

## Global spatio-temporal ERA5 precipitation downscaling to km and sub-hourly scale using generative AI

Luca Glawion, Julius Polz, Harald Kunstmann, Benjamin Fersch, Christian Chwala

### Angaben zur Veröffentlichung / Publication details:

Glawion, Luca, Julius Polz, Harald Kunstmann, Benjamin Fersch, and Christian Chwala. 2025. "Global spatio-temporal ERA5 precipitation downscaling to km and sub-hourly scale using generative AI." *npj Climate and Atmospheric Science* 8 (1): 219.  
<https://doi.org/10.1038/s41612-025-01103-y>.

<https://doi.org/10.1038/s41612-025-01103-y>

# Global spatio-temporal ERA5 precipitation downscaling to km and sub-hourly scale using generative AI

Check for updates

Luca Glawion<sup>1</sup>✉, Julius Polz<sup>1,2</sup>, Harald Kunstmann<sup>1,3</sup>, Benjamin Fersch<sup>1</sup> & Christian Chwala<sup>1</sup>

The spatial and temporal distribution of precipitation significantly impacts human lives. While reanalysis datasets provide consistent long-term global precipitation information that allows investigations of rainfall-driven hazards like larger-scale flooding, they lack the resolution to capture the high spatio-temporal variability of precipitation and miss intense local rainfall events. Here, we introduce SpateGAN-ERA5, the first deep learning-based spatio-temporal downscaling of precipitation data on a global scale. SpateGAN-ERA5 enhances ERA5 precipitation data from 24 km and 1 h to 2 km and 10 min, delivering high-resolution rainfall fields with realistic spatio-temporal patterns and accurate rain rate distribution, including extremes. Its computational efficiency enables the generation of a large ensemble of solutions, addressing uncertainties inherent to downscaling challenges and supports practical applicability for generating high-resolution precipitation data for arbitrary ERA5 time periods and regions on demand. Trained solely on data from Germany and validated in the US and Australia, considering diverse climates, including tropical rainfall regimes, SpateGAN-ERA5 demonstrates strong generalization, indicating robust global applicability. It fulfills critical needs for high-resolution precipitation data in hydrological and meteorological research.

Variations in precipitation critically influence society and ecosystems, affecting water resources, agriculture, and flood risks<sup>1–3</sup>. Climate change has already amplified precipitation variability, leading to more frequent and severe weather events<sup>4</sup>. Understanding and mitigating the impacts of precipitation extremes requires accurate historical records in a spatial and temporal resolution that captures the high variability of rainfall<sup>5–7</sup>. Observation-based rainfall products can only partially fulfill this requirement. Station networks have long records, but are not dense enough in most parts of the world<sup>8</sup> and thus lack spatial representativeness. In contrast, satellite rainfall products provide homogeneous spatial coverage but only have limited temporal coverage. In addition, they suffer from considerable errors due to their complex rainfall retrieval methods and exhibit spatial and temporal inhomogeneities<sup>9–11</sup>.

Assimilation of historical meteorological observations in first-principle-based physical simulations enables modeling of consistent, comprehensive, and long records of atmospheric conditions<sup>12</sup>. In the last decade, such reanalyses have accelerated scientific research in hydrological modeling<sup>13,14</sup>, flood prediction<sup>15</sup>, calculation of climate change-related costs<sup>16,17</sup>, or training data-driven weather forecasting models<sup>18–21</sup>.

However, existing global reanalyses still have significant limitations. The heterogeneous density of assimilated observations and the low spatio-temporal model resolution lead to uncertainties and biases<sup>12,22</sup>. In particular, the complex spatio-temporal structure of rainfall cannot be represented by the resolution of current reanalysis products, which leads to a significant underestimation of extreme values, which are crucial for impact analysis of severe weather events<sup>23–27</sup>. Running higher-resolution global reanalyses is currently not feasible due to the immense computational demand<sup>28–30</sup>.

Downscaling can be used to increase the spatial and temporal resolution of coarse-resolution global models, either dynamically, that is running a local-area high-resolution model, or by statistical post-processing. While dynamical downscaling is again limited by computational resources, statistical methods are computationally efficient and can be applied globally. However, traditional statistical approaches are not capable of generating realistic high-resolution rainfall fields with correct spatio-temporal patterns<sup>31</sup> and extreme values. Recently, advanced downscaling approaches leveraging deep neural networks have proven to be capable of this task. Successful applications have been shown for spatial and spatio-temporal

<sup>1</sup>Institute of Meteorology and Climate Research - Atmospheric Environmental Research (IMK-IFU), Campus Alpin, Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany. <sup>2</sup>Institute of Meteorology and Climate Research - Atmospheric Trace Gases and Remote Sensing (IMK-ASF), Karlsruhe Institute of Technology, Karlsruhe, Germany. <sup>3</sup>Institute of Geography, University of Augsburg, Augsburg, Germany. ✉e-mail: [luca.glawion@kit.edu](mailto:luca.glawion@kit.edu)

super-resolution<sup>32–36</sup>, and regional spatial downscaling<sup>37–41</sup>. Nevertheless, a skillful global sub-hourly, km-scale downscaling of precipitation data has remained a challenging problem.

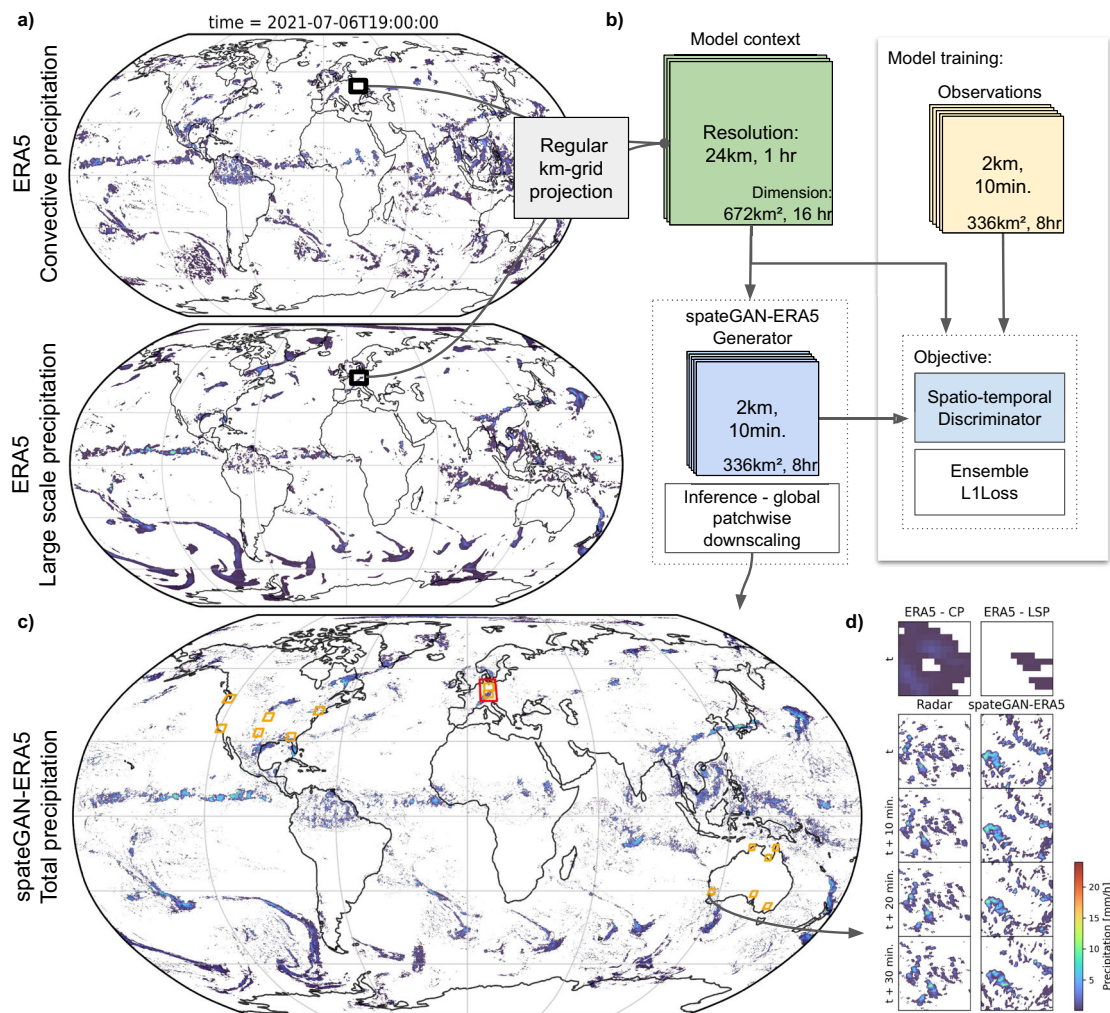
Here, we present spateGAN-ERA5, a conditional generative adversarial network for robust deep learning-based spatio-temporal downscaling of ERA5 precipitation data. Our model transforms hourly, 24 km (~0.25°) resolved ERA5 precipitation estimates into rainfields that resemble weather radar observations at a resolution of 10 min and 2 km. SpateGAN-ERA5 is trained on high-resolution quantitative precipitation estimates (QPE) from a gauge-adjusted and climatology-corrected weather radar product in Germany and is evaluated across three climatically diverse regions on the globe. The model generalizes well outside the training domain and enables computationally efficient global rainfall downscaling to a resolution that is fine enough to capture the spatio-temporal complexity of rainfall, especially for rainfall events with convective cells. It generates realistic extreme value distributions, spatial structures, and advection patterns, all in a well-calibrated ensemble that addresses the underdetermined nature of the downscaling problem. Thus, spateGAN-ERA5 significantly advances downscaling methodologies and opens up a wide field of possible scientific investigations in a variety of domains like hydrology, risk analysis, or agriculture.

