CliGAN: A structurally sensitive convolutional neural network model for statistical downscaling of precipitation from multi-model ensembles

Chiranjib Chaudhuri[1*] and Colin Robertson[1]

Wilfrid Laurier University, Department of Geography and Environmental Studies, Waterloo, Canada

Email: chiranjibchaudhuri@gmail.com

Abstract:

Despite numerous studies in statistical downscaling methodology, there remains a lack of methods which can downscale from precipitation modeled in global climate models to regional level high resolution gridded precipitation. This paper reports a novel downscaling method using a Generative Adversarial Network (GAN), CliGAN, which can downscale large-scale annual maximum precipitation given by simulation of multiple atmosphere-ocean global climate models (AOGCM) from Coupled Model Inter-comparison Project 6 (CMIP6) to regional-level gridded annual maximum precipitation data. This framework utilizes a convolution encoder-dense decoder network to create a generative network and a similar network to create a critic network. The model is trained using an adversarial training approach. The critic uses the Wasserstein distance loss function and the generator is trained using a combination of adversarial loss Wasserstein distance, structural loss with the multi-scale structural similarity index (MSSIM), and content loss with the Nash-Sutcliff Model Efficiency (NS). The MSSIM index allowed us to gain insight into the model's regional characteristics and shows that relying exclusively on point-based error functions, widely used in statistical downscaling, may not be enough to reliably simulate regional precipitation characteristics. Further use of structural loss functions within CNN-based downscaling methods may lead to higher quality downscaled climate model products.

1.0 Introduction:

Despite over twenty years of studies developing statistical downscaling methodology there remains a lack of methods which can downscale from AOGCM precipitation to regional level high resolution gridded precipitation (Tryhorn et. al. 2011, Chaudhuri et. al., 2017). Compared to other climate variables such as temperature or barometric pressure, precipitation is more fragmented in space and interactions of different atmospheric scales (local, meso, synoptic) and terrestrial features are more apparent in observed precipitation patterns. It is very difficult for continuous functions used in traditional statistical downscaling methods to simulate these types of local patterns. Recent advances in machine learning (ML) methods such as convolutional neural networks have started to address these long-standing issues (Shi et al. 2015). However, widely used loss functions such as Mean Absolute Error (MAE) and Nash-Sutcliffe efficiency (NS) consider overall simulation performance but ignore the spatial structure of precipitation; a key property if replicating observed local patterns is an objective (Plouffe et al. 2015). Narrowly

defined loss functions inhibit the potential of machine learning methods in downscaling, leading to poor performance of models at regional scales.

Synoptic scale climate variables are commonly simulated by Coupled Atmosphere-Ocean Global Climate Models (AOGCMs), which provide a numerical framework of climate systems based on the physio-chemical and biological characteristics of their components and feedback interactions (IPCC, 2007). Coupled AOGCMs are a computational framework which can simulate an estimate of the spatio-temporal behavior of different climatic variables under the effects of variable concentrations of greenhouse gases (GHG) in the atmosphere (Prudhomme et al. 2003). Physically-based representations of the physics and chemistry of the atmosphere and oceans make these models one of the most reliable tools for deriving future projections of meteorological variables (temperature, humidity, precipitation, wind speed, solar radiation, pressure, etc.). AOGCMs can simulate estimates of atmospheric variables which can be treated as possible representations of future climate (Smith et al. 2009).

AOGCMs degrade at higher spatial and temporal scales (Randall et al. 2007). AOGCMs typically run with a spatial resolution of 250 km to 600 km, the scale at which AOGCMs can capture synoptic-scale circulation patterns and correctly simulate smoothly varying fields, such as surface pressure. As the physical assumptions underlying the various parameterizations in AOGCMs target this scale of "variable resolving resolution", we can place a high level of confidence on the estimates over those scales. However, as we move from the synoptic scale into finer hydrologically relevant scales and analyze highly spatially heterogeneous fields such as precipitation, AOGCM skill quickly deteriorates (Hughes and Guttorp 1994). The coarse resolution of AOGCMs tends to distort the representation of regional variations of precipitation which, in turn, can alter the formation of site-specific precipitation conditions by affecting the sub grid-scale processes. Various assumptions in the parametrization of different processes, different resolutions of land covers and topography and their representations (Widmann et al. 2003), solution methods of different AOGCMs (FEM, FVM etc.) can affect the estimation of climatic variables. Failure of models in predicting highly variable processes driving, for example, daily precipitation, limits their utility in several applied and management settings (Trigo and Palutikof 2001; Brissette et al. 2006).

In order to study a climatic variable at hydrological or regional scales we need to reduce the scale of the outputs from climate models. The method used to reduce the scale of the AOGCM's output is broadly referred to as "Downscaling". As per the design and methodology downscaling procedures are broadly classified into two different types, namely dynamical downscaling and statistical downscaling. In dynamic downscaling the most common approach is to use a Regional Circulation Model (RCM) or Limited Area Model (LAM) which are designed to operate at a higher spatial resolution to simulate climatic variables of interest using AOGCM simulated fields as initial and boundary conditions (Brissette et al. 2006). But experimental design complexity and computational effort make this approach infeasible when multiple ensembles of AOGCMs are required in the study (Maurer 2007, Trzaska, Sylwia & Schnarr, Emilie. 2014).

Useful features of statistical downscaling are its simple architecture and less computational burden compared to the dynamic downscaling (Trzaska, Sylwia & Schnarr, Emilie. 2014). Statistical downscaling can produce synthetic variable of any prescribed length which makes it very popular in studies of climate change impacts (Maraun, D., et al. 2010). The Statistical downscaling draws the empirical relationships between the regional scale predictants (variables of interest) and predictors (AOGCMs) and constructs regional-scale atmospheric variable structure from large-scale simulated patterns. Hewitson and Crane (1992) explained the underlying

assumptions which created the basis of subsequent statistical downscaling methods. The robustness of statistical downscaling to study climate change impact of any region can partly be attributed to its methodology to incorporate historical observations which carry the location-specific climatic signature. The comparison of statistical and dynamic downscaling methods over Northern Canada by Dibike and Coulibaly (2005) showed that the biases in precipitation estimates were lower in statistical downscaling and the distributions of maximum and minimum temperatures were well estimated.

Statistical downscaling methods are broadly classified into three different categories based on their design of processing the predictors and predictants, (Murphy et al. 2004; Wilby and Harris 2006): (i) Weather generators, (ii) Weather typing and (iii) Transfer functions. Stochastic weather generators are essentially complex random number generators, which can be used to produce a synthetic series of data (Katz and Parlange 1996). This feature enables researchers to address natural variability when studying the impact of climate change. Brissette et al. 2007 classified weather generators into three types, (i) parametric (Richardson 1981; Kuchar 2004; Schoof et al. 2005; Hanson and Johnson 1998; Soltani and Hoogenboom 2003), (ii) semi-parametric or empirical (Semenov and Barrow 1997; Wilks and Wilby 1999) and (iii) non-parametric (Sharif and Burn 2006; Brandsma and Buishand 1998; Yates et al. 2003). Detailed discussion of these methods is beyond the scope of this paper but it should be noted one of the key advantages of weather generators is their ability to produce synthetic time series of climate data of desired length based on the statistical characteristics of the observed weather. Weather typing approaches (Brown and Katz 1995) involve clustering of regional scale meteorological variables and linking them with different classes of atmospheric circulation. Within this framework, future regional climate scenarios can be generated in two different ways: (i) by re-sampling from the observed variable distribution given the distribution of circulation pattern produced by a GCM, or (ii) by the Monte Carlo (MC) technique producing synthetic sequence of weather patterns and based on that sequence re-sampling from the archived data. The relative frequency of different types of weather patterns are then weighed to estimate the moments or the frequencies of the distribution for regional scale climate.

