engineering proceedings

Proceedings Paper

# Statistical Downscaling of Global Climate Models for Temperature Trend Analysis in Calgary †

**Mahdi Alipour [1*], Mohammad Bejani [2] and Arman Hosseinpour Salehi [1]**

1 Department of Civil and Environmental Engineering, University of Tabriz, mahdiap1397@gmail.com, armanhsalehi@gmail.com
2 Center of excellence in hydroinformatics, University of Tabriz, Tabriz, Iran bejani.mohammad96@gmail.com
* Correspondence: mahdiap1397@gmail.com
† Presented at the 4th International Electronic Conference on Applied Sciences, online conference 27 Oct–10 Nov 2023

**Abstract:** Climate change, particularly global warming, is a significant environmental issue that has gained widespread attention in recent decades. This study aimed to complement the model for the future by utilizing Global Climate Models (GCMs) data. The shallow-layered Artificial Neural Network (ANN) and deep-based Long Short-Term Memory (LSTM) network was applied to extract the historical temperature trend of the Calgary, Canada. Mutual Information (MI) was employed for screening purposes to ensure the quality of the input variables. The results of the study indicates that the LSTM model, which relied on the data screening method using MI, achieved RMSE of 0.01°C, DC of 0.93, a CC of 0.75 and a Bias of 1.89 have superiority over ANN method in Alberta region.

**Keywords:** climate change; temperature; statistical downscaling; artificial neural network; long short-term memory; mutual information

**Citation:** To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Published: date



## 1. Introduction

With the development of technology and industrialization of human societies, along with the increase in greenhouse gases in the past decades, the temperature of the Earth has risen and other climate parameters have changed. This phenomenon, known as climate change. Accurate prediction of these climate parameters is vital for proper planning in the future Statistical downscaling methods, including dynamic and interpolation approaches, are commonly utilized to simulate weather and climate at different scales, ranging from global to regional. These downscaling methods aim to bridge the spatial gap between the large-scale climate data and the specific requirements of local and regional applications. However, there are still challenges in applying statistical downscaling methods across different climate regions, ranging from cold regions to tropical regions. Understanding these challenges and addressing the gaps in the methodology is crucial for improving the performance of statistical downscaling and making it a preferred approach over other methods [1].Short and long-term weather forecasting is one of the fundamental challenges for researchers in the field of water resources and climate. To address this challenge, various tools such as atmospheric and oceanic General Circulation Models (GCMs), prediction scenarios, and downscaling models are used. By combining large-scale data from GCMs with small-scale synoptic data, these tools help predict and provide a more accurate trend of weather and climate in the future. Climate change encompasses various parameters such as temperature and precipitation. Having knowledge of the future changes in these parameters is vital for managing water resources in any given region.

Therefore, having temporal representation of these parameters in the future can be a useful tool for water resource management in any region [2]. To investigate climate change at the local and regional scale, it is necessary to use climate data and downscale them to a finer scale. The information available in climate data is at a large scale and not suitable for local and regional applications. Therefore, to assess the impacts of future climate at a smaller scale, there is a need to downscale these scenarios [3]. Among the statistical downscaling methods, downscaling through the use of generators, classification, and regression are more commonly utilized due to their practicality and low cost. These downscaling methods are used to simulate weather and climate in three categories: weather and climate generators, weather and climate classification, and regression-based methods[4]. [5]simulated the temperature and precipitation using GCM models, over two lakes in China until 2100. The results showed that in the future, temperature and precipitation fluctuations in these regions will be rise. Two methods, ANN and statistical downscaling model, were compared to downscale daily precipitation and temperature data in the Saguenay River basin in northern Quebec, Canada. The study concluded that ANNs perform better than the SDSM model in downscaling daily temperature data[6]. The long-term trend of temperature and its related factors in the North China region using EOF analysis and statistical dimensionality reduction, two temperature variables at the height of 850 hPa and the combination of geographic height and temperature at the height of 850 hPa have been used as predictors[7]. The results show that the temperature trend in July is related to the main EOF pattern of the two composite fields, and the statistical dimensionality reduction method is able to correctly reproduce the long-term temperature trend in North China. The use of these methods can help to better understand the temperature trend in different geographical areas, even if the main statistical models do not have enough accuracy in simulating the trend. Statistical downscaling of climate scenarios in Scandinavia is investigated by Bauren [8]. This method derives accurate and local climate information from discovery data by using statistical relationships of big data and local conditions. The importance of having accurate climate information to analyze the effects of climate change on the environment and societies makes it mandatory to use statistical downscaling in Scandinavian climate studies. The results show how the rate of warming in this region will change over the course of the 21st century and also show performance between different scenarios. The prediction of the future maximum and minimum temperature in the Boston Lake Basin has been made using SDSM models [9]. The models used data from the United States Centers for Environmental Prediction (NCEP) analysis and observations at four stations. The data series from the first thirty years (1961-1990) have been used to determine and calibrate the model, and the data series of the remaining ten years (1991-2001) have been used to validate the model. The results show that in the Boston Lake basin, the maximum and minimum temperatures will increase significantly in the future on a daily, monthly, seasonal and annual scale. The temperature increase in scenario A2 is higher than scenario B2, and the highest increase is observed in summer and the lowest increase is observed in winter.[10]

