forecasting

MDPI

Article

An In-Depth Look at Rising Temperatures: Forecasting with Advanced Time Series Models in Major US Regions

Kameron B. Kinast * and Ernest Fokoué

School of Mathematics and Statistics, Rochester Institute of Technology, Rochester, NY 14623, USA; epfeqa@rit.edu
* Correspondence: kbk7499@rit.edu

Abstract: With growing concerns over climate change, accurately predicting temperature trends is crucial for informed decision-making and policy development. In this study, we perform a comprehensive comparative analysis of four advanced time series forecasting models—Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing (ETS), Multilayer Perceptron (MLP), and Gaussian Processes (GP)—to assess changes in minimum and maximum temperatures across four key regions in the United States. Our analysis includes hyperparameter optimization for each model to ensure peak performance. The results indicate that the MLP model outperforms the other models in terms of accuracy for temperature forecasting. Utilizing this best-performing model, we conduct temperature projections to evaluate the hypothesis that the rates of change in temperatures are greater than zero. Our findings confirm a positive rate of change in both maximum and minimum temperatures, suggesting a consistent upward trend over time. This research underscores the critical importance of refining time series forecasting models to address the challenges posed by climate change and supporting the development of effective strategies to mitigate the impacts of rising temperatures. The insights gained from this work emphasize the need for continuous advancement in predictive modeling techniques to better understand and respond to the dynamics of climate change.

Keywords: time series forecasting; temperature trends; ARIMA; exponential smoothing; Multilayer Perceptron; Gaussian Processes

Citation: Kinast, K.B.; Fokoué, E. An In-Depth Look at Rising Temperatures: Forecasting with Advanced Time Series Models in Major US Regions. *Forecasting* **2024**, *6*, 815–838. https://doi.org/10.3390/forecast6030041

Academic Editors: Yang Yang and Wenlin Dai

Received: 16 July 2024
Revised: 14 September 2024
Accepted: 16 September 2024
Published: 18 September 2024

1. Introduction

For many decades, climate scientists have expressed growing concern over the impacts of climate change, particularly rising temperatures. The 2015 Paris Agreement, established at the United Nations Framework Convention on Climate Change (UNFCCC) conference, seeks to limit the global average temperature increase to below 2 °C above pre-industrial levels (1850–1900), with efforts to further restrict it to 1.5 °C [1]. Numerous studies have examined potential scenarios where the 1.5 °C or 2 °C thresholds are surpassed, assessing their wide-ranging impacts on sectors including the economy, water resources, and public health [2–5]. Rising temperatures pose significant risks to biodiversity, ecosystems, and human populations, with the degree of risk depending on various factors, such as extreme heat events, as highlighted by the Summary for Policymakers by the Intergovernmental Panel on Climate Change (IPCC) [6]. Given these critical concerns, it is crucial to continually advance our understanding and forecasting of temperature trends to inform effective climate action and policy decisions.

Despite the increase in global mean temperature, the magnitude of temperature changes varies across regions [7]. For example, research has shown that temperature and temperature extremes in China have risen more sharply than the global average [8]. Many studies have investigated these changes using various methods, including linear regression, statistical models, and the Coupled Model Intercomparison Project (CMIP), examining both national trends and regional variations within the United States. Researchers at the U.S. Global Change Research Program used linear regression to report a 0.7 °C increase

in annual average temperatures across the U.S. from 1986 to 2016 compared to 1901 to 1960 [9]. In a different study but a closer look, Chattopadhyay and Edwards applied nonparametric statistical tests to a time series of annual precipitation and air temperature in Kentucky (1950 to 2010), highlighting that only 3 out of 42 stations in Kentucky exhibited positive trends in temperature [10]. These findings emphasize the importance of region-specific weather studies in better understanding localized impacts and developing effective adaptation strategies.

While studies have highlighted regional variations in temperature trends, the assessment of changes in maximum and minimum temperatures remains ambiguous, with no evidence that trends in extremes and averages share the same patterns [11]. Lee, Li, and Lund developed statistical models incorporating extreme values and changepoint features to estimate trends, concluding that while minimum temperatures have increased significantly, maximum temperatures have risen more slowly on a monthly basis. Another study examining monthly maximum and minimum temperatures at over 10,000 stations in the U.S. found similar results, indicating that temperatures in the U.S. are not becoming more extreme [12]. The National Centers for Environmental Information (NCEI) developed the Climate at a Glance tool, offering near real-time analysis of monthly and annual global-scale temperatures using data from the National Oceanic and Atmospheric Administration's (NOAA) Global Surface Temperature Analysis [13]. Figure 1 presents the average monthly maximum and minimum temperatures in the United States from 1950 to 2022. While this tool does not include trend displays within the plots, a slight increase in temperature extremes over time is still evident. Given the need to evaluate regional trends within the U.S., this study proposes analyzing daily minimum and maximum temperatures in highly populated cities across the country.

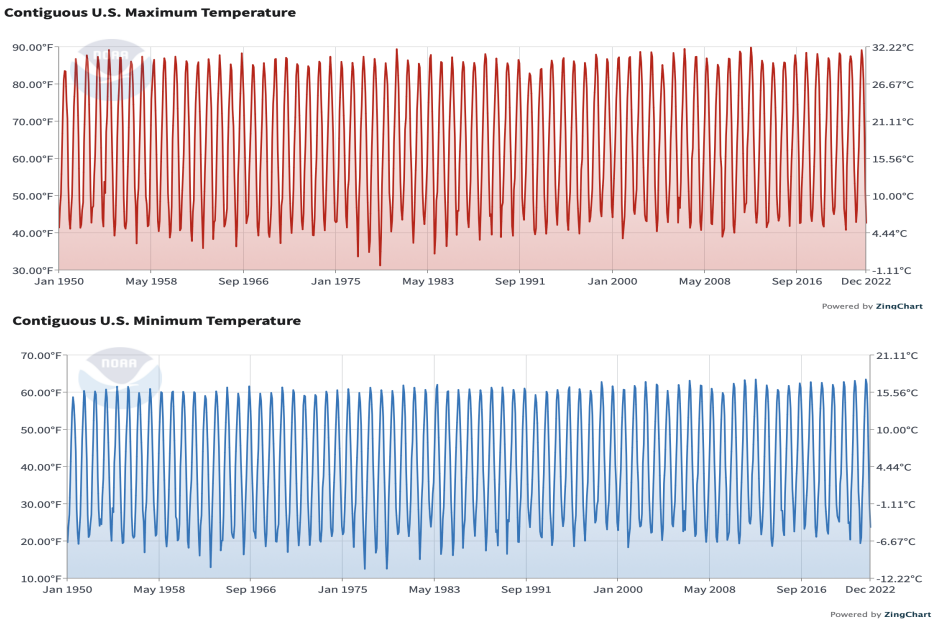

Figure 1. Line graphs of contiguous U.S. monthly maximum (**top**) and minimum (**bottom**) Temperatures from 1950 to 2022 [13].

Besides analyzing historic weather observations, scientists employ diverse machine learning and traditional approaches to forecast weather patterns. Numerous statistical techniques for time series temperature forecasting have been proposed in the literature. For instance, a research study developed a statistical forecasting technique using an Autoregressive Integrated Moving Average (ARIMA) model to predict future daily temperatures and precipitation in Chicago and Phoenix [14]. Another study examined the effectiveness of three exponential smoothing models in handling temperature data from three

cities in Texas, concluding that the Double Exponential Smoothing technique (DEST) is the most suitable [15]. Despite these advancements in forecasting techniques, there are limited comparative studies that evaluate time series forecasting models using identical weather datasets. For example, a study performed a comparative analysis of various weather forecasting models, including multiple linear regression, ARIMA, and artificial neural networks (ANN), across different studies. However, these studies did not utilize the same atmospheric science dataset, which limited direct comparisons of the models' performance [16].

Beyond the field of atmospheric sciences, several comparative studies have examined classical methods versus machine learning approaches for time series forecasting using the same dataset across different domains, such as econometrics. For instance, Faraway and Chatfield applied airline data to various neural network models, comparing them with forecasts from Box–Jenkins and Holt-Winters models, highlighting the potential for improved neural network performance with careful input selection [17]. Ahmed et al. conducted a comparative analysis of various machine learning models for time series forecasting using the M3 competition dataset [18]. They concluded that Multilayer Perceptron (MLP) and Gaussian Processes (GP) regression had the best model performance in this context. Examining diverse time series forecasting models is crucial for understanding and improving these models, thereby advancing predictive accuracy and enhancing decision-making processes.

As noted, there is a significant absence of comparative analyses on time series forecasting models utilizing temperature data. To address this gap, this study aims to provide a comprehensive review of the performance of several time series forecasting models using identical weather data. As highlighted by Ahmed et al., MLP and GP exhibit superior performance among machine learning models [18]. Additionally, ARIMA has been extensively applied in forecasting various climate time series, including air temperature, drought severity, and wind speed [14,19,20]. Exponential smoothing, a classical model in time series forecasting, remains relevant and is used to predict rainfall, greenhouse gas emissions, and other atmospheric variables [21,22]. Given the enduring relevance of classical models in time series analysis and forecasting, this study will assess the effectiveness of four widely used time series forecasting models:

1. Autoregressive Integrated Moving Average
2. Exponential Smoothing
3. Multilayer Perceptron
4. Gaussian Processes

Due to computational and data storage limitations, this study examines historical trends in maximum and minimum temperatures over a 20-year period (2003–2022) using the four previously mentioned time series forecasting models. This research uses the best-performing forecasting model to project temperatures for the subsequent eight years (2023–2030). In response to concerns about the impact of rising temperatures, the study tests the hypothesis that the rate of temperature change exceeds zero ($\beta_1 > 0$). It is important to note that this hypothesis is tested using data from a 28-year period, which includes the eight forecasted years.

The research will utilize Python [23] and Julia [24] to highlight the role of statistical computing in understanding climate change. First, an overview of the essential characteristics of the proposed statistical methods will be outlined. Next, these models will be applied to climate time series data, and the predictive performance of the proposed models will be summarized. Finally, the optimal statistical model will forecast temperatures for the next seven years, up to 2030.

2. Materials and Methods

2.1. Data and Study Areas

The temperature data used in this study were sourced from the NCEI Climate Data Online (CDO), a repository managed by the NOAA [25]. CDO serves as an archive of global

historical weather and climate data, providing comprehensive summaries of historical daily land surface observations from locations worldwide, spanning from 1 January 1950 to 31 December 2022.

The main focus of this study is forecasting temperatures in highly populated cities with diverse geographical features. The dataset includes two key time series measurements: maximum and minimum daily temperatures, initially recorded in Fahrenheit but converted to Celsius to maintain SI units. Given the dataset's extensive size, which encompasses nearly 75 years of historical data, and limitations in computation and data storage, a mini-batch approach was employed to concentrate on the most recent twenty years of data, specifically from 1 January 2003 to 31 December 2022.

During data preprocessing, missing values were replaced with median temperature values from the time series, as the data exhibited a heavily skewed distribution. The dataset was split into a training set and a testing set in an 80% to 20% ratio to facilitate comparative analysis of the time series forecasting models. This split resulted in the training set comprising approximately 16 years of observations, while the testing set encompassed the most recent four years, from 2019 to 2022. Temperature data will be forecasted for the next eight years, from 1 January 2023 to 31 December 2030, using the best-performing forecasting method identified.

To select highly populated cities, it is essential to consider the diversity of geographical features, such as coastal or mountainous regions. The Census Bureau subdivides the United States into four census regions—West, South, Midwest, and Northeast—each with two or more census divisions (Figure 2) [26].

