G.; et al. Analysis of the 2024 Hajj heat event and future temperature extremes in Mecca. *npj Nat. Hazards* **2025**, *2*, 107. [[CrossRef](#)]

7. Lazoglou, G.; Hadjinicolaou, P.; Sofokleous, I.; Bruggeman, A.; Zittis, G. Climate change and extremes in the Mediterranean island of Cyprus: From historical trends to future projections. *Environ. Res. Commun.* **2024**, *6*, 095020. [\[CrossRef\]](#)
8. Hadjinicolaou, P.; Tzyrkalli, A.; Zittis, G.; Lelieveld, J. Urbanisation and Geographical Signatures in Observed Air Temperature Station Trends Over the Mediterranean and the Middle East–North Africa. *Earth Syst. Environ.* **2023**, *7*, 649–659. [\[CrossRef\]](#)
9. Tzyrkalli, A.; Economou, T.; Lazoglou, G.; Constantinidou, K.; Hadjinicolaou, P.; Lelieveld, J. Urban Heat Island Trends in the Middle East and North Africa: A statistical approach. *Int. J. Climatol.* **2024**, *44*, 3998–4008. [\[CrossRef\]](#)
10. Flato, G.; Marotzke, J.; Abiodun, B.; Braconnot, P.; Chou, S.C.; Collins, W.; Cox, P.; Driouech, F.; Emori, S.; Eyring, V. Evaluation of Climate Models. In *Climate Change 2013: The Physical Science Basis*; Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., Eds.; Cambridge University Press: Cambridge, UK, 2014; pp. 741–866. [\[CrossRef\]](#)
11. Wuebbles, U.; Chen, D.; Facchini, D.; Frame, M.C.; Mahowald, D.; Winther, N.; Cubasch, J.G. Introduction. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., Eds.; Cambridge University Press: Cambridge, UK, 2013; pp. 119–158.
12. Maraun, D.; Widmann, M. *Statistical Downscaling and Bias Correction for Climate Research*; Cambridge University Press: Cambridge, UK, 2017. [\[CrossRef\]](#)
13. Hewitson, F.; Christensen, B.; Hulme, J.; von Storch, M.; Whetton, H.; Jones, P.; Mearns, R.; Fu, L.; Giorgi, C. Regional climate information: Evaluation and projections. In *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*; Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van der Linden, P.J., Dai, X., Maskell, K., Johnson, C.A., Eds.; Cambridge University Press: Cambridge, UK, 2001; pp. 583–638.
14. Chen, S.T.; Yu, P.S.; Tang, Y.H. Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *J. Hydrol.* **2010**, *385*, 13–22. [\[CrossRef\]](#)
15. Nourani, V.; Paknezhad, N.J.; Sharghi, E.; Khosravi, A. Estimation of prediction interval in ANN-based multi-GCMs downscaling of hydro-climatologic parameters. *J. Hydrol.* **2019**, *579*, 124226. [\[CrossRef\]](#)
16. Giorgi, F.; Mearns, L.O. Approaches to the simulation of regional climate change: A review. *Rev. Geophys.* **1991**, *29*, 191–216. [\[CrossRef\]](#)
17. Rummukainen, M. State-of-the-art with regional climate models. *Wiley Interdiscip. Rev. Clim. Change* **2010**, *1*, 82–96. [\[CrossRef\]](#)
18. Langendijk, G.S.; Halenka, T.; Hoffmann, P.; Adinolfi, M.; Campino, A.A.; Asselin, O.; Bastin, S.; Bechtel, B.; Belda, M.; Bushenkova, A.; et al. Towards better understanding the urban environment and its interactions with regional climate change—The WCRP CORDEX Flagship Pilot Study URB-RCC. *Urban Clim.* **2024**, *58*, 102165. [\[CrossRef\]](#)
19. Constantinidou, K.; Hadjinicolaou, P.; Tzyrkalli, A.; Zittis, G.; Lelieveld, J. Urban Heat Island in the Eastern Mediterranean and Middle East: Multi-City Approach in Regional Climate Modelling. *Int. J. Climatol.* **2026**, *46*, e70255. [\[CrossRef\]](#)
20. Wilby, R.L.; Wigley, T.M. Downscaling general circulation model output: A review of methods and limitations. *Prog. Phys. Geogr.* **1997**, *21*, 530–548. [\[CrossRef\]](#)
21. Wang, Y.; Leung, L.R.; McGregor, J.L.; Lee, D.K.; Wang, W.C.; Ding, Y.; Kimura, F. Regional Climate Modeling: Progress, Challenges, and Prospects. *J. Meteorol. Soc. Jpn. Ser. II* **2004**, *82*, 1599–1628. [\[CrossRef\]](#)
22. Wilby, R.L.; Charles, S.P.; Zorita, E.; Timbal, B.; Whetton, P.; Mearns, L.O. *Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods Document History*; Intergovernmental Panel on Climate Change: Geneva, Switzerland, 2004.
