

# Earth's Future

## RESEARCH ARTICLE

10.1029/2023EF004002

### Key Points:

- A generative adversarial network is shown to improve daily precipitation fields from a state-of-the-art Earth system model
- Biases in long-term temporal distributions are strongly reduced by the generative adversarial network
- Our network-based approach can be complemented with quantile mapping to improve precipitation fields further

### Supporting Information:

Supporting Information may be found in the online version of this article.

### Correspondence to:

P. Hess,  
[philipp.hess@tum.de](mailto:philipp.hess@tum.de)

### Citation:

Hess, P., Lange, S., Schötz, C., & Boers, N. (2023). Deep learning for bias-correcting CMIP6-class Earth system models. *Earth's Future*, 11, e2023EF004002. <https://doi.org/10.1029/2023EF004002>

Received 7 FEB 2023

Accepted 10 SEP 2023

© 2023 The Authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

## Deep Learning for Bias-Correcting CMIP6-Class Earth System Models

Philipp Hess<sup>1,2</sup> , Stefan Lange<sup>2</sup> , Christof Schötz<sup>2</sup> , and Niklas Boers<sup>1,2,3</sup> 

<sup>1</sup>Earth System Modelling, School of Engineering and Design, Technical University of Munich, Munich, Germany, <sup>2</sup>Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Potsdam, Germany, <sup>3</sup>Department of Mathematics, Global Systems Institute, University of Exeter, Exeter, UK

**Abstract** The accurate representation of precipitation in Earth system models (ESMs) is crucial for reliable projections of the ecological and socioeconomic impacts in response to anthropogenic global warming. The complex cross-scale interactions of processes that produce precipitation are challenging to model, however, inducing potentially strong biases in ESM fields, especially regarding extremes. State-of-the-art bias correction methods only address errors in the simulated frequency distributions locally at every individual grid cell. Improving unrealistic spatial patterns of the ESM output, which would require spatial context, has not been possible so far. Here, we show that a postprocessing method based on physically constrained generative adversarial networks (cGANs) can correct biases of a state-of-the-art, CMIP6-class ESM both in local frequency distributions and in the spatial patterns at once. While our method improves local frequency distributions equally well as gold-standard bias-adjustment frameworks, it strongly outperforms any existing methods in the correction of spatial patterns, especially in terms of the characteristic spatial intermittency of precipitation extremes.

**Plain Language Summary** Modeling the impacts of changing precipitation characteristics due to anthropogenic global warming requires accurate and realistic simulations. These are obtained from Earth system models (ESMs) that solve the governing equations of the atmosphere and oceans on discretized spatial grids, with a resolution of currently around 50–100 km. Therefore, small-scale physical processes that are crucial for precipitation cannot be simulated directly but have to be included as parameterizations of the resolved variables. This introduces biases that can be adjusted in a postprocessing step. However, current methods only consider local biases in temporal distributions over long periods and do not account for unrealistic spatial patterns on short time scales. Generative adversarial networks are designed to translate unpaired images from one domain to another. Here, we use them to generate realistic global high-resolution precipitation fields from ESM simulations and show that the results improve upon those obtained with a state-of-the-art bias correction methodology. The results show strongly improved spatial patterns with realistic intermittent characteristics and a similar skill in correcting temporal biases.

## 1. Introduction

Precipitation is a crucial climate variable, and changing amounts, frequencies, or spatial distributions have potentially severe ecological and socioeconomic impacts. With global warming projected to continue in the coming decades, assessing the impacts of changes in precipitation characteristics is an urgent challenge (Boyle & Klein, 2010; IPCC, 2021; Wilcox & Donner, 2007). Climate impact models are designed to assess the impacts of global warming on, for example, ecosystems, crop yields, vegetation and other land-surface characteristics, infrastructure, water resources, or the economy, in general (Kotz et al., 2022), using the output of climate or Earth system models (ESMs) as input. Especially for reliable assessments of the ecological and socioeconomic impacts, accurate ESM precipitation fields to feed the impact models are therefore crucial.

ESMs are integrated on spatial grids with finite resolution. The resolution is limited by the computational resources that are necessary to perform simulations on decadal to centennial time scales. Current state-of-the-art ESMs have a horizontal resolution on the order of 100 km, in exceptional cases going down to 50 km. Smaller-scale physical processes that are relevant for the generation of precipitation operate on scales below the size of individual grid cells. These can, therefore, not be resolved explicitly in ESMs and have to be included as

parameterizations of the resolved prognostic variables. These include droplet interactions, turbulence, and phase transitions in clouds that play a central role in the generation of precipitation.

The limited grid resolution hence introduces errors in the simulated precipitation fields, leading to biases in short-term spatial patterns and long-term summary statistics. These biases need to be addressed prior to passing the ESM precipitation fields to impact models. In particular, climate impact models are often developed and calibrated with input data from reanalysis data rather than ESM simulations. These reanalyses are created with data assimilation routines and combine various observations with high-resolution weather models. They hence provide a much more realistic input than the ESM simulations and statistical bias correction methods are necessary to remove biases in the ESM simulations output and to make them more similar to the reanalysis data for which the impact models are calibrated. Quantile mapping (QM) is a standard technique to correct systematic errors in ESM simulations. QM estimates a mapping between distributions from historical simulations and observations that can thereafter be applied to future simulations in order to provide more accurate simulated precipitation fields to impact models (Cannon et al., 2015; Déqué, 2007; Gudmundsson et al., 2012; Tong et al., 2021).

State-of-the-art bias correction methods such as QM are, however, confined to address errors in the simulated frequency distributions locally, that is, at every grid cell individually. Unrealistic spatial patterns of the ESM output, which would require spatial context, have so far not been addressed by postprocessing methods. For precipitation, this is particularly important because it has characteristic high intermittency not only in time but also in its spatial patterns. Multivariate bias correction approaches have recently been developed, aiming to improve spatial dependencies (Cannon, 2018; Vrac, 2018). However, these approaches are typically only employed in regional studies, as the input dimension becomes too large for global high-resolution ESM simulations. Moreover, such methods have been reported to suffer from instabilities and overfitting, while differences in their applicability and assumptions make them challenging to use (François et al., 2020).

Artificial neural networks from computer vision and image processing have been successfully applied to various tasks in Earth system science, ranging from weather forecasting (Bi et al., 2023; Rasp & Thuerey, 2021; Ravuri et al., 2021; Weyn et al., 2020; Zhang et al., 2023), to postprocessing numerical weather prediction fields (Grönquist et al., 2021; Price & Rasp, 2022; Wang et al., 2023), by extracting spatial features with convolutional layers (LeCun et al., 2015). Such deep learning-based postprocessing methods can be roughly grouped into two categories, using either paired or unpaired training data. Wang and Tian (2022), for instance, train a supervised convolutional neural network on paired data to correct temperature biases in the output of a ESM from the Climate Model Intercomparison Project phase 6 (CMIP6). On the other hand, Ravuri et al. (2021) use a generative machine learning model for highly skillful precipitation nowcasting.

Here, we employ a recently introduced postprocessing method (Hess et al., 2022) that can be trained on unpaired data, based on a cycle-consistent adversarial network (CycleGAN) to consistently improve both local frequency distributions and spatial patterns of state-of-art CMIP6 ESM precipitation fields. Generative adversarial networks (Goodfellow et al., 2014) in particular have emerged as a promising architecture that produces sharp images that are necessary to capture the high-frequency variability of precipitation (L. Harris et al., 2022; Price & Rasp, 2022; Ravuri et al., 2021). GANs have been specifically developed to be trained on unpaired image data sets (Zhu et al., 2017). This makes them a natural choice for postprocessing the output of climate projections, which—unlike weather forecasts—are not nudged to follow the trajectory of observations; due to the chaotic nature of the atmosphere, small deviations in the initial conditions or parameters lead to exponentially diverging trajectories (Lorenz, 1996). As a result, numerical weather forecasts lose their deterministic forecast skill after approximately 2 weeks at most, and century-scale climate simulations do not agree with observed daily weather records. Indeed the task of climate models is rather to produce accurate long-term statistics that agree with observations.

We apply our GAN approach to correct global high-resolution precipitation simulations of the GFDL-ESM4 model (Krasting et al., 2018) as a representative CMIP6-class ESM. So far, GAN-based approaches have only been applied to postprocess ESM simulations either in a regional context (François et al., 2021) or to a very low-resolution global ESM (Hess et al., 2022). In the latter study, the aim was to improve precipitation fields from an efficient and computationally lightweight ESM, making the combined model a fast ESM with precipitation fields competitive with state-of-the-art models. However, the latter study did not show if the cGAN approach could—in addition to contributing to an efficient, lightweight ESM—also be used for bias-correcting state-of-the-art, CMIP6-class models for subsequent impact modeling.

Here, we show that, indeed, a suitably designed and trained cGAN can improve even the distributions and spatial patterns of precipitation fields from a state-of-the-art comprehensive ESM, namely GFDL-ESM4. In particular, in contrast to rather specific existing methods for postprocessing ESM output for climate impact modeling, we will show that the GAN approach is general and can readily be applied to different ESMs and observational data sets used as ground truth.

In order to ensure that the cGAN-based postprocessing does not violate physical conservation laws, we include a suitable physical constraint, enforcing that the cGAN-based transformations do not change the overall global sum of daily precipitation values; essentially, this ensures that precipitation is only spatially redistributed (see Section 2). By framing bias correction as an image-to-image translation task, our approach corrects both spatial patterns of daily precipitation fields on short time scales and temporal distributions aggregated over decadal time scales. We evaluate the skill in improving spatial patterns and temporal distributions against the gold-standard Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) bias-adjustment and statistical downscaling (ISIMIP3BASD) method framework (Lange, 2019), which relies strongly on QM.

