Climate Variable Downscaling with Conditional Normalizing Flows

Christina Winkler, Paula Harder, David Rolnick
{christina.winkler, paula.harder, david.rolnick}@mila.quebec

Abstract

Predictions of global climate models typically operate on coarse spatial scales due to the large computational costs of climate simulations. This has led to a considerable interest in methods for statistical downscaling, a similar process to super-resolution in the computer vision context, to provide more local and regional climate information. In this work, we apply conditional normalizing flows to the task of climate variable downscaling. This approach allows for a probabilistic interpretation of the predictions, while also capturing the stochasticity inherent in the relationships among fine and coarse spatial scales. We showcase its successful performance on an ERA5 water content dataset for different upsampling factors. Additionally, we show that the method allows us to assess the predictive uncertainty in terms of standard deviation from the fitted conditional distribution mean.

1 Introduction

In climate modeling, simulations are typically run at coarse spatial resolution due to computational constraints. However, it is often of interest to obtain accurate predictions about the earth's climate not only on global but also on local scales, for example to guide local adaptation to precipitation or temperature trends. To fill this gap, statistical downscaling methods have increasingly been used to derive high-resolution information from low-resolution input. Early works have used Convolutional Neural Networks (CNNs) for climate variable downscaling [2, 10, 7, 11, 4, 9, 3]. However, these purely deterministic methods fail to capture the ill-determined nature of the problem – for the same low-resolution image, there exist many possible fine-scale realizations. Capturing such stochasticity is important in order to improve the accuracy of local scale predictions. Therefore, Generative Adversarial Networks (GANs) have become a widely used method in super-resolution and climate variable downscaling [**Chen_undated-ck**, 14, 15, 5, 12, 6, 15, 1]. However, such methods lack latent-space encoders and are known to suffer from mode collapse. It is hard to assess whether they are overfitting or generalizing. In climate variable downscaling, we require estimating a density as close to the true high-resolution pixel distribution as possible, as high-frequency details are of main importance. In this field, recently exact likelihood methods such as diffusion models have been applied [13] for climate variable downscaling by leveraging theory from optimal transport. First, the data is debiased and then a diffusion model is used for upsampling. In [5], the authors use normalizing flows for aligning the latent variables to a reference representation after performing statistical downscaling using a GAN architecture. Current state-of-the-art work [17] uses Fourier neural operators to learn a mapping between high and low-resolution climate data for arbitrary resolution downscaling.

In this work, we introduce the use of Conditional Normalizing Flows [16] (CNFs) for stochastic climate variable downscaling. They are particularly desirable, since we can tractably compute likelihood values, their sampling procedure is efficient, and we are able to assess predictive uncertainty due to its probabilistic interpretation. Unlike other generative models where predictive uncertainty is

often computed over an ensemble of different runs of weight initializations, or using techniques such as dropout, we are able to directly evaluate the predictive uncertainty of the CNF by computing the standard deviation from the fitted distribution mean.

Contributions   Our main contributions can be summarized as follows:

- We show for the first time how to apply conditional normalizing flows to the task of climate variable downscaling.

- We verify that CNF makes it possible to evaluate predictive uncertainty, by computing uncertainty maps from the standard deviation of sampled outputs.

2 Background

Normalizing flows represent a function as the composition of simpler invertible functions $f(\mathbf{z}) = f^K \circ f^{K-1} \circ \cdots \circ f^1(\mathbf{z})$ which yield the transformed random variables $\mathbf{z}^K \leftarrow ... \leftarrow \mathbf{z}^1 \leftarrow \mathbf{z}^0$ as intermediates after applying the transformations f^1 through f^k. The functions $f^k : \mathbb{R}^d \longmapsto \mathbb{R}^d$ are defined such that $f(\mathbf{z}^0) = \mathbf{y}$ with $\mathbf{y} \in \mathbb{R}^d$. All transformations f^k are invertible and differentiable, making it possible to computing the Jacobian determinant. Then, by applying the *Change of Variables Formula* we can model the density:

$$p_y(\mathbf{y}) \quad = p_\mathbf{z}(f(\mathbf{y})) \left| \det \frac{\partial f(\mathbf{y})}{\partial \mathbf{y}} \right| \tag{1}$$

where $\mathbf{y}$ is our input data at training time mapping to latent variable $\mathbf{z}$. This allows us to formulate a model for the marginal likelihood $p_\mathbf{y}(\mathbf{y})$ that can be computed tractably and optimized on the negative log-likelihood. We propose to learn conditional likelihoods using conditional normalizing flows [16] for the task of super-resolution on climate data. Take as input the low-resolution image $\mathbf{x} \in \mathcal{X}$ and as target the high-resolution image $\mathbf{y} \in \mathcal{Y}$. We learn a distribution $p_{Y|X}(\mathbf{y}|\mathbf{x})$ using a conditional prior $p_{Z|X}(\mathbf{z}|\mathbf{x})$ and a mapping $f_\phi : \mathcal{Y} \times \mathcal{X} \to \mathcal{Z}$, which is bijective in $\mathcal{Y}$ and $\mathcal{Z}$. The likelihood of this model is then defined as:

$$p_{Y|X}(\mathbf{y}|\mathbf{x}) = p_{Z|X}(\mathbf{z}|\mathbf{x}) \left| \frac{\partial \mathbf{z}}{\partial \mathbf{y}} \right| = p_{Z|X}(f_\phi(\mathbf{y}, \mathbf{x})|\mathbf{x}) \left| \frac{\partial f_\phi(\mathbf{y}, \mathbf{x})}{\partial \mathbf{y}} \right|. \tag{2}$$

Notice that the difference between Equations 1 and 2 is that all distributions are conditional, and the flow has a conditioning argument of $\mathbf{x}$. The generative process or in our case super-resolving an image from $\mathbf{x}$ to $\mathbf{y}$ is described by first sampling $\mathbf{z} \sim p_{Z|X}(\mathbf{z}|\mathbf{x})$ from a simple base density with its parameters conditioned on $\mathbf{x}$ (for us this is a diagonal Gaussian) and then passing it through a sequence of invertible mappings $f_\phi^{-1}(\mathbf{z}; \mathbf{x})$ to obtain a predicted super-resolved image $\hat{\mathbf{y}}$.

3 Experiments

ERA5 Hourly Water Content Dataset: This reanalysis dataset measures Total Column Water (TWC) provided in $\frac{kg}{m^2}$. It describes the vertical integral of the total amount of atmospheric water content, that is, cloud water, water vapor, and cloud ice, but not precipitation. We use the same water content dataset as described in [6] who perform physically consistent downsampling to create the low-resolution image counterparts. The dataset includes 40,000 training samples, with 10,000 for validation and 10,000 for testing. Similar as before, for preprocessing, we transform the input data values Z by $X = \frac{Z - \min Z}{\max Z - \min Z}$ such that they lie within range [0,1].

Experimental setup: For all experiments, we train the conditioned spatio-temporal flow with a learning rate of 2e-4 using a step-wise learning rate scheduler with a decay rate of 0.5 after every 200,000th parameter update step. We used the Adam optimizer [8] with exponential moving average and coefficients of running averages of gradients and its square are set to $\beta = (0.9, 0.99)$. We train the model with an architecture of 3 scales and 2 flow steps per scale for 35 epochs.

(a) *Ground truth*   (b) *Super-resolved samples*   (c) *Absolute Error*
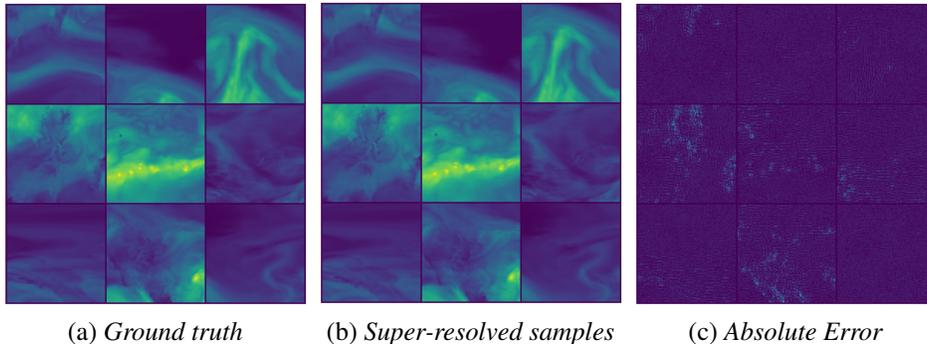
Figure 1: Super resolution results on the ERA5 water content TCW test data for $2 \times$ upsampling. Samples are taken from the CNF $x_{hr} \sim p(x_{hr}|x_{lr})$ with $\tau = 0.8$. *Best viewed electronically.*

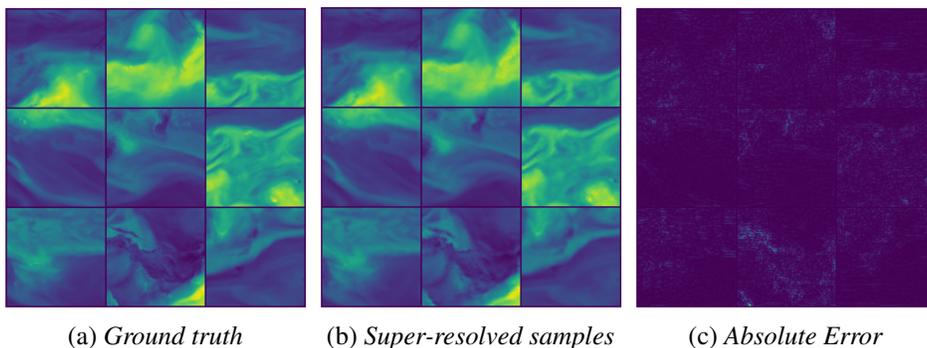(a) *Ground truth*   (b) *Super-resolved samples*   (c) *Absolute Error*

