Enabling Smart Dynamical Downscaling of Extreme Precipitation Events with Machine Learning

Xiaoming Shi[1,2]

[1]Division of Environment and Sustainability, Hong Kong University of Science and Technology
[2]Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology

Key Points:

- Dynamical downscaling at $\sim 1\,\mathrm{km}$ resolution produces reliable estimations of extreme rainfall but is computationally expensive.
- Machine learning (ML) makes smart dynamical downscaling (SDD) possible, where ML models filter out irrelevant large-scale patterns.
- We demonstrate that SDD can be enabled by support vector machines or deep neural networks, of which the latter performs better.

Corresponding author: Xiaoming Shi, `shixm@ust.hk`

Abstract

The projection of extreme convective precipitation by global climate models (GCM) exhibits significant uncertainty due to the coarse resolution of GCMs, which cannot resolve fine-scale processes. Direct dynamical downscaling (DDD) of regional climate at convection-permitting resolutions ($\sim 1\,\mathrm{km}$) provides valuable insight into the potential changes in extreme precipitation, but its computational expense is formidable. Here we document the effectiveness of machine learning in enabling smart dynamical downscaling (SDD), which performs downscaling only for a small subset of GCM data. Trained with reanalysis and satellite data for three Asian cities, support vector machines can filter out approximately 87% to 94% of circulation data, which are irrelevant to extremes. Deep convolutional neural networks, trained with larger data sets, can filter out more than 97% of circulation data and in the selected subset, retrieve 72% to 81% of the circulation patterns responsible for extreme events (rain intensity higher than the 99th percentile).

Plain Language Summary

Climate scientists use supercomputers to simulate the climate and predict how it may change under global warming. Extreme precipitation, which can disrupt the society by causing disasters like floods and landslides, is of great interest in climate studies. However, replicating severe rainstorms on a supercomputer, especially those in tropical and subtropical areas, is not easy because those rainstorms often contain fine-scale details that cannot be represented confidently without an extensive amount of computational resource. If we use computationally cheap computer models to simulate those rainstorms, we obtain results with substantial uncertainties. If we use computationally expensive ones, we cannot simulate many scenarios, and thus cannot be confident about the results. The power of machine learning in pattern recognition is here used to help modelers use their computational resources more efficiently. Instead of simulating all kinds of weather events, including unimportant ones, at high resolutions, we use machine learning algorithms to search coarse resolution climate data for those large-scale weather patterns that are more likely to cause severe rainstorms. Then modelers can make more efficient use of supercomputing resources by simulating impactful weather events only and advance our understanding of extreme precipitation.

1 Introduction

Extreme precipitation events often disrupt the society by causing disasters such as floods and landslides. Thus, predicting the response of precipitation extremes to global warming is crucial for our adaptation to climate change. However, predicting such changes is not straightforward because the performance of numerical simulation of extreme precipitation events is sensitive to model resolution (Li et al., 2018; Van Der Wiel et al., 2016), while grid spacings of current-generation climate models are still at a coarse $\sim 1^\circ$ resolution in the horizontal. Previous studies have demonstrated that to accurately predict future changes in extreme precipitation events, especially those associated with severe convection, it is necessary to resolve local storm dynamics with kilometer-scale resolutions, which are the so-called convection-permitting resolution (Kendon et al., 2014, 2017). Such high model resolution is necessary not only because of the small spatial scale of convective cells, but also because the essential roles played by the interaction between convection and large-scale dynamics, air-sea coupling, and topographic forcing in determining the intensity of extreme events (Nie et al., 2016; Kendon et al., 2017; Rainaud et al., 2017).

Modelers have been attempting to refine the resolution of global climate models, but the highest resolution so far is only $\sim 25\,\mathrm{km}$ (Haarsma et al., 2016). A direct dynamical downscaling (DDD) approach has been adopted in the regional climate simulations at convection-permitting resolutions and valuable findings have been obtained due to improved representation of fine-scale processes (Prein et al., 2015). For example, Prein et al. (2017) found that under the RCP8.5 scenario, the strengthening of precipitation intensity and the expansion of impact area will combine to give 80% increases in the total precipitation volume of mesoscale-convective systems in the US. However, DDD at the convection-permitting resolution has a very high demand on computational resources (Prein et al., 2015).

Is there a way to avoid the expensive computational cost of long-term DDD but still allow a convection-permitting resolution to be used? This question is the core problem we want to address in this study. When our concern is not the mean climate but instead a special kind of weather (e.g., extreme precipitation), we can save a tremendous amount of computational resources if we do not have to perform the DDD for every day of an extended period. In this study, we harness the power of machine learning to fulfill the goal of selecting a small subset of GCM data for the dynamic downscaling of extreme precipitation events. We call this modeling strategy smart dynamical downscaling (SDD).

