Bayesian Multi-Head Convolutional Neural Networks with Bahdanau Attention for Forecasting Daily Precipitation in Climate Change Monitoring

Firas Gerges[1], Michel C. Boufadel[2], Elie Bou-Zeid[3], Ankit Darekar[1], Hani Nassif[4], and Jason T. L. Wang[1]

[1] Department of Computer Science, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA
{fg92,apd8,wangj}@njit.edu

[2] Center for Natural Resources, Department of Civil and Environmental Engineering, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA
boufadel@njit.edu

[3] Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544, USA
ebouzeid@princeton.edu

[4] Department of Civil and Environmental Engineering,Rutgers University – New Brunswick, Piscataway, NJ 08854, USA
nassif@soe.rutgers.edu

Abstract. General Circulation Models (GCMs) are established numerical models for simulating multiple climate variables, decades into the future. GCMs produce such simulations at coarse resolution (100 to 600 km), making them inappropriate to monitor climate change at the local regional level. Downscaling approaches are usually adopted to infer the statistical relationship between the coarse simulations of GCMs and local observations and use the relationship to evaluate the simulations at a finer scale. In this paper, we propose a novel deep learning framework for forecasting daily precipitation values via downscaling. Our framework, named Precipitation CNN or PCNN, employs multi-head convolutional neural networks (CNNs) followed by Bahdanau attention blocks and an uncertainty quantification component with Bayesian inference. We apply PCNN to downscale the daily precipitation above the New Jersey portion of the Hackensack-Passaic watershed. Experiments show that PCNN is suitable for this task, reproducing the daily variability of precipitation. Moreover, we produce local-scale precipitation projections for multiple periods into the future (up to year 2100).

Keywords: Machine Learning · Convolutional Neural Networks · Statistical Downscaling · Climate Change.

1 Introduction

Climate change refers to the phenomena where regional and global climate patterns change over time. Following the industrial revolution, greenhouse gas emissions from human activities are the primary driver of climate change [4]. This anthropogenic climate change is severely impacting communities' infrastructure and ecosystems, causing sea level rise, species extinction, in addition to disturbance in the operation of key infrastructures, such as bridges and power supply. Although national and international policies and research centers are mainly focused on the changes in global climate patterns, climate change involves regional/local climate variability as well. Subsequently monitoring climate change at the local regional level is of utmost importance. General Circulation Models (GCMs) are numerical models for simulating the physical processes taking place on land and ocean surfaces, as well as in the atmosphere. The very large spatial resolutions of GCMs make them too coarse to monitor climate change at a smaller scale. As such, to analyze and project regional climatic changes, one would need to perform spatial downscaling of GCM outputs to desired finer scales. Spatial downscaling, in the context of GCMs, refers to enhancing the coarse spatial resolution of the GCM simulations, where the simulations are downscaled either to a local weather station level, to a finer grid resolution, or to a local region level (say a watershed). Multiple attempts exist in the literature that leverage machine learning for downscaling climate variables, such as temperature [5, 25], wind [6, 12], and precipitation [20, 25]. Machine learning methods used in the previous studies include classical techniques such as support vector regression (SVR) [9], random forests (RF) [24], decision trees (DT) [27], and multi-layer perceptron (MLP) [1], as well as deep learning techniques such as convolutional neural networks (CNNs) [18, 23]. With the GCM simulations, we present a novel deep learning method, named PCNN, for the downscaling of daily precipitation. PCNN employs a multi-head CNN framework, with embedded self-attention and stacked Bahdanau attention layers, that aims to implicitly capture the spatial relationship between multiple GCM grid points. Moreover, we adopt the Monte-Carlo dropout sampling technique [13, 21, 22] to quantify aleatoric and epistemic uncertainties. The CNNs used in the previous studies differ from ours in that they attempt to downscale a GCM grid to multiple grid locations at the same time where the output is also a grid. Furthermore, the previous CNNs adopt one type of simulations. In contrast, our PCNN downscales a GCM grid to a local region/area while employing multiple types of simulations (e.g., temperature, humidity, wind, and so on). As a case study, we apply PCNN to downscale the daily precipitation over the Hackensack-Passaic watershed in New Jersey. The main contributions of our work are summarized below.

- We develop a multi-head CNN framework (PCNN) with embedded self-attention and stacked Bahdanau attention blocks to downscale/forecast the daily precipitation in a local region (i.e., the area above the New Jersey portion of the Hackensack-Passaic watershed).
- We incorporate the Monte-Carlo dropout sampling technique into our framework to quantify the aleatoric and epistemic uncertainties.

- Experimental results obtained from forecasting/downscaling the daily precipitation in the local region show that our framework is suitable for the downscaling task.
- We apply our trained deep learning model to future climate simulations to produce local-scale projections for multiple periods into the future (up to year 2100).

