Environmental Data Science (2026), 5: e37, 1–15
doi:10.1017/eds.2026.10066

CAMBRIDGE UNIVERSITY PRESS

APPLICATION PAPER

IPSL-AID: generative diffusion models for climate downscaling from global to regional scales

Kishanthan Kingston[1], Olivier Boucher[1], Freddy Bouchet[2], Pierre Chapel[2], Rosemary Eade[1], Jean-Francois Lamarque[3], Redouane Lguensat[1] and Kazem Ardaneh[1]

[1]Climate Modeling Center, Sorbonne University, CNRS, IPSL, Paris, France
[2]LMD, CNRS, ENS/PSL, Sorbonne University, IPSL, Paris, France
[3]Climate Modeling and Analysis, LLC, Superior, CO, USA
Corresponding author: Kazem Ardaneh; Email: kardaneh@ipsl.fr

Received: 13 September 2026 UTC; **Revised:** 13 September 2026 UTC; **Accepted:** 13 September 2026 UTC

Keywords: climate downscaling; diffusion models; generative models; machine learning; uncertainty quantification

Abstract

Effective adaptation and mitigation strategies for climate change require high-resolution projections to inform strategic decision-making. Conventional global climate models, which typically operate at resolutions of 150 to 200 km, lack the capacity to represent essential regional processes. IPSL-AID is a global to regional downscaling tool based on a denoising diffusion probabilistic model designed to address this limitation. Trained on ERA5 reanalysis data, it generates 0.25° resolution fields for temperature, wind, and precipitation using coarse inputs and their spatiotemporal context. It also models probability distributions of fine-scale features to produce plausible scenarios for uncertainty quantification. The model accurately reconstructs statistical distributions, including extreme events, power spectra, and spatial structures. This work highlights the potential of generative diffusion models for efficient climate downscaling with uncertainty estimation. Climate downscaling generally comprises two components—super-resolution and bias correction. This work focuses on the former.

Impact Statement

IPSL-AID is a generative tool designed for climate downscaling—from the global to the regional scale—developed through a collaboration between computer scientists, AI/ML specialists, and climate scientists. This interdisciplinary team offers a multidimensional perspective that balances computational efficiency, statistical rigor, and physical consistency, hence making the work accessible to researchers across these fields.

1. Introduction

Anthropogenic climate change poses substantial risks to critical socioeconomic sectors, such as agriculture, forestry, energy, and water supply. The development of effective adaptation and mitigation strategies requires accessible, high-resolution (HR) climate projections to anticipate impacts at both local and

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (http://creativecommons.org/licenses/by-nc-sa/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use.

regional scales. General circulation models (GCMs), which are widely used in climate research, typically operate at spatial resolutions of approximately 150–200 km (IPCC, 2023). This coarse resolution is insufficient to capture essential fine-scale processes, especially those shaped by topography, land–sea contrast, and surface heterogeneity, all of which are vital to regional climate dynamics and extreme weather events. Therefore, downscaling climate model outputs to finer resolutions is necessary to provide relevant climate information at the local level.

Weather and climate downscaling methods are typically classified into two categories (Hewitson and Crane, 1996; Wilby and Wigley, 1997): (i) Dynamic downscaling uses nested regional climate models (RCMs) to explicitly simulate physical processes at finer spatial scales (Giorgi and Mearns, 1991; Vrac et al., 2012; Giorgi and Gutowski, 2015). However, the high computational demands of this approach limit its feasibility for large ensembles or long simulation periods. (ii) Statistical downscaling establishes empirical relationships between large-scale climate predictors, such as pressure fields and circulation indices, and local-scale variables (Wilby et al., 2004; Vrac et al., 2012; Maraun and Widmann, 2017). However, this approach has limited capacity to capture complex nonlinear interactions and to maintain multivariate physical consistency within the generated fields.

Recent studies have shown that RCMs can be emulated with relatively low computational cost (Doury et al., 2023; Rampal et al., 2024). Artificial intelligence-based methods for weather prediction, including Pangu-Weather (Bi et al., 2023), GraphCast (Lam et al., 2023), and FourCastNet (Pathak et al., 2022), have demonstrated significant potential (Koldunov et al., 2024). These data-driven models facilitate efficient, automated downscaling by learning to incorporate fine-scale details from reanalysis datasets such as ERA5. However, these approaches are generally deterministic and may introduce systematic biases. Critically, they often lack a robust framework for quantifying the uncertainty inherent in the downscaling process, which is essential for comprehensive risk assessment. Recent works in downscaling have highlighted the potential of diffusion models (Rampal et al., 2024). CorrDiff, introduced by Mardani et al. (2025), is a residual corrective diffusion model that downscales 25 km ERA5 inputs to 2 km resolution over Taiwan, demonstrating strong performance in reconstructing spatial spectra and synthesizing radar reflectivity. Similarly, Watt and Mansfield (2024) used a diffusion model to downscale 2° ERA5 fields to 0.25° over the continental United States, exhibiting superior performance compared to U-Net and linear interpolation. While these studies confirm the promise, they are constrained to fixed regional domains. Moreover, neither study addressed precipitation downscaling, a critical variable for impact assessment, and both relied on region-specific training that cannot be generalized to new geographic areas without retraining.

This article introduces IPSL AI Downscaling (IPSL-AID), a tool for global to regional downscaling using denoising diffusion probabilistic models (DDPMs; Ho et al., 2020; Song et al., 2020a, 2020b; Karras et al., 2022) with several key innovations. First, we propose a *random block sampling strategy* that enables training the model on global data with limited GPU resources, thereby delivering a model applicable to arbitrary regions without retraining, representing a step toward foundation models for downscaling. Second, the model architecture processes *multiple time steps simultaneously*, enabling temporally coherent evaluations. Third, the model jointly downscales *multiple atmospheric variables* (here, temperature, winds, and precipitation) using a flexible design to include more variables. Fourth, it provides a *comprehensive evaluation*, including deterministic metrics, distributional fidelity, spectral recovery, and probabilistic skills. Our results show that the model accurately reconstructs fine-scale variability and extreme events, while its generative nature enables uncertainty quantification—essential for risk assessment. This work addresses the super-resolution component of global-to-regional climate downscaling, providing a framework applicable to climate projections, including CMIP6 and future simulation sets.

2. Methodology overview

IPSL-AID is based on DDPMs. These models are trained to iteratively add and remove noise from data, thereby learning to reconstruct underlying fine-scale fields. For downscaling, we condition the model on

coarse fields and their spatiotemporal context to generate multiple plausible realizations. While IPSL-AID encompasses the full design space of diffusion models introduced by Karras et al. (2022), including variance-preserving, variance-expanding, and improved DDPM formulations, this work focuses on the Elucidated Diffusion Model (EDM). The core architecture is a U-Net implemented within the EDM framework. Supplementary Appendix A provides a technical description of the model architecture, training procedure, loss function, sampling algorithm, and overall workflow.