Machine learning has been increasingly used in geoscience in recent years. In the atmospheric science community, it has applied to real-time nowcasting (Han et al., 2017; McGovern et al., 2017), physical parameterization (Brenowitz & Bretherton, 2019; Gagne et al., 2020), and weather forecast (Weyn et al., 2019; Chattopadhyay et al., 2020). Here we harness the power of machine learning in pattern recognition to enable SDD. Because previous studies have suggested that the exact strengthening rate of extreme precipitation mostly relies on dynamic, instead of thermodynamic, response to warming (Lenderink & Van Meijgaard, 2008; Shi & Durran, 2016; Pfahl et al., 2017), such an SDD approach is urgently needed for fully utilizing the information from GCM simulations to explore the effect of changes in large-scale eddy circulations on extreme precipitation.

2 Data and Methods

2.1 Reanalysis and Satellite Data

In this study, we train machine learning models with reanalysis data of circulation and satellite data of precipitation. We use the NCEP/NCAR Global Reanalysis Products (Kalnay et al., 1996) to represent the state of the atmospheric circulation. This data set has $2.5° \times 2.5°$ horizontal resolution. Though the original data set is available on 17 pressure levels, we use only the lowest 8 vertical levels between 1000 hPa to 300 hPa in this study. The variables we use to depict the large-scale circulation include 7 three dimensional variables: geopotential height, relative humidity, temperature, u- and v-components of horizontal wind, vertical (pressure) velocity, and vorticity. For the training of support vector machine (SVM) models, we also included 3 additional single-level variables — surface pressure, tropopause pressure, and precipitable water. The temporal resolution of the reanalysis data is 6 hours. The period we used from the reanalysis data is June 2000 to May 2019.

The precipitation data we used is the final precipitation, Level 3 data of the Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (GPM IMERG; Huffman et al., 2019). This data set has 0.1° spatial resolution and 30 min temporal resolution originally. We used the data set between the period of June 2000 to May 2019. Because the reanalysis data has 6-hour temporal resolution, we average the GPM data in time to get the mean precipitation rate in the 6-hour intervals between two consecutive time slices of reanalysis data. To make it represent impactful events, we also average the precipitation data spatially to obtain coarse-grained data on a $0.5° \times 0.5°$ grid.

The precipitation data from reanalysis was not used in this study because they only represent precipitation from large-scale circulation and have a significant bias. Supplementary Figure S1 shows the histogram of precipitation intensities and the temporal distribution of extreme precipitation events for the grid point nearest to Hong Kong in reanalysis and GPM data. From Figure S1a, it is evident that the reanalysis data significantly underestimate the rain rates of intense precipitation events. Figure S1b suggests reanalysis cannot represent the correct timing of extreme events either. The satellite data suggest that extreme precipitation events peak in May and June, consistent with rain gauge observation (Su et al., 2019). In contrast, reanalysis data exhibit peak season of extreme rainfall from July to September.

Our application of machine learning focused on the area surrounding three Asian cities, Hong Kong (HK), Manila (MN), and Singapore (SG). In the training of support vector machines (SVM), the input to one SVM are the circulation data (7 three-dimensional variables and 3 two-dimensional variables) in a $15° \times 15°$ region centered at one of those three cities. Each time slice of the reanalysis data is categorized as producing "significant rain" or "no significant rain", "light rain" or "heavy rain", based GPM precipitation in the next 6 hours at the $0.5° \times 0.5°$ cell centered at the same city. The SVMs were trained to predict the correct categories from the large-scale circulation data.

In our application of convolutional neural networks (CNN), we identified three "useful" regions surrounding each of those three cities. Those regions share similar dynamic characteristics of rainy weather and gave us larger data sets for training CNNs. 6-hourly precipitation data of each $0.5° \times 0.5°$ cell within a "useful" regions are used to categorize the corresponding time (the beginning of 6-hour intervals) and location as producing "extreme rainfall" or "non-extreme rainfall". The input data for CNNs are the circulation data (7 three-dimensional variables) in $15° \times 15°$ square regions centered at each of the $0.5° \times 0.5°$ rainfall data cells within a "useful" region.

2.2 Support Vector Machine (SVM)

SVM is a machine learning model for binary classification problems. At its core, SVMs find a hyperplane in the feature space of data and separate points in the feature space into two different groups. The hyperplane in feature space is defined as the set of points $\mathbf{x}$ satisfying

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \tag{1}$$

where the vector $\mathbf{w}$ and scalar b for the best hyperplane are determined by an optimization procedure which maximizes the margin between two classes in feature space. For a linearly separable problem, $\mathbf{w}$ and b are entirely determined by those sample points that are closest to the best hyperplane. Those sample points are called support vectors. When the data are not linearly separable, one can use a soft margin technique to allow a small number of instances to be misclassified.