2 Problem Formulation and Data Collection

2.1 Problem Formulation

To perform statistical downscaling, one would need to relate large-scale GCM simulations of weather patterns to local observations. In this study, we aim to downscale the daily precipitation values using coarse-resolution simulations of multiple climate variables (temperature, heat flux, humidity, wind, and so on). To capture the different interactions of weather patterns, we extract the GCM simulations from multiple grid points, surrounding the area of interest. Fig. 1 shows the New Jersey portion of the Hackensack-Passaic watershed (highlighted in orange color), for which we attempt to downscale precipitation. The GCM grid points we use in our downscaling are represented by black circles. We select a total of $7 \times 7 = 49$ points, covering around 1.5 million km2 surrounding the watershed area. We formulate the downscaling task as a regression problem, and we aim to use GCM-simulated climate variables from the 49 grid points, treated as input features, to forecast the daily precipitation, treated as the label, above the watershed.

2.2 Local Observations

Our study area, highlighted in orange color in Fig. 1, consists of the New Jersey portion of the Hackensack-Passaic watershed. We aim to downscale the average daily precipitation over the whole area instead of a single station. We extracted the daily precipitation values from the meteorological data set (NCEI Accession 0129374) provided by the National Centers for Environmental Information (NCEI), of the National Oceanic and Atmospheric Administration (NOAA) [19]. These data are provided as a grid, with a 1/16 degree resolution. We selected the grid points contained within the watershed and computed their average for each day from January 1st, 1950, to December 31st, 2005 (daily data for 56 years). These local observations are used as labels (ground truth) in our study.

2.3 GCM Simulations

There are multiple GCMs included in CMIP5. We opted-in to select the CM3 model of the Geophysical Fluid Dynamics Laboratory (GFDL) of NOAA [11]. GFDL CM3 simulations are at grid with a 2 deg $\times$ 2.5 deg resolution (220 km $\times$ 270 km). For each of the 49 grid points shown in Fig. 1, we extract,

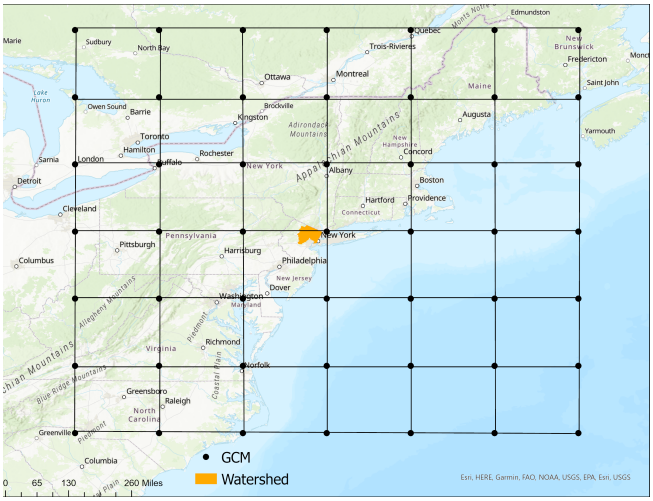

Fig. 1. GCM grid points (black circles) selected for downscaling. The area highlighted in orange color depicts the New Jersey portion of the Hackensack-Passaic watershed, for which we attempt to downscale/forecast daily precipitation.

from the CM3 model, 26 climate variables, which are listed in Table 1. As such, we have a total of $49 \times 26 = 1{,}274$ input features. We retrieved the daily data from January 1st, 1950, to December 31st, 2005, obtaining a total of 20,418 data records. Moreover, we retrieved the simulations for the periods between 2030-2040, 2060-2070, and 2090-2100, which are to be used for the long-term local-scale projections. We selected these future periods because we have the corresponding simulations available.

3 The PCNN Framework

We present a novel deep learning framework, named PCNN. This framework aims to apply convolution neural networks (CNNs) to each climate variable independently, where each climate variable is represented by a matrix containing values for all the 49 grid points, followed by a sequence of Bahdanau attention blocks. We apply dropout steps across the PCNN framework to perform uncertainty quantification using the Monte-Carlo dropout sampling technique. The main architecture of our deep learning framework is illustrated in Fig. 2.

3.1 Multi-Head Convolutional Neural Networks

Convolution neural networks (CNNs) [7, 8] have gained popularity for their performance in computer vision and image analysis. CNNs are based on a grid-like topology [10], and consist of a set of convolution and pooling layers. Convolution is a specialized matrix operation, which is the core of CNNs, used to extract local

Table 1. Large-scale Climate Variables Extracted from GCM Simulations.