In this study, 14 different geographical patterns were used as inputs, aims to simulate the historical temperature data using Long-short Term Memory (LSTM) and Artificial Neural Network (ANN) models utilizing large-scale GCMs and assess the suitability of these methods in simulating purposes.

## 2. Method and Materials

### *2.1. Study area and data set*

The study area for this research is Calgary, located in the province of Alberta, Canada. Calgary is situated at coordinates 51.0447 degrees north and 114.0719 degrees west, with an elevation of 1100 meters above sea level. The city experiences cold winters and mild to warm summers, with an average annual temperature of around 9 degrees Celsius.

Annual precipitation in Calgary is approximately 425 millimeters, distributed throughout the year. The land use distribution in Calgary is as follows: agricultural activities occupy about 30% of the total area, while forested areas constitute around 20%, Urban areas cover approximately 40% of the study region, consisting of residential, commercial, and industrial zones. Barren land, such as rocky terrain or areas with limited vegetation, accounts for roughly 5% of the total area. The remaining 5% includes water bodies, wetlands, parks, and protected areas. In terms of the socioeconomic context, Calgary has a diverse economy with key sectors including oil and gas, technology, finance, tourism, and transportation. Figure 1 shows the study site with the elevation map. The data used in this research is from the Can_ESM5, which is from www.canada.ca. This information includes variables such as climatic, humidity, temperature and pressure type predictors for the period from 1979 to 2014, as well as observational data from www.acis.alberta.ca.

