https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

CCdownscaling: an open-source Python package for multivariable statistical climate model downscaling V1.0

Andrew D. Polasky[1], Jenni L. Evans[1,2], and Jose D. Fuentes[1]

[1]Department of Meteorology and Atmospheric Science, The Pennsylvania State University
[2]Institute for Computational and Data Sciences, The Pennsylvania State University, University Park

Correspondence: Jenni L. Evans (jle7@psu.edu)

Abstract. Statistical downscaling methods provide an essential bridge between low resolution global climate models and localized information needed by decision makers. As the demand for localized climate information continues to grow to make projections for a wide variety of applications, the need for software that can provide this sort of downscaled data grows with it. The CCdownscaling package described in the article provides a number of downscaling methods, including Self Organizing Maps, as well as a number of evaluation metrics for assessing downscale model skill. In this article, we describe the features of the CCdownscaling package, and show an example use case for downscaling temperature and precipitation. It is open-source and freely available for use in generating downscaled projections.

1 Introduction

General circulation models (GCMs) provide the best available estimates for the state of the Earth's climate under a range of future greenhouse gas emissions scenarios. The GCMs typically run on horizontal spatial scales ranging from 1.0 to 2.5 degrees (Taylor et al., 2012). These spatial scales are greater than the footprint of the human-scale activities that are crucial for understanding the impacts of climate change (Radić and Clarke, 2011; Taylor et al., 2012). Therefore, it is often necessary to generate downscaled climate variables to the spatial scales of interest (Maraun et al., 2010) .

Precipitation is of particular importance for projecting impacts of climate change on human activity, but its spatial and temporal characteristics remain a challenge for GCMs to accurately portray. Even where GCMs do capture the precipitation dynamics at the desired spatial scale, the values of precipitation represent an areal average on the model grid scale, thereby missing the localized extreme precipitation events (Gervais et al., 2014). These characteristics create the so-called "drizzle" effect in GCMs, and result in projections with too many days of moderate precipitation and too few with no or extreme precipitation (Stephens et al., 2010; Mehran et al., 2014; Koutroulis et al., 2016). Correcting these model biases is crucial for determining accurate estimates of the impacts of climate change, especially for accurately predicting changes to the frequency and severity of droughts and flooding (Camici et al., 2014; Quintero et al., 2018; Ahmadalipour et al., 2017).

Statistical downscaling is one option for addressing these spatial-scale issues, and for bridging the gap between the information provided by the GCMs, and the information needed by decision makers (Robinson and Finkelstein, 1991; Fowler et al., 2007). Statistical downscaling methods use empirical relationships between the outputs of GCMs and more localized data to create a

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

model for scaling variables from the large to small dimensions. This approach can then be applied to GCM outputs for future climate scenarios to create localized estimates of conditions under different radiative forcing scenarios.

Numerous methods have been used for statistical downscaling, ranging widely in complexity. Some of the earlier downscaling methods include bias-correction approaches to correct, scale, and produce outputs from GCMs to the local conditions (Karl et al., 1990; Murphy, 1999). Subsequent models, based on a range of numerical techniques, include the use of artificial neural networks (Hewitson and Crane, 1996; Ahmed et al., 2015), stochastic weather generators (Wilks, 1999; Kilsby et al., 2007), and constructed analogs (Abatzoglou and Brown, 2012; Pierce et al., 2014). These methods have improved the availability and reliability of downscaled climate projections.

While downscaling is a widely used approach for estimating climate changes at given locales (Maraun and Widmann, 2018), there are limited easy-to-use computer software options for producing and evaluating tailored downscaled climate projections. Some existing options include the Statistical Downscaling Model (SDSM, Wilby and Dawson, 2013), which uses a conditional weather generator model, and the DownscaleR software package (Bedia et al., 2020), which provides several bias correction, linear regression, and analog methods. The SDSM provides only a single downscaling approach, thereby limiting its adaptability to specific applications.

To our knowledge, there is no currently available software package for the Self-Organizing Map (SOM) downscaling method. Therefore, the software package described in this article, named CCdownscaling (Climate Change Downscaling), serves the purpose of filling this gap while also providing a framework for incorporating additional downscaling methods and associated evaluation metrics. Downscaling methods are useful for ascertaining the impacts of climate changes at given locales, but must be tailored for use in specific applications. These applications range widely and their execution places a range of different requirements on the downscaled product. There currently exist a limited number of software packages for downscaling, typically each involving a single downscaling approach. The CCdownscaling package described in this article is intended to readily and easily apply climate downscaling methods.

2 Description of the CCdownscaling package

With the CCdownscaling package, we provide an open-source framework, incorporating a number of downscaling approaches we have found valuable. The code has been designed to easily accommodate new methods as desired by future users. CCdownscaling package provides a framework for using many common machine learning tools as downscaling methods, to allow users to leverage existing and ongoing advances in machine learning when approaching downscaling problems. The package is written in Python because it is widely used in both the Atmospheric Science and Machine Learning communities, making it an ideal choice for wide distribution and application. In particular, the package allows users to leverage the powerful scikit-learn (Pedregosa et al., 2011) and TensorFlow (Abadi et al., 2015) machine learning libraries to use common machine learning approaches for downscaling.

The CCdownscaling package provides several point-based downscaling methods, as well as metrics for evaluating the skill of different methods on several variables important for different downscaling applications. The SOM downscaling method

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

(initially described in Hewitson and Crane (2006)) is not currently available in any publicly available software package. We have incorporated the SOM downscaling in CCdownscaling within a flexible framework. This means that CCdownscaling can be easily extended to new machine learning methods, to allow for future additions to the downscaling package, and integrate with commonly used existing machine learning frameworks. This will allow for easy integration of future machine learning techniques. Using the scikit-learn package (Pedregosa et al., 2011), we provide a number of additional machine learning algorithms for comparison as downscaling methods, including random forest and multiple linear regression models. In short, the goals of this paper are to:

– Develop an easy to use Python package implementing the SOM algorithm, alongside other downscaling methods.
– Develop a set of metrics for evaluating the reliability of downscaling methods.
– Demonstrate the use of the methods and metrics in the CCdownscaling package on an example downscaling use case.

