

Generative Diffusion Outperforms Quantile Mapping for Climate Downscaling Across Scales

Kaushik Muduchuru

KaushikReddy.Muduchuru@zalf.de

Leibniz Centre for Agricultural Landscape Research (ZALF)

Amit Kumar Srivastava

Leibniz Centre for Agricultural Landscape Research (ZALF)

Krishnagopal Halder

Leibniz Centre for Agricultural Landscape Research (ZALF)

Manmeet Singh

Western Kentucky University

Ivonne Anders

Deutsches Klimarechenzentrum GmbH (DKRZ)

Senthold Asseng

Technical University of Munich, Digital Agriculture, HEF World Agricultural Systems Center

Luis Alfredo Pires Barbosa

Leibniz Centre for Agricultural Landscape Research (ZALF)

Kazi Jahidur Rahaman

Leibniz Centre for Agricultural Landscape Research (ZALF)

Wenzhi Zeng

Hohai University

Thomas Gaiser

University of Bonn

Frank Ewert

Leibniz Centre for Agricultural Landscape Research (ZALF)

Article

Keywords:

Posted Date: September 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10600756/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Generative Diffusion Outperforms Quantile Mapping for Climate Downscaling Across Scales

Kaushik Muduchuru^{1*}, Amit Kumar Srivastava¹, Krishnagopal Halder¹, Manmeet Singh², Ivonne Anders³, Senthod Asseng⁴, Luis Alfredo Pires Barbosa¹, Kazi Jahidur Rahaman¹, Wenzhi Zeng^{5,6,7}, Thomas Gaiser⁸, Frank Ewert^{1,8}

¹Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Str. 84, 15374, Müncheberg, Germany

²Earth, Environmental and Atmospheric Sciences, Western Kentucky University, Bowling Green, KY, United States of America

³Deutsches Klimarechenzentrum GmbH (DKRZ), Datamanagement, Hamburg, 20156, Germany

⁴Technical University of Munich, Department of Life Science Engineering, Digital Agriculture, HEF World Agricultural Systems Center, Freising, 85354, Germany

⁵College of Agricultural Science and Engineering, Hohai University, Nanjing, 210098, China

⁶Key Laboratory of Agricultural Low-Altitude Technology Innovation and Integration, Ministry of Agriculture and Rural Affairs, Nanjing 211100, China

⁷Jiangsu Province Engineering Research Center for Agricultural Soil–Water Efficient Utilisation, Carbon Sequestration and Emission Reduction, Nanjing 211100, China

⁸Institute of Crop Science and Resource Conservation, University of Bonn, Bonn, 53111, Germany

Corresponding Author:

Kaushik Muduchuru

Research Area 4 “Simulation & Data Science”

Leibniz Centre for Agricultural Landscape Research (ZALF)

Eberswalder Str. 84

15374 Müncheberg

Germany

Email: KaushikReddy.Muduchuru@zalf.de; kaushik.reddy.m@gmail.com

Ph: +49 15510 527370

Abstract:

Climate impact assessment at field and catchment scales depends on localised extremes, yet global climate models (~100 km) smooth over the terrain-driven variability, spatial structure, and inter-variable dependencies through which those extremes and compound events arise. Widely used statistical methods such as ISIMIP3BASD correct coarse-scale distributions but operate point-by-point, disrupting spatial coherence and daily co-variability. We present CasCorrDiff (Cascading Correction Diffusion), a generative diffusion framework that downscales in two supervised stages — 100 km to 10 km, then 10 km to 1 km — bridging a 100× resolution gap while preserving spatial covariance, inter-variable coupling, and physical event sequencing. Evaluated over Germany (2020–2024) against station and gridded references, CasCorrDiff outperforms ISIMIP3BASD and interpolated baselines in daily synchronisation (maximum temperature $r = 0.99$, NSE = 0.98) and reproduces observed mesoscale power spectra (10–200 km) and extreme indices, including the temperature Wasserstein distance ($W = 0.37$), consecutive dry days, and joint temperature–precipitation co-variability. It underestimates the 95th percentile of precipitation and loses skill in complex terrain. Because inference is decoupled from the multi-decade baselines that quantile mapping requires, a trained model downscales individual days on demand.

Introduction

While climate change is traditionally quantified through global mean temperature shifts, its primary ecological and economic damages arise from altered climate variability and localised extreme events^{1,2}. Heatwaves, droughts, and high precipitation events are changing in frequency, intensity, and spatial organisation, and these transformations of the underlying statistical distribution drive severe, often irreversible shocks to agricultural yields, ecosystems, and hydrological stability^{3–5}. Interannual variability alone accounts for 30% to 50% of crop yield variances globally⁶. Increasingly, the most damaging risks are compound extremes, involving co-occurring extreme events such as concurrent heat and drought or sequential transitions from wet to dry conditions. The joint occurrence of these extremes amplifies systemic vulnerability far beyond the impact of any single hazard^{7,8}. Simulating these risks requires climate data that is simultaneously local, resolving changing weather at the kilometre scales at which impacts are realised, and distributionally faithful, reproducing not only mean states but the variances, spatial structure, and multi-variable dependencies through which compound events arise. Global Climate Models (GCMs) cannot provide this. The GCM coarse grids (~100 km) smooth over the convective processes, terrain-driven heterogeneity, and higher-order statistical moments that govern local extremes^{9,10}. Bridging this scale gap through downscaling is therefore a prerequisite for translating global projections into locally actionable climate information¹¹.

To bridge this scale mismatch, two primary methodological paradigms exist: dynamical downscaling and statistical downscaling. Dynamical downscaling employs high-resolution regional climate models to simulate local physics explicitly compared to GCMs, but its higher computational demand severely restricts its deployment across multi-model ensembles or long-term horizons. Conversely, statistical downscaling offers a highly efficient alternative by establishing empirical relationships between large-scale predictors and local observations. Frameworks like ISIMIP3BASD¹², which leverages

parametric quantile mapping^{13,14}, have commodified bias-adjusted data for global impact modelling. The widespread utility of this trend-preserving approach is demonstrated by its adoption as the core processing engine for large-scale impact risk assessment pipelines, regional climate networks like EURO-CORDEX, global crop modelling initiatives under AgMIP, and operational workflows for extreme-event attribution^{15–17}. Yet, these traditional approaches carry severe limitations: because they typically adjust variables independently at individual grid points, they disrupt inter-variable dependencies, spatial coherence, and physical event sequencing, making them ill-suited for risk assessments^{18,19}.

Recently, deep learning has emerged to overcome these multi-dimensional limitations²⁰. Generative Adversarial Networks (GANs) have demonstrated skill in resolving fine-scale spatial structures^{21,22}, but they are frequently undermined by training instabilities and mode collapse, often failing to faithfully reproduce the full stochastic distribution of climate variables. Diffusion models have emerged as a promising alternative to GANs. They offer more stable training and a principled probabilistic framework for modelling complex data distributions^{23,24}. At their core, diffusion models gradually transform noise into structured outputs, making them particularly well suited for capturing uncertainty and variability—both central features of the climate system. Compared to GANs, they are less prone to issues such as mode collapse and training instability, which makes them attractive for climate applications. Several recent studies have already demonstrated their potential. For instance, Mardani et al., (2025)²⁵ applied diffusion models to kilometre-scale downscaling of surface weather variables such as wind, air temperature, and precipitation, conditioned on large-scale atmospheric fields. Similar approaches have been used in fluid-dynamics settings, where they successfully capture complex spatiotemporal behaviour driven by synoptic-scale conditions^{26,27}. They have also shown promise in learning convective organisation patterns, which is key for extreme precipitation events^{25,28}.