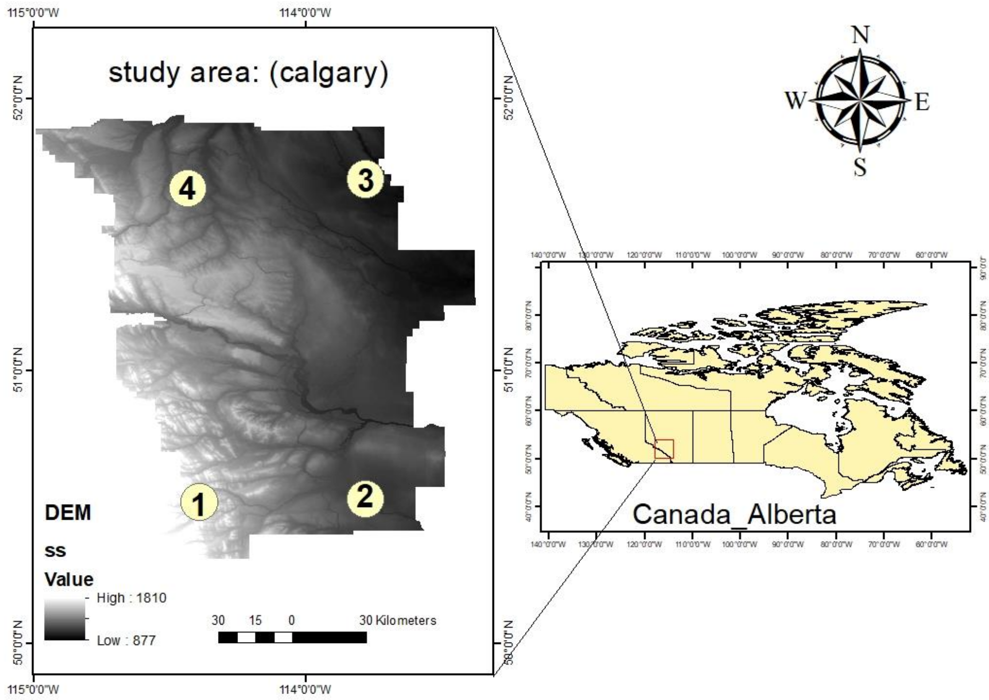


Figure 1. study area and the surrounded grid points

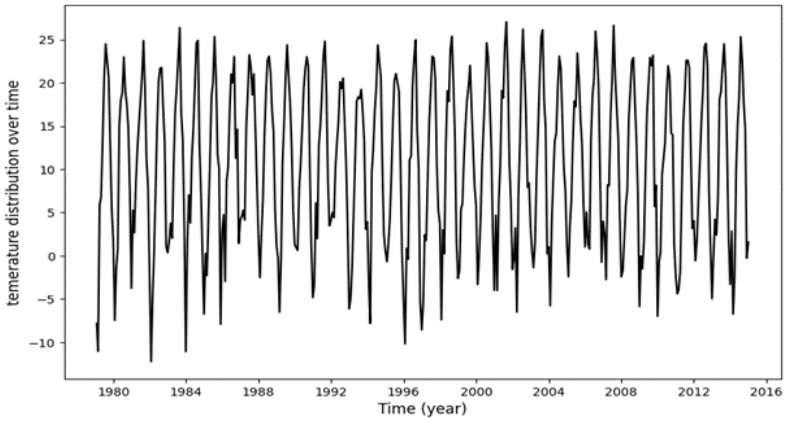


**Figure 2.** Historical trend of temperature at the statistical period**.**

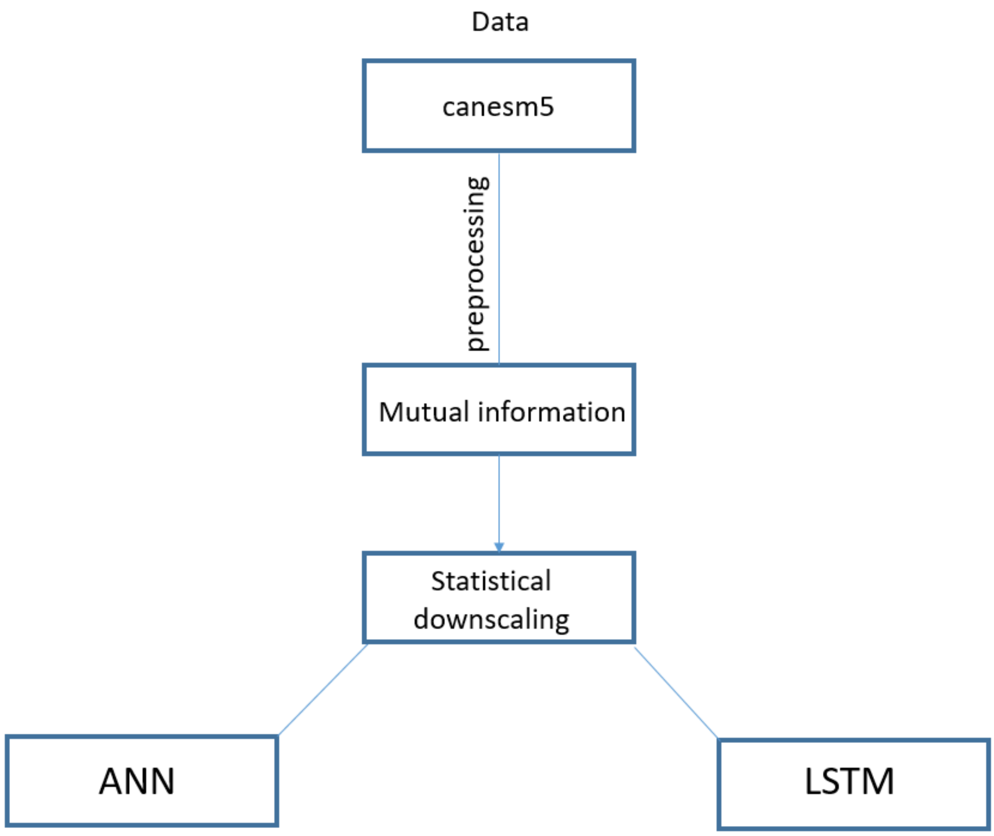


**Figure 3.** The flowchart of the proposed methodology .

*2.2. Data Preprocessing and Model Calibration*

Data preprocessing plays a crucial role in increasing the accuracy and reliability of statistical downscaling models. MI was utilized as a preprocessing technique to screen the input variables and ensure their quality. MI helps identify the relevant predictors and their relationships with the temperature of the study area.