3 Example Case

In this section we provide an example use case to demonstrate the use of the different downscaling methods and evaluation metrics provided in the CCdownscaling package. We downscale precipitation and daily maximum temperature from the Global Summary of Day (GSOD, National Climatic Data Center, 2020) dataset for O'Hare Airport in Chicago, Illinois. Reanalysis data is taken from the National Center for Environmental Prediction (NCEP, Kalnay et al., 1996) reanalysis 2 for relative humidity (at 850 hPa), air temperature (850 hPa), geopotential height (500 hPa), and zonal and meridonal wind components (700 hPa). This data is used to train the downscaling methods included in the CCdownscaling software package, and evaluate the results using the various metrics described below. Results shown below for this example, and the required data sets and Python code are provided in the GitHub repository (https://github.com/drewpolasky/CCdownscaling). The period from 1976 to 1999 is used to train the downscaling methods, which are then tested on the years from 2000 to 2005. In the following sections we discuss the methods and evaluation metrics in the context of this example case.

4 Downscaling Methods

The CCdownscaling package incorporates several downscaling methods, all conforming to a common framework for integration into downscaling workflows. The Application Programming Interface (API) mirrors the scikit-learn setup for machine learning models, which allows the easy integration of scikit-learn methods, and ease of use for those already familiar with scikit-learn. This approach allows for the development, testing, and comparison of different downscaling methods in a shared framework, as displayed in Fig. 1.

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

4.0.1 SOM Downscaling

Self-Organizing Maps (SOMs) are an unsupervised machine learning method for mapping a complex set of inputs onto a two-dimensional map of nodes, each representing a typical pattern observed in the input data (Kohonen, 1990). For downscaling, SOMs can be used to identify characteristic synoptic scale weather patterns, and relate those patterns to the observed local conditions (Hewitson and Crane, 2006). The SOM can then be used with GCM projections to explore changes to the frequency of these patterns in future climate scenarios (Gibson et al., 2016). SOMs have been broadly used for downscaling, particularly of precipitation, in regions such as South Africa (Hewitson and Crane, 2006), Florida (Sinha et al., 2018), and the Midwest United States (Polasky et al., 2021).

The SOM method begins by creating a set of nodes arranged in a two dimensional grid. Each of these nodes is defined by a vector, matching the length of the training data. To train the SOM, each element of the training data (in our case, coming from the reanalysis) is compared to the SOM nodes. The node whose vector is nearest (in Euclidean distance) to the training element is selected as the Best Match Unit (BMU),

$$BMU = min(W_v - I(t)) \tag{1}$$

Where W_v is the weight vector for node v, I is the input dataset, and t is the index of the training element. The BMU vector is then incrementally updated towards the training element, as are the neighboring nodes to the BMU. Each node weight vector is updated

$$W_v(s+1) = W_v(s) + \Theta(u,v)\alpha \times (I(t) - W_v(s)) \tag{2}$$

where s is the current iteration of the training, α is the learning rate of the SOM, and Θ is the neighborhood function, governing how much the update effects the nodes near to the BMU. The value of Θ decreases exponentially the distance of the node to BMU. Adjusting the neighboring nodes in addition to the BMU has the effect of sorting similar nodes to be near to one another in the map. The overall update rate is governed by α. With each successive pass through the dataset, α is decreased to more rapidly converge to a stable map.

Once the SOM has been trained, each day in the training data can be placed on the map by finding the BMU. For each node of the SOM, this gives a set of days corresponding to that pattern. The station observations for those days can be combined to create a probability function of local values for each SOM node. To create new downscaled projections, GCM data can be mapped onto the SOM, matching each day to its BMU. The probability function of the downscaling target variable can then be sampled from, producing the downscaled value for that day. By iterating through the days included in the GCM data sets, we produce a daily downscaled time series for the given location. The SOMs were implemented in Python using the TensorFlow library (Abadi et al., 2015), adapted from the open-source tensorflow-som project (Gorman, 2019).

The SOM method has the advantage of providing insight into the weather patterns giving rise to specific downscaled outcomes, through the patterns detected by the individual nodes of the SOM. In the O'Hare Airport example, the highest

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

precipitation nodes fall in the center of the top two rows of the SOM (Fig. 2). Two of these nodes for example, (0,4) and (1,2), both have high precipitation, but very different temperature patterns (Fig. 3). Node (1,2) appears to be a warm summer day, and the high precipitation likely corresponds to summer-type convection. The (0,4) node is colder, with a strong northwest/southeast temperature gradient, and the precipitation is likely driven by mid-latitude cyclones.

4.0.2 Scikit-Learn Downscaling Methods

Scikit-Learn is a widely used machine learning library for Python, that provides a wide range of machine learning tools (Pedregosa et al., 2011). These methods are easily adapted for use as downscaling tools, and this has been demonstrated for a random forest model for the O'Hare Airport example. Random forests are a widely used machine learning approach, that have been successfully applied to a wide range of problems (Breiman, 2001). Their adaptability and ease of use have led to random forest being a go-to method in machine learning (Biau and Scornet, 2016). Random forests have been used for downscaling temperature and precipitation in a variety of locations (Hutengs and Vohland, 2016; Sa'adi et al., 2017; Pang et al., 2017; Polasky et al., 2021).

The CCdownscaling package provides a framework for extending scikit-learn provided methods to better suit downscaling problems. An example of this functionality is given in two_step_random_forest model, which adapts the standard random forest from scikit-learn to have a classifier to initially split dry/precipitation days, then a second regressor model to predict the amount of precipitation for the wet days. This model addresses issues of the random forest producing too many days of middling precipitation, and too few dry days, similar to the undownscaled GCMs. The two step model outperforms the regular random forest for precipitation (Table 1).

4.0.3 Quantile Mapping

Quantile mapping is a commonly used approach for downscaling, especially for temperature (Maraun, 2013). Unlike the other methods included in this software package, quantile mapping is inherently single variable. It takes the approach of comparing the difference in value between the quantile ranks of an initial and final distribution. A transformation is calculated between these ranks, which can then be applied to a new input distribution, to create a downscaled output that better matches the observed distribution of events.

Quantile mapping is effective at bias correcting GCMs outputs, but suffers from several drawbacks. The first of these is variance inflation, where the variance of the final downscaled output is higher than that of the observed distribution (Maraun, 2013). Additionally quantile mapping relies on the fields of the target variable from the GCMs. This can be an issue, particularly in the case of precipitation, where the representation of the processes underlying the output values are not properly resolved. Other methods, which can make use of variables that are better captured in the GCMs, are likely to yield superior results of the downscaling in these circumstances. Nonetheless, quantile mapping provides a straightforward bias correction approach, and is a useful point of comparison for other downscaling methods.