Perhaps the most popular approach within statistical downscaling methods are Transfer functions, which is a regression-based framework (Crane and Hewitson 1998; Cannon and Whitfield 2002; Wilby et al. 2002; Tripathi et al. 2006). The method consists of developing a direct functional relationship between global large-scale variables (predictors) and local regional scale variables (predictants) through statistical fitting. One method can differ from the other on the choice of mathematical functions, predictor variables and procedure of deriving relationship. There have been several studies focused on the application of neural networks (Zorita and von Storch 1999; Wilby et al. 1998), regression-based methods (Wilby et al. 1998), support vector machine (Tripathi et al. 2006) and analog methods (Gutierrez et al. 2004) in statistical downscaling. Artificial Neural Networks (ANN), due to their robust nature in capturing the nonlinear relationship between predictors and predictants, have gained wide recognition in climate modeling community (Crane and Hewitson 1998; Wilby et al. 1998; Haylock et. al. 2006; Tran Anh et. al. 2019).

More recently, the use of machine learning and data science has increased in various fields owing to their superior performance and robust methods/software implementations. Specifically, Convolution Neural Network (CNN) modelling has gained wide popularity because of lower computational requirements compared to the dense networks through extraction of spatial information through kernel filters (LeCun et. al. 1998). There are a wide range of applications of

CNNs including image recognition (Krizhevsky et. al. 2017), image segmentation (Ronneberger et. al. 2015) and satellite image change detection (Wang et al. 2019). Image super-resolution (Dong et. al. 2014) is another method where CNNs have been applied to increase the resolution of an image, an application analogous to the climate model downscaling context. Vandal et. al. 2017 used a generalized stacked super resolution convolutional neural network to downscale daily precipitation over the US. Despite several limitations in their experiments the result shows the efficiency and robustness of the approach over other methods in predicting extremes. Cheng et. al. 2020 has also recently introduced a novel residual dense block (RDB) into the Laplacian pyramid super-resolution network (LapSRN) to generate high-resolution precipitation forecast. Onishi et. al. 2019 used super-resolution techniques to simulate high-resolution urban micrometeorology, while Ji et al. 2020 proposed several CNN-based architectures to forecast high-resolution precipitation. Underlying all of these models is the treatment of two-dimensional fields such as climate model outputs and gridded observations as analogous to non-geographic images which makes CNNs an ideal candidate as transfer functions in statistical downscaling.

The recent advancements of machine learning strategies have developed several algorithms for image super resolution, for example attention-based training (Ben et. al. 2020, Wenjie et. al 2020). One such recent advance, Generative Adversarial Training (GAN) was proposed by Goodfellow et. al. (2014) where two networks compete in a zero-sum game has been used widely to train deep neural networks in recent years. This method provides a superior training of network and can produce outputs which look superficially close to reality to human observers and address the gradient problem in an intuitive way. Stengel et. al. 2020 used adversarial learning to downscale wind and solar output from several AOGCM climate scenarios to regional level high-resolution. Cheng et. al. 2020 also used adversarial learning to downscale precipitation. Their result shows promising performance of generative adversarial network in downscaling climate data.

In this paper, we develop a novel downscaling method using GAN which can downscale an ensemble of large-scale annual maximum precipitation given by several AOGCMs to the regional-level gridded annual maximum precipitation. The objectives of our study are following;

1. Develop a methodology to downscale large-scale precipitation, given by several AOGCMs, to regional-scale precipitation by statistical downscaling using Convolution Neural Network and Generative Adversarial Training.
2. Propose a novel loss function which is a combination of content loss, structural loss, and adversarial loss which improves the prediction of global and regional qualities of the downscaled precipitation
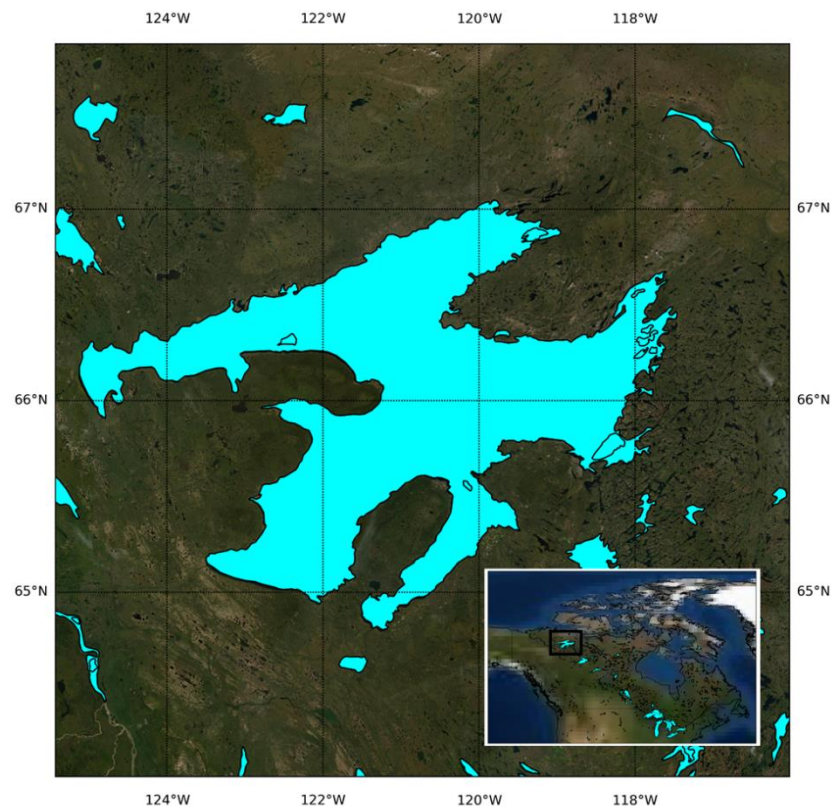
2. Methods
2.1 Study Area and Datasets:
The Tsá Túé Biosphere Reserve is situated on Great Bear Lake in Northwest Territories, Canada. Great Bear Lake is the last large pristine arctic lake, covering an area over $31,000$ km^2. The rate of environmental change in northern Canada has been widely documented, including landform transformation (Lantz and Kokelj 2008), shrub expansion in the tundra (Wilcox et al. 2019), permafrost thaw and related hydrological impacts (Quinton et al. 2011), wildlife declines (Vors and Boyce 2009) and changes affecting local communities and culture (Ford et al. 2015). Furthermore, scarce weather data due to sparse observational stations makes the estimation of the change through in-situ observation more difficult. It is evident the local communities in Canada's

north need to adapt and prepare for the anticipated impacts of climate change with help of the climatic data at hydrologically-relevant scales. This sets up a critical need for better climate impact prediction and estimation at regional scales. Figure 1 shows Great Bear Lake and its adjacent regions within Tsá Túé Biosphere Reserve which is used a downscaling region of interest in this study.

Figure 1: Study domain over the Tsa Tue Bio-reserve, locate in Northwest Territories, Canada. In the inset the location is shown in the box in perspective of entire Canada.

In order to develop a downscaling methodology, we employed annual maximum daily precipitation as our target variable. Annual maximum daily precipitation data during 1950-2010 were extracted from the NRCANMET daily gridded precipitation dataset (Hutchinson et al. 2009). While gridded data derived from a sparse observation network in northern Canada is not ideal validation data, these are the best available for the region and should replicate regional patterns. Wilby et al. (1999) and Wetterhall et al. (2005) in their studies discussed the properties of any predictor in statistical downscaling, they should be, (1) reliably simulated by GCMs, (2) readily available from the archives of GCM outputs, and (3) strongly correlated with the surface variables of interest (rainfall in the present case). The simulated rainfall by GCMs contains the dynamical information of the atmosphere as well as the information of the effects of climate change on rainfall, through different physical parameterizations. We therefore utilized the AOGCM simulated precipitation as a predictor of the observed (i.e., gridded) precipitation.

Nine AOGCMs from the CMIP6 archive, were selected for this study (Table 1). Different AOGCMs have different grid resolutions and many are represented in Gaussian grids, which

makes it hard to treat them within a common mathematical framework. To overcome this difficulty, we interpolated precipitation from different AOGCMs onto 10km resolution grids.

