Machine learning emulation of a local-scale UK climate model

Henry Addison
Department of Computer Science
University of Bristol
Bristol, UK
`henry.addison@bristol.ac.uk`

Elizabeth Kendon
Met Office Hadley Centre
Exeter, UK

Suman Ravuri
DeepMind
London, UK

Laurence Aitchison*
Department of Computer Science
University of Bristol
Bristol, UK

Peter AG Watson*
School of Geographical Sciences
University of Bristol
Bristol, UK

* equal contribution

Abstract

Climate change is causing the intensification of rainfall extremes. Precipitation projections with high spatial resolution are important for society to prepare for these changes, e.g. to model flooding impacts. Physics-based simulations for creating such projections are very computationally expensive. This work demonstrates the effectiveness of diffusion models, a form of deep generative models, for generating much more cheaply realistic high resolution rainfall samples for the UK conditioned on data from a low resolution simulation. We show for the first time a machine learning model that is able to produce realistic samples of high-resolution rainfall based on a physical model that resolves atmospheric convection, a key process behind extreme rainfall. By adding self-learnt, location-specific information to low resolution relative vorticity, quantiles and time-mean of the samples match well their counterparts from the high-resolution simulation.

1 Introduction

Climate change is predicted to cause intensification of heavy rainfall extremes in the UK [1], [2]. Understanding rainfall at a local ($\sim$ 2km) scale is important for better adapting to these changes - for example, to predict where flooding might occur, [3]. Machine learning (ML) techniques can emulate high-resolution simulations using low-resolution inputs to generate local projections [4], [5]. This means expensive projections from a climate model could be complemented with further, cheaper samples with realistic spatial and temporal structure enabling better understanding of the uncertainty of high-resolution precipitation. This is particularly useful for extreme events which are poorly sampled by high-resolution simulations due to their infrequent nature.

Numerical simulations of physical processes of the climate can be used to create projections of rainfall but they are extremely expensive leading to trade-offs. At one end of the scale are global climate models (GCMs). These produce global-covering projections. Many variants exist with large ensemble sizes which allow experimenting with different scenarios and uncertainty in the underlying physics and model. The trade-off is these models restrict their resolution to grids of 25km or more, which is too coarse to provide actionable insight [6].

At the other end of the simulation scale are dynamical downscaling models like the Met Office used to produce 2.2km local projections in their UKCP18 dataset [7]. These used a convection-permitting model (CPM), a specialised type of regional climate model (RCM), which had a fine enough resolution to produce local-scale projections that are useful to decision makers. The trade-off is a small ensemble size (12, all following a single climate change scenario following a single climate change scenario, SSP5-8.5) with reduced spatial and temporal domain (UK and Ireland for 60 years) [2].

An alternative approach to produce high resolution projections is called statistical downscaling. This approach attempts to fit a statistical relationship between a lower-resolution set of climate variables and higher-resolution values. They are much less computationally expensive than the dynamical downscaling approach of RCMs.

Many approaches have been tried based on both traditional statistics and machine learning. No method has proven better than the others in general [4], [6], [8]. Gutiérrez *et al.* [6] assess the skill of 45 different methods for precipitation alone for a single European collaborative experiment, though the better-performing methods relevant to this project's problem are mainly based on generalized linear models (GLMs) or nearest neighbours (methods of analogs). These only covered point-statistics at individual observation stations and cannot produce the full spatial structure of a rain field.

Doury *et al.* [5] use U-Net [9] as the basis for their emulation of a 12km RCM based on GCM inputs over parts of France and Spain. Vandal *et al.* [4] found no clear best option when comparing traditional approaches with some off-the-shelf machine learning approaches (Multi-task Sparse Structure Learning and Autoencoder Neural Networks) for daily and extreme precipitation over the Northeastern United States. These systems are deterministic: once trained, the same conditioning GCM input will produce the high-resolution output. Without a probabilistic element, these models struggle to predict the small-scale detail of precipitation [10]. This work uses diffusion models [11]–[13] (aka score-based generative models), a recent development in deep generative models, to solve these problems with statistical downscaling.