3. Data, sampling strategy, and metrics

The model is trained using four ERA5 variables (Hersbach et al., 2020): 2-m temperature (t2m), 10-m zonal wind (u10), 10-m meridional wind (v10), and precipitation (tp) at 0.25° resolution. To balance temporal variability and computational cost, we select four random time steps per day.

Training HR global downscaling models is computationally intensive and requires substantial memory and GPU resources. To address this challenge, a random spatial block sampling strategy was developed. During each training epoch, s spatial blocks of size 144×360 are generated, with block centers placed randomly (Figure 1). The longitude of each block is treated as periodic, while the latitude is constrained within valid global boundaries. Several values for the number of spatial blocks per epoch ($s = 6$, 9, and 12) were tested, using $s = 12$ was identified as an effective balance between computational efficiency and spatial diversity. For evaluation and inference, 20 fixed blocks corresponding to the ERA5 resolution of 1440×721 were used for global predictions. Our random block sampling strategy fundamentally differs from NVIDIA's patch-based multidiffusion approach (Brenowitz et al., 2025). Rather than training models in parallel on patches extracted from fixed domains using a Message Passing Interface (MPI) across multiple GPUs, IPSL-AID trains a single model using random blocks sampled across the globe. This enables direct application to arbitrary regions without retraining while retaining scalability to global inference.

A coarse-up (CU) procedure based on bilinear interpolation is used to separate large-scale and fine-scale components of the ERA5 fields. Let $\mathbf{y}^{\mathrm{HR}}$ be a HR field. The HR field is first reduced to a coarse resolution of 16×32 ($\approx 2.25°$ grid, comparable to typical GCMs resolution) by bilinear interpolation, which eliminates small-scale variability while preserving large-scale structure. The resulting coarse field is then scaled back to the original HR grid using the same interpolation method, yielding a smooth CU

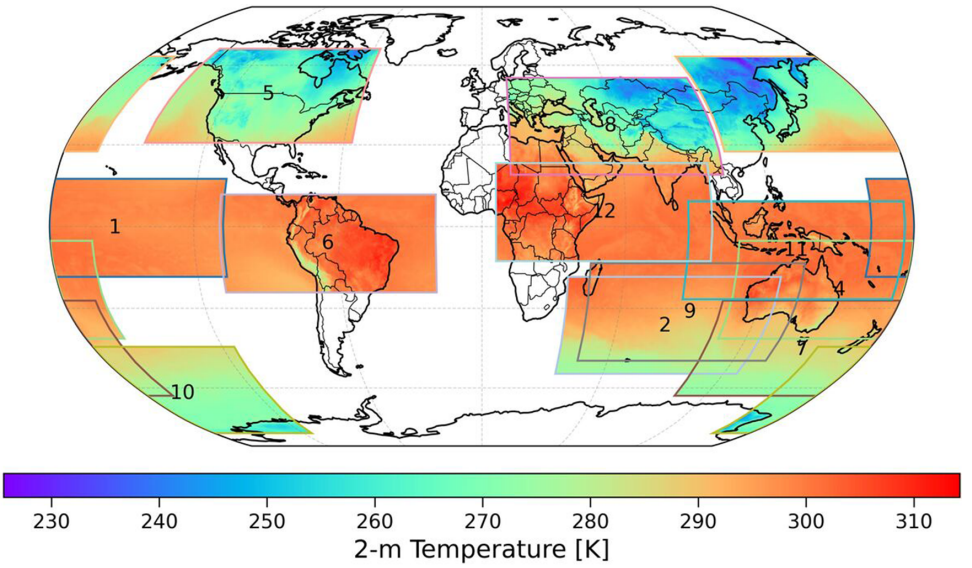


Figure 1. *An example of randomly sampled spatial blocks used during training of the global model.*

approximation, $\mathbf{y}^{\mathrm{CU}}$, unlike Mardani et al. (2025), who typically rely on mean regression models. The residual field, $\mathbf{R} = \mathbf{y}^{\mathrm{HR}} - \mathbf{y}^{\mathrm{CU}}$, represents the fine-scale information lost during the upscaling process and serves as the training target for the diffusion model.

A conditional diffusion model is trained to generate fine-scale residual fields given CU fields and auxiliary spatiotemporal conditioning. The CU fields are provided as spatial conditioning inputs and are augmented with geographical variables, including latitude, longitude, topography (z), and land-sea mask (LSM). Temporal information is incorporated using cosine–sine representations of the day of year and hour of day (time embedding). The trained model generates residual fields ($\mathbf{R}'$) conditioned on the CU inputs and spatiotemporal context. The final HR prediction is obtained by adding the generated residual to the CU field ($\mathbf{y}^{\mathrm{CU}} + \mathbf{R}'$), which preserves large-scale consistency and enhances fine-scale details. The workflow is shown in Supplementary Appendix A.5. We set the optimal parameters of the sampler following an ablation study (Supplementary Appendix B.1). We trained a U-Net model as a baseline using the L2 loss function to predict the HR fields, using the network described in Supplementary Appendix A.2. The inputs include the CU fields, static geographical variables, and temporal embeddings. The diffusion model has 92,140,548 trainable parameters, compared with 92,135,940 for the U-Net baseline. The models were trained on data from 2015 to 2019, validated on 2020 data, and tested on 2021 data. Training ran for 100 epochs with a batch size of 70, using four NVIDIA A100 GPUs (64 GB each) on the LEONARDO supercomputer. Training, evaluation, and inference for the diffusion model took approximately 6 days.

As the current work focuses on the super-resolution step of downscaling, each HR prediction is compared with its corresponding ERA5 reference using pointwise metrics, complemented by distributional diagnostics. The primary metrics include the Mean Absolute Error $\mathrm{MAE} = \frac{1}{N}\sum_{\mathrm{n}=1}^{\mathrm{N}} |y_{\mathrm{n}} - \hat{y}_{\mathrm{n}}|$, the Normalized MAE $\mathrm{NMAE} = \sum_{\mathrm{n}=1}^{\mathrm{N}} |y_{\mathrm{n}} - \hat{y}_{\mathrm{n}}| / \sum_{\mathrm{n}=1}^{\mathrm{N}} |y_{\mathrm{n}}|$, the Root Mean Square Error $\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{\mathrm{n}=1}^{\mathrm{N}} (y_{\mathrm{n}} - \hat{y}_{\mathrm{n}})^2}$, the coefficient of determination $R^2 = 1 - \sum_{\mathrm{n}=1}^{\mathrm{N}} (y_{\mathrm{n}} - \hat{y}_{\mathrm{n}})^2 / \sum_{\mathrm{n}=1}^{\mathrm{N}} (y_{\mathrm{n}} - \overline{y})^2$, and the Continuous Ranked Probability Score $\mathrm{CRPS}(F_{\mathrm{n}}, y_{\mathrm{n}}) = \int_{-\infty}^{\infty} \mathrm{d}\hat{y} \left(F_{\mathrm{n}}(\hat{y}) - \mathbf{1}(\hat{y} - y_{\mathrm{n}})\right)^2$. The CRPS is calculated using 100 time steps, for which the EDM sampler is run 10 times. Here, y_{n} and $\hat{y}_{\mathrm{n}}$ represent the HR fields from the ERA5 and predicted values, $\overline{y}$ is the mean of the truth, N is the number of samples, and F is the predictive cumulative distribution function.

