

---

# Multivariate climate downscaling with latent neural processes

---

Anna Vaughan<sup>1</sup> Nicholas D. Lane<sup>1</sup> Michael Herzog<sup>2</sup>

## Abstract

Statistical downscaling is a vital tool in generating high resolution projections for climate impact studies. This study applies convolutional latent neural processes to multivariate downscaling of maximum temperature and precipitation. In contrast to existing downscaling methods, this model is shown to produce spatially coherent predictions at arbitrary locations specified at test time, regardless of whether training data are available at these points.

## 1. Introduction

Generating high resolution climate projections is vital for assessing risks under different climate scenarios. Unfortunately, the computational requirements of modern earth system models limit the resolution of simulations. For this reason, raw model output is routinely post-processed using statistical methods to downscale results to higher spatial resolution (Maraun & Widmann, 2018).

Given shortcomings of using traditional statistical techniques for downscaling on many metrics (Widmann et al., 2019; Gutiérrez et al., 2019; Hertig et al., 2019; Maraun et al., 2019), there has been considerable work on applying modern deep learning architectures to this problem. While bias correction (Cannon, 2011; Bürger et al., 2012) and unsupervised methods using normalising flows (Groenke et al., 2020) have been proposed, the majority of studies apply perfect prognosis models. In these methods the aim is to learn a transfer function mapping from low resolution predictors to high resolution outputs. This function is trained on observational data then applied to the coarse resolution model output (Maraun & Widmann, 2018). Architectures previously applied to this task include convolutional neural networks (Vandal et al., 2017; Wang et al., 2021; Liu et al., 2020; Baño-Medina et al., 2020), autoencoders (Vandal et al., 2019), generative adversarial networks (Chaudhuri & Robertson, 2020), multilayer perceptrons (MLPs; Cannon, 2008) and long short-term memory networks (Misra et al., 2018).

In these studies the transfer function maps from a set of coarse resolution predictors to a fixed set of locations, typi-

cally on a grid, where training data are available. The model is then fitted by minimising a loss function between observations and model predictions at each point. This raises the question of how to make predictions at locations where training data are not available. Wang et al. (2021) used transfer learning to apply a convolutional neural network trained to map low resolution gridded predictors to high resolution gridded observations in one region to make skillful predictions in another. Though this allows predictions to be made at other locations, the resolution of the output grid remains fixed by the resolution of the training data. An alternative approach is to learn a transfer function from the low resolution predictors to a stochastic process (i.e a distribution over spatial functions of the variable being downsampled). This has advantages that the predicted stochastic process can be queried at any spatial location at test time, and uncertainty in the predictions is quantified. Vaughan et al. (2021) applied convolutional conditional neural processes (convCNP; Gordon et al., 2019) to downscaling temperature and precipitation at 86 locations in Europe. Unfortunately, the convCNP model is of little practical use for climate impact studies as it cannot produce spatially coherent samples (Dubois et al., 2020). A further limitation is that the convCNP model is only suitable for univariate downscaling. In many applications, for example agriculture and wildfire prediction, it is necessary to accurately model the joint distribution of multiple variables.

In this study we build on the work of Vaughan et al. (2021) to develop a new model for multivariate downscaling using a convolutional latent neural process (convLNP; Foong et al., 2020) to predict a multivariate stochastic process describing spatial fields of temperature and precipitation given low resolution predictor fields. This model jointly downscale daily maximum temperature and precipitation, quantifies uncertainty in predictions, generates coherent spatial samples and can generate predictions at arbitrary locations at test time.

For this work the convLNP model is trained to downscale gridded low-resolution ERA-5 reanalysis data to station observations over Germany, with performance evaluated at held out locations. To our knowledge this is the first attempt at multivariate downscaling by learning a transfer function to a stochastic process. Our findings indicate that this model is suitable for generating multivariate projections of temperature and precipitation suitable for use in climate