Here we evaluate and compare the Cascading Correction Diffusion model (CasCorrDiff; Fig. 1), a generative framework that treats fine-scale downscaling not as a deterministic mapping, but as a stochastic sampling process capable of resolving fine-scale climate distributions. Building on the residual-correction principle of CorrDiff²⁵, CasCorrDiff decomposes downscaling into a deterministic mean state, resolved by regression, and a stochastic sub-grid residual, generated by diffusion process^{23,24}. To establish the real-world efficacy of generative models for climate downscaling, we directly benchmark CasCorrDiff against ISIMIP3BASD, the gold-standard statistical bias-adjustment framework widely adopted across global impact networks (e.g., AgMIP, EURO-CORDEX). While ISIMIP3BASD relies on point-wise quantile mapping that disrupts spatial power spectra and inter-variable dependencies, CasCorrDiff overcomes these multi-dimensional limitations through a novelty two-stage cascade that bridges a $100 \times$ resolution jump (from ~ 100 km synoptic fields to 1 km local grids). By decomposing the task into an intermediate 10 km stage (supervised against MSWX) for macro-scale systematic bias correction, and a final 1 km stage (supervised against HYRAS) for terrain-driven conditional refinement, each stage learns the physical controls at its native scale. We evaluate whether this architecture restores the spatial covariance, daily variability, and station-scale alignment that traditional point-based methods might not necessarily capture. We compare CasCorrDiff against ISIMIP3BASD in capturing spatial power spectra and weather extremes while fully decoupling the inference from multi-decade historical baselines.

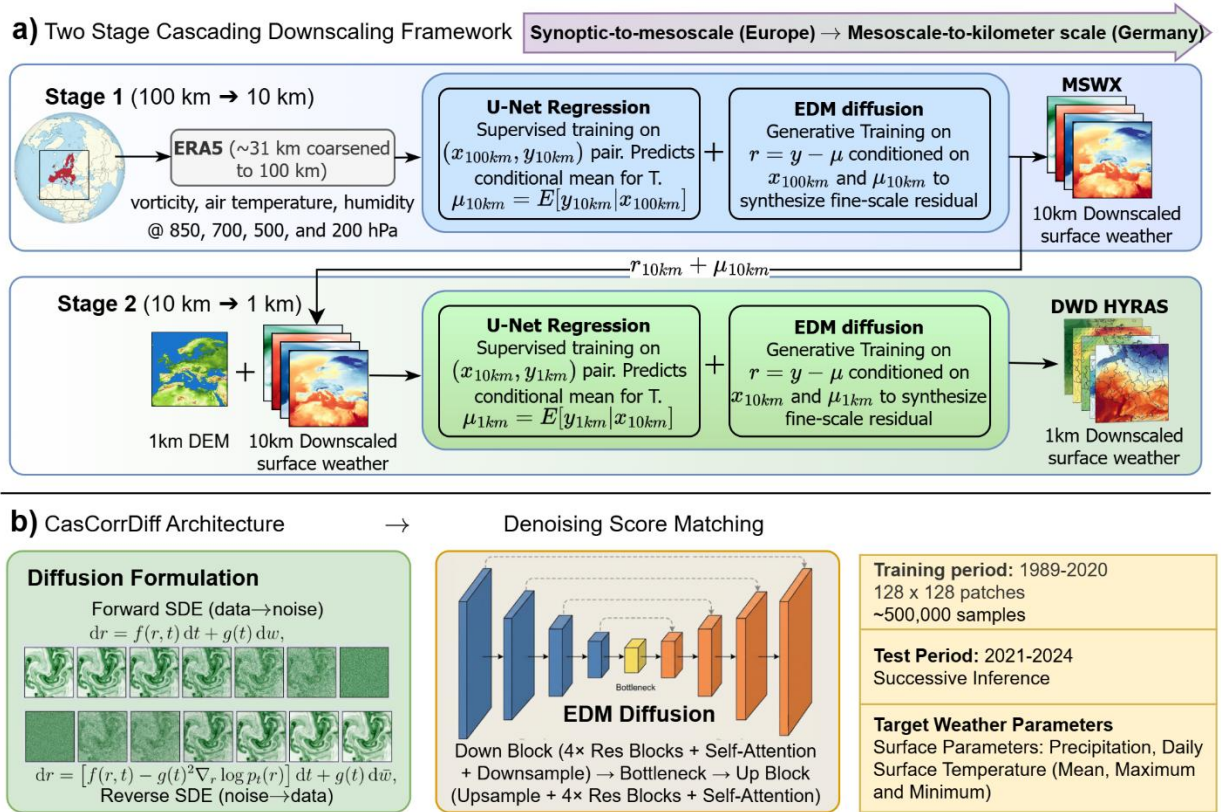


Figure 1: Overview of the hybrid Cascading Corrective Diffusion (CasCorrDiff) climate downscaling framework. a) Workflow schematic illustrating the two-stage progressive resolution enhancement. The coarse ERA5 global input (100 km) is downscaled in Stage 1 to 10 km (conditioned by upper-air variables; target: MSWX) to capture large-scale patterns. Stage 2 refines this output to a kilometre-scale (1 km) grid, heavily conditioned by a high-resolution Digital Elevation Model (DEM) and utilising HYRAS/DWD as the observational ground truth. Data specifics regarding training periods, test/validation windows, and variable sets (surface and upper-air) are summarised at the bottom. b) Detailed view of the CorrDiff architecture used at each stage. (Top-Left) The Elucidated Diffusion Model (EDM), showing Forward (data-to-noise) and Reverse (noise-to-data) stochastic differential equations (SDEs), illustrates the generative sampling process. (Middle) A deterministic U-Net Regression module predicts the conditional mean (capturing large-scale deterministic bias-correction), conditioned by coarse fields and, in Stage 2, by the DEM. (Top-Right) A second stochastic U-Net in the EDM performs residual generation (score matching), synthesising high-resolution details from the noisy residual, conditioning information, and the time embedding.

Results

Seasonal performance and daily weather variability of the downscaling framework

Evaluating the downscaling pipeline against independent in-situ records from the Global Historical Climatology Network Daily (GHCNd) station network establishes the model baseline across diverse seasonal synoptic regimes. Daily maximum near-surface air temperatures along with accumulated precipitation are well-resolved, with CasCorrDiff season-model pairs clustered in the upper skill quadrant ($r > 0.65$, $NSE > 0.40$). A clear hierarchy emerges when ranking performance across individual seasons, reflecting how the generative framework handles different structural features of the climate system. For daily maximum temperature, the skill rank follows the sequence Autumn (SON) > Spring (MAM) > Winter (DJF) > Summer (JJA), with 1 km CasCorrDiff reaching its peak performance in autumn ($r = 0.988$, $NSE = 0.976$) and spring ($r = 0.984$, $NSE = 0.968$). This indicates that the cascading framework

achieves its highest skill during stable autumn and spring synoptic setups, while complex boundary-layer interactions and higher surface heating during peak summer (JJA; $r = 0.972$, $NSE = 0.941$) present the highest relative modelling error.

Conversely, precipitation performance exhibits a completely different seasonal behaviour, ranking as Winter (DJF) > Spring (MAM) > Autumn (SON) > Summer (JJA). The model excels at capturing stable, frontally driven cold-season precipitation at 1 km (DJF: $r = 0.778$, $NSE = 0.599$; MAM: $r = 0.768$, $NSE = 0.566$), but encounters lower relative skill during summer (JJA: $r = 0.670$, $NSE = 0.445$). This drop reflects the challenges of resolving highly localised, short-duration convective cells during peak summer and handling rapid convective transitions. Importantly, even during these physically challenging seasons, the cascading framework significantly outperforms the standard ERA5 bicubic interpolated baseline (which drops to $r = 0.604$, $NSE = 0.284$ in JJA). Where standard bicubic interpolation suffers from chronic spatial smoothing, severely underestimates localised extreme volumes, and fails to capture true sub-grid stochastic variance, CasCorrDiff successfully recovers localised intensity distributions and structural details from ERA5. The transitions from the native 100 km ERA5 resolution down to the 10 km intermediate scale, and subsequently to the final 1 km local scale, yield a progressive increase in temporal synchronisation and magnitude calibration across all seasons. This spatial refinement is most pronounced in the Nash–Sutcliffe Efficiency (NSE) score distributions. The stepping mechanism achieves a localised gain in skill rather than a spatial redistribution of systemic error, verifying the physical reliability of the cascading approach prior to its application to uncalibrated global climate model projections.