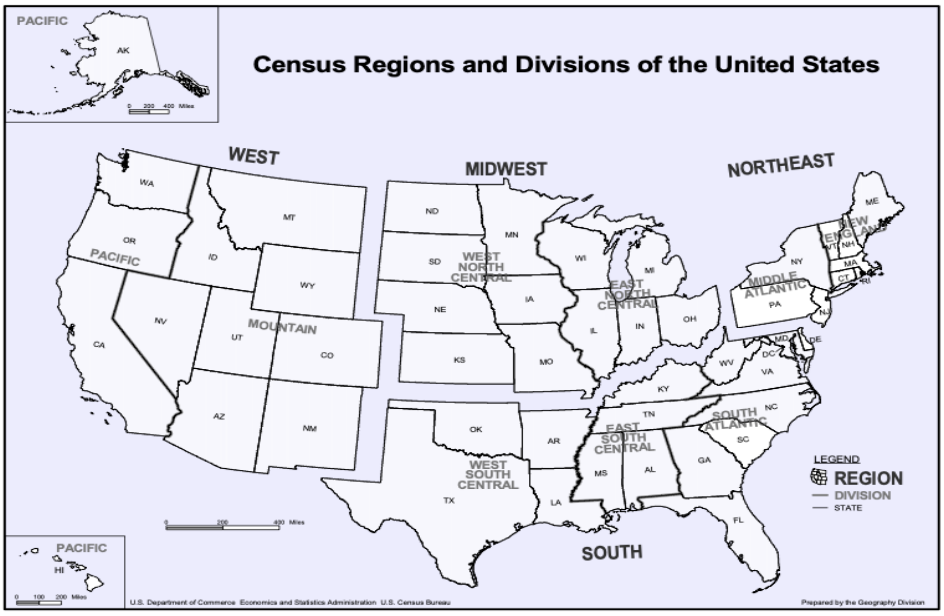

Figure 2. Map of the census regions and divisions of the United States [26].

To effectively represent the regions, at least one site was selected from each region. This study includes data from five airport locations across the United States. The chosen airport sites, which represent various regions, were as follows:

- Houston
- Chicago
- Boston
- San Francisco
- Miami

The Köppen-Geiger Climate Classification system categorizes regions into classes and subclasses based on climate variations, such as precipitation and air temperatures [27].

According to the current map, Houston is classified as Cfa, indicating a humid subtropical climate characterized by hot, long summers with evenly distributed precipitation throughout the year. Chicago is classified as Dfa, representing a hot-summer humid continental climate with hot and humid summers and severe winters. While Boston shares the same classification as Chicago, its coastal location makes it susceptible to hurricane seasons. San Francisco is classified as Csb, a dry-summer subtropical climate, also known as a Mediterranean climate, featuring dry, hot summers and mild winters. Miami is classified as Am, a tropical monsoon climate, experiencing short, dry winters and hot, humid summers. These classification symbols and their descriptions are detailed in several research papers [27–29].

2.2. Time Series Forecasting Models

In this research, Julia is the primary programming language used to build time series forecasting models for application to temperature datasets. However, Julia lacks the necessary packages for constructing MLP models. Therefore, Python, specifically the PyTorch package [30], is utilized to develop the MLP framework and apply the temperature datasets to the model. For the other three time series forecasting models, several Julia packages are employed, including Forecast, StateSpaceModels, and GaussianProcesses [31–33].

The computing device used for this research is designed to ensure efficient processing and data handling. The JupyterLab environment operates on a Tanzu Kubernetes cluster, deployed on a physical Dell cluster. This setup includes 128 CPU cores, 2 TB of memory, and 224 TB of NVMe disk storage, providing the necessary resources for computational tasks. Each CPU core supports 64 threads, with a base clock speed of 2.8 GHz and the ability to boost up to 3.7 GHz, ensuring robust performance during data processing and model training.

2.2.1. Autoregressive Integrated Moving Average

The Autoregressive Integrated Moving Average (ARIMA) model, popularized by Box and Jenkins [34], is a combination of three components: the autoregressive (AR) model, the integrated (I) method, and the moving average (MA) model. For a stationary time series, the Autoregressive Moving Average (ARMA) model can be expressed as:

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}. \quad (1)$$

In this equation, X_t represents the value of the time series at time t, determined by its past values $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$, with $\phi_1, \phi_2, \ldots, \phi_p$ as autoregressive coefficients. It also incorporates past prediction errors $\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-q}$, with $\theta_1, \theta_2, \ldots, \theta_q$ as moving average coefficients [35]. The constant term μ accounts for a non-zero mean in the time series, while ϵ_t represents the error term at time t. The integrated method addresses non-stationarity in the time series through differencing. The ARIMA model is typically denoted as ARIMA (p, d, q), where p is the order of the AR model, d is the degree of differencing, and q is the order of the MA model.

As previously mentioned, the ARIMA model is a form of the Box–Jenkins model, a mathematical model that uses regression analysis to fit and forecast time series data. The Box–Jenkins method involves an iterative three-step process for developing a forecasting model for the time series:

1. Identification stage: Ensuring that the time series is stationary. The Augmented Dickey–Fuller (ADF) test determines whether the time series is stationary. The autocorrelation function (ACF) and the partial autocorrelation function (PACF) are used to identify the appropriate AR, MA, or ARMA models. The ACF assesses the correlation across all lags or intervals between observations, while the PACF focuses on the direct correlation at specific lags, accounting for the influence of intermediate lags.
2. Estimation Stage: Estimating the model parameters using goodness-of-fit tests to determine p, d, and q. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are commonly used for model selection:

$$\text{AIC} = 2k - 2\ln(L) \text{ and BIC} = k\ln(n) - 2\ln(L). \quad (2)$$

The AIC measures the information value of the model using maximum likelihood estimates, denoted as L, and the number of parameters in the model, denoted as k. The BIC is similar to the AIC but includes a larger penalty term for the number of observations, denoted as n.

3. Forecasting Stage: Using the fitted model to forecast future values of time series.

2.2.2. Exponential Smoothing

Exponential smoothing (ETS) is a classical time series forecasting method that estimates future values based on weighted averages of past observations, with the assumption that more recent observations carry greater significance than older ones. Initially proposed by Brown in 1956 [36], ETS has since expanded and refined into various exponential smoothing models. For instance, Holt and Winters extended Brown's simple exponential smoothing to create the Triple Exponential Smoothing (TES) model, also known as the Holt–Winters method [37]. The TES incorporates additional components to account for trend and seasonality. The forecasting equation for TES can be expressed as:

$$F_{t+n} = L_t + nT_t + S_{t-p+n}, n = 1, 2, 3, \ldots \quad (3)$$

where L_t represents the level component, nT_t denotes the trend component, S_{t-p+n} represents the seasonal component, and n represents the number of future periods ahead of the current time period t. Given the presence of seasonal patterns observed in maximum and minimum temperatures, this study employs the TES model as the primary exponential smoothing technique.

2.2.3. Multilayer Perceptron

The Multilayer Perceptron (MLP) is a widely used type of feedforward neural network in machine learning applications. It consists of an input layer, one or more hidden layers with densely connected neurons utilizing nonlinear activation functions, and an output layer. The training algorithm for an MLP typically involves backpropagation, which adjusts the network's weights to minimize the difference between predicted and actual outputs of training data [38].

In this study, the design of the MLP model is a simple, straightforward structure consisting of three essential layers: an input layer, a single hidden layer, and an output layer. The design of this model prioritizes simplicity to address computation and data storage limitations. The MLP architecture begins with an input layer that captures a specific number of lagged observations of the temperature variable, which determines the number of neurons in the input layer. Determining the number of input neurons is based on the frequency of shifts and trends observed in the temperature dataset, with further details provided in the Results section (Section 3.1.3). Following the input layer, the hidden layer consists of four neurons. Given the constraints of the study's limited computational resources, the number of neurons was selected to balance model simplicity and computational efficiency. A smaller number of neurons helps reduce overfitting and computational demands while allowing the model enough flexibility to capture meaningful patterns in the temperature data. The MLP model concludes with a single output neuron in the output layer, which generates forecasting future temperature values.

The model uses the Rectified Linear Unit (ReLU) activation function, which sets any negative values to 0. ReLU is chosen for its computational efficiency and ability to introduce a simple nonlinear transformation. Additionally, several research studies on forecasting air temperatures using deep learning neural networks have employed the ReLU activation function due to its effectiveness in capturing complex patterns while maintaining simplicity in computation [39]. To mitigate overfitting in the MLP model, a regularization technique known as dropout is implemented. Dropout randomly ignores selected neurons

during training, which reduces the model's reliance on specific neuron weights. A common dropout rate of $p = 0.2$ is used in each hidden layer, as applied in this study.

An optimizer is crucial for enhancing the MLP model's performance by optimizing hyperparameters to minimize losses. Adaptive Moment Estimation, or Adam for short, is one of the most commonly stochastic optimization algorithms for neural networks. Adam maintains exponentially decaying averages of past squared gradients based on the first and second moments of the loss function gradients [40]. This study employs Adam with a learning rate of 0.001, a common default for many optimization algorithms. The learning rate determines the step size for updating the model's weights based on the computed gradients.

The MLP model includes a loss function to estimate the model's losses, allowing the updating of parameters to minimize the losses. Given that the MLP model predicts real values, it addresses a regression predictive modeling problem. Consequently, the default loss function employed in this study is the Mean Squared Error (MSE).

2.2.4. Gaussian Processes

Gaussian Processes (GP) are highly effective tools for time series analysis, offering a flexible and nonparametric method to model complex and nonlinear relationships in data. A GP is essentially a collection of random variables where any finite subset has a joint Gaussian distribution [41]. In the context of time series analysis, a GP can be interpreted as a probability distribution over the set of all possible time series functions, with the mean and covariance functions defining the properties of the GP.

Assuming that the underlying function generating the time series values at time t is denoted by f_t, GP regression models the function f_t as a Gaussian process with a mean function μ_t and covariance function Σ_t. In this study, the mean function is set to zero, which is a standard choice in time series forecasting due to the uncertainty about the magnitude and direction of the trend [42]. Suppose the time series X_t is given with the observed values up to time t, i.e., $\mathcal{D}_t = X_1, X_2, \ldots, X_t$. The conditional distribution of X_{t+1} given $\mathcal{D}_t$ is then:

$$p(X_{t+1} \mid \mathcal{D}_t) \quad = \quad \mathcal{N}(0, \Sigma_{t+1}). \tag{4}$$

The posterior covariance is computed as:

$$\Sigma_{t+1} \quad = \quad k(X_{t+1}, X_{t+1}) - k_{t+1,t}^{\top}(K_t + \sigma^2 I)^{-1} k_{t+1,t}. \tag{5}$$

Here, $k_{t+1} = [k(X_1, X_{t+1}), k(X_2, X_{t+1}), \ldots, k(X_t, X_{t+1})]^{\top}$, K_t is the $t \times t$ matrix with entries $K_{i,j} = k(X_i, X_j)$, σ^2 is the noise variance, and I is the identity matrix. The kernel functions $k(X_i, X_j)$ specify the correlation between the predicted and observed time series values.

GPs often employ a kernel composition approach, combining multiple kernels through addition or multiplication to shape the resulting distribution. In this study, the GP kernel composition is formulated as:

$$k(X_i, X_j) = k_{Noise}(X_i, X_j) + k_{Periodic}(X_i, X_j) + k_{RBF}(X_i, X_j) \tag{6}$$

Here, the noise kernel, $k_{Noise}(X_i, X_j)$, captures the noise aspects of the time series, the periodic kernel, $k_{Periodic}(X_i, X_j)$, models seasonal patterns, and the RBF (Radial Basis Function) kernel, $k_{RBF}(X_i, X_j)$, accounts for nonlinear trends [43]. This proposed composition allows the GP to flexibly model various components of the time series, integrating noise characteristics, periodic effects, and nonlinear trends into its predictive model.

2.2.5. Model Selection Criterion

This study evaluates model performance using testing datasets to identify the optimal time series forecasting model. Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are key metrics for forecast accuracy, which provide more meaningful and accurate

evaluations when forecasts are on the same scale [44]. Given that forecast temperatures are consistently measured, the performance of the four different time series forecasting models is assessed using RMSE (Equation (7)) and MAE (Equation (8)).

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(X_{t,i} - \hat{X}_{t+1,i})^2}{N}} \tag{7}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|X_{t,i} - \hat{X}_{t+1,i}| \tag{8}$$

In these equations, $X_{t,i}$ represents the actual observed values, $\hat{X}_{t+1,i}$ represents the predicted values at time $t+1$, and N represents the number of observations in the testing dataset. These statistical measures provide a quantitative assessment of the accuracy and performance of the forecasting models by quantifying the differences between the predicted and actual values.