Table 1: Features of different input AOGCMs

AOGCM	Institution	Grid Type	Horizontal dimension (Lon/Lat)	Vertical Levels
BCC ESM,	Beijing Climate Center	T42	128x64	26
CAN ESM5	Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada	T63	128x64	49
CESM2	National Center for Atmospheric Research	0.9x1.25 finite volume grid	288x192	70
CNRM CM6.1	Centre National de Recherches Meteorologiques	T127	256x128	91
CNRM ESM2	Centre National de Recherches Meteorologiques	T127	256x128	91
GFDL CM4	Geophysical Fluid Dynamics Laboratory	C96	360x180	33
HAD GEM3	Met Office Hadley Centre	N96	192x144	85
MRI	Meteorological Research Institute, Tsukuba, Ibaraki 305-0052, Japan	TL159	320x160	80
UK ESM1	Met Office Hadley Centre	N96	192x144	85

2.2 Downscaling Method:

We train a generating network G_{θ_G} which can transform coarse resolution AOGCM extreme precipitation (P_{GCM}) to regional scale fine resolution extreme precipitation (P_{obs}). The technical difficulties of the problem involve increment of the resolution, bias correction, and regional scale precipitation feature corrections. The number of training years y=1, ..., n then the θ_G is obtained through solving the minimization of the downscaling total loss function l^D;

$$\widehat{\theta_G} = \frac{argmin}{\theta_G} \frac{1}{n} \sum_{y=1}^{n} l^D \left(G_{\theta_G}(P_{GCM}^y), P_{obs}^y \right) \quad (1)$$

We designed the downscaling total loss as weighted combination of several components. We discuss these components in detail later.

2.2.1 Adversarial training:

We used Wasserstein GAN (WGAN) (Arjovsky et. al. 2017) as discriminative network . It uses the Earth-Mover (also called Wasserstein-1) distance W(q, p), which is informally defined as the minimum cost of transporting mass in order to transform the distribution q into the distribution p (where the cost is mass times transport distance).

Formally, the game between the generator G_{θ_G} and the discriminator D_{θ_D} is a minimax objective. The WGAN value function is constructed using the Kantorovich-Rubinstein (Villani 2008) duality to obtain;

$$\frac{min}{G_{\theta_G}} \frac{max}{D_{\theta_D} \in D} E\left[D_{\theta_D}(P_{obs})\right] - E\left[D_{\theta_D}\left(G_{\theta_G}(P_{GCM})\right)\right] \qquad (2)$$

Where, D is the set of *1-Lipschitz* functions. To enforce the Lipschitz constraint on the WGAN, Arjovsky et. al. (2017) propose to clip the weights of the WGAN to lie within a compact space [−c, c]. The set of functions satisfying this constraint is a subset of the *k-Lipschitz* functions for some k which depends on c and the WGAN architecture.

2.2.3 Downscaling Total Loss:
The Total loss is designed as a linear combination of 3-loss components: adversarial loss, content loss, and structural loss.

$$Total\ loss = adversarial\ loss + content\ loss + structural\ loss$$

Adversarial Loss:
The adversarial loss is given by the approximate earth moving distance between the observation and generator output. It is given by;

$$Adversarial\ Loss = D_{\theta_D}\left(G_{\theta_G}(P_{GCM})\right) \qquad (3)$$

Content Loss:
The Nash–Sutcliffe model efficiency coefficient (NSE) is widely popular loss estimate which assess the predictive ability of any hydrology model. We have taken (1-NSE) as content loss;

$$Content\ Loss = \frac{\sum_{y=1}^{n}\left(G_{\theta_G}(P_{GCM}^{y}) - P_{obs}^{y}\right)^2}{\sum_{y=1}^{n}\left(P_{obs}^{y} - \frac{1}{n}\sum_{y=1}^{n}P_{obs}^{y}\right)^2} \qquad (4)$$

This loss can take range 0 to ∞. 0 loss means perfect match with the observation. The loss value 1 means that the model is as accurate as the average of the observed values. The loss greater than 1 signifies that the observed mean is a better predictor than the model.

Structural Loss:

The overall low error may not conserve the regional relevant details common to high resolution precipitation. To address this problem and conserve the regional details we introduced a structural loss function. Multi-Scale Structural Similarity Index (MSSIM) (Wang et. al. 2003) compares the luminance, contrast, and structure of two 2D field. We compared this metric in multiple scale to preserve the regional scale structure of the precipitation at those scales. We defined our loss based on structural dissimilarity (DSSIM) which is defined by (1-MSSIM)/2;

$$Structural\ Loss = \frac{1}{2}(1 - MSSIM) = \frac{1}{2}\left(1 - l_M^{\alpha_M} \prod_{i=1}^{M} c_i^{\beta_i} s_i^{\gamma_i}\right) \qquad (5)$$

Where the highest scale or the window size is given by M. α_i, β_i, and γ_i are weights of luminance, contrast, and structure components respectively at scale *i*. We used α=β=γ=1. For any given scale the luminance comparison is given by;

$$l = \frac{2\mu_{obs}\mu_G + c_1}{\mu_{obs}^2 + \mu_G^2 + c_1} \qquad (6)$$

The contrast comparison is given by;

$$c = \frac{2\sigma_{obs}\sigma_G + c_2}{\sigma_{obs}^2 + \sigma_G^2 + c_2} \qquad (7)$$
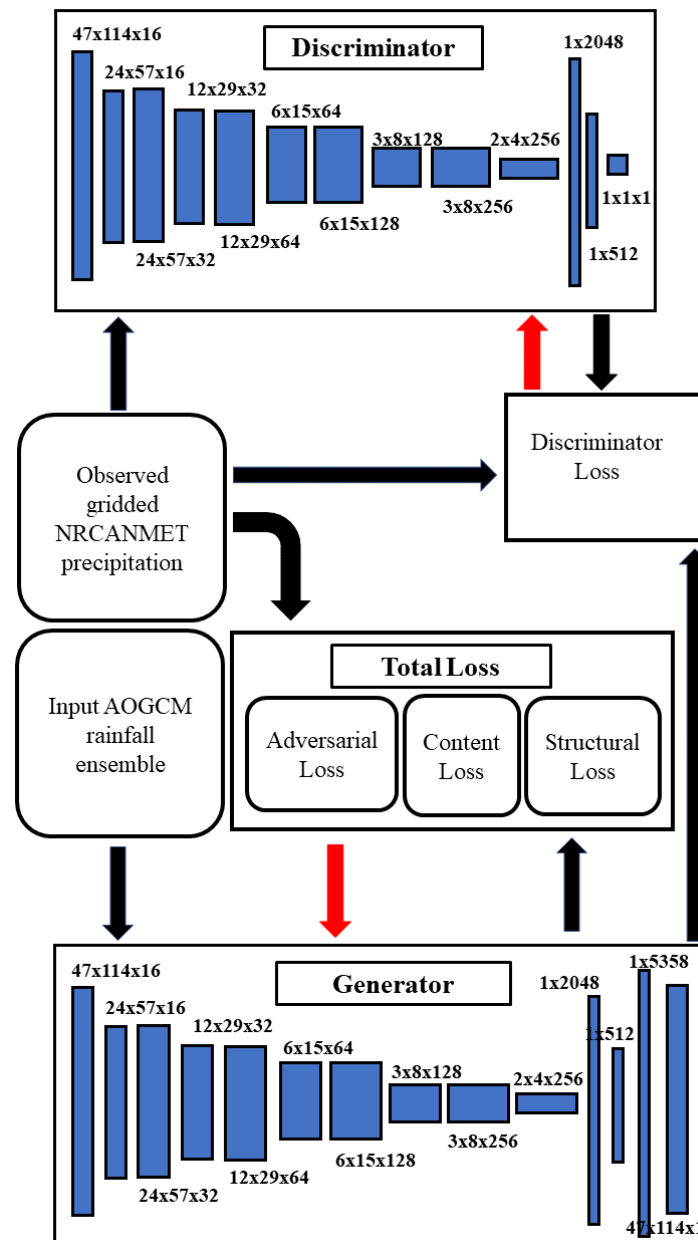
and the structure comparison is given by;

$$s = \frac{\sigma_{obs,G} + c_3}{\sigma_{obs}\sigma_G + c_3} \qquad (8)$$

Where μ_{obs} and μ_G are the mean of observation and generated values respectively, σ_{obs} and σ_G are standard deviation of observation and generated values respectively, and $\sigma_{obs,G}$ is the co-variance between observation and generated values. c_1, c_2, and c_3 are small constants to stabilize the divisions with weak denominator. We used $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$, $c_3 = \frac{c_2}{2}$ where $k_1 = 0.01$, $k_2 = 0.03$ and L is the dynamic range of the input which we have taken as 100.