Climate Variable	Description	Unit
clt	Total cloud fraction	%
$hfls$	Surface upward latent heat flux	W/m2
$hfss$	Surface upward sensible heat flux	W/m2
hus_{250}	Specific humidity at 250 hPa	-
hus_{500}	Specific humidity at 500 hPa	-
hus_{850}	Specific humidity at 850 hPa	-
$huss$	Specific humidity at near-surface	-
pr	Precipitation	Kg/m2/s
psl	Sea level pressure	Pa
rhs	Relative humidity at near-surface	%
$sfcWind$	Daily mean wind speed at near-surface	m/s
ta_{250}	Air temperature at 250 hPa	K
ta_{500}	Air temperature at 500 hPa	K
ta_{850}	Air temperature at 850 hPa	K
tas	Air temperature at near-surface	K
ua_{250}	Eastward wind at 250 hPa	m/s
ua_{500}	Eastward wind at 500 hPa	m/s
ua_{850}	Eastward wind at 850 hPa	m/s
uas	Eastward wind at near-surface	m/s
va_{250}	Northward wind at 250 hPa	m/s
va_{500}	Northward wind at 500 hPa	m/s
va_{850}	Northward wind at 850 hPa	m/s
vas	Northward wind at near-surface	m/s
zg_{250}	Geopotential height at 250 hPa	m
zg_{500}	Geopotential height at 500 hPa	m
zg_{850}	Geopotential height at 850 hPa	m

patterns (learnable features) of the corresponding feature group. The convolution layer applies filters to the input features using a set of kernels. A pooling layer often follows each convolution layer and aims to extract the patterns by focusing on the maximum, minimum, or average-based statistical summary of the neighborhood. In our architecture, we utilize a self-attention layer after each pooling layer, which allows our network to learn to choose a subset of the pooling output by giving selective attention to the features. On each input matrix representing a climate variable, we apply a sequence of convolution-pooling-attention two times, where the first pooling layer applies maximum pooling, and the second is average pooling. These sequences are applied independently to each climate variable where the sequences are followed by a dense layer. The outputs of the dense layer are concatenated and fed to the next component. We refer to our architecture as "multi-head," which is not related to the multi-head attention mechanism but rather the multiple input heads of our model (reminiscent of the "multi-convolutional heads" [14, 16]). Specifically, we have 26 climate variables, so there are totally 26 input heads. Each climate variable is represented by a 7 $\times$

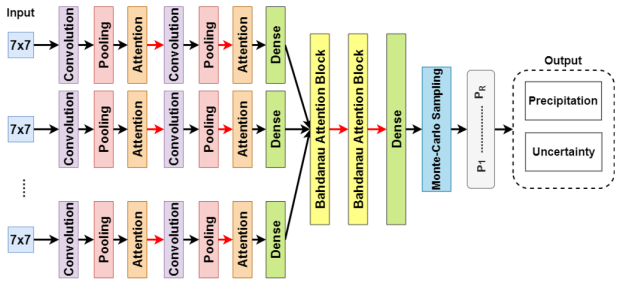

Fig. 2. Architecture of the proposed PCNN model. The input consists of the 26 7×7 matrices corresponding to the 26 climate variables considered in the study. Each 7×7 matrix contains the values of the corresponding climate variable taken from the 49 grid points shown in Fig. 1. Each matrix is independently fed to a sequence of convolution, pooling, and self-attention layers, followed by a dense layer. Red arrows represent dropout steps. The outputs of the 26 independent sequences are concatenated and fed to a sequence of Bahdanau attention layers. By utilizing the Monte-Carlo sampling technique, the proposed model outputs R prediction samples (P_o for $1 \leq o \leq R$). The mean of the R prediction samples is the predicted daily precipitation. Our model also outputs the aleatoric and epistemic uncertainty values associated with the input data and the model.

7 matrix corresponding to the 49 grid points in Fig. 1. As shown in Fig. 2, each input matrix is fed to a different sub-model, each starting with a convolution layer.

3.2 Bahdanau Attention Blocks

The Bahdanau attention methodology [2] was originally proposed to improve the performance of conventional encoder-decoder models. The main difference between the Bahdanau attention and the conventional attention is that the former aims to use a variable-length vector instead of the fixed-length one, to enhance the translation performance of the models. Similar to other attention approaches, the components of the Bahdanau attention include hidden decoder states s, context vectors c, weights a, attention scores e, as well as an annotation vector h. In the Bahdanau attention, the encoder uses the input sequence to generate the annotation sets h_i, which are combined with the hidden decoder state of the previous step s_{t-1}, and fed to an alignment model to evaluate the attention scores $e_{t,i}$. These attention scores are normalized into wights $a_{t,i}$, which are used to evaluate the context vector c_t. Similar to other attention methodologies, the context vector is used with the previous target hidden state to produce the final output. In our PCNN framework, we adopt two Bahdanau attention blocks followed by a dense layer as shown in Fig. 2 to improve the learning capability of the proposed model.