3. Results

Four different time series forecasting models were applied to forecast the maximum and minimum temperatures of five highly populated cities in the United States, using a training set from 2003 to 2018 and a testing set from 2019 to 2022. This section is divided into three main parts: the first examines the optimal hyperparameters for each model through various tests and optimizations; the second provides a comparative analysis of the forecasting performance of the four models, evaluating each technique's effectiveness in temperature forecasting; and the third projects temperatures from 2023 to 2030, offering insights into future trends based on the chosen models.

3.1. Hyperparameters Optimization

3.1.1. Autoregressive Integrated Moving Average

Prior to developing the ARIMA model, it is critical to ensure that the temperature variables are stationary to avoid misleading results, a common challenge in time series analysis. The ADF test, as shown in Table 1, is used to check for stationarity. The *p*-values for the temperature time series across all five cities are well below the standard significance level of 0.05. Therefore, we reject the null hypothesis, indicating that the time series are stationary. This implies that there is no need to add a differencing order to the time series data.

Table 1. Results of the Augmented Dickey–Fuller (ADF) test for stationarity on minimum and maximum temperatures, based on raining datasets of five cities.

Variables	Measures	Houston	Chicago	Boston	San Francisco	Miami
TMAX	ADF Values	−9.16	−7.74	−8.28	−13.76	−10.53
	p-values	$<1 \times 10^{-14}$	$<1 \times 10^{-10}$	$<1 \times 10^{-12}$	$<1 \times 10^{-24}$	$<1 \times 10^{-18}$
TMIN	ADF Values	−9.08	−7.50	−6.58	−10.41	−11.80
	p-values	$<1 \times 10^{-14}$	$<1 \times 10^{-10}$	$<1 \times 10^{-8}$	$<1 \times 10^{-17}$	$<1 \times 10^{-21}$

The ACF and PACF are critical statistical measures for analyzing time series data and identifying the appropriate ARIMA model type. Figure 3 illustrates the ACF and PACF plots for Houston's maximum and minimum temperatures, with a maximum of 25 lags. The plots include black straight and dashed lines representing the thresholds for the 99% and 95% confidence intervals, respectively. ACF or PACF values falling outside these intervals suggest statistical significance and correlation. The PACF plots use a step-wise approach with two estimation methods: the blue bars represent the 'step' method, which estimates AR parameters for an expanding model, while the red bars represent the 'real' method, which provides the true PACF by removing the linear effects of previous lags [31]. The ACF plots indicate a slow decline with increasing lags, whereas the PACF plots show a few significant

lags. Therefore, the patterns indicate that Houston's temperatures are well-suited for a pure AR model. The ACF and PACF plots for the other four cities exhibit similar behaviors, leading this study to focus solely on the pure AR model, as presented in Appendix A.1.

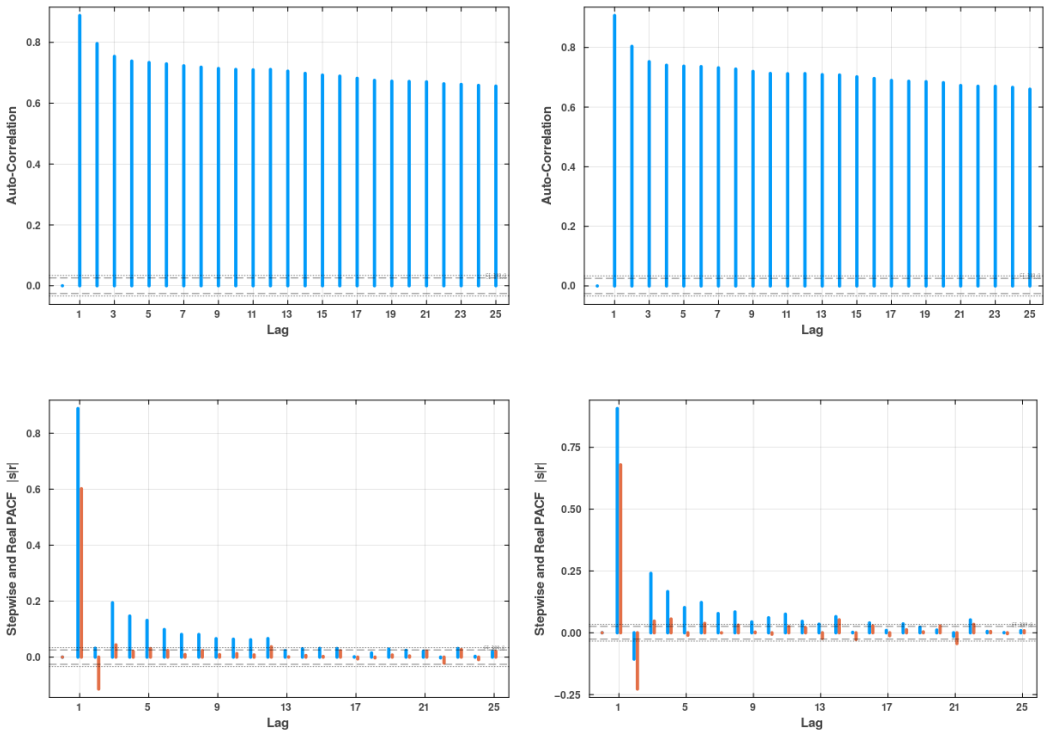

Figure 3. Autocorrelation Function (ACF) (**top**) and Partial Autocorrelation Function (PACF) (**bottom**) for Houston's maximum (**right**) and minimum (**left**) temperatures, based on the training dataset.

In this study, a simple optimization method was employed to determine the optimal lag length for the AR model of the time series observations, evaluated using the AIC and BIC (Equation (2)). According to Table 2, it is evident that the optimal lag order of the pure AR model ranges from lag 80 to 110 based on the AIC score for all five cities, except for San Francisco's maximum temperatures, which have an optimal lag order of 19. Although it may seem unusual for San Francisco to have a lower lag order than other cities, this is likely due to the unique characteristics of the maximum temperature data for San Francisco compared to other cities.

Table 2. Model order selection for Autoregressive Integrated Moving Average (ARIMA) models of maximum and minimum temperatures across five cities.

Variables	City	AIC Score	BIC Score	Model
TMAX	Houston	3.38	20,558.26	ARIMA (88,0,0)
	Chicago	4.04	24,473.82	ARIMA (109,0,0)
	Boston	4.06	24,590.41	ARIMA (108,0,0)
	San Francisco	3.08	18,412.60	ARIMA (19,0,0)
	Miami	2.30	14,700.06	ARIMA (100,0,0)
TMIN	Houston	3.28	20,070.57	ARIMA (101,0,0)
	Chicago	3.66	22,267.02	ARIMA (106,0,0)
	Boston	3.23	19,856.57	ARIMA (108,0,0)
	San Francisco	1.95	12,845.68	ARIMA (97,0,0)
	Miami	2.55	16,098.39	ARIMA (106,0,0)

3.1.2. Exponential Smoothing

As stated in Section 2.2.2, the TES model requires trend and seasonal components to perform forecasting. This study used the seasonal-trend decomposition procedure based on loess (STL) to identify the trend and seasonality [45]. Figure 4 shows the time series decomposition, where the STL method divides the time series observations into data (top), trend (first middle), seasonal (second middle), and remainder (bottom). The decomposition reveals that the time series has seasonal components with a periodicity occurring every year from 2003 to 2018. However, no trend is present in either the maximum or minimum temperatures. The observed seasonal and trend components in Houston's daily temperatures are similar to those in the other four cities, as detailed in Appendix A.2.

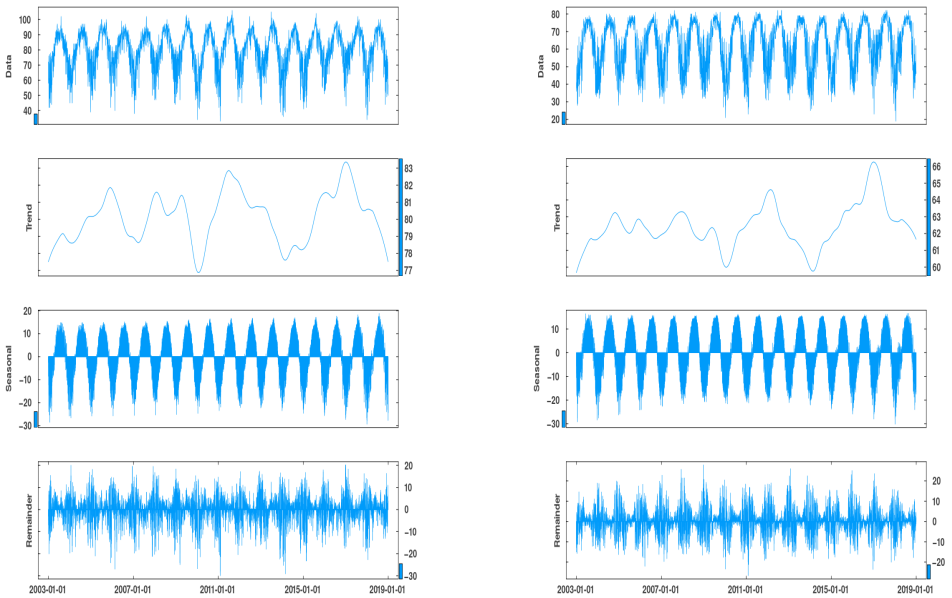

Figure 4. Plots of time series decomposition for Houston's maximum (**right**) and minimum (**left**) temperatures, based on the training dataset.

This study utilizes the StateSpaceModels package in Julia to fit the ETS model with temperature data. This model includes boolean parameters for trend consideration, damped trends, and an integer parameter for the number of periods as seasonal components [32]. Given the seasonal components revealed by the time series decomposition and the 16-year training set (2003 to 2018), the ETS model, specifically the Holt–Winters method, is employed with a periodicity of 16 seasons without considering a trend.

3.1.3. Multilayer Perceptrons

Prior to training the MLP model, the training set is split into sequences of three observations, which serve as input neurons for the model. As shown in the data section of Figure 4, temperatures can fluctuate rapidly over a few days rather than showing a consistent trend. By splitting the training set into sequences of three observations, the model can better capture short-term variations or shifts in temperatures occurring over short intervals, thus improving its predictive accuracy.

The MLP model is trained using a batch size, which specifies the number of samples processed before updating the model's internal parameters. In this study, both the training and testing datasets are divided into batches of 12 samples for iterative parameter updates. Given that the MLP model employs the Adam optimizer, the batch size is a crucial hyperparameter for training models with gradient-based optimizers. A smaller batch size is preferable when computational memory is limited, as it can also help prevent overfitting. During training, the algorithm iterates over the entire dataset multiple times, a process known as epochs. To ensure effective minimization of the loss function, the model undergoes 2000 epochs of training to optimize its parameters.

3.1.4. Gaussian Processes

In the previous section, we established the kernel composition (Equation (6)) for the GP model. Optimal hyperparameter values for each kernel were determined through iterative trials, aiming to minimize the RMSE, as detailed in Table 3. During these trials, adjustments were made to avoid issues such as infinite sine values in the periodic kernel. The selected optimal hyperparameters, listed in the first row of Table 3, are as follows: a noise kernel with a standard deviation of 1.0; a periodic kernel with a length scale of 0.0, a standard deviation of 1.0, and a periodicity of 1.0; and an RBF kernel with a length scale of 4.0 and a standard deviation of 0.0.

Table 3. Trials and errors with values for kernel composition in Gaussian Processes (GP), based on Houston's maximum temperatures.