4.0.4 Code Setup

All the downscaling models in CCdownscaling conform to a common API, making the addition of new methods to a downscaling use case straightforward. In the O'Hare Airport example, we train four different models, all called using the same format, where the model is created, then trained (fit), and final downscaling output is generated (predict):

```
som = som_downscale.som_downscale(som_x = 7, som_y = 5, batch = 512, alpha = 0.1, epochs = 50)
rf_two_part = correction_downscale_methods.two_step_random_forest()
random_forest = sklearn.ensemble.RandomForestRegressor()
qmap = correction_downscale_methods.quantile_mapping()

#train
som.fit(training_data, train_hist, seed = 1)
random_forest.fit(training_data, train_hist)
rf_two_part.fit(training_data, train_hist)
qmap.fit(rean_precip_train, train_hist)

#generate outputs from the test data
som_output = som.predict(test_data)
random_forest_output = random_forest.predict(test_data)
rf_two_part_output = rf_two_part.predict(test_data)
qmap_output = qmap.predict(rean_precip_test)
```

5 Validation Methods

In addition to providing a range of downscaling methods, this package also provides a number of different evaluation metrics, to compare methods and assess the suitability of a downscaling output for a given task. Depending on the goals of the downscaling and even the use case, different downscaling applications may best be assessed using different evaluation criteria. For example, estimating drought conditions will require accurate estimation of average precipitation and temperature over longer time frames, while projecting for flooding situations will require accurate representation of large precipitation events, and temporal correlation for multi-day events. In this section, we describe a number of evaluation metrics, and show the results of these metrics on the O'Hare Airport example for a range of downscaling methods.

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

5.0.1 SOM Training Metrics

The SOM method includes two specialized training metrics: quantization and topological error. These are commonly used metrics for assessing the training characteristics of a model setup, and should be used when tuning the hyperparameters of the model.

Quantization error (QE) refers to the average distance between the day vectors assigned to a node, and the characteristic vector of the node (Kohonen, 1990). Smaller values of QE represent reduced spread within each cluster. As the size of the map increases, this value will naturally decrease, as days are increasingly subdivided between clusters. A common tactic for settling on a map size is to look for the “elbow" where the decrease in QE slows down, with the increase in map size (Céréghino and Park, 2009). In Fig. 4, this occurs at the 5×7 map size, and that map size was selected for further analyses.

Topographical Error (TE) is calculated as the percentage of input vectors whose second-best matching units are adjacent to their best match units. This is a measure of how well the topology of the original dataset is being preserved in the lower-dimension space of the SOM (Uriarte and Martín, 2005). TE generally increases as the size of the map increases, with more opportunities for non-adjacent nodes to match an input vector (Fig. 5). These two metrics combine to help choose an optimal size for the SOM, balancing the gains in specificity against the increasing complexity of the model.

5.0.2 Distribution Tests

One of the key characteristics for a downscaling method is determining the skill of a downscaling approach is how well it reproduces the historical distribution of events. Precipitation, in particular, is commonly spread over too many days with small amounts of precipitation in GCMs projections, reducing the number of dry days and large precipitation events (Fig. 6). These metrics are especially valuable when trying to understand the shifts in climate, and the frequency of different event types under climate change scenarios (Perkins et al., 2013; Polasky et al., 2021).

The PDF skill score measures the similarity between two probability density functions (PDFs), by calculating the difference between observed and modeled counts within each bin of a histogram:

$$S_{score} = \sum_{1}^{n} min(Z_m, Z_o) \tag{3}$$

where S_{score} is the skill score, Z_m and Z_o are the frequency of modeled and observed values in a given bin, respectively, and n is the number of bins used to calculate the PDF (Perkins et al., 2007). PDF skill score provides an easily-interpretable score for the similarity between the observed and downscaled distribution of values.

The Kolomogorov-Smirnov (KS) test provides a non-parametric statistical test to evaluate the likelihood that two samples are drawn from the same underlying distribution (Massey Jr., 1951). The KS test provides both a test statistic, which can be used to asses the similarity of two distributions (where smaller values are more similar), and a probability that the given distributions are drawn from the same underlying probability distribution. Tables 1 and 2 show the KS scores and probabilities for the different methods for the O’Hare Airport example for precipitation and maximum air temperature, respectively.

© Author(s) 2022. CC BY 4.0 License.

In most cases, PDF score and KS test statistic are strongly correlated (Brands et al., 2012). An exception occurs when the distribution includes a large number of near zero values, as in the case of daily precipitation. In this instance, PDF score can struggle due to the kernel density smoothing applied in the calculation (Brands et al., 2012). The effect of this difference can be seen in the O'Hare Airport example. For temperature, the SOM and random forest methods score similarly to one another in both PDF score and KS statistic (Table 2). By contrast, for precipitation, the PDF statistic for the two methods is similar, while the SOM far outperforms the random forest on KS statistic. Visually, we can see that the SOM much better matches the observed distribution (Fig. 6), indicating that the KS test statistic better represents the relative skill of the two downscaling methods for precipitation.

5.0.3 Error Metrics

Error metrics are common methods for evaluating model skill, evaluating the day by day differences in downscaled and observed values for the testing period. These direct measure of difference are most useful when comparing methods that seek to recreate the specific conditions on each individual day, such as the random forest model. Much of the goal with downscaling, and with any climate modeling, is to understand the full range and frequency of events occurring, rather than to predict the correct event on the specific day. For this reason, the error metrics are often less useful for assessing downscaling model skill than the distribution tests. Nonetheless, these measures can provide useful comparisons, especially between similar classes of models.

The methods described above should only be applied when comparing downscaled outputs calculated from reanalysis data to the ground-truth observations. Data sets coming from GCMs are unlikely to match the daily variation of the historical records, and the results may not represent the documented conditions. Commonly used error metrics such as mean squared error and bias are included in the package. The results for the O'Hare Airport example can be found in Table 1 for daily precipitation, and Table 2 for maximum temperature. The results for SOM and linear regression models for temperature show the differences between the error and distribution metrics. The SOM outperforms the linear model on the PDF and KS scores, as it provides a better performance of matching the overall distribution of events, but scores worse in terms of RMSE and bias, because it does not match the specific day-to-day variations as well as the linear model.

5.1 Autocorrelation

Autocorrelation provides a metric for temporal similarity within time series data. It is calculated by taking the correlation between a timeseries and a time-lagged copy of the same timeseries, for a number of different time lags. This can be particularly important for applications such as flooding, where understanding the likelihood of multi-day large rainfall events is crucial for projecting the frequency of major flooding events in a given climate scenario. In the case of O'Hare Airport, the autocorrelation (of both the observed and downscaled datasets) falls off rapidly, with most of the downscaled methods, apart from the SOM and Qmap, overestimating the observed correlation at a lag of one day (Fig. 7).