## Results

### Generative spatio-temporal downscaling of global ERA5 precipitation

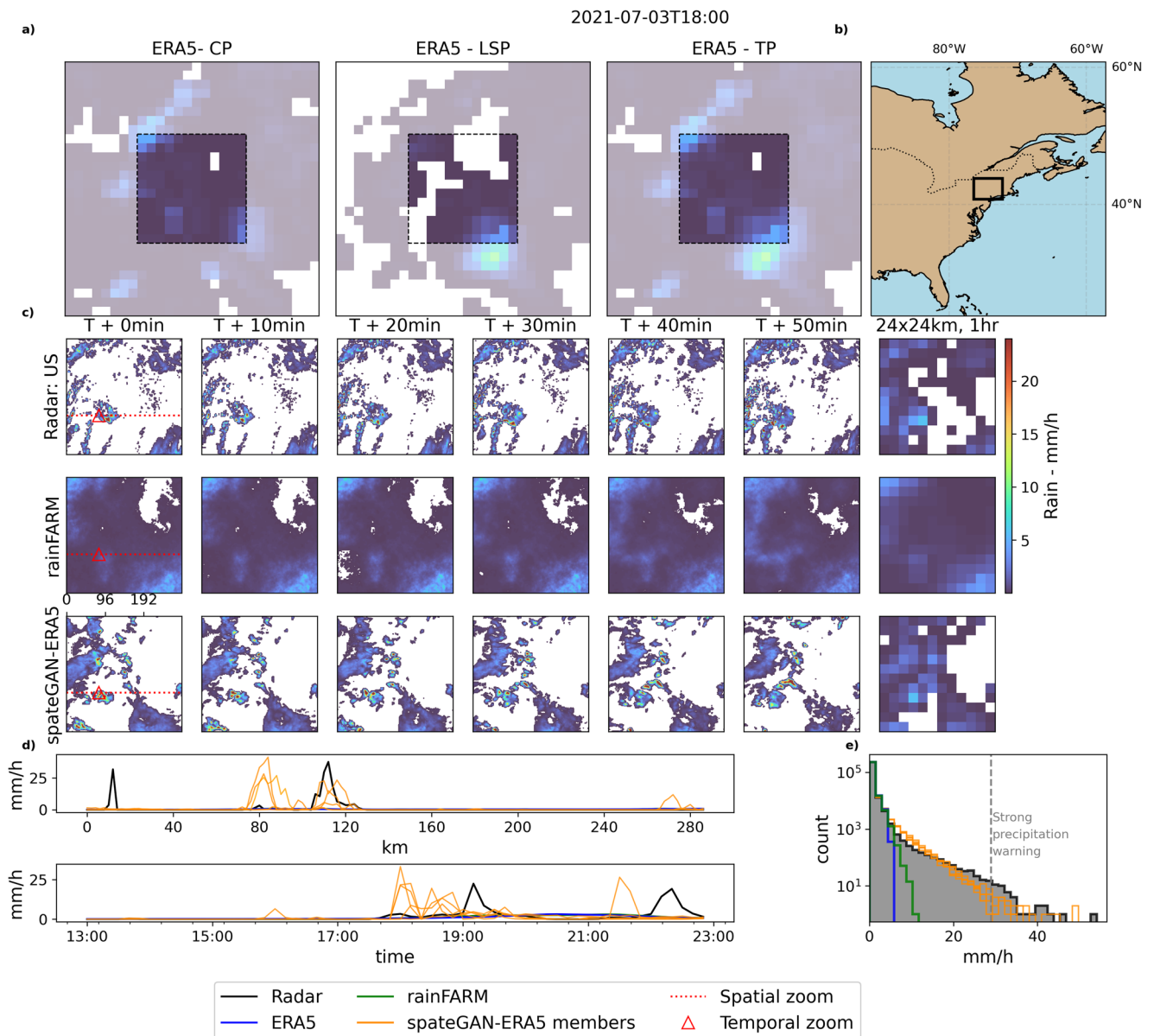
For global downscaling of ERA5 precipitation data we use a conditional generative adversarial network (cGAN) with ERA5 convective (CP) and large-scale precipitation (LSP) as the coarse condition and gauge-adjusted weather radar data as the high-resolution reference (see Fig. 1b). The downscaling of hourly ERA5 precipitation fields with a spatial resolution of 24 km is performed by a generator model producing a field with a 12-times higher spatial and a 6-times higher temporal resolution. Specifically, the generator processes CP and LSP input patches with a size of 28 by 28 grid cells and 16 time steps. To provide more contextual information, the input is four times the domain size of the actual downscaled area (see Fig. 2a).

A main feature of GANs is the custom learnable loss function (the discriminator). In our model, this enables the generation of realistic fields that fulfill a wide range of statistical and structural criteria for precipitation. The applied neural network architecture extends the spateGAN model established for a weather radar video-super-resolution approach<sup>32</sup> and is described in the “Model description” section. The model is trained using high-quality gauge-adjusted weather radar data provided by the



**Fig. 1 | Model and evaluation area overview for spatio-temporal downscaling of global ERA5 precipitation estimates.** spateGAN-ERA5 can transform coarse-resolution ERA5 rainfields shown in (a) into high-resolution rainfields (2 km, 10 min.) as they would be observed by a gauge-adjusted radar product regarding their spatial structures and rain rate distribution. b Schematic of the model operating on patchwise downscaling of km gridded convective and large-scale ERA5 precipitation variables in a probabilistic manner. c Global downscaling predictions

enabled by patch stitching provide continuous rainfields (full resolution map shown in ancillary files Fig. A1). From the area marked by the red box, patches are drawn and used for model training. Downscaling performance is evaluated using radar observations as a comparison from the regions marked by the orange boxes. d Detailed highlight shows the resolved resolution in time and space, and the comparison with the Australian weather radar observations.



**Fig. 2 | Case study of performance on a challenging precipitation event starting on 03.07.21 in the US with observed convective cells.** **a** Model input patches consisting of larger ERA5 data, i.e., the convective and large-scale precipitation contribution to the total precipitation sum. **b** Location of the radar observation. **c** Observations, spateGAN-ERA5 predictions, and rainFARM downscaling for the target domain in 10-min increments from  $t$  to  $t + 50$  min. and as a coarsened version

approximating ERA5 resolution. **d** 1-D cutouts showing spateGAN-ERA5 ensemble members for a specific pixel along the temporal dimension (top panel) and a horizontal cross-section for one time step (bottom panel). **e** Distribution for temporally aggregated data with 2 km and 1 h resolution, including maps shown in (c) and the previous and following 6 h. A severe precipitation warning threshold of the German Weather Service is set at 25 mm/h.

German Meteorological Service (DWD) from the years 2009–2020<sup>42</sup>. Details on the adversarial training procedure are given in the “Training and model selection” section. The model is efficient, fast, and small enough to run on a single NVIDIA-Tesla-V100 GPU by downscaling one patch in 0.04 s in inference mode. Data-parallel training on 4 A100 80 GB GPUs took 3 days Table 1.

Global fields are produced by stitching overlapping high-resolution patches (see the “Data preparation” section). We evaluate the downscaling skill by comparison to weather radar data from the year 2021 in three different countries (Germany, USA, Australia) that cover a wide range of climatic conditions (see Fig. 1c). Performance is compared to the stochastic rainfall downscaling method rainFARM, which is based on the interpolation of the power spectrum to smaller scales<sup>43,44</sup>, and to trilinear interpolation as a simple baseline method (see the “Reference methods” section).

### Case study

We select a variety of meteorologically interesting events (Fig. 2 and Supplementary Figs. 5–9) to showcase the spatio-temporal downscaling performance of spateGAN-ERA5 and how this overcomes the inherent limitations of ERA5 precipitation data.

Here, we focus on the event in Fig. 2 showing convective cells in the United States as observed by the MRMS dataset, which are at a scale known not to be resolvable by ERA5<sup>45</sup>. Even when compared to coarsened radar observations at ERA5 resolution, ERA5 shows a too-low variance (see Fig. 2e) with an underestimation of extreme values (see Supplementary Information Section 1.5). Being able to reconstruct such small-scale rainfall cells is of particular interest to improve ERA5 precipitation estimates in regions and seasons with a high amount of convective precipitation, such as the tropics and extratropics<sup>22</sup>.

**Table 1 | Overview of used rainfall observation datasets**

	RADKLIM-YW	MRMS	Rainfields 3	
Country	Germany	US	Australia	Australia - tropics
Training period	2009–2020	–	–	–
Model selection	01–06.2021	–	–	–
Evaluation	07–12.2021	07–12.2021	07–12.2021	01–03.2021
Rain gauge adjusted the radar	✓	–	✓	✓
Source	42,74	75,76	77	

We consider the first week of each month for model selection and evaluation.

SpateGAN-ERA5 is able to reconstruct convective rainfall fields with small-scale structures and plausible rain rates, including heavy local rainfall. The rain cells show temporal continuity, hardly allowing for a qualitative differentiation between observed and predicted rainfields (see video V1 in ancillary files). Predicted rainfall may occur at a misplaced spatial or temporal position, but with a magnitude similar to the associated radar observation (see Fig. 2d). This misplacement is not solely due to the underdetermined nature of the downscaling problem but also reflects differences between ERA5 and radar data on a coarser scale. The probabilistic nature of spateGAN-ERA5 accounts for such uncertainties, but is also constrained by the contextual information provided by ERA5. For example, the predicted ensemble shows greater variability in intensity than in spatial or temporal localization.

RainFARM fails to reconstruct small-scale convective cells, overestimates the spatial extent of rainfall, and underestimates extremes. By design, rainFARM mostly coincides with ERA5 at the coarse resolution, limiting spatio-temporal disaggregation. This leads to only slightly more granular rainfall fields than using simple interpolation techniques.

**Skillful representation of extreme values**

To get a more complete picture of the extreme value statistics of spateGAN-ERA5, we analyze data from different climatic regions in the US, Germany, and Australia (see the “Evaluation” section), including severe tropical rainfall events in Australia, highlighted in Supplementary Information Section 1.4.

The fractions skill score (FSS) (Fig. 3a) shows that only for the smallest rain rate threshold and up to a spatial scale of 16 km the interpolated ERA5 and rainFARM rainfields have a higher location accuracy than a single ensemble member of spateGAN-ERA5. Considering an increased spatial scale or ensemble of predictions, the generative model consistently outperforms the other methods across all rain rate thresholds. For intense rainfall larger than 5 mm/h, spateGAN-ERA5 is the only model with acceptable skill. The relative improvement in terms of  $\Delta$ mFSS when considering spateGAN-ERA5 as a downscaling technique instead of an interpolated ERA5 is highest for Australia, the dataset where interpolation has the lowest absolute mFSS. This is followed by Germany, the training region, evaluated over an out-of-sample time period, and the US (see Supplementary Table 1). For the tropic dataset,  $\Delta$ mFSS is slightly below the US, however, the overall skill is highest. This indicates a strong ability of the model to generalize well outside its training domain.

The distributions shown in Fig. 3b further support spateGAN-ERA5’s capability in predicting plausible extreme values. Predictions generally follow the reference’s lognormal distribution for Australia, the tropics (see Supplementary Fig. 10b), and Germany, which is physically reasonable<sup>46,47</sup>. Different characteristics in the US are physically more implausible and thus likely due to systematic errors in the non-gauge-adjusted MRMS radar data. Overall, spateGAN-ERA5 underestimates the frequency of strong precipitation for Australia and the US and overestimates it for Germany. Since spateGAN-ERA5 follows the average precipitation amount of ERA5 (see the “Model description” section), this is in agreement with the biases of the individual evaluation datasets as shown in Supplementary Table 1.