Kernel Composition	RMSE Values
N(1) + P(0,1,1) + RBF(4,0)	12.51
N(1) + P(0,1,1)	12.81
P(0,1,1) + RBF(4,0)	12.81
N(1) + P(0,1,2) + RBF(4,0)	Failed due to infinite sin values
N(1) + P(0,1,0) + RBF(4,0)	Failed due to infinite sin values
N(1) + P(0,1,1) + RBF(2,0)	12.80
N(1) + P(0,1,1) + RBF(5,0)	12.83
N(1) + P(0,1,1) + RBF(4,4)	76
N(1) + P(0,1,0) + RBF(4,4)	76
N(1) + P(0,1,1) + RBF(3,0)	12.86
N(1) + P(1,1,1) + RBF(4,0)	Failed due to infinite sin values

3.2. Comparative Forecasting Performance

The forecasting results of the four models, along with the testing set of observed values from the NCEI's CDO, are presented in Figure 5 for maximum temperatures and Figure 6 for minimum temperatures. The x-axis of the figures represents the dates from 1 January 2019 to 31 December 2022, corresponding to the size of the testing set. Note that the maximum and minimum temperatures are measured in Celsius. The figures highlight that the predicted values of the MLP model closely align with the actual observations, in contrast to the predictions from the other three models.

Tables 4 and 5 display the model performance metrics for the four models. Based on RMSE and MAE, it is evident that the MLP model consistently demonstrates superior forecasting accuracy for both maximum and minimum temperatures across the five cities. The overall ranking of the four models based on RMSE and MAE is as follows: MLP, ARIMA, ETS, and GP. However, when specifically assessing the performance in Houston and Boston's maximum temperatures, the GP model outperforms the ETS model.

Table 4. Performance metrics for forecasting maximum temperatures across five cities.

Model	Measures	Houston	Chicago	Boston	San Francisco	Miami
ARIMA	RMSE	11.24	16.75	13.74	8.62	5.45
	MAE	8.90	14.10	11.33	6.69	4.29
ETS	RMSE	28.14	30.28	25.54	14.62	7.11
	MAE	25.63	25.09	20.87	12.07	5.92
MLP	RMSE	6.09	8.24	8.25	4.99	3.54
	MAE	4.42	6.33	6.53	3.81	2.29
GP	RMSE	12.51	-	17.94	-	-
	MAE	10.36	-	15.24	-	-

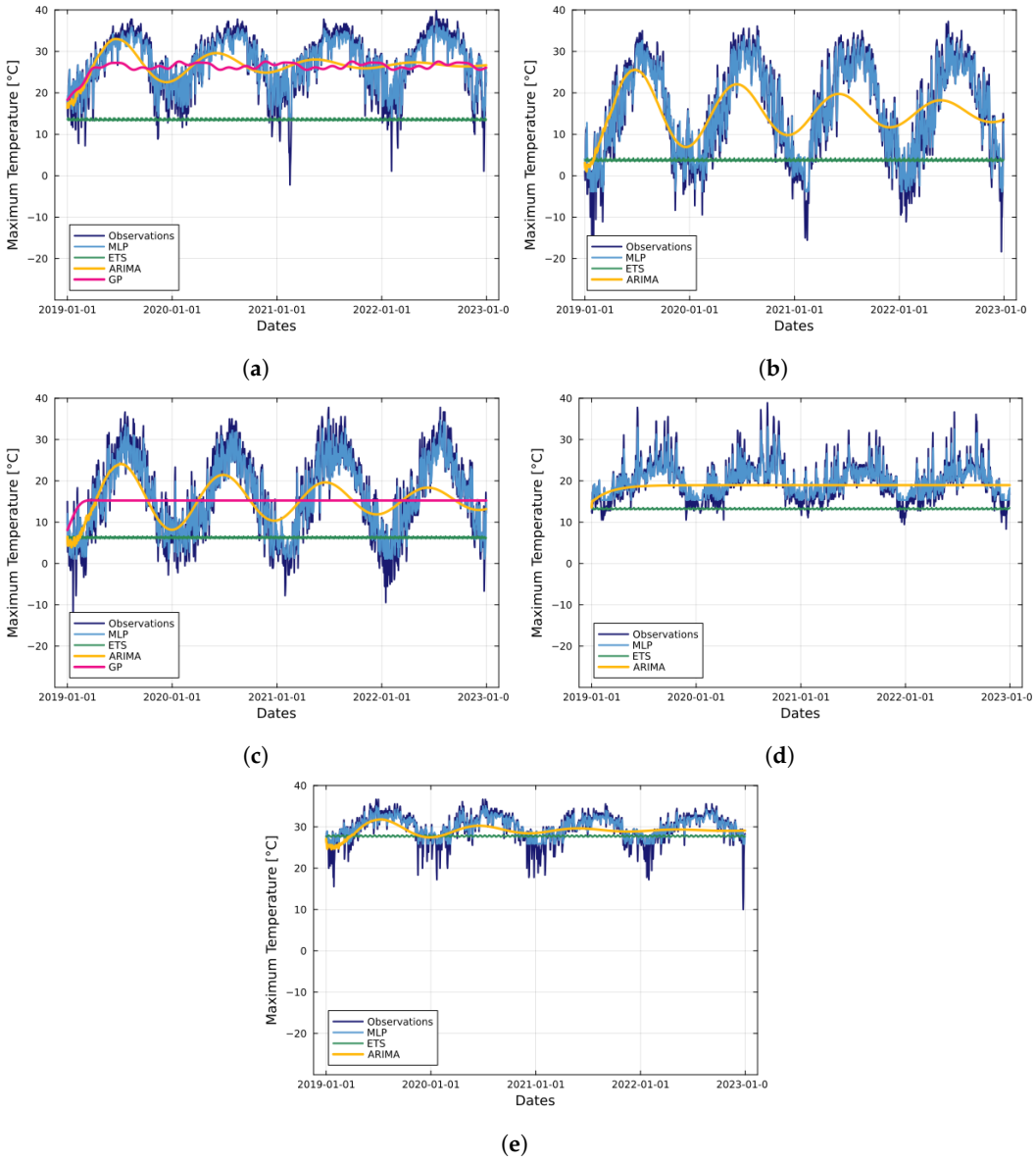

Figure 5. Forecasting results for maximum temperatures (in Celsius) across five cities: (**a**) Houston, (**b**) Chicago, (**c**) Boston, (**d**) San Francisco, and (**e**) Miami. Each subfigure displays the observed values from the testing set along with predictions from four time series forecasting models: MLP, ARIMA, ETS, and GP.

Table 5. Performance metrics for forecasting minimum temperatures across five cities.

Model	Measures	Houston	Chicago	Boston	San Francisco	Miami
ARIMA	RMSE	11.60	16.75	11.59	5.50	6.82
	MAE	9.65	12.90	9.43	4.53	5.48
ETS	RMSE	21.50	23.67	22.52	9.41	7.32
	MAE	18.65	19.29	18.48	8.00	5.56
MLP	RMSE	5.70	6.56	5.42	2.81	4.00
	MAE	4.00	5.01	4.07	2.13	2.80
GP	RMSE	-	-	-	-	-
	MAE	-	-	-	-	-

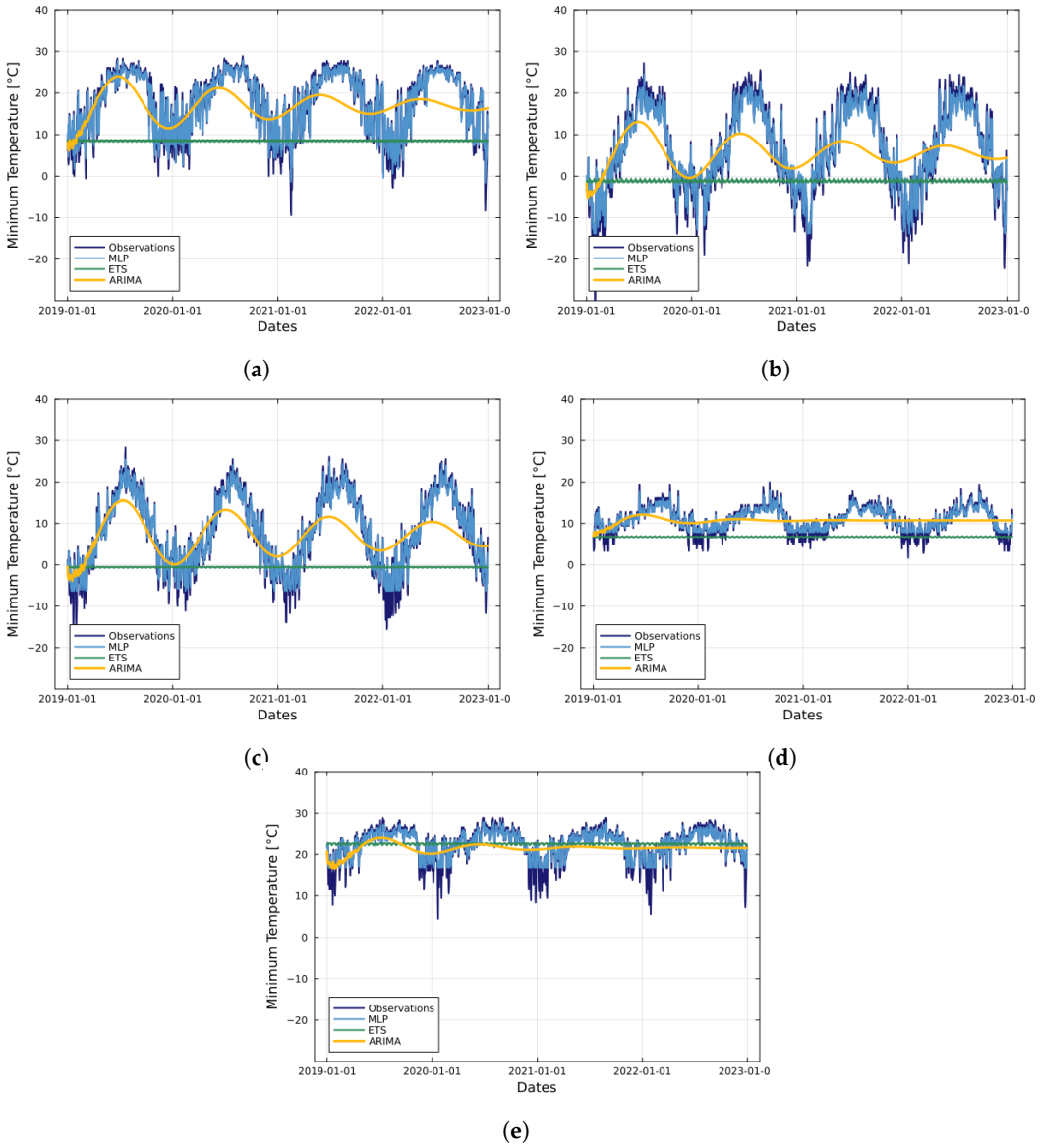

Figure 6. Forecasting results for minimum temperatures (in Celsius) across five cities: (**a**) Houston, (**b**) Chicago, (**c**) Boston, (**d**) San Francisco, and (**e**) Miami. Each subfigure displays the observed values from the testing set along with predictions from four time series forecasting models: MLP, ARIMA, ETS, and GP.

3.3. Temperature Projections

The analysis of the figures and the evaluation of model performance indicate that the MLP model outperforms the other three forecasting models. Figures 7 and 8 display the projections of maximum and minimum temperatures for all five cities using the MLP model. These figures cover the period from 1 January 2003 to 31 December 2030, incorporating observed data from 2003 to 2022 and forecasting temperatures for the subsequent 8 years. As with previous figures, the maximum and minimum temperatures are measured in Celsius.

Each graph in Figures 7 and 8 includes a regression equation displayed in the bottom left corner of each subfigure. The regression equations show that the coefficient β is consistently greater than 0, indicating a positive rate of change in temperatures. Notably, the maximum temperature time series for Chicago and Boston has the highest β value, while Chicago exhibits the highest β value for the minimum temperature time series. Additionally, the amplitude of periodic variation in maximum and minimum temperatures is significantly reduced compared to the actual observed values.

