Environmental Data Science (2026), 5: e25, 1–12
doi:10.1017/eds.2026.10046

CAMBRIDGE UNIVERSITY PRESS

APPLICATION PAPER

Super-resolving coarse-resolution weather forecasts with flow matching

Aymeric Delefosse[1], Anastase Charantonis[1,2] and Dominique Béréziat[3]

[1]ARCHES, INRIA, Paris, France
[2]SU, Inria, Centre Inria de Sorbonne Université, Paris, France
[3]CNRS, LIP6, Sorbonne Université, Paris, France
Corresponding author: Aymeric Delefosse; Email: aymeric.delefosse@inria.fr

Received: 30 May 2026 UTC; **Revised:** 30 May 2026 UTC; **Accepted:** 04 June 2026 UTC

Keywords: deep learning; downscaling; generative model; super-resolution; weather forecasting

Abstract

Machine learning-based weather forecasting models now surpass state-of-the-art numerical weather prediction systems, but training and operating these models at high spatial resolution remains computationally expensive. We present a modular framework that decouples forecasting from spatial resolution by applying learned generative super-resolution as a post-processing step to coarse-resolution forecast trajectories. We formulate super-resolution as a stochastic inverse problem, using a residual formulation to preserve large-scale structure while reconstructing unresolved variability. The model is trained with flow matching exclusively on reanalysis data and is applied to global medium-range forecasts. We evaluate (i) design consistency by re-coarsening super-resolved forecasts and comparing them to the original coarse trajectories and (ii) high-resolution forecast quality using standard ensemble verification metrics and spectral diagnostics. Results show that super-resolution preserves large-scale structure and variance after re-coarsening, introduces physically consistent small-scale variability, and achieves competitive probabilistic forecast skill at 0.25° resolution relative to an operational ensemble baseline while requiring only a modest additional training cost compared with end-to-end high-resolution forecasting.

Impact Statement

This article shows how learned super-resolution can enhance the spatial detail of coarse-resolution weather forecasts without retraining forecasting models at operational resolution. The approach reduces computational requirements, making high-resolution weather products more accessible in research and operational settings with limited resources.

1. Introduction

In recent years, machine learning (ML)-based weather forecasting has advanced significantly, with models now matching or surpassing traditional numerical weather prediction (NWP) systems in forecast skill and inference efficiency (Bi et al., 2023; Lam et al., 2023; Lang et al., 2026). This progress is attributable to the development of increasingly powerful architectures and the availability of high-quality,

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 licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/eds.2026.10046 Published online by Cambridge University Press

multi-decadal reanalysis datasets, such as ERA5 (Hersbach et al., 2020). This new generation of models enables rapid generation of large ensembles, but training remains computationally demanding: multi-terabyte, high-resolution datasets impose severe data throughput and memory constraints, which typically require access to large-scale computational resources (Bonev et al., 2025).

At the same time, models trained at coarser spatial resolutions, where data volume and memory requirements are far more manageable, demonstrate that competitive forecast skill can still be achieved (Nguyen et al., 2024; Verma et al., 2024; Couairon et al., 2026). This reflects the scale dependence of atmospheric predictability: for medium-range forecasts, skill is primarily constrained by large-scale atmospheric dynamics (Anthes et al., 1985; Moore, 2023). However, this reduced-cost regime remains insufficient to resolve the spatial scales that govern regional weather. These are necessary to represent the organization and intensity of the phenomena that drive local impacts, including heavy precipitation, strong winds, and tropical cyclones.

This motivates exploring alternative strategies that decouple spatial resolution from model training cost. Inspired by advances in computer vision, a natural direction is to employ learned super-resolution (SR) methods (Dong et al., 2016; Ledig et al., 2017; Saharia et al., 2023), also commonly referred to as downscaling in the climate sciences. Rather than retraining or fine-tuning a forecasting model at higher resolution, SR aims to learn a mapping from coarse to fine spatial scales. This approach offers complementary advantages in modularity and efficiency: the forecasting and the downscaling models can be developed and optimized independently. In this framework, forecasting operates at low resolution, while super-resolution recovers subgrid-scale variability. So far, data-driven downscaling has primarily focused on regional climate applications (Watt and Mansfield, 2024; Hess et al., 2025; Mardani et al., 2025), typically operating on individual states and a limited set of target variables, without explicit consideration of forecast trajectories. Notable exceptions include SwinRDM (Chen et al., 2023), which combines a coarse-resolution recurrent forecast model with a diffusion-based upsampling step, demonstrating the viability of cascading forecasting and generative downscaling for high-resolution weather prediction. More recently, Climate in a Bottle (Brenowitz et al., 2025) integrates generative downscaling within a climate modeling pipeline to produce kilometer-scale forecasts.

In this work, we apply learned generative super-resolution to global medium-range forecast trajectories produced by coarse-resolution data-driven models. Using ArchesWeatherGen (Couairon et al., 2026) as a representative forecasting system, we assess whether super-resolution can reconstruct physically meaningful fine-scale structure while remaining consistent with the underlying large-scale dynamics, thereby supporting its feasibility as a modular and computationally affordable pathway toward higher-resolution forecasting. The remainder of the paper is structured as follows. We first introduce the problem formulation and the proposed super-resolution framework. We then describe the generative model and its application to forecast trajectories, followed by the experimental setup, evaluation methodology, and results. Finally, we discuss the implications, limitations, and possible directions for future work.

2. Methods

2.1. Problem formulation

We consider the problem of reconstructing a high-resolution atmospheric trajectory consistent with a given coarse-resolution forecast. Let $\mathbf{x}_t$ denote the global weather state at time t. In our framework, $\mathbf{x}_t$ consists of six upper-air variables (temperature, geopotential, specific humidity, wind components U, V, and W) sampled on a latitude-longitude grid at 13 pressure levels and four surface variables (2 m temperature, mean sea-level pressure, and 10 m wind components U and V).