Furthermore, in nonlinear classification problems, it is common to use a kernel function to replace dot product for operating the optimization algorithm in a transformed feature space implicitly. In our application, we used the Gaussian radial basis function,

$$G(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma^2}\right) \tag{2}$$

where $\sqrt{2}\sigma$ is called kernal scale. Besides σ, the other hyperparameter for training an SVM is the box constraint which appears in the soft margin formula and decides the tolerance level of misclassification by the SVM.

We used MATLAB R2019a in our study to train SVMs with 10-fold cross-validation. The hyperparameters were tuned using Bayesian optimization.

2.3 Convolutional Neural Network (CNN)

In its essence, a neural network transforms the signal from one layer of neurons to the next through a linear transformation and the use of a nonlinear activation function,

$$\mathbf{z}^{[k]} = \mathbf{W}^{[k]}\mathbf{a}^{[k-1]} + \mathbf{b}^{[k]}\,, \qquad \mathbf{a}^{[k]} = g^{[k]}(\mathbf{z}^{[k]}) \tag{3}$$

where $\mathbf{a}^{[k]}$ is the activation of Layer k, $\mathbf{W}^{[k]}$ is a weight matrix, and $\mathbf{b}^{[k]}$ is a bias vector. $g^{[k]}$ is a non-linear activation function. For Layer 0, the activation $\mathbf{a}^{[0]}$ is simply the vector of input data $\mathbf{x}$. A fully connected layer in a deep neural network connects every neuron in the previous layer to every neuron in the current layer. A convolution layer, by contrast, has multiple filters, which are used to convolve a sub-block of the activation data from the previous layer and connect that subset of neurons in the previous layer to a neuron in the current layer.

We again used MATLAB R2019a to train the CNNs in this study. 75% of input data were used to train the models and 25% used for cross-validation. The stochastic gradient descent with momentum (SGDM) method was used to find the optimal weights and bias of the CNNs.

2.4 Performance Metrics

In the training of SVMs and CNNs, algorithms try to achieve the highest classification accuracy. However, because extreme precipitation events are only a small fraction of the entire data set, the accuracy of trained models always appears intuitively high. Thus, in our discussion below, we report the performance of trained models in terms of precision and recall.

The precision is usually defined as the number of true positive instances divided by the number of true and false positive instances. The recall is usually defined as the number of true positive instances divided by the number of true positive and false negative cases. However, because the distribution of rain intensity is continuous and we want to ensure the machine learning models can retrieve most of the extreme precipitation events, it is often necessary to train a machine learning model using a relatively low percentile of precipitation rate as the criterion for categorizing data, but evaluate the effectiveness of the trained model in retrieving the extreme events defined with a different, higher percentile. Thus, we adopt the following notation for precision and recall,

$$P_y^{\mathrm{M}} = \frac{|\{r > r_y\} \cap \{r' > r_y\}|}{|\{r' > r_y\}|}\,, \tag{4}$$

$$R_y^{\mathrm{M}} = \frac{|\{r > r_y\} \cap \{r' > r_y\}|}{|\{r > r_y\}|}\,, \tag{5}$$

where P_y^{M} and R_y^{M} are the precision and recall of the model M when precipitation rates greater than the y-th percentile of rain rate, r_y, are labeled as positive. r_y may be *different* from the actual threshold used in categorizing data when training the model M. $\{r > r_y\}$ represent the set of instances for which real precipitation rates (r) are higher than r_y, and $\{r' > r_y\}$ is the set of instances for which the model M predicts their precipiation rates (r') are greater than r_y. r' was not computed by the machine learning models explicitly, but rather the condition, $r' > r_y$, was judged by the classification model M.

3 Results

3.1 Dual SVM Model

We first attempted to select instances for extreme events by training a pair of SVMs. The first SVM (SVM1) tells whether the circulation data of a time slice can produce "sig-