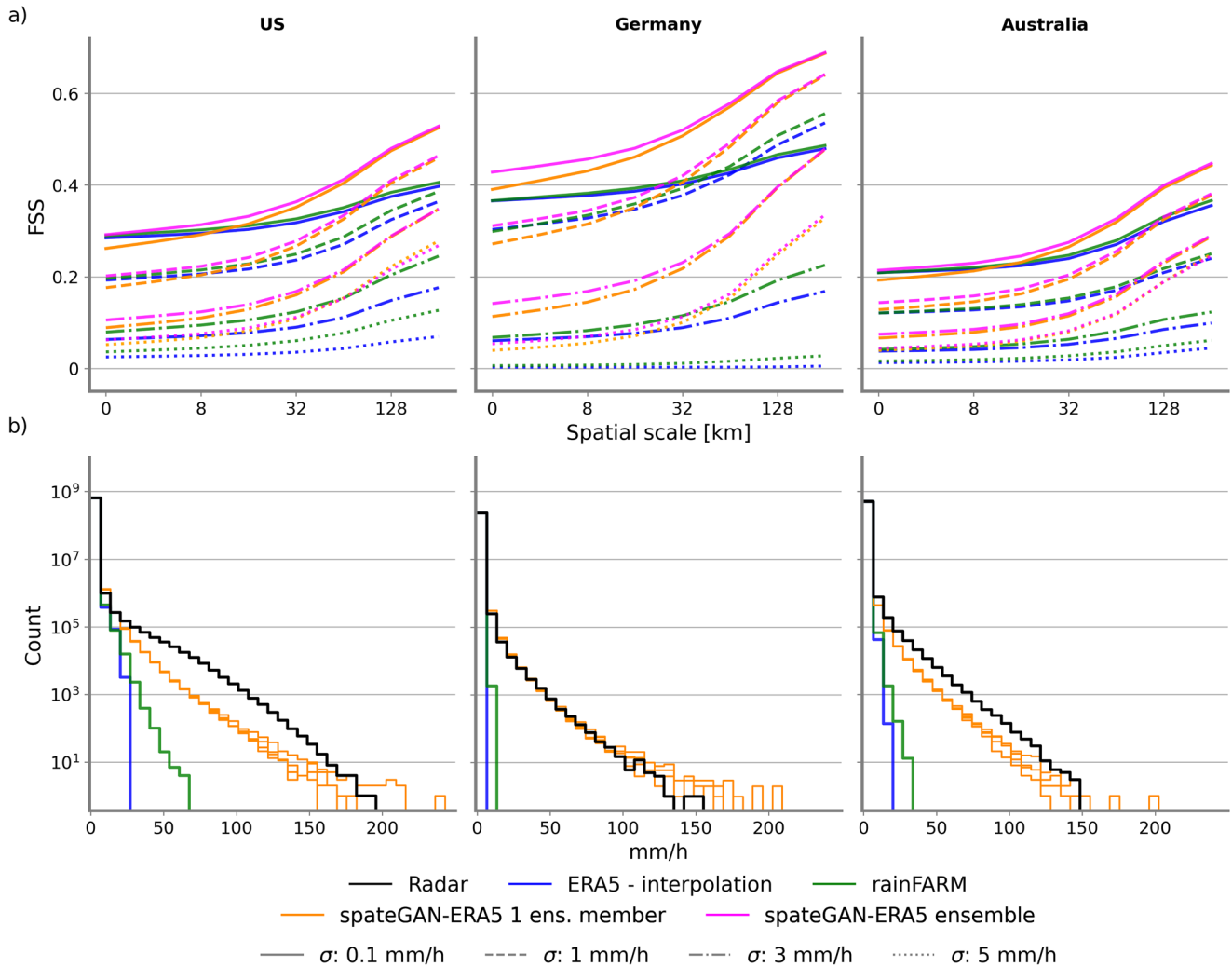
In terms of a pixel-wise deterministic skill (MAE and RMSE), ERA5 interpolation and rainFARM show the best results (Supplementary Table 1). However, this is mainly due to their tendency to produce smoother rainfields with dampened extreme values, avoiding a double penalty for misplaced small-scale events. Since we aim for sharp probabilistic estimates, we use these scores with caution. The superior CRPS shows that spateGAN-ERA5 predictions have the highest ensemble skill. The ensemble quality, important for a correct representation of extremes, is analyzed by rank histograms (Supplementary Fig. 1). It shows a well-calibrated ensemble with a slight under-dispersive tendency for spateGAN-ERA5 and an unfavorable heavy under-dispersive tendency for rainFARM.

**Spatial plausibility of highly resolved rainfall fields**

Spatial and temporal patterns of rainfall are the tangible result of the physical processes that drive precipitation formation and evolution in the atmosphere<sup>48,49</sup>. Accurately reconstructing these patterns presents a considerable challenge, especially when using data-driven models, which lack a priori knowledge of the underlying atmospheric physics<sup>50,51</sup>. These models must learn to reproduce sharp gradients, coherent advection structures, and multi-scale variability from limited, coarsely-resolved, and potentially biased training data. We consider weather radar observations as a sufficient reference to allow for the statistical analysis of such spatio-temporal patterns. The qualitative assessment of the “Generative spatio-temporal downscaling of global ERA5 precipitation” section suggests that spateGAN-ERA5 predictions are hardly distinguishable from real radar observations, while ERA5 interpolation produces blurry rainfields. Visually, rainFARM only slightly improves over the interpolation. To quantify this observation, we chose radial averaged power spectral density (RAPSD). As a measure of anisotropy, a key aspect of specific spatial patterns often caused by horizontal advection<sup>49</sup>, we define the linear eccentricity in terms of spatial autocorrelation in the “Evaluation” section.

This analysis uses a subset of each evaluation dataset, described in the “Data preparation” section, focusing on cases with greater consistency between ERA5 and radar observations. For the RAPSD (shown in Fig. 4a), spateGAN-ERA5 largely replicates the power spectrum of the radar observations in Germany, with slight deviations at the smallest wavelengths close to the target resolution. In the US, Australia, and the tropics (see Supplementary Fig. 10c), an underestimation of all wavelengths is apparent. These discrepancies in RAPSD can be traced back to the mean field biases of ERA5, which are stronger for more extreme events<sup>22</sup>, and by design, not corrected by spateGAN-ERA5. When focusing solely on spatial characteristics and disregarding a multiplicative bias, the normalized RAPSD shows an almost perfect alignment between predictions and observations for all datasets.

ERA5 interpolation produces overly smoothed rainfields, resulting in a considerably lower RAPSD and normalized RAPSD for shorter wavelengths. RainFARM slightly improves the power spectrum, increasing the amplitude for wavelengths between the ERA5 resolution of 24 km up to the final 2 km resolution. However, the method introduces a physically unrealistic jump in the power spectrum at 24 km<sup>47,52</sup>. The temporal power spectrum density shows a similar behavior of all methods for the temporal dimension (Supplementary Fig. 2).



**Fig. 3 | Investigation of the downscaling distribution reconstruction skill for the evaluation datasets in Germany, the US, and Australia in 2021. a** Fractions skill score (FSS) for thresholds 0.1, 1, 3, and 5 mm/h and a temporal scale of 1 h. We

report the slightly improved probabilistic ensemble FSS for rainFARM. **b** Distribution comparison showing multiple spateGAN-ERA5 ensemble members.

Linear eccentricity is analyzed in Fig. 4b and illustrated for a single field in Supplementary Fig. 3. The spateGAN-ERA5 distribution of the score is close to the observations while rainFARM stays similar to the ERA5 interpolation, providing rainfields that are highly autocorrelated for a large spatial lag. SpateGAN-ERA5 produces small-scale features that resemble the radar observations in terms of size, orientation, and eccentricity (see Supplementary Fig. 4).

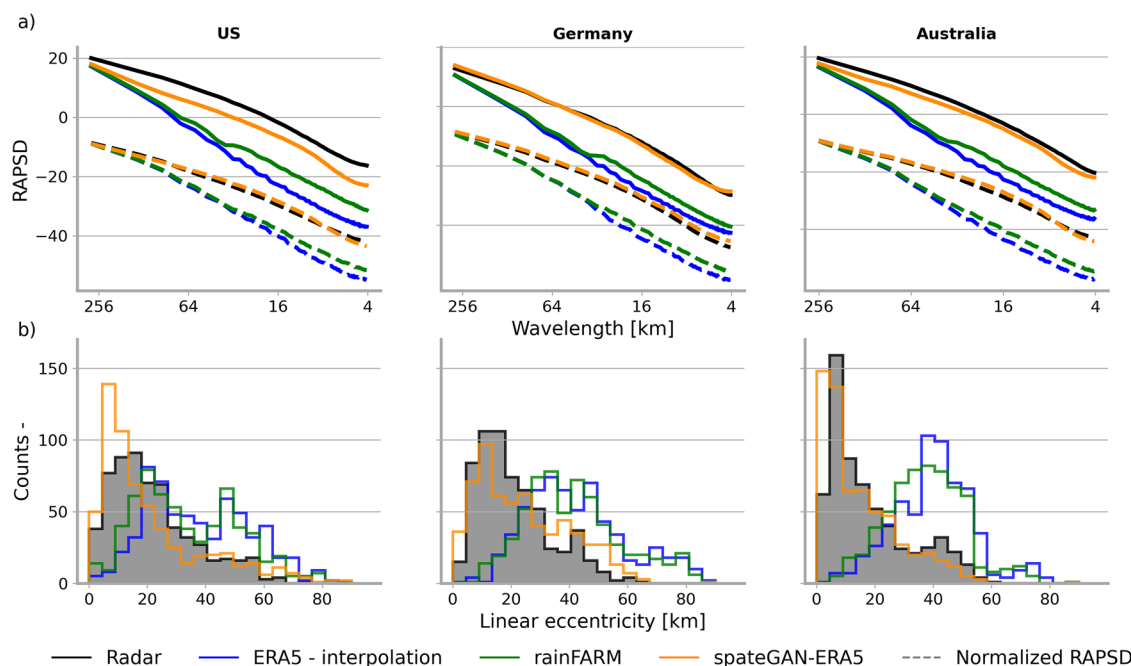
### Discussion

A critical issue in atmospheric sciences is the extent to which deep learning models trained for a specific region can generalize to other regions. Additionally, discrepancies between modeled and observed data distributions persist, particularly in free-running climate simulations, which can become entirely decoupled from observations beyond a certain lead time. In the context of downscaling, a widely adopted approach involves training super-resolution models that do not rely on perfectly matched input-output pairs. However, if the synthetically coarsened training data deviate significantly from the actual climate model output distribution, this mismatch can lead to a degradation in model performance during inference<sup>37</sup>.