2.1.1. Coarse-resolution forecasts

Let $\mathbf{x}_t^{\mathrm{LR}}$ denote a low-resolution (LR) atmospheric state at time t. A forecasting model f_θ is trained to predict the future state at a fixed lead time δ,

$$\widehat{\mathbf{x}}_{t+\delta}^{\mathrm{LR}} = f_\theta\left(\mathbf{x}_t^{\mathrm{LR}}\right), \tag{1}$$

https://doi.org/10.1017/eds.2026.10046 Published online by Cambridge University Press

where δ is set to $\delta = 24\,\mathrm{h}$ throughout this work. Forecast trajectories of length T are obtained by applying the model autoregressively,

$$\widehat{\mathbf{x}}^{\mathrm{LR}}_{t+k\delta} = f_\theta\left(\widehat{\mathbf{x}}^{\mathrm{LR}}_{t+(k-1)\delta}\right), 1 \le k \le T. \tag{2}$$

Our objective is to construct corresponding high-resolution (HR) forecasts $\left(\widehat{\mathbf{x}}^{\mathrm{HR}}_{t+k\delta}\right)_{1 \le k \le T}$.

2.1.2. The coarse-graining operator

The relationship between resolutions is formalized through a known regridding operator C. For any high-resolution state $\mathbf{x}^{\mathrm{HR}}_t$, the associated coarse representation is given by $\mathbf{x}^{\mathrm{LR}}_t = C\left(\mathbf{x}^{\mathrm{HR}}_t\right)$. By construction, C removes information associated with unresolved spatial scales and therefore defines a lossy, non-invertible mapping. The super-resolution problem can thus be viewed as learning an approximate inverse operator C^{-1}_ϕ such that $\widehat{\mathbf{x}}^{\mathrm{HR}}_t = C^{-1}_\phi\left(\mathbf{x}^{\mathrm{LR}}_t\right)$. A design objective is that reapplying the coarse-graining operator to the reconstructed high-resolution fields approximately recovers the original coarse state, $C\left(\widehat{\mathbf{x}}^{\mathrm{HR}}_t\right) \approx \mathbf{x}^{\mathrm{LR}}_t$. This reflects the intent to preserve the large-scale organization of the coarse state while enriching it with unresolved structure.

2.1.3. Residual formulation

To structure this inverse problem, we adopt a residual decomposition formally analogous to a Reynolds decomposition, in which a field is expressed as the sum of a resolved component and a fluctuating part (Reynolds, 1895). Let $\mathcal{I}$ denote a deterministic interpolation operator (e.g., bicubic) mapping coarse-resolution fields to the high-resolution grid, and define $\uparrow\mathbf{x}^{\mathrm{LR}}_t = \mathcal{I}\left(\mathbf{x}^{\mathrm{LR}}_t\right)$. By construction, this interpolated field preserves the large-scale structure of the coarse state but contains no genuine small-scale variability, that is, $C\left(\mathcal{I}\left(\mathbf{x}^{\mathrm{LR}}_t\right)\right) \approx \mathbf{x}^{\mathrm{LR}}_t$. The high-resolution state can then be written as $\mathbf{x}^{\mathrm{HR}}_t = \uparrow\mathbf{x}^{\mathrm{LR}}_t + \mathbf{r}_t$, where $\mathbf{r}$ represents the unresolved high-frequency residual. We train a neural network g_ϕ to predict this residual, $\widehat{\mathbf{r}}_t = g_\phi\left(\uparrow\mathbf{x}^{\mathrm{LR}}_t\right)$, yielding the final reconstruction

$$\widehat{\mathbf{x}}^{\mathrm{HR}}_t = \uparrow\mathbf{x}^{\mathrm{LR}}_t + \widehat{\mathbf{r}}_t. \tag{3}$$

This formulation concentrates model capacity on reconstructing fine-scale structure while explicitly preserving the large-scale information already present in the coarse field. Similar residual strategies have proven effective in climate downscaling (Li et al., 2024; Lopez-Gomez et al., 2025; Mardani et al., 2025).

2.1.4. Stochastic inverse modeling

From a physical standpoint, the inverse problem is inherently non-unique: a coarse forecast constrains the large-scale atmospheric flow but does not uniquely determine fine-scale structure. Subgrid variability admits many realizations compatible with the same resolved state. Accordingly, our objective is not to recover a single deterministic field but to learn a stochastic inverse mapping by modeling the conditional distribution of unresolved residuals, $p\left(\mathbf{r}_t \middle| \uparrow\mathbf{x}^{\mathrm{LR}}_t\right)$. Although we operate on forecast trajectories, we do not explicitly condition the inverse model on temporal history. At 24-hour lead times and 1.5° resolution, fine-scale organization is primarily constrained by the large-scale flow at that time, making additional temporal conditioning weakly informative while increasing model complexity. We empirically verify this in Supplementary Section C.2.

2.2. Super-resolution model

We adopt a conditional generative super-resolution approach inspired by SR3 (Saharia et al., 2023), adapted to the residual formulation and to global weather fields. Instead of the original SR3 U-Net backbone, we use a 3D Swin U-Net transformer architecture as in ArchesWeather (Couairon et al., 2026), which better scales to high-resolution fields (Crowson et al., 2024). To model the stochasticity of

https://doi.org/10.1017/eds.2026.10046 Published online by Cambridge University Press

unresolved scales, we train the model using flow matching (Lipman et al., 2023), learning the conditional distribution $p\left(\mathbf{r}_t \middle| \uparrow \mathbf{x}_t^{\mathrm{LR}}\right)$. Further architectural details are provided in Supplementary Section A.2.

2.2.1. *Application to forecasts*

We apply the super-resolution operator as a post-processing step to coarse-resolution forecast states for all lead times (Figure 1a). Given a coarse-resolution forecast trajectory $\left(\widehat{\mathbf{x}}_{t+k\delta}^{\mathrm{LR}}\right)_{1 \le k \le T}$, the corresponding high-resolution fields are obtained independently at each lead time as

$$\widehat{\mathbf{x}}_{t+k\delta}^{\mathrm{HR}} = C_\phi^{-1}\left(\widehat{\mathbf{x}}_{t+k\delta}^{\mathrm{LR}}\right), 1 \le k \le T. \tag{4}$$

In the following, we consider ArchesWeatherGen as the coarse-resolution forecasting framework, which is augmented by the proposed super-resolution approach.