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

6 Conclusions

The accelerating need for reliable, highly localized data for climate change scenarios has led to the development of software packages such as CCdownscaling that can readily and reliably provide information tailored to individual sites or regions. CCdownscaling is a new software package providing options for downscaling approaches and tools in a manner that the user can readily apply and iterate upon to meet the needs and requirements of the desired downscaled climate change projections. CCdownscaling provides SOM based downscaling, along with the extension of traditional machine learning tools to downscaling, as well as a variety of metrics for assessing the skill and utility of downscaled outputs. These tools, along with the example given, provide a framework for creating and evaluating new downscaled products, tailored to a location and use case.

7 Code and Data Availability

The current version of model is available from the project website: https://github.com/drewpolasky/CCdownscaling under the MIT licence. The exact version of the model used to produce the results used in this paper is archived on Zenodo (DOI: 10.5281/zenodo.6506660), as are input data and scripts to run the model and produce the plots for all the simulations presented in this paper: https://zenodo.org/record/6506677.

Acknowledgements. Thanks to Zach Moon for assistance in making the code more usable, and for helpful comments on the manuscript. Thanks as well to the Penn State Institute for Computational and Data Sciences for computational resources and support. This work was supported by the National Science Foundation under grant 1639342.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, https://www.tensorflow.org/, software available from tensorflow.org, 2015.

Abatzoglou, J. T. and Brown, T. J.: A comparison of statistical downscaling methods suited for wildfire applications, International journal of climatology, 32, 772–780, 2012.

Ahmadalipour, A., Moradkhani, H., and Svoboda, M.: Centennial drought outlook over the CONUS using NASA-NEX downscaled climate ensemble, International Journal of Climatology, 37, 2477–2491, 2017.

Ahmed, K., Shahid, S., Haroon, S. B., and 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, 1325–1341, 2015.

Bedia, J., Baño-Medina, J., Legasa, M. N., Iturbide, M., Manzanas, R., Herrera, S., Casanueva, A., San-Martín, D., Cofiño, A. S., and Gutiérrez, J. M.: Statistical downscaling with the downscaleR package (v3. 1.0): contribution to the VALUE intercomparison experiment, Geoscientific Model Development, 13, 1711–1735, 2020.

Biau, G. and Scornet, E.: A random forest guided tour, Test, 25, 197–227, 2016.

Brands, S., Gutiérrez, J. M., Herrera, S., and Cofiño, A.: On the use of reanalysis data for downscaling, Journal of Climate, 25, 2517–2526, 2012.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Camici, S., Brocca, L., Melone, F., and Moramarco, T.: Impact of climate change on flood frequency using different climate models and downscaling approaches, Journal of Hydrologic Engineering, 19, 04014 002, 2014.

Céréghino, R. and Park, Y.-S.: Review of the self-organizing map (SOM) approach in water resources: commentary, Environmental Modelling & Software, 24, 945–947, 2009.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C.: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, International Journal of Climatology: A Journal of the Royal Meteorological Society, 27, 1547–1578, 2007.

Gervais, M., Tremblay, L. B., Gyakum, J. R., and Atallah, E.: Representing extremes in a daily gridded precipitation analysis over the United States: Impacts of station density, resolution, and gridding methods, Journal of Climate, 27, 5201–5218, 2014.

Gibson, P. B., Perkins-Kirkpatrick, S. E., and Renwick, J. A.: Projected changes in synoptic weather patterns over New Zealand examined through self-organizing maps, International Journal of Climatology, 36, 3934–3948, 2016.

Gorman, C.: tensorflow-som, https://github.com/cgorman/tensorflow-som, 2019.

Hewitson, B. C. and Crane, R. G.: Climate downscaling: techniques and application, Climate Research, 7, 85–95, 1996.

Hewitson, B. C. and Crane, R. G.: Consensus between GCM climate change projections with empirical downscaling: precipitation downscaling over South Africa, International Journal of Climatology, 26, 1315–1337, https://doi.org/10.1002/joc.1314, 2006.

Hutengs, C. and Vohland, M.: Downscaling land surface temperatures at regional scales with random forest regression, Remote Sensing of Environment, 178, 127–141, https://doi.org/10.1016/j.rse.2016.03.006, 2016.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 77, 437–472, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Karl, T. R., Wang, W.-C., Schlesinger, M. E., Knight, R. W., and Portman, D.: A method of relating general circulation model simulated climate to the observed local climate. Part I: Seasonal statistics, Journal of Climate, 3, 1053–1079, 1990.

Kilsby, C. G., Jones, P., Burton, A., Ford, A., Fowler, H. J., Harpham, C., James, P., Smith, A., and Wilby, R.: A daily weather generator for use in climate change studies, Environmental Modelling & Software, 22, 1705–1719, https://doi.org/10.1016/j.envsoft.2007.02.005, 2007.

Kohonen, T.: The self-organizing map, Proceedings of the IEEE, 78, 1464–1480, https://doi.org/10.1109/5.58325, 1990.

Koutroulis, A. G., Grillakis, M., Tsanis, I., and Papadimitriou, L.: Evaluation of precipitation and temperature simulation performance of the CMIP3 and CMIP5 historical experiments, Climate Dynamics, 47, 1881–1898, https://doi.org/https://doi.org/10.1007/s00382-015-2938-x, 2016.

Maraun, D.: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue, Journal of Climate, 26, 2137–2143, 2013.

Maraun, D. and Widmann, M.: Statistical downscaling and bias correction for climate research, Cambridge University Press, 2018.

Maraun, D., Wetterhall, F., Ireson, A., Chandler, R., Kendon, E., Widmann, M., Brienen, S., Rust, H., Sauter, T., Themeßl, M., et al.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Reviews of geophysics, 48, 2010.

Massey Jr., F. J.: The Kolmogorov-Smirnov test for goodness of fit, Journal of the American statistical Association, 46, 68–78, 1951.

Mehran, A., AghaKouchak, A., and Phillips, T. J.: Evaluation of CMIP5 continental precipitation simulations relative to satellite-based gauge-adjusted observations, Journal of Geophysical Research: Atmospheres, 119, 1695–1707, 2014.

Murphy, J.: An evaluation of statistical and dynamical techniques for downscaling local climate, Journal of Climate, 12, 2256–2284, https://doi.org/10.1175/1520-0442(1999)012<2256:AEOSAD>2.0.CO;2, 1999.