Power Spectral Density (PSD) analysis is used to evaluate the model's ability to reconstruct spatial variability across scales, from large-scale structures to smaller scale features. Probability Density Functions (PDFs) on logarithmic scales enable comparisons of statistical distributions and the evaluation of extreme-value representation. Distributional consistency is further quantified using the Kullback–Leibler (KL) divergence $\mathrm{KL}(p\|q) = \int \mathrm{d}y p(y) \log \frac{p(y)}{q(y)}$, where p and q denote the reference and predicted PDFs, respectively. Additionally, the Pearson correlation coefficient $\rho_{\mathrm{p,q}} = \frac{\mathrm{cov}(p,q)}{\sigma_{\mathrm{p}}\sigma_{\mathrm{q}}}$ is computed between the predicted and true fields to assess the similarity of their spatial patterns and overall structure, where p and q denote the flattened prediction and truth, respectively. The qth quantile of a sample a is $Q(q) = (1-g)\,y[j] + g\,y[j+1]$, with y sorted, $j = \lfloor (n-1)q \rfloor$, and $g = (n-1)q - j$.

For the diffusion model, we also evaluated rank histograms, which assess how well the predicted ensemble represents the truth, and the spread–skill ratio (SSR), which quantifies the consistency between ensemble spread and prediction error. Given an ensemble of size M (we used $M = 10$), the sorted predictions are $\hat{y}_{(1)} \le \hat{y}_{(2)} \le \cdots \le \hat{y}_{(M)}$. The rank r of reference y_{n} relative to the ensemble is given by $r = \sum_{\mathrm{m}=1}^{\mathrm{M}} \mathbf{1}(\hat{y}_{\mathrm{m}} \le y_{\mathrm{n}})$, so that $r \in \{0, \ldots, M\}$. The rank histogram is obtained by plotting the histogram of r over all prediction–truth pairs. The SSR is defined as $\mathrm{SSR} = \frac{\sigma_{\mathrm{ens}}}{\mathrm{RMSE}}$, where the ensemble spread is given by $\sigma_{\mathrm{ens}} = \sqrt{\frac{M+1}{M}} \sqrt{\frac{1}{N}\sum_{\mathrm{n}=1}^{\mathrm{N}} \frac{1}{M}\sum_{\mathrm{m}=1}^{\mathrm{M}} \left(\hat{y}_{\mathrm{n}}^{(\mathrm{m})} - \overline{\hat{y}}_{\mathrm{n}}\right)^2}$, and $\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{\mathrm{n}=1}^{\mathrm{N}} \left(\overline{\hat{y}}_{\mathrm{n}} - y_{\mathrm{n}}\right)^2}$. Here, $\hat{y}_{\mathrm{n}}^{(\mathrm{m})}$ denotes the m-th ensemble member (with $m = 1, \ldots, M$), $\overline{\hat{y}}_{\mathrm{n}} = \frac{1}{M}\sum_{\mathrm{m}=1}^{\mathrm{M}} \hat{y}_{\mathrm{n}}^{(\mathrm{m})}$ is the ensemble mean, y_{n} is the reference value, and N is the number of samples. The corrective factor in the definition of σ_{ens} ensures the optimal value of the SSR is 1, as explained in Fortin et al. (2014).

4. Results and discussion

Figure 2 shows surface plots for the downscaled variables: t2m, u10, v10, and tp. Three representative members of a 10-member ensemble, the ensemble standard deviation, and a zoomed-in over South America are shown in Supplementary Appendix B.2 and Supplementary Appendix B.3, respectively. Visually, the model successfully reconstructs the fine-scale spatial details of the HR fields. Table 1 summarizes the performance of the CU input, U-Net, the diffusion model, and the diffusion ensemble mean. For t2m, the prediction achieves an MAE of 0.44 ± 0.02 K, an RMSE of 0.71 ± 0.04 K, and an R^2 of 1.00. The corresponding ensemble mean improves these scores, with an MAE of 0.32 ± 0.01 K, an RMSE of 0.52 ± 0.01 K, and an R^2 of 1.00. The wind components are predicted with similar accuracy. For u10, the prediction yields an MAE of 0.52 ± 0.02 m s^{-1}, an RMSE of 0.79 ± 0.03 m s^{-1}, and an R^2 of 0.98, while the ensemble mean achieves an MAE of 0.40 ± 0.01 m s^{-1}, an RMSE of 0.62 ± 0.03 m s^{-1}, and an R^2 of 0.988 ± 0.002. Similarly, for v10, the prediction has an MAE of 0.50 ± 0.01 m s^{-1}, an RMSE of 0.74 ± 0.03 m s^{-1}, and an R^2 of 0.98, while the ensemble mean yields an MAE of 0.38 ± 0.01 m s^{-1}, an RMSE of 0.59 ± 0.03 m s^{-1}, and an R^2 of 0.985 ± 0.002. For tp, the prediction shows an average MAE of 0.10 ± 0.00 mm hr^{-1}, which appears very low; however, the NMAE of 0.67 ± 0.02 indicates a larger relative error. The ensemble mean slightly improves the absolute error, with an MAE of 0.055 ± 0.001 mm hr^{-1} and an RMSE of 0.234 ± 0.001 mm hr^{-1}, while the NMAE remains relatively high at 0.53 ± 0.01 and R^2 reaches 0.68 ± 0.01. The discrepancy between MAE and NMAE arises from the variable's skewed distribution, characterized by a high frequency of zero values, which makes absolute error metrics less informative. Given its temporal and spatial variability, precipitation is one of the most complex fields to downscale; thus, the emphasis is on overall quality (see the following discussions).

The CRPS metric is 0.25 ± 0.01 K for t2m, 0.29 ± 0.01 m s^{-1} for u10, 0.28 ± 0.01 m s^{-1} for v10, and 0.04 mm hr^{-1} for tp. Watt and Mansfield (2024) reported CRPS values of 0.254 K (t2m), 0.224 m s^{-1} (u10), and 0.232 m s^{-1} (v10) over the United States, whereas Mardani et al. (2025), for the Taiwan domain, report higher values for extremes: 0.55 K, 0.86 m s^{-1}, and 0.95 m s^{-1}, respectively. The low CRPS for precipitation suggests that the model's predicted distribution is confident, particularly in correctly capturing the prevalence of dry (zero-precipitation) conditions. Overall, the diffusion model yields physically consistent (an intervariable correlation analysis that is presented in Supplementary Appendix B.6) and statistically accurate downscaling, with the ensemble mean improving deterministic performance.