Statistical downscaling of the full grid (rather than just a selection of observation stations) can be considered as similar to the super-resolution problem of natural images. The grid squares of the GCM and CPM outputs are like pixels with their intensity corresponding to climate variables such as daily rainfall (albeit not just integer values between 0 and 255). Common approaches for state-of-the-art sample generation, probability density estimation and super-resolution of natural images rely on deep generative models: diffusion models, Generative Adversarial Networks (GANs) [14], Variational Autoencoders (VAEs) [15], and autoregressive (AR) models like PixelRNN [16]. Diffusion models have shown competitive results in the natural image domain [17] and offer desirable trade-offs in sample sharpness (VAEs tend to produce blurry samples) and sample diversity (GANs can suffer from mode collapse [18]) and sampling cost (computationally cheaper than AR and much cheaper than a simulation). For more details on diffusions models see Appendix A.

Deep generative models have been successfully applied to problems in weather and climate such as short-term forecasting (nowcasting) of sequences of rainfall radar fields [10] and downscaling rainfall [19]. Commonly these use GANs but they can be difficult to train and their limitations imply they might underestimate the probability of extreme events. Diffusion models instead model the full data distribution and should not suffer these same problems. To our knowledge this work is the first application of diffusion models in the domain of climate downscaling.

2   Method

Dataset The Met Office's UKCP18 dataset contains both the low-resolution GCM global 60km projections [20] and high-resolution CPM local 2.2km projections [7]. Coarse relative vorticity at 850hPa is chosen as the conditioning input since it is a good predictor of high resolution rainfall [21]. For practical purposes the target data cover a 64x64 region of an 8.8km grid (2.2km grid coarsened 4x). This is small enough to fit on available GPU, fine enough to be useful and covers a large enough area for samples to be interpretable. For more details of the data see Appendix B.

Model The emulator is a score-based generative model based on NCSN++[13], adapted to allow conditional training and sampling. For sampling the Euler-Maruyama method is used to solve the reverse SDE. Location-specific parameters were added in order to tie pixels to the underlying

physical location of each pixel. This allows the model to learn relevant features for each location which effect rainfall that may not be available from the coarse relative vorticity inputs alone. Two models were trained: one based on just the simulation data as the conditioning inputs and one with an 8-channel feature map of location-specific parameters that are learnt from the data stacked on top of the vorticity conditioning tensor. Implementation is available on Github: `https://github.com/henryaddison/score_sde_pytorch`, which is a fork of `https://github.com/yang-song/score_sde_pytorch` [13].
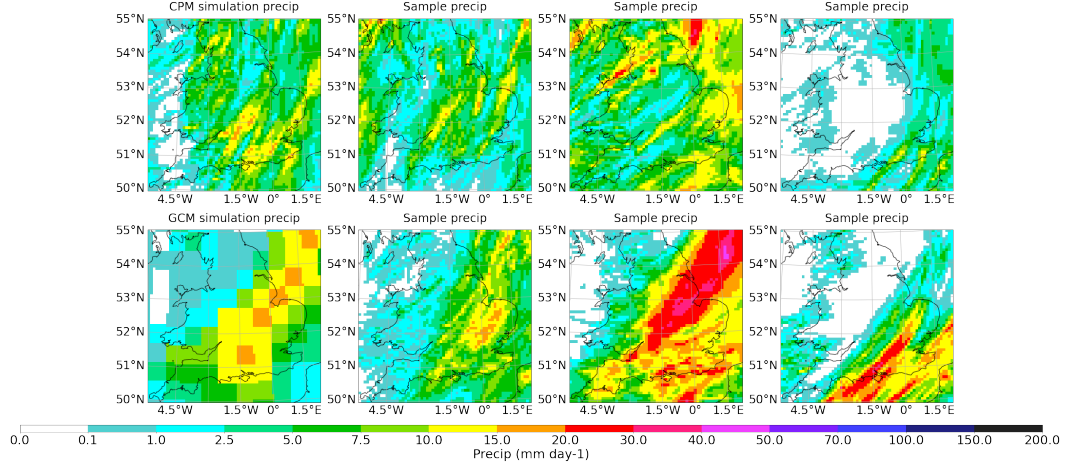
3   Results

Figure 1: Example samples conditioned on data from CPM (top) and GCM (bottom). First column is the rainfall from the simulation of the conditioning input. Columns 2 to 4 are the samples from the ML model. The highly variable nature of rainfall means samples should not match the simulation rainfall but represent the full range of plausible rainfall for the given low resolution conditions, of which the simulation output is a single example.