Across the complete evaluation domain, model bias evolves systematically across the pipeline hierarchy, reflecting distinct physical and methodological mechanisms (Figs. S2–3). For precipitation, coarse ERA5 (~100 km) exhibits a chronic positive bias (+9.5%), driven by the inability of coarse grids to resolve local convective dynamics. The intermediate 10 km downscaling stage absorbs much of this large-scale systemic error, shifting the bias slightly negative to −8.8% (Fig. S2b). Applying statistical quantile mapping (ISIMIP3BASD) further minimises the mean bias to −3.7% (Fig. S2d); however, this distribution-matching approach achieves low mean error at the expense of day-to-day temporal synchronisation. In contrast, the final 1 km CasCorrDiff model maintains a comparable bias (−7.6%; Fig. S2c) while successfully restoring local spatial texture and variability, demonstrating that intermediate downscaling handles bulk bias correction while the 1 km generative stage refines fine-scale detail. Daily maximum temperature follows a much simpler trajectory: coarse ERA5 already exhibits only a minor cold bias (−6.0%), which is progressively eliminated through intermediate downscaling down to the final 1 km CasCorrDiff output, culminating in near-zero bias (−0.6%; Figs. S3b–c).

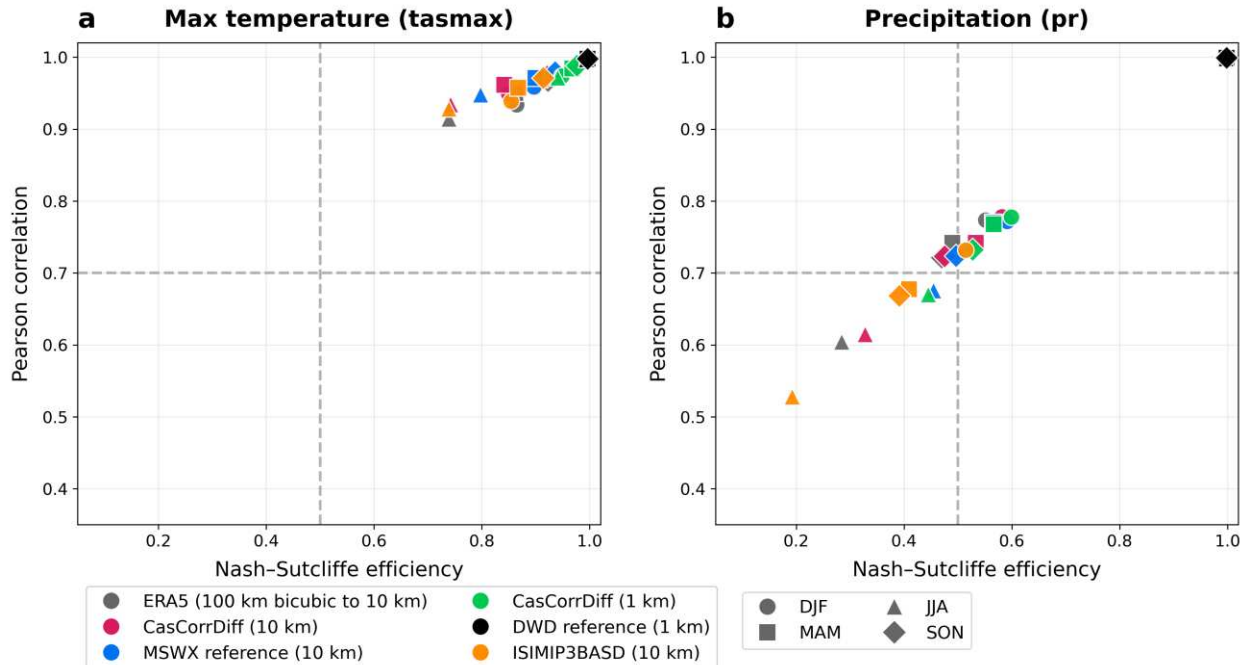


Figure 2. Scatter plots of Pearson correlation coefficient (r) versus Nash-Sutcliffe Efficiency (NSE) across all seasons (markers) and datasets (colors). Comparisons are shown for (a) daily maximum temperature (tasmax) and (d) precipitation (pr). The evaluated datasets include ERA5 (100 km resolution), CasCorrDiff downscaled predictions (10 km and 1 km), ISIMIP3BASD downscaled dataset and the DWD HYRAS reference datasets

Improved representation of probability distributions and mesoscale structures

Structural performance of the downscaling pipeline from synoptic to sub-mesoscale is analysed with the spatial power spectral densities of precipitation and daily maximum temperature for the test period (2020–2024). Radially symmetric spectra, derived via two-dimensional fast Fourier transforms (2D FFT) and azimuthal averaging across time steps and prediction ensembles, provide a robust characterisation of spatial variability from synoptic to local scales (1000 km to 2 km).

The 10 km CasCorrDiff configuration for precipitation (Fig. 3a) recovers substantial spatial variance across the entire mesoscale regime (10–200 km). When cascaded to the 1 km local scale, the diffusion model exhibits a marked upward shift in power, with performance that closely matches the DWD HYRAS observations across the entire mesoscale spectrum. As expected, the ERA5 dataset (100 km), though it performs well in capturing synoptic-scale structures, significantly deviates at the mesoscale, reflecting the smoothed mesoscale structures. ISIMIP3BASD statistical disaggregation underrepresents fine-scale variance even though a spatial multivariate rotational correction is applied at every coarse grid. The energy at fine scales is higher than in the 10 km CasCorrDiff configuration and the MSWX observation, reflecting a tendency to inflate fine-scale features. The daily maximum temperature spectra (Fig. 3b) reveal a structurally similar hierarchy, characterised by smoother spectral slopes, reflecting the higher spatial coherence of temperature relative to stochastic precipitation fields. The 1 km CasCorrDiff predictions demonstrate exceptional performance in capturing sub-grid variance, converging perfectly with the DWD reference truth down to the km scale. Across both variables, a systematic ordering in effective spatial fidelity emerges:

DWD Truth (1 km) $\approx$ CasCorrDiff 1 km > MSWX / CasCorrDiff 10 km > ISIMIP3BASD > ERA5 Input

This persistent hierarchy confirms that the generative framework injects physically realistic spatial structures rather than merely interpolating or statistical disaggregating coarse-scale fields.

To verify performance beyond central tendencies, a quantile–quantile difference analysis evaluates distributional biases across the 5th to 95th percentiles. This diagnosis assesses the marginal distribution of the pooled space-time sample rather than evaluating spatial structures or daily variability individually. Using this metric, for example, the 95th percentile from a single empirical CDF represents the "wettest 5% of all cell-days" rather than the tail behaviour of any single station over time.

For precipitation (Fig. 3c), In the upper tail, a striking divergence emerges. ERA5 exhibits a severe, monotonically increasing positive bias that accelerates to +1.12 mm/day at the 95th percentile, reflecting the artificial drizzle effect common to coarse-grid parameterisations. ISIMIP3BASD minimises the pooled marginal tail error best among the downscaled products, tracking close to perfect agreement with a minor residual underestimation of −0.19 mm/day at q_{95} . The 10 km CasCorrDiff configuration matches this performance closely, landing at −0.21 mm/day. Conversely, the 1 km CasCorrDiff variant exhibits a distinct non-linear trajectory: it shows a slight positive bias across mid-range quantiles (+0.51 mm/day at q_{50}) before dropping monotonically into a genuine dry bias of −1.10 mm/day at the 95th percentile—an underestimation representing roughly 10% of DWD's true q_{95} value.

Because this quantile analysis evaluates an unaligned, domain-wide pool, it is insulated from localised spatio-temporal phase errors (e.g., a convective event a few kilometres or days away from its target). Instead, the upper-tail deficit of the 1 km model points to a structural, macro-scale constraint within the generative framework. This signature closely matches a large-scale spectral deficit identified in the model's energy spectrum (where the ratio of CasCorrDiff to DWD energy drops to ≈ 0.6 within the 20–200 km mesoscale band). A model that under-produces coherent, large-scale wet anomalies will natively yield a thinned upper marginal tail while still generating highly realistic fine-scale local textures. What CasCorrDiff buys over traditional parametric methods like ISIMIP3BASD is therefore not an optimisation of the marginal distribution—which quantile mapping targets directly by design—but a massive improvement in capturing high-frequency, sub-grid spatial variance.

For daily maximum temperature (Fig. 3d), the bias structures remain bit-identical to prior configurations, revealing a highly stable thermodynamic signature. ERA5 exhibits a distinct warm bias across the majority of the distribution (+0.2 to +0.45°C), which drops sharply into a cold bias (−0.25°C) only at the extreme 95th percentile. ISIMIP3BASD displays a persistent cold bias across all quantiles, intensifying to −0.3°C at the upper tail.