3. Experiments

3.1. *Training and evaluation data*

Training pairs $(\mathbf{x}^{\mathrm{LR}}, \mathbf{x}^{\mathrm{HR}})$ are constructed by regridding ERA5 data at 0.25° resolution to 1.5° using the first-order conservative scheme. This remapping ensures that fluxes remapped to the coarse grid represent the same net energy and mass exchange as on the fine grid (Jones, 1999). This setup corresponds to a 6 × downscaling task in both latitude and longitude. We use data from 1979 to 2018 for training, 2019 for validation, and 2020 for testing.

3.1.1. *Zero-shot application to forecasts*

Although the super-resolution model is trained exclusively on reanalysis pairs, we apply it to model forecast outputs in a zero-shot manner. This is possible because ArchesWeatherGen is trained to sample physically plausible coarse-resolution trajectories from the ERA5 data distribution. ArchesWeatherGen produces stochastic ensemble members such that multiple forecast realizations correspond to the same reanalysis state. As a result, no unique supervised mapping from forecast outputs to ERA5 exists, whereas training exclusively on ERA5-based pairs yields a well-posed inverse regridding task that is independent of forecast model errors.

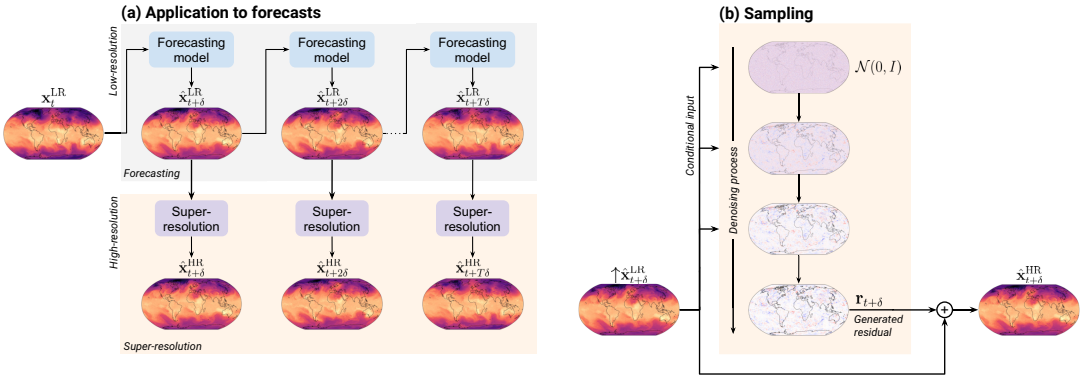


Figure 1. *Overview of the super-resolution procedure. (a) Super-resolution is applied independently to each coarse-resolution forecast state produced by the base forecasting model. (b) Sampling procedure: starting from Gaussian noise, a flow-matching model generates a high-resolution residual conditioned on the interpolated coarse-resolution forecast, which is added to obtain the final high-resolution prediction.*

https://doi.org/10.1017/eds.2026.10046 Published online by Cambridge University Press

3.2. Evaluation methodology

Our experimental evaluation is organized around two complementary objectives: (i) validating the consistency of the super-resolution operator with respect to the underlying coarse-resolution forecasts and (ii) assessing the quality of the resulting high-resolution forecasts in a standard benchmarking setting. In both sets of experiments, each high-resolution ensemble member is obtained by super-resolving the corresponding low-resolution ArchesWeatherGen trajectory.

3.2.1. Design validation

We first assess the impact of super-resolution on coarse-resolution forecasts produced by ArchesWeatherGen at 1.5° resolution. To this end, we compare re-coarsened super-resolved forecasts against the original ArchesWeatherGen trajectories using pattern correlation, activity ratio, and normalized RMSE. Re-coarsening is performed using the same conservative regridding operator C used to construct the coarse-resolution training data. Pattern correlation is computed as the Pearson correlation coefficient over the latitude-longitude grid, while the activity ratio and normalized RMSE quantify changes in resolved-scale variance and amplitude, respectively. Precise definitions of each metric are provided in Supplementary Section B.1.

3.2.2. Forecasting benchmark

In a second set of experiments, we evaluate the quality of super-resolved forecasts within a standard global weather forecasting framework. Following the WeatherBench 2 protocol (Rasp et al., 2024) and the experimental setup of ArchesWeatherGen, we compare post-processed high-resolution forecasts against established baselines, including IFS ENS (ECMWF, 2019), GenCast (Price et al., 2025), and a naive bicubic interpolation of coarse-resolution ArchesWeatherGen forecasts to the target resolution, which serves as the no-added-value baseline. The exact formula for each metric is presented in Supplementary Section B.2.

4. Results

4.1. Design validation

We assess the consistency of the super-resolution design by comparing re-coarsened super-resolved forecasts to the original coarse-resolution trajectories produced by ArchesWeatherGen at a representative lead time of 1 day (Figure 2). This table-based view allows a broader coverage of variables and pressure levels while focusing on a representative short-range lead time. Across all considered surface and upper-air variables, spatial correlations remain extremely close to unity, indicating that the large-scale spatial structure is preserved after super-resolution and re-coarsening. Activity ratios stay near one, ruling out both spurious variance injection and excessive smoothing. Normalized RMSE values remain uniformly small, confirming that reinjecting super-resolved states induces only minor, well-controlled deviations from the original coarse trajectories. Taken together, these diagnostics demonstrate that the proposed super-resolution approach leaves the large-scale state essentially unchanged while introducing only limited and physically consistent subgrid-scale modifications.

4.2. Quality of super-resolved forecasts

We now evaluate the quality of super-resolved forecasts using standard ensemble verification metrics as well as physical consistency diagnostics.