Quantifying the “realisticness” of spatial precipitation patterns is a key problem in current research (Ravuri et al., 2021). We use spatial spectral densities and the fractal dimension of spatial patterns as a measure to quantify the similarity of intermittent and unpaired precipitation fields. We will show that our GAN can learn to recognize spatial patterns using convolutional layers and strongly improves the characteristic intermittency in spatial precipitation patterns. We will also show that our GAN, combined with a subsequent application of the ISIMIP3BASD routine, leads to the best overall performance in bias-correcting comprehensive ESM fields.

## 2. Methods

### 2.1. Training Data

We use global fields of daily precipitation with a horizontal resolution of  $1^\circ$  from the GFDL-ESM4 ESM (Krasting et al., 2018) historical simulation and the W5E5v2 reanalysis product (WFDE5 over land merged with ERA5 over the ocean [W5E5 v2.0] [Dataset], 2021) as observation-based ground truth. The W5E5v2 data set is based on the ERA5 (Hersbach et al., 2020) reanalysis and has been bias adjusted using the Global Precipitation Climatology Centre (GPCC) full data monthly product v2020 (Schneider et al., 2011) over land and the Global Precipitation Climatology Project (GPCP) v2.3 data set (Huffman et al., 1997) over the ocean to further improve the precipitation statistics of ERA5. Both data sets have been regridded to the same  $1^\circ$  horizontal resolution using bilinear interpolation following Beck et al. (2019). We split the data set into three periods for training (1950–2000), validation (2001–2003), and testing (2004–2014). This corresponds to 8,030 samples for training, 1,095 for validation, and 4,015 for testing. During preprocessing, the training data are log transformed with  $\tilde{x} = \log(x + \epsilon) - \log(\epsilon)$  with  $\epsilon = 0.0001$ , following Rasp and Thuerey (2021), to account for zeros in the transform. The data are then normalized to the interval  $[-1, 1]$  following Zhu et al. (2017).

### 2.2. Cycle-Consistent Generative Adversarial Networks

This section gives a brief overview of the GAN used in this study. We refer to Hess et al. (2022) and Zhu et al. (2017) for a more comprehensive description and discussion. Generative adversarial networks learn to generate images that are nearly indistinguishable from real-world examples through a two-player game (Goodfellow et al., 2014). In this setup, a first network  $G$ , the so-called generator, produces images with the objective of fooling a second network  $D$ , the discriminator, which has to classify whether a given sample is generated (“fake”) or drawn from a real-world data set (“real”). Mathematically this can be formalized as

$$G^* = \min_G \max_D \mathcal{L}_{\text{GAN}}(D, G), \quad (1)$$

with  $G^*$  being the optimal generator network. The loss function  $\mathcal{L}_{\text{GAN}}(D, G)$  can be defined as

$$\mathcal{L}_{\text{GAN}}(D, G) = \mathbb{E}_{y \sim p_y(y)} [\log(D(y))] + \mathbb{E}_{x \sim p_x(x)} [\log(1 - D(G(x)))], \quad (2)$$

where  $p_y(y)$  is the distribution of the real-world target data. Samples from the ESM simulation data distribution, here denoted as  $p_x(x)$ , are used as inputs by  $G$  to produce realistic images. The cGAN (Zhu et al., 2017) consists of two generator–discriminator pairs, where the generators  $G$  and  $F$  learn inverse mappings between two domains

$X$  and  $Y$ . This allows defining an additional cycle-consistency loss that constrains the training of the networks, that is,

$$\begin{aligned} \mathcal{L}_{\text{cycle}}(G, F) = & \mathbb{E}_{x \sim p_x(x)} [\|F(G(x)) - x\|_1] \\ & + \mathbb{E}_{y \sim p_y(y)} [\|G(F(y)) - y\|_1]. \end{aligned} \quad (3)$$

It measures the error caused by a translation cycle of an image to the other domain and back. Further, an additional loss term is introduced to regularize the networks to be close to an identity mapping with

$$\begin{aligned} \mathcal{L}_{\text{ident}}(G, F) = & \mathbb{E}_{x \sim p_x(x)} [\|F(x) - x\|_1] \\ & + \mathbb{E}_{y \sim p_y(y)} [\|G(y) - y\|_1]. \end{aligned} \quad (4)$$

The regularization forces the generator networks to output a similar image as given as input to the network. This is in line with the idea that the generator should only improve small-scale features of the image, that is, change the style, while preserving the large-scale features, that is, the overall content of the image. In practice, the log likelihood loss can be replaced by a mean squared error loss to facilitate more stable training. Further, the generator loss is reformulated to be minimized by inverting the labels, that is,

$$\begin{aligned} \mathcal{L}_{\text{Generator}} = & \mathbb{E}_{x \sim p_x(x)} [(D_X(G(x)) - 1)^2] \\ & + \mathbb{E}_{y \sim p_y(y)} [(D_Y(F(y)) - 1)^2] \\ & + \lambda \mathcal{L}_{\text{cycle}}(G, F) + \tilde{\lambda} \mathcal{L}_{\text{ident}}(G, F), \end{aligned} \quad (5)$$

where  $\lambda$  and  $\tilde{\lambda}$  are set to 10 and 5, respectively, following Zhu et al. (2017). The corresponding loss term for the discriminator networks is given by

$$\begin{aligned} \mathcal{L}_{\text{Discriminator}} = & \mathbb{E}_{y \sim p_y(y)} [(D_Y(y) - 1)^2] + \mathbb{E}_{x \sim p_x(x)} [(D_Y(G(x)))^2] \\ & + \mathbb{E}_{x \sim p_x(x)} [(D_X(x) - 1)^2] + \mathbb{E}_{y \sim p_y(y)} [(D_X(F(y)))^2]. \end{aligned} \quad (6)$$

The weights of the generator and discriminator networks are then optimized with the ADAM (Kingma & Ba, 2014) optimizer using a learning rate of  $2e^{-4}$  and updated in an alternating fashion. We train the network for 350 epochs and a batch size of 1, saving model checkpoints every other epoch. We evaluate the checkpoints on the validation data set to determine the best model instance.

### 2.3. Network Architectures

Both the generator and discriminator have fully convolutional architectures. The generator uses ReLU activation functions, instance normalization, and reflection padding. The discriminator uses leaky ReLU activations with slope 0.2 instead, together with instance normalization. For a more detailed description of the cGAN approach, we refer to our previous study (Hess et al., 2022). The general network architectures in this study are similar, but with an increase in the number of residual layers in the generator network (the constrained and unconstrained networks) from 6 to 7.

The final layer of the generator can be constrained to preserve the global sum of the input, that is, by rescaling

$$\tilde{y}_i = y_i \frac{\sum_i^{N_{\text{grid}}} x_i}{\sum_i^{N_{\text{grid}}} y_i}, \quad (7)$$

where  $x_i$  and  $y_i$  are grid cell values of the generator input and output, respectively, and  $N_{\text{grid}}$  is the number of grid cells. The generator without this constraint will be referred to as unconstrained in this study. The global physical constraint enforces that the global daily precipitation sum is not affected by the cGAN postprocessing and hence remains identical to the original value from the GFDL-ESM4 simulations. This is motivated by the observation that large-scale average trends in precipitation follow the Clausius-Clapeyron relation (Traxl et al., 2021), which is based on thermodynamic relations and hence can be expected to be modeled well in GFDL-ESM4.

### 2.4. Quantile Mapping-Based Bias Adjustment

We compare the performance of our cGAN-based method to the bias-adjustment method ISMIP3BASD v3.0.1 (Lange, 2019, 2022) that has been developed for phase 3 of the ISMIP3 (Frieler et al., 2017; Warszawski

et al., 2014). This state-of-the-art bias-adjustment method is based on a trend-preserving QM framework. It represents a very strong baseline for comparison as it has been developed prior to this study and used not only in ISIMIP3 but also to prepare many of the climate projections that went into the Interactive Atlas produced as part of the sixth assessment report of Working Group 1 of the Intergovernmental Panel on Climate Change (IPCC, <https://interactive-atlas.ipcc.ch/>).

In the most basic form of QM, cumulative distribution functions (CDFs) of the historical simulation  $F_{\text{mod,hist}}(x)$  and observations  $F_{\text{obs,hist}}(x)$  are fitted to historical observations and simulations, which can then be applied to bias-adjust future simulations, that is,

$$\hat{x}_{\text{mod,proj}}(t) = F_{\text{obs,hist}}^{-1} \left[ F_{\text{mod,hist}} \left( x_{\text{mod,proj}}(t) \right) \right], \quad (8)$$

where  $x_{\text{mod,proj}}(t)$  is the model variable at a given grid cell and time instance from a climate projection simulation, and  $\hat{x}_{\text{mod,proj}}(t)$  is the quantile-mapped estimate. The CDFs used for QM can be either empirical or parametric.

In the present study, we apply parametric QM to the GFDL-ESM4 simulation and empirical QM to the constrained cGAN output. The parametric QM is carried out using ISIMIP3BASD v3.0.1, with settings as specified in Lange (2019) for precipitation, that is, Bernoulli-gamma distributions are used for the CDFs and trends in wet-day precipitation quantiles are preserved multiplicatively in most cases. The empirical QM is carried out in the same way as the parametric case, with the difference that empirical instead of parametric distributions are used. In both cases (empirical and parametric QM), CDFs are fitted and mapped for each grid cell and day of the year separately, using data from a window of 31 days width centered on the given day of the year for CDF fitting. We also tested empirical QM for the GFDL-ESM4 simulation and parametric QM for the constrained cGAN output, but those choices led to slightly worse results.

To evaluate the methods in this study, we define the grid cell-wise bias as the difference in long-term averages,

$$\text{Bias}(\hat{y}, y) = \frac{1}{T} \sum_{t=1}^T \hat{y}_t - \frac{1}{T} \sum_{t=1}^T y_t, \quad (9)$$

where  $T$  is the number of time steps;  $\hat{y}_t$  and  $y_t$  the modeled and observed precipitation, respectively, at time step  $t$ .