23. Maraun, D.; Wetterhall, F.; Ireson, A.M.; Chandler, R.E.; Kendon, E.J.; Widmann, M.; Brienen, S.; Rust, H.W.; Sauter, T.; Themel, M.; et al. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.* **2010**, *48*, RG3003. [\[CrossRef\]](#)
24. Sachindra, D.A.; Huang, F.; Barton, A.F.; Perera, B.J. Multi-model ensemble approach for statistically downscaling general circulation model outputs to precipitation. *Q. J. R. Meteorol. Soc.* **2014**, *140*, 1161–1178. [\[CrossRef\]](#)
25. Koutroumanou-Kontosi, K.; Cartalis, C.; Philippopoulos, K.; Agathangelidis, I.; Polydoros, A. A Methodology for Bridging the Gap between Regional- and City-Scale Climate Simulations for the Urban Thermal Environment. *Climate* **2022**, *10*, 106. [\[CrossRef\]](#)
26. Liu, Y.; Yang, Y.; Jing, W.; Yue, X. Comparison of Different Machine Learning Approaches for Monthly Satellite-Based Soil Moisture Downscaling over Northeast China. *Remote Sens.* **2017**, *10*, 31. [\[CrossRef\]](#)
27. Tripathi, S.; Srinivas, V.V.; Nanjundiah, R.S. Downscaling of precipitation for climate change scenarios: A support vector machine approach. *J. Hydrol.* **2006**, *330*, 621–640. [\[CrossRef\]](#)
28. Anandhi, A.; Srinivas, V.V.; Nanjundiah, R.S.; Kumar, D.N. Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine. *Int. J. Climatol.* **2008**, *28*, 401–420. [\[CrossRef\]](#)
29. Devak, M.; Dhanya, C.T.; Gosain, A.K. Dynamic coupling of support vector machine and K-nearest neighbour for downscaling daily rainfall. *J. Hydrol.* **2015**, *525*, 286–301. [\[CrossRef\]](#)
30. Schoof, J.T.; Pryor, S.C.; Schoof, J.T. Downscaling Temperature and Precipitation: A Comparison of Regression-Based Methods and Artificial Neural Networks. *Part Phys. Environ. Geogr. Commons Int. J. Climatol.* **2001**, *21*, 773–790. [\[CrossRef\]](#)

31. Ahmed, K.; Shahid, S.; Haroon, S.B.; Wang, X.J. Multilayer perceptron neural network for downscaling rainfall in arid region: A case study of Baluchistan, Pakistan. *J. Earth Syst. Sci.* **2015**, *124*, 1325–1341. [[CrossRef](#)]
32. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
33. Vandal, T.; Kodra, E.; Ganguly, A.R. Intercomparison of machine learning methods for statistical downscaling: The case of daily and extreme precipitation. *Theor. Appl. Climatol.* **2019**, *137*, 557–570. [[CrossRef](#)]
34. Pan, B.; Hsu, K.; AghaKouchak, A.; Sorooshian, S. Improving Precipitation Estimation Using Convolutional Neural Network. *Water Resour. Res.* **2019**, *55*, 2301–2321. [[CrossRef](#)]
35. González-Abad, J.; Gutiérrez, J.M. Are Deep Learning Methods Suitable for Downscaling Global Climate Projections? An Intercomparison for Temperature and Precipitation over Spain. *Artif. Intell. Earth Syst.* **2025**, *4*, 240121. [[CrossRef](#)]
36. Balmaceda-Huarte, R.; Baño-Medina, J.; Olmo, M.E.; Bettolli, M.L. On the use of convolutional neural networks for downscaling daily temperatures over southern South America in a climate change scenario. *Clim. Dyn.* **2024**, *62*, 383–397. [[CrossRef](#)]
37. Rampal, N.; Hobeichi, S.; Gibson, P.B.; Baño-Medina, J.; Abramowitz, G.; Beucier, T.; González-Abad, J.; Chapman, W.; Harder, P.; Gutiérrez, G.M. Enhancing Regional Climate Downscaling through Advances in Machine Learning. *Artif. Intell. Earth Syst.* **2024**, *3*, 230066. [[CrossRef](#)]
38. Dorji, S.; Herath, S.; Mishra, B.K. Future Climate of Colombo Downscaled with SDSM-Neural Network. *Climate* **2017**, *5*, 24. [[CrossRef](#)]
39. Li, D.; Zou, L.; Zhou, T. Extreme Climate Event Changes in China in the 1.5 and 2 °C Warmer Climates: Results From Statistical and Dynamical Downscaling. *J. Geophys. Res. Atmos.* **2018**, *123*, 10215–10230. [[CrossRef](#)]
40. Struglia, M.V.; Anav, A.; Antonelli, M.; Calmanti, S.; Catalano, F.; Dell'Aquila, A.; Pichelli, E.; Pisacane, G. Impact of spatial resolution on multi-scenario WRF-ARW simulations driven by the CMIP6 MPI-ESM1-2-HR global model: A focus on precipitation distribution over Italy. *Geosci. Model Dev.* **2025**, *18*, 6095–6116. [[CrossRef](#)]
41. Teixeira, M.; Carvalho, D.; Rocha, A.; Pereira, S.C. Climate change impacts on human health in Portugal based on downscaled CMIP6 future climate scenarios: An analysis using thermal comfort indices. *Environ. Res. Clim.* **2026**, *5*, 025004. [[CrossRef](#)]