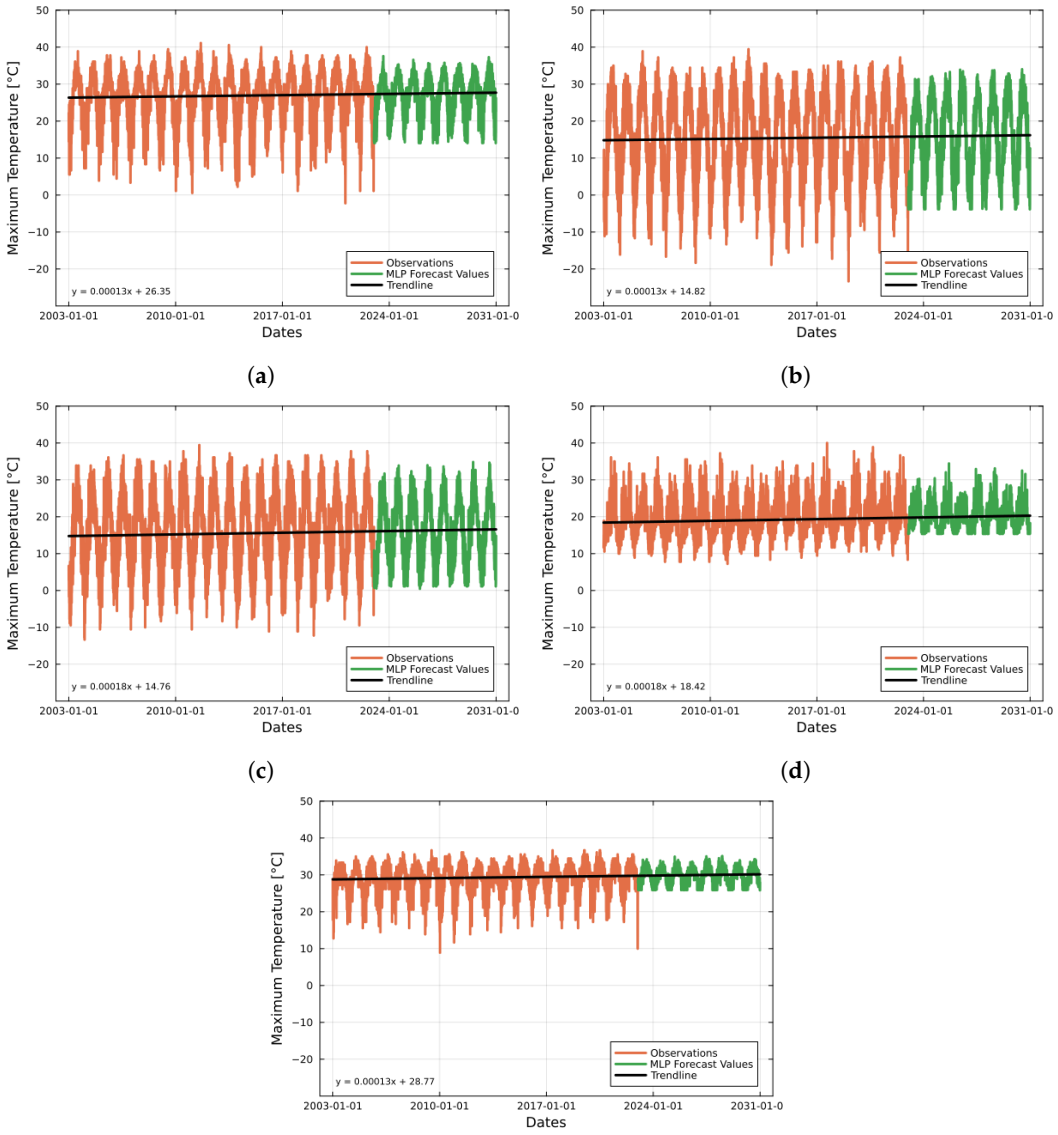

Figure 7. Projected maximum temperatures (in Celsius) for five cities: (**a**) Houston, (**b**) Chicago, (**c**) Boston, (**d**) San Francisco, and (**e**) Miami, using the MLP model. Each subfigure displays observed and forecasted temperatures.

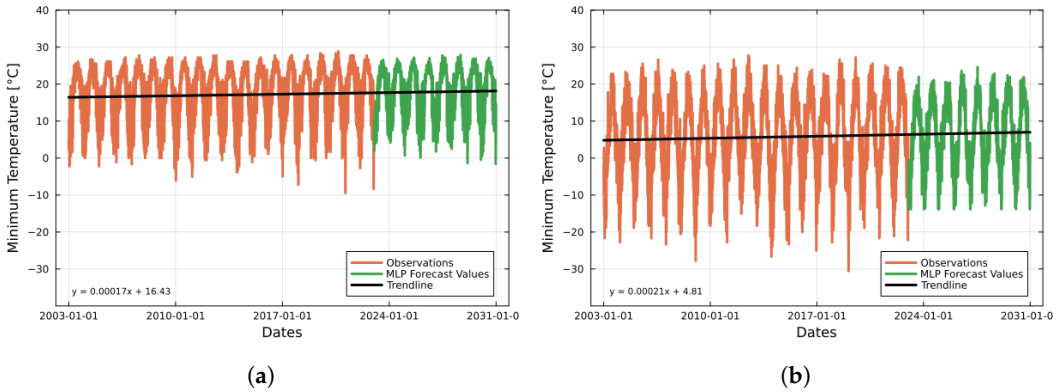

Figure 8. *Cont.*

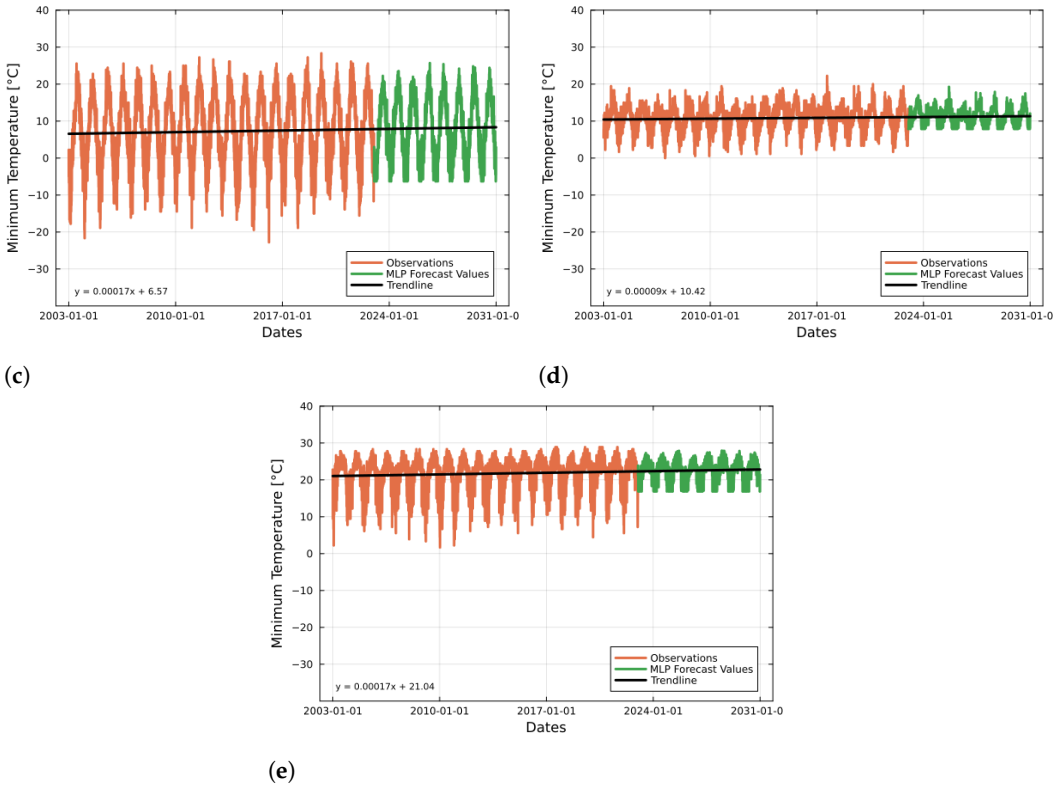

Figure 8. Projected minimum temperatures (in Celsius) for five cities: (**a**) Houston, (**b**) Chicago, (**c**) Boston, (**d**) San Francisco, and (**e**) Miami, using the MLP model. Each subfigure displays observed and forecasted temperatures.

4. Discussion and Conclusions

The time series forecasting models used in this study are based on foundational concepts in machine learning and statistical techniques. While these methods have been refined and evolved over time, the core principles and algorithms remain rooted in traditional advancements. Despite rapid advancements in time series forecasting, understanding the application and suitability of each method to specific datasets and their characteristics remains crucial.

In this paper, we explored the feasibility of four time series forecasting models—ARIMA, ETS, MLP, and GP—for temperatures in five highly populated U.S. cities. Each city's temperature range is influenced by geographic features such as its coastal or mountainous location. We observed significant seasonal components but no trends in temperature across all cities. Unlike the other three time series forecasting models, the parameter values of ARIMA varied in each city, reflecting the diverse temperature ranges, while the other three models were adjusted based on temperature characteristics.

The MLP model demonstrated the highest accuracy for predicting maximum and minimum temperatures across all cities, based on RMSE and MAE. The ARIMA model performed second best, followed by ETS and GP. However, as presented in the results, the MLP model struggled with capturing extreme temperatures and the magnitude of temperature amplitude. For instance, the MLP model had difficulty capturing extreme low temperatures in Miami, as shown in Figures 5 and 6.

While the MLP model demonstrated the best performance, the other three models exhibited poor forecasting results. The pure AR model performed well initially, but the amplitude of the predicted values declined in later stages. This reduction in amplitude could be due to several factors, such as AR coefficients approaching zero, overfitting, or inherent noise in the data. The ETS model, on the other hand, appeared nearly constant with a very small amplitude when zoomed in, which might be due to the chosen number of periods. Although the time series decomposition suggested 16 periods, the dataset

consisted of daily observations, so a seasonal period of 365 might be more suitable for capturing annual patterns. Due to the computation and data storage limitations, running the ETS model with 365 periods was not feasible.

Upon review of these tables and figures, it is clear that the fixed kernel composition of the GP model failed to produce valid predictions for the maximum and minimum temperatures, except for the maximum temperature of Houston and Boston. This is due to instability and ill-conditioning in GP models, which can lead to infinite values in the sine function of the periodic kernel and result in errors during matrix inversions, rendering the fixed kernel composition ineffective for these cities [46]. As mentioned, if we consider the maximum temperatures of only two cities (Houston and Boston), the GP model outperformed the ETS model. Based on the figures and tables, we can conclude that MLP and ARIMA models effectively captured the periodic characteristics of temperature variables, unlike the ETS and GP models.

While this paper suggests that the MLP model is the most effective for time series forecasting of temperatures, it is important to acknowledge certain limitations in the model development. This study is limited by its focus on a single structure, optimization method, and activation function for the MLP model. Alternative approaches, such as the Levenberg–Marquardt (LM) algorithm, are utilized by some researchers for error-correction learning in time series analysis involving periodic components [47,48]. Additionally, the study is constrained by the use of a single fixed kernel composition for the GP model. There are over 50 potential kernel combinations for forecasting time series; however, the Julia programming language currently lacks an optimization method to determine the best kernel composition or to identify optimal hyperparameters based on the dataset.

Limited access to systems with higher computational power is a significant limitation of this study. Time series forecasting models, particularly ETS, MLP, and GP, have optimal hyperparameters that may require advanced computing resources, such as supercomputers, for more efficient forecasting. Additionally, increasing the dataset size might enhance model performance, but the constraints on computational storage made this unfeasible. Consequently, this research focused on a 20-year segment of the dataset due to available computational resources. Access to high-performance computing systems could potentially yield better results.

Finally, the ARIMA, ETS, and GP models were implemented using Julia, whereas the MLP model was developed with Python. Julia, being a relatively newer programming language compared to Python, is still evolving, with continuous innovation and the development of new models and packages for time series forecasting. In contrast, Python is a well-established language with a wide array of packages. These differences in the maturity and ecosystems of the two languages may have influenced the performance and optimization capabilities of the models.

It is important to note that time series forecasting models primarily rely on historical data and do not account for the underlying physical processes, such as thermodynamics or fluid dynamics, that influence weather patterns. Additionally, these models assume constant impacts of climate change, which limits their ability to predict future conditions that might diverge significantly from past trends, such as potential reductions in greenhouse gas emissions. Therefore, it is crucial to acknowledge the uncertainties inherent in time series forecasting and the influence of uncontrolled variables when projecting future weather patterns.