## 2.5. Evaluating Extreme Event Statistics

We are particularly interested in the cGAN's performance in correcting the characteristics of extreme rainfall. To evaluate and benchmark its performance to improve temporal statistics, we compare the representation of extreme event statistics in the reanalysis W5E5v2 against the raw GFDL-ESM4 model output and the bias-corrected data, in terms of return values and waiting times.

### 2.5.1. Estimating Return Values

The probability of a precipitation event  $X$  to exceed a threshold  $t \in \mathbb{R}$  is given by  $p = 1 - P(X \leq t)$ , where  $P(X \leq t)$  is the CDF. The return time, that is, the average number of time steps until an event exceeds the threshold  $t$ , is then given by  $r = 1/p$  and  $t$  is called the return value. The aim in the following is to construct the function  $t = U(r)$  that maps a return time  $r$  to a corresponding return value  $t$ .

We employ a peaks-over-threshold method: denote by  $X_{(1)} \leq \dots \leq X_{(n)}$  the local precipitation events ordered by magnitude. Choose  $k = \lceil 0.05n \rceil$  so that  $X_{(n-k+1)}, \dots, X_{(n)}$  exceed the 95th percentile of the empirical distribution of local precipitation events. Following classical results from extreme value theory, for example, de Haan and Ferreira (2006), we can use these excesses to estimate the tail of their distribution. Here, we apply the so-called moment estimator (de Haan & Ferreira, 2006, Theorem 4.3.1 and Equations 3.5.2, 3.5.9, and 4.2.4):

$$M_j := \frac{1}{k} \sum_{i=0}^{k-1} (\log(X_{(n-i)}) - \log(X_{(n-k)}))^j, \quad (10)$$

$$\hat{y} := M_1 + 1 - \frac{1}{2} \left( 1 - \frac{M_1^2}{M_2} \right)^{-1}, \quad (11)$$

$$\hat{\sigma} := \frac{1}{2} X_{(n-k)} M_1 \left( 1 - \frac{M_1^2}{M_2} \right)^{-1}, \quad (12)$$

$$\hat{U}(r) := X_{(n-k)} + \hat{\sigma} \frac{\left( \frac{rk}{n} \right)^{\hat{\gamma}} - 1}{\hat{\gamma}}. \quad (13)$$

The advantage of using the moment estimator over the also common maximum likelihood approach is that the former is more computationally efficient and thus more suitable for the high-resolution fields in our study. Using Equation 13, we can determine the return value for a given, large return time and a given grid cell to compare the different postprocessing methods with regard to their capability of improving these features.

### 2.5.2. Waiting Time Distributions

Besides return values for a specific return time, that is, the average waiting time between consecutive events, the *distribution* of waiting times is another key characteristic of extremes and crucial for hydrological impact modeling. We compare the waiting time distributions of the reanalysis and modeled precipitation data by first applying a threshold, defined as the 95th percentile of the local empirical distribution, to the precipitation times series. Counting the number of days between consecutive precipitation events then determines the waiting times. We compare the empirical distributions of the waiting times for each grid cell between the W5E5v2 reanalysis and the different postprocessing methods using the Wasserstein-1 distance between distributions, defined as

$$W_1(p_x, p_y) = \int_{\mathbb{R}} |F_x(z) - F_y(z)| dz, \quad (14)$$

where  $p_x$  and  $p_y$  are the distributions of the modeled waiting times and of the reanalysis (W5E5v2) target waiting times, respectively, and  $F_x$  and  $F_y$ , their respective CDFs.

### 2.6. Evaluating Spatial Patterns

Quantifying how realistic spatial precipitation fields are, is an ongoing research question in itself, which has become more important with the application of deep learning to weather forecasting and postprocessing. In these applications, neural networks often achieve error statistics and skill scores competitive with physical models, while the output fields can, at the same time, show unphysical characteristics, such as blurring or excessive smoothing. Ravuri et al. (2021) compare the spatial intermittency, which is characteristic of precipitation fields, using the radially averaged power spectral density (RAPSD) computed from the spatial fields; in the latter study, the RAPSD-based quantification was complemented by interviews with a large number of meteorological experts.

Power spectra provide a useful tool for studying patterns across wide ranges of scales (D. Harris et al., 2001) by applying a Fourier transform that describes the spatial signal in terms of frequencies. The power spectrum of a 2D field  $f(x, y)$  can be defined as (Ruzanski & Chandrasekar, 2011)

$$P(f) = |F(k_x, k_y)|^2,$$

where

$$F(k_x, k_y) = \frac{1}{MN} \sum_{x=-M/2}^{M/2-1} \sum_{y=-N/2}^{N/2-1} f(x, y) e^{-i2\pi \left( \frac{k_x x}{M} + \frac{k_y y}{N} \right)},$$

and  $M$  and  $N$  are the dimensions of the field. The radially averaged power spectrum is then found by averaging the power spectrum  $P(f)$  in polar coordinates in angular direction as a function of the radius  $k = \sqrt{k_x^2 + k_y^2}$ , such that

$$\bar{P}_r(k) = \frac{1}{N_r(k)} \sum_{i=1}^{N_r(k)} P(k_i), \quad (15)$$

where  $N_r(k)$  is the number of frequency samples and the spatial radial wavelength  $\lambda = \Delta/k$  with  $\Delta$  is the grid spacing of the field in cartesian coordinates.

We propose here the fractal dimension of precipitation fields as a new metric to quantify how realistic precipitation patterns from generative neural networks are. The fractal dimension has been used in the past already to study how precipitation patterns change across scales (Lovejoy et al., 1987). We compute the fractal dimension via the box-counting algorithm (Husain et al., 2021; Lovejoy et al., 1987; Meisel et al., 1992). The box-counting algorithm divides the image into  $N_{\text{sq}}$  squares of linear size  $s$  that cover the boundary of binary patterns. We thus first transform the continuous precipitation data into binary fields by applying a global quantile threshold (see Figure 8 for an example). The size of the squares, that is, the scale of measurement, is then reduced iteratively by a factor of 2 and the number of squares or “boxes” is counted at each scale. The fractal dimension (FD) can then be determined from the slope of a polynomial fit of the resulting log–log scaling of  $N_{\text{sq}}$  as a function of the measured scale with

$$\text{FD} = \frac{\log(N_{\text{sq}})}{\log(1/s)}. \quad (16)$$

### 3. Results

We evaluate our cGAN method on two different tasks and time scales. First, the correction of daily rainfall frequency distributions at each grid cell locally, aggregated from decade-long time series. Second, we quantify the ability to improve spatial patterns on daily time scales. Our approach is compared to the raw GFDL-ESM4 model output, as well as to the ISIMIP3BASD methodology applied to the GFDL-ESM4 output.

#### 3.1. Temporal Distributions

We compute global histograms of relative precipitation frequencies using daily time series (Figure 1a). The GFDL-ESM4 model overestimates frequencies in the tail, namely for events above 50 mm/day (i.e., the 99.7th percentile). Our GAN-based method, as well as ISIMIP3BASD and the cGAN-ISIMIP3BASD combination, corrects the histogram to match the W5E5v2 ground truth equally well, as can also be seen in the absolute error of the histograms (Figure 1b).

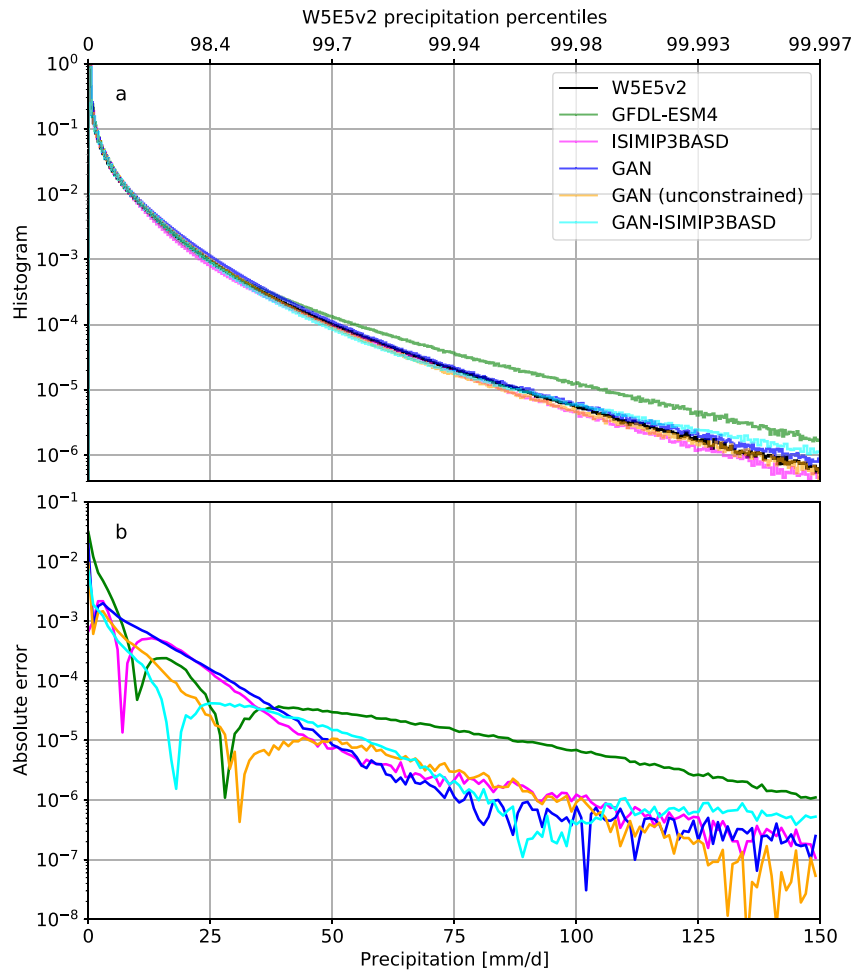
Comparing the differences in long-term averages of precipitation per grid cell (Figure 2 and Section 2), large biases are apparent in the GFDL-ESM4 model output, especially in the tropics. The double-peaked Inter-Tropical Convergence Zone (ITCZ) bias is visible. The double-ITCZ bias can also be inferred from the latitudinal profile of the precipitation mean in Figure 3.