National Climatic Data Center: Global Summary of Day, https://catalog.data.gov/dataset/global-surface-summary-of-the-day-gsod, 2020.

Pang, B., Yue, J., Zhao, G., and Xu, Z.: Statistical downscaling of temperature with the random forest model, Advances in Meteorology, 2017, https://doi.org/https://doi.org/10.1155/2017/7265178, 2017.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

Perkins, S., Pitman, A., and Sisson, S.: Systematic differences in future 20 year temperature extremes in AR4 model projections over Australia as a function of model skill, International journal of climatology, 33, 1153–1167, 2013.

Perkins, S. E., Pitman, A., Holbrook, N., and McAneney, J.: Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions, Journal of climate, 20, 4356–4376, https://doi.org/https://doi.org/10.1175/JCLI4253.1, 2007.

Pierce, D. W., Cayan, D. R., and Thrasher, B. L.: Statistical downscaling using localized constructed analogs (LOCA), Journal of Hydrometeorology, 15, 2558–2585, 2014.

Polasky, A. D., Evans, J. L., Fuentes, J. D., and Hamilton, H. L.: Statistical climate model downscaling for impact projections in the Midwest United States, International Journal of Climatology, 2021.

Quintero, F., Mantilla, R., Anderson, C., Claman, D., and Krajewski, W.: Assessment of changes in flood frequency due to the effects of climate change: Implications for engineering design, Hydrology, 5, 19, 2018.

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

Radić, V. and Clarke, G. K.: Evaluation of IPCC models' performance in simulating late-twentieth-century climatologies and weather patterns over North America, Journal of Climate, 24, 5257–5274, https://doi.org/https://doi.org/10.1175/JCLI-D-11-00011.1, 2011.

Robinson, P. and Finkelstein, P.: The development of impact-oriented climate scenarios, Bulletin of the American Meteorological society, 72, 481–490, 1991.

Sa'adi, Z., Shahid, S., Chung, E.-S., and bin Ismail, T.: Projection of spatial and temporal changes of rainfall in Sarawak of Borneo Island using statistical downscaling of CMIP5 models, Atmospheric Research, 197, 446–460, https://doi.org/10.1016/j.atmosres.2017.08.002, 2017.

Sinha, P., Mann, M. E., Fuentes, J. D., Mejia, A., Ning, L., Sun, W., He, T., and Obeysekera, J.: Downscaled rainfall projections in South Florida using self-organizing maps, Science of The Total Environment, 635, 1110 – 1123, https://doi.org/https://doi.org/10.1016/j.scitotenv.2018.04.144, 2018.

Stephens, G. L., L'Ecuyer, T., Forbes, R., Gettelmen, A., Golaz, J.-C., Bodas-Salcedo, A., Suzuki, K., Gabriel, P., and Haynes, J.: Dreary state of precipitation in global models, Journal of Geophysical Research: Atmospheres, 115, https://doi.org/https://doi.org/10.1029/2010JD014532, 2010.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bulletin of the American Meteorological Society, 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.

Uriarte, E. A. and Martín, F. D.: Topology preservation in SOM, International journal of applied mathematics and computer sciences, 1, 19–22, 2005.

Wilby, R. L. and Dawson, C. W.: The statistical downscaling model: insights from one decade of application, International Journal of Climatology, 33, 1707–1719, 2013.

Wilks, D. S.: Multisite downscaling of daily precipitation with a stochastic weather generator, Climate Research, 11, 125–136, 1999.

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

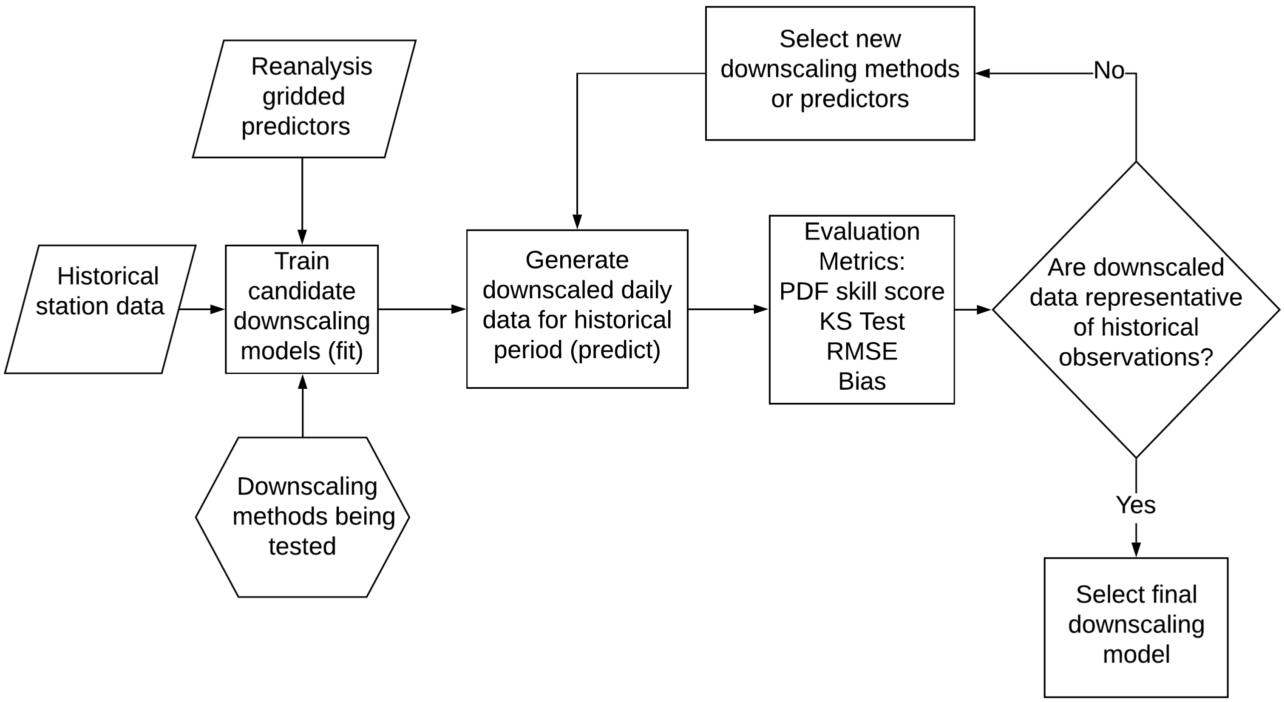

Figure 1. This flowchart provides the sequence of the downscaling process. Downscaling methods are trained and tested using the different evaluation metrics described in the text. In parenthesis are the names of the functions used for a given step in the downscaling process. The figure is adapted from Polasky et al. (2021).