Moreover, time series analysis includes a wide range of methodologies that extend beyond the scope of this study, and statistical models continue to advance. Future research could include exploring additional machine learning methods for climate time series forecasting, such as Long Short-Term Memory (LSTM) networks, Support Vector Machines (SVM), and Convolutional Neural Networks (CNN) [49–51]. Expanding datasets and incorporating various forecasting models could uncover optimal methods for different scenarios, suggesting that no single model is universally applicable.

According to the temperature projections, the performance of the MLP model supports the hypothesis that temperatures exhibit a positive rate of change with $\beta > 0$. Although the rate of change may be small, it shows a consistent upward trend over time. This study analyzes a 28-year period, which is considerably shorter than the IPCC's 150-year period analysis from the pre-industrial era to the present. Consequently, the rate of increase in maximum and minimum temperatures might differ if a longer time frame, including the pre-industrial era, were considered.

However, the rate of change in maximum and minimum temperatures varies across the five cities. For example, the rate of change in minimum temperature is 0.00009 in San Francisco, while in Chicago, it is 0.00021. Additionally, the rate of change differs between the maximum and minimum temperatures across the five cities. In San Francisco, for instance, the minimum temperature shows a much lower rate of change compared to the maximum temperature. Although the rate of change is small, these differences suggest that policy targets may not be universally applicable across all regions, as the indicators and relationships between global temperature changes and their regional impacts differ [52].

While this paper focuses on a subset of time series forecasting techniques, many other methods effective for different weather patterns were not included. It is crucial to continue innovating and improving these models to address climate challenges and support decision-making processes regarding climate policies. Future research plans include expanding comparative studies by incorporating additional time series forecasting models and improving hyperparameter optimization. By broadening the scope of comparative studies, we better understand the range of available forecasting models and identify those best suited for specific weather patterns and characteristics, considering factors such as geographic features and population.

Author Contributions: Conceptualization, K.B.K. and E.F.; methodology, K.B.K. and E.F.; software, K.B.K.; validation, K.B.K.; formal analysis, K.B.K.; investigation, K.B.K.; resources, K.B.K. and E.F.; data curation, K.B.K.; writing—original draft preparation, K.B.K.; writing—review and editing, K.B.K. and E.F.; visualization, K.B.K.; supervision, E.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are available upon request.

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

Acronyms

The following acronyms are used in this manuscript:

ADF	Augmented Dickey–Fuller Test
ACF	Auto-Correlation Function
AIC	Akaike Information Criterion
ARIMA	Autoregressive Integrated Moving Average
BIC	Bayesian Information Criterion
ETS	Exponential Smoothing
GP	Gaussian Processes
MAE	Mean Absolute Error
MSE	Mean Squared Error
MLP	Multilayer Perceptron
PACF	Partial Autocorrelation Function
ReLu	Rectified Linear Unit
RMSE	Root Mean Squared Error
TES	Triple Exponential Smoothing

Appendix A

Appendix A.1. ACF and PACF Plots for Maximum and Minimum Temperatures of the Four Remaining Cities

Figures A1–A4 illustrates the ACF and PACF plots for maximum and minimum temperatures in Chicago, Boston, San Francisco, and Miami, in that order. These plots are configured with a maximum lag order of 25, and the black straight and dashed line represent the thresholds for the 99% and 95% confidence interval, respectively. The PACF plots employ a step-wise approach using two estimation methods. The blue bars depict the 'step' method, which estimates AR parameters for an expanding model. In contrast, the red bars represent the 'real' method, which provides the true PACF by accounting for the linear effects of previous lags [31]. Similar to the ACF and PACF plots for Houston's temperatures (Figure 3), the ACF plots exhibit a slow decline, while the PACF plots exhibit a few significant lags at early stages. Consequently, the temperature datasets for all four cities are also well-suited for a pure AR model.

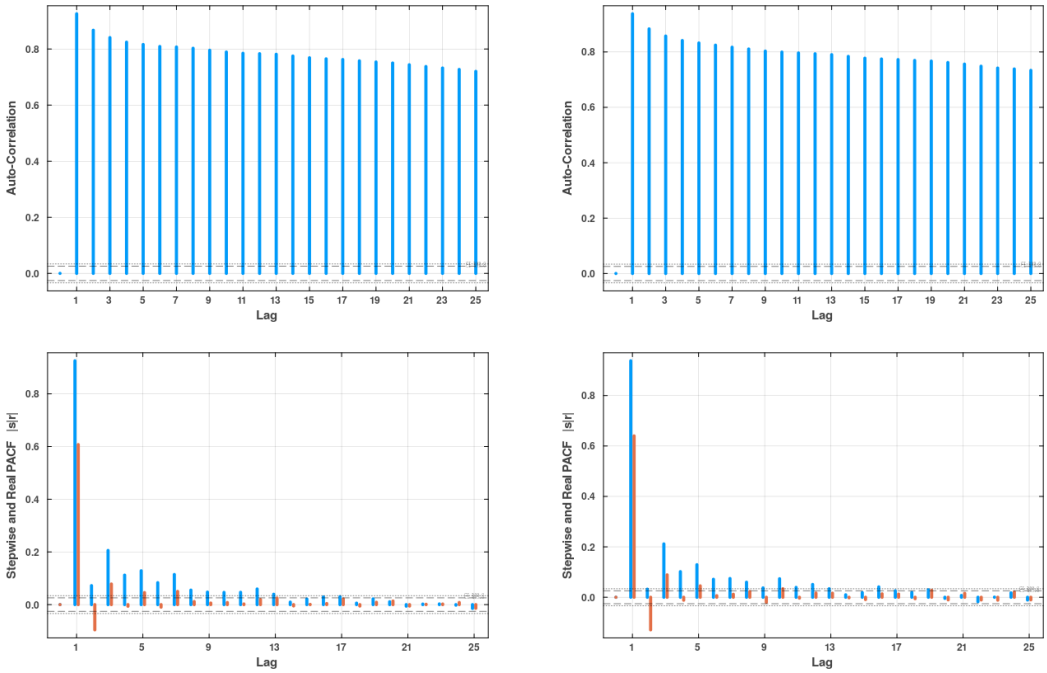

Figure A1. Autocorrelation Function (ACF) (**top**) and Partial Autocorrelation Function (PACF) (**bottom**) for Chicago's maximum (**right**) and minimum (**left**) temperatures, based on the training dataset.

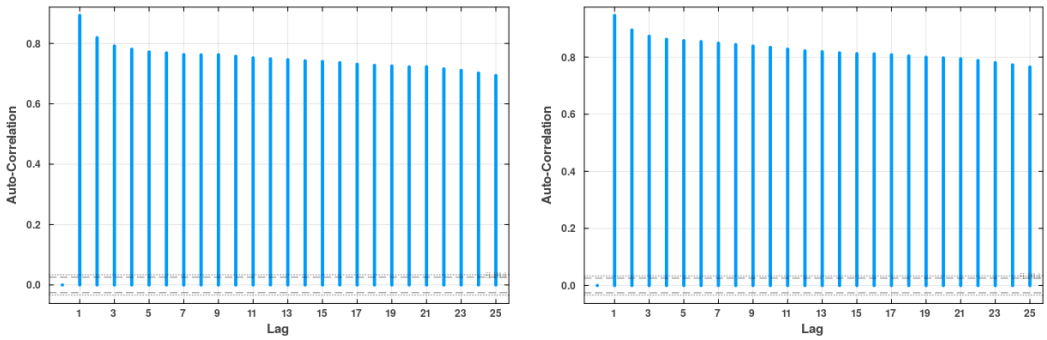

Figure A2. *Cont.*

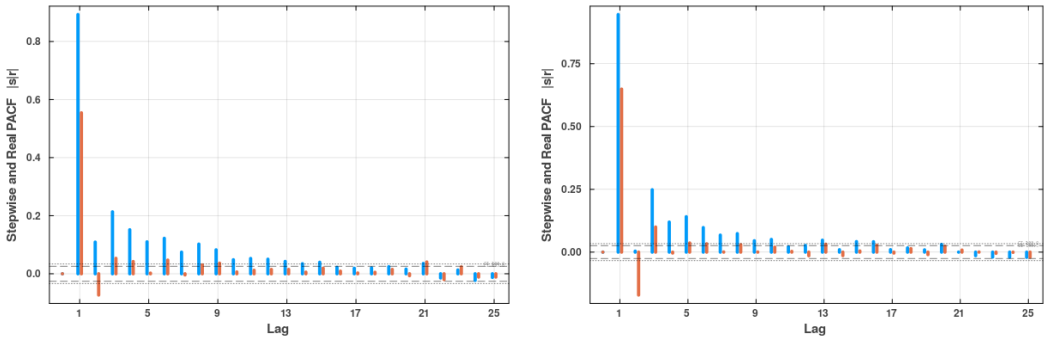

Figure A2. Autocorrelation Function (ACF) (**top**) and Partial Autocorrelation Function (PACF) (**bottom**) for Boston's maximum (**right**) and minimum (**left**) temperatures, based on the training dataset.

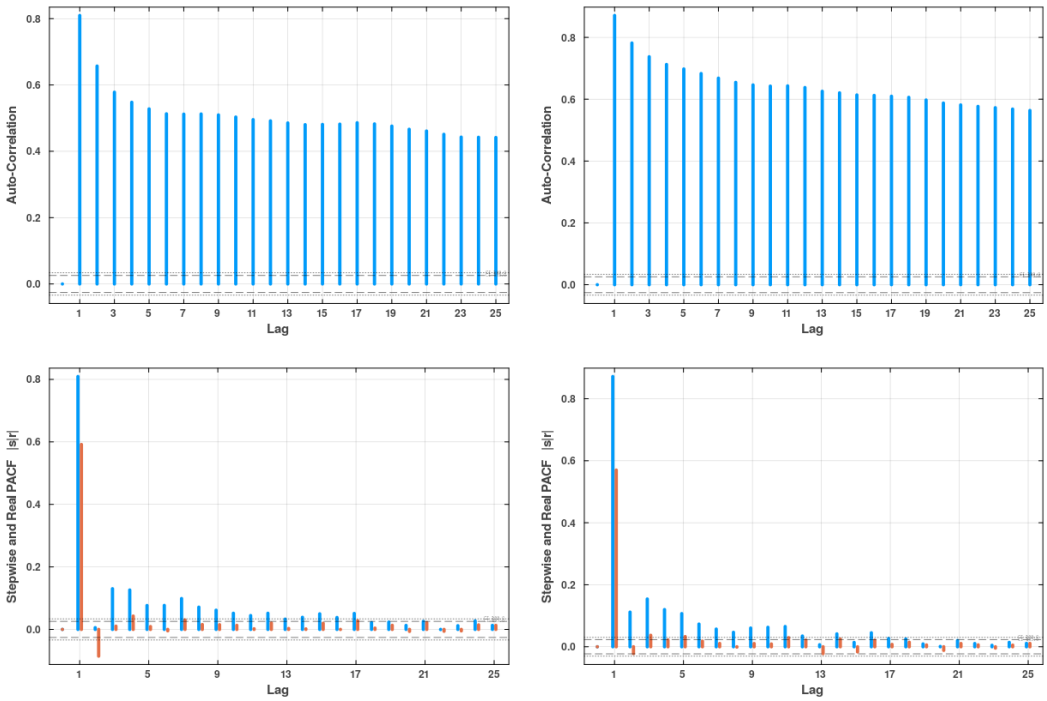

Figure A3. Autocorrelation Function (ACF) (**top**) and Partial Autocorrelation Function (PACF) (**bottom**) for San Francisco's maximum (**right**) and minimum (**left**) temperatures, based on the training dataset.