Table 1 summarizes the annual biases shown in Figure 2 as absolute averages and additionally for the four seasons. The cGAN alone reduces the annual bias of the GFDL-ESM4 model by 38.7%. The unconstrained cGAN performs better than the physically constrained one, with bias reductions of 50.5%. As expected, the ISIMIP3BASD gives even better results for correcting the local mean since it is specifically designed to transform the local frequency distributions accurately. It is, therefore, notable that applying the ISIMIP3BASD procedure on the constrained cGAN output improves the postprocessing further, leading to a local bias reduction of the mean by 63.6%, compared to ISIMIP3BASD with 59.4%. For seasonal time series, the order in which the methods perform is the same as for the annual data.

Besides the error in the mean, we also compute differences in the 95th percentile for each grid cell, shown in Figure S1 of the Supporting Information S1 and as mean absolute errors (MAEs) in Table 1. Also, in the case of heavy precipitation values, we find that ISIMIP3BASD improves upon the cGAN but that combining cGAN and ISIMIP3BASD leads to the best agreement of the locally computed quantiles.

#### 3.2. Extremes

We evaluate the error of the return values for extreme events with a 10-year return time (see Figure 4). The GFDL-ESM4 model shows a strong overprediction of the return values in the tropics. This is also reflected in the high global MAE over the return value differences of 49.9. The ISIMIP3BASD method is able to reduce the MAE to 34.37 but shows a tendency to underestimate the return values. Both the unconstrained and the constrained cGAN alone perform remarkably well, reducing the MAE to 30.53 and 31.44, respectively. The constrained cGAN shows a slight overestimate of return values in the midlatitudes. Combining the cGAN with the ISIMIP3BASD

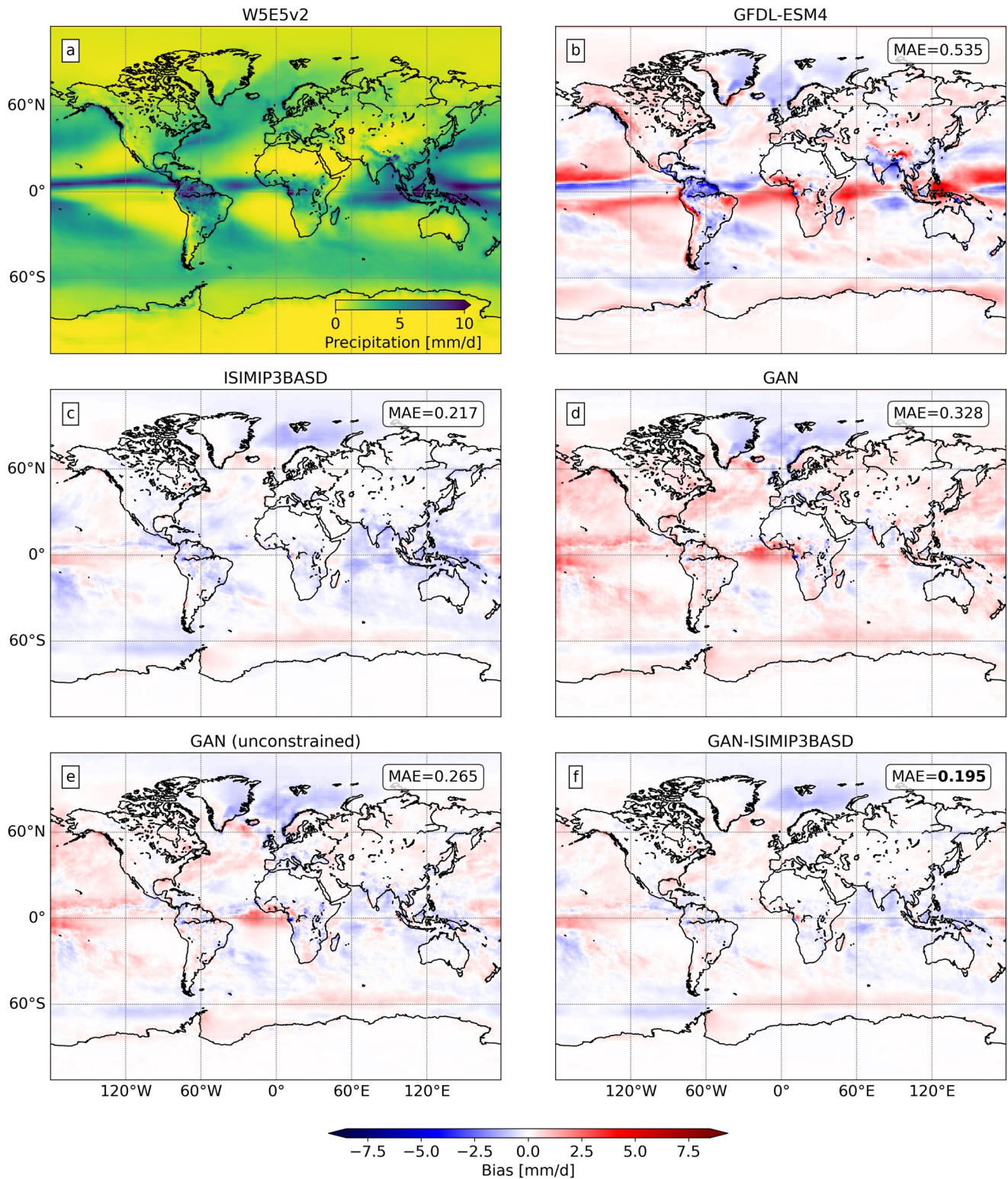


**Figure 1.** Histograms of relative precipitation frequencies over the entire globe and test period (2004–2014). (a) The histograms are shown for the W5E5v2 ground truth (black), GFDL-ESM4 (green), ISIMIP3BASD (magenta), cGAN (blue), unconstrained cGAN (orange), and the (constrained) cGAN-ISIMIP3BASD combination (cyan). (b) Distances of the histograms to the W5E5v2 ground truth are shown for the same models as in (a). Percentiles corresponding to the W5E5v2 precipitation values are given on the second x-axis at the top. Note that GFDL-ESM4 overestimates the frequencies of strong and extreme rainfall events. All compared methods show similar performance in correcting the local frequency distributions.

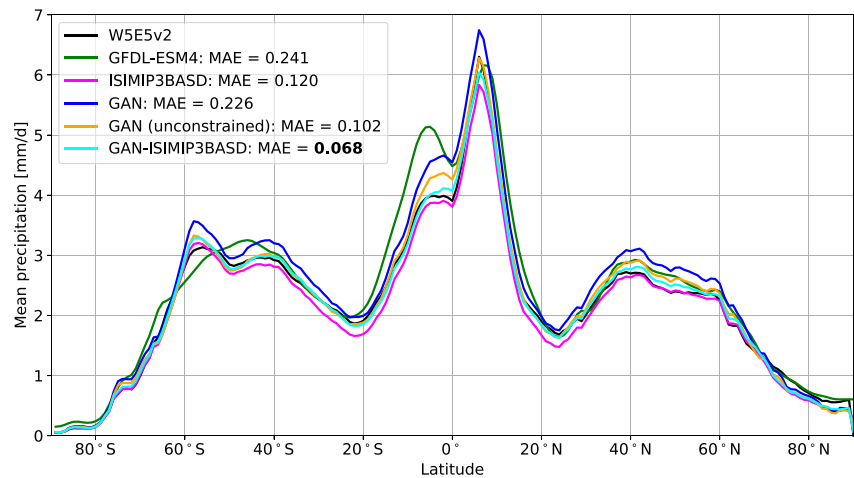
method shows an overall MAE of 38.27 and does not lead to improved performance (Figure 4). We compare latitude profiles of the return value errors, where particularly the cGAN-based postprocessings show a significant error reduction in the tropics (Figure S5 in Supporting Information S1).

To measure the error in the waiting times between extremes, we report the Wasserstein distance between the W5E5v2 reanalysis and the modeled waiting time distributions of precipitation events above the 95th percentile in Figure 5. Notably, the GFDL-ESM4 model shows a large distance to the ground truth over continental regions, especially in North America and East Asia, which is reduced by all employed postprocessing methods. The cGAN-ISIMIP3BASD method improves upon the other approaches with a global MAE of 11.26 in the Wasserstein distance, compared to 17.79 in the case of GFDL-ESM4, 11.43 for ISIMIP3BASD alone, and the constrained and unconstrained cGANs with MAEs of 12.38 and 12.4, respectively. A comparison with the number of events in Figure 5a shows that the distance in the distributions remains comparably large after correction for regions where the number of precipitation events over the 10-year test period is particularly low. Hence, the error in these regions might be caused by either the models themselves or the insufficient amount of precipitation events. We compare latitude profiles of the Wasserstein distances, which shows a general reduction of distances in the raw GFDL-ESM4 simulation through the postprocessing methods, with the best performance given by cGAN-ISIMIP3BASD (Figure S6 in Supporting Information S1).





**Figure 2.** Bias in the long-term average precipitation over the entire test set between the W5E5v2 ground truth (a) and GFDL-ESM4 (b), ISIMIP3BASD (c), cGAN (d), unconstrained cGAN (e), and the (constrained) cGAN-ISIMIP3BASD combination (f). The mean absolute error w.r.t. the W5E5v2 ground truth is shown in the upper right corner.