2.2.4 Networks:
The generative network is built in the style of convolution encoder and dense decoder. We used LeakyReLU with momentum 0.2 to activate the hidden layers and used strided convolution for downsampling. ReLU has been used in the last layer of the Generator. We used L2-norm regularization with weights 1e-5 on model weights and biases to prevent over-fitting.

Figure 2: Schematic of the Generative Adversarial Network

Like the Generative network, the discriminative network has convolution encoder with strided convolution downsampler. We used the LeakyReLU with momentum 0.2 as activation in hidden layer and ReLU activation is used in the output layer. Batch normalization with momentum 0.8 is used after every convolutional layers to stabilize the training. The discriminative network computes the approximate earth moving distance (Wasserstein distance) between the observed and predicted precipitation. Following the recommendation of WGAN we created a compact support for layer weights by clipping them between [-1, 1]. The schematic for the networks and training procedure are shown in Figure 2.
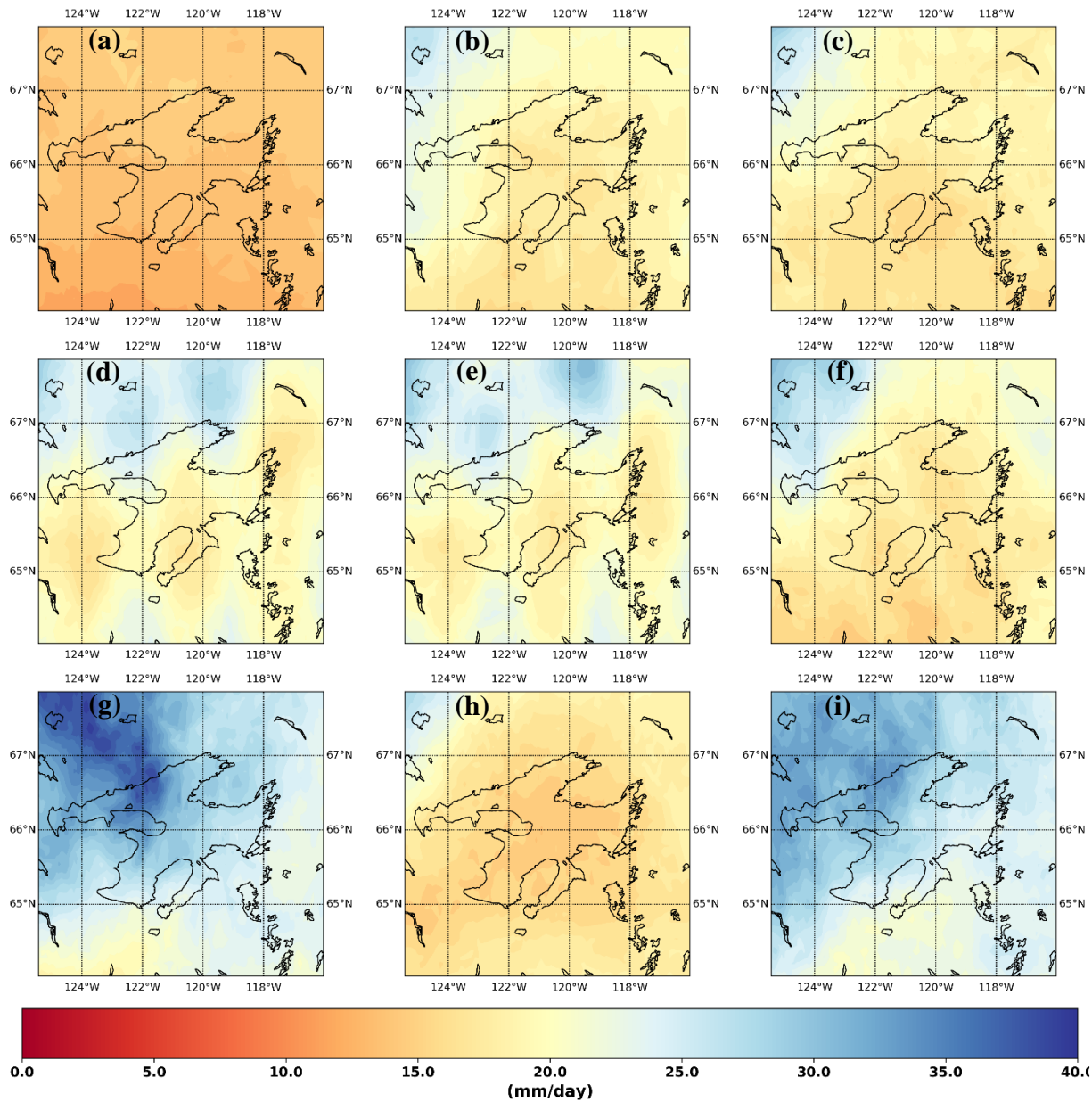
2.2.5 Training details

Model training was performed on the NVIDIA Quadro 4000 GPU. The l_2-norm regularization was been used on both weight and bias of Generator to prevent over-fitting. Furthermore, we employed random sub-sampling strategy to train the network to capture the stochasticity of the entire population. The entire dataset is temporally split into random training and testing set for each training epoch. The network is trained using the training set and the performance is tested on testing set. We used 66%/33% training/testing data divide which resulted into 41 temporal points for training and 20 temporal points for testing. We avoided any kind of rescaling of the input or output data and allowed the model to learn and evolve the climate signal present in the different model datasets and observations. The Generator was optimized using Adam first-order gradient-based optimization (Kingma et. al. 2014) based on L_{inf} norm (Adamax) (Ruder et. al. 2016) with initial learning rate m=0.02 and β_1 decay 0.5. We wanted the discriminator to compute the gradient based on an adaptive window near present iteration and didn't want to use all the previous generator results of the gradient information. Therefore, we used a moving window update of gradient in Adam (Adadelta) (Zeiler et. al. 2012) to train the discriminator Wasserstein distance computer. The model is trained for 10,000 iterations which is found to be more than adequate for simulating realistic high-resolution precipitation patterns.

To compare the performance of Generative adversarial training we trained two more models. In the 1st alternative network traditional mean absolute error has been used to train the generator instead of combination of loss applied in the CliGAN. We also trained a 2nd alternative model inspired from the Chaudhuri et. al. 2017. In this model, first 10 principal components with largest eigenvalues from each of the nine climate models are mapped to the 20 principal components (~99% variance explained) of the observation using a single layer neural network. The overfitting of the network is prevented using 0.5 neuron drop out rate between the layers. More details on the climate model eigen spectra, observation eigen spectra, and training performance of the network is given in the reference section of the paper.

3.0 Results and Discussion

The median annual maximum rainfall generated by the various AOGCMs examined in this study varied considerable (Figure 3), indicating significant inter-model uncertainty and need for an ensemble approach (Chen et. al.2014). One of the major challenges in downscaling these precipitation variables to any region is to correct these biases (or differences) according to the observed pattern (Manzanas et. al. 2018). An ideal downscaling method not only generates a higher resolution regional level realization of the climatic variable but should also reproduce both patterns and magnitudes. The models examined here show the need for downscaling methods that can produce a spatial and temporally coherent observed climatic distribution.

Figure 3 : Temporal median of AOGCM annual maximum precipitation of (a) BCC ESM, (b) CAN ESM5, (c) CESM2, (d) CNRM CM6.1, (e) CNRM ESM2, (f) GFDL CM4, (g) HAD GEM3, (h) MRI, and (i) UK ESM1.