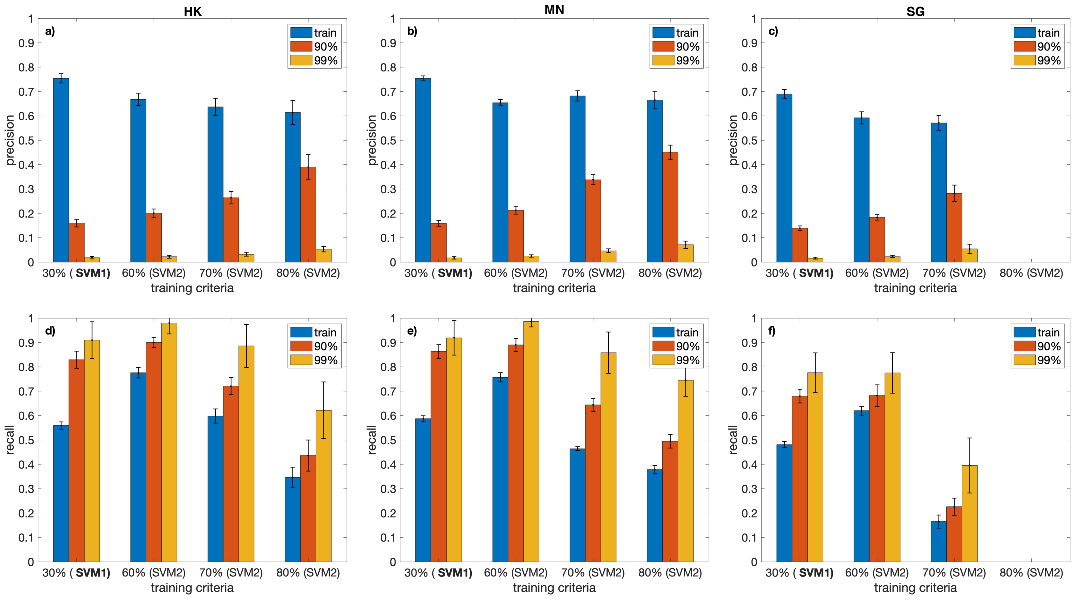

Figure 1. Precision (a–c) and recall (d–f) of the trained SVMs. a) and d) are the SVMs for Hong Kong (HK), b) and e) for Manila (MN), c) and f) for Singapore (SG). The SVMs were trained for the thresholds indicated below the horizontal axis, but their performance is evaluated against the training criteria and the 90th and 99th percentiles of rain rates.

nificant" rainfall or not. It was trained with the 30th percentile of rain rates as the threshold for "significant" rainfall. The subset of circulation data which SVM1 predicts to produce significant rain is then adopted by the second SVM (SVM2), which uses a higher percentile (60th, 70th, or 80th) as its criterion for "extremes". We found that this dual-SVM strategy can yield higher precision and recall than using one SVM to predict "extremes" directly.

Figure 1 shows the performance of the Dual SVM model trained with the data for the three cities, HK, MN, and SG. The precision of SVM1 for its training criteria, P_{30}^{SVM1}, is around 0.7, and the recall of SVM1 for its training criteria, R_{30}^{SVM1}, is between 0.48 to 0.59. These recall rates are not very high. However, if we target retrieve precipitation event with rain rates higher than the 90th and 99th percentiles, we can find that the corresponding recall rates, R_{90}^{SVM1} and R_{99}^{SVM1}, are between 0.82 to 0.92 for HK and MN, and between 0.69 to 0.79 for SG. Thus, the trained SVM1 is effective in retrieving the majority of extreme precipitation events. It should be noted that because we did not include rain rates lower than $0.05\,\mathrm{mm\,h^{-1}}$ in calculating the percentiles, Thus SVM1 eliminates much more than 30% circulation data from all time slices. Precipitation rates in HK, MN, and SG only exceed the corresponding 30th percentiles in 14.5%, 28.9%, and 29.4%, respectively, of the 19 years.

Figure 1 also shows the performance of SVM2 with respect to training criteria, as well as for real extreme events defined by the 90th and 99th percentiles. For SG, we were unable to obtain a converged solution when the training criterion was set as the 80th percentile. The precision of SVM2 when evaluation criteria are the 90th and 99th percentiles increases as the training criteria increase, a natural result of the narrowing mismatch between training and evaluation criteria. The recall of SVM2 decreases as the training criteria increases. A higher training threshold means that we can filter out more "irrelevant" instances. However, it also increases our chance of losing actual extreme events