Our own analysis highlights the significant discrepancy between coarsened radar observations and ERA5 precipitation, both in terms of extreme value distributions and spatio-temporal structures. This mismatch stems, partly, from the limited convective parameterization schemes in

numerical modeling<sup>22</sup>. We show that a carefully designed training sampling scheme, which can be described as training on loosely paired images, results in a high downscaling performance. This involves selecting ERA5 model input samples that closely match their corresponding observation targets, and by choosing Germany as the primary training region, showing a relatively high agreement between reanalysis data and targets. While tested on reanalysis data, this idea of training a model is generic and can, e.g., be applied to train on loosely paired images from nudged climate simulations and observation data, thereby facilitating the downscaling of traditional climate model scenarios, which then can be downscaled in inference.

To evaluate the generalization capabilities of spateGAN-ERA5, we tested its performance on spatial domains and time periods not included in the training data. This is particularly relevant given that high-quality meteorological observations with fine spatial and temporal resolution are only sparsely available. Training a downscaling model to represent the high variability of precipitation on a global scale is, therefore, inherently challenging using observations alone. Our findings indicate that spateGAN-ERA5 exhibits robust performance even outside the training region, demonstrating its ability to reconstruct precipitation fields in climatologically distinct environments. In particular, the model also performs well in tropical regions such as northern Australia, where high-intensity rainfall is dominated by convective processes that differ fundamentally from the precipitation dynamics of mid-latitude regions like Germany. In some cases,



**Fig. 4 | Spatial characteristic scores for a subset of the evaluation datasets in Germany, the US, and Australia in 2021 (see description in the “Data preparation” section). a** Mean radially averaged power spectral density (RAPSD) and mean

normalized radially averaged power spectral density (dashed line). **b** Distribution of linear eccentricity of the 2D autocorrelation representation (0.5 Pearson correlation coefficient ellipse).

it is even exceeding its performance within the training domain, depending on the evaluation metric. This suggests a strong generalization capacity and can therefore be used on an extended scale, providing a global precipitation product with improved rainfall distribution characteristics. By leveraging the full historical record of the ERA5 reanalysis dataset, extending back to 1940, spateGAN-ERA5 is able to provide high-resolution precipitation reconstructions for an unprecedented time record, which is a significant advancement over conventional precipitation datasets.

To evaluate the quality of the downscaled precipitation fields, we consider a variety of spatial structures and pixel-wise scores and conduct an event-based analysis to reflect the diverse variability and characteristics of rainfall. Given that precipitation is inherently difficult to model due to its high variability and intermittency, we could show spateGAN-ERA5’s ability to disaggregate and reconstruct the statistical properties of rainfields across temporal and spatial scales, with plausible extreme values that are completely missing within the initial low-resolution input data. To generate realistic rain events from this data, a mere extrapolation of the spatial power spectrum of ERA5 as performed by rainFARM proved to be insufficient for the given problem. SpateGAN-ERA5, as a generative model, shows high structural similarities between its predictions and the reference datasets, as can be shown by evaluating the RAPSD, temporal PSD, and linear eccentricity in the “Spatial plausibility of highly resolved rainfall fields” section. Furthermore, the probabilistic, yet computationally efficient, method explicitly accounts for downscaling-related uncertainties when refining precipitation fields in space and time. The model architecture and training methodology are designed to be adaptable, making it applicable to other precipitation datasets and resolutions, thereby serving as a versatile tool for various scientific and operational applications.

As shown in the “Case study” section and Fig. 2e, spateGAN-ERA5 is the only presented method that is able to reconstruct a distribution similar to the observations, with predictions of larger rainfall intensities in the severe weather warning range. This demonstrates its potential for enabling more accurate hydrological modeling, particularly in flood risk assessments, where detailed precipitation fields are essential for simulating extreme rainfall events and their impacts. The ability to generate high-resolution precipitation maps several orders of magnitude faster

than traditional dynamical downscaling methods addresses critical needs in meteorological and hydrological research. In the context of climate impact studies, spateGAN-ERA5 facilitates improved assessments of long-term precipitation trends and variability, helping to refine projections of extreme events under different climate scenarios. Its ability to reconstruct convective rainfall events, which are often missing in traditional climate model outputs, makes it particularly useful for assessing localized hazards, such as flash floods, and informing disaster risk management strategies.

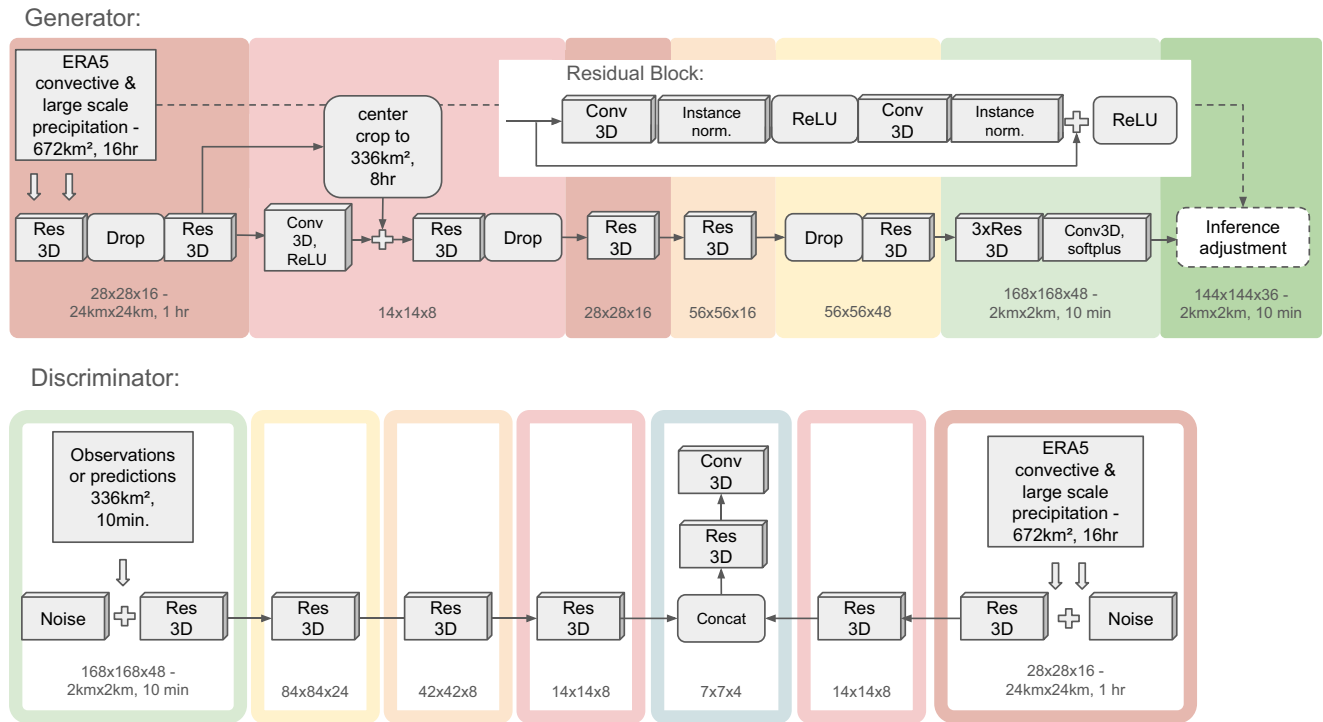
### Methods

SpateGAN-ERA5 performs spatio-temporal downscaling of ERA5 precipitation estimates, increasing the resolution from 24 km and 1 h to 2 km and 10 min. The model receives input patches of the ERA5 variables convective and large-scale precipitation of size 16 h × 672 km × 672 km and performs the downscaling for a centered domain of 8 h × 336 km × 336 km. We trained the model in Germany, where a consistently high-resolution and high-quality reference dataset is available through the gauge-adjusted and climatology-corrected radar product RADKLIM-YW provided by the German Meteorological Service, and where a high agreement between ERA5 precipitation and observation data can be shown (see Fig. 3)<sup>22</sup>. Global downscaling is achieved by downscaling and stitching overlapping patches.

### Model description

We build on the successful precipitation video-super-resolution approach, spateGAN<sup>32</sup>, consisting of a generator, trained in an adversarial manner with a discriminator model. The main ERA5 downscaling generator model (see Fig. 5) comprises four consecutive components that make use of 3D-convolutional residual blocks (Res3D) to capture spatio-temporal dependencies.

First, the ERA5 convective and large-scale precipitation input data are processed on their initial resolution. Second, it passes a UNET-like down-sampling and skip connection with an added cropping operation. This allows the model to process data at multiple resolutions, consider global and local features, and focus on the target domain at an early model stage. Third, the spatial and temporal resolution of the input data is successively



**Fig. 5 | Schematic overview of the spateGAN-ERA5 generator and discriminator model architecture using Residual Blocks (Res3D).** During inference, boundary regions are removed, and an ERA5 mean field bias adjustment is applied.

increased, and the structures of the rainfields are refined by 4 upsampling blocks, including bilinear and linear interpolation and Res3D blocks. Finally, three subsequent Res3D blocks adjust fine-scale structures and limit the prediction range to positive values using a Softplus activation function.

Temporally constant dropout ( $p = 0.2$ ) at three different generator depths introduces scale and rain event-dependent perturbation at low, mid, and high frequencies and enables spatio-temporally continuous probabilistic downscaling. The perturbation in combination with the ensemble loss supports the model in reconstructing the missing tail of the ERA5 precipitation distribution.

During inference mode, i.e., for evaluation and global prediction, we apply three additional operations. First, we freeze the dropout seed for each produced ensemble member, which improves the spatio-temporal consistency of the rainfields compared to a random perturbation in space and time. Second, we cut the outermost edges (24 km and 1 h) to remove boundary effects. For global predictions, this routine differs slightly, as described in the “Data preparation” section. Third, we apply a patchwise mean field bias correction to the predictions<sup>53</sup>, by multiplying the average predicted rainfall by a single value to match the average rainfall amount of the associated ERA5 input patch. This ensures that the provided ERA5 precipitation amount is preserved and that the model can be applied in regions where the ERA5 bias strongly deviates from the training distribution.