Figure 2: Super resolution results on the ERA5 TCW water content test data for $4 \times$ upsampling. Samples are taken from the CNF $x_{hr} \sim p(x_{hr}|x_{lr})$ with $\tau = 0.8$. *Best viewed electronically.*

3.1 Qualitative Evaluation

Figure 1 and 2 display super-resolution results predicted by the conditional normalizing flow on the hourly water content and daily temperature datasets for upsampling factors of 2 and 4 respectively. The method is able to generalize over images in the test set, where each test sample conveys very different water content distributions. However, in regions with high intensity values, there is greater absolute error in predicting the correct pixel values than for regions with low intensity values. This may arise simply because the same percentage error results in a larger absolute error in such regions.

3.2 Quantitative Evaluation

Table 1 shows the quantitative results of our method compared to a GAN architecture and bicubic interpolation. We added a perceptual Mean Squared Error loss between the predictions and ground truth image to improve sample quality. It can be seen that the generative approach outperforms the bicubic baseline. For the two times upsampling task, the super-resolution GAN outperforms the CNF on all metrics except the Continuous Ranked Probability Score (CRPS).

Table 1: CNF evaluated on MAE, RMSE and CRPS on the held out ERA5 water content test set. We compare our method to bicubic interpolation and a GAN.

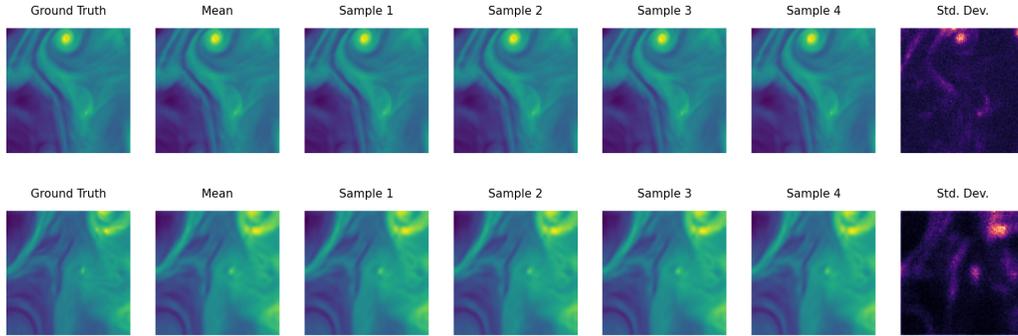| Model Type	TCW $2\times$ upsampling			TCW $4\times$ upsampling		
	MAE	RMSE	CRPS	MAE	RMSE	CRPS
Bicubic	6.96 $\pm$1.88	8.44 $\pm$2.09	-	6.90 $\pm$3.04	8.37 $\pm$3.23	-
CNF	**5.22** $\pm$1.86	**5.72** $\pm$1.98	**0.0150** $\pm$0.0092	**5.26** $\pm$1.86	**5.80** $\pm$1.99	**0.0174** $\pm$0.0118
GAN	5.27 $\pm$ 1.85	5.81 $\pm$ 1.97	0.0373 $\pm$ 0.0280	5.33 $\pm$1.84	5.90 $\pm$1.95	0.0454 $\pm$0.0393

Figure 3: The top row depicts the ground truth, conditional mean, different high-resolution realizations for one low-resolution image and computed standard deviation from the conditional mean for a $2\times$ upsampling factor. The bottom row displays the same experiment for an upsampling factor of $4\times$.

3.3 Sample Uncertainty

One of the main advantages of normalizing flows is the ability to generate multiple samples for one initial condition. In our case, this would mean generating multiple high-resolution realizations for the same low-resolution image. Figure 3 visualizes the standard deviation computed across twenty samples from the model for one low-resolution image. For convenience, we plotted only four predicted samples to compare with. It can be seen that in areas of high variance and finer texture regions, the standard deviation is generally higher. In applications such as flood risk estimation, this may be highly advantageous, since we deliberately want to have a model which is able to capture anomalies in the water content distribution.

4   Conclusion

In this work, we have shown the successful application of conditional normalizing flows to climate variables. The proposed method provides the advantage of density estimation and efficient sampling, and is able to model the stochasticity inherent in the relationships among fine and coarse spatial scales of climate variables. Additionally, we have shown that the method allows us to compute uncertainty maps in terms of standard deviation computed from the distribution mean.