The statistical fidelity of the downscaled fields was evaluated by jointly analyzing density scatter plots, PDFs, and PSDs of the model predictions, the HR ERA5 reference, and the CU inputs (Figure 3). The first row displays density scatter plots comparing the diffusion model predictions with the HR reference across all variables. Each panel presents the joint distribution of values, with colors representing the logarithm of sample density and the one-to-one reference line. Quantitative metrics, including R^2, MAE, and RMSE, are also given. For the t2m and (u10, v10) components, the predictions show a strong linear relationship with the reference fields, as indicated by points tightly clustered along the diagonal and R^2 values near 0.99. Small vertical lines in the density plot for u10 and v10 indicate that some true values correspond to predicted values with the opposite sign. These discrepancies likely stem from challenges in downscaling coastal winds influenced by land–sea breezes (Figure 4) and affect only a small fraction of datapoints. For precipitation, although the scatter indicates increased spread, particularly at higher intensities, the model substantially sharpens near the center, reduces underestimation, and more closely aligns with the one-to-one relationship, producing the same zero-precipitation points.

The model exhibits a strong capacity to replicate the full statistical distribution of t2m. The predicted mean ($\mu = 278.89$ K) and standard deviation ($\sigma = 21.33$ K) closely match those of the ground truth ($\mu = 278.91$ K, $\sigma = 21.32$ K), whereas the CU input displays a slightly reduced variance ($\sigma = 21.18$ K). This high level of distributional agreement is quantitatively supported by a KL divergence of 0.00 and a Pearson correlation of 1.00. The log-PDF curves of the prediction and reference are nearly indistinguishable across the t2m range, demonstrating accurate reproduction of both central tendencies and distribution tails. The model's performance for the (u10, v10) components is comparably robust. For the u10, both the mean ($\mu_{\text{pred}} = -0.04$ m s^{-1}, $\mu_{\text{truth}} = -0.04$ m s^{-1}) and variance ($\sigma_{\text{pred}} = 5.63$ m s^{-1}, $\sigma_{\text{truth}} = 5.64$ m s^{-1}) are

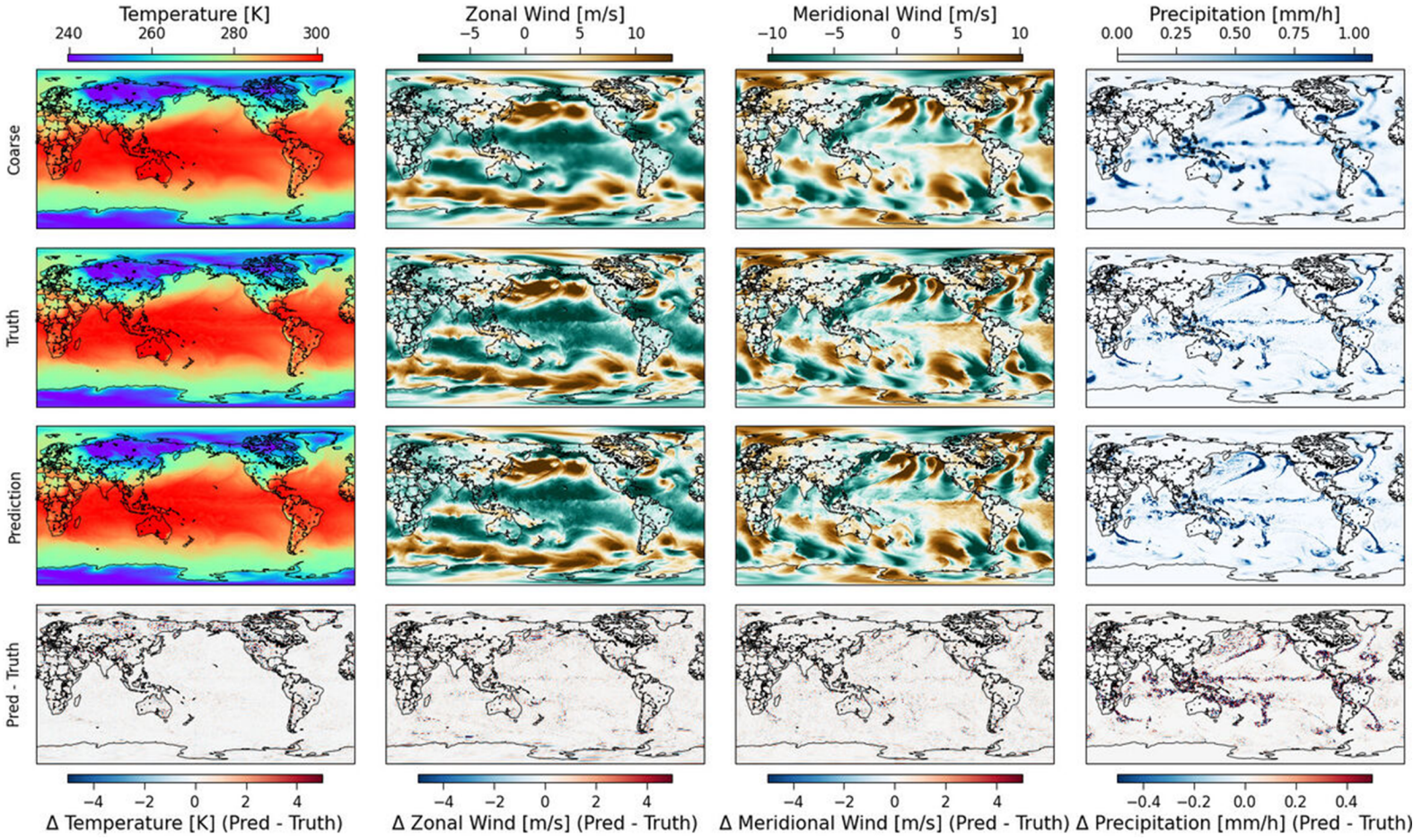


Figure 2. *Surface plots are shown for t2m, u10, v10, and tp (columns 1 to 4). The rows, from top to bottom, represent the CU ERA5 input, the HR ERA5 reference, the model prediction, and the difference between the prediction and the HR reference. All panels correspond to 2021-01-01 06:00 UTC.*

Table 1. *Performance comparison of diffusion, ensemble diffusion, U-Net, and CU input*

	t2m (K)				u10 (m s^{-1})				v10 (m s^{-1})				tp (mm hr^{-1})			
Metric	Diff	Ens.	U-Net	CU	Diff	Ens.	U-Net	CU	Diff	Ens.	U-Net	CU	Diff	Ens.	U-Net	CU
MAE/NMAE↓	0.44	0.32	0.31	0.79	0.52	0.40	0.35	0.82	0.50	0.38	0.34	0.78	0.67	0.53	0.49	0.76
RMSE↓	0.71	0.52	0.50	1.40	0.79	0.62	0.55	1.26	0.74	0.59	0.53	1.18	0.28	0.23	0.21	0.30
R^2↑	1.00	1.00	1.00	1.00	0.98	0.98	0.99	0.95	0.98	0.98	0.99	0.94	0.50	0.68	0.72	0.48