The overall design of the generator is memory efficient and can be run in inference on smaller GPUs (10 GB per sample). This allows the application of our model by a broad research community, not only those with access to the latest-generation GPUs with large memory.

The discriminator (see Fig. 5) is trained simultaneously with the generator. Its inputs are the temporal sequences of high-resolution prediction or observation, as well as the coarse-resolution context provided to the generator. Its training objective is to decide if the high-resolution field is real or artificially generated. The loss function is binary cross-entropy. Within the model, the high-resolution and low-resolution data are treated separately, and as a first step, Gaussian noise (mean = 1, std. 0.05) is added to the input data to prevent the model from learning to distinguish rainfields based on

quantization characteristics. A series of Res3D blocks then processes the data and extracts spatio-temporal features. The coarse and high-resolution inputs are concatenated at a late stage to encourage a comparison based on latent features extracted on multiple resolutions. The discriminator model is thereby used as a powerful dynamical loss function for the generator, which learns to discriminate structure- and distribution-related rainfall characteristics.

**Model details**

For downsampling operations, the skip connection of the Res3D blocks includes a 3D-convolutional layer with a kernel size of 1 and instance normalization to harmonize the dimensions. All remaining convolutional layers in the networks use a kernel size of 3.

The generator uses 3D reflection padding in all layers with 3D convolution, and the discriminator uses zero padding. Except for the first, second, and last Res3D Block of the generator, the first Res3D block of the high-resolution discriminator path, and the two Res3D blocks of the low-resolution discriminator path, we apply instances of normalized convolutions. For the generator, we use a feature dimension of 96. For the discriminator, the high-resolution features are 128, 128, 128, and 64, and the low-resolution features are 64, 32. After concatenating, the final Res3D Block decreases the features to 64, which are compressed to 1 within the last 3D-convolutional layer.

The specific model architecture stems from an iterative optimization process that started during our investigations for precipitation video-super-resolution in ref. 32 and was further developed for the task of ERA5 precipitation downscaling. Thereby, we also tried, e.g., state-of-the-art vision transformer network layers as a generator, which did not result in a performance improvement and led us to stick to the well-proven 3D Residual layers. In general, an extensive hyperparameter optimization is desirable. Due to the computational complexity, long training runs, and limited computational resources, we could not test all possible parameter combinations and therefore cannot state that spateGAN-ERA5 provides the best possible results. We would rather invite the research community to build on our work and further improve the downscaling of precipitation data.

**Objective function**

As an objective function, we use a well-known stepwise adversarial training strategy<sup>54,55</sup>.

The discriminator D receives the ERA5 context X and target observations Y or predictions  $\hat{Y}$  of the generator and is trained to minimize the binary cross-entropy loss

$$\mathcal{L}_D = -\mathbb{E}_{X,Y}[\log D(X, Y)] - \mathbb{E}_{X,\hat{Y}}[\log(1 - D(X, \hat{Y}))] \quad (1)$$

The generator loss includes an adversarial loss

$$\mathcal{L}_{GAN}(G) = -\mathbb{E}_X[\log D(X, G(X))] \quad (2)$$

and an ensemble L1-loss defined as

$$\mathcal{L}_{L1}(G) = \left| Y - \frac{1}{3} \sum_{i=1}^3 \hat{Y}_i \right|, \quad (3)$$

which compares high-resolution targets to the ensemble mean prediction of 3 members  $\hat{Y}_1, \hat{Y}_2, \hat{Y}_3$ . This ensures that the predictions remain close to the ground truth while reducing the double penalty of small convective cells or heavy precipitation misplaced during training.

The total generator loss is

$$\mathcal{L}_G = \mathcal{L}_{GAN}(G) + \mathcal{L}_{L1}(G) \quad (4)$$

**Training and model selection**

The model is trained for  $2 \times 10^5$  adversarial training steps. The learning rate is  $1 \times 10^{-4}$  for the generator and  $2 \times 10^{-4}$  for the discriminator and uses AdamW optimizer<sup>56</sup> with  $\beta_1 = 0.0$  and  $\beta_2 = 0.999$  (Discriminator:  $\beta_1 = 0.0$  and  $\beta_2 = 0.5$ ). We employ data-parallel training on 3 Nvidia A100 GPUs with 80 GB of memory each for 4 days. The batch size is set to 9 per training step. In inference mode, downscaling 1 patch takes 0.04 s on one A100 GPU.

We save all model weights after every 250 training steps and identify the best generator training state by downscaling and evaluating the independent model selection dataset 4.7. We select the final model by calculating the average of the ensemble FSS (meFSS) of the thresholds 0.1, 1, 3, 5, and 8 mm/h, spatial scales 1, 4, 8, 16, 32, 64, and 128 km, and temporal scale of 1 h. This considers the ensemble quality and location accuracy for different categories of rainfall intensities, independent of the heavily skewed distribution of rainfall.

**Evaluation**

For evaluation, we verify the performance of the downscaling methods using a set of quantitative scores since no single metric is capable of capturing the complexity of highly resolved rainfields. We calculate the *root mean square error (RMSE)* is a pixel-wise error computed for a single predicted ensemble member:

$$RMSE = \sqrt{(Y - \hat{Y}_i)^2} \quad (5)$$

The *continuous ranked probability score (CRPS)*<sup>57</sup> measures the prediction accuracy by accounting for the ensemble spread and bias. The Cumulative Density Function (CDF) of the predicted ensemble at a specific point and time step ( $\hat{F}(xt)$ ) is compared to the observed rainfall y.

$$CRPS(\hat{F}, y) = \int_{-\infty}^{\infty} (\hat{F}(xt) - 1(xt \geq y))^2 dx \quad (6)$$

$$1(xt \geq y) \mapsto \begin{cases} 0 & : \quad xt < y \\ 1 & : \quad xt \geq y \end{cases} \quad (7)$$

We report the CRPS as the average CRPS for each dataset. For deterministic methods (ensemble size of 1), i.e., for interpolated ERA5, this score reduces to the *mean absolute error (MAE)*.

$$MAE = |(Y - \hat{Y}_i)| \quad (8)$$

The *fractions skill score (FSS)*<sup>58,59</sup> is defined as

$$FSS = 1 - \frac{(f_{\hat{Y}} - f_Y)^2}{f_{\hat{Y}}^2 + f_Y^2}, \quad (9)$$

where  $f_Y$  (resp.  $f_{\hat{Y}}$ ) is the fraction of pixels within a spatial and temporal ( $s, t$ ) neighborhood that exceed a certain observed (resp. predicted) rainfall intensity threshold ( $\sigma$ ). The averaging is performed over the respective neighborhoods of all locations and time steps of each evaluation dataset. For the ensemble FSS, the fraction of ensemble members exceeding ( $\sigma$ ) is considered.

We calculate the mean FSS (mFSS) or mean ensemble FSS (meFSS) of a set of different scales ( $s = 0, 4, 8, 16, 32, 64, 128, 256$  km,  $t = 1$  h) and thresholds ( $\sigma = 0.1, 1, 3, 5$  mm/h). The  $\Delta$ mFSS is the relative deviation of the meFSS of rainFARM and spateGAN-ERA5 to the mFSS of interpolated ERA5, expressed as a percentage, and illustrates the performance benefits when considering an alternative downscaling method instead of pure interpolation. For data on ERA5 resolution, the mFSS considers spatial scales of 0, 24, 96, and 192 km and rain thresholds of 0.1, 1, 3, and 5 mm/h.

The *radially averaged power spectral density (RAPSD)* and *power spectral density (PSD)*<sup>60,61</sup> measure how power is distributed across spatial and temporal frequencies. The temporal PSD acts thereby as an indicator for plausible advection. The RAPSD is calculated for single images using the PySTEPS<sup>62</sup> implementation and is averaged for each evaluation dataset. The PSD is calculated along the temporal dimension for each pixel and for each week of the evaluation datasets and is afterwards averaged for each dataset. Additionally, we report the normalized RAPSD and PSD, where the power spectrum of each image or time sequence is normalized so that it sums to one.

We use *rank histograms*<sup>63,64</sup> to validate the variability and reliability of an ensemble of probabilistic rainfall predictions. For each pixel and time step of the evaluation datasets, 100 ensemble predictions are considered in increasing order, and the normalized rank  $r$  of the actual observation value is determined. Perfectly calibrated ensembles show a uniformly distributed  $r$ , where predictions and observations stem from the same distribution.

We investigate the spatial anisotropy of rainfields by calculating the autocorrelation of single images of observations and predictions for spatial lags from 0 to 60 km in x and y direction<sup>65</sup>. We estimate an ellipse from the 0.5 Pearson Correlation Coefficient (PCC) counterline for each individual autocorrelation field and retrieve the variables' length of major axis  $a$  and length of minor axis  $b$  to determine the *linear eccentricity*

$$ecc_1 = \sqrt{a^2 - b^2}, \quad (10)$$

*eccentricity*

$$ecc = \sqrt{1 - \frac{b^2}{a^2}}, \quad (11)$$

*and size*

$$size = \sqrt{a * b}. \quad (12)$$

Furthermore, we compute the *orientation* of the ellipse, i.e., of the major axis, in degrees.

We define the *BIAS* as

$$BIAS = \frac{\bar{Y} - \bar{X}}{\bar{Y}} \quad (13)$$

where  $\bar{X}$  is the average predicted precipitation amount of each evaluation region and  $\bar{Y}$  is the average observed rainfall.

During evaluation, spateGAN-ERA5 downscales patches that overlap in the temporal dimension to generate a continuous sequence of temporally consistent rainfields, by keeping the central 2 h of each patch. For the case study videos, a linear blending approach is applied to 1 h overlapping periods, with weights decaying from 1 to 0, effectively smoothing out minor remaining temporal discontinuities in the predictions.

In total, the probabilistic model performance is evaluated using 100 ensemble members for calculating rank histograms and CRPS shown in Supplementary Fig. 1 and Supplementary Table 1. For the ensemble FSS and meFSS, we calculate only 6 members since the score converges at a small ensemble size. For the presented evaluation, rain rates smaller than 0.01 of all compared datasets are set to zero.

## Datasets

The model input and, therefore, the only dataset required for applying spateGAN-ERA5 are the convective and large-scale variables from the ERA5 reanalysis. The model is trained using gauge-adjusted and climatology-corrected radar data in Germany. We use two additional radar datasets for evaluation from the United States and Australia to test the model's ability for generalization outside of its training distribution. Even if it seems obvious at first to include data from the US and Australia for model training, we have deliberately refrained from doing so. Pure radar observations can be highly error-prone and do not match the quality of a sophisticated, gauge-adjusted, and climatologically corrected product such as RADKLIM-YW. Due to the lack of high-resolution data availability, we use radar observations to get an indication of spateGAN-ERA5's generalization capabilities.