42. Ni, X.; Chang, Y.; Bai, T.; Liu, P.; Song, H.; Wang, F.; Jin, M. Projections of Urban Heat Island Effects Under Future Climate Scenarios: A Case Study in Zhengzhou, China. *Remote Sens.* **2025**, *17*, 2660. [[CrossRef](#)]
43. Peel, M.C.; Finlayson, B.L.; McMahon, T.A. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 1633–1644. [[CrossRef](#)]
44. Camera, C.; Bruggeman, A.; Hadjinicolaou, P.; Pashiardis, S.; Lange, M.A. Evaluation of interpolation techniques for the creation of gridded daily precipitation ($1 \times 1 \text{ km}^2$); Cyprus, 1980–2010. *J. Geophys. Res. Atmos.* **2014**, *119*, 693–712. [[CrossRef](#)]
45. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [[CrossRef](#)]
46. Eyring, V.; Bony, S.; Meehl, G.A.; Senior, C.A.; Stevens, B.; Stouffer, R.J.; Taylor, K.E. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **2016**, *9*, 1937–1958. [[CrossRef](#)]
47. Mauritsen, T.; Bader, J.; Becker, T.; Behrens, J.; Bittner, M.; Brokopf, R.; Brovkin, V.; Claussen, M.; Crueger, T.; Esch, M.; et al. Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO₂. *J. Adv. Model. Earth Syst.* **2019**, *11*, 998–1038. [[CrossRef](#)] [[PubMed](#)]
48. Adachi, S.A.; Tomita, H. Methodology of the Constraint Condition in Dynamical Downscaling for Regional Climate Evaluation: A Review. *J. Geophys. Res. Atmos.* **2020**, *125*, e2019JD032166. [[CrossRef](#)]
49. Ferreiro-Lera, G.-B.; Penas, Á.; del Río, S. Unveiling Deviations from IPCC Temperature Projections through Bayesian Downscaling and Assessment of CMIP6 General Circulation Models in a Climate-Vulnerable Region. *Remote Sens.* **2024**, *16*, 1831. [[CrossRef](#)]
50. Müller, W.A.; Jungclaus, J.H.; Mauritsen, T.; Baehr, J.; Bittner, M.; Budich, R.; Bunzel, F.; Esch, M.; Ghosh, R.; Haak, H.; et al. A Higher-resolution Version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *J. Adv. Model. Earth Syst.* **2018**, *10*, 1383–1413. [[CrossRef](#)]
51. Kusaka, H.; Kimura, F. Coupling a single-layer urban canopy model with a simple atmospheric model: Impact on urban heat island simulation for an idealized case. *J. Meteorol. Soc. Jpn.* **2004**, *82*, 67–80. [[CrossRef](#)]
52. Koutroumanou-Kontosi, K.; Cartalis, C.; Hadjinicolaou, P.; Constantinidou, K.; Agathangelidis, I.; Lelieveld, J. The Urban Thermal Environment of a Climate Change Hotspot: High-Resolution Modelling of Nicosia, Cyprus. *Earth Syst. Environ.* **2025**, *9*, 1691–1714. [[CrossRef](#)]
53. Jerez, S.; López-Romero, J.M.; Turco, M.; Lorente-Plazas, R.; Gómez-Navarro, J.J.; Jiménez-Guerrero, P.; Montávez, J.P. On the Spin-Up Period in WRF Simulations Over Europe: Trade-Offs Between Length and Seasonality. *J. Adv. Model. Earth Syst.* **2020**, *12*, e2019MS001945. [[CrossRef](#)]
54. Fowler, H.J.; Blenkinsop, S.; Tebaldi, C. Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. *Int. J. Climatol.* **2007**, *27*, 1547–1578. [[CrossRef](#)]
55. Gutiérrez, J.M.; San-Martín, D.; Brands, S.; Manzanar, R.; Herrera, S. Reassessing Statistical Downscaling Techniques for Their Robust Application under Climate Change Conditions. *J. Clim.* **2013**, *26*, 171–188. [[CrossRef](#)]

56. Manzananas, R.; Lucero, A.; Weisheimer, A.; Gutiérrez, J.M. Can bias correction and statistical downscaling methods improve the skill of seasonal precipitation forecasts? *Clim. Dyn.* **2018**, *50*, 1161–1176. [\[CrossRef\]](#)
57. Hanssen-Bauer, I.; Achberger, C.; Benestad, R.E.; Chen, D.; Førland, E.J. Statistical downscaling of climate scenarios over Scandinavia. *Clim. Res.* **2005**, *29*, 255–268. [\[CrossRef\]](#)
58. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.; Gruber, B.; Lafourcade, B.; Leitão, P.J.; et al. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2013**, *36*, 27–46. [\[CrossRef\]](#)
59. Agatonovic-Kustrin, S.; Beresford, R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J. Pharm. Biomed. Anal.* **2000**, *22*, 717–727. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Kingma, D.P.; Ba, J.L. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015.