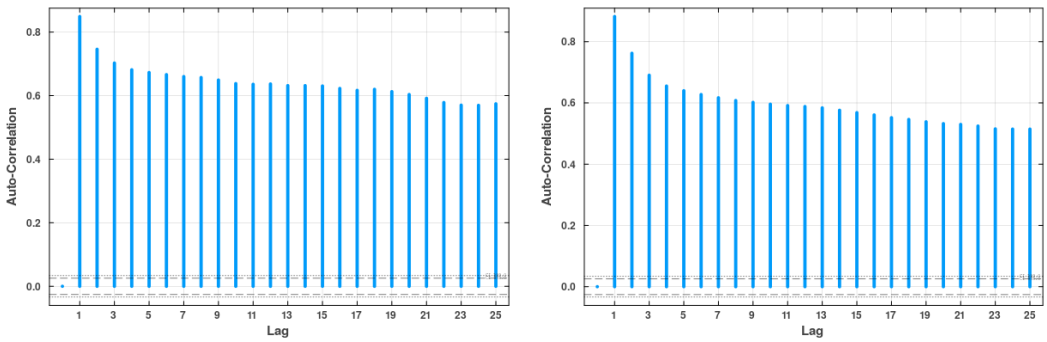

Figure A4. *Cont.*

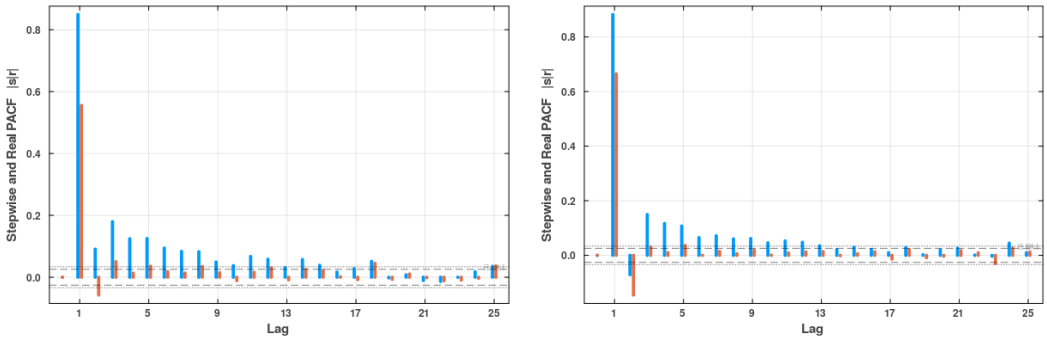

Figure A4. Autocorrelation Function (ACF) (**top**) and Partial Autocorrelation Function (PACF) (**bottom**) for Miami's maximum (**right**) and minimum (**left**) temperatures, based on the training dataset.

Appendix A.2. Time Series Decomposition Plots for Maximum and Minimum Temperatures of the Four Remaining Cities

Figures A5–A8 show the time series decomposition of maximum and minimum temperatures for Chicago, Boston, San Francisco, and Miami, in that order. Each plot divides the time series observations into data (top), trend (first middle), seasonal (second middle), and remainder (bottom). Similar to the time series decomposition plot for Houston (Figure 4), all plots clearly indicate the presence of seasonal components and the absence of a trend component.

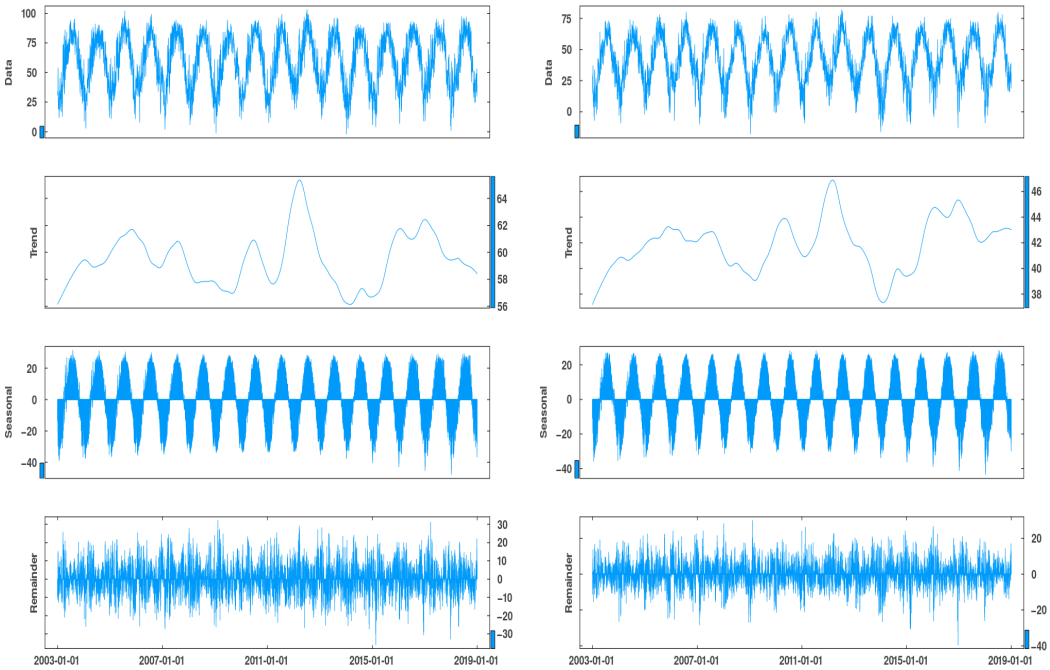

Figure A5. Plots of time series decomposition for Chicago's maximum (**right**) and minimum (**left**) temperatures, based on the training dataset.

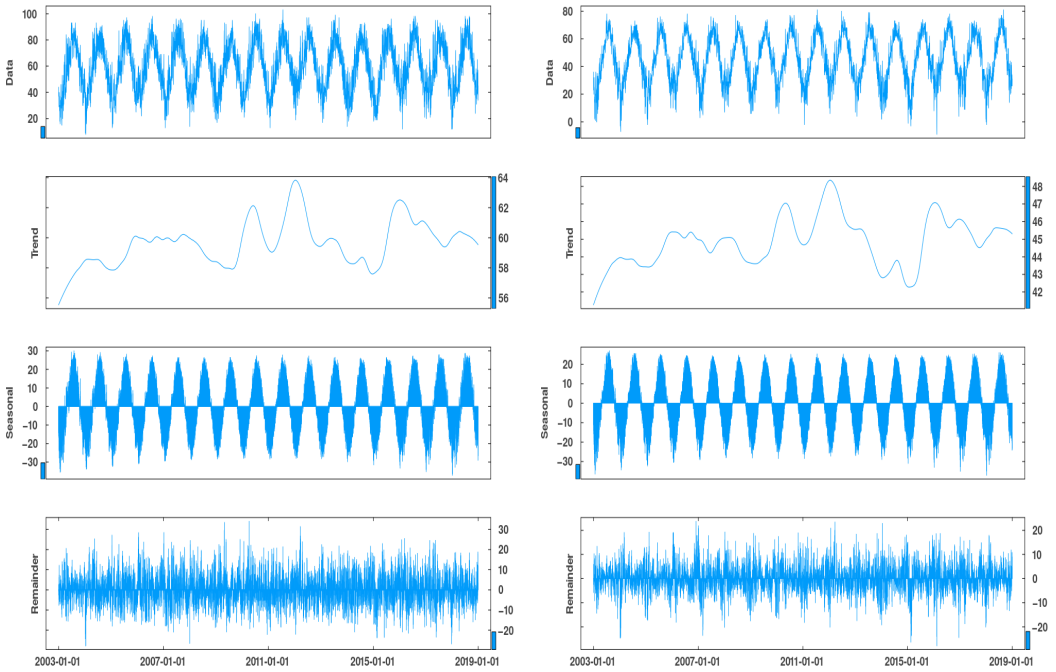

Figure A6. Plots of time series decomposition for Boston's maximum (**right**) and minimum (**left**) temperatures, based on the training dataset.

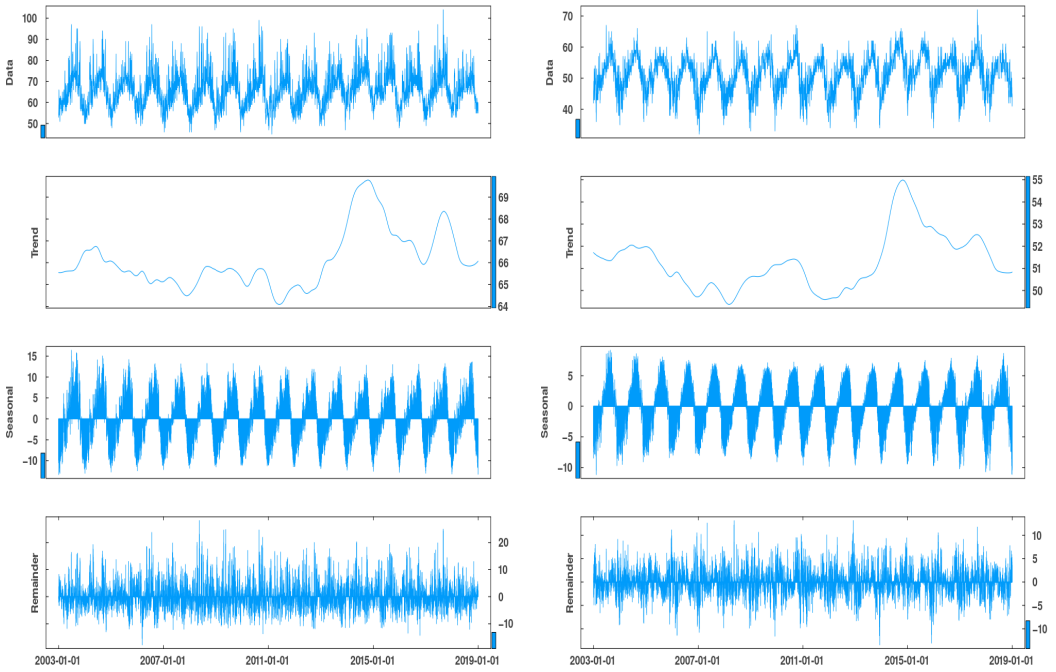

Figure A7. Plots of time series decomposition for San Francisco's maximum (**right**) and minimum (**left**) temperatures, based on the training dataset.

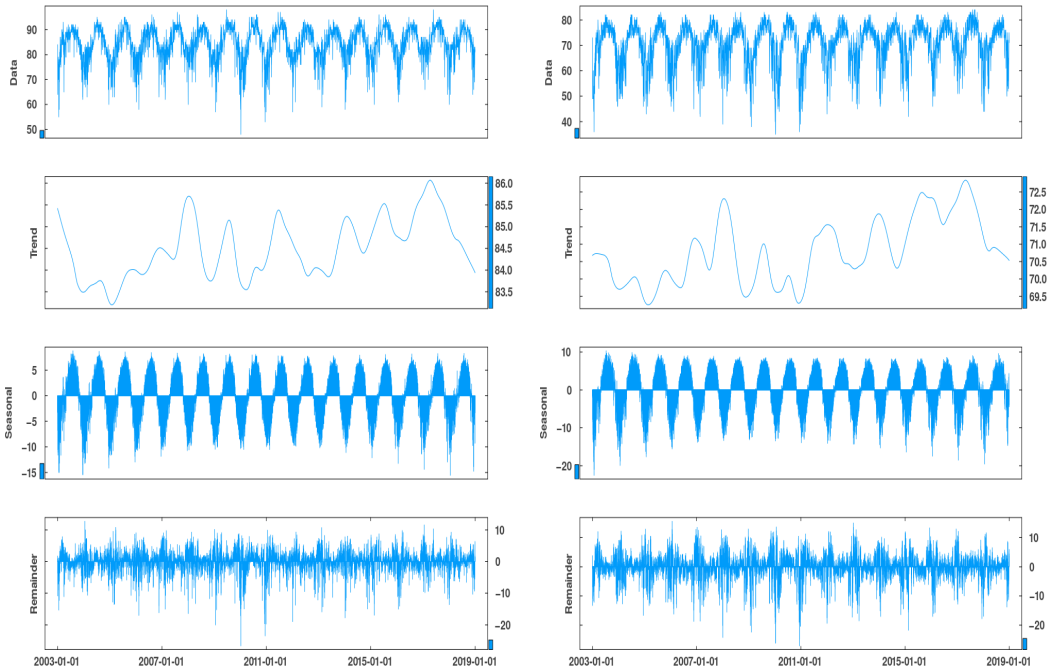

Figure A8. Plots of time series decomposition for Miami's maximum (**right**) and minimum (**left**) temperatures, based on the training dataset.