*2.3. Statistical Downscaling Methods*

In order to extract the historical temperature trend of Calgary, statistical downscaling methods based on ANN and LSTM networks were employed. These advanced machine learning methods were used as simulation tools to capture complex relationships and patterns in the input data, making them suitable for climate modeling. ANN models consist of interconnected artificial neurons that can learn from the data and make predictions. They are effective in capturing nonlinear relationships and have been widely used in various fields, including climate modeling. In this study, the ANN model was employed alongside the LSTM network to downscale large-scale data from GCMs to a finer scale suitable for local and regional applications. LSTM networks, on the other hand, are a type of recurrent neural network (RNN) that can process and predict sequences of data. LSTM networks have the ability to capture long-term dependencies in the input data and are particularly suited for time series analysis. In the context of climate modeling, LSTM networks can effectively capture temporal patterns and trends in temperature data.Both the ANN and LSTM models were calibrated and trained using historical temperature data in Calgary. Hyperparameters of downscaling models done using the following setup: Relu activation function, Epochs=50, batch size 32, 80% train and 20% test set. The LSTM model was configured with four hidden layers, while the ANN model had one hidden layer. The calibration process included training the models with the selected inputs and their performance. By utilizing both ANN and LSTM models in the statistical downscaling process, this study aimed to capture both the complex nonlinear relationships and the temporal patterns in the historical temperature data of Calgary.

### *2.4. Evaluation criteria*

To evaluate the accuracy of the utilized models, evaluation criteria are used: determination of error criteria (RMSE), determination of coefficient (DC), average error (Bias) and correlation coefficient (CC).

RMSE measures the overall error of the model by calculating the square root of the difference between the estimate and the target.

DC represents the best model in the observed explanation, so that it shows the highest fit.

The bias measures the difference between the estimate and the target and provides the systematic error or bias of the model.

CC evaluates the strength and linear relationship between model output and observed data.

Using these criteria, the performance of ANN and LSTM models has been observed against the data.

The mathematical equations behind evaluation criteria denote as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{R_i} - \mathrm{Z_i})^2}{\mathrm{N}}} \quad (1)$$

$$\mathrm{CC} = \frac{\sum(x-\bar{x}).(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2.(y-\bar{y})^2}} \quad (2)$$

$$\mathrm{DC} = 1 - \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{Z_i} - \mathrm{R_i})^2}{\sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{Z_i} - \bar{\mathrm{Z}})^2} \quad (3)$$

where yi is the estimated value, oi is the target, o ̅is the mean value of the targets and n is the sample. Considering the CC criteria, the X indicates actual value input data, $\bar{X}$ denoted as the mean value of input data, Y refers to observed value and $\bar{Y}$ indicates mean value of observed data.

## 3. Results

This study, aimed to use GCM data as input for a statistical downscaling method based on LSTM and ANN combined with MI screening. Methodology for simulating historical temperatures GCM data were subjected to MI screening to identify ten climate parameters that show a significant non-linear relationship with temperature in the study area. These parameters are selected based on the non-linear and significant relationships they have with the temperature trend. Non-linear relationships based on MI show that the change in a parameter does not have a constant or proportional effect on the temperature trend. Instead, these relationships may show complex patterns and interactions, where the effect of a single parameter on temperature may vary across different ranges and conditions (See Table 1).