References

[1] Chiranjib Chaudhuri and Colin Robertson. "CliGAN: A Structurally Sensitive Convolutional Neural Network Model for Statistical Downscaling of Precipitation from Multi-Model Ensembles". In: *Water* 12.12 (2020). ISSN: 2073-4441. DOI: 10.3390/w12123353. URL: https://www.mdpi.com/2073-4441/12/12/3353.

[2] Zhihao Chen et al. "Physics-informed generative neural network: an application to troposphere temperature prediction". In: *Environmental Research Letters* 16.6 (May 2021), p. 065003. DOI: 10.1088/1748-9326/abfde9. URL: https://dx.doi.org/10.1088/1748-9326/abfde9.

[3] Jianxin Cheng et al. "ResLap: Generating High-Resolution Climate Prediction Through Image Super-Resolution". In: *IEEE Access* 8 (2020), pp. 39623–39634. DOI: 10.1109/ACCESS.2020.2974785.

[4] Andrew Geiss and Joseph C. Hardin. "Radar Super Resolution Using a Deep Convolutional Neural Network". In: *Journal of Atmospheric and Oceanic Technology* 37.12 (2020), pp. 2197–2207. DOI: https://doi.org/10.1175/JTECH-D-20-0074.1. URL: https://journals.ametsoc.org/view/journals/atot/37/12/jtech-d-20-0074.1.xml.

[5] Brian Groenke, Luke Madaus, and Claire Monteleoni. "ClimAlign: Unsupervised statistical downscaling of climate variables via normalizing flows". In: (Aug. 2020). arXiv: 2008.04679 [cs.CV].

[6] Paula Harder. "Generating physically-consistent high-resolution climate data with hard-constrained neural networks". In: *AAAI 2022 Fall Symposium: The Role of AI in Responding to Climate Challenges*. 2022. URL: https://www.climatechange.ai/papers/aaaifss2022/5.

[7] Nidhin Harilal, Mayank Singh, and Udit Bhatia. "Augmented Convolutional LSTMs for Generation of High-Resolution Climate Change Projections". In: *IEEE Access* 9 (2021), pp. 25208–25218. DOI: 10.1109/ACCESS.2021.3057500.

[8]   Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015. 2014. URL: `http://arxiv.org/abs/1412.6980`.

[9]   Yumin Liu, Auroop R. Ganguly, and Jennifer Dy. "Climate Downscaling Using YNet: A Deep Convolutional Network with Skip Connections and Fusion". In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD '20. Virtual Event, CA, USA: Association for Computing Machinery, 2020, pp. 3145–3153. ISBN: 9781450379984. DOI: `10.1145/3394486.3403366`. URL: `https://doi.org/10.1145/3394486.3403366`.

[10]   U. Mital et al. "Downscaled hyper-resolution (400 m) gridded datasets of daily precipitation and temperature (2008–2019) for the East–Taylor subbasin (western United States)". In: *Earth System Science Data* 14.11 (2022), pp. 4949–4966. DOI: `10.5194/essd-14-4949-2022`. URL: `https://essd.copernicus.org/articles/14/4949/2022/`.

[11]   Yingkai Sha et al. "Deep-Learning-Based Gridded Downscaling of Surface Meteorological Variables in Complex Terrain. Part I: Daily Maximum and Minimum 2-m Temperature". In: *Journal of Applied Meteorology and Climatology* 59.12 (2020), pp. 2057–2073. DOI: `https://doi.org/10.1175/JAMC-D-20-0057.1`. URL: `https://journals.ametsoc.org/view/journals/apme/59/12/jamc-d-20-0057.1.xml`.

[12]   Ashutosh Kumar Singh, Adrian Albert, and Brian White. "Downscaling Numerical Weather Models with GANs". In: 2019. URL: `https://api.semanticscholar.org/CorpusID:226785468`.

[13]   Zhong Yi Wan et al. "Debias Coarsely, Sample Conditionally: Statistical Downscaling through Optimal Transport and Probabilistic Diffusion Models". In: (May 2023). arXiv: `2305.15618 [cs.LG]`.

[14]   Xintao Wang et al. "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks". In: *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. Sept. 2018.

[15]   Campbell D. Watson et al. *Investigating two super-resolution methods for downscaling precipitation: ESRGAN and CAR*. 2020. arXiv: `2012.01233 [physics.ao-ph]`.

[16]   Christina Winkler et al. *Learning Likelihoods with Conditional Normalizing Flows*. 2019. arXiv: `1912.00042 [cs.LG]`.

[17]   Qidong Yang et al. *Fourier Neural Operators for Arbitrary Resolution Climate Data Downscaling*. 2023. arXiv: `2305.14452 [cs.LG]`.