https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

Generative reconstruction of high-resolution historical climate fields from station observations

Jie Chao[1,2], Qi Liu[1], Hongming Yan[3], Zhenlu Liu[2], Jin Xu[2], Shuojie Gao[2], Zihang Han[4], Hang Su[5], Ziniu Xiao[1,2], Anlai Sun[6], and Baoxiang Pan[2]

[1]Joint Laboratory of Fengyun Remote Sensing, School of Earth and Space Sciences, University of Science and Technology of China, Hefei, China
[2]State Key Laboratory of Earth System Numerical Modeling and Application, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
[3]Yunnan Provincial Climate Center, Kunming, China
[4]Hubei Key Laboratory of Intelligent Yangtze and Hydroelectric Science, China Yangtze Power Co., Ltd. Hubei, China
[5]State Key Laboratory of Atmospheric Environment and Extreme Meteorology, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
[6]Beijing Huayun Shinetek Science and Technology Co., Ltd., Beijing, China

Correspondence: Qi Liu (qliu7@ustc.edu.cn), Anlai Sun (rissal616@sohu.com), Baoxiang Pan (panbaoxiang@lasg.iap.ac.cn)

Abstract. Reconstructing complete, high-resolution daily meteorological fields over multidecadal periods from station observations is severely underdetermined: stations are sparse, unevenly distributed, and temporally inconsistent, especially over complex terrain. Because this ill-posedness requires prior assumptions, the choice of prior determines how well extremes are preserved. Conventional products, whether interpolation, multi-source fusion, or reanalysis, rely on prescribed background-error covariances that attenuate localized extremes as coverage declines. We introduce Generative REconstruction (GRE), a sequential assimilation framework that replaces prescribed covariances with a generative prior learned from satellite-era gridded fields. The prior, a diffusion model trained jointly on precipitation, pressure, wind, and temperature, encodes multivariate spatial structure including heavy-tailed extremes. At each step, posterior sampling conditions the prior on available stations and a forecast background, producing an ensemble whose spread reflects observational coverage. Applied over Southwest China (1951–2024), GRE yields daily 0.1° ensemble reconstructions of all four variables. Conditioned on 80% of stations, GRE achieves $R = 0.93$ and RMSE $= 5.8$ mm/day for daily precipitation versus 0.69 and 7.1 for the baseline product, recovering rather than smoothing localized peaks. The joint prior propagates observational information across variables, constraining fields not directly observed. Over three decades, reconstructed precipitation extremes remain consistent with the reference climatology while ensemble spread contracts as gauges densify, evidencing calibrated long-term uncertainty. The forecast background adds most value under sparse coverage and persistent synoptic conditions, maintaining bounded physically consistent uncertainty throughout the early record.

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

1 Introduction

Understanding long-term regional climate variability requires high-resolution, gridded meteorological fields extending back decades before modern observational networks matured. Over Southwest China, where monsoon dynamics interact with the complex topography of the Tibetan Plateau margin, the Yunnan Plateau, and the Sichuan Basin (Xu et al., 2008), such records are essential for diagnosing climate extremes (Zhang et al., 2008; Zhai et al., 2005) and evaluating climate model fidelity (Flato et al., 2014); they also underpin the design criteria of long-lived infrastructure such as hydropower systems, flood defenses, and water-resource management schemes (Liu et al., 2022). Yet for much of the mid-20th century, in situ observations are sparse, unevenly distributed, and limited to few locations, making spatially complete, temporally consistent, high-resolution reconstruction a severely under-determined challenge (Schneider, 2001; Chao et al., 2025).

Conventional gridded products address this gap through three methodologically distinct strategies, interpolation, multi-source fusion, and reanalysis, each limited by a different mechanism, yet all converging on a common emergent behavior: attenuation of localized extremes as observational coverage decreases. Interpolation-based products (Wu and Gao, 2013; Yatagai et al., 2012; Hu et al., 2025) honor station records but impose prescribed covariance structures that, by construction, redistribute intensity across smoothing neighborhoods, damping terrain-modulated gradients and spatially concentrated extremes. Multi-source fusion products such as the China Meteorological Forcing Dataset (CMFD, He et al. 2020) achieve $0.1°$ resolution by merging satellite retrievals, reanalysis fields, and station data, but their reliance on satellite-era observations means that estimates before 1979, if available, may be less directly constrained and therefore less reliable, and the statistical relationships calibrated during the satellite period may not transfer faithfully to earlier decades when station density alone must constrain the solution. Dynamical reanalyses, such as ERA5 (Hersbach et al., 2020), ensure physical consistency across variables but operate at $\sim 0.25°$ resolution, too coarse to resolve orographic precipitation gradients, and their skill degrades as assimilated observations thin before the satellite era. In each case, the loss of amplitude fidelity worsens precisely when it matters most: during extreme events in the early instrumental record, where stations are fewest and the scientific need for accurate intensity estimates is greatest.

Machine learning offers an alternative strategy: rather than prescribing spatial covariance, learn it from trustworthy data (Pan et al., 2019; Luo et al., 2026). Meteorological fields exhibit spatiotemporal covariability rooted in governing dynamics and quasi-stationary boundary conditions (Chao et al., 2025; Yang et al., 2025); a model that captures this structure encodes an implicit physical prior, so that reconstructing a field from sparse observations becomes posterior inference under a learned distribution rather than deterministic interpolation with prescribed smoothing (Reichstein et al., 2019). Deterministic inpainting has demonstrated skill surpassing conventional statistical methods for global ocean temperature reconstruction (Kadow et al., 2020, 2024), but produces single best-estimate fields that cannot represent the ambiguity inherent in severely underdetermined problems. Probabilistic diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020; Kingma et al., 2021; Song et al., 2021) address this gap on two levels: as generative models, they learn the full joint distribution over spatial fields, enabling principled posterior sampling conditioned on arbitrary observation configurations (Song et al., 2022; Chung et al., 2023); as score-based denoisers, their progressive noise schedule captures spatial structure from large-scale gradients to fine-scale extremes, a multi-

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

scale property well matched to meteorological fields shaped by both synoptic circulation and local orography. These strengths have been demonstrated in nowcasting (Nai et al., 2024; Wan et al., 2026), ensemble forecasting (Price et al., 2025; Nai et al., 2025; Ou et al., 2025), stochastic downscaling (Pan et al., 2023; Mardani et al., 2025; Wang et al., 2026), climate reconstruction with ensemble uncertainty (Ling et al., 2024; Chao et al., 2025), temporal assimilation through coupled forecasting (Yang et al., 2025), and station-scale data assimilation (Huang et al., 2024; Manshausen et al., 2025).