**Figure 3.** Precipitation averaged over longitudes and the entire test set period from the W5E5v2 ground truth (black) and GFDL-ESM4 (green), ISIMIP3BASD (magenta), cGAN (blue), unconstrained cGAN (orange), and the (constrained) cGAN-ISIMIP3BASD combination (cyan). To quantify the differences between the shown lines, we show their mean absolute error w.r.t. the W5E5v2 ground truth in the legend. These values are different from the ones shown in Table 1 as the average is taken here over the longitudes without their absolute value. The cGAN-ISIMIP3BASD approach shows the lowest error.

### 3.3. Spatial Patterns

We compare the ability of the cGAN to improve spatial patterns based on the W5E5v2 ground truth, against the GFDL-ESM4 simulations and the ISIMIP3BASD method applied to the GFDL-ESM4 simulations. To model realistic precipitation fields, the characteristic spatial intermittency needs to be captured accurately.

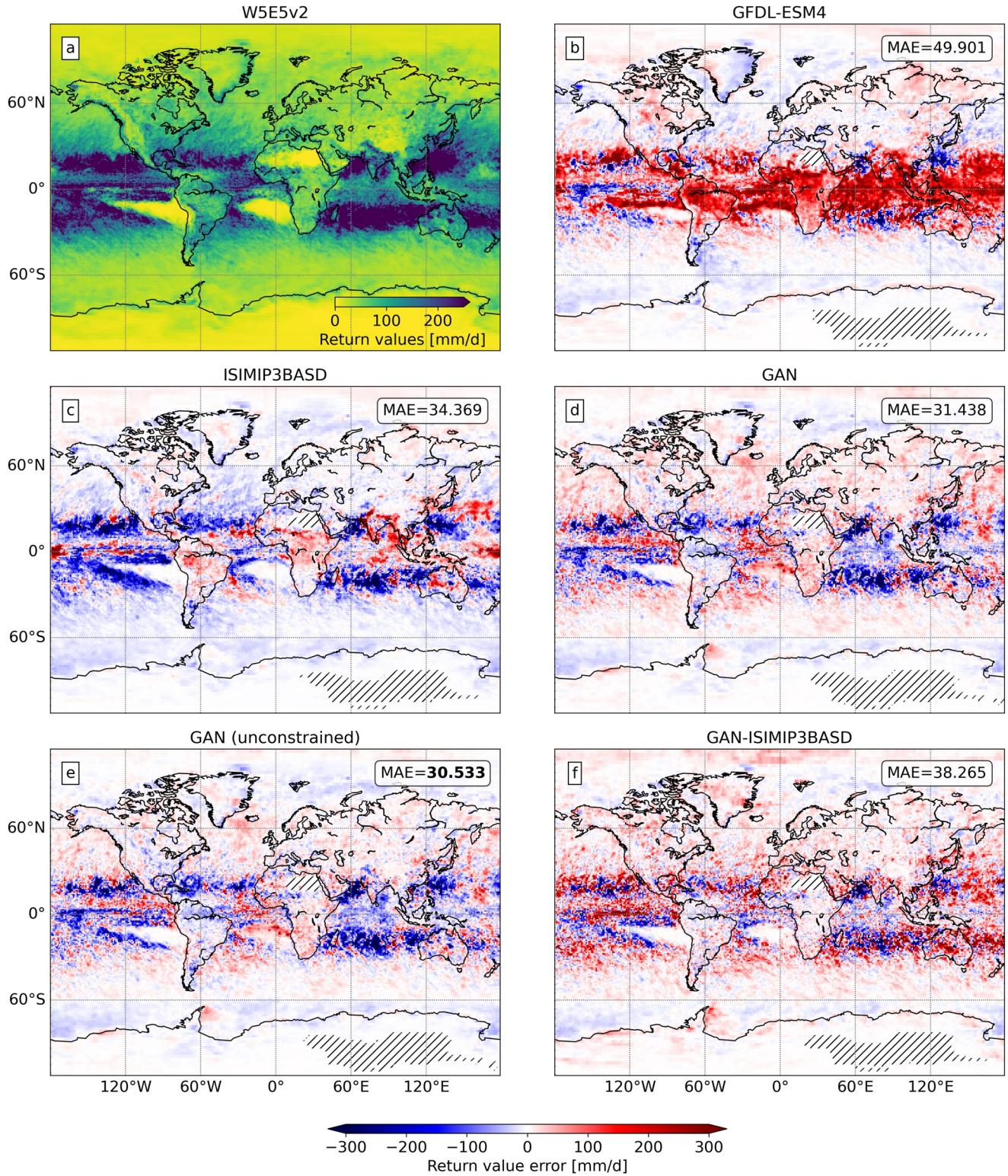
We compute the RAPSD of global precipitation fields, averaged over the test set for each method (Figure 6a). GFDL-ESM4 shows noticeable deviations from W5E5v2 in the RAPSD (Figure 6b). Our cGAN can correct these over the entire range of wavelengths, closely matching the W5E5v2 ground truth. Improvements over ISIMIP3BASD are especially pronounced in the range of high frequencies (low wavelengths), which are responsible for the intermittent spatial variability of daily precipitation fields. Adding the physical constraint to the cGAN does not affect the ability to produce realistic RAPSD distributions. After applying ISIMIP3BASD to the cGAN-processed fields, most of the improvements generated by the cGAN are retained, as shown by the cGAN-ISIMIP3BASD results.

**Table 1**

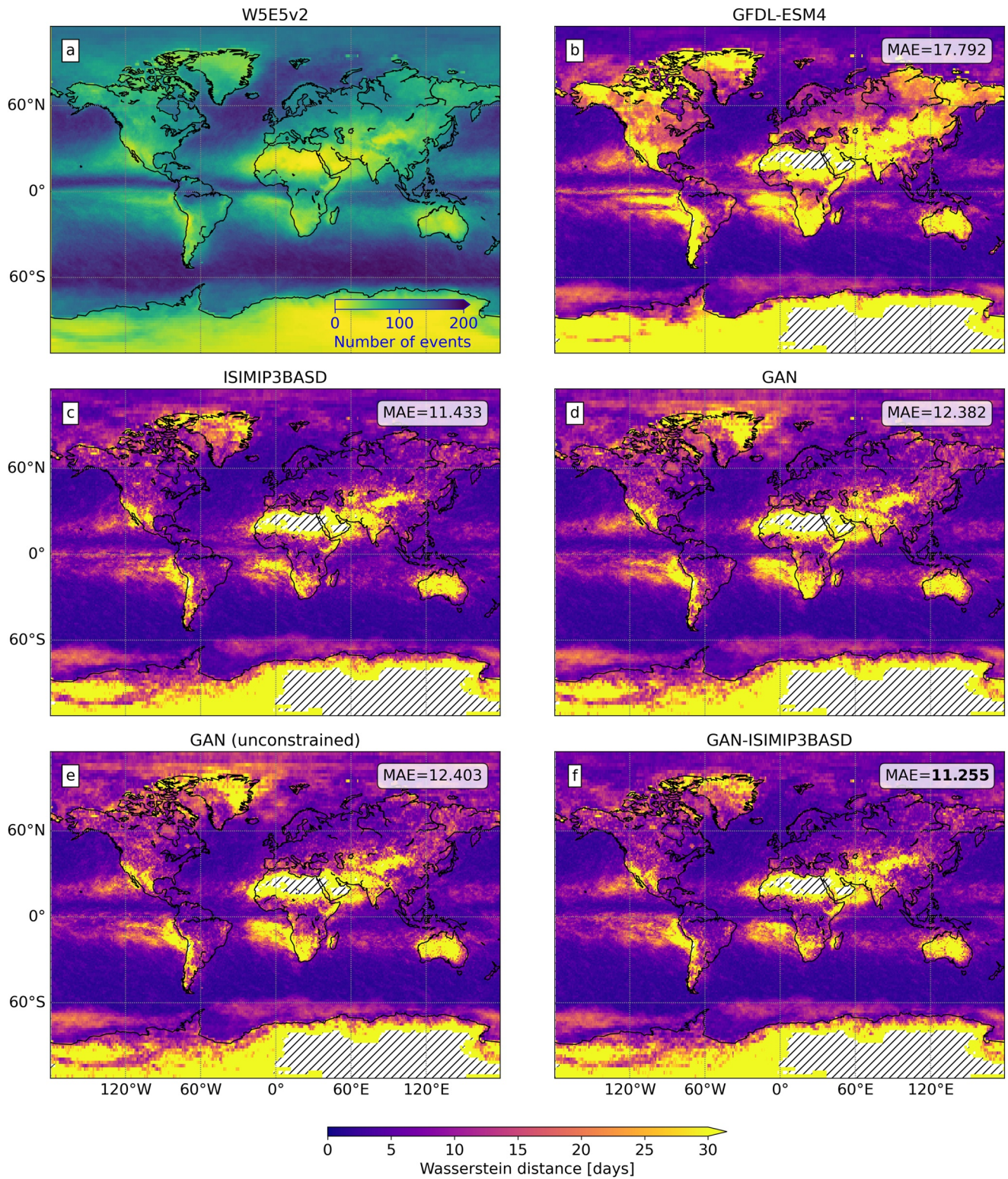
*The Globally Averaged Absolute Value of the Grid Cell-Wise Difference in the Long-Term Precipitation Average, As Well As the 95th Percentile, Between the W5E5v2 Ground Truth and GFDL-ESM4, ISIMIP3BASD, cGAN, Unconstrained cGAN, and the (Constrained) cGAN-ISIMIP3BASD Combination for Annual and Seasonal Time Series (in mm/day)*