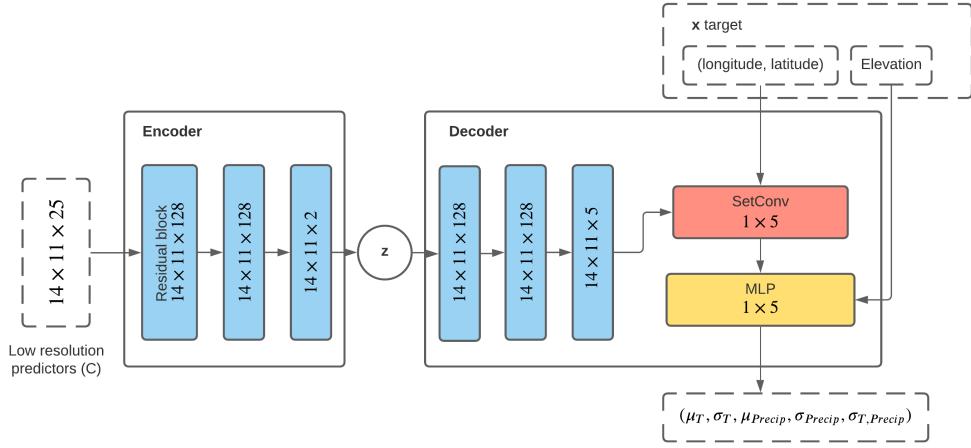


Figure 1. convLNP model architecture.

impact studies.

## 2. Methodology

**Idea:** Use a latent neural process [Garnelo et al. \(2018\)](#) to learn a transfer function mapping a set of daily gridded low resolution climate predictors to a multivariate stochastic process indexed by longitude and latitude. Unlike the convCNP model used by [Vaughan et al. \(2021\)](#), the latent neural process is able to model correlations between locations [\(Dubois et al., 2020\)](#). As translation equivariance is an important inductive bias in multisite downscaling [\(Baño-Medina & Gutiérrez, 2019\)](#), we use a convolutional neural process model [\(Foong et al., 2020\)](#), which incorporates translation equivariance into the latent neural process.

**Experiment:** ERA-Interim reanalysis data [\(Dee et al., 2011\)](#) at 0.75 degree resolution is downscaled to weather station data from the European Climate and Assessment Dataset [\(Klok & Klein Tank, 2009\)](#) in a region from 6 to 16 degrees longitude and 47 to 55 degrees latitude (Figure 2a). This region is chosen as Germany has a high density of station observations and complex topography in the Alps.

Predictors are the 0.75 degree resolution longitude-latitude grids of 25 variables from the reanalysis dataset. Atmospheric predictors are surface level mean and maximum temperature, wind and precipitation. Tropospheric predictors are humidity temperature and winds at 850hPa, 700hPa and 500hPa. Invariant predictors are longitude, latitude, sub-gridscale orography angle, anisotropy and standard deviation and geopotential. Day of year is also included represented by periodic transforms to capture seasonal variation. The training set consists of reanalysis predictors from 1979-

2002 together with observations from 397 stations (Figure 2a). The held out validation set consists of reanalysis predictors from 2003-2008 and observations from 19 stations held out from the training set (Figure 2b), assessing the ability of the model to generalise to unseen times and locations. These 19 validation stations are selected from locations used in the VALUE downscaling intercomparison experiment as they provide high quality observations in regions where statistical downscaling models are typically less accurate [\(Maraun et al., 2015\)](#).

**ConvLNP architecture** Using a convLNP the predictive distribution of maximum temperature and precipitation  $\mathbf{y}^{(t)}$  at  $T$  locations  $\mathbf{x}^{(t)}$  specified by longitude, latitude, and elevation coordinates is modelled as

$$p_{\theta}(\{\mathbf{y}^{(t)}\}_{t=1}^T | \{\mathbf{x}^{(t)}\}_{t=1}^T, C) = \int p_{\theta}(\mathbf{z}|C) \prod_t \mathcal{N}(\mathbf{y}^{(t)}; \Sigma^{(t)}(\mathbf{x}^{(t)}, \mathbf{z}; C), \mu^{(t)}(\mathbf{x}^{(t)}, \mathbf{z}; C)) dz$$

Where  $C$  is the set of low resolution predictors on longitude-latitude grids,  $\mathbf{z}$  is a latent variable and  $\mathcal{N}(\mathbf{y}; \Sigma, \mu)$  is a multivariate Gaussian distribution with mean  $\mu$  and covariance  $\Sigma$ .

The convLNP is implemented in two stages: an encoder parameterizing  $p_{\theta}(\mathbf{z}|C)$  mapping the low resolution predictors to a distribution over latent variable  $\mathbf{z}$  followed by a decoder parameterizing the predictive distribution  $p_{\theta}(\mathbf{y}|\mathbf{x}, \mathbf{z})$  over  $\mathbf{y}$  at location  $\mathbf{x}$  given  $\mathbf{z}$ .