Despite these progresses, existing spatial reconstructions mostly operate at monthly resolution (Ling et al., 2024; Kadow et al., 2024), and temporal assimilation frameworks target weather-to-seasonal timescales (Yang et al., 2025). Few studies have produced daily, multivariate, high-resolution climate reconstructions from sparse stations over multi-decadal periods. Such reconstructions must simultaneously capture the coupled spatiotemporal variability across meteorological variables and remain robust when observations are sparse, unevenly distributed, or temporally inconsistent, conditions that characterize the early instrumental record. Finally, they must provide calibrated uncertainty estimates that reflect the varying strength of observational constraints.

Motivated by these challenges, we propose Generative REconstruction (GRE). GRE combines a probabilistic diffusion model, a deterministic forecast model, and diffusion posterior sampling (DPS; Chung et al. 2023) into a data-driven analogue of sequential data assimilation (Kalnay, 2003; Evensen, 2003), replacing prescribed background-error statistics with a generative prior learned from physically consistent gridded data. We apply GRE to produce daily $0.1°$ fields of precipitation, surface pressure, wind speed, and temperature over Southwest China (15°–35°N, 90°–110°E) since 1951, yielding multivariate-coherent, spatiotemporally consistent, and uncertainty-aware reconstructions that remain robust under the sparse and temporally varying observational constraints of the early instrumental record.

2 Data and Methods

GRE is a data-driven analogue of sequential data assimilation, coupling two independently trained components at inference (Figure 1). A diffusion model (Section 2.2, Figure 1a) learns the joint spatial distribution of meteorological fields from a satellite-era derived gridded product (CMFD v2.0; see Section 2.1), encoding cross-variable spatial structure as a generative prior. A forecast model (Section 2.3, Figure 1b) propagates the previous day's reconstructed state forward, providing temporal context that constrains the current step, particularly where and when stations are absent. The posterior sampling procedure (Section 2.4, Figure 1c) fuses the prior with two sources of constraint—direct station observations and the forecast background—through reverse-diffusion steps, producing ensemble reconstructions whose spread reflects remaining uncertainty. This cycle repeats daily: each day's reconstruction conditions on its own observations and the preceding forecast, then advances through the forecast model to initialize guidance for the next day.

2.1 Study Domain and Data

GRE operates on a $0.1° \times 0.1°$ grid (200×200 points) over $15°$–$35°$N, $90°$–$110°$E, reconstructing daily fields of precipitation (P), surface pressure (sp), 10-m wind speed (U_{10m}), and 2-m temperature (T_{2m}) for 1951–2024. Both the diffusion model

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

and forecast model are trained on CMFD v2.0 (He et al., 2020, 2024), a derived gridded product that is itself constructed by fusing satellite retrievals, reanalysis fields, and station observations through terrain-aware interpolation and elevation-based corrections. Training is restricted to the satellite era (1981–2020), when CMFD is most directly constrained by observations. The diffusion prior is trained on daily fields over the full 1981–2020 window; the forecast model is trained on consecutive daily pairs from 1981–2015 and validated on 2016–2020. Station observations from the China Meteorological Administration (CMA), obtained through the China Meteorological Data Service Centre (China Meteorological Data Service Centre, 2026), provide daily P, U_{10m}, and T_{2m} for conditioning during reconstruction. The station network is unevenly distributed, concentrated in lower-elevation basins and river valleys with substantially sparser coverage over the western highlands and the Tibetan Plateau margin; the number of reporting stations also grows markedly over time, from a sparse network in the early 1950s to a considerably denser one by the 1980s (Fig. S1 in the Supplement). Surface pressure (sp) is unavailable from CMA stations and is therefore constrained solely by the learned multivariate prior and forecast guidance. The diffusion model is additionally conditioned on the digital elevation model (DEM) from the GEBCO_2026 Grid (GEBCO Bathymetric Compilation Group 2026, 2026) and solar angle (SA) as static channels encoding topographic and seasonal radiative forcing.

2.2 Diffusion-Based Spatial Prior

The spatial prior is a score-based diffusion model (Sohl-Dickstein et al., 2015; Ho et al., 2020; Kingma et al., 2021; Song et al., 2021) that learns the joint distribution of meteorological fields by constructing a continuous bijective mapping between the data distribution and a standard Gaussian. Defining $\mathbf{x}_0 \in \mathbb{R}^{C\times H\times W}$ ($C = 4$ variables, $H = W = 200$ grid points) as a multi-variable daily meteorological field over the study domain, the forward stochastic differential equation (SDE):

$$d\mathbf{x}_\tau = f_\tau(\mathbf{x}_\tau)\,d\tau + g_\tau\,d\mathbf{w}_\tau, \quad \tau \in [0,1], \tag{1}$$