4.2.1. Summary ensemble metrics

Figure 3 summarizes the performance of the different models using key ensemble metrics, computed in their fair versions: Ensemble Mean RMSE, CRPS, Energy Score (ES), Brier Score, and the spread-skill ratio. Brier Scores are evaluated at symmetric distribution tails by averaging scores at the corresponding lower and upper quantile thresholds (e.g., 1% and 99%). All metrics are expressed as relative improvements with respect to IFS ENS. Scores are averaged over both headline upper-air variables (Z500, Q700, T850, U850, V850) and the four surface variables (T2M, SP, U10M, V10M).

https://doi.org/10.1017/eds.2026.10046 Published online by Cambridge University Press

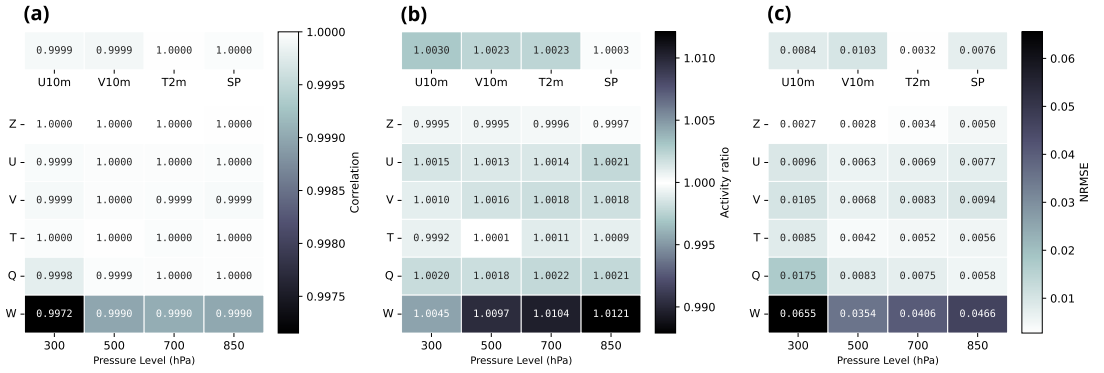


Figure 2. *Comparison of re-coarsened super-resolved forecasts with the original ArchesWeatherGen coarse-resolution trajectories at 1-day lead time. Shown are (a) spatial correlation, (b) activity ratio, and (c) normalized RMSE, for surface variables and selected pressure levels. Color intensity encodes the deviation from the ideal reference: values approaching white indicate near-perfect agreement (unity for correlation and activity ratio, zero for NRMSE), while darker shades correspond to larger departures from the original coarse-resolution trajectories.*

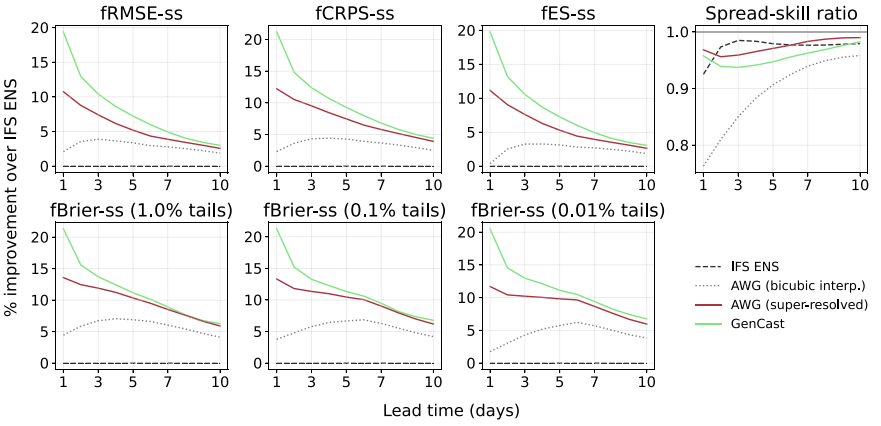


Figure 3. *Global ensemble forecast skill at 0.25°. Relative improvement over IFS ENS for fair (f) ensemble skill-score (ss) metrics (Ensemble Mean RMSE, CRPS, Energy Score, Brier Score, spread-skill ratio), averaged over WeatherBench headline variables.*

Overall, GenCast exhibits the strongest average performance at short to medium lead times (1–7 days). However, ArchesWeatherGen remains competitive and shows comparable skill at longer lead times (7–10 days), particularly for wind components and specific humidity. This behavior is consistent with the increasing dominance of large-scale dynamics at extended lead times. In addition, ArchesWeatherGen displays a more favorable spread-skill ratio across all lead times. While the ensemble remains slightly under-dispersive, this under-dispersion is less pronounced than for GenCast.

4.2.2. Per-variable CRPS analysis

While the previous results focus on metrics averaged across variables, we now examine performance at the level of individual physical variables. Figure 4 reports CRPS skill scores by variable and lead time. We focus on the CRPS, which assesses the quality of the predicted marginal distributions and, when

https://doi.org/10.1017/eds.2026.10046 Published online by Cambridge University Press

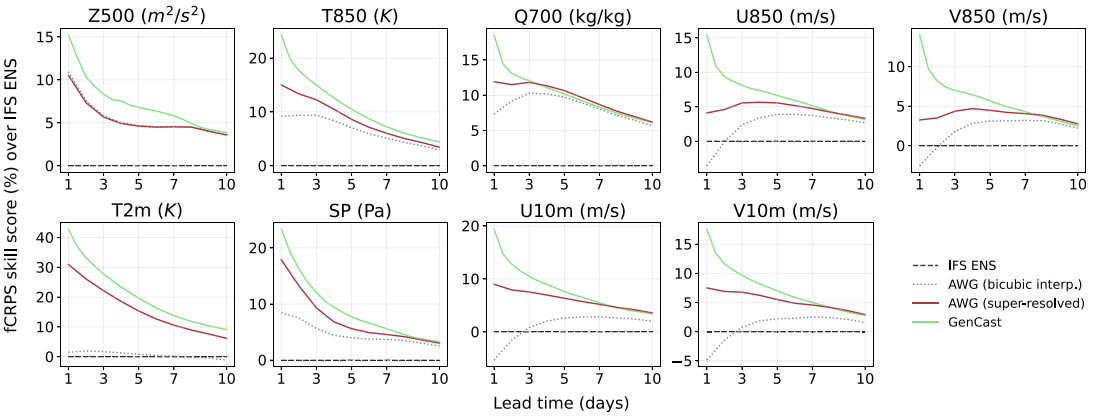


Figure 4. *Per-variable fair CRPS skill scores relative to IFS ENS at 0.25°. Results are shown for GenCast, and ArchesWeatherGen (AWG) after bicubic interpolation and learned super-resolution, as a function of lead time.*

evaluated over long lead times, implicitly penalizes physically inconsistent trajectories in autoregressive models.