A schematic of the model architecture is shown in Figure 1. In the encoder, the low resolution predictor grids are passed

through a convolutional network consisting of three residual blocks. This outputs predictions of  $(\mu, \sigma)$  parameterizing an independent Gaussian latent variable at each gridpoint.

A sample from these Gaussian distributions is used as input to the decoder, and passed through a further three residual blocks to output predictions of the five parameters specifying a joint Gaussian distribution of temperature and precipitation: the mean for temperature and precipitation ( $\mu_T$ ,  $\mu_{Precip}$ ), the variance for temperature and precipitation ( $\sigma_T$ ,  $\sigma_{Precip}$ ) and the covariance ( $\sigma_{T,Precip}$ ). These gridded predictions are transformed to the required location  $\mathbf{x}$  using a set convolution layer (Gordon et al., 2019). Finally, the five predicted parameters together with the elevation at  $\mathbf{x}$  are passed through a MLP, adjusting for the effect of subgrid-scale elevation. The model outputs predictions of the parameters specifying a multivariate Gaussian over temperature and precipitation at location  $\mathbf{x}$ . For details of the model architecture and training, see Appendix A.

**Baseline:** As to our knowledge this is the first attempt at multivariate downscaling by learning a transfer function to a stochastic process, there is no directly comparable baseline from previous work. We instead construct a simple transfer learning baseline aiming to answer the following two questions:

1. Is the skill of the convLNP model at unseen locations any higher than training a single-site model and using transfer learning to make predictions at other locations?
2. Does the multivariate convLNP predict intervariable correlations more accurately than applying separate univariate models?

The baseline model consists of two independent multi-layer perceptrons for maximum temperature and precipitation, each of which take spatial predictors from the low resolution grid together with the elevation at  $\mathbf{x}$  as input, and output predictions of maximum temperature or precipitation at that location. In a comparison of downscaling models for stations over Europe (Gutiérrez et al., 2019), many of the best performing models use joined principal components (PCs) of the predictor fields as input to capture spatial coherence. We therefore take the leading 19 joined principal components (explaining  $> 95\%$  of the variance) of the predictor fields as input to the MLP model. For details of the baseline architecture and training see Appendix B.

### 3. Results

Models are evaluated on multiple metrics. Marginal aspects are compared using mean absolute error (MAE) and pearson (spearman) correlation for maximum temperature (precipitation) between model predictions and observations.

Table 1. Comparison of results for the convLNP and baseline models on the validation set.

METRIC	CONVLNP	BASELINE
MAXIMUM TEMPERATURE		
MAE (C)	<b>1.95</b>	2.33
PEARSON	<b>0.96</b>	<b>0.96</b>
CMD	<b>1.29E-4</b>	6.43E-4
DOF BIAS	<b>0.02</b>	0.16
PRECIPITATION		
MAE (C)	<b>2.51</b>	2.64
SPEARMAN	<b>0.64</b>	<b>0.64</b>
CMD	<b>0.02</b>	0.06
DOF BIAS	<b>1.79</b>	2.24
MULTIVARIATE		
CORRELATION BIAS	<b>0.08</b>	0.15

These metrics are averaged over the 19 validation stations. Following (Widmann et al., 2019), two metrics are calculated specifically to evaluate how well spatial correlations are reproduced. The correlation matrix distance (CMD; Herdin, 2005) measures the similarity of two correlation matrices, with a value of zero indicating that the correlation structure is identical up to a scaling factor and a value of one indicating that the correlation structures are very different. The second spatial metric, bias in degrees of freedom (DOF) compares the predicted to observed spatial degrees of freedom (Widmann et al., 2019). Finally, predictions of inter-variable correlations between maximum temperature and precipitation are assessed by the absolute bias in spearman correlation between the variables (Maraun & Widmann, 2018) in model predictions compared to observations.

Results for each model on the validation set are shown in Table 1. The convLNP model outperforms or equals the performance of the baseline transfer learning model on all metrics. For maximum temperature, the MAE is 1.95C for the convLNP compared to 2.33C for the baseline. Pearson correlations are 0.96 for both models. Similarly for precipitation, the MAE is lower for the convLNP model at 2.51mm compared to 2.64mm for the baseline, with equal spearman correlations of 0.64.