progressively transforms $\mathbf{x}_0$ into Gaussian noise $\mathbf{x}_1 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$, where f_τ and g_τ are the pre-defined drift and diffusion coefficients and $\mathbf{w}_\tau$ is a standard Wiener process. We adopt a variance-preserving formulation in which f_τ and g_τ are set by a log signal-to-noise-ratio (log-SNR) schedule that yields closed-form noising coefficients $\alpha(\tau)$ and $\sigma(\tau)$ (Sect. S1 in the Supplement). This process admits a time-reversed SDE (Anderson, 1982):

$$d\mathbf{x}_\tau = \left[f_\tau(\mathbf{x}_\tau) - g_\tau^2\,\nabla_{\mathbf{x}_\tau}\log p_\tau(\mathbf{x}_\tau \mid \mathbf{c})\right]d\bar{\tau} + g_\tau\,d\bar{\mathbf{w}}_\tau, \tag{2}$$

where $d\bar{\tau}$ denotes reverse time, $\mathbf{c} = [\mathrm{DEM};\mathrm{SA}]$ encodes static topographic and solar forcing attributes. Solving this reverse SDE from $\mathbf{x}_1 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ generates samples from the learned distribution. In practice, numerical SDE solvers with discretized time steps are used (Song et al., 2021). The only unknown, the score function $\nabla_{\mathbf{x}_\tau}\log p_\tau(\mathbf{x}_\tau \mid \mathbf{c})$, is approximated by a neural network s_θ trained via denoising score matching (Vincent, 2011):

$$s_\theta^* = \arg\min_{s_\theta}\,\mathbb{E}_{\tau,\mathbf{x}_0,\mathbf{x}_\tau\sim p_{0\tau}(\mathbf{x}_\tau\mid\mathbf{x}_0)}\left\|s_\theta(\mathbf{x}_\tau,\tau \mid \mathbf{c}) - \nabla_{\mathbf{x}_\tau}\log p_{0\tau}(\mathbf{x}_\tau \mid \mathbf{x}_0)\right\|^2. \tag{3}$$

As the noise schedule progresses from fine to coarse perturbations, the learned score captures spatial structure across scales, from synoptic-scale gradients at high noise levels to mesoscale features at low noise levels, a multi-scale property well suited