The diffusion configurations substantially mitigate these errors. The 10 km CasCorrDiff product yields the most stable baseline, flatlining near perfect agreement (+0.05°C) across the entire distribution. The 1 km CasCorrDiff product tracks the zero line closely but displays high-frequency quantile oscillations between the 5th and 70th percentiles before dipping at the extreme tail. This behaviour underscores the trade-off of hyper-resolution modelling: while the 1 km framework successfully resolves the full spatial probability distribution essential for regional applications, its unaligned local variance requires careful consideration when validated against point-scale instruments for the extreme tails.

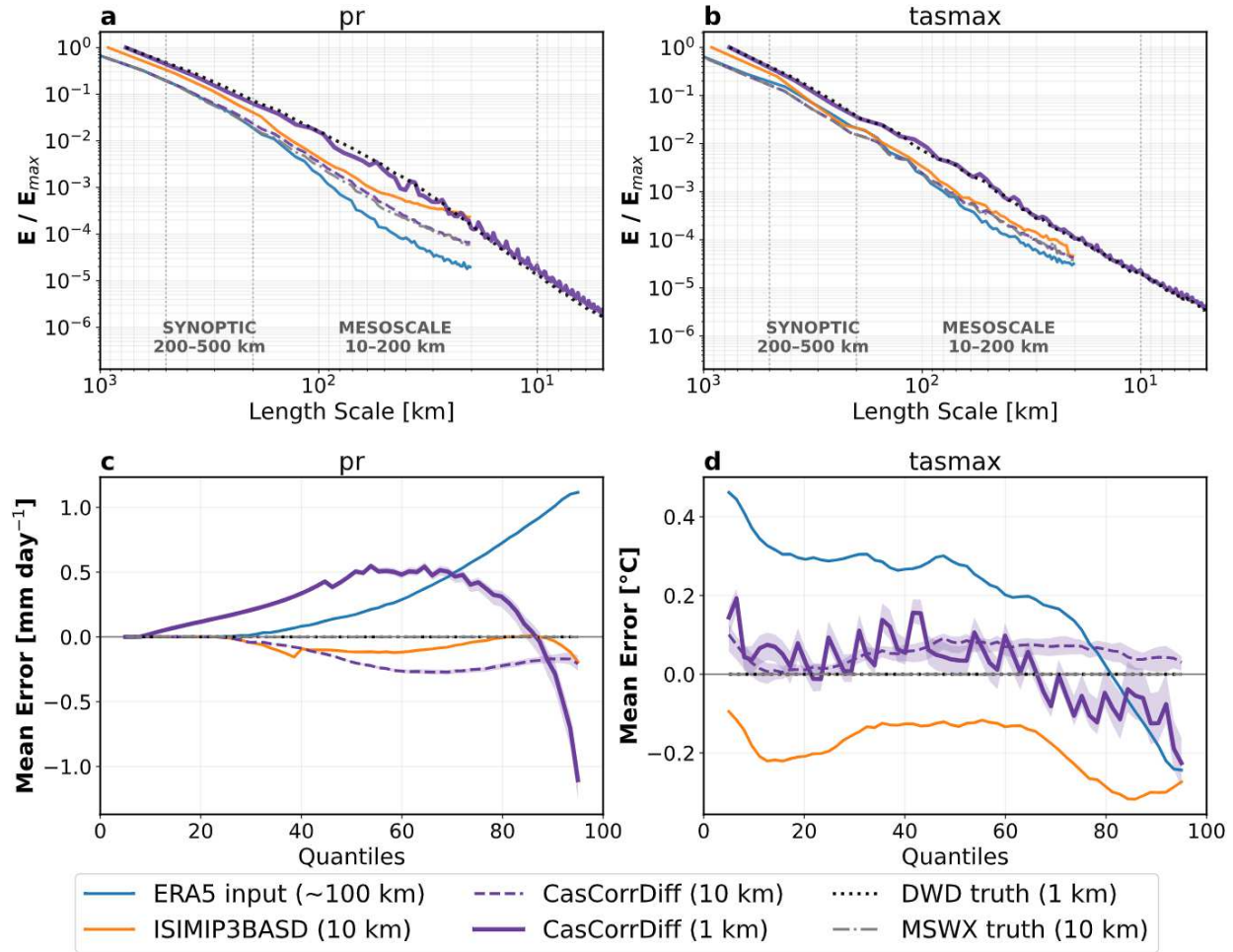


Figure 3. Scale-dependent spatial power density and distributional consistency. Spatial power density over Germany (2020–2024) for (a) precipitation and (b) maximum daily temperature. Spectra are computed via two-dimensional fast Fourier transforms (2D FFT) and radial averaging of spatial fields. Gray dotted lines delineate the mesoscale (10–200 km) and synoptic (200–500 km) regimes. Quantile-quantile mean error analysis for (c) precipitation and (d) maximum daily temperature. Differences between model and reference observations are evaluated across the 5th to 95th percentiles to capture broad distributions. Shaded regions denote the prediction interval for the eight ensembles. CasCorrDiff predictions (10 km and 1 km) are compared against ERA5 coarse reanalysis input (~100 km), ISIMIP3BASD statistical downscaling (10 km). Reference baselines are established using DWD for 1-km methods and MSWX for 10-km methods.

Spatio-temporal Representation of Climate Extremes

Distributional diagnostics above establish that the marginal distributions and scale-dependent spatial variance are well reproduced, but neither reveals whether extremes occur at right locations or how long they persist and co-occurrence (joint extreme events) are preserved, the characteristics that most directly govern climate impacts such as drought, water availability, heat stress and compound extremes. Hence an evaluation of the spatial-temporal representation of temperature and precipitation extremes directly, combining station-based metrics with gridded extreme indices is pertinent for establishing CasCorrDiff framework for impact assessments.

Daily maximum temperature, being continuous rather than event-based (e.g. precipitation), is evaluated with the Wasserstein distance, which estimates the discrepancy between the modelled and observed distributions as a temporal field at each station, in place of the event-based critical success index used for precipitation; this whole-distribution measure is complemented by the extreme bias, the difference between modelled and observed 95th-percentile daily maximum temperature. The spatial-mean Wasserstein distance is lowest for CasCorrDiff at 1 km ($W = 0.37$), whereas CasCorrDiff at 10 km is expectedly higher ($W=1.31$) showing that the distributional gain is realised specifically in the final 1 km cascade (Fig. 4a–d). ISIMIP3BASD is higher compared to CasCorrDiff at 10 km. Agreement is best over northwestern Germany, while southern Germany shows the largest discrepancies for hot extremes, likely reflecting complex topography and land–atmosphere interactions. All datasets underestimate high daily maximum temperature, but the 1 km prediction has by far the smallest extreme bias ($\text{Bias}_{95} = -0.24^\circ\text{C}$). The 10km CasCorrDiff prediction has extreme bias of -1.4°C which is better than -1.6°C for both ERA5 and ISIMIP3BASD (Fig. 4e–h). Regionally, these biases are smallest over northwestern Germany and strongest for hot days in the south, mirroring the Wasserstein pattern, and are computed over 2020–2024 against collocated GHCND station observations. The analogous station-based precipitation detection skill follows the same ordering (Fig. S8): the critical success index is highest for CasCorrDiff at 1 km (0.63) and declines through the CasCorrDiff 10 km, ERA5 and ISIMIP3BASD configurations (0.61–0.59), with the weakest wet-day detection and the strongest negative bias in extremes over the drier, more intermittent northeast; ERA5 shows the strongest underestimation of precipitation extremes there, whereas ISIMIP3BASD overcorrects them. CasCorrDiff at 1 km thus most accurately represents both the distribution and the extremes of daily maximum temperature, although hot days in southern Germany remain a persistent challenge.