61. Chandra, B.; Sharma, R.K. Deep learning with adaptive learning rate using laplacian score. *Expert Syst. Appl.* **2016**, *63*, 1–7. [\[CrossRef\]](#)
62. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [\[CrossRef\]](#)
63. Mitra, A.K.; Iyengar, G.R.; Durai, V.R.; Sanjay, J.; Krishnamurti, T.N.; Mishra, A.; Sikka, D.R. Experimental real-time multi-model ensemble (MME) prediction of rainfall during monsoon 2008: Large-scale medium-range aspects. *J. Earth Syst. Sci.* **2011**, *120*, 27–52. [\[CrossRef\]](#)
64. Kumar, A.; Mitra, A.K.; Bohra, A.K.; Iyengar, G.R.; Durai, V.R. Multi-model ensemble (MME) prediction of rainfall using neural networks during monsoon season in India. *Meteorol. Appl.* **2012**, *19*, 161–169. [\[CrossRef\]](#)
65. L., W.B. The generalization of ‘student’s’ problem when several different population variances are involved. *Biometrika* **1947**, *34*, 28–35. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Baño-Medina, J.; Manzananas, R.; Gutierrez, J.M. Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geosci. Model Dev.* **2020**, *13*, 2109–2124. [\[CrossRef\]](#)
67. Bhakare, S.; Gesso, S.D.; Venturini, M.; Zardi, D.; Trentini, L.; Matiu, M.; Petitta, M. Intercomparison of Machine Learning Models for Spatial Downscaling of Daily Mean Temperature in Complex Terrain. *Atmosphere* **2024**, *15*, 1085. [\[CrossRef\]](#)
68. Roukounakis, N.; Varotsos, K.V.; Katsanos, D.; Lemesios, I.; Giannakopoulos, C.; Retalis, A. High Resolution WRF Modelling of Extreme Heat Events and Mapping of the Urban Heat Island Characteristics in Athens, Greece. *Sustainability* **2023**, *15*, 16509. [\[CrossRef\]](#)
69. Chen, F.; Yang, X.; Zhu, W. WRF simulations of urban heat island under hot-weather synoptic conditions: The case study of Hangzhou City, China. *Atmos. Res.* **2014**, *138*, 364–377. [\[CrossRef\]](#)
70. Baño-Medina, J.; Manzananas, R.; Cimadevilla, E.; Fernandez, J.; Gonzalez-Abad, J.; Cofinó, A.S.; Gutierrez, J.M. Downscaling multi-model climate projection ensembles with deep learning (DeepESD): Contribution to CORDEX EUR-44. *Geosci. Model Dev.* **2022**, *15*, 6747–6758. [\[CrossRef\]](#)
71. Gumus, V.; Moçayd, N.E.; Seker, M.; Seaid, M. Evaluation of future temperature and precipitation projections in Morocco using the ANN-based multi-model ensemble from CMIP6. *Atmos. Res.* **2023**, *292*, 106880. [\[CrossRef\]](#)
72. Huang, J.; Huang, Z.; Liu, W. Combining the WRF model and LCZ scheme to assess spatiotemporal variations of thermal comfort in Shenzhen’s built-up areas. *Sustain. Cities Soc.* **2025**, *122*, 106252. [\[CrossRef\]](#)
73. Rahmani, N.; Sharifi, A. Urban heat dynamics in Local Climate Zones (LCZs): A systematic review. *Build. Environ.* **2025**, *267*, 112225. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.