3.3 Uncertainty Quantification

In many real-world applications, uncertainty quantification is important [17, 26]. Uncertainty occurs in different components (or steps) within a deep neural network. We are particularly interested in quantifying two uncertainty types: aleatoric, which portrays the intrinsic randomness in the input data, and epistemic, which portrays the uncertainty of the deep learning model itself. We incorporate the Monte-Carlo dropout sampling technique to quantify the two uncertainties, following the methodology in [13, 15]. This methodology employs the Bayes' theorem $P(W \mid D) = (P(D \mid W)P(W))/P(D)$ to calculate the probability $P(W)$ over the network weights W. However, evaluating the posterior probability was shown to be a difficult task, and we could opt-in to leverage the parameterized variational distribution $q_\theta(W)$ over the weights (variational inference [3]). This could be achieved by using dropout mechanisms across the network during training. Still, one would need to minimize the relative entropy of $q_\theta(W)$, which could be done by using the Adam optimizer, with the cross-entropy loss function. It is standard to use dropout mechanisms within deep learning models to tackle the problem of over-fitting. The methodology behind dropout is to randomly (following a certain rate) drop neurons in selected layers. Using dropout during training would enable better generalization, and subsequently better performance on unseen data. To quantify the uncertainties in our case, we apply the dropout mechanisms during testing and perform the prediction R times (where R is set to 50 in our study) to produce R Monte-Carlo samples for each test case. We calculate the mean and variance over the R samples and leverage them to quantify the aleatoric and epistemic uncertainties [13, 15].

4 Experiments and Results

4.1 Experimental Setup

We followed the 80/20 split procedure, in which, 80% of the data, from January 1950 to August 1994 with 16,334 data records, are used to train and calibrate our model, and 20% of the data, from September 1994 to October 2005 with 4,084 data records, are used for testing. We adopt multiple performance metrics that are often employed in the precipitation downscaling literature. These are the root mean squared error (RMSE) reported in millimeter (mm), the same unit as precipitation, mean absolute error (MAE) reported in the same unit as precipitation (mm), and error in daily mean (Mean Bias) which denotes the absolute difference (reported in %) between the mean of the predicted and observed precipitation values (over the evaluation period), viz: Mean Bias = | observed mean – predicted mean| / observed mean x 100%. We also compute the error in daily standard deviation (SD Bias), which depicts the absolute difference (reported in %) between the standard deviation of the predicted and observed precipitation values (over the evaluation period), viz: SD Bias = | observed SD – predicted SD| / observed SD × 100%. We used 20% of the training data for hyperparameter tuning. Table 2 summarizes the hyperparameters and their corresponding values that gave the best results, and that are used to configure PCNN.

Table 2. Hyperparameters Used by PCNN.

Hyperparameter	Value	Description
Epochs	2000	Number of epochs
Batch Size	64	Size of each batch
Activation (Conv)	ReLU	Activation function in the convolution layers
Filters	32	Number of filters in convolution layers
Optimizer	Adam	Optimization algorithm used
Loss Function	MSE	Loss function used

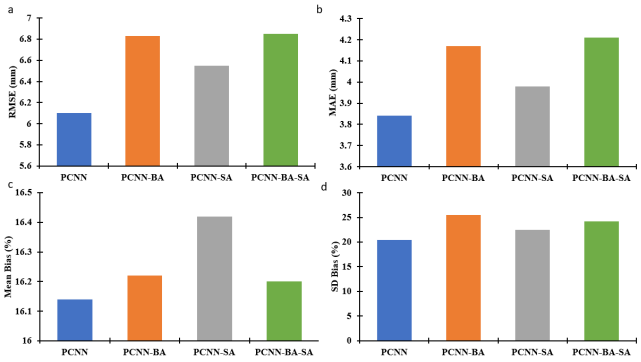

Fig. 3. Results of the ablation studies, in terms of a) RMSE reported in mm, b) MAE reported in mm, c) Mean Bias reported in %, and d) Standard Deviation (SD) Bias reported in %.

4.2 Ablation Studies

To assess the contribution of each component of our model, we turned off the Bayesian inference in our model and compared the performance of four variants: (i) PCNN which refers to the original model; (ii) PCNN-BA which refers to the model without the Bahdanau attention layers; (iii) PCNN-SA which refers to the model without the self-attention layers; and (iv) PCNN-BA-SA which refers to the model without the Bahdanau attention and self-attention layers. We report in Fig. 3 the RMSE, MAE, Mean Bias, and Mean SD for the four variants. PCNN has better performance than the other three variants, indicating the importance of both the Bahdanau attention and self-attention components.

4.3 Comparative Studies

Next, we compared the performance of PCNN with that of widely used machine learning (ML) methods in the downscaling literature: support vector regression (SVR) [9], random forests (RF) [24], decision trees (DT) [27], multilayer perceptron (MLP) [1], and convolutional neural networks (CNN) [25]. Since the related ML methods cannot quantify uncertainties, we again turned off the Bayesian inference in our model when comparing with the related ML methods. Notice that