Season	Percentile	GFDL-ESM4	ISIMIP3-BASD	%	cGAN	%	cGAN (unconst.)	%	cGAN-ISIMIP3-BASD	%
Annual	–	0.535	0.217	59.4	0.328	38.7	0.265	50.5	<b>0.195</b>	<b>63.6</b>
DJF	–	0.634	0.321	49.4	0.395	37.7	0.371	41.5	<b>0.308</b>	<b>51.4</b>
MAM	–	0.722	0.314	56.5	0.419	42.0	0.378	47.6	<b>0.285</b>	<b>60.5</b>
JJA	–	0.743	0.289	61.1	0.451	39.3	0.357	52.0	<b>0.280</b>	<b>62.3</b>
SON	–	0.643	0.327	49.1	0.409	36.4	0.362	43.7	<b>0.306</b>	<b>52.4</b>
Annual	95th	2.264	1.073	52.6	1.415	37.5	1.213	46.4	<b>0.945</b>	<b>58.3</b>
DJF	95th	2.782	1.496	46.2	1.725	38.0	1.655	40.5	<b>1.432</b>	<b>48.5</b>
MAM	95th	2.948	1.482	49.7	1.805	38.8	1.661	43.7	<b>1.337</b>	<b>54.6</b>
JJA	95th	2.944	1.366	53.6	1.852	37.1	1.532	48.0	<b>1.247</b>	<b>57.6</b>
SON	95th	2.689	1.495	44.4	1.741	35.3	1.592	40.8	<b>1.366</b>	<b>49.2</b>

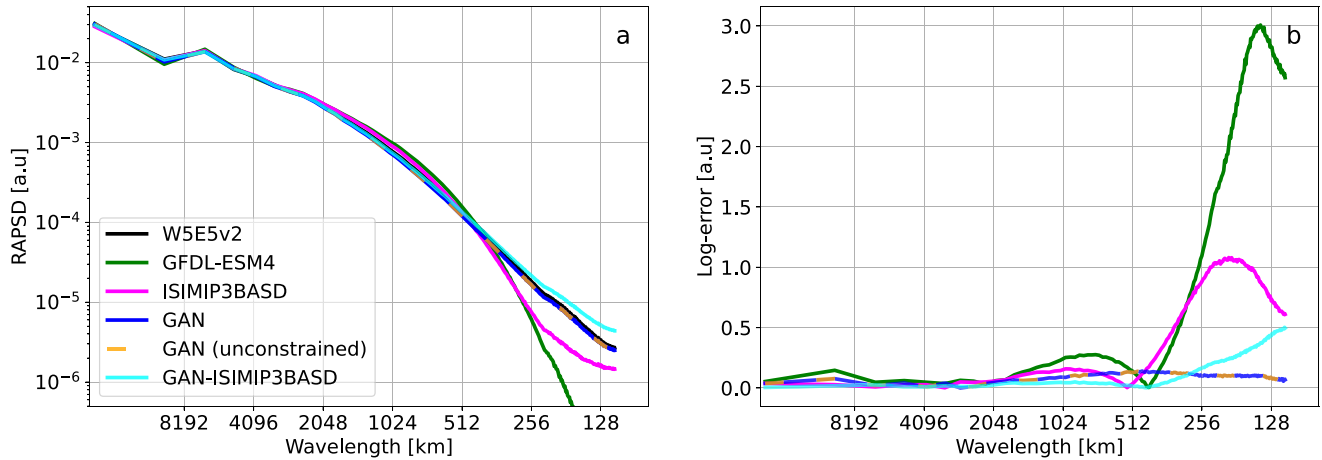
*Note.* The relative improvement over the raw GFDL-ESM4 climate model output is shown as percentages for each method. The best performance is indicated in bold font.



**Figure 4.** The return values for 10-year return times are shown for the W5E5v2 ground truth (a) and the respective errors for GFDL-ESM4 (b), ISIMIP3BASD (c), the cGAN (d), the unconstrained cGAN (e), and the (constrained) cGAN-ISIMIP3BASD combination (f). The global mean absolute error (MAE) with respect to the W5E5v2 ground truth is shown in the top right corner of each panel, with the unconstrained cGAN exhibiting the lowest error. The hatched area indicates regions that were excluded from the analysis due to insufficient events.

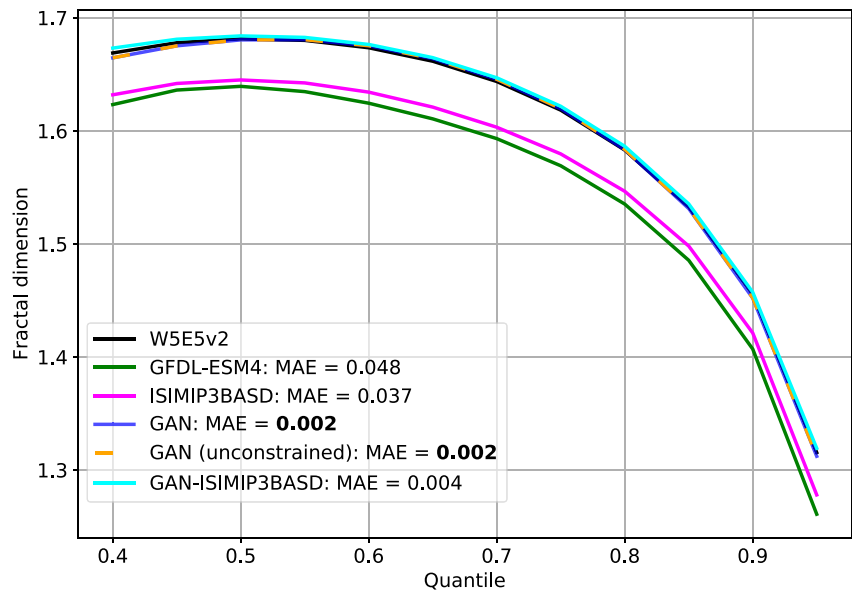


**Figure 5.** The number of precipitation events above the 95th percentile W5E5v2 ground truth (a) and the relative Wasserstein-1 distance between the waiting time distributions of the ground truth and the GFDL-ESM4 (b), ISIMIP3BASD (c), the constrained cGAN (d), the unconstrained cGAN (e), and the (constrained) cGAN-ISIMIP3BASD combination (f). The global mean absolute error with respect to the W5E5v2 ground truth is shown in the top right corner of each panel, where the cGAN-ISIMIP3BASD method achieves the lowest error. The hatched area indicates regions that were excluded from the analysis due to insufficient events.

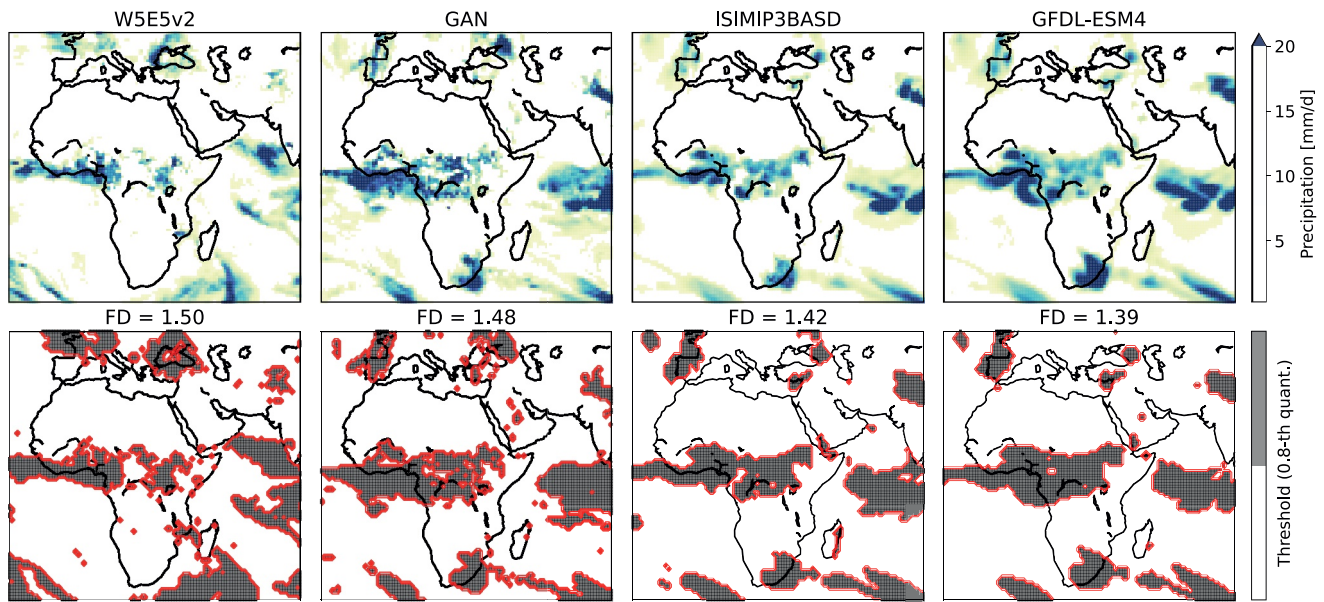


**Figure 6.** Radially averaged power spectral densities (RAPSDs) of the spatial precipitation fields. (a) The RAPSDs are shown as an average over all samples in the test set, for the W5E5v2 ground truth (black) and GFDL-ESM4 (green), ISIMIP3BASD (magenta), cGAN (blue), unconstrained cGAN (orange, dashed), and the (constrained) cGAN-ISIMIP3BASD combination (cyan). (b) The absolute error of the log-transformed spectra w.r.t. the ground truth is shown to highlight the differences. The cGANs and W5E5v2 ground truth agree so closely that they are indistinguishable. In contrast to ISIMIP3BASD, the cGAN can correct the intermittent spectrum accurately over the entire range down to the smallest wavelengths.