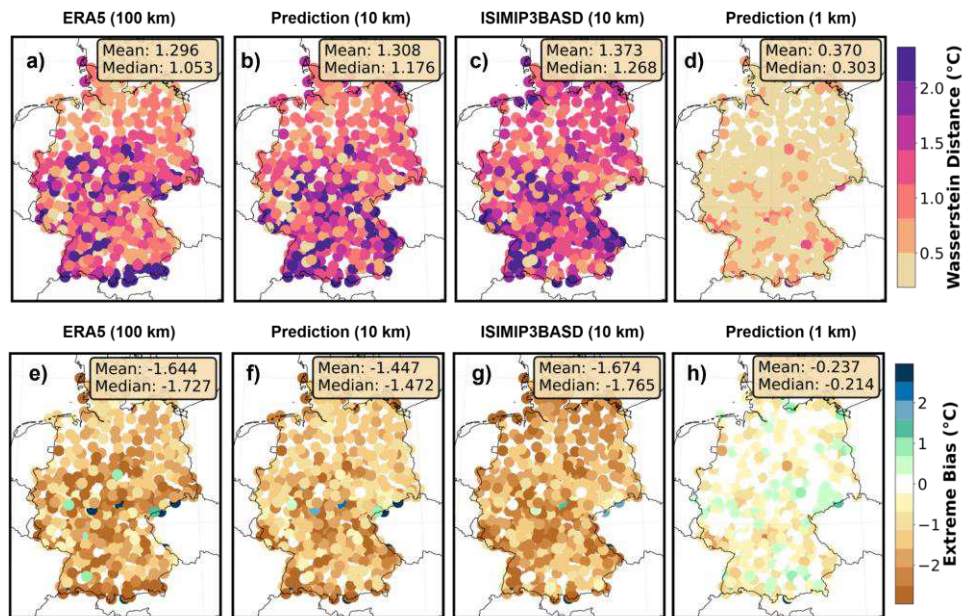


Figure 4. Spatial distribution of temporal metrics for daily maximum temperature across station locations. (a–d) Wasserstein distance ($^\circ\text{C}$), measuring the optimal transport distance between the modelled and observed distributions at each station; lower values indicate closer agreement between the distributions. (e–h) Extreme temperature bias, calculated as the difference between modelled and observed 95th percentile daily maximum temperature ($^\circ\text{C}$) for ERA5 (100 km; a,e), CasCorrDiff Prediction (10 km;

b,f), ISIMIP (10 km; c,g), and Prediction (1 km;); positive (negative) values indicate overestimation (underestimation) of warm extremes. Metrics are computed over the full evaluation period (2020–2024) using collocated GHCNd station observations.

To assess dry extreme conditions over a larger, continuous domain, we compare gridded fields against the MSWX reference at 10 km over 2020–2024, which enables evaluation beyond the German station network and the limited extent of the 1 km DWD HYRAS dataset. Maximum consecutive dry days (CDD; days with precipitation $< 1 \text{ mm day}^{-1}$) provide a spatially continuous event-persistence measure (Fig. 5). Observations show longer dry spells in northeastern Germany and shorter durations in the west and south (Fig. S7), consistent with the regional precipitation climatology. Bicubic interpolated ERA5 underestimates CDD across most regions, reflecting overly frequent precipitation that interrupts longer dry spells, whereas CasCorrDiff at 10 km reproduces the spatial distribution of CDD with biases comparable to the ISIMIP3BASD, though discrepancies remain in drier regions. The corresponding wet-day frequency, highest in western and southern Germany and declining toward the northeast, is broadly reproduced (Fig. S6), but ERA5 overestimates occurrence where observed frequencies are higher, while both downscaling methods overcorrect toward drier conditions (Fig. S9; CasCorrDiff, $-0.5 \text{ days yr}^{-1}$; ISIMIP3BASD, $-0.2 \text{ days yr}^{-1}$) and ISIMIP3BASD additionally shows locally patchy behaviour (Fig. S9g). Over northern Germany both downscaling methods jointly overestimate wet-day frequency and CDD, indicating an amplified distribution—excessively dry at low quantiles and overly wet at high quantiles—that may exacerbate the region's hydroclimatic contrasts, plausibly reflecting the strong scale dependence of precipitation extremes that is difficult to capture from 100 km conditioning alone. Their joint temperature–precipitation behaviour is likewise reproduced by CasCorrDiff (Fig. S10), which captures the observed co-variability more closely than the narrower ERA5 distribution or the moderately over-dispersed ISIMIP3BASD, though a residual underestimation of hot–wet conditions persist from ERA5.

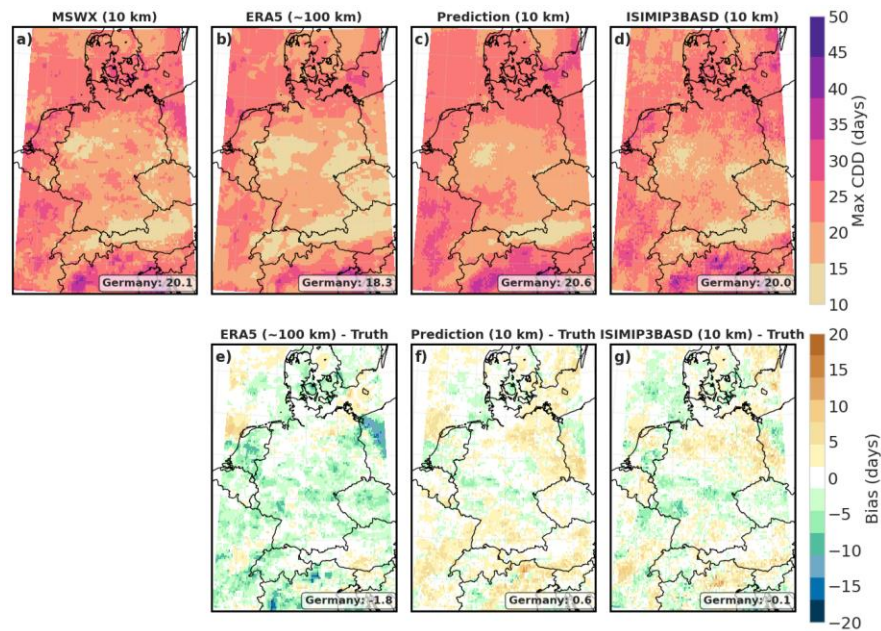


Figure 5. Spatial distribution of maximum Consecutive Dry Days (CDD; precipitation $< 1 \text{ mm/day}$) over Germany (2020–2024). The top row (a–d) displays absolute CDD values (days) for the MSWX reference, ERA5 reanalysis, CasCorrDiff prediction

(10 km), and ISIMIP3BASD (10 km). The bottom row (e–g) illustrates the model bias (days) relative to the MSWX reference. Area-weighted mean values for Germany are indicated in the bottom right of each panel.

Discussion

This work highlights contrasting philosophies between deep learning downscaling and traditional statistical bias-correction approaches, particularly for precipitation and daily maximum temperature, representing a major paradigm shift in regional climate downscaling. A central architectural advantage of CasCorrDiff is its cascading multi-step structure which enables massive resolution jumps possible. In this work we achieved a $100\times$ downscaling from coarse atmospheric inputs down to kilometre scale information. The evaluation reveals a clear functional division within the cascade: the intermediate 10 km stage performs the bulk of macro-scale systematic bias correction, while the 1 km stage functions as a conditional generator, injecting fine-scale spatial details without altering domain-averaged statistics. For daily precipitation, downscaling raw ERA5 input ($NSE = 0.284$) via CasCorrDiff to 10 km substantially improves daily synchronization, while the final 1 km refinement further elevates predictive skill ($NSE = 0.599$). For daily maximum temperature, the 1 km cascade achieves near-perfect station-scale alignment ($NSE \geq 0.941$; Figs. S2-3, Fig. 2).

The CasCorrDiff framework contrasts sharply with the statistical downscaling method, the ISIMIP3BASD method¹², which pairs coarse-scale quantile mapping with stochastic multipoint downscaling (e.g., matrix-based spatial rotation). While ISIMIP3BASD attempts to restore fine-scale variability^{18,29} during disaggregation, it struggles to maintain physical consistency across complex regional domains, leading to lower skill, most notably for summer daily precipitation ($NSE = 0.19$ vs. $NSE = 0.44$ for CasCorrDiff at 1 km). This architectural difference explains why ISIMIP3BASD posts competitive 1D precipitation quantile scores³⁰ while failing to reproduce spatial energy spectra: by fitting point-by-point distributions, it matches marginal statistics but leaves spatial variance across coarse neighbouring cells unconstrained. Operating on a large patch-based framework, CasCorrDiff integrates upper-air synoptic channels alongside high-resolution terrain features to capture mesoscale processes like orographic lift. As a learned stochastic model, it balances marginal corrections with realistic spatial variance, matching observed power spectra across mesoscale ranges. Beyond spatial fidelity, this framework alters operational demands: while ISIMIP3BASD requires multi-decade historical baselines to construct empirical cumulative distribution functions for every run, CasCorrDiff's inference phase is completely decoupled from historical time series, allowing it to downscale an isolated day on demand once trained. This behaviour is consistent with the role of diffusion-based models in generating high-resolution detail conditioned on coarser inputs^{25,31,32}.