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

Figure 2. Heat map of the SOM nodes for O'Hare Airport. The color represents the frequency of occurrence of that node, while the number in each box is the average precipitation on the days in that node cluster.

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

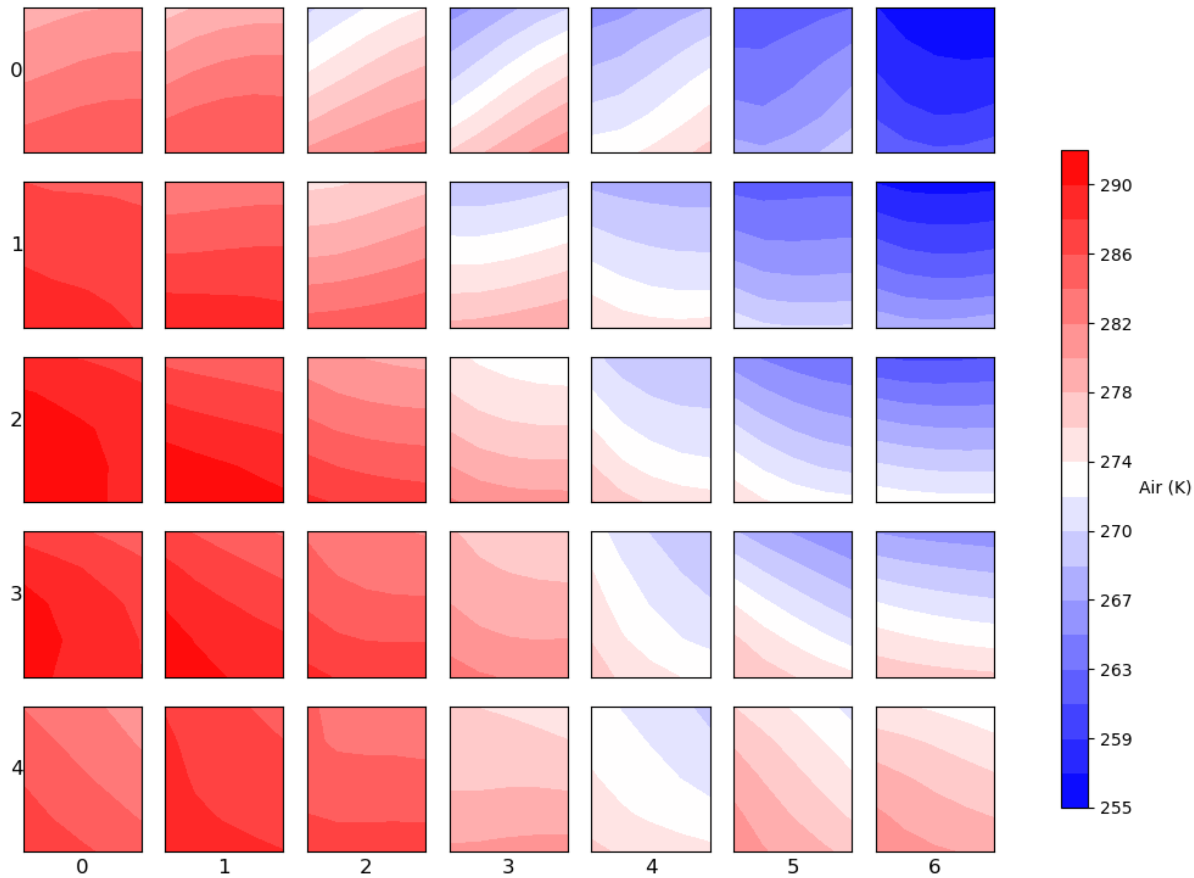

Figure 3. Average daily 850mb air temperature (K) for each node in the 5 × 7 SOM for the O'Hare Airport example. The strongest distinction is between primarily summer days on the left, and cold winter days on the right, with more intermediate patterns in between. There is also a separation on positively/negatively tilted temperature gradients moving from top to bottom.