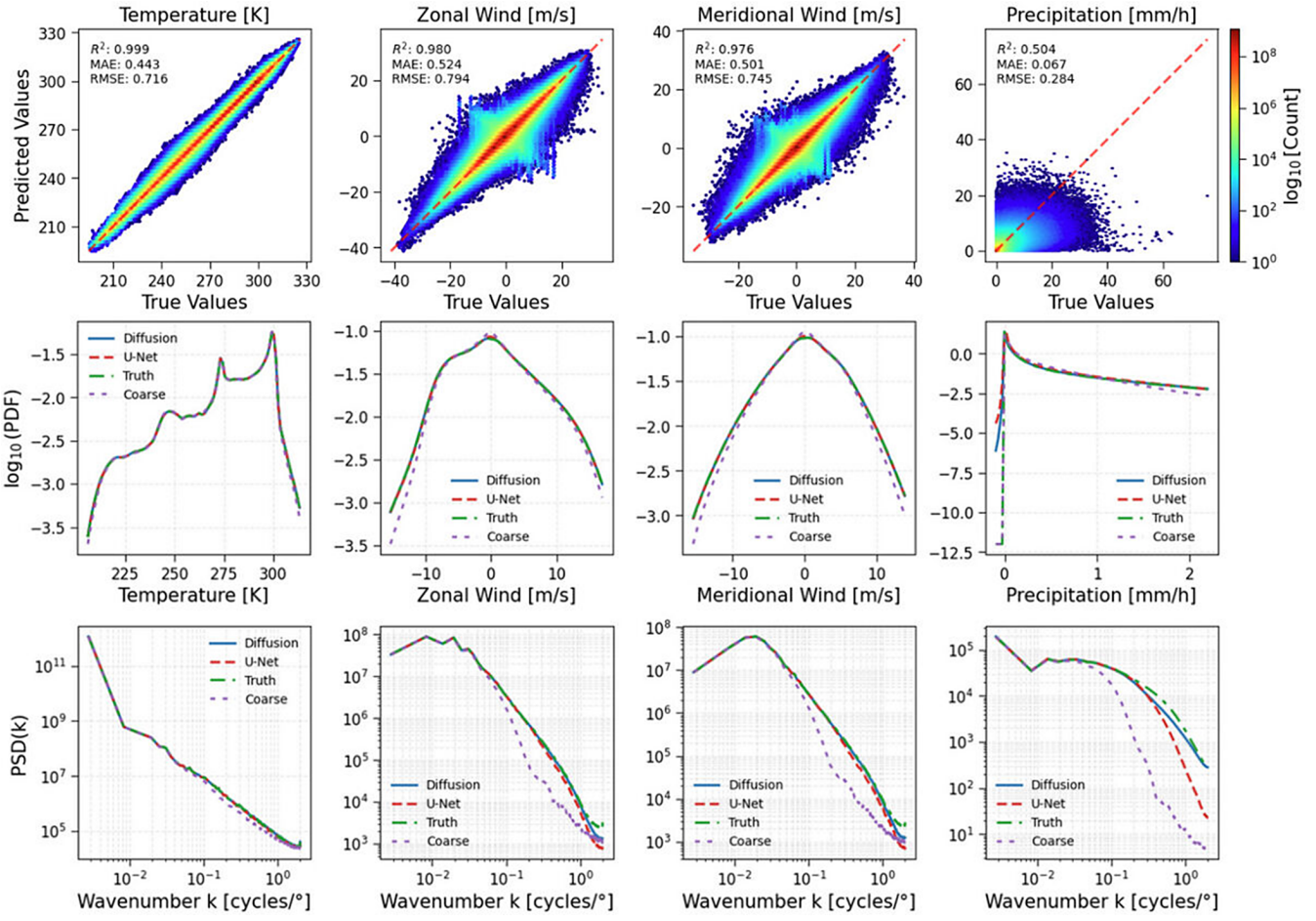


Figure 3. *Model evaluation for the complete 2021 evaluation dataset. Columns represent t2m, u10, v10, and tp, respectively. Rows display density scatter plots (top), PDFs (middle), and PSDs (bottom).*

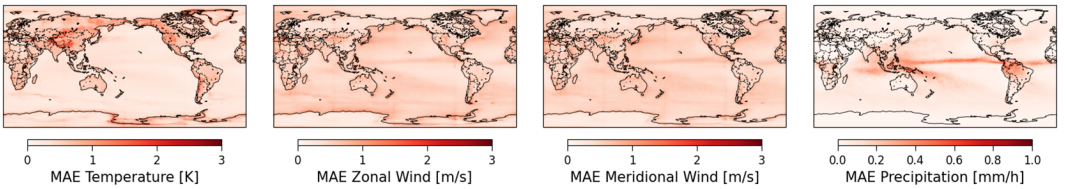


Figure 4. *Spatial distribution of the MAE, averaged over the 2021 dataset, for the downscaled variables: t2m, u10, v10, and tp.*

accurately captured, while the CU input underestimates variability ($\sigma = 5.26$ m s^{-1}). The predicted distribution aligns closely with the reference, as proved by a KL divergence of 0.00 and a Pearson correlation of 0.99. The v10 shows similar results, with the model recovering the true variance ($\sigma_{\text{pred}} = 4.77$ m s^{-1}, $\sigma_{\text{truth}} = 4.78$ m s^{-1}) from the smoothed CU field ($\sigma = 4.39$ m s^{-1}), while maintaining a KL divergence of 0.00 and a Pearson correlation of 0.99. Precipitation prediction presents significant challenges due to its intermittency, non-Gaussian distribution, and localized extremes. Nevertheless, the model demonstrates substantial improvement over the CU input, with a KL divergence of 0.001 and a Pearson correlation of 0.73. The log-PDF indicates that the predictions more accurately capture the frequency of light-to-moderate precipitation and partially recover the tail of extreme events, which are suppressed in the CU field.

The U-Net baseline reproduces statistical distributions effectively (KL divergence and Pearson correlations of approximately 10^{-4} and 0.99, respectively). However, its performance declines for precipitation with a KL divergence of 0.3. Analysis of residual fields (Supplementary Appendix B.4) shows that the U-Net generates PDFs that are more peaked and narrower than those from the diffusion model across variables, indicating underestimation of fine-scale variability and extremes.