Figure 1 shows example high resolution samples from a model conditioned on relative vorticity in both modes of operation: sampling based on conditioning input of coarsened CPM data or GCM data. For each model, 3 samples were generated for each of the 4,320 timestamps in the validation set. Note that the aim is for the model to capture the full range of rainfall that is possible given the low resolution conditions described by the relative vorticity. Therefore samples should look realistic and cover a range that plausibly includes the rainfall from the simulation but they shouldn't all match it since the simulation output is just one possible occurrence of a highly variable phenomenon.
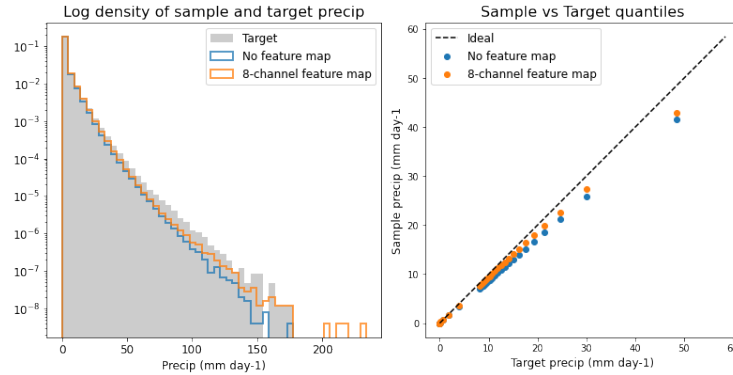

Figure 2: Per-pixel CPM-driven sample distribution: Log density (left) and Q-Q plot (right). Quantiles plotted are in three 9-member ranges of centiles: 10th-90th, 91st-99th, 99.1th-99.9th

3

Figure 2 shows the per-pixel Q-Q and log density plots of the sample and target distributions. These are based on samples conditioned on coarsened-CPM inputs for the two models variants. Both plots show the ML models match the target distribution well. The sample quantiles all match the target quantiles well and to be out by a constant scale factor of about 10%. Results using GCM inputs are similar in quality (Appendix C).
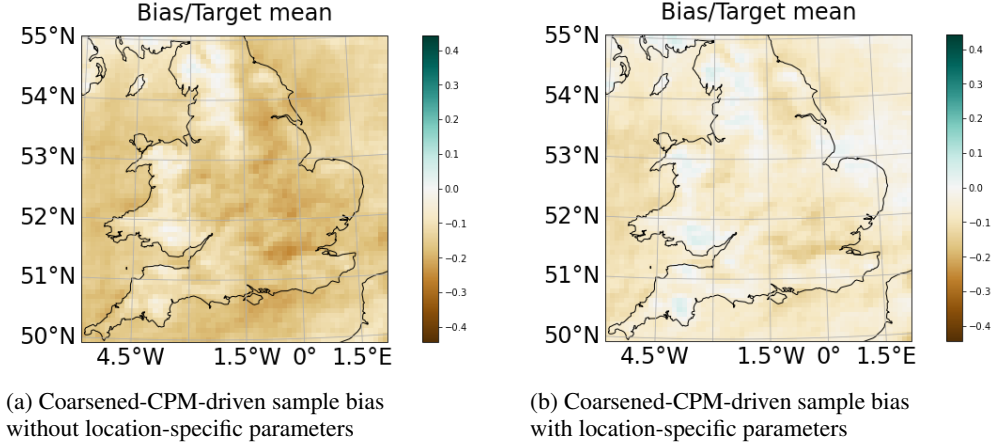

(a) Coarsened-CPM-driven sample bias without location-specific parameters

(b) Coarsened-CPM-driven sample bias with location-specific parameters

Figure 3: Mean-normalized bias using coarsened-CPM-based inputs with and without the self-learnt feature map

Figure 3 shows the normalized bias at each pixel. This is defined as sample mean (over time and all samples) minus target mean (over time) divided by the target mean (over time). The two versions represent without location-specific parameters (3a) and with self-learnt location-specific parameters (3b). The biases at each pixel are mostly negative so that on average the ML model is slightly under-predicting rainfall. Comparison between the two models show a reduction to the bias by adding the location-specific parameters. This matches the improvement in quantile results with the feature map (Figure 2) and again the bias of this version lies within about 10% of the target mean at each pixel.

4 Conclusion

High-resolution projections of precipitation are needed to more effectively adapt to future changes in precipitation. However, to make these solely with numerical climate simulations is very expensive. A diffusion model designed for natural images has been adapted to create the first emulator of CPM rainfall. As a generative model, it can cope with the stochastic nature of rainfall and also produce more samples without needing new inputs. The result is cheaper, high resolution rainfall samples with realistic spatial structure based on low resolution relative vorticity.