ArchesWeatherGen achieves better CRPS skill scores than IFS ENS across all headline variables and lead times by 7.41% on average. Compared to GenCast, ArchesWeatherGen exhibits slightly worse performance for lead times of 1–6 days, particularly for wind variables, but achieves similar or better CRPS scores at lead times of 7–10 days. A notable exception is specific humidity, for which ArchesWeatherGen converges to GenCast-level performance at earlier lead times (3–10 days). Relative to bicubic interpolation, the super-resolution approach yields substantially improved probabilistic skill across nearly all variables, with the exception of geopotential. For this variable, bicubic interpolation performs comparatively well due to its intrinsically smooth and large-scale nature. This contrast highlights that the gains from learned super-resolution are most pronounced for variables with richer small-scale variability. We further provide a comparison of fair CRPS skill scores at 1.5° and 0.25° resolutions in Supplementary Section C.3.

4.2.3. Physical consistency and spectral properties

Beyond pointwise probabilistic scores, we assess the physical realism of the generated forecasts through power spectral analysis (Figure 5). Power spectra are computed for selected variables at 1-day and 10-day lead times, with energy averaged across ensemble members. The dashed vertical line marks the transition between scales resolved by the low-resolution forecast (1.5°) and those introduced by super-resolution (0.25°). Both generative models recover substantially more fine-scale energy than bicubic interpolation, demonstrating improved physical consistency at small spatial scales. ArchesWeatherGen exhibits a slightly slower decay of energy at the finest resolved scales compared to GenCast, indicating a better recovery of small-scale variability, although the differences remain moderate. Overall, the spectral analysis confirms that super-resolved forecasts preserve large-scale structure while consistently recovering fine-scale variability.

4.3. Case study: Hurricane Teddy

To complement the quantitative evaluation, we present a case study of Hurricane Teddy, a major Atlantic storm that reached Category 4 intensity in September 2020. It produced one of the most organized deep convective structures of that season. At its peak, Teddy exhibited a well-defined eye, vigorous spiral rainbands, and a broad moisture envelope that spanned much of the subtropical North

https://doi.org/10.1017/eds.2026.10046 Published online by Cambridge University Press

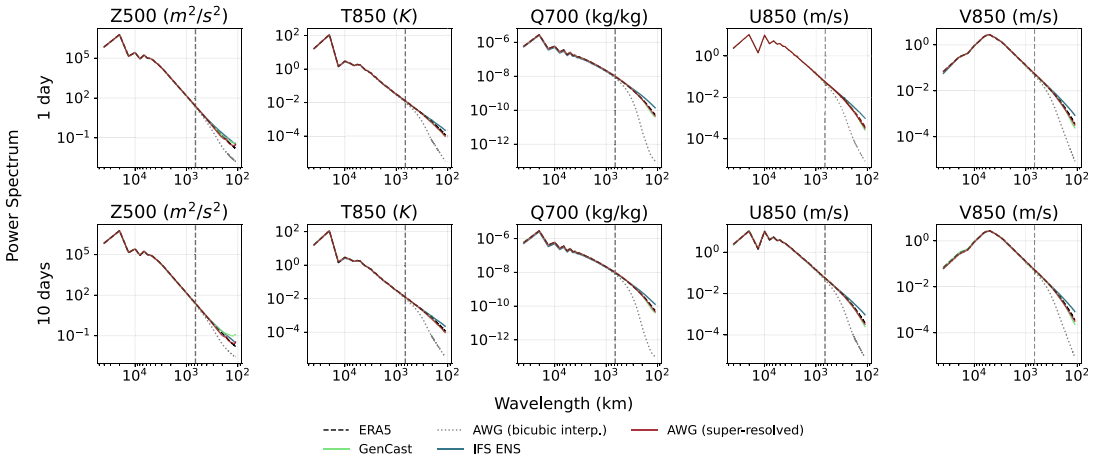


Figure 5. *Comparison of power spectra across models and lead times. Power spectra of Z500, T850, Q700, U850, and V850 at 1-day (top) and 10-day (bottom) lead times. For each model and wavelength, the energy at that wavelength is averaged across samples. The dashed vertical line marks the transition between low- and high-resolution scales (1.5° to 0.25°).*

Atlantic. Figure 6 showcases ArchesWeatherGen super-resolved humidity fields at a 1-day lead time, initialized on September 20, 2020.

At 1.5° resolution, the low-resolution ensemble members (b–d) accurately capture the large-scale moisture distribution and the hurricane's position, aligning closely with the ERA5 reference (a). Ensemble spread is visible in the representation of the cyclone core: it is less well-defined in the worst-performing member (d) than in the median (c) and best (b) members, while the synoptic-scale flow and storm placement remain consistent across all members. Super-resolution is then applied to each low-resolution member (f–n). At 0.25°, the SR model reconstructs fine-scale structures, including cyclonic moisture spirals, indications of a dry intrusion on the western flank, and sharper outer gradients. For each member, multiple stochastic draws produce distinct realizations that differ in the positioning of small-scale moisture filaments while preserving the large-scale organization inherited from the conditioning low-resolution state. Importantly, this behavior holds consistently across lead times: Supplementary Section C.5 shows that analogous fine-scale structures are recovered at 7-day lead time, demonstrating that the SR model produces physically plausible reconstructions throughout the forecast horizon. It should be noted that such stochastic variability does not necessarily improve forecast skill. Since the SR model is conditioned on the low-resolution input, large-scale errors are not corrected and propagate to the super-resolved outputs. Probabilistic scores may, therefore, penalize the ensemble if it is misaligned with the reference, despite exhibiting realistic fine-scale variability. Nevertheless, by representing physically plausible subgrid structures, super-resolution enables the exploration of internal variability and provides a more complete depiction of uncertainty at fine scales, which is particularly relevant for intense systems such as tropical cyclones. To further quantitatively evaluate the model's capacity beyond open-ocean dynamics, we extend the analysis to a topographically complex region during the passage of Storm Alex over the Alps (Supplementary Section C.6).