**Table 1.** MI-screened GCM data.

| Temp(1) | Temp(2) | Temp(3) | Temp(4) | Shum(4) | S850(4) | Shum(3) | S850(3) | S850(2) | Shum(2) |
|---|---|---|---|---|---|---|---|---|---|

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| mean | 2.45 | 2.95 | 1.23 | -0.24 | 0.12 | 0.12 | 0.13 | 0.12 | 0.12 | 0.13 |
| std | 9.40 | 10.42 | 10.66 | 9.38 | 0.07 | 0.07 | 0.08 | 0.07 | 0.06 | 0.06 |
| min | -19.20 | -22.32 | -25.23 | -24.85 | 0.02 | 0.02 | 0.01 | 0.01 | 0.02 | 0.02 |
| 25% | -5.08 | -5.85 | -7.62 | -7.46 | 0.07 | 0.07 | 0.06 | 0.06 | 0.07 | 0.07 |
| 50% | 0.87 | 1.81 | 0.82 | -1.36 | 0.10 | 0.10 | 0.10 | 0.09 | 0.10 | 0.11 |
| 75% | 10.83 | 12.58 | 11.53 | 8.51 | 0.18 | 0.18 | 0.19 | 0.17 | 0.17 | 0.17 |
| max | 20.12 | 21.77 | 18.71 | 16.26 | 0.32 | 0.32 | 0.34 | 0.32 | 0.32 | 0.33 |

Based on the MI screening results, the selected climate parameters were used as input for LSTM and ANN models. Based on the Table 2, the results show that the LSTM model have better ability in recording the historical temperature trend compared to the ANN model. The superiority of the LSTM model can be attributed to its deep layers and frequent connections. Deep layers allow the model to learn complex patterns and relationships in the data, while frequent connections enable error correction and integration of information from previous layers. In contrast, the ANN model, with a hidden layer, is limited in its ability to capture complex relationships. Therefore, the LSTM-based network is a more effective method for simulating historical temperature trends. The findings of this study show the effectiveness of the LSTM model as a reference model for the purposes of simulation and impact assessment in the Calgary region. The LSTM model's ability to accurately capture historical temperature trends can provide valuable insights for climate-related studies and decision-making processes. The graph of LSTM model results is shown in Figure 3.

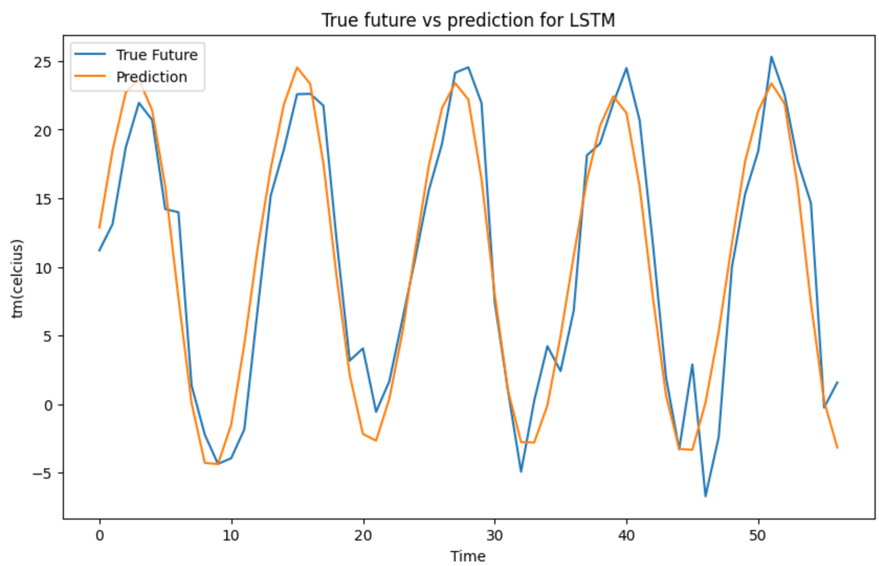


**Figure 3.** Diagram of LSTM model results**.**

**Table 2.** Evaluation criteria results.

| Model | CC | | RMSE (C°) | | DC | | Bias | |
|---|---|---|---|---|---|---|---|---|
| | Train | Validation | Train | Validation | Train | Validation | Train | Validation |
| LSTM | 0.75 | 0.69 | 0.01 | 0.03 | 0.93 | 0.89 | 1.89 | 2.2 |
| ANN | 0.62 | 0.54 | 1.2 | 1.5 | 0.78 | 0.71 | 2.01 | 2.54 |

## 4. Conclusions

This study considered the LSTM model, combined with data screening using MI, show better performance in extracting historical temperature trends compared to the ANN model. LSTM model obtained higher accuracy compared to ANN model resulting in higher DC, CC and lower RMSE and bias. The LSTM-based network shows potential as a simulation and impact assessment tool for the Calgary region. However, it is important to acknowledge the limitations and uncertainties associated with the models, data, and methods used in this research. These uncertainties must be effectively addressed to policymakers and stakeholders to make informed decisions about climate change adaptation and mitigation strategies.