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

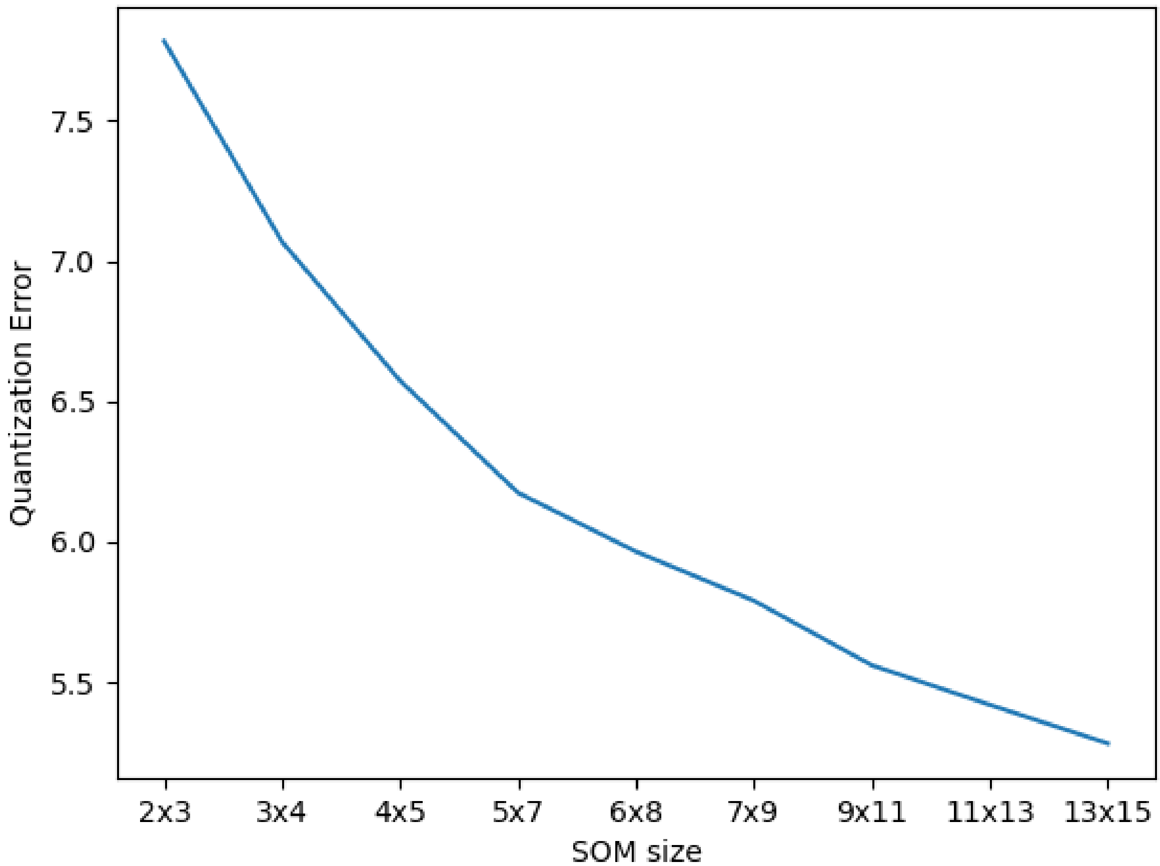

Figure 4. Quantization error for SOM grids of increasing size. Quantization error decreases as the number of nodes increases, as the specificity of each node increases. Choosing the correct size of SOM grid is a trade off between the increasing complexity of the graph, and thus potential for overfitting, and the decreasing quantization error with larger maps. For this example, a grid size of 5x7 was chosen, as representing the “elbow" where rate of decrease in the QE slows down.

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

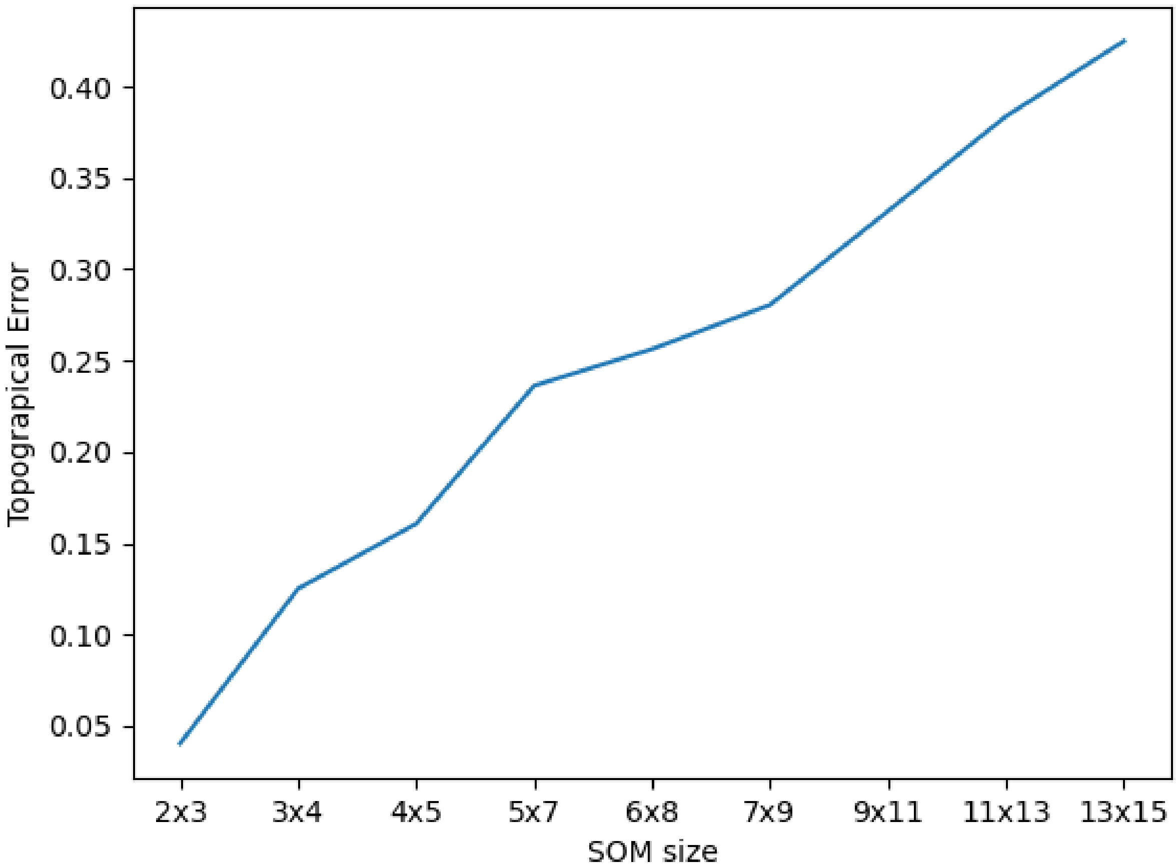

Figure 5. Topographic error for increasing SOM grids sizes for the same example as Fig. 4. TE increases monotonically with the size of the grid.