**ERA5 dataset.** The ERA5 reanalysis provides global, hourly model data spanning the past 70 years<sup>12,66</sup>. It integrates observational data with numerical model predictions through advanced data assimilation techniques, resulting in a high-quality benchmark dataset. For precipitation, the ERA5 4D-var system assimilates hourly NCEP stage IV gauge-adjusted weather radar precipitation information over the US<sup>67,68</sup>. In this study, we used the years 2009–2021, where ERA5 aligns with the available radar data. We utilize the variables convective and large-scale precipitation of hourly ERA5 data as input for spatio-temporal downscaling. Including additional variables as input, such as wind components, temperature, pressure level, etc., did not enhance overall performance in the presented setup.

We do not use finer resolved ERA5-land precipitation estimates, since they lack valuable scale-related information<sup>29,69,70</sup>, exclude oceans and coastal areas, and have a higher release latency<sup>71</sup>.

Despite the known limitations of ERA5 precipitation estimates, which include spatially heterogeneous quality, biases<sup>12,22</sup>, a tendency to smooth out local extremes due to the coarse resolution of 0.25° and 1 h<sup>72</sup>, and limitations in modeling convective events<sup>45,73</sup>, the product is most commonly used in environmental research.

**RADKLIM-YW Germany.** For training, model selection, and part of the validation of spateGAN-ERA5, we use the gauge-adjusted and climatology-corrected weather radar product RADKLIM-YW provided by the German Meteorological Service (DWD) as target data<sup>42,74</sup>.

This product is a composite of precipitation information from a network of 16 C-band weather radars. It is adjusted by approximately 1000 rain gauges that are homogeneously distributed in Germany with a density of one gauge per 330 km<sup>2</sup>. In addition to the RADOLAN gauge adjustment, effects like range-dependent underestimation and beam blockage are covered by an additional climatological correction.

The grid extent is 900 km × 1100 km in polar stereographic projection, covering almost the entire Germany and its surrounding border regions, with a resolution of 1 km × 1 km and a temporal resolution of 5 min. Each grid cell represents a 5 min. rainfall sum with a quantization of 0.01 mm. Regions not covered by the 150 km measurement radii of the radars or missing measured values are marked with “NaNs.” For our investigation, we used data on the provided km grid, coarsened to 2 km and 10 min. resolution. We use the years 2009–2020 for model training, the first half of the year 2021 for model selection, and the second half for evaluation, preventing data leakage and testing for generalization abilities. For evaluation, we select two fixed locations of the size 336 km × 336 km, highlighted in Fig. 1, covering almost the entire country.

**Multi-Radar Multi-Sensor System (MRMS) United States.** For validation purposes, we use the radar composite from the Multi-Radar Multi-Sensor (MRMS) system comprising 146 WSR-88D radars covering the US and 30 Canadian radars<sup>75,76</sup>. Climatic conditions in the United States have a high variability, ranging from continental, subtropical, and Mediterranean to tropical.

The MRMS dataset we use covers the time period from July to December 2021 and is not gauge-adjusted. Alternative gauge-adjusted QPE products are not available at a sub-hourly resolution and, therefore, are not suitable for most parts of our analysis.

MRMS covers the region from 20° to 55° latitude North and 130° to 60° longitude West with a resolution of 0.01° in both latitude and longitude directions. The temporal resolution is 2 min. We select 6 regions exhibiting a high radar quality and covering different climatic regions of the country (see Fig. 1, yellow boxes). For evaluation, we regrid the radar observations of each 6 locations to their associated regular km UTM projection and downsample them to 2 km and 10 min. resolution. Each location has a domain size of 336 km × 336 km.

**Australian Radar Network.** We additionally use quantitative precipitation estimates from the Australian operational radar network<sup>77</sup>.

We select data from 6 different C-band weather radars, covering subtropical regions across the country, for the period from July to December 2021 (see Fig. 1). The individual locations have a radar coverage of 150 km and are selected by considering less beam blockage, data availability, and homogeneous distribution. 3 of these radar sites operate Doppler radars. The QPE is gauge-adjusted but strongly depends on the availability of the heterogeneously distributed rain gauge observations<sup>78</sup>. An increased bias between ERA5 and the Australian radar was visible, and the radar quality may be a larger factor than in Germany (see Supplementary Table 1). The product has a spatial grid resolution of 0.5 km × 0.5 km using an Albers Conical Equal Area projection and a temporal interval between 5, 6, and 10 min. For evaluation, we downsample the observations to 2 km and 10 min. resolution. Due to the smaller radar coverage, each location has a domain size of 280 km × 280 km.

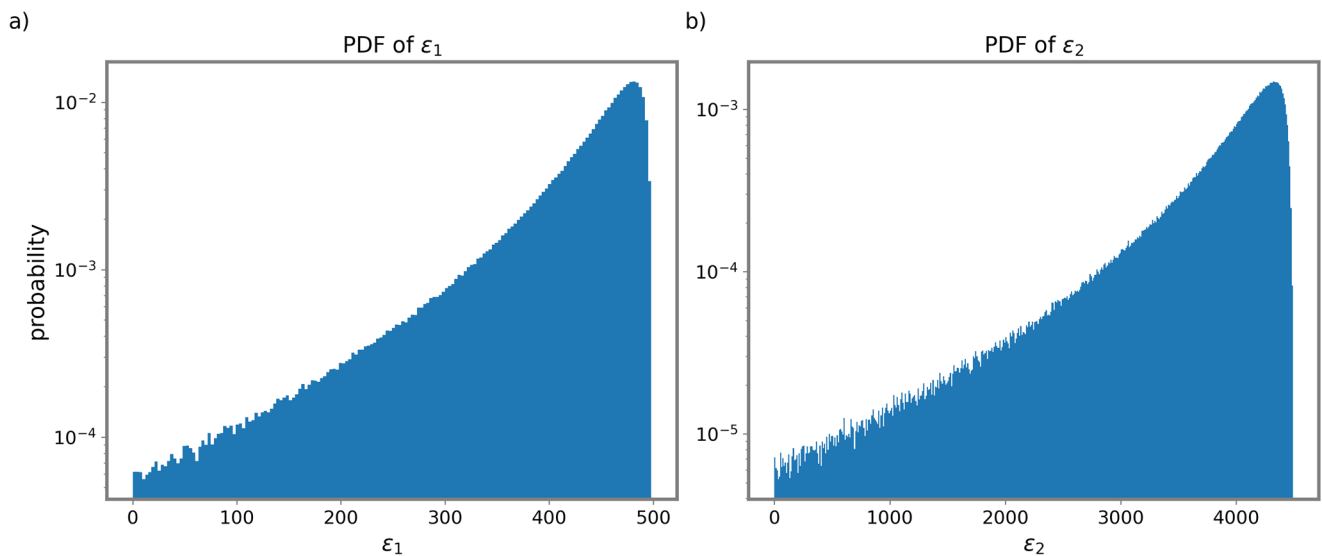
Additionally, we select two of the northernmost radar observation stations, located in Darwin and Weipa, Australia, and use the time period from January to March 2021 to cover tropical rainfall regimes in our separate investigation shown in the Supplementary Information X.

## Data preparation

Observation data and ERA5 precipitation estimates are adjusted to be used for model training, selection, evaluation, and global inference as described below.

**Training and model selection dataset.** For training, we draw random target samples from RADKLIM-YW, each with 48 continuous radar observation time steps and a size of 168 × 168 pixels, i.e., 8 h and 336 km × 336 km. The associated model input is received by first interpolating ERA5 data to the target grid and afterwards downsampling the extracted patches to 24 km and 1 h to approximate the initial resolution.

Since most of the time, little to no rain falls in the training region of Germany, we apply a subsampling routine, selecting only samples with a



**Fig. 6 | Probability density functions (PDF) of the dynamic thresholds used in the subsampling routine. a** PDF of rainfall intensity threshold  $\epsilon_1$  and **b** PDF of total precipitation amount threshold  $\epsilon_2$ .

sufficient amount of wet pixels and total precipitation in both input and target to avoid learning from data that contains little to no rain and fewer wet pixels. For each randomly drawn sample, the following conditions must be fulfilled by the ERA5 input  $X$  and the RADKLIM-YW observation  $y$ :

1.  $X$  and  $y$  do not contain missing values
2. The 66th quantiles of the pixel values in  $X$  and  $y$  exceed  $\epsilon_1$ , where  $\epsilon_1 = | - 50\epsilon' + 500|$  and where  $\epsilon'$  is drawn from *Lognormal* (0, 1).
3.  $\sum_{h,w,t} X > \epsilon_2$  and  $\sum_{h,w,t} y > \epsilon_2$ , where  $\epsilon_2 = | - 450\epsilon + 4500|$  and where  $\epsilon$  is drawn from *Lognormal* (0, 1).

The distribution of the thresholds  $\epsilon_1$  and  $\epsilon_2$  is shown in Fig. 6 and roughly reflects the inverse probability of drawing samples that match the given thresholds. The resulting number of observation samples contained in the training data is about 20,000 (850 GB). During training, we apply standard data augmentation<sup>79</sup> in the form of a rotation (90° or 270°) or reflection (vertical or horizontal) to every alternate sample passed to the model, increasing sample diversity and reducing directional biases, particularly for Germany with dominant westerly wind patterns.

For model selection, we randomly draw additional samples and apply them to the subsampling routine. We select 1000 samples from the temporally independent time period (January–June 2021). We adjust the average rainfall of the targets of this dataset, using a scalar multiplication, so that it matches the average rainfall of the corresponding ERA5 data. This supports the identification of a model state that tends to modify the average precipitation of the ERA5 input samples less drastically and allows the model to be applied outside the training region.

**Evaluation dataset.** SpateGAN-ERA5 is evaluated using a temporally and spatially independent dataset. The evaluation period contains every first week of the months of July–December 2021 and every first week of the months of January–March 2021 for evaluating tropical rainfall events in Australia. The data is sampled using fixed patch locations in the US, Germany, and Australia, highlighted in Fig. 1. For the associated ERA5 samples, the data are projected to the observation grid and afterwards interpolated to 24 km resolution. The domain size is 672 km and includes the previous and following 8 h of the evaluation observation time period.

To analyze the spatial characteristics of the predicted rainfields, i.e., radially averaged power spectral density and anisotropy, we select a subset of each evaluation dataset that exhibits greater consistency between ERA5 and radar observations. This subset includes cases where interpolated ERA5 achieves an mFSS score exceeding 0.2.