The model's ability to recover spatial variability across scales is evaluated by analyzing the PSDs of the generated fields and comparing them with the HR reference and the CU input, as illustrated in the third row of Figure 3. For all variables, the diffusion model prediction closely matches the reference spectrum throughout most of the resolved wavenumber range, extending into smaller scale regimes. The CU input shows significant depletion at high wavenumbers, reflecting the loss of small-scale variability due to spatial averaging. In contrast, the predicted spectra maintain both the spectral slope and amplitude, indicating the model's capacity to reproduce physically consistent small-scale variability. U-Net spectra generally follow the large-scale truth data but exhibit earlier attenuation at higher wavenumbers, indicating a reduced ability to capture small-scale variability compared to the diffusion model. In the case of precipitation, the recovery of high-wavenumber energy is especially pronounced, signifying the restoration of fine-scale structures associated with convective and localized rainfall processes, which remain parameterized at the 25 km resolution. Only minor deviations are observed at the highest wavenumbers, where all spectra are affected by numerical noise and resolution constraints.

The model's performance in predicting extreme values was quantified using quantile–quantile analysis of distribution upper tails (Table 2). For t2m, both diffusion model and the U-Net reconstruction closely match observed quantiles from 90th to 99.5th percentile. The diffusion model shows perfect agreement with truth (309.21 versus 309.21 K at $q = 0.995$), while U-Net slightly underestimates the highest quantiles (309.16 versus 309.21 K). CU input underestimates high temperatures, with discrepancies reaching 0.91 K at the 99.5th percentile. For wind components, both models slightly underestimate extreme wind speeds. The diffusion model remains close to truth (15.80 versus 15.81 m s^{-1} at $q = 0.995$ for u10), while U-Net shows larger underestimation (15.72 m s^{-1}). Similar behavior occurs for v10. This shows both models reproduce most wind distribution, though capturing extreme events remains challenging. CU input performs worse, underestimating u10 extremes by 0.70 m s^{-1} at the 99.5th percentile. For precipitation, the diffusion model slightly underestimates moderate extremes but overestimates most extreme events (2.45 versus 2.43 mm hr^{-1} at $q = 0.995$). The U-Net underestimates the upper tail more

Table 2. *Extreme true, diffusion model, U-Net, and CU input quantiles for the downscaled variables*

	t2m (K)				u10 (m s^{-1})				v10 (m s^{-1})				tp (mm hr^{-1})			
Quantile (q)	Truth	Diff	U-Net	CU	Truth	Diff	U-Net	CU	Truth	Diff	U-Net	CU	Truth	Diff	U-Net	CU
0.90	300.09	300.07	300.08	300.01	7.61	7.60	7.56	7.09	6.10	6.09	6.06	5.64	0.20	0.21	0.21	0.27
0.95	301.20	301.17	301.14	300.97	10.24	10.24	10.19	9.71	7.93	7.91	7.89	7.31	0.47	0.49	0.49	0.49
0.975	302.57	302.57	302.52	302.13	12.33	12.33	12.27	11.77	9.60	9.57	9.55	8.82	0.92	0.94	0.89	0.75
0.99	306.40	306.41	306.35	305.55	14.48	14.47	14.40	13.85	11.62	11.58	11.56	10.66	1.71	1.73	1.56	1.13
0.995	309.21	309.21	309.16	308.30	15.81	15.80	15.72	15.11	13.04	13.00	12.97	11.94	2.43	2.45	2.15	1.45

remarkably (2.15 versus 2.43 mm hr^{-1} at $q = 0.995$). CU input overestimates lower precipitation quantiles but significantly underestimates the highest quantiles, with values below two-thirds of ground truth at $q = 0.995$.

Spatial error distributions indicate regional variation in model performance as illustrated in Figure 4. The MAE for t2m is higher at high latitudes, particularly near the poles, and in areas with complex topography, such as the Himalayas and the Andes. The (u10, v10) components also show increased errors in coastal regions and along steep topographic gradients, likely attributable to the modulation of surface winds by land–sea contrasts and orographic influences. In contrast, precipitation errors are most pronounced in the tropics, especially within the Intertropical Convergence Zone and over tropical rainforests, where intense and localized rainfall events pose significant challenges for downscaling. The analogous maps illustrating the relative error are provided in Supplementary Appendix B.5.

The rank histograms shown in Figure 5 are generally close to uniform, indicating that the diffusion model ensemble is reasonably well calibrated overall. However, the temperature histogram exhibits a slight upward slope with a negative bias, suggesting that the truth often lies above the ensemble members, indicating a tendency toward underestimation. The u10 histogram is fairly flat but shows a mild positive bias, pointing to a slight overestimation. In contrast, the v10 histogram is nearly uniform with negligible bias, indicating very good calibration. The precipitation histogram displays a noticeable spike at the lowest rank and a slight skew, suggesting that the truth frequently falls below the ensemble range, which implies overestimation and potential under-dispersion.

Figure 6 presents density plots of ensemble spread versus RMSE. The red dashed line represents the ideal 1:1 relationship (SSR = 1), which corresponds to a perfectly calibrated ensemble. All variables display SSR values slightly below 1, indicating a mild overall under-dispersion. The highest sample density is concentrated at low RMSE and low spread for all variables, which reflects strong predictive skill. Deviations from the 1:1 line become increasingly evident as the RMSE values increase. For u10, v10, and tp, the spread remains relatively small as RMSE increases, suggesting under-dispersion in higher error regimes. t2m exhibits a similar, though less pronounced, pattern.

5. Conclusions and limitations

This work introduces IPSL-AID, a diffusion-based generative model developed for global-to-regional climate downscaling, and presents its super-resolution component. IPSL-AID reconstructs HR fields of temperature, winds, and precipitation from coarse input data and their spatiotemporal context. The model is trained on ERA5 reanalysis using a novel random block sampling strategy, which allows for global applicability without the need for retraining in new regions. A key advantage of this approach is its probabilistic framework. By learning the underlying distribution of residuals at fine scales, IPSL-AID generates multiple plausible HR realizations from a single coarse input. This capability enables robust uncertainty quantification through ensemble spread analysis, rank histograms, and spread–skill ratio diagnostics.

Quantitative evaluation indicates that IPSL-AID achieves low deterministic errors (MAE of 0.44 K for T2m, 0.52 m s^{-1} for u10, and 0.50 m s^{-1} for v10) and accurately represents the statistical distribution of the variables. The model reproduces spatial variability across scales and captures extreme events, with quantile estimates closely matching observations up to the 99.5th percentile. IPSL-AID demonstrates robust performance across diverse atmospheric variables and regions while maintaining physical consistency, as shown by intervariable correlation analysis that reveals realistic multivariate dependencies, such as temperature–wind patterns over oceans and continents. The ensemble mean further enhances deterministic scores, reducing MAE to 0.32 K for T2m and 0.40 m s^{-1} for u10.