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

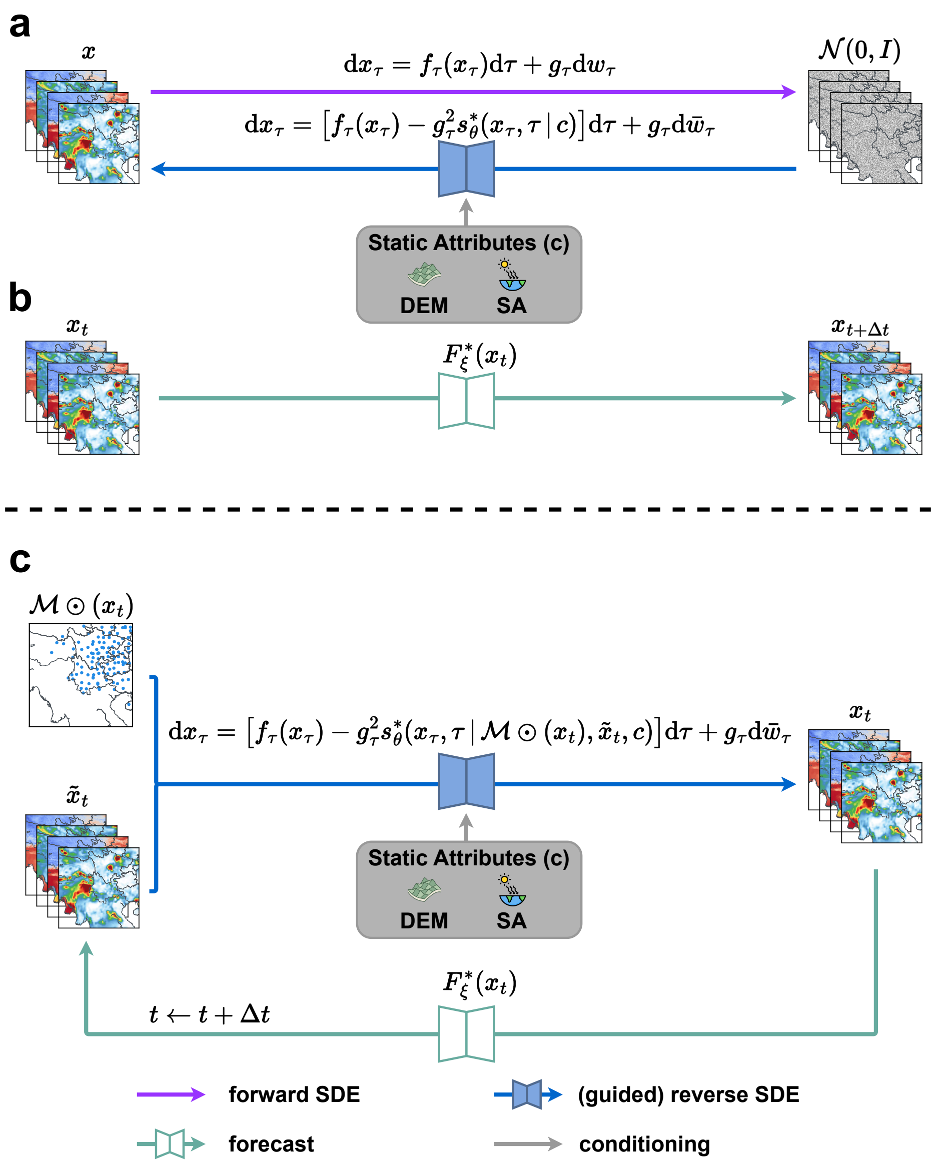


Figure 1. Overview of GRE, structured as a data-driven sequential assimilation cycle with two phases. **Training (a, b):** two models are independently trained on CMFD v2.0 (1981–2020). The diffusion model **(a)** learns the probability distribution of multivariate weather fields, conditioned on topography (DEM) and seasonal solar forcing (SA), by training a neural network to map continuously between a standard Gaussian and the target distribution, a bijective transformation whose reversal requires only the learned score function s_θ^* (Sohl-Dickstein et al., 2015; Ho et al., 2020; Kingma et al., 2021; Song et al., 2021). The forecast model **(b)** learns how these fields evolve from one day to the next. **Reconstruction (c):** for each day t, the diffusion model generates candidate weather fields consistent with the learned prior; these candidates are then steered toward agreement with the day's station observations $\mathcal{M} \odot (\mathbf{x}_t)$ and with the forecast $\tilde{\mathbf{x}}_t$ carried forward from the previous day's analysis, producing the reconstructed field $\mathbf{x}_t$. The forecast model then advances this analysis to provide the background for day $t + \Delta t$, and the cycle repeats, mirroring the forecast–analysis loop of classical sequential assimilation, but with the learned prior replacing prescribed background-error statistics.

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

to meteorological fields shaped by both large-scale circulation and local orography. The model is trained on daily four-variable fields from CMFD v2.0 (1981–2020), with DEM and SA concatenated as additional input channels (see Sect. S1 in the Supplement for network architecture, training strategy, and implementation details).

2.3 Forecast Model for Temporal Propagation

Station observations on any single day provide only sparse spatial coverage, but observations from preceding days carry additional information about the current meteorological state. GRE exploits this temporal dependence by propagating the previous day's reconstructed field forward through a deterministic forecast model trained on consecutive daily field pairs $(\mathbf{x}_t, \mathbf{x}_{t+\Delta t})$ from CMFD v2.0 by minimizing the mean squared prediction error:

$$F_\xi^* = \arg\min_{F_\xi} \mathbb{E}_t \left\| F_\xi(\mathbf{x}_t) - \mathbf{x}_{t+\Delta t} \right\|^2, \tag{4}$$

where t is calendar time, $\Delta t = 1$ day, ξ^* is optimal parameter of the forecasting model F. At inference, the forecast $\tilde{\mathbf{x}}_{t+\Delta t} = F_\xi^*(\mathbf{x}_t)$ provides a useful but limited role: MSE optimization yields the conditional mean; the forecast tends to smooth fine-scale variability and under-represent extremes, deficiencies that would be unacceptable in a standalone reconstruction system. However, within GRE's framework, these limitations are absorbed by the diffusion prior, which supplies the realistic spatial structure and non-Gaussian variability that the forecast lacks. The forecast instead serves as a soft temporal anchor: it biases the reconstruction toward dynamically plausible states without rigidly imposing its own smoothed estimate. This division of labor (temporal context from the forecast, spatial realism from the prior, observational fidelity from station conditioning) is formalized in the posterior sampling procedure described next (see Sect. S2 in the Supplement for model configuration, training strategy, and implementation details).

2.4 Sequential Reconstruction via Posterior Sampling

During reconstruction, GRE processes one day at a time following the forecast–analysis cycle of sequential data assimilation: the forecast model propagates yesterday's reconstruction forward to form a background estimate $\tilde{\mathbf{x}}_t = F_\xi^*(\mathbf{x}_{t-\Delta t})$, and diffusion posterior sampling (DPS; Chung et al. 2023) then corrects this background using station observations, producing the current day's analysis. This cycle repeats daily, with each analysis feeding the next day's forecast. Importantly, GRE can cold-start at any point in the record: when no forecast is available, reconstruction relies solely on the diffusion prior and station observations, requiring no temporal context or spin-up. When a forecast is available, it enters as an additional soft constraint that improves the reconstruction but is not essential. This flexibility is valuable for long-term climate reconstruction, where temporal gaps and record boundaries are common.

Specifically, starting from Gaussian noise $\mathbf{x}_1 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$, the reverse SDE is solved with a guided score:

$$d\mathbf{x}_\tau = \Big[f_\tau(\mathbf{x}_\tau) - g_\tau^2\, s_\theta^*\big(\mathbf{x}_\tau, \tau \mid \mathcal{M}_t \odot (\mathbf{x}_t), \tilde{\mathbf{x}}_t, \mathbf{c}\big) \Big] d\bar{\tau} + g_\tau\, d\bar{\mathbf{w}}_\tau, \tag{5}$$

Here s_θ^* denotes the *effective* guided score. The network itself evaluates only the unconditional score $s_\theta^*(\mathbf{x}_\tau, \tau \mid \mathbf{c})$, while the observation and forecast constraints enter through external DPS gradient corrections rather than through the network's

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

inputs. In this expression, $\mathcal{M}_t \odot (\mathbf{x}_t)$ denotes the station observations on day t, $\tilde{\mathbf{x}}_t$ denotes the forecast when available, and $\mathbf{c} = [\mathrm{DEM}; \mathrm{SA}]$ denotes the static attributes. At each reverse-diffusion step, the denoised field estimate $\tilde{\mathbf{x}}_0(\mathbf{x}_\tau)$ is computed from the unconditional score via Tweedie's formula (Efron, 2011):

$$\tilde{\mathbf{x}}_0(\mathbf{x}_\tau) = \frac{1}{\alpha(\tau)}\left[\mathbf{x}_\tau + \sigma^2(\tau)\, s_\theta(\mathbf{x}_\tau, \tau \mid \mathbf{c})\right], \tag{6}$$

and two DPS corrections are then applied sequentially: first, a forecast-guidance gradient steers the sample toward the background $\tilde{\mathbf{x}}_t$; second, an observation-guidance gradient enforces consistency with station measurements:

$$\mathbf{x}_\tau \leftarrow \mathbf{x}_\tau - \lambda_f \nabla_{\mathbf{x}_\tau} \left\|\tilde{\mathbf{x}}_0(\mathbf{x}_\tau) - \tilde{\mathbf{x}}_t\right\|^2, \tag{7}$$

$$\mathbf{x}_\tau \leftarrow \mathbf{x}_\tau - \lambda_o \nabla_{\mathbf{x}_\tau} \left\|\mathcal{M}_t \odot \tilde{\mathbf{x}}_0(\mathbf{x}_\tau) - \mathcal{M}_t \odot (\mathbf{x}_t)\right\|^2, \tag{8}$$

where λ_f and λ_o are forecast- and observation-guidance weights (see Sect. S1 in the Supplement for the hyperparameters). Applying observation guidance last ensures that where station data exist, the reconstruction is anchored to measurements rather than to the forecast. The corrections are imposed at fixed intervals along the reverse trajectory, while the remaining steps follow the unconditional reverse dynamics. The full procedure is summarized in Algorithm 1.