5. Discussion

We investigated learned super-resolution as a modular extension to global weather forecasting models trained at coarse spatial resolution. By decoupling forecasting and downscaling, our approach enables the reconstruction of physically plausible fine-scale structure while preserving the large-scale dynamics learned at low resolution, without retraining forecasting models at operational resolution. Despite its

https://doi.org/10.1017/eds.2026.10046 Published online by Cambridge University Press

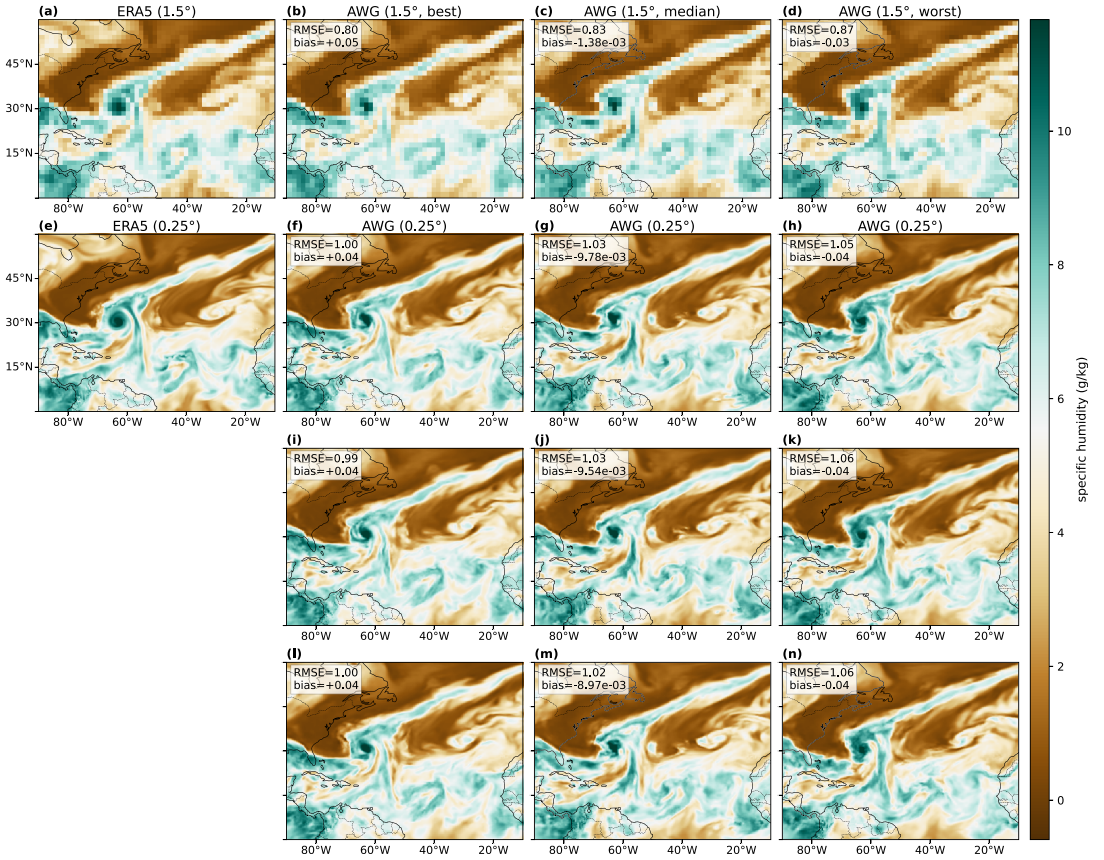


Figure 6. *Super-resolution of specific humidity at 700 hPa over the North Atlantic (Hurricane Teddy, 21 September 2020 12:00 UTC, +24 h lead time). ERA5 regridded to 1.5° (a) and at native 0.25° (e) serve as low- and high-resolution ground truth, respectively. AWG ensemble forecasts at 1.5° are shown for the best, median, and worst members by RMSE (b–d). Each column of the high-resolution block (f,i,l), (g,j,m), (h,k,n) shows three independent stochastic SR draws at 0.25° for the corresponding LR member. RMSE and bias (g kg^{-1}) against the appropriate ERA5 reference are reported in the top-left corner of each panel.*

modest computational cost, this strategy achieves competitive forecast performance relative to substantially more expensive end-to-end high-resolution forecasting models.

5.1. Generality and scope

Although demonstrated using the ArchesWeatherGen framework, the proposed approach is not tied to a specific forecasting architecture. In principle, it can be applied to any forecasting model provided that (i) its coarse-resolution outputs represent physically realistic large-scale flow and (ii) its temporal sampling is consistent with the dynamics it resolves. In this work, we focused on daily (24 h) forecast trajectories, a setting in which the method reliably reconstructs fine-scale spatial structure while preserving the underlying coarse-resolution forecast trajectory at the resolved time scale. Extending such daily trajectories to sub-daily (e.g., hourly) outputs would require a dedicated temporal super-resolution component, which we leave for future work. Moreover, by building on probabilistic forecasts rather than deterministic forecasts, the approach naturally decouples large- and fine-scale uncertainties, allowing each to be modeled and sampled independently. Extending the method to deterministic, smoothed inputs would instead require learning a direct or unpaired mapping from forecast outputs to high-resolution reference data (Wan et al., 2023; Keisler et al., 2026).

https://doi.org/10.1017/eds.2026.10046 Published online by Cambridge University Press

5.2. Broader implications

Beyond medium-range weather forecasting, our framework naturally extends to sub-seasonal and climate applications, where large-scale dynamics are generally well represented but fine-scale variability remains under-resolved. In this context, super-resolution models could be trained on high-resolution climate simulations or reanalysis-derived datasets and applied in a zero-shot manner to coarse global climate models, as demonstrated in related work (Hess et al., 2025). A particularly promising avenue is the use of global or limited-area kilometer-scale models, such as ICON (Hohenegger et al., 2023), as sources of high-resolution reference data.