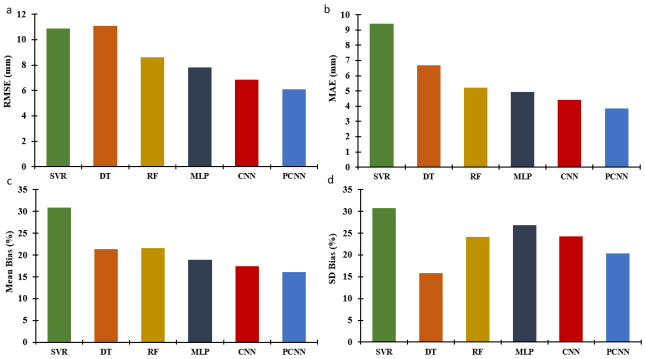

Fig. 4. Performance comparison between PCNN and the related ML methods in terms of a) RMSE reported in mm, b) MAE reported in mm, c) Mean Bias reported in %, and d) Standard Deviation (SD) Bias reported in %.

the CNN in the downscaling literature is tailored for multiple input types but without the attention mechanisms, and therefore the CNN in the downscaling literature is equivalent to the PCNN-BA-SA used in our ablation studies. Fig. 4 presents the results of the comparative studies. It can be seen from the figure that PCNN outperforms the other ML methods in terms of RMSE, MAE and Mean Bias. We note that decision trees (DT) produced the closest daily standard deviation (SD) to the observed data.

4.4 Uncertainty Quantification Results

The use of the Monte-Carlo dropout sampling technique allowed us to quantify the aleatoric and epistemic uncertainty when making predictions. Fig. 5 presents predicted mean daily precipitation within each month for the evaluation period, as well as the obtained epistemic and aleatoric uncertainties. More uncertainties come from the data than from our model. For reference, the average aleatoric (data) and epistemic (model) uncertainties are 1.41 and 0.86 respectively. These values depict that the model and data uncertainties are relatively low over the evaluation period. One could further reduce these uncertainties by better tuning the hyperparameters of the model (for model uncertainty) and refining the dataset (for data uncertainty). The aleatoric uncertainty is data-dependent, as such, when applying PCNN on a different location, a different value for such uncertainty would arise. This is due to the fact that the bias and noise in the GCM simulations are location-dependent.

4.5 Daily Variability

An important goal of downscaling precipitation is to reproduce the daily variability in addition to calculating the accuracy of daily predictions. Performance metrics such as RMSE and MAE are related to the accuracy of daily predictions

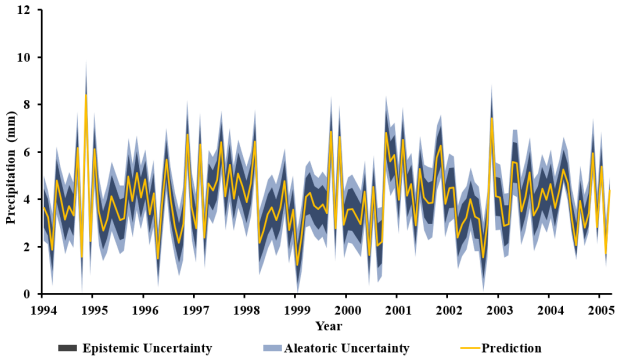

Fig. 5. Predicted monthly averages of daily precipitation as well as the epistemic and aleatoric uncertainties for the 1994-2005 period.

and are important when comparing machine learning (ML) methods to identify the most suitable one. However, to assess the true suitability of that identified ML method, one would need to analyze its performance in reproducing the daily variability and calculating certain precipitation measures, such as mean wet and dry spell lengths. Mean wet spell length is computed as the average number of consecutive days with precipitation (i.e., precipitation $\geq$ 1 mm). Similarly, mean dry spell length is the average number of consecutive days without precipitation (i.e., precipitation $<$ 1 mm). In climatology, a total precipitation of 1 mm is often the cutoff to classify days as wet or dry. Fig. 6 presents the mean wet and dry spell lengths (and standard deviations as error bars), as extracted from PCNN prediction results, as well as from NOAA observations. These results demonstrate the ability of PCNN to reproduce wet and dry spells, with relatively low bias errors (0.68% and 7.3% respectively). Here, Wet Bias error = |observed mean wet spell length – predicted mean wet spell length| / observed mean wet spell length $\times$ 100% and Dry Bias error = |observed mean dry spell length – predicted mean dry spell length| / observed mean dry spell length $\times$ 100%. Moreover, we investigate the ability of our PCNN framework to reproduce the probability distribution of precipitation. Fig. 7 shows the cumulative distribution functions (CDF) for daily precipitation as obtained from the PCNN prediction results and the NOAA observations respectively. It is apparent that the probability distribution of the prediction results from PCNN is tightly close to that of the NOAA observations. To further assess the ability of PCNN to reproduce variability, we cluster precipitation values into ranges and compute the number of days (frequency) predicted within each range. For example, PCNN predicted a precipitation value between 0 and 2 mm for 2520 days, and as such, the range 0-2 will have a frequency of 2520 days. Fig. 8 compares the frequency (i.e., the number of days) of each precipitation range between those predicted by PCNN (gray bars) and the NOAA observations (yellow bars) where the histograms are log scaled for visibility. It can be seen from Fig. 8 that our PCNN

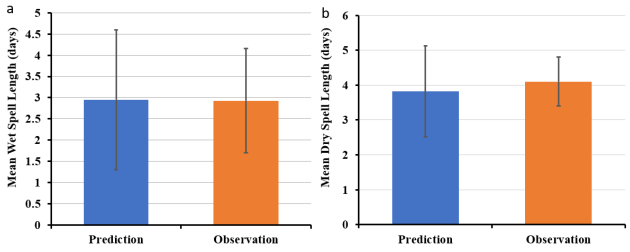

Fig. 6. Mean a) wet and b) dry spell length as observed (orange) and predicted using PCNN (blue) where error bars represent standard deviations. Here the mean wet (dry, respectively) spell length is the average number of consecutive days with (without, respectively) precipitation over the 1994-2005 period.