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

Algorithm 1 Generative REconstruction

Input:

- Pre-trained diffusion model with score $s_\theta(\mathbf{x}_\tau, \tau \mid \mathbf{c}) \approx \nabla_{\mathbf{x}_\tau} \log p_\tau(\mathbf{x}_\tau \mid \mathbf{c})$
- Pre-trained forecast model F_ξ^*
- Station observations $\{\mathcal{M}_t \odot (\mathbf{x}_t)\}_{t=t_1}^{t_T}$ (may be absent on some days)
- Static attributes $\mathbf{c} = [\mathrm{DEM}; \mathrm{SA}]$

Output:

- Ensemble of reconstructed fields $\{\mathbf{x}_t^{(n)}\}_{t=t_1}^{t_T}, n = 1, \ldots, N_{\text{ens}}$

Parameters:

- guidance weights λ_f, λ_o
- warm-start noise level τ^*
- number of reverse steps K

```
for t ← t_1 to t_T do
    for n ← 1 to N_ens do
        // Part A: Initialize latent variable
        if t is first day of month (no forecast available) then
            τ_start ← 1  Sample x_{τ_start} from N(0, I)  // Cold start: uninformative latent
        else
            τ_start ← τ*  Obtain x_{τ_start} by solving forward SDE from τ = 0 to τ*, starting from x_{τ=0} = x̃_t^(n)  // Warm start:
                latent encodes forecast information
        // Part B: Guided reverse SDE
        for k ← K to 1 do
            Compute denoised estimate x̃_0(x_{τ_k}) via Tweedie's formula  Take unconditional reverse SDE step using s_θ(x_{τ_k}, τ_k | c)
            // Forecast guidance (background correction)
            if forecast x̃_t^(n) is available then
                x_{τ_{k-1}} ← x_{τ_{k-1}} − λ_f ∇_{x_{τ_k}} ‖x̃_0(x_{τ_k}) − x̃_t^(n)‖²
            // Observation guidance (analysis correction)
            if station observations M_t are available on day t then
                x_{τ_{k-1}} ← x_{τ_{k-1}} − λ_o ∇_{x_{τ_k}} ‖M_t ⊙ (x̃_0(x_{τ_k})) − M_t ⊙ (x_t)‖²
        x_t^(n) ← x_{τ_0}
        // Part C: Propagate forward for next day
        x̃_{t+Δt}^(n) ← F_ξ^*(x_t^(n))
```

---

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

3 Results

We evaluate GRE in five steps that move from its underlying spatial prior to the full reconstruction system under realistic observational constraints. Section 3.1 examines whether the diffusion prior captures the multivariate spatial climatology of the study domain, a prerequisite for every later result, by comparing the statistics of unconditionally generated fields with the training climatology. Section 3.2 then asks whether the conditioned reconstruction honors station observations and reproduces localized extremes, using an observation ablation in which available stations are progressively withheld and skill is measured at the held-out sites. Section 3.3 investigates the added value of multivariate coupling and temporal continuity, isolating how cross-variable physical relationships encoded in the diffusion prior and temporal information propagated by the forecast jointly constrain the reconstruction when observations are removed. Section 3.4 then evaluates the network-wide impact of the forecast background, quantifying how the propagated temporal information contributes across the full station network and how its effect varies with station density and seasonal conditions. Finally, Section 3.5 extends the evaluation to the multi-decadal scale, comparing reconstructed daily precipitation extremes and means against the reference product over 1951–1981 and examining how ensemble uncertainty co-evolves with the historical growth of the gauge network.

3.1 Multivariate Spatial Structure Learned by the Diffusion Prior

A reconstruction prior is physically useful only if the fields it generates share the climatological and multivariate statistical structure of the region, rather than merely resembling individual days. We assess this directly by drawing $10{,}000$ fields from the prior through sampling the reverse SDE, conditioning only on the static attributes $\mathbf{c} = [\mathrm{DEM}; \mathrm{SA}]$, with no station observations and no forecast guidance. We thereafter compare the statistics of this generated ensemble with the CMFD training climatology. Because the generated fields are never constrained by observations, any agreement in their aggregate statistics must originate from spatial and cross-variable structure encoded in the prior.

Figure 2 summarizes the comparison for multivariate statistics. For single-grid-point statistics (Figure 2a–h), the generated climatological mean and standard deviation of precipitation match CMFD with pattern correlation coefficients (PCCs) of 0.92 and 0.88, while the maximum precipitation and maximum 10-m wind speed show PCCs of 0.86 and 0.94, respectively. The first and second Empirical Orthogonal Functions (EOF1, EOF2) of precipitation are also well reproduced with PCCs of 0.89 and 0.95 (Figures 2i–l). The smoother fields are captured even more faithfully: surface pressure, 10-m wind speed, and 2-m temperature all reach PCC ≥ 0.95 for the mean and the leading modes (Figs. S2–S4 in the Supplement). Also, the multivariate dependence structure is well preserved: the inter-variable correlations in GRE closely match those in CMFD across all variable pairs (Fig. S5 in the Supplement). The radially averaged power spectra of the generated fields (Figures 2m–p) overlap those of CMFD across all four variables, from synoptic wavenumbers down to the grid scale, showing that the prior reproduces the distribution of spatial variance across scales rather than oversmoothing fine-scale structure.