Further improvements are possible too. Adding conditioning inputs based on temperature and humidity should provide more information than relative vorticity at 850hPa alone. The extrapolation properties of the ML model have not been tested to see how it copes with unseen, very extreme conditions well beyond the 99.9th centile. The performance of the model when transferred to generate samples based on GCM inputs could be improved by transforming the GCM variables to better match the distribution of the CPM variables. Further generalization to ensembles of other climate models will also be valuable, leading to the potential to use the ML model to downscale climate models for which physics-based downscaling has not been implemented. There may also be value in generating sequences rather than single snapshots of rainfall and higher temporal frequencies than daily.

References

[1] M. G. Donat, A. L. Lowry, L. V. Alexander, P. A. O'Gorman and N. Maher, "More extreme precipitation in the world's dry and wet regions," *Nature Climate Change*, vol. 6, pp. 508–513, 2016. DOI: `10.1038/nclimate2941`.

[2] E. J. Kendon, G. Fosser, J. Murphy *et al.*, "Ukcp convection-permitting model projections: Science report," 2019. [Online]. Available: `https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP-Convection-permitting-model-projections-report.pdf`.

[3] N. Schaller, J. Sillmann, M. Muller *et al.*, "The role of spatial and temporal model resolution in a flood event storyline approach in western norway," *Weather and Climate Extremes*, vol. 29, p. 100 259, 2020, ISSN: 2212-0947. DOI: `ARTN10025910.1016/j.wace.2020.100259`.

[4] T. Vandal, E. Kodra and A. R. Ganguly, "Intercomparison of machine learning methods for statistical downscaling: The case of daily and extreme precipitation," *Theoretical and Applied Climatology*, vol. 137, no. 1-2, pp. 557–570, 2018. DOI: `10.1007/s00704-018-2613-3`.

[5] A. Doury, S. Somot, S. Gadat, A. Ribes and L. Corre, "Regional climate model emulator based on deep learning: Concept and first evaluation of a novel hybrid downscaling approach," *Climate Dynamics*, 2022, ISSN: 0930-7575 1432-0894. DOI: `10.1007/s00382-022-06343-9`. [Online]. Available: `https://doi.org/10.1007/s00382-022-06343-9`.

[6] J. M. Gutiérrez, D. Maraun, M. Widmann *et al.*, "An intercomparison of a large ensemble of statistical downscaling methods over europe: Results from the value perfect predictor cross-validation experiment," *International Journal of Climatology*, vol. 39, no. 9, pp. 3750–3785, 2019. DOI: `10.1002/joc.5462`.

[7] Met Office Hadley Centre, *UKCP18 Local Projections at 2.2km Resolution for 1980-2080*, Centre for Environmental Data Analysis, 2019. [Online]. Available: `https://catalogue.ceda.ac.uk/uuid/d5822183143c4011a2bb304ee7c0baf7` (visited on 09/06/2021).

[8] H. J. Fowler, S. Blenkinsop and C. Tebaldi, "Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling," *International Journal of Climatology*, vol. 27, no. 12, pp. 1547–1578, 2007. DOI: `10.1002/joc.1556`.

[9] O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention, Pt Iii*, vol. 9351, pp. 234–241, 2015, ISSN: 0302-9743. DOI: `10.1007/978-3-319-24574-4_28`.

[10] S. Ravuri, K. Lenc, M. Willson *et al.*, "Skillful precipitation nowcasting using deep generative models of radar," 2021. DOI: `arxiv:2104.00954`.

[11] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," in *Advances in Neural Information Processing Systems*, vol. 32. [Online]. Available: `https://proceedings.neurips.cc/paper/2019/hash/3001ef257407d5a371a96dcd947c7d93-Abstract.html`.

[12] J. Ho, A. Jain and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851. [Online]. Available: `https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html`.

[13] Y. Song, J. Sohl-Dickstein, Diederik, A. Kumar, S. Ermon and B. Poole, "Score-based generative modeling through stochastic differential equations," in *ICLR*, 2021. DOI: `arxiv:2011.13456`.

[14] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza *et al.*, "Generative adversarial nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, MIT Press, 2014, pp. 2672–2680.

[15] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2014.

[16] A. Van Den Oord, N. Kalchbrenner and K. Kavukcuoglu, "Pixel recurrent neural networks," in *33rd International Conference on Machine Learning, ICML 2016*, vol. 4, 2016, pp. 2611–2620.