In addition to the high-resolution observations and predictions, we evaluate the performance of the individual downscaling methods and datasets on a coarser resolution, approximately that of ERA5 (see Supplementary Information Section 1.5). Therefore, we average the observations and predictions of the evaluation datasets to a spatial resolution of 24 km using 2D average pooling and aggregate the temporal dimension to 1 h resolution.

**Generation of global fields.** We define a processing pipeline for producing seamless global high-resolution precipitation maps from a deep learning model that operates on patchwise downscaling.

First, ERA5 data on its original lat-lon grid is segmented into patches. Each patch covers a regular spatial extent of 672 km × 672 km. We calculate the necessary ERA5 lat-lon coordinates to maintain these patches with the required spatial extent by using the Haversine formula. To simplify the process, the latitude center coordinate of each patch is used to determine the longitudinal extent. Resulting spatial distortions in the longitude directions can be neglected due to the small patch sizes. In comparison to the evaluation and training datasets, where ERA5 is regridded onto a regular kilometer grid using the radar observation projection or UTM projection, this is a more efficient method for global high-resolution mapping. The patches are designed to overlap, such that the target prediction domain of 336 km × 336 km overlaps by approximately 10% in both latitude and longitude directions.

The generated patches are then interpolated onto a regular grid with dimensions of 672 km × 672 km using nearest neighbor interpolation. This data has an approximate resolution of 24 km and enters the SpateGAN-ERA5 model as input data. Downscaling of patches on a km-grid ensures that the model receives data that does not exhibit any latitude-dependent spatial distortion of physical properties. After downscaling, SpateGAN-ERA5 applies a mean field bias adjustment. Due to extensive areas of uncertain, low-intensity rainfall in the ERA5 dataset - particularly over ocean regions - all ERA5 rain rates below 0.1 mm/h are set to zero for this adjustment. The resulting downscaled high-resolution patches are seamlessly interpolated onto a global latitude-longitude grid with a resolution of 0.018°, which corresponds to approximately 2 km at the equator.

To combine the individual overlapping patches, a linear weighting (decaying from 1 to 0 while approaching the border of the patch) is applied in the overlapping regions. This blending process ensures smoother transitions between patches, aiming for continuous large-scale rainfall field circulation (see Supplementary Fig. 13).

## Reference methods

RainFARM is a statistical downscaling approach implemented in the PySTEPS package<sup>62</sup>. It produces small-scale variability by a stochastic process that estimates and extends the spectral slope from each coarse input patch with an estimated scaling factor while preserving key statistical properties. Most importantly, rainFARM produces an isotropic spatial distribution and preserves the rainfall amount when aggregated to the initial resolution.

RainFARM, therefore, serves as a suitable baseline method. Similar to our deep learning approach, it does not rely on additional input data such as atmospheric variables or orography. It was specifically developed for meteorological-scale downscaling, has been successfully applied in various downscaling studies across different contexts<sup>33,37,80</sup> and allows for the generation of multiple ensemble members.

In our study, we apply spatial downscaling of ERA5 total precipitation using the advanced spectral rainFARM algorithm<sup>44</sup>, followed by temporal interpolation. The probabilistic downscaling is conducted using a different fixed random seed for the stochastic component of the method.

For this particular problem, the performance was better than applying the combined spatio-temporal downscaling operation described in ref. 43. Downscaling and aggregating the individual ERA5 variables, convective and large-scale precipitation separately, lead to negligible differences. Unlike spateGAN-ERA5, rainFARM downscales patches of the whole ERA5 input domain of 672 km × 672 km and 16 h, and is afterward cropped to match the domain of the radar observations from the evaluation datasets.

Trilinear interpolation of ERA5 total precipitation, in both space dimensions and the time dimension, serves as a simple baseline where the ERA5 rainfall information can be compared to the high-resolution radar observation without an artificial generation of small-scale features. We interpolate the projected ERA5 data on the coarse km grid described in the “Data preparation” section.

## Data availability

The results and model of this study are produced by publicly available datasets ERA5<sup>12</sup>, RADKLIM-YW<sup>42</sup>, MRMS<sup>75,76</sup>, and the Australian operational radar network<sup>77</sup>. The ERA5 dataset can be downloaded from <https://cds.climate.copernicus.eu/>. The Australian observations can be accessed from <https://thredds.nci.org.au/thredds/catalog/rq0/rainfields3/catalog.html>.

## Code availability

The study was conducted using several open-source frameworks, including PyTorch<sup>81</sup> (<https://pytorch.org/>) and pySTEPS (<https://github.com/pySTEPS/pySTEPS>). Maps were produced using cartopy (<https://scitools.org.uk/cartopy>). The spateGAN-ERA5 model, implemented and optimized in a Python framework, is available at [https://github.com/LGlawion/spateGAN\\_ERA5](https://github.com/LGlawion/spateGAN_ERA5).

Received: 25 February 2025; Accepted: 27 May 2025;

Published online: 15 June 2025

## References

- Gherardi, L. A. & Sala, O. E. Effect of interannual precipitation variability on dryland productivity: a global synthesis. *Glob. Change Biol.* **25**, 269–276 (2019).
- Kotz, M., Levermann, A. & Wenz, L. The effect of rainfall changes on economic production. *Nature* **601**, 223–227 (2022).
- Ray, D. K., Gerber, J. S., MacDonald, G. K. & West, P. C. Climate variation explains a third of global crop yield variability. *Nat. Commun.* **6**, 5989 (2015).
- Zhang, W., Zhou, T. & Wu, P. Anthropogenic amplification of precipitation variability over the past century. *Science* **385**, 427–432 (2024).
- Berne, A., Delrieu, G., Creutin, J.-D. & Obled, C. Temporal and spatial resolution of rainfall measurements required for urban hydrology. *J. Hydrol.* **299**, 166–179 (2004).
- Ochoa-Rodriguez, S. et al. Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: a multi-catchment investigation. *J. Hydrol.* **531**, 389–407 (2015).
- Tamm, O., Saaremaa, E., Rakkema, K., Jaagus, J. & Tamm, T. The intensification of short-duration rainfall extremes due to climate change - need for a frequent update of intensity-duration-frequency curves. *Clim. Serv.* **30**, 100349 (2023).
- Kidd, C. et al. So, how much of the Earth’s surface is covered by rain gauges? *Bull. Am. Meteorol. Soc.* **98**, 69–78 (2017).
- Guo, H. et al. Inter-comparison of high-resolution satellite precipitation products over Central Asia. *Remote Sens.* **7**, 7181–7211 (2015).
- Guo, H. et al. Early assessment of integrated multi-satellite retrievals for global precipitation measurement over China. *Atmos. Res.* **176**–**177**, 121–133 (2016).
- Bai, L., Shi, C., Li, L., Yang, Y. & Wu, J. Accuracy of CHIRPS satellite-rainfall products over mainland China. *Remote Sens.* **10**, 362 (2018).
- Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
- Gebrechorkos, S. H. et al. Global-scale evaluation of precipitation datasets for hydrological modelling. *Hydrol. Earth Syst. Sci.* **28**, 3099–3118 (2024).
- Tarek, M., Brissette, F. P. & Arsenault, R. Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrol. Earth Syst. Sci.* **24**, 2527–2544 (2020).
- Nearing, G. et al. Global prediction of extreme floods in ungauged watersheds. *Nature* **627**, 559–563 (2024).
- Kotz, M., Levermann, A. & Wenz, L. The economic commitment of climate change. *Nature* **628**, 551–557 (2024).
- Lenton, T. M. et al. Quantifying the human cost of global warming. *Nat. Sustain.* **6**, 1237–1247 (2023).
- Lam, R. et al. Learning skillful medium-range global weather forecasting. *Science* **382**, 1416–142 (2023).
- Pathak, J. et al. FourCastNet: a global data-driven high-resolution weather model using adaptive Fourier neural operators. *In Proc. Platform for Advanced Scientific Computing Conference (PASC)* (2023).
- Lang, S. et al. AIFS – ECMWF’s data-driven forecasting system. Preprint at <http://arxiv.org/abs/2406.01465> (2024).
- Kochkov, D. et al. Neural general circulation models for weather and climate. *Nature* <https://www.nature.com/articles/s41586-024-07744-y> (2024).
- Lavers, D. A., Simmons, A., Vamborg, F. & Rodwell, M. J. An evaluation of ERA5 precipitation for climate monitoring. *Q. J. R. Meteorol. Soc.* **148**, 3152–3165 (2022).
- Volosciuk, C., Maraun, D., Semenov, V. A. & Park, W. Extreme precipitation in an atmosphere general circulation model: impact of horizontal and vertical model resolutions. *J. Clim.* **28**, 1184–1205 (2015).
- Seo, D. J. Real-time estimation of rainfall fields using radar rainfall and rain gage data. *J. Hydrol.* **208**, 37–52 (1998).
- Sun, X., Mein, R. G., Keenan, T. D. & Elliott, J. F. Flood estimation using radar and raingauge data. *J. Hydrol.* **239**, 4–18 (2000).
- Aleshina, M. A., Semenov, V. A. & Chernokulsky, A. V. A link between surface air temperature and extreme precipitation over Russia from station and reanalysis data. *Environ. Res. Lett.* **16**, 105004 (2021).
- Ben-Bouallegue, Z. et al. The Rise of Data-Driven Weather Forecasting: A First Statistical Assessment of Machine Learning-Based Weather Forecasts in an Operational-Like Context. *Bull. Am. Meteorol. Soc.* **105**, E864–E883 (2024).
- Tabari, H. et al. Local impact analysis of climate change on precipitation extremes: are high-resolution climate models needed for realistic simulations? *Hydrol. Earth Syst. Sci.* **20**, 3843–3857 (2016).
- Zandler, H., Haag, I. & Samimi, C. Evaluation needs and temporal performance differences of gridded precipitation products in peripheral mountain regions. *Sci. Rep.* **9**, 15118 (2019).