Together, these diagnostics establish that the prior encodes the region's multivariate spatial climatology, providing a physically grounded constraint wherever and whenever station observations are absent.

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

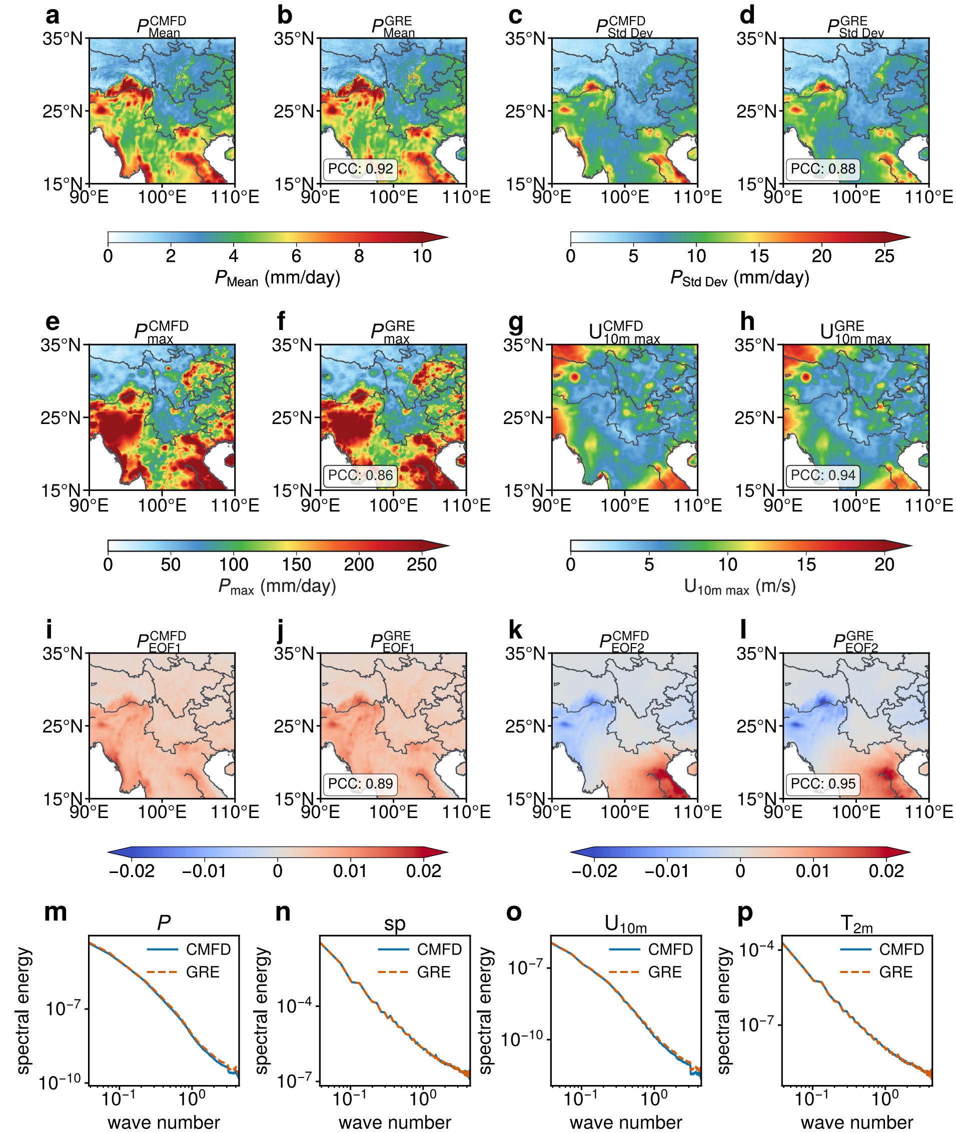


Figure 2. Comparison of multivariate spatial statistics between the CMFD and unconditionally GRE samples. **(a–h)** Grid-point-wise statistics: **(a,b)** climatological mean precipitation (P) for CMFD and GRE; **(c,d)** climatological standard deviation of precipitation P for CMFD and GRE; **(e,f)** maximum precipitation ($P_{\max}$) for CMFD and GRE; **(g,h)** maximum 10-m wind speed ($U_{10\mathrm{m},\max}$) for CMFD and GRE. Each GRE panel is annotated with its pattern correlation coefficient (PCC) against the corresponding CMFD field. **(i–l)** Spatial pattern comparison: **(i,j)** first Empirical Orthogonal Function (EOF1) of precipitation (P) for CMFD and GRE; **(k,l)** second Empirical Orthogonal Function (EOF2) of precipitation (P) for CMFD and GRE. **(m–p)** Radially averaged power spectra (spectral energy versus wavenumber) for **(m)** precipitation (P), **(n)** surface pressure (sp), **(o)** 10-m wind speed ($U_{10\mathrm{m}}$), and **(p)** 2-m temperature ($T_{2\mathrm{m}}$): CMFD (solid) and GRE (dashed).

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

3.2 Reconstructing Historical Extremes

Having established that the prior encodes the region's multivariate spatial climatology, we now evaluate the full sequential assimilation system, in which the prior is conditioned on station observations and coupled with the forecast model to produce daily reconstructions. We focus on precipitation (Figure 3), the most intermittent and heavy-tailed of the four variables, and therefore the one that most severely tests whether the framework preserves localized intensity rather than smoothing it away. The practical value of a historical reconstruction depends on whether it captures the intensity and location of extreme events, precisely the features that conventional products smooth as the station network thins toward the early record. To mimic the sparser networks of earlier decades, we condition GRE on a randomly selected 50% ($\mathrm{GRE}_{50\%}$, Figure 3e,h) or 80% ($\mathrm{GRE}_{80\%}$, Figure 3f,i) of each day's reporting stations and score the reconstruction at the remaining withheld sites. Because the historical network changes from day to day, the selection is applied to each day's available set, producing a different set of conditioning and evaluation stations daily. We benchmark against CMFD over the same withheld sites.