Of particular interest in this study is how well the models capture the spatial distribution of the downscaled variables. For both maximum temperature and precipitation, the convLNP outperforms the baseline on the CMD and DOF bias metrics. Further insight into the ability of the model to reproduce spatial fields is gained by querying the predicted stochastic process at 0.05 degree resolution over the domain. Figure 2 shows an example of convLNP predictions for maximum temperature (d,e,f) and precipitation (g,h,i) compared to the low resolution reanalysis predictors of these

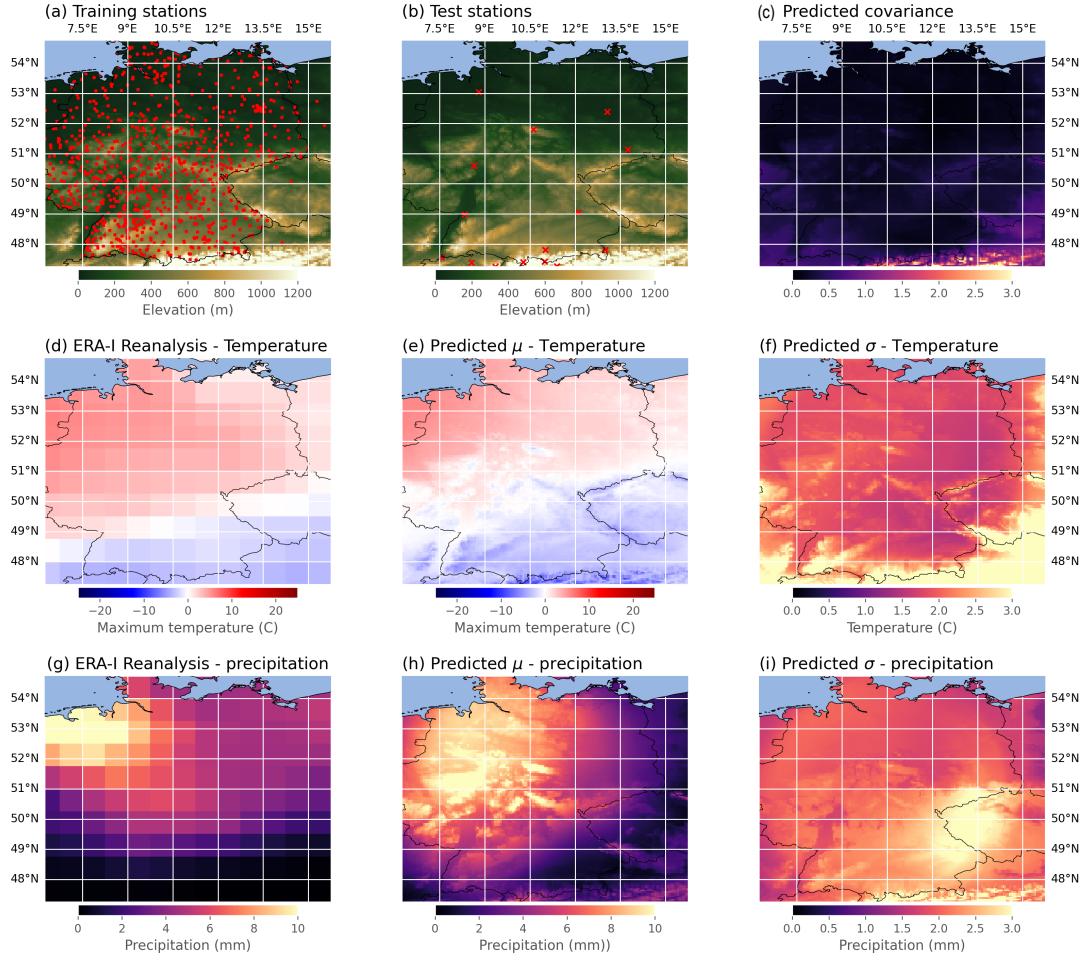


Figure 2. convLNP model predictions for 12/01/2003, showing (a) training stations and elevation (b) validation stations and elevation, (c) predicted maximum temperature-precipitation covariance, and low resolution reanalysis, predicted mean and predicted variance for maximum temperature (d,e,f) and precipitation (g,h,i).

fields. For both variables, large scale features are consistent with the low resolution input. For maximum temperature, predicted variance (model uncertainty) is highest in regions with complex topography and areas where training data are sparse, as expected. For precipitation, higher uncertainty is seen towards the edges of the area of predicted precipitation, especially in elevated regions. Locations at greater elevation are predicted to be colder and receive higher precipitation.