[17] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," in *Advances in Neural Information Processing Systems*, vol. 34, 2021. [Online]. Available: `https://proceedings.neurips.cc/paper/2021/file/49ad23d1ec9fa4bd8d77d02681df5cfa-Paper.pdf`.

[18] A. Grover, M. Dhar and S. Ermon, "Flow-gan: Combining maximum likelihood and adversarial learning in generative models," *Thirty-Second Aaai Conference on Artificial Intelligence / Thirtieth Innovative Applications of Artificial Intelligence Conference / Eighth Aaai Symposium on Educational Advances in Artificial Intelligence*, vol. 32, no. 1, pp. 3069–3076, 2018.

[19] J. Leinonen, D. Nerini and A. Berne, "Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–13, 2020.

[20] Met Office Hadley Centre, *UKCP18 Global Projections at 60km Resolution for 1900-2100*, Centre for Environmental Data Analysis, 2018. [Online]. Available: https://catalogue.ceda.ac.uk/uuid/97bc0c622a24489aa105f5b8a8efa3f0 (visited on 09/06/2021).

[21] S. C. Chan, E. J. Kendon, N. Roberts, S. Blenkinsop and H. J. Fowler, "Large-scale predictors for extreme hourly precipitation events in convection-permitting climate simulations," *Journal of Climate*, vol. 31, no. 6, pp. 2115–2131, 2018, ISSN: 0894-8755. DOI: 10.1175/Jcli-D-17-0404.1.

A Diffusion Models

Probabilistic models assume that observed data, such as high-resolution rainfall over the UK, is drawn from an unknown distribution $p^*(\mathbf{x})$. A conditional model such as high-resolution rainfall conditioned on coarse GCM inputs $p^*(\mathbf{x}|\mathbf{g})$ can also be considered but for simplicity this section will stick with the context free version.

Song *et al.* [13] combine earlier approaches [11], [12] into a single framework called Score-Based Generative Models with Stochastic Differential Equations (SDEs). The idea is to imagine a diffusion process $\{\mathbf{x}(t)_{t=0}^T\}$ modelled by an SDE:

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)d\mathbf{w} \tag{1}$$

When run forward a sample, $\mathbf{x}(0)$, from the data distribution, p_0, is gradually perturbed over time into a sample from a final noise distribution, p_T, somewhat like how a structured gas sample will gradually diffuse randomly across a room. The final distribution is chosen as something tractable for sampling, usually a Gaussian.

More interesting for us is running the diffusion process backwards:

$$\mathrm{d}\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - g(t)^2 \nabla_\mathbf{x} \log p_t(x)]dt + g(t)d\bar{\mathbf{w}} \tag{2}$$

By solving this, samples from p_T (which is easy by design) can be converted into samples from the original data distribution. This requires two steps: calculating the score, $\nabla_\mathbf{x} \log p_t(x)$, and then applying numerical approaches to solve Equation 2.

The score is estimated as a neural net $s_\theta(\mathbf{x}, t)$ where θ are determined by minimizing:

$$\mathbb{E}_t\{\lambda(t)\mathbb{E}_{\mathbf{x}(0)}\mathbb{E}_{\mathbf{x}(0)|\mathbf{x}(t)} \left[||s_\theta(\mathbf{x}(t), t) - \nabla_{\mathbf{x}(t)} \log p_{0t}(\mathbf{x}(t)|\mathbf{x}(0))||_2^2\right\} \tag{3}$$

where λ is a positive weighting function that is chosen along with f and g.

Song *et al.* [13] summarize three approaches for solving the reverse SDE. General-purpose numerical methods can be used to find approximate solutions to the SDE. Predictor-Corrector sampling takes this a step further by using making use of estimated score at each timestep to apply a correction to the sample estimated at that timestep by the general purpose solver. Alternatively the problem can be reformulated as a deterministic process without affecting the trajectory probabilities and in turn solved using an ODE solver.

B Dataset

The Met Office's UKCP18 datasets for both the low-resolution GCM global 60km projections [20] and high-resolution CPM local 2.2km projections [7]. For the whole UK and Ireland domain, the high-resolution grid is 484 x 606 while the coarse grid is 17 x 23. For practical purposes this study limits the geographical domain to the area covered by a 64x64 area of an 8.8km grid (so 2.2km high-resolution coarsened 4x) centred on Birmingham. This covers a large enough area (England and Wales) to make sample interpretation easier while also boosting the resolution of the 60km projections. The temporal frequency is daily. Data are available hourly for three 20-year periods (1981-2000, 2021-2040, 2061-2080). The default CPM ensemble member and its corresponding GCM were chosen. The dataset is split into a training set (70%, n=15,120) and a validation set (20%, n=4,320) with a 10% test set reserved once fine-tuning is complete.