While CasCorrDiff significantly outperforms standard linear interpolation and statistical baselines in capturing wet-day frequencies, dry-spell persistence, and day-to-day variability, persistent regional biases remain, particularly in topographically varied terrain and areas with higher precipitation intermittency. Furthermore, fine-scale precipitation extremes shift mechanistically across resolutions (from stratiform-dominated at coarse scales to convective-dominated at local scales), which can complicate station-scale extreme representations when coarse inputs lack the resolution to trigger mesoscale dynamics. In high-relief regions, both methods introduce a persistent negative wet-day bias that, when compounded with historical drying trends, risks exaggerating future drought projections and distorting downstream

hydrological and agricultural impact assessments. Finally, while CasCorrDiff handles joint temperature–precipitation interactions far better than raw driver data, accurately simulating rare compound extremes remains an ongoing challenge, underscoring the need for improved physically guided conditioning in future generative models.

Methods

Climate and topographic characteristics

Germany spans over a highly heterogeneous geographical area (357,000 km²) that serves as a useful testing ground for synoptic to kilometre-scale climate downscaling. Topographically we can divide Germany into three structural zones: the low-lying North German Plain, Central Uplands and the high-elevation Northern Alps along the southern border, rising above 2,500 m above sea level. This topographical gradient induces distinct regional climate regimes. The northern and northwestern regions experience a predominantly maritime climate characterized by mild winters and cool summers with year-round precipitation driven by Atlantic synoptic systems. Within this overarching climate, eastern Germany stretching from the Mecklenburg Lake District through Brandenburg to Saxony-Anhalt stands out as an increasingly dry macro-region (Fig. S7a). In this area, moisture-depleted westerly winds behind upland rain shadows combine with coarse, sandy soils to drive annual rainfall below 550 mm. Progressing south-eastward, the climate transitions into a more continental regime with higher seasonal temperature ranges and summer convective storms. The complex terrain of the Central Uplands and the Alps further creates strong localised microclimates governed by sharp elevation-dependent lapse rates and orographic effects. Resolving this substantial subgrid-scale spatial variance is critical for agricultural and hydrological risk assessments, but it remains fundamentally unresolved within the coarse horizontal grids of conventional global climate models.

Observational baseline and independent validation networks

To map these kilometre-scale climatic structures, this study relies on high-resolution gridded observations provided by the Deutscher Wetterdienst (DWD). The primary target dataset is the HYRAS meteorological reference archive, which provides daily, 1 km gridded surface fields across Germany from 1951 to the present³³. HYRAS covers near-surface air temperature (mean, maximum, minimum), total accumulated precipitation, downwelling shortwave solar radiation, and relative humidity. European-wide MSWX³⁴ dataset at 10km allows for downscaling for the intermediate stage. To evaluate model performance without grid-interpolation artefacts, we utilize independent point observations from NOAA’s Global Historical Climatology Network Daily (GHCNd) database³⁵. These land-based station records provide a true point-scale baseline to assess how well each downscaling framework reproduces local sub-grid stochastic variance and extreme values. See section S1 for further details on datasets.

Stage 1 downscaling: synoptic to mesoscale (100 km → 10 km)

The first step of the cascading framework learns the transition from large synoptic-scale circulations down to intermediate mesoscale weather patterns over the historical period 1989–2017.

The predictor inputs for Stage 1 consist of large-scale atmospheric fields extracted from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis³⁶. To emulate the structural scale of typical global climate models from the Coupled Model Intercomparison Project (CMIP), the native ~31 km ERA5 fields are line-interpolated to a standardized 1° grid (~100 km). The daily aggregated inputs provided to the Stage 1 deterministic U-Net include near-surface variables: total precipitation (mm day^{-1}), downwelling shortwave solar radiation (W m^{-2}), relative humidity (%), and daily mean, maximum, and minimum near-surface air temperatures (K).

To constrain the network with upper-air dynamics known to drive regional weather extremes over Central Europe, these surface variables are augmented with multi-level dynamic and thermodynamic predictors at the 850, 700, 500, and 200 hPa pressure surfaces, specifically: zonal (u) and meridional (v) wind vectors, air temperature, specific humidity, and geopotential height³⁷ (Table S1). Offline land-surface reanalysis variants, such as ERA5-Land, are purposely excluded to avoid inheriting pre-existing model dependencies or land-surface parameterization biases³⁸.

The corresponding output of Stage 1 is a 10 km (0.1°) resolution multi-variable weather field. The training target for this stage is the MSWX dataset. By integrating high-resolution satellite-derived microwave and infrared records with gauge observations, MSWX provides an accurate representation of intermediate mesoscale structures and convective event envelopes, acting as the bridge in our spatial cascade.

Stage 2 downscaling: mesoscale to local scale (10 km → 1 km)

The second stage refines the spatial structure from the intermediate 10 km mesoscale down to the final 1 km local scale, focusing on resolving topographically driven microclimates.

The inputs provided to Stage 2 consist of the 10 km meteorological fields output from Stage 1 (or matching historical 10 km MSWX fields during training). To guide the localised mapping, these atmospheric drivers are tightly coupled with high-resolution static boundary features, including a 1 km Digital Elevation Model (DEM) and derived land-surface characteristics (slope, aspect, and vegetation fractional cover). The inclusion of these boundary conditions allows the Stage 2 deterministic U-Net to learn localised physical relationships, such as how regional air masses interact with mountain ridges and valley networks.

The corresponding target output of Stage 2 is the 1 km gridded DWD HYRAS dataset. By separating the learning task into these two distinct stages, Stage 1 isolates the mapping of fluid atmospheric dynamics across scales, while Stage 2 specializes in anchoring those intermediate weather fields to fixed, high-resolution terrestrial characteristics.

Cascading correction diffusion model formulation

To resolve sub-grid scale uncertainty and prevent the spatial smoothing typical of standard neural network regressions, the Cascading Correction Diffusion model (CasCorrDiff) integrates a stochastic component at each stage using a residual learning framework^{25,28,39}.

At each downscaling tier, the coarser input field is upsampled to the target spatial grid using bilinear interpolation for continuous variables, and first-order conservative remapping for precipitation to maintain mass conservation across scales. The deterministic U-Net first estimates the conditional mean state of the target variable from these upsampled inputs.

A second, stochastic U-Net is then embedded within an Elucidated Diffusion Model (EDM) framework⁴⁰ to model only the residual difference (x_0) between the deterministic conditional mean estimate and the true high-resolution observations. By isolating sub-grid stochastic variance, this formulation simplifies the loss landscape and forces the generative model to focus entirely on synthesising missing fine-scale spatial textures and turbulent fields.

Within each cascading block, the high-resolution conditional residual distribution $p(x_0|y)$ is resolved through a continuous-time score-based generative framework^{23,24}. The forward diffusion process systematically perturbs the target high-resolution residual field into isotropic Gaussian noise over a continuous time horizon $t \in [0, T]$, governed by the stochastic differential equation:

$$dx = f(t)xdt + g(t)dw$$

where $f(t)$ represents the drift coefficient, $g(t)$ denotes the diffusion coefficient, and w represents a standard multi-dimensional Wiener process. Generative synthesis is achieved by integrating the corresponding reverse-time SDE from the noise horizon T back to the physical data state 0:

$$dx = [f(t)x - g(t)^2 \nabla_x \log p_t(x|y)]dt + g(t)d\bar{w}$$

where $\bar{w}$ represents a backward Wiener process. The score function, $\nabla_x \log p_t(x|y)$, describes the gradient of the log-probability density at time step t and is approximated by our stochastic U-Net (ϵ_θ), optimized via a conditional denoising score matching objective:

$$\min_{\theta} \mathbb{E}_{t, x_0, \epsilon} [\lambda(t) \|\epsilon_\theta(x_t, t, y) - \epsilon\|^2]$$

where $\epsilon \sim \mathcal{N}(0, I)$, and $\lambda(t)$ is a time-dependent weighting profile optimized to balance sample quality across diverse spatial frequencies⁴⁰. The reverse SDE is solved using a second-order Heun-based numerical sampler over 35 steps, minimizing the number of necessary network evaluations while preventing structural blur.