The final metric to consider is how well the model predicts intervariable correlations. The bias in intervariable correlations is substantially lower for the convLNP model compared to the baseline. Predicted covariance is largest in regions with complex topography (Figure 2c).

## 4. Conclusion

We have presented a new approach for multivariate climate downscaling using a convLNP. This model outperforms a transfer learning baseline, and reproduces spatially realistic high resolution fields of maximum temperature and precipitation with accurate intervariable correlations.

The convLNP model has a number of advantages over existing machine learning downscaling methods:

**Application to new domains** the convLNP can be queried at new locations at test time and generate spatially coherent predictions.

**Uncertainty quantification** the model robustly quantifies uncertainty in predictions.

**Multivariate downscaling** the model predicts a joint distribution of multiple variables, making it suitable for application to impact studies relying on accurate inter-variable correlations.

Multiple areas remain for future work. Further verification is required with an ensemble of baselines, particularly quantifying how well extreme values are represented and exploring whether predictions are physically consistent. The convLNP model will then be applied to climate impact studies where multivariate projections are required, with an initial focus on local wildfire risk projections.

## References

- Baño-Medina, J. and Gutiérrez, J. M. The importance of inductive bias in convolutional models for statistical downscaling. In *Proceedings of the 9th International Workshop on Climate Informatics: CI*, 2019.
- Baño-Medina, J., Manzanas, R., and Gutiérrez, J. M. Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geoscientific Model Development*, 13(4):2109–2124, 2020.
- Bürger, G., Murdock, T., Werner, A., Sobie, S., and Cannon, A. Downscaling extremes—an intercomparison of multiple statistical methods for present climate. *Journal of Climate*, 25(12):4366–4388, 2012.
- Cannon, A. J. Probabilistic multisite precipitation downscaling by an expanded bernoulli–gamma density network. *Journal of Hydrometeorology*, 9(6):1284–1300, 2008.
- Cannon, A. J. Quantile regression neural networks: Implementation in r and application to precipitation downscaling. *Computers & geosciences*, 37(9):1277–1284, 2011.
- Chaudhuri, C. and Robertson, C. Cligan: A structurally sensitive convolutional neural network model for statistical downscaling of precipitation from multi-model ensembles. *Water*, 12(12):3353, 2020.
- Chollet, F. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.
- Dee, D. P., Uppala, S. M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al. The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, 137(656):553–597, 2011.
- Dubois, Y., Gordon, J., and Foong, A. Y. Neural process family. <http://yanndubs.github.io/Neural-Process-Family/>, September 2020.
- Foong, A. Y., Bruinsma, W. P., Gordon, J., Dubois, Y., Requeima, J., and Turner, R. E. Meta-learning stationary stochastic process prediction with convolutional neural processes. *arXiv preprint arXiv:2007.01332*, 2020.
- Garnelo, M., Schwarz, J., Rosenbaum, D., Viola, F., Rezende, D. J., Eslami, S., and Teh, Y. W. Neural processes. *arXiv preprint arXiv:1807.01622*, 2018.
- Gordon, J., Bruinsma, W. P., Foong, A. Y., Requeima, J., Dubois, Y., and Turner, R. E. Convolutional conditional neural processes. *arXiv preprint arXiv:1910.13556*, 2019.