For a second way to quantify how realistic the simulated and postprocessed precipitation fields are, with a focus on high-frequency spatial intermittency, we investigate the fractal dimension (Edgar & Edgar, 2008) of the lines separating grid cells with daily rainfall sums above and below a given quantile threshold (see Section 2). For a sample and qualitative comparison of precipitation fields over the African continent, see Figure 8, and additionally for central Asia and South America (Figures S2 and S3 in Supporting Information S1). The daily spatial precipitation fields are first converted to binary images using a quantile threshold. The respective quantiles are determined from the precipitation distribution over the entire test set period and globe. The mean of the fractal dimension computed with box counting (see Section 2; Husain et al., 2021; Lovejoy et al., 1987; Meisel et al., 1992) for each time slice is then investigated (Figure 7). Both the GFDL-ESM4 simulations themselves and the results of applying the ISIMIP3BASD



**Figure 7.** The fractal dimension (see Section 2) of binary global precipitation fields is compared as averages for different quantile thresholds. Results are shown for the W5E5v2 ground truth (black) and GFDL-ESM4 (green), ISIMIP3BASD (magenta), cGAN (blue), unconstrained cGAN (orange, dashed), and the (constrained) cGAN-ISIMIP3BASD combination (cyan). The cGAN can accurately reproduce the fractal dimension of the W5E5v2 ground truth spatial precipitation fields over all quantile thresholds, clearly outperforming the ISIMIP3BASD baseline.



**Figure 8.** Qualitative comparison of precipitation fields at the same date (31st May 2004). The global fields have been cropped to a region over the African continent to better visualize the small-scale representation of precipitation (see Figures S2 and S3 in Supporting Information S1 for more examples). The top row shows the daily precipitation fields for the W5E5v2 ground truth, cGAN, ISIMIP3BASD, and GFDL-ESM4. The bottom row shows the respective contour lines of binary fields used to compute the fractal dimension (FD) given in the titles.

postprocessing to them exhibit spatial patterns with a lower fractal dimension than the W5E5v2 ground truth, implying too low spatial intermittency. In contrast, the cGAN translates spatial fields simulated by GFDL-ESM4 in a way that results in closely matching fractal dimensions over the entire range of quantiles.

#### 4. Discussion

Postprocessing climate projections is a fundamentally different task from postprocessing weather forecast simulations (Hess et al., 2022). In the latter case, data-driven postprocessing methods, for example, based on deep learning, minimize differences between paired samples of variables such as spatial precipitation fields (Hess & Boers, 2022). Beyond time scales of a few days, however, the chaotic nature of the atmosphere leads to exponentially diverging trajectories. For climate or ESM output, there is no observation-based ground truth to directly compare to. We, therefore, frame the postprocessing of ESM projections with applications for subsequent impact modeling in mind as an image-to-image translation task with unpaired samples. We focus on precipitation in this study since it is highly relevant for impact assessment in the context of anthropogenic climate change, and because it is arguably the most difficult variable to model due to the high degree of intermittency both in time and in space. However, our method can be extended to multiple variables in a straightforward manner by including them as additional channels in the image.

To this end, we apply a recently developed postprocessing method based on physically constrained GANs to global simulations of a state-of-the-art, high-resolution ESM from the CMIP6 model ensemble, namely the GFDL-ESM4 (Krasting et al., 2018; O'Neill et al., 2016). We evaluate our method against the gold-standard bias correction framework ISIMIP3BASD. Our model can be trained on unpaired samples that are characteristic of climate simulations. It is able to correct the ESM simulations in two regards: temporal distributions over long time scales, including extremes in the distributions' tails, as well as spatial patterns of individual global snapshots of the model output. The latter is not possible with established methods. Our GAN-based approach is designed as a general framework that can readily apply to different ESMs and observational target data sets. This contrasts existing bias-adjustment methods that are often tailored to specific applications.

We chose to correct precipitation because it is arguably one of the hardest variables to represent accurately in ESMs. So far, GANs have only been applied to regional studies or low-resolution global ESMs (François et al., 2021; Hess et al., 2022). The GFDL-ESM4 model simulations are hence chosen in order to test if our

cGAN approach would lead to improvements even when postprocessing global high-resolution simulations of one of the most complex and sophisticated ESMs to date. In the same spirit, we evaluate our approach against a very strong baseline given by the state-of-the-art bias correction framework ISIMIP3BASD, which is based on a trend-preserving QM method (Lange, 2019).

Comparing long-term summary statistics, our method yields histograms of relative precipitation frequencies that closely agree with corresponding histograms from reanalysis data (Figure 1). This means that the extremes in the far end of the tail are accurately captured, with similar skill to the ISIMIP3BASD baseline that is mainly designed for this task. Differences in the grid cell-wise long-term average show that the cGAN skillfully reduces biases (Figure 2); in particular, the often reported double-peaked ITCZ bias of the GFDL-ESM4 simulations, which is a common feature of most climate models (Tian & Dong, 2020), is strongly reduced (Figure 3). The ISIMIP3BASD method—since it is specifically designed for this—produces slightly lower biases for grid-cell-wise averages than the cGAN; we show that combining both methods by first applying the cGAN and then the ISIMIP3BASD procedure leads to the overall best performance. Evaluating the representation of extremes in the bias-adjusted simulations shows that the cGAN outperforms the other models in terms of capturing the magnitude of return values for events with a 10-year return time. We note that applying the ISIMIP3BASD method on cGAN output decreases the performance for 10-year return values. We are able to improve waiting time distributions of precipitation events above the 95th percentile, with the combined cGAN-ISIMIP3BASD method performing better than the other methods. The fact that the combined cGAN-ISIMIP3BASD approach performs better in the latter case could be attributed to the larger set of events available for estimating an accurate mapping of the quantiles. The estimation might be more challenging for the much rarer extremes that only occur once in 10 years on average, given the 20 years of training data.

Regarding the correction of spatial patterns of the modeled precipitation fields, we compare the spectral density and fractal dimensions of the spatial precipitation fields. Our results show that, indeed, only the cGAN can capture the characteristic spatial intermittency of precipitation closely (Figures 6 and 7). We believe that the measure of fractal dimension is also relevant for other fields, such as nowcasting and medium-range weather forecasting, where blurriness in deep learning-based predictions is often reported (Ravuri et al., 2021) and needs to be further quantified.

Postprocessing methods for climate projections have to be able to preserve the trends that result from the nonstationary dynamics of the Earth system on long-time scales. We have therefore introduced the architecture constraint of preserving the global precipitation amount every day in the climate model output (Hess et al., 2022). We find that this does not affect the quality of the spatial patterns that are produced by our GAN method. However, the skill of correcting mean error biases is slightly reduced by the constraint. This can be expected in part as the constraint is constructed to follow the global mean of the ESM (see Figure S3 in Supporting Information S1). Hence, biases in the global ESM mean can influence the constrained cGAN. This also motivates our choice to demonstrate the combination of the constrained cGAN with the QM-based ISIMIP3BASD procedure since it can be applied to future climate scenarios, making it more suitable for actual applications than the unconstrained architecture. By construction, the ISIMIP3BASD method is trend preserving; hence, combining it with the constrained cGAN also produces realistic trends (see Figure S4 in Supporting Information S1). In our experiments, we also applied QM before the cGAN processing but did not see any notable improvements compared to the cGAN alone.

There are several directions to further develop or approach. The architecture employed here has been built for equally spaced two-dimensional images. Extending the GAN architecture to perform convolutions on the spherical surface, for example, using graph neural networks, might lead to more efficient and accurate models. Moreover, GANs are comparably difficult to train, which could make it challenging to identify suitable network architectures. Using large ensembles of climate simulations could provide additional training data that could further improve the performance. Another straightforward extension of our method would be the inclusion of further input variables or the prediction of additional high-impact physical variables, such as near-surface temperatures that are also important for regional impact models.

### Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

## Data Availability Statement

The Python code for processing and analyzing the data, together with the PyTorch Lightning (Falcon, 2019) code, is available in GAN training code [Software] (2023). The ISIMIP3BASD code in Lange (2022) is used for this study. The W5E5 data from WFDE5 over land merged with ERA5 over the ocean (W5E5 v2.0) [Dataset] (2021) are used and the GFDL-ESM4 data are available in Cmpip6 gfdl-esm4 model data [Dataset] (n.d.).

## Acknowledgments

NB and PH acknowledge funding by the Volkswagen Foundation, as well as the European Regional Development Fund (ERDF), the German Federal Ministry of Education and Research and the Land Brandenburg for supporting this project by providing resources on the high-performance computer system at the Potsdam Institute for Climate Impact Research. NB and CS acknowledge funding by the European Union's Horizon 2020 Research and Innovation Programme under Grant 820970. NB acknowledges funding under the Marie Skłodowska-Curie Grant 956170, as well as from the Federal Ministry of Education and Research under Grant 01LS2001A. SL acknowledges funding from the European Union's Horizon 2022 Research and Innovation Programme under Grant 101081193 Optimal High Resolution Earth System Models for Exploring Future Climate Changes (OptimESM). Open Access funding enabled and organized by Projekt DEAL.