Shortcomings, uncertainties and policy aspects:

Despite the promising results obtained from the LSTM model, there are several shortcomings and uncertainties that should be considered. First, this study focused solely on temperature trends, and analysis of other climate parameters such as precipitation was not included. Future research could consider incorporating additional variables to provide a more comprehensive understanding of the effects of climate change. Second, this study assumes stationarity of climate data, which may not be true for future climate conditions. It is necessary to consider the non-stationarity of climate variables when making long-term predictions. Furthermore, the accuracy and reliability of downscaling models depend on the quality and resolution of input data from GCMs. Uncertainties associated with the ability of GCMs to capture regional climate patterns and processes must be acknowledged. Additionally, the downscaling models used in this study provide insights into historical trends, but may have limitations when predicting future climate scenarios.

From a policy perspective, the findings of this study can help make informed decisions about climate change adaptation and mitigation strategies in the Calgary region. The improved accuracy of downscaling models can help policymakers and stakeholders understand local temperature trends and their potential implications for different sectors, such as agriculture, water resources, and infrastructure planning. However, policymakers should also consider the uncertainties and limitations associated with the models and data used in this study. Strong and comprehensive assessments, combining multiple information sources and considering different climate scenarios are essential for effective policy formulation and implementation.

Future research directions:

This study paves the way for further research on statistical downscaling of climate variables in the Calgary region. Future studies could investigate the downscaling of other climate parameters, such as precipitation, to provide a more comprehensive understanding of the local effects of climate change. In addition, incorporating more advanced machine learning techniques or ensemble modeling approaches can increase the accuracy and robustness of downscaling models. Addressing the non-stationarity of climate variables and considering the uncertainties associated with GCM forecasts should also be prioritized. In addition, expanding the temporal and spatial domain of analysis and incorporating a longer period of observational data can lead to more reliable and robust results.

## References

1. Hay, L.E., R.L. Wilby, and G.H. Leavesley, *A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States 1.* JAWRA Journal of the American Water Resources Association, 2000. **36**(2): p. 387-397.
2. Grotch, S.L. and M.C. MacCracken, *The use of general circulation models to predict regional climatic change.* Journal of climate, 1991. **4**(3): p. 286-303.
3. Wilby, R.L., C.W. Dawson, and E.M. Barrow, *SDSM—a decision support tool for the assessment of regional climate change impacts.* Environmental Modelling & Software, 2002. **17**(2): p. 145-157.
4. Murphy, J., *An evaluation of statistical and dynamical techniques for downscaling local climate.* Journal of climate, 1999. **12**(8): p. 2256-2284.
5. Zhu, B., et al., *CMIP5 projected changes in temperature and precipitation in arid and humid basins.* Theoretical and Applied Climatology, 2019. **136**: p. 1133-1144.

6. Tripathi, S., V. Srinivas, and R.S. Nanjundiah, *Downscaling of precipitation for climate change scenarios: a support vector machine approach.* Journal of hydrology, 2006. **330**(3-4): p. 621-640.
7. Fan, L., C. Fu, and D. Chen, *Long-term trend of temperature derived by statistical downscaling based on EOF analysis.* Acta Meteorologica Sinica, 2011. **25**(3): p. 327-339.
8. Hanssen-Bauer, I., et al., *Statistical downscaling of climate scenarios over Scandinavia.* Climate Research, 2005. **29**(3): p. 255-268.
9. Bing, Q. and J. Jiahu, *Analysis of trents of future temperature in the Bosten Lake Basin based on a statistical downscaling model.* Journal of Resources and Ecology, 2010. **1**(3): p. 268-273.
10. Baño-Medina, J., R. Manzanas, and J.M. Gutiérrez, *Configuration and intercomparison of deep learning neural models for statistical downscaling.* Geoscientific Model Development, 2020. **13**(4): p. 2109-2124.