- Groenke, B., Madaus, L., and Monteleoni, C. Climalign: Unsupervised statistical downscaling of climate variables via normalizing flows. In *Proceedings of the 10th International Conference on Climate Informatics*, pp. 60–66, 2020.
- Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., Rössler, O., Wibig, J., Wilcke, R., Kotlarski, S., 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*, 39(9):3750–3785, 2019.
- Herdin, M., Czink, N., Ozcelik, H., and Bonek, E. Correlation matrix distance, a meaningful measure for evaluation of non-stationary mimo channels. In *2005 IEEE 61st Vehicular Technology Conference*, volume 1, pp. 136–140. IEEE, 2005.
- Hertig, E., Maraun, D., Bartholy, J., Pongracz, R., Vrac, M., Mares, I., Gutiérrez, J. M., Wibig, J., Casanueva, A., and Soares, P. M. Comparison of statistical downscaling methods with respect to extreme events over europe: Validation results from the perfect predictor experiment of the cost action value. *International Journal of Climatology*, 39(9):3846–3867, 2019.
- Klok, E. and Klein Tank, A. Updated and extended european dataset of daily climate observations. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 29(8):1182–1191, 2009.
- Liu, Y., Ganguly, A. R., and Dy, J. 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*, pp. 3145–3153, 2020.
- Maraun, D. and Widmann, M. *Statistical downscaling and bias correction for climate research*. Cambridge University Press, 2018.
- Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., Wibig, J., Huth, R., and Wilcke, R. A. Value: A framework to validate downscaling approaches for climate change studies. *Earth's Future*, 3(1):1–14, 2015.
- Maraun, D., Huth, R., Gutiérrez, J. M., Martín, D. S., Dubrovsky, M., Fischer, A., Hertig, E., Soares, P. M., Bartholy, J., Pongrácz, R., et al. The value perfect predictor experiment: evaluation of temporal variability. *International Journal of Climatology*, 39(9):3786–3818, 2019.
- Misra, S., Sarkar, S., and Mitra, P. Statistical downscaling of precipitation using long short-term memory recurrent neural networks. *Theoretical and applied climatology*, 134(3):1179–1196, 2018.
- Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., and Ganguly, A. R. Deepsd: Generating high resolution climate change projections through single image super-resolution. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1663–1672, 2017.
- Vandal, T., Kodra, E., and Ganguly, A. R. Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation. *Theoretical and Applied Climatology*, 137(1):557–570, 2019.
- Vaughan, A., Tebbutt, W., Hosking, J. S., and Turner, R. E. Convolutional conditional neural processes for local climate downscaling. *Geoscientific Model Development Discussions*, pp. 1–25, 2021.
- Wang, F., Tian, D., Lowe, L., Kalin, L., and Lehrter, J. Deep learning for daily precipitation and temperature downscaling. *Water Resources Research*, pp. e2020WR029308, 2021.
- Widmann, M., Bedia, J., Gutiérrez, J. M., Bosshard, T., Hertig, E., Maraun, D., Casado, M. J., Ramos, P., Cardoso, R. M., Soares, P. M., et al. Validation of spatial variability in downscaling results from the value perfect predictor experiment. *International Journal of Climatology*, 39(9):3819–3845, 2019.

## A. convLNP model

### A.1. Architecture

The encoder consists of three residual blocks with depth separable convolutions (Chollet, 2017). Each block consists of two layers of ReLU nonlinearities followed by a 2D convolutional layer with a kernel size of 3 and 128 intermediate channels. The latent variable  $\mathbf{z}$  has 64 channels.

The decoder consists of a further three residual blocks with architecture identical to above. This is followed by a set convolution (Gordon et al., 2019), where the outputs of the parameters on the low resolution grid are used as weights for an exponentiated-quadratic kernel with learnable length scale to make predictions at target location  $\mathbf{x}$ .

Resulting predictions of parameters are then concatenated with the elevation coordinate and passed through a MLP with four hidden layers, each with 64 neurons, and ReLU activations.

## A.2. Training

The convLNP is trained for 100 epochs using Adam, with a learning rate of  $5 \times 10^{-4}$ . As the log-predictive likelihood is not analytically tractable, we instead minimise the neural process maximum likelihood objective (Foong et al., 2020), defined by

$$\hat{\mathcal{L}}_{NPML} = \log\left(\frac{1}{L} \sum_{l=1}^L \prod_{t=1}^T p_{\theta}(\mathbf{y}^{(t)} | \mathbf{x}^{(t)}, \mathbf{z}_l)\right)$$

Where  $\mathbf{z}_l \sim p_{\theta}(\mathbf{z}|C)$ . The number of samples  $L$  from the latent variable is set to 24.

## B. Baseline model

The baseline MLP has 4 hidden layers each with 64 units and ReLU nonlinearities. Both the maximum temeprature and precipitation models are trained for 100 epochs using Adam with a learning rate of  $1 \times 10^{-3}$  minimising the mean squared error.