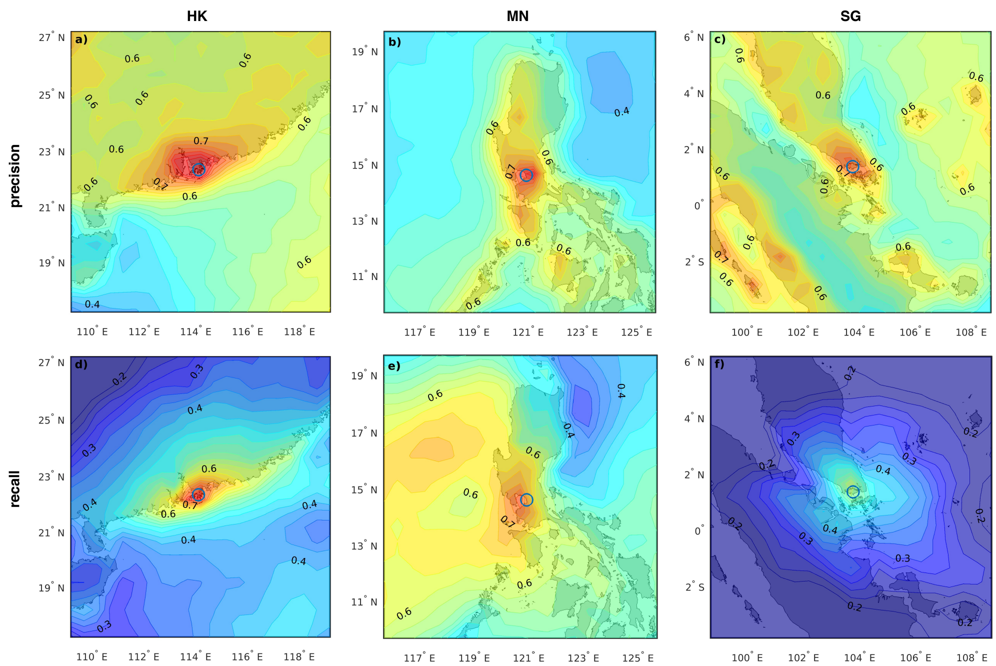

Figure 2. Precision (upper row) and recall (lower row) for the SVM1's which were trained for HK, MN, and SG but are applied to areas surrounding those cities here.

due to misclassification. Based on Fig. 1, the SVM2 trained with the 70th percentile of rain rates are the most balanced models for applications. If we target to retrieve extreme events defined by the 99th percentile in the selection, the SVM1 and the SVM2 trained with the 70th percentile can yield combined recall rates of $R_{99}^{\mathrm{SVM1}} R_{99}^{\mathrm{SVM2}} = 0.81,\ 0.79,$ and $0.31,$ for HK, MN, and SG, respectively.

As mentioned above, the number of time slices which produced significant rainfall is smaller than 30% (15% for HK) of the entire 19-year period of our data. If we use the SVM2 trained with the 70th percentile as the training criterion, SVM2 can eliminate approximately another 4/7 data from the subsets selected by SVM1. Thus, the Dual SVM models, in total, can eliminate approximately more than 87% (94% for HK) of circulation data from extreme events candidates. That is a significant saving of computational cost in dynamic downscaling. However, it still means that we need to "waste" a notable fraction of our computation to ensure most extreme events are kept by the Dual SVM models. The unsatisfactory performance of the Dual SVM model in SG data suggests we cannot obtain a very reliable subset of data if we want to study deep tropical extreme rainfall. Can we overcome this difficulty with other machine learning algorithms? The answer is positive. In Section 3.3, we will demonstrate that the use of deep neural network can yield significantly better performance than the Dual SVM models.

3.2 Useful Areas

Large data set is needed for deep learning to prevent over-fitting. When we focus on only one point on a map, the availability of observation data is limited. In the training of Dual SVM models above, we used 19 years of 6-hourly data, including 27,756 time slices. The areas surrounding the place of interest should be in a similar climate regime. Therefore, we evaluate the similarity of extreme event dynamics by applying the SVM1's

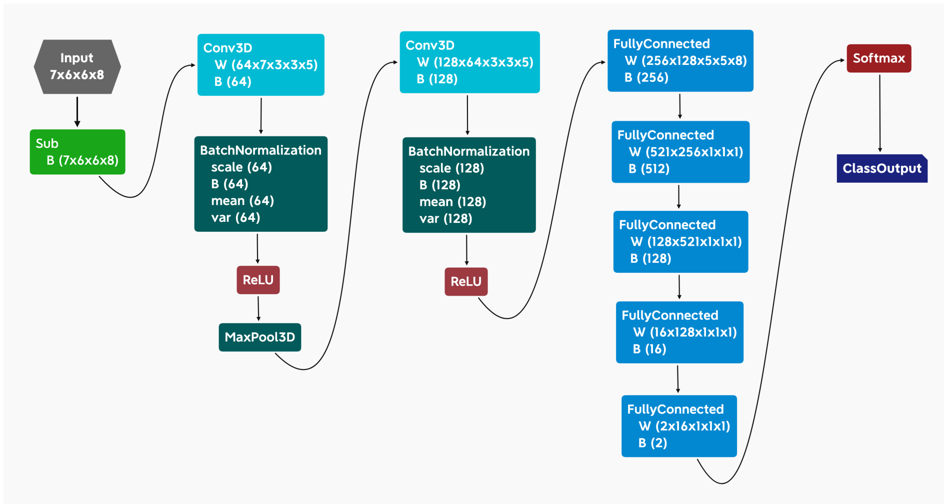

Figure 3. Structure of the CNNs trained in this study. The size of one input "image" is 7 (channels/variables) $\times$ 6 (height/latitude) $\times$ 6 (width/longitude) $\times$ 8 (depth/altitude). The first convolution layer has 64 filters and the second has 128 filters. The size of filters is 3 (height) $\times$ 3 (width) $\times$ 5 (depth). The number of neurons in the five fully connected layers is 256, 512, 128, 16, and 2, respectively.

trained above to regions surrounding HK, MN, and SG, the data of which, except right at the three cells of those three cities, were not seen during training SVM1.