## References

- Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Van Dijk, A. I., et al. (2019). MSWEP V2 global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bulletin of the American Meteorological Society*, 100(3), 473–500. <https://doi.org/10.1175/BAMS-D-17-0138.1>
- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619(7970), 533–538. <https://doi.org/10.1038/s41586-023-06185-3>
- Boyle, J., & Klein, S. A. (2010). Impact of horizontal resolution on climate model forecasts of tropical precipitation and diabatic heating for the TWP-ICE period. *Journal of Geophysical Research*, 115, D23113. <https://doi.org/10.1029/2010JD014262>
- Cannon, A. J. (2018). Multivariate quantile mapping bias correction: An  $N$ -dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dynamics*, 50(1), 31–49. <https://doi.org/10.1007/s00382-017-3580-6>
- Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*, 28(17), 6938–6959. <https://doi.org/10.1175/JCLI-D-14-00754.1>
- Cmpip6 gfdl-esm4 model data. (n.d.). Cmpip6 gfdl-esm4 model data [Dataset]. Retrieved from <https://esgf-node.llnl.gov/projects/cmpip6/>
- de Haan, L., & Ferreira, A. (2006). *Extreme value theory*. Springer New York. <https://doi.org/10.1007/0-387-34471-3>
- Déqué, M. (2007). Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. *Global and Planetary Change*, 57(1–2), 16–26. <https://doi.org/10.1016/j.gloplacha.2006.11.030>
- Edgar, G. A., & Edgar, G. A. (2008). *Measure, topology, and fractal geometry* (Vol. 2). Springer.
- Falcon, W. (2019). *PyTorch Lightning*. GitHub Repository. Retrieved from <https://github.com/PyTorchLightning/pytorch-lightning>
- François, B., Thao, S., & Vrac, M. (2021). Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks. *Climate Dynamics*, 57(11), 3323–3353. <https://doi.org/10.1007/s00382-021-05869-8>
- François, B., Vrac, M., Cannon, A. J., Robin, Y., & Allard, D. (2020). Multivariate bias corrections of climate simulations: Which benefits for which losses? *Earth System Dynamics*, 11(2), 537–562. <https://doi.org/10.5194/esd-11-537-2020>
- Frieler, K., Lange, S., Piontek, F., Reyer, C. P., Schewe, J., Warszawski, L., et al. (2017). Assessing the impacts of 1.5°C global warming—Simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b). *Geoscientific Model Development*, 10(12), 4321–4345. <https://doi.org/10.5194/gmd-10-4321-2017>
- GAN training code [Software]. (2023). Zenodo Repository. Retrieved from <https://doi.org/10.5281/zenodo.8203912>
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (Vol. 27).
- Grönquist, P., Yao, C., Ben-Nun, T., Dryden, N., Dueben, P., Li, S., & Hoefler, T. (2021). Deep learning for post-processing ensemble weather forecasts. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194), 20200092. <https://doi.org/10.1098/rsta.2020.0092>
- Gudmundsson, L., Bremnes, J. B., Haugen, J. E., & Engen-Skaugen, T. (2012). Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations—A comparison of methods. *Hydrology and Earth System Sciences*, 16(9), 3383–3390. <https://doi.org/10.5194/hess-16-3383-2012>
- Harris, D., Foufoula-Georgiou, E., Droegemeier, K. K., & Levit, J. J. (2001). Multiscale statistical properties of a high-resolution precipitation forecast. *Journal of Hydrometeorology*, 2(4), 406–418. [https://doi.org/10.1175/1525-7541\(2001\)002<0406:MSPOAH>2.0.CO;2](https://doi.org/10.1175/1525-7541(2001)002<0406:MSPOAH>2.0.CO;2)
- Harris, L., McRae, A. T., Chantry, M., Dueben, P. D., & Palmer, T. N. (2022). A generative deep learning approach to stochastic downscaling of precipitation forecasts. arXiv preprint arXiv:2204.02028.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- Hess, P., & Boers, N. (2022). Deep learning for improving numerical weather prediction of heavy rainfall. *Journal of Advances in Modeling Earth Systems*, 14, e2021MS002765. <https://doi.org/10.1029/2021MS002765>
- Hess, P., Drüke, M., Petri, S., Strnad, F. M., & Boers, N. (2022). Physically constrained generative adversarial networks for improving precipitation fields from Earth system models. *Nature Machine Intelligence*, 4(10), 828–839. <https://doi.org/10.1038/s42256-022-00540-1>
- Huffman, G. J., Adler, R. F., Arkin, P., Chang, A., Ferraro, R., Gruber, A., et al. (1997). The Global Precipitation Climatology Project (GPCP) combined precipitation dataset. *Bulletin of the American Meteorological Society*, 78(1), 5–20. [https://doi.org/10.1175/1520-0477\(1997\)078<0005:TGPCPG>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<0005:TGPCPG>2.0.CO;2)
- Husain, A., Reddy, J., Bisht, D., & Sajid, M. (2021). Fractal dimension of coastline of Australia. *Scientific Reports*, 11(1), 6304. <https://doi.org/10.1038/s41598-021-85405-0>
- IPCC. (2021). In V. Masson-Delmotte, et al. (Eds.), *Climate change 2021: The physical science basis. Contribution of Working Group I to the sixth assessment report of the intergovernmental Panel on Climate Change*. Cambridge University Press. Retrieved from <https://www.ipcc.ch/report/sixth-assessment-report-working-group-i/>
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Kotz, M., Levermann, A., & Wenz, L. (2022). The effect of rainfall changes on economic production. *Nature*, 601(7892), 223–227. <https://doi.org/10.1038/s41586-021-04283-8>
- Krasting, J. P., John, J. G., Blanton, C., McHugh, C., Nikonov, S., Radhakrishnan, A., et al. (2018). *NOAA-GFDL GFDL-ESM4 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. <https://doi.org/10.22033/ESGF/CMIP6.1407>
- Lange, S. (2019). Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0). *Geoscientific Model Development*, 12(7), 3055–3070. <https://doi.org/10.5194/gmd-12-3055-2019>
- Lange, S. (2022). ISIMIP3BASD [Software]. Zenodo. <https://doi.org/10.5281/zenodo.6758997>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lorenz, E. N. (1996). Predictability: A problem partly solved. In *Proc. seminar on predictability* (Vol. 1).



- Lovejoy, S., Schertzer, D., & Tsonis, A. (1987). Functional box-counting and multiple elliptical dimensions in rain. *Science*, 235(4792), 1036–1038. <https://doi.org/10.1126/science.235.4792.1036>
- Meisel, L., Johnson, M., & Cote, P. (1992). Box-counting multifractal analysis. *Physical Review A*, 45(10), 6989–6996. <https://doi.org/10.1103/physreva.45.6989>
- O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 9(9), 3461–3482. <https://doi.org/10.5194/gmd-9-3461-2016>
- Price, I., & Rasp, S. (2022). Increasing the accuracy and resolution of precipitation forecasts using deep generative models. In *International Conference on Artificial Intelligence and Statistics* (pp. 10555–10571).
- Rasp, S., & Thuery, N. (2021). Data-driven medium-range weather prediction with a Resnet pretrained on climate simulations: A new model for WeatherBench. *Journal of Advances in Modeling Earth Systems*, 13, e2020MS002405. <https://doi.org/10.1029/2020MS002405>
- Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., et al. (2021). Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878), 672–677. <https://doi.org/10.1038/s41586-021-03854-z>
- Ruzanski, E., & Chandrasekar, V. (2011). Scale filtering for improved nowcasting performance in a high-resolution X-band radar network. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6), 2296–2307. <https://doi.org/10.1109/tgrs.2010.2103946>
- Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., & Ziese, M. (2011). GPCP full data reanalysis version 6.0 at 0.5°: Monthly land-surface precipitation from rain-gauges built on GTS-based and historic data. *GPCC Data Report*, 10, 585.
- Tian, B., & Dong, X. (2020). The double-ITCZ bias in CMIP3, CMIP5, and CMIP6 models based on annual mean precipitation. *Geophysical Research Letters*, 47, e2020GL087232. <https://doi.org/10.1029/2020GL087232>
- Tong, Y., Gao, X., Han, Z., Xu, Y., Xu, Y., & Giorgi, F. (2021). Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods. *Climate Dynamics*, 57(5), 1425–1443. <https://doi.org/10.1007/s00382-020-05447-4>
- Traxl, D., Boers, N., Rheinwalt, A., & Bookhagen, B. (2021). The role of cyclonic activity in tropical temperature-rainfall scaling. *Nature Communications*, 12(1), 6732. <https://doi.org/10.1038/s41467-021-27111-z>
- Vrac, M. (2018). Multivariate bias adjustment of high-dimensional climate simulations: The Rank Resampling for Distributions and Dependencies ( $R^2D^2$ ) bias correction. *Hydrology and Earth System Sciences*, 22(6), 3175–3196. <https://doi.org/10.5194/hess-22-3175-2018>
- Wang, F., & Tian, D. (2022). On deep learning-based bias correction and downscaling of multiple climate models simulations. *Climate Dynamics*, 59(11–12), 3451–3468. <https://doi.org/10.1007/s00382-022-06277-2>
- Wang, F., Tian, D., & Carroll, M. (2023). Customized deep learning for precipitation bias correction and downscaling. *Geoscientific Model Development*, 16(2), 535–556. <https://doi.org/10.5194/gmd-16-535-2023>
- Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe, J. (2014). The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): Project framework. *Proceedings of the National Academy of Sciences of the United States of America*, 111(9), 3228–3232. <https://doi.org/10.1073/pnas.1312330110>
- Weyn, J. A., Durran, D. R., & Caruana, R. (2020). Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *Journal of Advances in Modeling Earth Systems*, 12, e2020MS002109. <https://doi.org/10.1029/2020MS002109>
- WFDE5 over land merged with ERA5 over the ocean (W5E5 v2.0) [Dataset]. (2021). ISIMIP Repository. Retrieved from <https://doi.org/10.48364/ISIMIP.342217>
- Wilcox, E. M., & Donner, L. J. (2007). The frequency of extreme rain events in satellite rain-rate estimates and an atmospheric general circulation model. *Journal of Climate*, 20(1), 53–69. <https://doi.org/10.1175/jcli3987.1>
- Zhang, Y., Long, M., Chen, K., Xing, L., Jin, R., Jordan, M. I., & Wang, J. (2023). Skilful nowcasting of extreme precipitation with NowcastNet. *Nature*, 619, 526–532. <https://doi.org/10.1038/s41586-023-06184-4>
- Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2223–2232).