Despite its strengths, IPSL-AID exhibits certain limitations. Performance degrades in regions with complex topography (e.g., Himalayas, Andes), coastal zones, and areas characterized by intense convective precipitation, where errors remain elevated. For precipitation, the normalized MAE of 0.53 indicates a substantial relative error, despite a low absolute error, reflecting the inherent challenge of downscaling non-Gaussian fields. The rank histograms reveal systematic tendencies: a slight

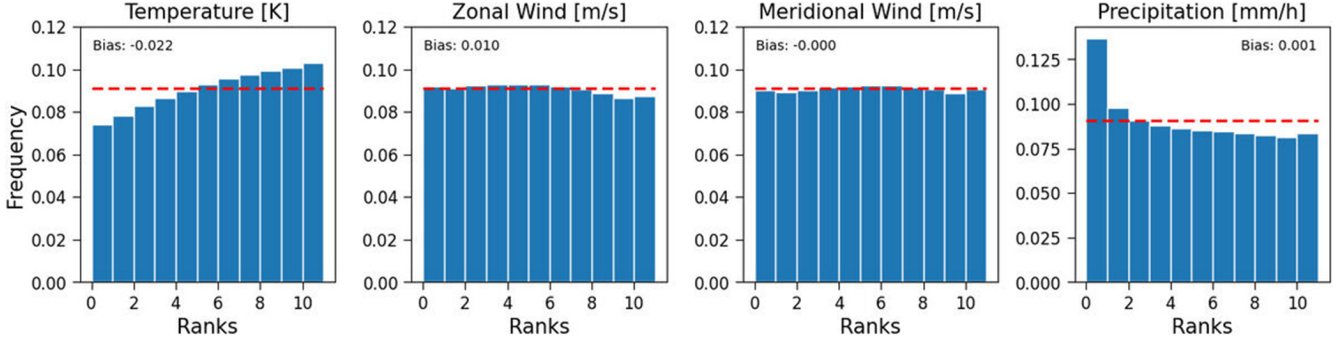


Figure 5. *Rank histograms for the 10-member diffusion ensemble. Columns show t2m, u10, v10, and tp. Red dashed lines mark the expected uniform frequency.*

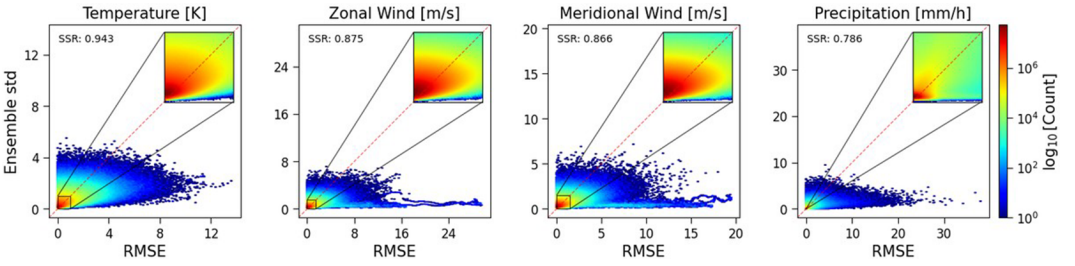


Figure 6. *Density plots of ensemble spread versus RMSE for t2m, u10, v10, and tp. The red dashed line indicates the 1:1 relationship corresponding to perfect spread–skill (SSR = 1). In each panel, the inset shows a zoomed-in view containing approximately 99% of the data points.*

underestimation for temperature, a mild overestimation for u10 winds, and under-dispersion for precipitation extremes. Spread–skill analysis indicates under-dispersion in higher error regimes, suggesting the ensemble may not fully capture uncertainty in challenging conditions. Additionally, block boundary artifacts remain visible during global inference, and the five-year training period is short compared to the multi-decadal timescales typically required for climate studies. The computational demands of global training—six days on four A100 GPUs—may also limit accessibility for some research groups.

Several avenues for future development are planned. First, we aim to extend the model to additional atmospheric variables and apply the framework to downscale future projections from CMIP6 models. This will require developing and integrating appropriate bias correction or debiasing strategies (Wan et al., 2023, 2025). To address computational limitations, distributed multi-GPU training strategies will be explored to accommodate longer training periods (e.g., 40 years) while maintaining global coverage. For block boundary artifacts, several mitigation approaches are under investigation: (i) an overlap-block method with cosine-weighted blending during inference, (ii) explicit boundary regularization in the loss function to penalize discontinuities along seam lines, (iii) frequency domain spectral loss to suppress characteristic artifact frequencies, and (iv) adaptive Fourier domain filtering as a post-processing step.

Finally, we plan to incorporate additional conditioning variables (e.g., soil moisture and surface fluxes) and explore alternative diffusion formulations to improve performance for precipitation extremes. The apparent contradiction between low CRPS and high NMAE for precipitation reflects the different aspects they measure: CRPS captures probabilistic calibration, whereas NMAE reflects the difficulty of deterministic prediction for a non-Gaussian variable; with conditioning limited to topography, strong deterministic skill is not achieved.

Open peer review. To view the open peer review materials for this article, please visit http://doi.org/10.1017/eds.2026.10066.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/eds.2026.10066.

Acknowledgments. We acknowledge the EuroHPC Joint Undertaking for awarding this project access to the EuroHPC supercomputer LEONARDO, hosted by CINECA (Italy) and the LEONARDO consortium through an EuroHPC Development Access call (EHPC-DEV-2024D10-070, EHPC-DEV-2025D08-095).

Author contribution. Conceptualization: All authors contributed equally. Data curation: K.K.; Formal analysis: All authors contributed equally. Funding acquisition: J.F.L, O.B, and K.A.; Investigation: All authors contributed equally. Methodology: All authors contributed equally. Project administration: O.B. and K.A.; Resources: O.B. and K.A.; Software: K.A. and K.K.; Supervision: J.F.L., O.B., and F.B.; Validation: K.K., K.A., and P.C.; Visualization: K.K., K.A., and P.C.; Writing—original draft: K.A.; Writing—review and editing: All authors contributed equally. All authors have reviewed and approved the final manuscript.

Competing interests. The authors declare none.

Data availability statement. The IPSL-AID source code is publicly available on GitHub at https://github.com/kardaneh/IPSL-AID. The specific code version, along with the pre-trained model checkpoints, configuration scripts, and the 2021 ERA5 dataset used for inference in this study, are archived and publicly accessible on Zenodo at https://doi.org/10.5281/zenodo.19200763.

Ethics statement. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Funding statement. Project S24JROI011 supported this work from the Grantham Foundation (To K.A and O.B). R.L. acknowledges support from Agence Nationale de la Recherche – France 2030 through the PEPR TRACCS programme under grant number ANR-22-EXTR-0011 (LOCALISING).

Provenance. This article is part of the Climate Informatics 2026 proceedings and was accepted in Environmental Data Science on the basis of the Climate Informatics peer review process.

References

Bi K, **Xie L**, **Zhang H**, **Chen X**, **Gu X and 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.