5.3. Efficiency and limitations

A key advantage of the proposed approach is computational efficiency. Training the full pipeline—deterministic forecasting, generative forecasting, and super-resolution—is on the order of 30 A100-days in total. Super-resolution training represents a relatively small fraction of this cost, requiring approximately 7 A100-days, compared to about 23 A100-days for training ArchesWeatherGen. At inference time, generating a full 10-day high-resolution forecast takes about 8 minutes on a single NVIDIA V100 32GB GPU, which can become significant when applied at scale, such as for long test sets or large ensembles. This computational cost motivates future work on speeding up inference through distillation or consistency-based generative modeling (Stock et al., 2025).

5.4. Conclusion

Overall, this work demonstrates that learned super-resolution provides a physically grounded and computationally efficient pathway toward higher-resolution weather and climate prediction. By separating large-scale prediction from fine-scale reconstruction, the proposed framework offers flexibility, interpretability, and scalability across forecasting horizons and application domains. Our results demonstrate that this approach achieves competitive forecast skill at operational resolution with substantially reduced computational cost and that recovered small-scale variability exhibits realistic spectral characteristics.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/eds.2026.10046.

Data availability statement. The ERA5 reanalysis data used in this study are publicly available from the Copernicus Climate Data Store (https://doi.org/10.24381/cds.adbb2d47). The code used in this study is available at https://doi.org/10.5281/zenodo.19355356.

Acknowledgments. The authors thank Guillaume Couairon, David Landry, and Pierre Garcia for fruitful discussions. The authors used Claude (Anthropic) to assist in drafting portions of this manuscript. All content was reviewed and revised by the authors.

Author contribution. Conceptualization: A.D. and A.C. Data curation: A.D. Formal analysis: A.D. Methodology: A.D. Software: A.D. Supervision: A.C. and D.B. Validation: A.D., A.C., and D.B. Visualization: A.D. Writing—original draft: A.D. Writing—review and editing: A.C. and D.B. All authors approved the final submitted draft.

Provenance statement. 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.

Competing interests. The authors declare none.

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

Funding statement. This work was supported by the French National Research Agency (ANR) and the Ministry of Higher Education and Research. This project was granted HPC resources by GENCI at IDRIS under the allocation AD011016443 on the Jean-Zay supercomputer.

References

Anthes RA, **Kuo YH**, **Baumhefner DP**, **Errico RM and Bettge TW** (1985) Predictability of mesoscale atmospheric motions. *Advances in Geophysics 28*(PB), 159–202. https://doi.org/10.1016/S0065-2687(08)60188-0.

https://doi.org/10.1017/eds.2026.10046 Published online by Cambridge University Press

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.

Bonev B, **Kurth T**, **Mahesh A**, **Bisson M**, **Kossaifi J**, **Kashinath K**, **Anandkumar A**, **Collins WD**, **Pritchard MS and Keller A** (2025) FourCastNet 3: A Geometric Approach to ProAbabilistic Machine-Learning Weather Forecasting at Scale. arXiv:2507.12144 [cs.LG].

Brenowitz ND, **Ge T**, **Subramaniam A**, **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:2505.06474 [physics.ao-ph].

Chen L, **Du F**, **Hu Y**, **Wang Z and Wang F** (2023) SwinRDM: Integrate SwinRNN with diffusion model towards high-resolution and high-quality weather forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence 37*(1), 322–330. https://doi.org/10.1609/aaai.v37i1.25105.

Couairon G, **Singh R**, **Charantonis A**, **Lessig C and Monteleoni C** (2026) ArchesWeatherGen: Skillful and compute-efficient probabilistic weather forecasting with machine learning. *Science Advances 12*(17), eadx2372. https://doi.org/10.1126/sciadv.adx2372.

Crowson K, **Baumann SA**, **Birch A**, **Abraham TM**, **Kaplan DZ and Shippole E** (2024) Scalable High-Resolution Pixel-Space Image Synthesis with Hourglass Diffusion Transformers. In Proceedings of the 41st International Conference on Machine Learning (ICML 2024). Vienna, Austria: Proceedings of Machine Learning Research, Vol. *235*, 9550–9575. July 21–27, 2024.

Dong C, **Loy CC**, **He K and Tang X** (2016) Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence 38*(2), 295–307. https://doi.org/10.1109/TPAMI.2015.2439281.

ECMWF (2019) IFS documentation CY46R1 – part V: Ensemble prediction system. In *IFS Documentation* CY46R1 – Part V: Ensemble Prediction System. In IFS Documentation CY46R1. Shinfield Park: European Centre for Medium-Range Weather Forecasts. https://doi.org/10.21957/38yug0cev.

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 J-N** (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society 146*(730), 1999–2049. https://doi.org/10.1002/qj.3803.

Hess P, **Aich M**, **Pan B and Boers N** (2025) Fast, scale-adaptive and uncertainty-aware downscaling of earth system model fields with generative machine learning. *Nature Machine Intelligence 7*(3), 363–373. https://doi.org/10.1038/s42256-025-00980-5.

Hohenegger C, **Korn P**, **Linardakis L**, **Redler R**, **Schnur R**, **Adamidis P**, **Bao J**, **Bastin S**, **Behravesh M**, **Bergemann M**, **Biercamp J**, **Bockelmann H**, **Brokopf R**, **Brüggemann N**, **Casaroli L**, **Chegini F**, **Datseris G**, **Esch M**, **George G**, **Giorgetta M**, **Gutjahr O**, **Haak H**, **Hanke M**, **Ilyina T**, **Jahns T**, **Jungclaus J**, **Kern M**, **Klocke D**, **Kluft L**, **Kölling T**, **Kornblueh L**, **Kosukhin S**, **Kroll C**, **Lee J**, **Mauritsen T**, **Mehlmann C**, **Mieslinger T**, **Naumann AK**, **Paccini L**, **Peinado A**, **Praturi DS**, **Putrasahan D**, **Rast S**, **Riddick T**, **Roeber N**, **Schmidt H**, **Schulzweida U**, **Schütte F**, **Segura H**, **Shevchenko R**, **Singh V**, **Specht M**, **Stephan CC**, **von Storch JS**, **Vogel R**, **Wengel C**, **Winkler M**, **Ziemen F**, **Marotzke J and Stevens B** (2023) ICON-sapphire: Simulating the components of the earth system and their interactions at kilometer and subkilometer scales. *Geoscientific Model Development 16*(2), 779–811. https://doi.org/10.5194/gmd-16-779-2023.