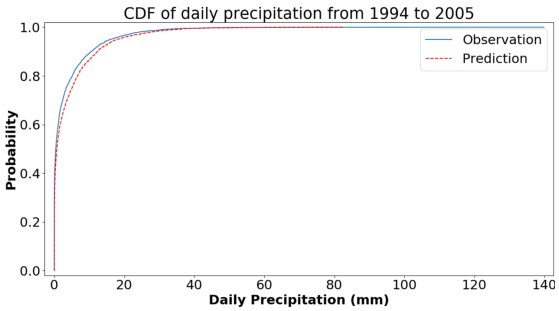

Fig. 7. Cumulative distribution function (CDF) for the daily precipitation using PCNN (dashed red line). The blue line represents the observed cumulative distribution function.

framework reproduces the frequency distribution well, which supports our claim that PCNN is suitable for the downscaling task.

4.6 Long-Term Projections

Heavy precipitation has multiple impacts on the environment, leading to crop damage, increased flooding rate, as well as soil erosion. Furthermore, runoff from precipitation can wash pollutants into water bodies, affecting the water quality. As such, long-term projections of local-scale daily precipitation is crucial for water quality and risk management. Our experimental results demonstrate the suitability of PCNN for the downscaling of daily precipitation in the Hackensack-Passaic watershed. As such, we re-trained our PCNN model using the 20,418 data records obtained from the period between 1950 and 2005 and applied the trained model to produce future projections. In particular, we used the GCM simulations for the periods between 2031-2040, 2061-2070, and 2091-2100 where we have the data available for these periods. We downscaled the precipitation for each period independently where the RCP8.5 was the radiative heat scenario

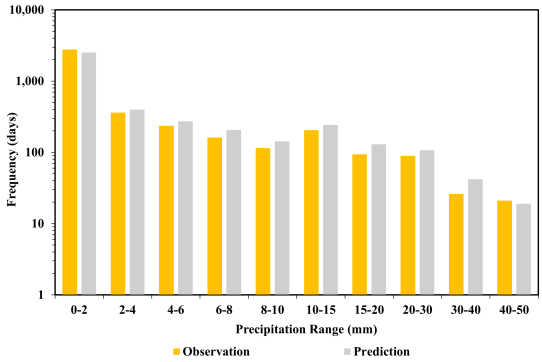

Fig. 8. Frequency histograms (log scaled) of daily precipitation (mm) values predicted by the PCNN framework (gray) and obtained from NOAA observaions (yellow) for the 1994-2005 period. The X-axis shows the precipitation ranges used to compute the frequency. For instance, a histogram for the 0-2 range represents all those days with precipitation between 0 mm and 2 mm.

used in the study. Fig. 9 shows the a) mean daily precipitation, b) mean wet spell length, c) mean dry spell length, and d) average of the number of wet days (with precipitation *geq* 1 mm) annually for each period where red error bars represent standard deviations. We note the large standard deviations of the mean daily precipitation values, which are expected given that the observed daily precipitation values from NOAA in the period between 1950 and 2005 have a standard deviation of 6.9 mm. The reported mean and standard deviation are for all the days within a study period (10 years each). As such, the large standard deviation depicts that the PCNN model is not predicting a near-mean value for each day, but with a variability that resembles that of the training period. A low standard deviation would imply that most days have precipitation values close to the mean daily precipitation over the 10-year period, which would be wrong, given the large number of dry days, as well as days with high rainfall rate, in the period between 1950 and 2005. Fig. 10 shows the mean daily precipitation within each month for the selected future periods, as well as the epistemic and aleatoric uncertainties. We note that the increasing trend of the mean daily precipitation from 2031 to 2100 agree with existing climate change studies which argue that more rain and snow are expected in the future. This fact is further supported by Fig. 9d which shows an increasing trend in the average number of wet days annually. We note that an increasing trend in mean daily precipitation might not be visible when examining each 10-years period independently (see Fig. 5 and each period in Fig. 10). However, such a trend will become apparent when analyzing a relatively longer period (e.g., 2031-2100) as shown in Fig. 10.