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

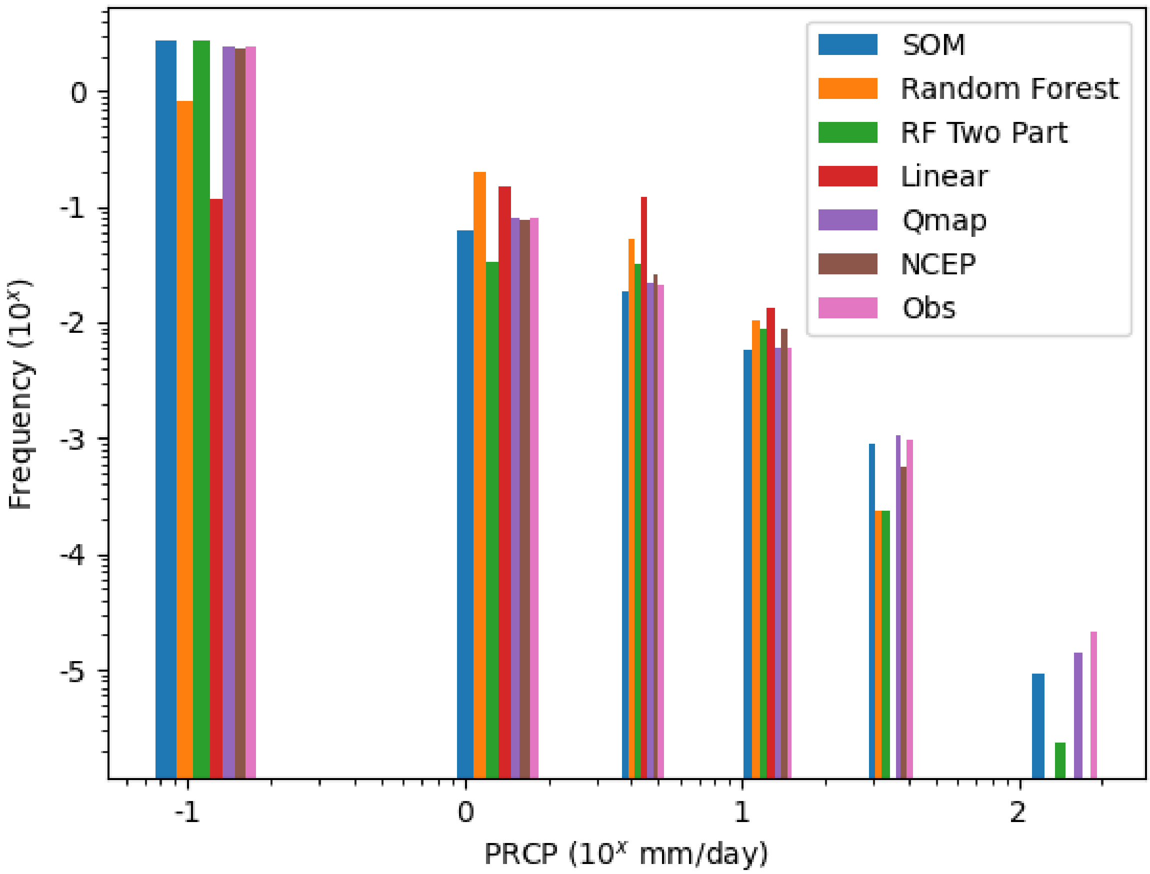

Figure 6. Histogram of precipitation amounts for O'Hare Airport for SOM and random forest downscaling, as well as the uncorrected NCEP precipitation values, and the station observations. The NCEP values underestimate particularly the large precipitation events. The two part random forest downscaling corrects this to some degree, but the SOM does a significantly better job.

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

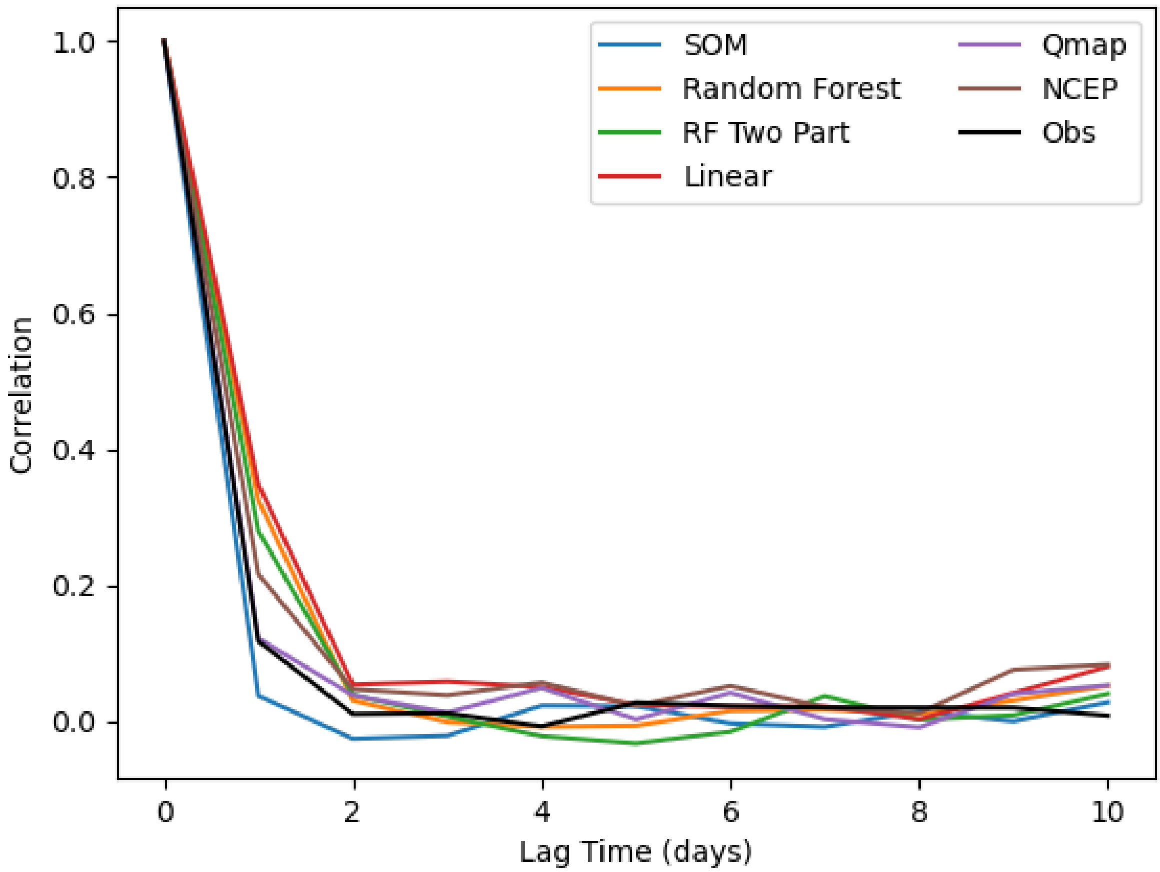

Figure 7. Auto correlation for the different downscaling methods for precipitation at O'Hare airport.

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

Table 1. Results scores for several different downscaling methods using the O'Hare airport example precipitation data. There are noticeable difference between the SOM and Random Forest (RF) scores, with the SOM showing a much higher RMSE than the RF but a higher PDF skill score, indicating that while the SOM does not match the individual days as well as the Random Forest model, it provides a better match of the distribution of events. The NCEP data come from the NCEP reanalysis precipitation for the grid point nearest to O'Hare Airport, without any downscaling applied.

	PDF Skill Score	KS Test Statistic	RMSE	Bias
SOM	0.993	0.03	11.03	0.04
RF	0.974	0.64	8.57	-0.48
RF Two Part	0.971	0.08	8.51	0.45
Qmap	0.993	0.17	9.97	-0.18
NCEP	0.986	0.65	8.88	0.12

https://doi.org/10.5194/egusphere-2022-282
Preprint. Discussion started: 6 May 2022
© Author(s) 2022. CC BY 4.0 License.

Table 2. As in Table 2, but for daily maximum temperature. The SOM and Quantile mapping (Qmap) methods have somewhat higher PDF skill scores, while the Random Forest does the best job of matching the day-to-day values, with the lowest RMSE.

	PDF Skill Score	KS Test Statistic	RMSE	Bias
SOM	0.959	0.02	8.0	0.256
RF	0.935	0.04	3.24	-0.36
Qmap	0.957	0.03	2.76	-0.25