Figure 2 shows the performance of the SVM1's in relevant regions. What might be surprising is that the spatial distribution of precision and recall exhibit some dependence on the terrain. What is most impressive is the case of HK. In Figure 2a and d, precision and recall are maximized along the South China coast, suggesting that precipitation events in HK are significantly affected by coastal location and topography.

Good precision and recall in Figure 2 indicate the applicability of the model trained at a city only to its surrounding areas, thereby the similarity in the dynamics of extreme events. We can define "useful" areas as those exhibiting relatively high precision and recall. For HK and MN, we select grid points at which $P_{30}^{\mathrm{SVM1}} > 0.60$ and $R_{30}^{\mathrm{SVM1}} > 0.50$. For SG, we select grid points at which $P_{30}^{\mathrm{SVM1}} > 0.45$ and $R_{30}^{\mathrm{SVM1}} > 0.35$. Using those "useful" areas allows us to enlarge the training data set for HK, MN, and SG to 1.2, 1.3, and 1.4 million instances, respectively.

3.3 Convolutional Neural Networks

The structure of the CNNs trained in this study is shown in Fig. 3. It is a series network with 2 three-dimensional convolution layers and 5 fully connected layers. The convolution layers are followed by batch normalization layers, rectified linear unit (ReLU) layers, and a three-dimensional max-pooling layer (for the first convolution layer) before

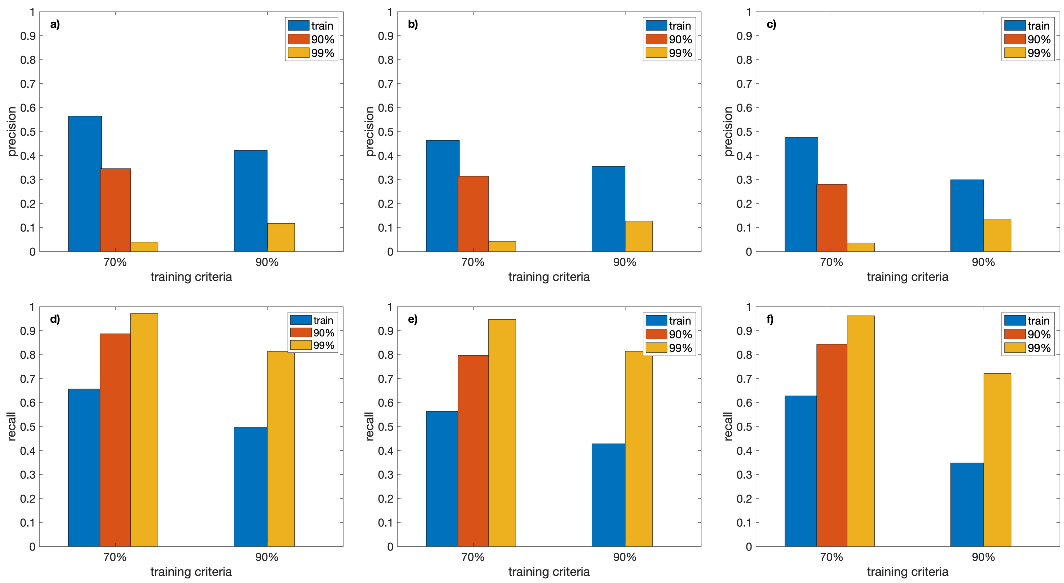

Figure 4. Precision (a–c) and recall (d–f) of the trained CNNs (RxNet). a) and d) are the RxNet for Hong Kong (HK), b) and e) for Manila (MN), c) and f) for Singapore (SG). The CNNs were trained for the thresholds indicated below the horizontal axis but their performance is evaluated against the training criteria and the 90th and 99th percentiles of rain rates.

connecting the fully connected layers. Training the CNNs is much more time-consuming on a multi-core CPU than training the SVMs. However, training the CNNs on a GPU is fast and can finish 100 epochs of iteration within a day. We trained our CNNs with two criteria for categorizing "extremes", the 70th and 90th percentiles of rain rates, and we call them RxNet70 (RN70) and RxNet90 (RN90). RxNet70 was trained for 100 epochs of iteration, RxNet90 was trained for 60 epochs. The latter were trained for fewer iterations because we found RxNet90 appears to converge faster than RxNet70, probably because the number of "extreme" events is significantly less in training the RxNet90's than that in training RxNet70's.