The network models are trained on historical data spanning 1989–2017. To ensure robust evaluation under shifting climate conditions, we apply a strict chronological data split: 2018–2019 data are used exclusively for validation loss monitoring and hyperparameter tuning, while 2020–2024 data are withheld as a fully independent test set²⁰.

The U-Net architectural backbone

The core computational building block for both the deterministic and stochastic modules in this study is a two-dimensional U-Net architecture⁴¹. This encoder-decoder framework with skip connections is inherently optimized for geophysical downscaling tasks because its symmetric hierarchical structure allows

the network to process large-scale contextual atmospheric patterns while simultaneously preserving pixel-aligned spatial details.

The down-sampling path (encoder) progressively reduces the spatial dimensions of the input fields through a series of convolutional layers while increasing feature channel density. Each encoder block consists of 3×3 convolutions, group normalization, and sigmoid-weighted linear unit (SiLU) activations, supplemented by self-attention layers at lower spatial resolutions to capture non-local teleconnections. This compression path isolates abstract, large-scale synoptic features such as pressure gradients and regional thermal advection.

The up-sampling path (decoder) systematically mirrors the encoder, progressively reconstructing the target high-resolution spatial grid. At each up-sampling step, high-resolution feature maps from the encoder are transferred directly to the decoder via skip connections. These connections bypass the bottlenecks of the network, preventing the loss of fine-grained spatial information.

For the deterministic module of our framework, the U-Net takes coarse atmospheric fields alongside static surface boundary features as inputs, and outputs a single, smooth deterministic prediction representing the conditional mean state of the target climate variable. For the stochastic module, the U-Net is modified to accept a noisy residual field and a scalar time-step embedding as additional inputs, outputting a denoising score vector that directs the generation of sub-grid atmospheric textures.

Benchmarking and statistical verification metrics

To benchmark CasCorrDiff, we implement the Inter-Sectoral Impact Model Intercomparison Project framework (ISIMIP3BASD)¹². Unlike our generative approach, ISIMIP3BASD sequentially decouples bias correction and spatial downscaling. First, parametric quantile mapping is performed at the native resolution of the climate input to adjust systematic distributional errors relative to the observations while strictly preserving the underlying long-term climate change trend. For a selected variable x , the trend-preserving mapping is formalised as:

$$x_{\text{adj}} = G^{-1}(F_{\text{obs}}(x))$$

where F and G represent the cumulative distribution functions (CDFs) of the model and observational baselines, respectively. Variable-specific distribution profiles are selected to enforce physical boundaries (e.g., Gamma for precipitation, Gaussian for temperature/radiation, and Beta for bounded relative humidity). To enforce multi-scale physical mass balance, a spatial aggregation constraint is explicitly locked:

$$\langle x_{\text{down}} \rangle = x_{\text{coarse}}$$

where $\langle \cdot \rangle$ indicates the spatial average over the coarse grid domain. Spatial downscaling to the targeted 10 km grid is subsequently completed via a multivariate bias-correction technique (Cannon, 2018), which iteratively applies orthogonal random rotations in the spatial space combined with independent univariate quantile mapping.

Model evaluation is conducted over the independent test window (2020–2024) across all 800+ stations. A minimum threshold of 10 valid paired daily observations is required per station to compute the metrics.

To evaluate spatial intermittency and diagnose spatial smoothing, we compute the Radially Averaged Power Spectral Density (RAPSD)²¹. The 2D power spectrum of a spatial field $X(x, y)$ is derived via discrete Fourier coefficients:

$$P(k_x, k_y) = \frac{1}{MN} |\hat{X}(k_x, k_y)|^2$$

where M and N represent the grid boundaries. The 2D power spectrum is azimuthally integrated to provide a 1D scale-dependent energy distribution:

$$\text{RAPSD}(k_r) = \frac{1}{2\pi} \int_0^{2\pi} P(k_r \cos \phi, k_r \sin \phi) d\phi$$

where $k_r = \sqrt{k_x^2 + k_y^2}$ is the radial wavenumber.

Temporal synchronization and continuous grid skill are validated using the Nash–Sutcliffe Efficiency (Nash and Sutcliffe, 1970):

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (x_{\text{obs},i} - x_{\text{pred},i})^2}{\sum_{i=1}^n (x_{\text{obs},i} - \bar{x}_{\text{obs}})^2}$$

where $x_{\text{obs},i}$ and $x_{\text{pred},i}$ represent observed and downscaled daily series, and $\bar{x}_{\text{obs}}$ is the stationary station mean.

Full distributional alignment across the entire range of values is quantified via the first-order Wasserstein distance:

$$W_1(F, G) = \int_{-\infty}^{\infty} |F(x) - G(x)| dx$$

where F and G are the empirical CDFs of the observations and predictions. W_1 maintains the physical unit of the variable under assessment.

Precipitation state occurrence and spatial displacement fidelity are validated using the Critical Success Index⁴², calculated using a binary wet-day threshold ($\geq 1.0 \text{ mm day}^{-1}$):

$$\text{CSI} = \frac{\text{Hits}}{\text{Hits} + \text{Misses} + \text{False Alarms}}$$

The framework's extreme verification capacity is isolated at the 95th percentile (Bias_{p95}):

$$\text{Bias}_{p95} = x_{\text{pred},p95} - x_{\text{obs},p95}$$

where $x_{\text{pred},p95}$ and $x_{\text{obs},p95}$ are the 95th percentile thresholds of the downscaled and observed distributions, respectively.

Code Availability

The CasCorrDiff code is available open-source on GitHub (<https://github.com/Kaushikreddym/CasCorrDiff>) and archived on Zenodo (DOI: 10.5281/zenodo.19879687). The software is licensed under the MIT License. The datasets required for training are downloaded, cleaned, and post-processed using “climdata” python package (<https://kaushikreddym.github.io/climdata/>). The ISIMIP3BASD method codes are available from Lange et al., (2019) and vendorized in climdata.

Data Availability

All primary datasets used in this study are publicly available from their respective open repositories: ERA5 reanalysis from the ECMWF Copernicus Climate Data Store, MSWX from GloH2O, HYRAS reference grids from the Deutscher Wetterdienst (DWD) Climate Data Center, GHCNd station records from NOAA NCEI (see Table S1 for download links).

Acknowledgement

This research is supported by German Research Foundation (DFG) - “FAIRagro” project number 501899475. We gratefully acknowledge the financial support of the Leibniz Association under LL-SYSTAIN (Grant Labs-2024-IOR). We would also like to acknowledge the project entitled "Supporting Agroecological Transitions of Agroforestry Landscapes using an Innovative Transdisciplinary Assessment Framework" (AGRECO4CAST). The project is Co-funded by the European Union: “AGROECOLOGY Call 1- AGRECO4CAST (ID132)”. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority (funding ID for Germany: 031B1610A) can be held responsible for them. We also acknowledge PARAM-VC project (funding ID for Germany: 01DQ25001A), under Indo-German Science & Technology Centre (IGSTC) 2+2 call, funded jointly by the Department of Science and Technology (DST) under the Government of India and the Federal Ministry of Research, Technology and Space of Germany (BMFTR), and “SUSTAINCrop” (HORIZON-EUROPE, funding ID: 0672EMSM).

References

1. Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M. & García-Herrera, R. The Hot Summer of 2010: Redrawing the Temperature Record Map of Europe. *Science* **332**, 220–224 (2011).
2. Seneviratne, S. I. *et al.* Weather and Climate Extreme Events in a Changing Climate. in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Masson-Delmotte, V. *et al.*) 1513–1766 (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021). doi:10.1017/9781009157896.013.
3. Dietz, S., Rising, J., Stoerk, T. & Wagner, G. Economic impacts of tipping points in the climate system. *J. Agric. Econ.* **118**, e2103081118 (2021).
4. Albrich, K., Rammer, W. & Seidl, R. Climate change causes critical transitions and irreversible alterations of mountain forests. *Glob. Change Biol.* **26**, 4013–4027 (2020).
5. Cook, B. I., Mankin, J. S. & Anchukaitis, K. J. Climate change and drought: From past to future. *Curr. Clim. Change Rep.* **4**, 164–179 (2018).
6. 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).
7. Raymond, C. *et al.* Understanding and managing connected extreme events. *Nat. Clim. Change* **10**, 611–621 (2020).