References

1. United Nations. *Framework Convention on Climate Change*; Adoption of the Paris Agreement. 2015. Available online: https://unfccc.int/sites/default/files/l09r01.pdf (accessed on 16 July 2024).
2. Pretis, F.; Schwarz, M.; Tang, K.; Haustein, K.; Allen, M.R. Uncertain impacts on economic growth when stabilizing global temperatures at 1.5 °C or 2 °C warming. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2018**, *376*, 20160460. [CrossRef] [PubMed]
3. Jacob, D.; Kotova, L.; Teichmann, C.; Sobolowski, S.P.; Vautard, R.; Donnelly, C.; Koutroulis, A.G.; Grillakis, M.G.; Tsanis, I.K.; Damm, A.; et al. Climate Impacts in Europe Under +1.5 °C Global Warming. *Earth's Future* **2018**, *6*, 264–285. [CrossRef]
4. Döll, P.; Trautmann, T.; Gerten, D.; Schmied, H.M.; Ostberg, S.; Saaed, F.; Schleussner, C.F. Risks for the global freshwater system at 1.5 °C and 2 °C global warming. *Environ. Res. Lett.* **2018**, *13*, 044038. [CrossRef]
5. Ebi, K.L.; Hasegawa, T.; Hayes, K.; Monaghan, A.; Paz, S.; Berry, P. Health risks of warming of 1.5 °C, 2 °C, and higher, above pre-industrial temperatures. *Environ. Res. Lett.* **2018**, *13*, 063007. [CrossRef]
6. Intergovernmental Panel on Climate Change (IPCC). Summary for policymakers. In *Global Warming of 1.5 °C: IPCC Special Report on Impacts of Global Warming of 1.5 °C above Pre-Industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty*; Masson-Delmotte, V., Zhai, P., Pörtner, H.O., Roberts, D., Skea, J., Shukla, P., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R.S.C., et al., Eds.; Cambridge University Press: Cambridge, UK, 2022; pp. 1–24. [CrossRef]
7. Arnell, N.W.; Lowe, J.A.; Lloyd-Hughes, B.; Osborn, T.J. The impacts avoided with a 1.5 °C climate target: A global and regional assessment. *Clim. Chang.* **2018**, *147*, 61–76. [CrossRef]
8. Shi, C.; Jiang, Z.H.; Chen, W.L.; Li, L. Changes in temperature extremes over China under 1.5 °C and 2 °C global warming targets. *Adv. Clim. Chang. Res.* **2018**, *9*, 120–129. [CrossRef]
9. Vose, R.S.; Easterling, D.R.; Kunkel, K.E.; LeGrande, A.N.; Wehner, M.F. Temperature changes in the United States. In *Climate Science Special Report: Fourth National Climate Assessment*; Wuebbles, D.J., Fahey, D.W., Hibbard, K.A., Dokken, D.J., Stewart, B.C., Maycock, T.K., Eds.; U.S. Global Change Research Program: Washington, DC, USA, 2017; Volume 1, pp. 185–206. [CrossRef]
10. Chattopadhyay, S.; Edwards, D.R. Long-Term Trend Analysis of Precipitation and Air Temperature for Kentucky, United States. *Climate* **2016**, *4*, 10. [CrossRef]
11. Lee, J.; Li, S.; Lund, R. Trends in Extreme U.S. Temperatures. *J. Clim.* **2014**, *27*, 4209–4225. [CrossRef]
12. Kunkel, K.E.; Vose, R.S.; Stevens, L.E.; Knight, R.W. Is the monthly temperature climate of the United States becoming more extreme? *Geophys. Res. Lett.* **2015**, *42*, 629–636. [CrossRef]
13. National Centers for Environmental Information. Climate at a Glance. Available online: https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/ (accessed on 12 July 2024).
14. Lai, Y.; Dzombak, D.A. Use of the Autoregressive Integrated Moving Average (ARIMA) Model to Forecast Near-Term Regional Temperature and Precipitation. *Weather Forecast.* **2020**, *35*, 959–976. [CrossRef]
15. Muhamad, N.S.; Din, A.M. Exponential smoothing techniques on daily temperature level data. In Proceedings of the 6th International Conference on Computing and Informatics, Kuala Lumpur, Malaysia, 25–27 April 2017; Volume 217, pp. 62–68.

16. Medar, R.; Angadi, A.B.; Niranjan, P.Y.; Tamase, P. Comparative study of different weather forecasting models. In Proceedings of the 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India, 1–2 August 2017; pp. 1604–1609. [CrossRef]
17. Faraway, J.; Chatfield, C. Time Series Forecasting with Neural Networks: A Comparative Study Using the Air Line Data. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1998**, *47*, 231–250. [CrossRef]
18. Ahmed, N.K.; Atiya, A.F.; Gayar, N.E.; El-Shishiny, H. An Empirical Comparison of Machine Learning Models for Time Series Forecasting. *Econ. Rev.* **2010**, *29*, 594–621. [CrossRef]
19. Kumar, N.; Middey, A. Extreme climate index estimation and projection in association with enviro-meteorological parameters using random forest-ARIMA hybrid model over the Vidarbha region, India. *Environ. Monit. Assess.* **2023**, *195*, 380. [CrossRef] [PubMed]
20. Shukur, O.B.; Lee, M.H. Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA. *Renew. Energy* **2015**, *76*, 637–647. [CrossRef]
21. Dankwa, P.; Cudjoe, E.; Amuah, E.E.Y.; Kazapoe, R.W.; Agyemang, E.P. Analyzing and forecasting rainfall patterns in the Manga-Bawku area, northeastern Ghana: Possible implication of climate change. *Environ. Chall.* **2021**, *5*, 100354. [CrossRef]
22. Fatima, S.; Ali, S.S.; Zia, S.S.; Hussain, E.; Fraz, T.R.; Khan, M.S. Forecasting carbon dioxide emission of Asian countries using ARIMA and simple exponential smoothing models. *Int. J. Econ. Environ. Geol.* **2019**, *10*, 64–69.
23. Van Rossum, G.; Drake, F.L. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, USA, 2009.
24. Bezanson, J.; Karpinski, S.; Shah, V.B.; Edelman, A. Julia: A Fast Dynamic Language for Technical Computing. *arXiv* **2012**, arXiv:1209.5145.
25. National Centers for Environmental Information. Climate Data Online. Available online: https://www.ncdc.noaa.gov/cdo-web/ (accessed on 19 June 2024).
26. United States Census Bureau. Geographic Levels. 2021. Available online: https://www.census.gov/programs-surveys/economic-census/guidance-geographies/levels.html (accessed on 19 June 2024).
27. Beck, H.E.; Zimmermann, N.E.; McVicar, T.R.; Vergopolan, N.; Berg, A.; Wood, E.F. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci. Data* **2018**, *5*, 180214. [CrossRef]
28. Kottek, M.; Grieser, J.; Beck, C.; Rudolf, B.; Rubel, F. World Map of the Köppen? Geiger climate classification updated. *Meteorol. Z.* **2006**, *15*, 259–263. [CrossRef]
29. 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]
30. Ansel, J.; Yang, E.; He, H.; Gimelshein, N.; Jain, A.; Voznesensky, M.; Bao, B.; Bell, P.; Berard, D.; Burovski, E.; et al. PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation. In Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'24), La Jolla, CA, USA, 27 April–1 May 2024; Volume 2. [CrossRef]
31. Urbano, F. Forecast.jl: A Julia Package for Time Series Forecasting. 2020. Available online: https://github.com/viraltux/Forecast.jl (accessed on 22 August 2024).
32. Saavedra, R.; Bodin, G.; Souto, M. StateSpaceModels.jl: A Julia Package for time series Analysis in a State-Space Framework. *arXiv* **2020**, arXiv:1908.01757.
33. Fairbrother, J.; Nemeth, C.; Rischard, M.; Brea, J.; Pinder, T. GaussianProcesses.jl: A Nonparametric Bayes Package for the Julia Language. *J. Stat. Softw.* **2022**, *102*, 1–36. [CrossRef]
34. Box, G.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*; Holden-Day: San Francisco, CA, USA 1976.
35. Shumway, R.H.; Stoffer, D.S. ARIMA models. In *Time Series Analysis and Its Applications: With R Examples*; Springer International Publishing: Cham, Switzerland, 2017; pp. 75–163. [CrossRef]
36. Brown, R. *Exponential Smoothing for Predicting Demand*; Little: Cambridge, MA, USA, 1956.
37. Holt, C.C. Forecasting seasonals and trends by exponentially weighted moving averages. *Int. J. Forecast.* **2004**, *20*, 5–10. [CrossRef]
38. Popescu, M.C.; Balas, V.E.; Perescu-Popescu, L.; Mastorakis, N. Multilayer perceptron and neural networks. *WSEAS Trans. Circuits Syst.* **2009**, *8*, 579–588.
39. Park, I.; Kim, H.S.; Lee, J.; Kim, J.H.; Song, C.H.; Kim, H.K. Temperature Prediction Using the Missing Data Refinement Model Based on a Long Short-Term Memory Neural Network. *Atmosphere* **2019**, *10*, 718. [CrossRef]
40. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.
41. Quadrianto, N.; Kersting, K.; Xu, Z. Gaussian Process. In *Encyclopedia of Machine Learning*; Springer: Boston, MA, USA, 2010; pp. 428–439. [CrossRef]
42. Roberts, S.; Osborne, M.; Ebden, M.; Reece, S.; Gibson, N.; Aigrain, S. Gaussian processes for time series modelling. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2013**, *371*, 20110550. [CrossRef]
43. Corani, G.; Benavoli, A.; Zaffalon, M. Time Series Forecasting with Gaussian Processes Needs Priors. In *Machine Learning and Knowledge Discovery in Databases, Applied Data Science Track*; Dong, Y., Kourtellis, N., Hammer, B., Lozano, J.A., Eds.; Springer: Cham, Switzerland, 2021; pp. 103–117. [CrossRef]
44. Hyndman, R.J. Measuring forecast accuracy. In *Business Forecasting: Practical Problems and Solutions*; Wiley: Hoboken, NJ, USA, 2014; pp. 177–183.

45. Cleveland, R.B.; Cleveland, W.S.; Terpenning, I. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *J. Off. Stat.* **1990**, *6*, 3–73.
46. Marchildon, A.L.; Zingg, D.W. A Non-Intrusive Solution to the Ill-Conditioning Problem of the Gradient-Enhanced Gaussian Covariance Matrix for Gaussian Processes. *J. Sci. Comput.* **2023**, *95*, 65. [CrossRef]
47. Mohammadi, B.; Mehdizadeh, S.; Ahmadi, F.; Lien, N.T.T.; Linh, N.T.T.; Pham, Q.B. Developing hybrid time series and artificial intelligence models for estimating air temperatures. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 1189–1204. [CrossRef]
48. Ustaoglu, B.; Cigizoglu, H.; Karaca, M. Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. *Meteorol. Appl.* **2008**, *15*, 431–445. [CrossRef]
49. Xiao, C.; Chen, N.; Hu, C.; Wang, K.; Gong, J.; Chen, Z. Short and mid-term sea surface temperature prediction using time series satellite data and LSTM-AdaBoost combination approach. *Remote Sens. Environ.* **2019**, *233*, 111358. [CrossRef]
50. Paniagua-Tineo, A.; Salcedo-Sanz, S.; Casanova-Mateo, C.; Ortiz-García, E.; Cony, M.; Hernández-Martín, E. Prediction of daily maximum temperature using a support vector regression algorithm. *Renew. Energy* **2011**, *36*, 3054–3060. [CrossRef]
51. Zhang, Z.; Dong, Y. Temperature Forecasting via Convolutional Recurrent Neural Networks Based on time series Data. *Complexity* **2020**, *2020*, 3536572. [CrossRef]
52. Seneviratne, S.I.; Rogelj, J.; Sèfèrian, R.; Wartenburger, R.; Allen, M.R.; Cain, M.; Millar, R.J.; Ebi, K.L.; Ellis, N.; Hoegh-Guldberg, O.; et al. The many possible climates from the Paris Agreement's aim of 1.5 °C warming. *Nature* **2018**, *558*, 41–49. [CrossRef]