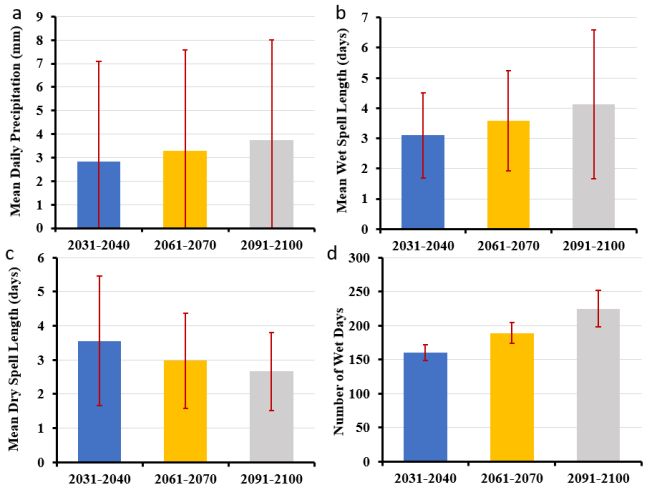

Fig. 9. Projected a) mean daily precipitation, b) mean wet spell length, c) mean dry spell length, and d) average of the number of wet days annually for three future periods where red error bars represent standard deviations. The standard deviations of the projected daily precipitation values are large, which means that we don't have a near-mean value for all the days within each period. This is expected given that the observed standard deviation on the training data is 6.9 mm.

5 Conclusions

In this paper, we develop a novel deep learning framework (PCNN) based on convolution neural networks and Bahdanau attention for the statistical downscaling of daily precipitation. Our framework employs a multi-head model, consisting of convolution neural networks with self-attention, as well as stacked Bahdanau attention layers. Moreover, we incorporate the Monte-Carlo dropout sampling technique to quantify the aleatoric and epistemic uncertainties. We trained the PCNN model to downscale the daily precipitation above the New Jersey portion of the Hackensack-Passaic watershed by using the coarse-resolution simulations from the GCM as the input to the model. Experiments show that PCNN outperforms closely related machine learning methods and is able to reproduce the daily variability of precipitation. This variability is depicted by the mean daily precipitation, mean dry and wet spell lengths, the cumulative distribution function, as well as the frequency distribution of daily precipitation values. Our results were obtained by utilizing an 80/20 split procedure. We also conducted additional experiments with five-fold cross validation and obtained similar results. We then trained the PCNN model by data from 1950 to 2005 and applied the trained model to project the long-term daily precipitation using future GCM simulations for the periods between 2031-2040, 2061-2070, and 2091-2100 respectively. We reported the projected statistics for each future period (mean, wet and dry spell lengths, number of wet days).

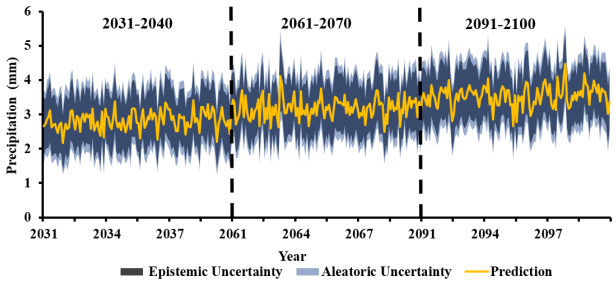

Fig. 10. Projections of precipitation as produced by PCNN with the aleatoric and epistemic uncertainties for the periods 2031-2040, 2061-2070, and 2091-2100 respectively. The periods are separated by dashed lines.

Acknowledgements This work was supported by the Bridge Resource Program (BRP) from the New Jersey Department of Transportation. We acknowledge the Working Group on Coupled Modelling of the World Climate Research Program, responsible for CMIP, and we thank the Geophysical Fluid Dynamics Laboratory of NOAA for producing and making available their model output via Earth System Grid Federation.

References

1. Ahmed, K., Shahid, S., Haroon, S.B., Xiao-Jun, W.: Multilayer perceptron neural network for downscaling rainfall in arid region: a case study of baluchistan, pakistan. Journal of Earth System Science **124**(6), 1325–1341 (2015)
2. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. In: 3rd International Conference on Learning Representations (2015)
3. Blei, D.M., Kucukelbir, A., McAuliffe, J.D.: Variational inference: A review for statisticians. Journal of the American statistical Association **112**(518), 859–877 (2017)
4. Fang, J., Zhu, J., Wang, S., Yue, C., Shen, H.: Global warming, human-induced carbon emissions, and their uncertainties. Science China Earth Sciences **54**(10), 1458–1468 (2011)
5. Gerges, F., Boufadel, M.C., Bou-Zeid, E., Nassif, H., Wang, J.T.L.: A novel deep learning approach to the statistical downscaling of temperatures for monitoring climate change. In: The 6th International Conference on Machine Learning and Soft Computing. ACM (2022). https://doi.org/https://doi.org/10.1145/3523150.3523151
6. Gerges, F., Boufadel, M.C., Bou-Zeid, E., Nassif, H., Wang, J.T.: A novel bayesian deep learning approach to the downscaling of wind speed with uncertainty quantification. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. pp. 55–66. Springer (2022). https://doi.org/https://doi.org/10.1007/978-3-031-05981-0˙5