Brenowitz ND, **Ge T**, **Subramaniam A**, **Manshausen P**, **Gupta A**, **Hall DM**, **Mardani M**, **Vahdat A**, **Kashinath K and Pritchard MS** (2025) Climate in a bottle: Towards a generative foundation model for the kilometer-scale global atmosphere. *arXiv preprint.* arXiv:2505.06474.

Doury A, **Somot S**, **Gadat S**, **Ribes A and Corre L** (2023) Regional climate model emulator based on deep learning: Concept and first evaluation of a novel hybrid downscaling approach. *Climate Dynamics 60*(5), 1751–1779.

Fortin V, **Abaza M**, **Anctil F and Turcotte R** (2014) Why should ensemble spread match the RMSE of the ensemble mean? *Journal of Hydrometeorology 15*(4), 1708–1713. https://doi.org/10.1175/JHM-D-14-0008.1.

Giorgi F and Gutowski WJ (2015) Regional dynamical downscaling and the CORDEX initiative. *Annual Review of Environment and Resources 40*(1), 467–490.

Giorgi F and Mearns LO (1991) Approaches to the simulation of regional climate change: a review. *Reviews of Geophysics 29*(2), 191–216.

Hersbach H, **Bell B**, **Berrisford P**, **Hirahara S**, **Horányi A**, **Muñoz-Sabater J**, **Nicolas J**, **Peubey C**, **Radu R**, **Schepers D**, **Simmons A**, **Soci C**, **Abdalla S**, **Abellan X**, **Balsamo G**, **Bechtold P**, **Biavati G**, **Bidlot J**, **Bonavita M**, **de Chiara G**, **Dahlgren P**, **Dee D**, **Diamantakis M**, **Dragani R**, **Flemming J**, **Forbes R**, **Fuentes M**, **Geer A**, **Haimberger L**, **Healy S**, **Hogan RJ**, **Hólm E**, **Janisková M**, **Keeley S**, **Laloyaux P**, **Lopez P**, **Lupu C**, **Radnoti G**, **de Rosnay P**, **Rozum I**, **Vamborg F**, **Villaume S and Thépaut JN** (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society 146*(730), 1999–2049.

Hewitson BC and Crane RG (1996) Climate downscaling: techniques and application. *Climate Research 7*(2), 85–95.

Ho J, **Jain A and Abbeel P** (2020) Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems 33*, 6840–6851.

Intergovernmental Panel on Climate Change (IPCC) (2023) Linking global-to-regional Climate Change. In *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press, pp. 1363–1512.

Karras T, **Aittala M**, **Aila T and Laine S** (2022) Elucidating the design space of diffusion-based generative models. *Advances in Neural Information Processing Systems 35*, 26565–26577.

Koldunov N, **Rackow T**, **Lessig C**, **Danilov S**, **Cheedela SK**, **Sidorenko D**, **Sandu I and Jung T** (2024) Emerging AI-based weather prediction models as downscaling tools. *arXiv preprint. arXiv:2406.17977*.

Lam R, **Sanchez-Gonzalez A**, **Willson M**, **Wirnsberger P**, **Fortunato M**, **Alet F**, **Ravuri S**, **Ewalds T**, **Eaton-Rosen Z**, **Hu W**, **Merose A**, **Hoyer S**, **Holland G**, **Vinyals O**, **Stott J**, **Pritzel A**, **Mohamed S and Battaglia P** (2023) Learning skillful medium-range global weather forecasting. *Science 382*(6677), 1416–1421.

Maraun D and Widmann M (2017) *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge: Cambridge University Press.

Mardani M, **Brenowitz N**, **Cohen Y**, **Pathak J**, **Chen C-Y**, **Liu C-C**, **Vahdat A**, **Nabian MA**, **Ge T**, **Subramaniam A**, **Kashinath K**, **Kautz J and Pritchard M** (2025) Residual corrective diffusion modeling for km-scale atmospheric downscaling. *Communications Earth & Environment 6*(1), 124.

Pathak J, **Subramanian S**, **Harrington P**, **Raja S**, **Chattopadhyay A**, **Mardani M**, **Kurth T**, **Hall D and Li Z**, **Azizzadenesheli K.**, **Hassanzadeh P**, **Kashinath K and Anandkumar A** (2022) Fourcastnet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. *arXiv preprint. arXiv:2202.11214*, Retrieved April 15, 2024, from http://arxiv.org/abs/2202.11214.

Rampal N, **Hobeichi S**, **Gibson PB**, **Baño-Medina J**, **Abramowitz G**, **Beucler T**, **González-Abad J**, **Chapman W**, **Harder P and Gutiérrez JM** (2024) Enhancing regional climate downscaling through advances in machine learning. *Artificial Intelligence for the Earth Systems 3*(2), 230066.

Song J, **Meng C and Ermon S** (2020a) Denoising diffusion implicit models. *arXiv preprint. arXiv:2010.02502*.

Song Y, **Sohl-Dickstein J**, **Kingma DP**, **Kumar A**, **Ermon S and Poole B** (2020b) Score-based generative modeling through stochastic differential equations. *arXiv preprint. arXiv:2011.13456*.

Vrac M, **Drobinski P**, **Merlo A**, **Herrmann M**, **Lavaysse C**, **Li L and Somot S** (2012) Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty assessment. *Natural Hazards and Earth System Sciences 12*(9), 2769–2784.

Wan ZY, **Baptista R**, **Chen Y-F**, **Anderson J**, **Boral A**, **Sha F and Zepeda-Núnez L** (2023) Debias coarsely, sample conditionally: statistical downscaling through optimal transport and probabilistic diffusion models. *arXiv preprint. arXiv:2305.15618*.

Wan ZY, **Lopez-Gomez I**, **Carver R**, **Schneider T**, **Anderson J**, **Sha F and Zepeda-Núñez L** (2025) Regional climate risk assessment from climate models using probabilistic machine learning. *arXiv preprint. arXiv:2412.08079*, Retrieved December 10, 2025, from http://arxiv.org/abs/2412.08079.

Watt RA and Mansfield LA (2024) Generative diffusion-based downscaling for climate. *arXiv preprint.* arXiv:2404.17752.

Wilby RL and Wigley TML (1997) Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography 21*(4), 530–548.

Wilby RL, **Charles SP**, **Zorita E**, **Timbal B**, **Whetton P and Mearns LO** (2004) Guidelines for use of climate scenarios developed from statistical downscaling methods, *Supporting material of the Intergovernmental Panel on Climate Change, available from the DDC of IPCC TGCIA 27.* https://api.semanticscholar.org/CorpusID:17449064

Cite this article: Kingston K, Boucher O, Bouchet F, Chapel P, Eade R, Lamarque J.-F., Lguensat R and Ardaneh K (2026). IPSL-AID: generative diffusion models for climate downscaling from global to regional scales. *Environmental Data Science*, 5: e37. doi:10.1017/eds.2026.10066