Figure 3a–c shows the reconstruction on 25 June 1951, the earliest and most sparsely observed day analyzed. CMFD (Figure 3b) reproduces a broad wet pattern but spreads the rainfall across the region, displacing and damping the observed peak. GRE conditioned on the available stations (Figure 3c) instead produces a spatially concentrated maximum collocated with the observed values: as the station data are assimilated, the diffusion prior fills in terrain-aligned gradients around them, and a localized extreme emerges rather than being smeared out. This behavior, intensity preserved rather than redistributed, is the qualitative signature that distinguishes a generative prior from covariance-based interpolation.

To move beyond a single case, we repeat the experiment over the full year 1951 and compare the reconstructed daily precipitation against the withheld observations. The maps of annual-mean RMSE at each station (Figure 3d–f) show that CMFD's largest errors concentrate near the plateau–basin transition, where orography modulates precipitation most strongly; $\mathrm{GRE}_{80\%}$ (Figure 3f,i) markedly reduces these errors. The scatter comparisons across all withheld station-days (Figure 3g–i) confirm this at the distribution level: $\mathrm{GRE}_{80\%}$ achieves $R = 0.93$ and RMSE $= 5.8$ mm/day, compared with $R = 0.69$ and RMSE $= 7.1$ mm/day for CMFD. Even $\mathrm{GRE}_{50\%}$, conditioned on only half of each day's stations (Figure 3e,h), attains a higher correlation than CMFD ($R = 0.78$ versus 0.69), a notable result given that CMFD was constructed using the full station network together with satellite and reanalysis data. Reconstruction quality thus improves steadily as more stations are available, and GRE retains better agreement with withheld observations than the conventional product even under the sparse conditions typical of the early instrumental record.

3.3 Added Value of Multivariate Coupling and Temporal Continuity

The stations reporting on a given day are not the only information available to the reconstruction. Two additional factors shape the analysis: the physical relationships among meteorological variables, for example, a rainfall extreme is typically accompanied by a pressure low, convergent winds, and a temperature drop, and the memory of previous days' weather carried forward by the forecast model. Since GRE's diffusion prior is trained on all four variables jointly, it encodes these cross-variable relationships. Meanwhile, the forecast propagates yesterday's analysis into today's background, providing temporal