- 561 8. Zscheischler, J. *et al.* A typology of compound weather and climate events. *Nat. Rev.*
 562 *Earth Environ.* **1**, 333–347 (2020).
- 563 9. Maraun, D. *et al.* Precipitation downscaling under climate change: Recent
 564 developments to bridge the gap between dynamical models and the end user. *Rev.*
 565 *Geophys.* **48**, (2010).
- 566 10. Sippel, S. & Otto, F. E. L. Beyond climatological extremes - assessing how the odds of
 567 hydrometeorological extreme events in South-East Europe change in a warming
 568 climate. *Clim. Change* **125**, 381–398 (2014).
- 569 11. Poncet, N. *et al.* Does a convection-permitting regional climate model bring new
 570 perspectives on the projection of Mediterranean floods? *Nat. Hazards Earth Syst. Sci.*
 571 **24**, 1163–1183 (2024).
- 572 12. Lange, S. Trend-preserving bias adjustment and statistical downscaling with
 573 ISIMIP3BASD (v1.0). *Geosci. Model Dev.* **12**, 3055–3070 (2019).
- 574 13. Hempel, S., Frieler, K., Warszawski, L., Schewe, J. & Piontek, F. A trend-preserving bias
 575 correction – the ISI-MIP approach. *Earth Syst. Dyn.* **4**, 219–236 (2013).
- 576 14. Frieler, K. *et al.* Understanding the weather signal in national crop-yield variability.
 577 *Earths Future* **5**, 605–616 (2017).
- 578 15. Jägermeyr, J. *et al.* Climate impacts on global agriculture emerge earlier in new
 579 generation of climate and crop models. *Nat. Food* **2**, 873–885 (2021).
- 580 16. Menz, C. Bias Adjusted CORDEX-EUR11 Simulations. World Data Center for Climate
 581 (WDCC) at DKRZ <https://doi.org/10.26050/WDCC/BADCORD-EUR11> (2023).

- 582 17. Bechtold, M. *et al.* A global hourly ISIMIP3 climate forcing dataset for impact modeling.
 583 Preprint at <https://doi.org/10.5194/essd-2026-227> (2026).
- 584 18. Maraun, D. Bias Correction, Quantile Mapping, and Downscaling: Revisiting the
 585 Inflation Issue. *J. Clim.* **26**, 2137–2143 (2013).
- 586 19. François, B., Vrac, M., Cannon, A. J., Robin, Y. & Allard, D. Multivariate bias corrections
 587 of climate simulations: which benefits for which losses? *Earth Syst. Dyn.* **11**, 537–562
 588 (2020).
- 589 20. Reichstein, M. *et al.* Deep learning and process understanding for data-driven Earth
 590 system science. *Nature* **566**, 195–204 (2019).
- 591 21. Ravuri, S. *et al.* Skilful precipitation nowcasting using deep generative models of radar.
 592 *Nature* **597**, 672–677 (2021).
- 593 22. Hess, P., Lange, S., Schötz, C. & Boers, N. Deep Learning for Bias-Correcting CMIP6-
 594 Class Earth System Models. *Earths Future* **11**, e2023EF004002 (2023).
- 595 23. Ho, J., Jain, A. & Abbeel, P. Denoising Diffusion Probabilistic Models. Preprint at
 596 <https://doi.org/10.48550/ARXIV.2006.11239> (2020).
- 597 24. Song, W. *et al.* Convolutional Neural Network, Res-Unet++, -Based Dispersion Curve
 598 Picking From Noise Cross-Correlations. *J. Geophys. Res. Solid Earth* **126**,
 599 e2021JB022027 (2021).
- 600 25. Mardani, M. *et al.* Residual corrective diffusion modeling for km-scale atmospheric
 601 downscaling. *Commun. Earth Environ.* **6**, 124 (2025).

- 602 26. Bischoff, T. & Deck, K. Unpaired Downscaling of Fluid Flows with Diffusion Bridges.
 603 *Artif. Intell. Earth Syst.* <https://doi.org/10.1175/AIES-D-23-0039.1> (2024)
 604 doi:10.1175/AIES-D-23-0039.1.
- 605 27. Wan, Z. Y. *et al.* Debias Coarsely, Sample Conditionally: Statistical Downscaling
 606 through Optimal Transport and Probabilistic Diffusion Models. Preprint at
 607 <https://doi.org/10.48550/ARXIV.2305.15618> (2023).
- 608 28. Pathak, J. *et al.* Kilometer-scale convection-allowing model emulation using generative
 609 diffusion modeling. *Sci. Adv.* **12**, eadv0423 (2026).
- 610 29. Von Storch, H. On the Use of “Inflation” in Statistical Downscaling. *J. Clim.* **12**, 3505–
 611 3506 (1999).
- 612 30. Cannon, A. J., Sobie, S. R. & Murdock, T. Q. Bias Correction of GCM Precipitation by
 613 Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and
 614 Extremes? *J. Clim.* **28**, 6938–6959 (2015).
- 615 31. Addison, H., Kendon, E., Ravuri, S., Aitchison, L. & Watson, P. A. Machine learning
 616 emulation of a local-scale UK climate model. Preprint at
 617 <https://doi.org/10.48550/ARXIV.2211.16116> (2022).
- 618 32. Hatanaka, Y., Glaser, Y., Galgon, G., Torri, G. & Sadowski, P. Diffusion Models for High-
 619 Resolution Solar Forecasts. Preprint at <https://doi.org/10.48550/ARXIV.2302.00170>
 620 (2023).
- 621 33. Kaspar, F. *et al.* Monitoring of climate change in Germany – data, products and services
 622 of Germany’s National Climate Data Centre. *Adv. Sci. Res.* **10**, 99–106 (2013).

- 623 34. Beck, H. E. *et al.* MSWX: Global 3-Hourly 0.1° Bias-Corrected Meteorological Data
 624 Including Near-Real-Time Updates and Forecast Ensembles. *Bull. Am. Meteorol. Soc.*
 625 **103**, E710–E732 (2022).
- 626 35. Menne, M. J. *et al.* The Global Historical Climatology Network hourly (GHCNh) dataset.
 627 in vol. 105 446924 (2025).
- 628 36. Hersbach, H. *et al.* The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049
 629 (2020).
- 630 37. Shahi, N. K., Zolina, O. & Gulev, S. K. Large-Scale Drivers of Widespread Extreme
 631 Precipitation in Europe: Insights from High-Resolution Regional Climate Model
 632 Simulations. *J. Hydrometeorol.* **26**, 221–238 (2025).
- 633 38. Muñoz-Sabater, J. *et al.* ERA5-Land: a state-of-the-art global reanalysis dataset for land
 634 applications. *Earth Syst. Sci. Data* **13**, 4349–4383 (2021).
- 635 39. Rampal, N., Gibson, P. B., Sherwood, S., Abramowitz, G. & Hobeichi, S. A Reliable
 636 Generative Adversarial Network Approach for Climate Downscaling and Weather
 637 Generation. *J. Adv. Model. Earth Syst.* **17**, e2024MS004668 (2025).
- 638 40. Karras, T., Aittala, M., Aila, T. & Laine, S. Elucidating the Design Space of Diffusion-
 639 Based Generative Models. Preprint at <https://doi.org/10.48550/ARXIV.2206.00364>
 640 (2022).
- 641 41. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical
 642 Image Segmentation. in *Medical Image Computing and Computer-Assisted*
 643 *Intervention – MICCAI 2015* (eds Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.)
 644 vol. 9351 234–241 (Springer International Publishing, Cham, 2015).

- 645 42. Schaefer, J. T. The Critical Success Index as an Indicator of Warning Skill. *Weather*
646 *Forecast.* **5**, 570–575 (1990).
647

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SIDownscalingDiffusionISIMIP.docx](#)