Jones PW (1999) First- and second-order conservative remapping schemes for grids in spherical coordinates. *Monthly Weather Review 127*(9), 2204–2210. https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2.

Keisler J, **Charantonis AA**, **Goude Y**, **Oueslati B and Monteleoni C** (2026) SerpentFlow: Generative Unpaired Domain Alignment via Shared-Structure Decomposition. arXiv:2601.01979 [cs.LG].

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. https://doi.org/10.1126/science.adi2336.

Lang S, **Alexe M**, **Clare MCA**, **Roberts C**, **Adewoyin R**, **Ben Bouallègue Z**, **Chantry M**, **Dramsch J**, **Dueben PD**, **Hahner S**, **Maciel P**, **Prieto-Nemesio A**, **O'Brien C**, **Pinault F**, **Polster J**, **Raoult B**, **Tietsche S and Leutbecher M** (2026) AIFS-CRPS: Ensemble forecasting using a model trained with a loss function based on the continuous ranked probability score. *npj Artificial Intelligence 2*(1), 18. https://doi.org/10.1038/s44387-026-00073-7.

Ledig C, **Theis L**, **Huszar F**, **Caballero J**, **Cunningham A**, **Acosta A**, **Aitken A**, **Tejani A**, **Totz J**, **Wang Z and Shi W** (2017) Photo-realistic single image super-resolution using a generative adversarial network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4681–4690.

Li L, **Carver R**, **Lopez-Gomez I**, **Sha F and Anderson J** (2024) Generative emulation of weather forecast ensembles with diffusion models. *Science Advances 10*(13), eadk4489. https://doi.org/10.1126/sciadv.adk4489.

Lipman Y, **Chen RTQ**, **Ben-Hamu H**, **Nickel M and Le M** (2023) Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*. Kigali, Rawanda: ICLR 2023. May 1–5, 2023. https://openreview.net/forum?id=PqvMRDCJT9t.

Lopez-Gomez I, **Wan ZY**, **Zepeda-Núñez L**, **Schneider T**, **Anderson J and Sha F** (2025) Dynamical-generative downscaling of climate model ensembles. *Proceedings of the National Academy of Sciences 122*(17), e2420288122. https://doi.org/10.1073/pnas.2420288122.

https://doi.org/10.1017/eds.2026.10046 Published online by 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. https://doi.org/10.1038/s43247-025-02042-5.

Moore BJ (2023) Flow dependence of medium-range precipitation forecast skill over California. *Weather and Forecasting 38*(5), 699–720. https://doi.org/10.1175/WAF-D-22-0081.1.

Nguyen T, **Shah R**, **Bansal H**, **Arcomano T**, **Maulik R**, **Kotamarthi V**, **Foster I**, **Madireddy S and Grover A** (2024) Scaling transformer neural networks for skillful and reliable medium-range weather forecasting. *Advances in Neural Information Processing Systems 37*, 68740–68771.

Price I, **Sanchez-Gonzalez A**, **Alet F**, **Andersson TR**, **El-Kadi A**, **Masters D**, **Ewalds T**, **Stott J**, **Mohamed S**, **Battaglia P**, **Lam R and Willson M** (2025) Probabilistic weather forecasting with machine learning. *Nature 637*(8044), 84–90. https://doi.org/10.1038/s41586-024-08252-9.

Rasp S, **Hoyer S**, **Merose A**, **Langmore I**, **Battaglia P**, **Russell T**, **Sanchez-Gonzalez A**, **Yang V**, **Carver R**, **Agrawal S**, **Chantry M**, **Ben Bouallegue Z**, **Dueben P**, **Bromberg C**, **Sisk J**, **Barrington L**, **Bell A and Sha F** (2024) WeatherBench 2: A benchmark for the next generation of data-driven global weather models. *Journal of Advances in Modeling Earth Systems 16*(6), e2023MS004019. https://doi.org/10.1029/2023MS004019.

Reynolds O (1895) IV. On the dynamical theory of incompressible viscous fluids and the determination of the criterion. *Philosophical Transactions of the Royal Society A* (186), 123–164. https://doi.org/10.1098/rsta.1895.0004.

Saharia C, **Ho J**, **Chan W**, **Salimans T**, **Fleet DJ and Norouzi M** (2023) Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence 45*(4), 4713–4726. https://doi.org/10.1109/TPAMI.2022.3204461.

Stock J, **Arcomano T and Kotamarthi R** (2025) Swift: An Autoregressive Consistency Model for Efficient Weather Forecasting. arXiv:2509.25631 [cs.LG].

Verma Y, **Heinonen M and Garg V** (2024) ClimODE: Climate and weather forecasting with physics-informed neural ODEs. In *The Twelfth International Conference on Learning Representations*. Vienna, Austria: ICLR 2024. May 7–11, 2024. https://openreview.net/forum?id=xuY33XhEGR.

Wan ZY, **Baptista R**, **Boral A**, **Chen Y-F**, **Anderson J**, **Sha F and Zepeda-Núñez L** (2023) Debias coarsely, sample conditionally: Statistical downscaling through optimal transport and probabilistic diffusion models. *Advances in Neural Information Processing Systems 36*, 47749–47763.

Watt RA and Mansfield LA (2024) Generative Diffusion-based Downscaling for Climate. arXiv:2404.17752 [physics.ao-ph].

Cite this article: Delefosse A, Charantonis A and Béréziat D (2026). Super-resolving coarse-resolution weather forecasts with flow matching. *Environmental Data Science*, 5: e25. doi:10.1017/eds.2026.10046

https://doi.org/10.1017/eds.2026.10046 Published online by Cambridge University Press