A key contribution in the CliGAN modelling framework is the development of a novel loss training method incorporating spatial structure of rainfall. To validate our use of the total loss functions, we experimented with different combinations of loss functions (Table 2). Each model is trained with different permutation of loss functions for 10,000 iterations and the median training and testing loss of the last 50 iterations is reported in the table. The Adversarial Loss and NS loss combination has content error of 0.015 and 0.142 and structural error of 0.025 and 0.100 for training and testing, respectively. The Adversarial Loss and MSSIM loss combination have content error of 0.070 and 0.110 and structural error of 0.019 and 0.025 for training and testing, respectively. The NS and MSSIM loss combination have content error of 0.011 and 0.774 and structural error of 0.024 and 0.283 for training and testing, respectively. The combination of all 3 Adversarial Loss, NS loss, and MSSIM loss have content error of 0.015 and 0.043 and structural

error of 0.024 and 0.033 for training and testing, respectively. It can also be noticed the Adversarial Loss plays a vital role is stabilizing the performance of the network (Figure 4). Even though the training results are similar magnitude for all the experiments, the inclusion of the adversarial loss function in the combination of loss functions, have produced testing results order of magnitude improved results for the testing set. The total loss combination is found to be producing balanced error for both content and structural error and performed stable for both training and testing. Thus, we go forward with this total loss combination. To validate the model, we trained the total loss combination for 20,000 iterations the content loss is found to be 0.011 and 0.020 and structural loss is found to be 0.021 and 0.017 for training and testing respectively. This justifies our choice of using this loss function.

Table 2: Performance of different objective function for the model. The median of last 50 iterations have been used to measure the performance.

Loss combination	Content Loss		Structural Loss	
	Train	Test	Train	Test
Adversarial + NS	0.015	0.142	0.025	0.100
Adversarial + MSSIM	0.070	0.110	0.019	0.025
NS + MSSIM	0.011	0.774	0.024	0.283
Adversarial + NS + MSSIM	0.015	0.043	0.024	0.033
Adversarial + NS + MSSIM LT	0.011	0.020	0.021	0.017

Figure 4: Different loss functions used for training and their training and testing error.
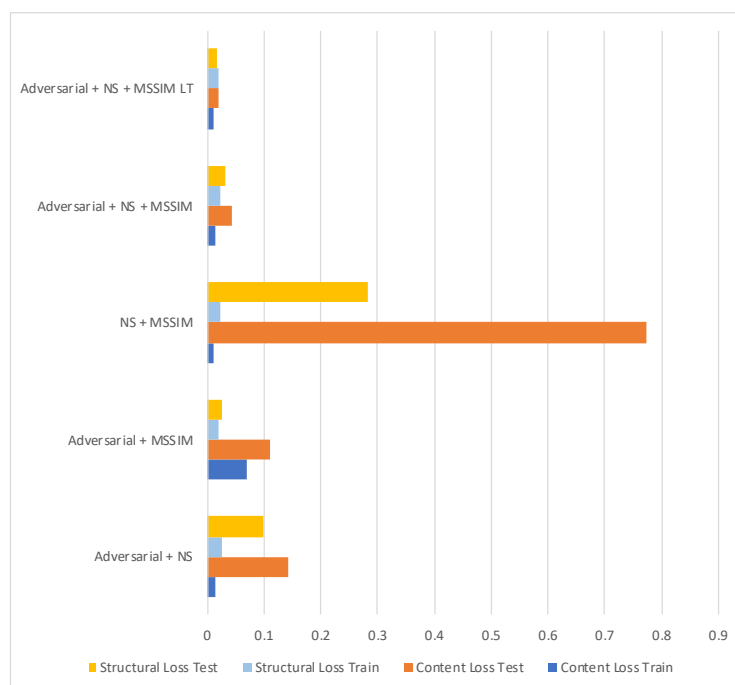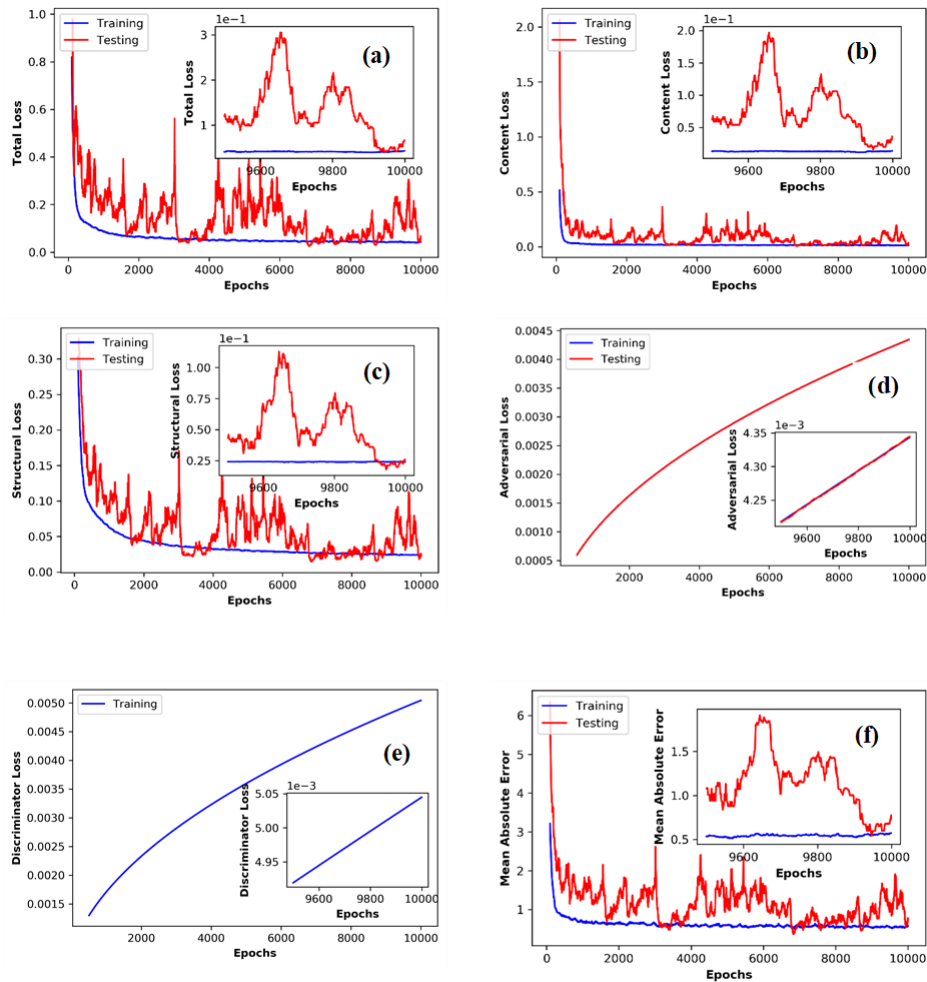
Figure 5 shows the trace of different training and testing losses for all the loss function combination error function for 10,000 iterations. For total, content, structural loss first 100 iteration results are discarded as part of generator warm up and for adversarial loss and discriminator loss first 500 iterations are discarded as part of discriminator spin up. Figure 5a shows the trace of the total loss. Considering this is a combination of 3 types of errors: adversarial error, content error, and structural error, it does not have any meaningful unit. Figure 5b shows the trace of content loss and Figure 5c shows the trace of the structural loss. Notice the stable training trace and fluctuating testing trace. However, overall, the testing error is showing decreasing trend for both content and structural loss. The adversarial loss keeps on increasing (Figure 5d) for both training and testing. This signifies the ability of the discriminator to distinguish generator output and observation. The training loss of the discriminator (Figure 5e) shows the increasing trace. However, it is marginally greater in magnitude than adversarial loss. This means the generator is doing a good job in emulating the observed patterns. Mean Absolute Error trace (Figure 5f) shows a pattern like the content loss.

Figure 5: (a) Total Loss of training, (b) Content Loss, (c) Structural Loss, (d) Adversarial Loss, (e) Discriminator Loss, (f) Mean Absolute Error of the training.

For final simulation we trained the model with all the available input data. This also enabled us to compare the model against other model effectively. For the total loss of the model, error continued to decrease even after 10,000 iterations (Figure S1a). The content loss stabilized around a value 1e-3 after 10,000 iterations (Figure S1b), while the structural loss stabilized around a value of 2.4e-2 (Figure S1c). The adversarial loss in Figure S1d keeps on increasing after 10,000 iterations indicating the discriminator is resolving the differences between the observed and predicted precipitation patterns. The discriminator error also keeps on increasing (Figure S1e), signifying good performance of the generator. However, the loss of the discriminator is still lower than the adversarial error. As an extra diagnostic we also tracked widely popular loss metric mean absolute error. Figure S1f shows the trace of mean absolute error which stabilized around a value of 0.1 mm/day after 10,000 iterations.