Figure 4 shows the performance of the trained RxNet70 and RxNet90. For training criteria, RxNet70 exhibits precision around 0.5 and recall around 0.6. When evaluated with higher percentiles (90th and 99th), its precision becomes lower than that for the training criterion, but its recall becomes high. R_{90}^{RN70} is between 0.8 and 0.9 for the three regions. R_{99}^{RN70} is above 0.95 for all three regions, which is much higher than the recall of the Dual SVM model mentioned above (< 0.81).

The trained RxNet90 exhibits precision between 0.3 to 0.4 for training criteria, which is lower than that of RxNet70. However, when evaluating its performance in retrieving the extreme events defined with the 99th percentile, P_{99}^{RN90} is significantly higher than P_{99}^{RN70}. The recall of RxNet90 for training criteria is slightly low, between 0.35 to 0.5 for the three regions. However, when being evaluated against the 99th percentile, R_{99}^{RN90} is 0.81, 0.81, and 0.72 for those three regions surrounding HK, MN, and SG, respectively. Those recall rates are decent, considering the rarity of the extreme events with rain rates higher than the 99th percentile.

The performance of the CNNs is significantly better than the Dual SVM models. The training of Dual SVM models failed to converge for SG when the training criterion was raised to the 80th percentile. By contrast, RxNet90 can reach a converged solution and has a decent recall for extreme events with rain rates higher than the 99th percentile. The optimal Dual SVM models can only filter out approximately 87% irrelevant instances of circulation data for MN and SG (94% for HK), but by contrast, applying RxNet90 to circulation data can filter out approximately more than 97% (98% for HK) of irrelevant instances. Meanwhile, the recall of RxNet90 for extreme events defined by the 90th and 99th percentiles is no less than that of the optimal Dual SVM models.

4 Conclusions

The sensitivity of tropical and subtropical extreme precipitation to global warming is highly uncertain. By constraining climate model results with satellite observation, O'Gorman (2012) estimated that the sensitivity of tropical extreme rainfall to global warming is 6–14 $\%\,\mathrm{K}^{-1}$, which contains a significant range of uncertainty. The high end of this estimate represents the result from not only thermodynamic scaling, which is approximately 7 $\%\,\mathrm{K}^{-1}$ due to the moistening of the atmosphere, but also from enhanced upward motions in extreme events. This dynamic strengthening is certainly possible, given that the high percentiles of convective available potential energy (CAPE) increase robustly in the tropics and subtropics of GCM simulations under warming (Singh et al., 2017). However, to what extent can the increase in CAPE be realized as ascending motions in extreme rainstorms is not precisely known.

To narrow the uncertainty in the estimation of future extreme precipitation, dynamic downscaling would give us the most reliable results. DDD is straightforward but prohibitively expensive regarding computational resource. Meanwhile, high degree of internal climate variability (Deser et al., 2012; Wallace et al., 2012) requires us to sample a reasonably large number of climate simulations if possible.

Here we demonstrated that machine learning can indeed enable SDD, in which only the large-scale patterns that have a high probability of producing extreme events are dynamically downscaled. In our study, the best performance was obtained when training CNNs using the 90th percentile of rain rates as the threshold for labeling "extremes". Because the distribution of precipitation intensities is continuous, it is unavoidable to have a significant number of misclassifications in the machine learning models. For this reason, we chose to train the machine learning model with a relatively low percentile (e.g., 90th percentile) as the categorizing criterion when we target to retrieve most of the extreme events defined by a higher percentile (e.g., 99th percentile). We found that trained deep neural network, RxNet90, is very effective in filtering out irrelevant large-scale circulation patterns and retaining the majority instances which are very likely to generate extreme events. We advocate using deep learning techniques to enable the SDD of extreme events in climate studies and advance our understanding of future climate.

Acknowledgments

The author acknowledges the support of the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. AoE/E-603/18) and thank Lin Su and Yongquan Qu for useful conversations. The reanalysis data was obtained from the Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory (https://rda.ucar.edu/datasets/ds090.0/). The GPM IMERG precipitation data was provided by the Goddard Earth Sciences Data and Information Services Center (GES DISC) (https://doi.org/10.5067/gpm/imerg/3b-hh/06).

References

Brenowitz, N. D., & Bretherton, C. S. (2019). Spatially extended tests of a neural network parametrization trained by coarse-graining. *J. Adv. Model. Earth*

Syst., *11*(8), 2728–2744.

Chattopadhyay, A., Nabizadeh, E., & Hassanzadeh, P. (2020). Analog forecasting of extreme-causing weather patterns using deep learning. *J. Adv. Model. Earth Syst.*, *12*(2), e2019MS001958.

Deser, C., Knutti, R., Solomon, S., & Phillips, A. S. (2012). Communication of the role of natural variability in future north american climate. *Nature Clim. Change*, *2*(11), 775–779.