It is not possible to train a model using GCM data as conditioning input because the weather conditions in GCM and CPM are different for the same timestamp [5]. As far are the CPM is concerned, the GCM stops at the edge of Europe so the two simulations diverge at times. Therefore, the dataset used in the training phase will the conditioning input data is from the CPM regridded and coarsened to match the output of the GCM and the target is 4x coarsened precipitation from the CPM. It is also better preferable to avoid using low resolution rainfall as a conditioning input but rather variables that are as well-represented in GCM as the CPM, such as wind, temperature and humidity. Based on work by Chan *et al.* [21], the first conditioning input variable was chosen to be relative vorticity (a

derivative of wind field which measures circulation in the atmosphere and correlated with how stormy or fair conditions are) at altitude corresponding to 850hPa. Once fitted, samples can then be drawn from the model either using conditioning inputs based on coarsened CPM data or data from a GCM.
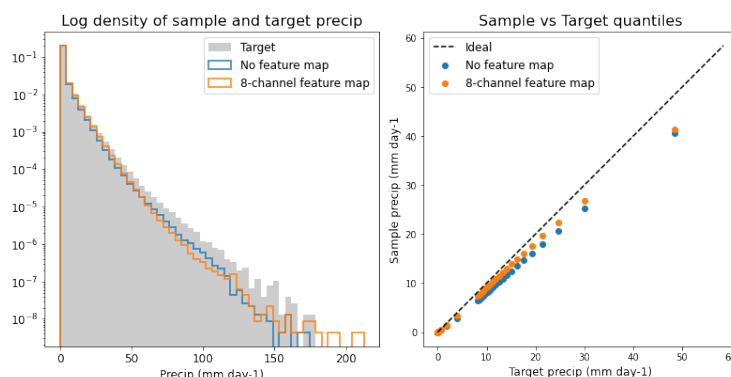
C   GCM-driven results

Figure 4: Per-pixel GCM-driven sample distribution: Log density (left) and Q-Q plot (right)

(a) GCM-driven sample bias without self-learnt feature map

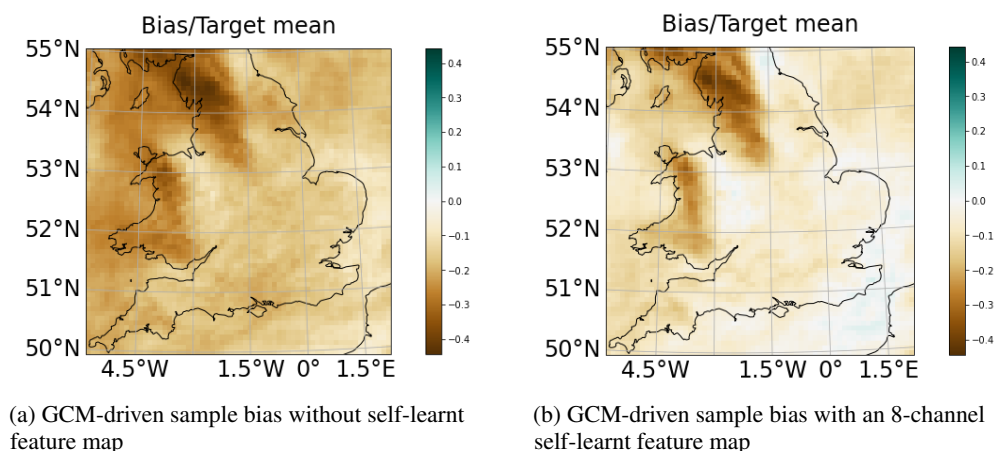(b) GCM-driven sample bias with an 8-channel self-learnt feature map

Figure 5: Mean-normalized bias using GCM inputs with and without the self-learnt feature map

Comparing the bias between sampling modes (CPM-based samples in 3b with GCM-based samples in 5b) shows a difference between the sample distributions depending on the source of the conditioning input. The CPM-based bias is smaller overall but there is also an inversion: CPM-based bias is smallest (and slightly positive) over upland areas like Snowdonia and the Lake District but GCM-based bias is much larger (and negative) other these same upland region.