Figure 6: Temporal median of annual maximum precipitation (a) Observed, and (b) Simulated.
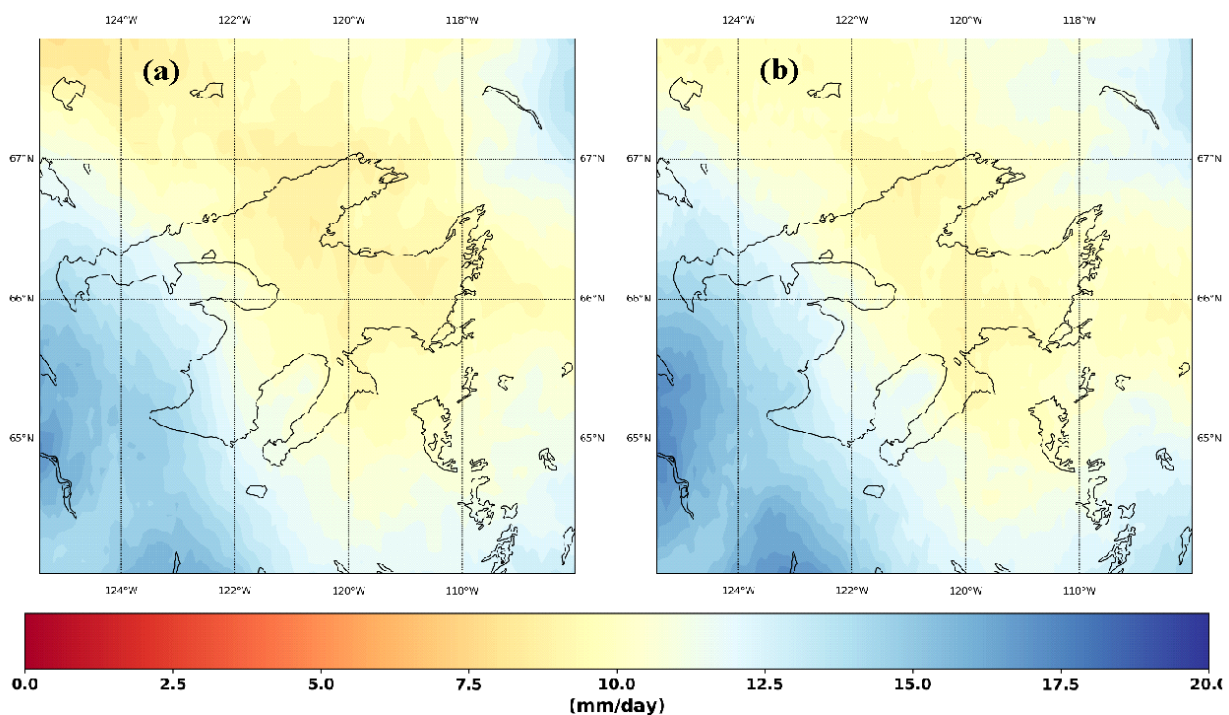

Figure 6 shows the temporal mean of observed (Figure 6a) and downscaled (Figure 6b) annual maximum precipitation. The regional patterns are well captured by the model. The observed pattern of high rainfall in the south western high region with orographic influences and the low rainfall over the lake is captured by the model. While more work is required to understand the performance of structural loss training of climate model outputs in different geographic contexts, the results here show that different processes are captured which yield more accurate overall output products. This is evident in the error diagnostics of the downscaling performance outlined in Figure 7. For the temporal mean absolute percent error (Figure 7a), maximum error is around 5%, however this error is mostly confined to the north-western part of the domain. We think this is due to the low observed rainfall without any special features in the models (which can again be attributed to absence of orographic or land cover features) in this region attributed to this high

percent error. Figure 7b shows the temporal correlation between the observed and downscaled precipitation. The minimum correlation is 0.995 which is well beyond acceptable limit. However, interestingly the correlation values show geometrical rectilinear orientation. We hypothesized that this is an artifact of the convolution filters used in the generator for downscaling, and thus this result should be interpreted with caution. Calculating p-values for the Kolmogorov-Smirnov test for equivalency of temporal distribution of observed and downscaled annual maximum (Figure 7c) indicate that we cannot reject the equivalency of the distributions between the observed and downscaled (null hypothesis) at 90% confidence limit.

Figure 7: Performance of the downscaling (a) Mean Absolute Percentage Error, (b) Temporal Correlation, and (c) P-value for Kolmogorov-Smirnov test of temporal distribution equivalency.
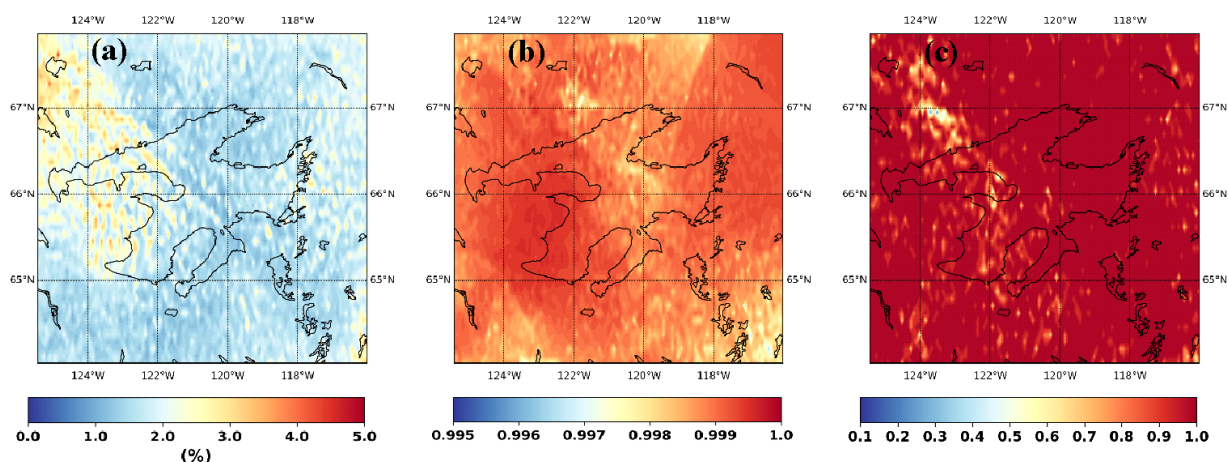

Table 3 shows the performance of different models and loss functions. The Generative adversarial network is compared against the model trained using traditional mean absolute error and another network with principal component mapping. The performance functions are Mean Absolute Error (MAE), NS coefficient, Correlation coefficient, and p-value of the two sample KS test. In all the performance measure the performance of GAN is superior followed by MAE based model. PCA based model performed poorly compared to the other model. However, all models performed reasonably good.

Table 3: Performance comparison of different downscaling models

Performance/Model	GAN	MAE	PCA
MAE	1.71	1.85	2.7
NS	0.996	0.995	0.991
Correlation	0.9987	0.9986	0.995
KS p-value	~1e-48	~1e-25	~1e-8

4.0 Conclusions

To our knowledge, this is the first attempt to downscale spatial fields of climate variables, utilizing CNN and adversarial training. The major conclusions from our study are summarized below;

1. Our framework can utilize diverse information present in different AOGCM simulations to create a spatially coherent field close to observation. The approach is similar in spirit to Reliability Ensemble Averaging (REA) proposed by Girogi et. al., 2001 within a CNN and adversarial training context.
2. Our model demonstrates good performance for downscaling extreme precipitation which is generally considered less predictable than mean climate/precipitation (Haylock et. al. 2006; Frei et al. 2006; Hundecha et. al. 2008).
3. The MSSIM index allowed us to get the insight into the model's regional characteristics and indicates the only point-based error functions which are widely used in statistical downscaling, may not be enough to reliably simulate the regional characteristics.
4. Further use of total loss function which is combination of adversarial, content, and structural loss within a CNN-based downscaling method may lead to higher quality downscaled products.
5. The adversarial loss is able to provide meaningful gradient to the weight optimization when traditional loss functions fails in near convergence variabilities.