Gagne, D. J., Christensen, H. M., Subramanian, A. C., & Monahan, A. H. (2020). Machine learning for stochastic parameterization: Generative adversarial networks in the lorenz'96 model. *J. Adv. Model. Earth Syst.*, *12*(3), e2019MS001896.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., … others (2016). High resolution model intercomparison project (highresmip v1. 0) for cmip6. *Geosci. Model Dev.*, *9*(11), 4185–4208.

Han, L., Sun, J., Zhang, W., Xiu, Y., Feng, H., & Lin, Y. (2017). A machine learning nowcasting method based on real-time reanalysis data. *J. Geophys. Res.*, *122*(7), 4038–4051.

Huffman, G., Stocker, E., Bolvin, D., Nelkin, E., & Jackson, T. (2019). *GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree × 0.1 degree V06, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC).* doi: 10.5067/GPM/IMERG/3B-HH/05

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., … others (1996). The ncep/ncar 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, *77*(3), 437–472.

Kendon, E. J., Ban, N., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., … Wilkinson, J. M. (2017). Do convection-permitting regional climate models improve projections of future precipitation change? *Bull. Amer. Meteor. Soc.*, *98*(1), 79–93.

Kendon, E. J., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., & Senior, C. A. (2014). Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nature Clim. Change*, *4*(7), 570–576.

Lenderink, G., & Van Meijgaard, E. (2008). Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nat. Geosci.*, *1*(8), 511–514.

Li, J., Chen, H., Rong, X., Su, J., Xin, Y., Furtado, K., … Li, N. (2018). How well can a climate model simulate an extreme precipitation event: A case study using the transpose-amip experiment. *J. Climate*, *31*(16), 6543–6556.

McGovern, A., Elmore, K. L., Gagne, D. J., Haupt, S. E., Karstens, C. D., Lagerquist, R., … Williams, J. K. (2017). Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bull. Amer. Meteor. Soc.*, *98*(10), 2073–2090.

Nie, J., Shaevitz, D. A., & Sobel, A. H. (2016). Forcings and feedbacks on convection in the 2010 pakistan flood: Modeling extreme precipitation with interactive large-scale ascent. *J. Adv. Model. Earth Syst.*, *8*(3), 1055-1072. doi: 10.1002/2016MS000663

O'Gorman, P. A. (2012). Sensitivity of tropical precipitation extremes to climate change. *Nat. Geosci.*, *5*(10), 697–700.

Pfahl, S., O'Gorman, P. A., & Fischer, E. M. (2017). Understanding the regional pattern of projected future changes in extreme precipitation. *Nature Clim. Change*, *7*(6), 423–427.

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., … others (2015). A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Rev. Geophys.*, *53*(2), 323–361.

Prein, A. F., Liu, C., Ikeda, K., Trier, S. B., Rasmussen, R. M., Holland, G. J., & Clark, M. P. (2017). Increased rainfall volume from future convective storms in

the us. *Nature Clim. Change*, *7*(12), 880–884.
Rainaud, R., Brossier, C. L., Ducrocq, V., & Giordani, H. (2017). High-resolution air–sea coupling impact on two heavy precipitation events in the western mediterranean. *Quart. J. Roy. Meteor. Soc.*, *143*(707), 2448–2462.
Shi, X., & Durran, D. (2016). Sensitivities of extreme precipitation to global warming are lower over mountains than over oceans and plains. *J. Climate*, *29*(13), 4779–4791.
Singh, M. S., Kuang, Z., Maloney, E. D., Hannah, W. M., & Wolding, B. O. (2017). Increasing potential for intense tropical and subtropical thunderstorms under global warming. *Proc. Natl. Acad. Sci. USA*, *114*(44), 11657–11662.
Su, L., Li, J., Shi, X., & Fung, J. C. (2019). Spatiotemporal variation in presummer precipitation over south china from 1979 to 2015 and its relationship with urbanization. *J. Geophys. Res.*, *124*(13), 6737–6749.
Van Der Wiel, K., Kapnick, S. B., Vecchi, G. A., Cooke, W. F., Delworth, T. L., Jia, L., ... Zeng, F. (2016). The resolution dependence of contiguous us precipitation extremes in response to co2 forcing. *J. Climate*, *29*(22), 7991–8012.
Wallace, J. M., Fu, Q., Smoliak, B. V., Lin, P., & Johanson, C. M. (2012). Simulated versus observed patterns of warming over the extratropical northern hemisphere continents during the cold season. *Proc. Natl. Acad. Sci. USA*, *109*(36), 14337–14342.
Weyn, J. A., Durran, D. R., & Caruana, R. (2019). Can machines learn to predict weather? using deep learning to predict gridded 500-hpa geopotential height from historical weather data. *J. Adv. Model. Earth Syst.*, *11*(8), 2680–2693.