30. Bjarke, N. R., Livneh, B., Barsugli, J. J., Pendergrass, A. G. & Small, E. E. Evaluating large-storm dominance in high-resolution GCMs and observations across the Western Contiguous United States. *Earth's Future* **12**, e2023EF004289 (2024).
31. Bytheway, J. L., Thompson, E. J., Yang, J. & Chen, H. Evaluation of the RainFARM statistical downscaling technique applied to IMERG over global oceans using passive aquatic listener in situ rain measurements. *J. Hydrometeorol.* **24**, 2351–2367 (2023).
32. Glawion, L., Polz, J., Kunstmann, H., Fersch, B. & Chwala, C. spateGAN: spatio-temporal downscaling of rainfall fields using a cGAN approach. *Earth Space Sci.* **10**, e2023EA002906 (2023).
33. Leinonen, J., Nerini, D. & Berne, A. Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. *IEEE Trans. Geosci. Remote Sens.* **59**, 7211–7223 (2021).
34. Serifi, A., Günther, T. & Ban, N. Spatio-temporal downscaling of climate data using convolutional and error-predicting neural networks. *Front. Clim.* **3**, 656479 (2021).
35. Kumar, B. et al. Deep learning-based downscaling of summer monsoon rainfall data over Indian region. *Theor. Appl. Climatol.* **143**, 1145–1156 (2021).
36. Kumar, B. et al. On the modern deep learning approaches for precipitation downscaling. *Earth Sci. Inform.* **16**, 1459–1472 (2023).
37. Harris, L., McRae, A. T. T., Chantry, M., Dueben, P. D. & Palmer, T. N. A generative deep learning approach to stochastic downscaling of precipitation forecasts. *J. Adv. Model. Earth Syst.* **14**, e2022MS003120 (2022).
38. Price, I. & Rasp, S. Increasing the accuracy and resolution of precipitation forecasts using deep generative models. Preprint at <http://arxiv.org/abs/2203.12297> (2022).
39. Harilal, N., Singh, M. & Bhatia, U. Augmented convolutional LSTMs for generation of high-resolution climate change projections. *IEEE Access* **9**, 25208–25218 (2021).
40. Rampal, N., Gibson, P. B., Sherwood, S. & Abramowitz, G. On the extrapolation of generative adversarial networks for downscaling precipitation extremes in warmer climates. *Geophys. Res. Lett.* **51**, e2024GL112492 (2024).
41. Rampal, N., Gibson, P. B., Sherwood, S., Abramowitz, G. & Hobeichi, S. A reliable generative adversarial network approach for climate downscaling and weather generation. *J. Adv. Model. Earth Syst.* **17**, e2024MS004668 (2025).
42. Winterrath, T. et al. Radar climatology (RADKLIM) version 2017.002; gridded precipitation data for Germany: Radar-based quasi gauge-adjusted five-minute precipitation rate (YW) [https://opendata.dwd.de/climate\\_environment/CDC/help/landing\\_pages/doi/landingpage\\_RADKLIM\\_RW\\_V2017.002-en.html](https://opendata.dwd.de/climate_environment/CDC/help/landing_pages/doi/landingpage_RADKLIM_RW_V2017.002-en.html) (2018).
43. Rebora, N., Ferraris, L., Hardenberg, J. V. & Provenzale, A. RainFARM: rainfall downscaling by a filtered autoregressive model. *J. Hydrometeorol.* **7**, 724–738 (2006).
44. D'Onofrio, D., Palazzi, E., Hardenberg, J. V., Provenzale, A. & Calmanti, S. Stochastic rainfall downscaling of climate models. *J. Hydrometeorol.* **15**, 830–843 (2014).
45. Kendon, E. J. et al. Do convection-permitting regional climate models improve projections of future precipitation change? *Bull. Am. Meteorol. Soc.* **98**, 79–93 (2017).
46. Bell, T. L. A space-time stochastic model of rainfall for satellite remote-sensing studies. *J. Geophys. Res. Atmos.* **92**, 9631–9643 (1987).
47. Crane, R. K. Space-time structure of rain rate fields. *J. Geophys. Res. Atmos.* **95**, 2011–2020 (1990).
48. Zick, S. E. & Matyas, C. J. A shape metric methodology for studying the evolving geometries of synoptic-scale precipitation patterns in tropical cyclones. *Ann. Am. Assoc. Geogr.* **106**, 1217–1235 (2016).
49. Schertzer, D. & Lovejoy, S. Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *J. Geophys. Res. Atmos.* **92**, 9693–9714 (1987).
50. Kashinath, K. et al. Physics-informed machine learning: case studies for weather and climate modelling. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **379**, 20200093 (2021).
51. Teufel, B. et al. Physics-informed deep learning framework to model intense precipitation events at super resolution. *Geosci. Lett.* **10**, 19 (2023).
52. Pegram, G. G. S. & Clothier, A. N. High resolution space-time modelling of rainfall: the “String of Beads” model. *J. Hydrol.* **241**, 26–41 (2001).
53. Hess, P., Drüke, M., Petri, S., Strnad, F. M. & Boers, N. Physically constrained generative adversarial networks for improving precipitation fields from Earth system models. *Nat. Mach. Intell.* **4**, 828–839 (2022).
54. Goodfellow, I. et al. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, Vol. 27 (eds Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. & Weinberger, K. Q.) <https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf> (Curran Associates, Inc., 2014).
55. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 5967–5976 (2017).
56. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. In *International Conference on Learning Representations* (2019).
57. Gneiting, T. & Raftery, A. E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **102**, 359–378 (2007).
58. Roberts, N. Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteorol. Appl.* **15**, 163–169 (2008).
59. Roberts, N. M. & Lean, H. W. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Weather Rev.* **136**, 78–97 (2008).
60. Harris, D., Fofoula-Georgiou, E., Droegemeier, K. K. & Levit, J. J. Multiscale statistical properties of a high-resolution precipitation forecast. *J. Hydrometeorol.* **2**, 406–418 (2001).
61. Sinclair, S. & Pegram, G. G. S. Empirical mode decomposition in 2-D space and time: a tool for space-time rainfall analysis and nowcasting. *Hydrol. Earth Syst. Sci.* **11**, 127–137 (2005).
62. Pulkkinen, S. et al. Pysteps: an open-source Python library for probabilistic precipitation nowcasting (v1.0). *Geosci. Model Dev.* **12**, 4185–4219 (2019).
63. Candille, G. & Talagrand, O. Evaluation of probabilistic prediction systems for a scalar variable. *Q. J. R. Meteorol. Soc.* **131**, 2131–2150 (2005).
64. Hamill, T. M. Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Weather Rev.* **129**, 550–560 (2001).
65. Foresti, L. et al. A quest for precipitation attractors in weather radar archives. *Nonlinear Process. Geophys.* **31**, 259–286 (2024).
66. Bell, B. et al. The ERA5 global reanalysis: preliminary extension to 1950. *Q. J. R. Meteorol. Soc.* **147**, 4186–4227 (2021).
67. Lopez, P. Direct 4D-var assimilation of NCEP stage IV radar and gauge precipitation data at ECMWF. *Mon. Weather Rev.* **139**, 2098–2116 (2011).
68. Lin, Y. & Mitchell, K. E. The NCEP stage II/IV hourly precipitation analyses: development and applications. In *Proc. 19th Conference Hydrology, American Meteorological Society, San Diego, CA, USA*, Vol. **10** (2005).
69. Chen, Y. et al. Spatial performance of multiple reanalysis precipitation datasets on the southern slope of central Himalaya. *Atmos. Res.* **250**, 105365 (2021).
70. Gomis-Cebolla, J., Rattayova, V., Salazar-Galán, S. & Francés, F. Evaluation of ERA5 and ERA5-Land reanalysis precipitation datasets over Spain (1951–2020). *Atmos. Res.* **284**, 106606 (2023).
71. Xu, J., Ma, Z., Yan, S. & Peng, J. Do ERA5 and ERA5-land precipitation estimates outperform satellite-based precipitation products? A comprehensive comparison between state-of-the-art model-based

- and satellite-based precipitation products over mainland China. *J. Hydrol.* **605**, 127353 (2022).
72. Bandhauer, M. et al. Evaluation of daily precipitation analyses in E-OBS (v19.0e) and ERA5 by comparison to regional high-resolution datasets in European regions. *Int. J. Climatol.* **42**, 727–747 (2022).
73. Beck, H. E. et al. MSWEP V2 Global 3-Hourly 0.1° precipitation: methodology and quantitative assessment. *Bull. Am. Meteorol. Soc.* **100**, 473–500 (2019).
74. Winterrath, T. et al. *Erstellung einer radargestützten Niederschlagsklimatologie*. Tech. Rep. 251, Deutscher Wetterdienst [https://www.dwd.de/DE/leistungen/pbfb\\_verlag\\_berichte/pdf\\_einzelbaende/251\\_pdf.pdf?\\_\\_blob=publicationFile&v=2](https://www.dwd.de/DE/leistungen/pbfb_verlag_berichte/pdf_einzelbaende/251_pdf.pdf?__blob=publicationFile&v=2) (2017).
75. Zhang, J. et al. Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: initial operating capabilities. *Bull. Am. Meteorol. Soc.* **97**, 621–638 (2016).
76. Smith, T. M. et al. Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: initial operating capabilities. *Bull. Am. Meteorol. Soc.* **97**, 1617–1630 (2016).
77. Seed, A., Curtis, M. & Velasco-Forero, C. AURA - operational radar rainfields 3 [https://pid.nci.org.au/doi/f0493\\_5520\\_4121\\_7835](https://pid.nci.org.au/doi/f0493_5520_4121_7835) (2022).
78. Chumchean, S., Seed, A. & Sharma, A. Correcting of real-time radar rainfall bias using a Kalman filtering approach. *J. Hydrol.* **317**, 123–137 (2006).
79. Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data* **6**, 60 (2019).
80. Terzago, S., Palazzi, E. & von Hardenberg, J. Stochastic downscaling of precipitation in complex orography: a simple method to reproduce a realistic fine-scale climatology. *Nat. Hazards Earth Syst. Sci.* **18**, 2825–2840 (2018).
81. Paszke, A. et al. Pytorch: An imperative style, high-performance deep learning library, *Advances in neural information processing systems* 32 (2019).

## Acknowledgements

This work was supported by funding from the Federal Ministry of Education and Research (BMBF) and the Helmholtz Research Field Earth Environment within the Innovation Pool Project SCENIC and ACTUATE and the HFMI project HCLimRep. Further support has been granted through the DFG research unit RealPEP (Grant Number: CH-1785/1-2). Furthermore, we acknowledge the support of the Deutsches Klimarechenzentrum (DKRZ) by providing high-performance cluster resources, granted by its Scientific Steering Committee (WLA) under project ID 1343.

## Author contributions

The conceptualization of the study was carried out by L.G., C.C., J.P., H.K., and B.F. Data curation was performed by L.G. and C.C., with funding acquisition led by H.K., B.F., and C.C. The analysis, software, and visualization were performed by L.G. The manuscript was written by L.G., J.P. and C.C.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41612-025-01103-y>.

**Correspondence** and requests for materials should be addressed to Luca Glawion.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025