These results hold considerable potential for fields investigating climate change impacts and climate processes. In many sectors, the scale mismatch between AOGCMs and spatial scales at the level of impacts (e.g., hydrological, ecological, economic) make studying direct linkages and outcomes of climate models difficult. Downscaling methods in general help to bridge this divide by making climate model information more meaningful and useful for scientists working across disciplines at more localized geographic scales. Our analysis shows that incorporating structure into the training of CNN-based statistical downscaling models reproduces observational gridded data for annual daily maximum rainfall in northern Canada, where station observational data is sparse and where climate change information is greatly needed. Further work developing and evaluating the CliGAN modelling framework in different geographic and thematic contexts is needed to realize the benefits to the wider scientific community.

References:

Arjovsky M., Chintala S., and Bottou L.; (2017). Wasserstein gan. arXiv preprint arXiv:1701.07875.

Ben Niu and Weilei Wen and Wenqi Ren and Xiangde Zhang and Lianping Yang and Shuzhen Wang and Kaihao Zhang and Xiaochun Cao and Haifeng Shen (2020). Single Image Super-Resolution via a Holistic Attention Network. arXiv 2008.08767.

Brandsma, T., and T. A. Buishand (1998), Simulation of extreme precipitation in the rhine basin by nearest-neighbour resampling, *Hydrology and Earth System Sciences*, *325*, 195–209.

Brissette, F., R. Leconte, and M. Khalili (2007), Efficient stochastic generation of multisite synthetic precipitation data, *Journal of Hydrology*, *345*, 121–133.

Brown, B. G., and R. W. Katz (1995), Regional analysis of temperature extremes:spatial analog for climate change?, *Journal of Climate*, *8*, 108–119.

Cannon, A. J., and P. H. Whitfield (2002), Downscaling recent streamflow conditions in british columbia, canada using ensemble neural network models, *Journal of Hydrology*, *259*, 136–151.

Chaudhuri et. al.; (2017). C. Chaudhuri, and R. Srivastava. A novel approach for statistical downscaling of future precipitation over the Indo-Gangetic Basin. Journal of hydrology, https://doi.org/10.1016/j.jhydrol.2017.01.024, 2017.

Cheng, J., Kuang, Q., Shen, C., Liu, J., Tan, X., & Liu, W. (2020). ResLap: Generating High-Resolution Climate Prediction Through Image Super-Resolution. IEEE Access, 8, 39623-39634.

Chen, H., Sun, J. and Chen, X. (2014), Projection and uncertainty analysis of global precipitation-related extremes using CMIP5 models. Int. J. Climatol., 34: 2730-2748. doi:10.1002/joc.3871

C. Prudhomme, D. Jakob, C. Svensson, (2003). Uncertainty and climate change impact on the flood regime of small uk catchments. J. Hydrol., 277 (2003), pp. 1-23

Crane, R. G., and B. C. Hewitson (1998), Doubled co2 precipitation changes for the susquehanna basin: down-scaling from the genesis general circulation model, *International Journal of Climatology*, *18*, 65–76.

Dibike, Y. B., and P. Coulibaly (2005), Hydrologic impact of climate change in the saguenay watershed: Comparison of downscaling methods and hydrologic models, *Journal of Hydrology*, *307*, 145–163.

Diederik P. Kingma, Jimmy Ba (2014) Adam: A Method for Stochastic Optimization. *Proceedings of the 3rd International Conference on Learning Representations (ICLR),*

Dong C., Loy C.C., He K., Tang X., (2014). Learning a Deep Convolutional Network for Image Super-Resolution. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8692. Springer, Cham

Fischlin A et al.; (2007).Climate Change 2007: Working Group II: Impacts, Adaptation and Vulnerability Cambridge Univ. Press, New York (2007)

Ford, J. D., McDowell, G., & Pearce, T. (2015). The adaptation challenge in the Arctic. Nature Climate Change, 5(12), 1046–1053. https://doi.org/10.1038/nclimate2723

Frei C, Schöll R, Fukutome S, Schmidli J, Vidale PL. 2006. Future change of precipitation extremes in Europe: intercomparison of scenarios from regional climate models. *Journal of Geophysical Research* **111**: D06105, DOI: 10.1029/2005JD005965.

Giorgi and Mearns; (2003). F. Giorgi, L.O. Mearns. Probability of regional climate change calculated using the reliability ensemble averaging (REA) method. Geophys. Res. Lett., 30 (1629) (2003), 10.1029/2003GL017130

Goodfellow et. al., 2014; I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Advances in Neural Information Processing Systems (NIPS), pages 2672–2680, 2014. 3, 4, 6

Gutierrez, J. M., A. S. Cofino, R. Cano, and M. A. Rodriguez (2004), Clustering methods for statistical downscaling in short-range weather forecasts, *Monthly Weather Review*, *132*, 2169–2183.

Hanson, C. L., and G. L. Johnson (1998), Gem (generation of weather elements for multiple applications): its application in areas of complex terrain, *Hydrology, Water Resources and Ecology in Headwaters*, *248*, 27–32.

Haylock MR, Cawley GC, Harpham C, Wilby RL, Goodess CM. 2006. Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios. *International Journal of Climatology* **26**: 1397–1415.

Hewitson, B. C., and R. G. Crane (1992), Regional-scale climate prediction from the giss gcm, *Palaeogeogr Palaeoclimatol Palaeoecol*, *97*, 249–267.

Hughes, J. P., and P. Guttorp (1994), A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena, *Water Resources Research*, *30(5)*, 1535–1546.

Hundecha Y, Bárdossy A. 2008. Statistical downscaling of extremes of daily precipitation and temperature and construction of their future scenarios. *International Journal of Climatology* **28**: 589–610.

Hutchinson, M. F., McKenney, D. W., Lawrence, K., Pedlar, J. H., Hopkinson, R. F., Milewska, E., & Papadopol, P. (2009). Development and Testing of Canada-Wide Interpolated Spatial Models of Daily Minimum–Maximum Temperature and Precipitation for 1961–2003. *Journal of Applied Meteorology and Climatology*, *48*(4), 725–741. https://doi.org/10.1175/2008JAMC1979.1

IPCC, 2007. Climate change 2007: impacts, adaptation and vulnerability. contribution of working group ii to the fourth assessment report of the intergovernmental panel on climate change. Cambridge Univ. Press, New York.

Ji, Y., Zhi, X., Tian, Y., Peng, T., Huo, Z., and Ji, L. (2020): Downscaling of Precipitation Forecasts Based on Single Image Super-Resolution, EGU General Assembly 2020, Online, 4–8 May 2020, EGU2020-8533, https://doi.org/10.5194/egusphere-egu2020-8533, 2020

Jianxin Cheng, Jin Liu, Zhou Xu, Chenkai Shen, Qiuming Kuang (2020). Generating High-Resolution Climate Prediction through Generative Adversarial Network. Procedia Computer Science, Volume 174, 2020, Pages 123-127, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.06.067.

Katz, R. W., and M. B. Parlange (1996), Mixtures of stochastic processes: applications to statistical downscaling, *Climate Research*, *7*, 185–193.

Krizhevsky A, Sutskever I, and Hinton. G E (2012) Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105,

Kuchar, L. (2004), Using wgenk to generate synthetic daily weather data for modelling of agricultural processes, *Mathematics and Computers in Simulation*, *17*, 69–75.

Lantz, T. C., & Kokelj, S. V. (2008). Increasing rates of retrogressive thaw slump activity in the Mackenzie Delta region, N.W.T., Canada. Geophysical Research Letters, 35(6). https://doi.org/10.1029/2007GL032433

Lecun Y., Bottou L., Bengio Y., and Haffner P., (1998) "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov.

Manzanas, R., Lucero, A., Weisheimer, A. *et al.* Can bias correction and statistical downscaling methods improve the skill of seasonal precipitation forecasts?. *Clim Dyn* **50,** 1161–1176 (2018). https://doi.org/10.1007/s00382-017-3668-z

Maraun, D., et al. (2010), Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, RG3003, doi:10.1029/2009RG000314.