7. Gerges, F., Shih, F., Azar, D.: Automated diagnosis of acne and rosacea using convolution neural networks. In: 2021 4th International Conference on Artificial Intelligence and Pattern Recognition. pp. 607–613 (2021)
8. Gerges, F., Shih, F.Y.: A convolutional deep neural network approach for skin cancer detection using skin lesion images. International Journal of Electrical and Computer Engineering **15**(8), 475–478 (2021)
9. Ghosh, S.: Svm-pgsl coupled approach for statistical downscaling to predict rainfall from gcm output. Journal of Geophysical Research: Atmospheres **115**(D22) (2010)
10. Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y.: Deep learning, vol. 1. MIT press Cambridge (2016)
11. Griffies, S.M., Winton, M., Donner, L.J., Horowitz, L.W., Downes, S.M., Farneti, R., Gnanadesikan, A., Hurlin, W.J., Lee, H.C., Liang, Z.: The gfdl cm3 coupled climate model: characteristics of the ocean and sea ice simulations. Journal of Climate **24**(13), 3520–3544 (2011)
12. Hu, W., Scholz, Y., Yeligeti, M., von Bremen, L., Schroedter-Homscheidt, M.: Statistical downscaling of wind speed time series data based on topographic variables. In: EGU General Assembly Conference Abstracts. pp. EGU21–12734 (2021)
13. Jiang, H., Jing, J., Wang, J., Liu, C., Li, Q., Xu, Y., Wang, J.T.L., Wang, H.: Tracing h alpha fibrils through bayesian deep learning. The Astrophysical Journal Supplement Series **256**(1), 20 (2021)
14. Khan, Z.N., Ahmad, J.: Attention induced multi-head convolutional neural network for human activity recognition. Applied Soft Computing **110**, 107671 (2021)
15. Kwon, Y., Won, J.H., Kim, B.J., Paik, M.C.: Uncertainty quantification using bayesian neural networks in classification: Application to biomedical image segmentation. Computational Statistics and Data Analysis **142**, 106816 (2020)
16. Linmans, J., van der Laak, J., Litjens, G.: Efficient out-of-distribution detection in digital pathology using multi-head convolutional neural networks. In: MIDL. pp. 465–478 (2020)
17. Liu, J.: Variable selection with rigorous uncertainty quantification using deep bayesian neural networks: Posterior concentration and bernstein-von mises phenomenon. In: International Conference on Artificial Intelligence and Statistics. pp. 3124–3132. PMLR (2021)
18. Liu, Z., Wan, M., Guo, S., Achan, K., Yu, P.S.: Basconv: Aggregating heterogeneous interactions for basket recommendation with graph convolutional neural network. In: Proceedings of the 2020 SIAM International Conference on Data Mining. pp. 64–72. SIAM (2020)
19. Livneh, B., Bohn, T.J., Pierce, D.W., Munoz-Arriola, F., Nijssen, B., Vose, R., Cayan, D.R., Brekke, L.: A spatially comprehensive, hydrometeorological data set for mexico, the us, and southern canada 1950–2013. Scientific data **2**(1), 1–12 (2015)
20. Misra, S., Sarkar, S., Mitra, P.: Statistical downscaling of precipitation using long short-term memory recurrent neural networks. Theoretical and Applied Climatology **134**(3), 1179–1196 (2018)
21. Myojin, T., Hashimoto, S., Ishihama, N.: Detecting uncertain bnn outputs on fpga using monte carlo dropout sampling. In: International Conference on Artificial Neural Networks. pp. 27–38. Springer (2020)
22. Myojin, T., Hashimoto, S., Mori, K., Sugawara, K., Ishihama, N.: Improving reliability of object detection for lunar craters using monte carlo dropout. In: International Conference on Artificial Neural Networks. pp. 68–80. Springer (2019)
23. Pan, X., Shi, J., Luo, P., Wang, X., Tang, X.: Spatial as deep: Spatial cnn for traffic scene understanding. In: 32nd AAAI Conference on Artificial Intelligence (2018)

24. Pang, B., Yue, J., Zhao, G., Xu, Z.: Statistical downscaling of temperature with the random forest model. Advances in Meteorology **2017** (2017)
25. Sun, L., Lan, Y.: Statistical downscaling of daily temperature and precipitation over china using deep learning neural models: Localization and comparison with other methods. International Journal of Climatology **41**(2), 1128–1147 (2021)
26. Wang, Y., Rocková, V.: Uncertainty quantification for sparse deep learning. In: International Conference on Artificial Intelligence and Statistics. pp. 298–308. PMLR (2020)
27. Xu, R., Chen, N., Chen, Y., Chen, Z.: Downscaling and projection of multi-cmip5 precipitation using machine learning methods in the upper han river basin. Advances in Meteorology **2020** (2020)