Matthew D. Zeiler (2012), ADADELTA: An Adaptive Learning Rate Method, arXiv:1212.5701

Maurer, E. (2007), Uncertainty in hydrologic impacts of climate change in the sierra nevada, california under two emissions scenarios, Climatic Change, 82, 309–325, doi: 10.1007/s10584-006-9180-9 r.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth (2004), Quantifying uncertainties in climate change from a large ensemble of general circulation model predictions, *Nature*, *430*, 768–772, doi: 10.1038/nature02771.

NRC; 2008. NRCEcological Impacts of Climate Change The National Academy Press, Washington, DC (2008), 10.17226/12491

Olaf Ronneberger, Philipp Fischer, and Thomas Brox, 2015. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. Springer,

Onishi, Ryo, Daisuke Sugiyama, and Keigo Matsuda (2019). "Super-Resolution Simulation for Real-Time Prediction of Urban Micrometeorology." SOLA 15.0 (2019): 178–182. Crossref. Web.

Prudhomme, C., D. Jakob, and C. Svensson (2003), Uncertainty and climate change impact on the flood regime of small uk catchments, *Journal of Hydrology*, *277*, 1–23.

Plouffe, C. C. F., Robertson, C. and Chandrapala, L. 2015. Comparing interpolation techniques for monthly rainfall mapping using multiple evaluation criteria and auxiliary data sources: A case study of Sri Lanka. Environmental Modelling and Software, 67:57-71

Quinton, W. L., Hayashi, M., & Chasmer, L. E. (2011). Permafrost-thaw-induced land-cover change in the Canadian subarctic: Implications for water resources. Hydrological Processes, 25(1), 152–158. https://doi.org/10.1002/hyp.7894

Randall, D., R. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Srinivasan, R. Stouffer, A. Sumi, and K. Taylor (2007), Cilmate models and their evaluation. in: Climate change 2007: The physical science basis, in Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M.Tignor, and H. Miller, Cambridge University Press,, Cambridge, United Kingdom and New York, NY, USA.

Richardson, C. (1981), Stochastic simulation of daily precipitation, temperature, and solar radiation, *Water Resources Research*, *17*, 182–190.

Ronneberger, Olaf; Fischer, Philipp; Brox, Thomas (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". arXiv:1505.04597.

Semenov, M. A., and E. M. Barrow (1997), Use of a stochastic weather generator in the development of climate change scenarios, *Climatic Change*, *35*, 397–414.

SHI, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., & WOO, W. (2015). Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 28* (pp. 802–810). Curran Associates, Inc. http://papers.nips.cc/paper/5955-convolutional-lstm-network-a-machine-learning-approach-for-precipitation-nowcasting.pdf

Schoof, T. Z., A. Arguez, J. Brolley, and J. J. O'Brien (2005), A new weather generator based on spectral properties of surface air temperatures, Agricultural and Forest Meteorology, 135, 241–251.

Sebastian Ruder (2016), An overview of gradient descent optimization algorithms, arXiv:1609.04747

Sharif, M., and D. H. Burn (2006), Simulating climate change scenarios using an improved k-nearest neighbor model, *Journal of hydrology*, *325*, 179–196.

Smith, R. L., C. Tebaldi, D. Nychka, and L. Mearns (2009), Bayesian modeling of uncertainty in ensembles of climate models, Journal of the American Statistical Association, 104(485), 97–116, doi:10.1198/jasa.2009.0007.

Soltani, A., and G. Hoogenboom (2003), A statistical comparison of the stochastic weather generators wgen and simmeteo, *Climate Research*, *24*, 215–230.

Stengel, K., Glaws, A., Hettinger, D., & King, R. N. (2020). Adversarial super-resolution of climatological wind and solar data. Proceedings of the National Academy of Sciences, 117(29), 16805-16815.

Tran Anh, D, Van, SP, Dang, TD, Hoang, LP. Downscaling rainfall using deep learning long short-term memory and feedforward neural network. *Int J Climatol*. 2019; 39: 4170– 4188. https://doi.org/10.1002/joc.6066

Trigo, R., and J. Palutikof (2001), Precipitation scenarios over iberia: a comparison between direct gcm output and different downscaling techniques, *Journal of Climate*, *14*, 4422–4446.

Tripathi, S., V. V. Srinivas, and R. S. Nanjundiah (2006), Downscaling of precipitation for climate change scenarios: a support vector machine approach, *Journal of Hydrology*, *330*, 621–640.

Tryhorn, L. and DeGaetano, A. (2011), A comparison of techniques for downscaling extreme precipitation over the Northeastern United States. Int. J. Climatol., 31: 1975-1989. doi:10.1002/joc.2208

Trzaska, Sylwia & Schnarr, Emilie. (2014). A Review of Downscaling Methods for Climate Change Projections.

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., & Ganguly, A. R. (2017, August). Deepsd: Generating high resolution climate change projections through single image super-resolution. In Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining (pp. 1663-1672).

Villani C.; 2008. Optimal transport: old and new, volume 338. Springer Science & Business Media, 2008.

Vors, L. S., & Boyce, M. S. (2009). Global declines of caribou and reindeer. Global Change Biology, 15(11), 2626–2633. https://doi.org/10.1111/j.1365-2486.2009.01974.x

Wang, Q., Yuan, Z., Du, Q., & Li, X. (2019). GETNET: A General End-to-End 2-D CNN Framework for Hyperspectral Image Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(1), 3–13. https://doi.org/10.1109/TGRS.2018.2849692

Wenjie Ai and Xiaoguang Tu and Shilei Cheng and Mei Xie (2020). Single Image Super-Resolution via Residual Neuron Attention Networks arXiv 2005.10455.

Wetterhall, F., S. Halldin, and C. Xu (2005), Statistical precipitation downscaling in central sweden with the analogue method, Journal of Hydrology, 306, 174–190.

Widmann, M., C. S. Bretherton, and S.-J. EP (2003), Statistical precipitation downscaling over the north-western united states using numerically simulated precipitation as a predictor, Journal of Climate, 16(5), 799–816.
Wilby, R. L., H. Hassan, and K. Hanaki (1998), Statistical downscaling of hydrometeorological variables using general circulation model output, *Journal of Hydrology*, *205*, 1–19.

Wilby, R. L., C. W. Dawson, and E. M. Barrow (2002), Sdsm – a decision support tool for the assessment of regional climate change impacts, *Environ Modell Softw*, *17*, 147–159.

Wilby, R. L., and I. Harris; (2006), A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the river thames, uk,, *Water Resources Research*, *42*, W02419, doi:10.1029/2005WR004065.

Wilcox, E. J., Keim, D., de Jong, T., Walker, B., Sonnentag, O., Sniderhan, A. E., Mann, P., & Marsh, P. (2019). Tundra shrub expansion may amplify permafrost thaw by advancing snowmelt timing. Arctic Science, 5(4), 202–217. https://doi.org/10.1139/as-2018-0028

Wilks, D. S., and R. L. Wilby (1999), The weather generation game: a review of stochastic weather models, *Progress in Physical Geography*, *23*, 329–357.

Yann LeCun and Yoshua Bengio, 1998. Convolutional networks for images, speech, and time series. In The handbook of brain theory and neural networks, Michael A. Arbib (Ed.). MIT Press, Cambridge, MA, USA 255-258.

Yates, D., S. Gangopadhyay, B. Rajagopalan, and K. Strzepek (2003), A technique for generating regional climate scenarios using a nearest-neighbour algorithm, *Water Resources Research*, *39*, 1199–1213.

Zorita, E., and H. von Storch (1999), The analog method - a simple statistical downscaling technique: comparison with more complicated methods, *Journal of Climate*, *12*, 2474–2489.

Z. Wang, E. P. Simoncelli and A. C. Bovik, (2003). Multiscale structural similarity for image quality assessment, The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 2003, pp.1398-1402Vol.2.

Declarations
Not Applicable

Funding
This work was funded in part by the Global Water Futures research programme.

Conflicts of interest/Competing interests
The authors declare that there are no competing interests.

Availability of data and material
Any data that support the findings of this study, not already publicly available, are available from the corresponding author, C. Chaudhuri, upon reasonable request.

Author Contribution
The idea behind this research was conceived, implemented, and written equally by both authors.

Consent to participate
Not Applicable

Consent for publication
Not Applicable

Code availability
The custom code used in this study, not already publicly available, are available from the corresponding author, C. Chaudhuri, upon reasonable request.