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

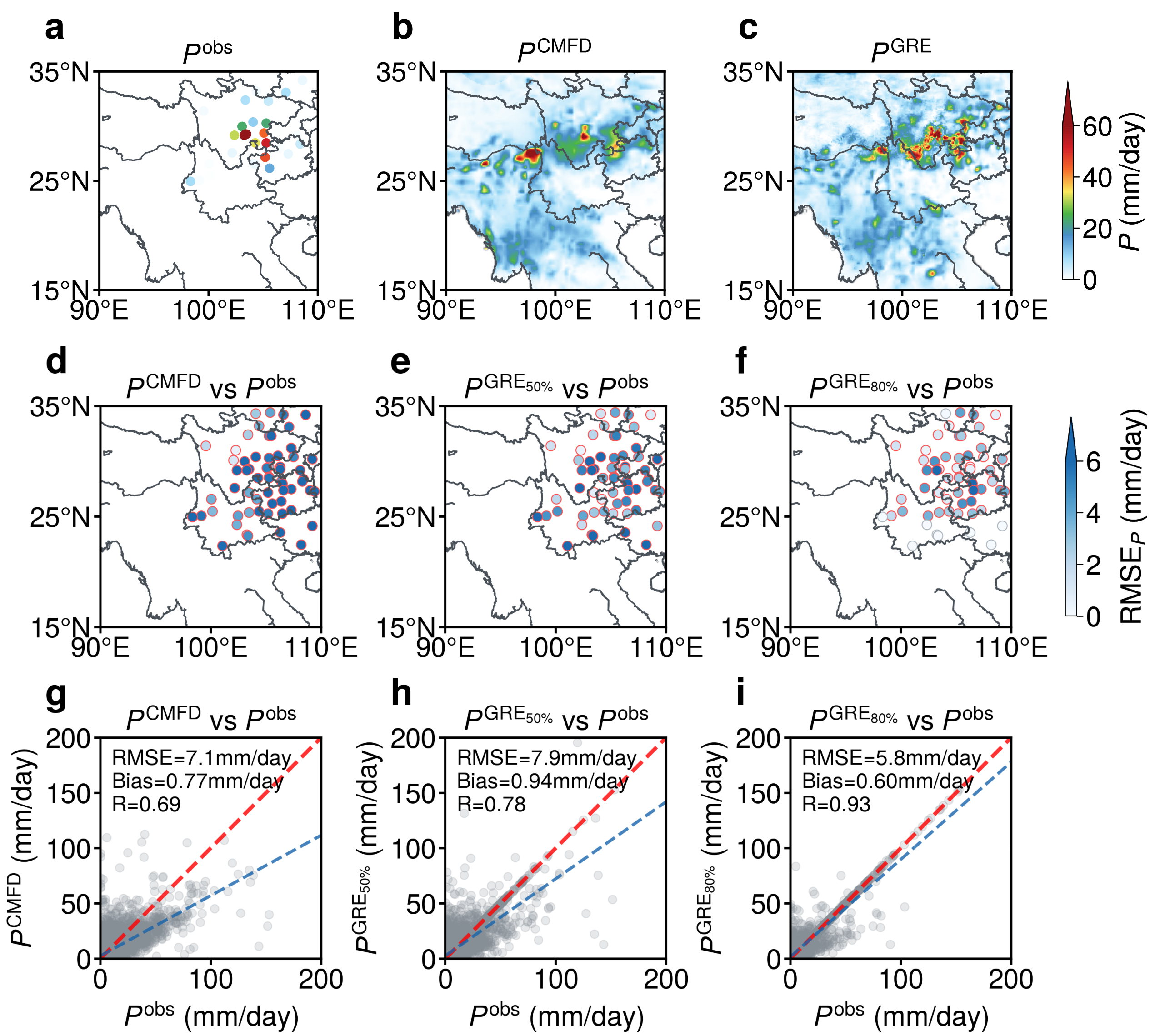


Figure 3. Observation constraint and extreme-event fidelity for 1951 precipitation. **(a–c)**: single case, 25 June 1951, the earliest and most sparsely observed day analyzed: **(a)** station observations (P^{obs}), **(b)** CMFD (P^{CMFD}), and **(c)** GRE conditioned on all stations (P^{GRE}). **(d–f)**: annual-mean (1951) station-level RMSE relative to observations, mapped at station locations within the evaluation region (red circle), for **(d)** CMFD, **(e)** GRE$_{50\%}$, and **(f)** GRE$_{80\%}$. **(g–i)**: scatter plot of observed versus reconstructed daily precipitation, with RMSE, correlation (R), and bias labeled, for **(g)** CMFD, **(h)** GRE$_{50\%}$, and **(i)** GRE$_{80\%}$. GRE$_{50\%}$ and GRE$_{80\%}$ are conditioned on 50% and 80% of each day's available stations and scored at the withheld stations.

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

continuity. Here we isolate how much each factor contributes by progressively removing current-day observations at a single grid point that records two major precipitation events (Figure 4).

We select a grid point that records two major precipitation events within a 45-day window (19 July to 1 September 1951): 74 mm/day on 31 July and 151 mm/day on 25 August. To isolate how different sources of information contribute to the reconstruction, we run three experiments that progressively remove current-day observations on the event days (Figure 4). In the first experiment (Figure 4a–d), all observed variables (P, U_{10m}, T_{2m}) are assimilated on the event days as on every other day. In the second (Figure 4e–h), precipitation is withheld on the event days while wind speed and temperature remain, so that the reconstruction of precipitation draws on cross-variable relationships and the forecast background. In the third (Figure 4i–l), no current-day observations are supplied on the event days, leaving the reconstruction to rely entirely on the forecast propagated from the previous day's analysis and on the prior. Surface pressure (sp) is never observed at this location in any experiment; its behavior therefore reveals how much constraint the other variables and the forecast provide indirectly.

When all station data are supplied (Figure 4a–d), the ensemble collapses onto the measurements for P, U_{10m}, and T_{2m}, with near-zero spread and error. Surface pressure, though unobserved, is kept to a modest spread of 1.8–2.6 hPa, anchored by its learned physical co-variability with the observed fields and by the forecast.

Removing precipitation on the event days widens the precipitation ensemble substantially: spread of 40.9 mm/day on 31 July and 97.8 mm/day on 25 August. Yet the ensemble mean still tracks the withheld peaks rather than collapsing to climatology (Figure 4e–h). The retained wind and temperature observations, together with the forecast, continue to inform precipitation through their shared physical relationships: heavy rainfall is typically accompanied by enhanced wind convergence and local cooling, and these signatures help the analysis recover the precipitation signal even though no rain gauge is assimilated. Wind speed and temperature, still observed, remain pinned to their measurements with zero spread. Surface pressure spread increases modestly (to 3.0–3.1 hPa), reflecting the loss of precipitation as an indirect anchor.

The contrast between the two events is informative. On 31 July (74 mm/day), cross-variable coupling reduces precipitation spread from 58.8 to 40.9 mm/day relative to the fully withheld case, a roughly 30% reduction. On 25 August (151 mm/day), the reduction is negligible (98.1 to 97.8 mm/day): the event is so far into the tail of the precipitation distribution that wind and temperature provide limited additional constraint. Cross-variable information thus helps most for moderate extremes and diminishes as events become exceptionally rare.

When no current-day observations are assimilated (Figure 4i–l), the reconstruction on the event days is driven entirely by the forecast background and the prior. Precipitation spread widens further (58.8 and 98.1 mm/day), and ensemble mean errors grow to 25.3 and 51.5 mm/day, large but still well below the event magnitudes of 74 and 151 mm/day, indicating that the forecast carries meaningful temporal information from the preceding days when observations were available. Wind speed and temperature, now also unobserved, develop nonzero but small spreads. Temperature is notably robust: even without any current-day observations, its absolute error remains ≤ 0.2 K, reflecting its smooth, persistent character that the forecast alone nearly captures. Precipitation, by contrast, is intermittent and heavy-tailed, and therefore much harder to reconstruct from temporal information alone.

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

Surface pressure spread roughly triples relative to the fully observed case (1.8–2.6 to 7.4–7.6 hPa). Because sp is never directly observed, this tripling measures the total indirect constraint lost when the other variables are removed: cross-variable coupling and the forecast together were holding the pressure uncertainty to under 3 hPa, and without those anchors the ensemble widens but remains bounded.

Across all three experiments, ensemble uncertainty grows in an orderly, variable-appropriate manner as observational support is removed. The heavy-tailed precipitation field absorbs most of the added uncertainty; the smoothly varying temperature stays well constrained even when unobserved; and the ensemble never diverges. This behavior, spread scaling with the strength of the available constraint while remaining bounded, is the hallmark of a well-calibrated ensemble and is essential for reconstructing the early instrumental record, where station reporting is both sparse and intermittent.

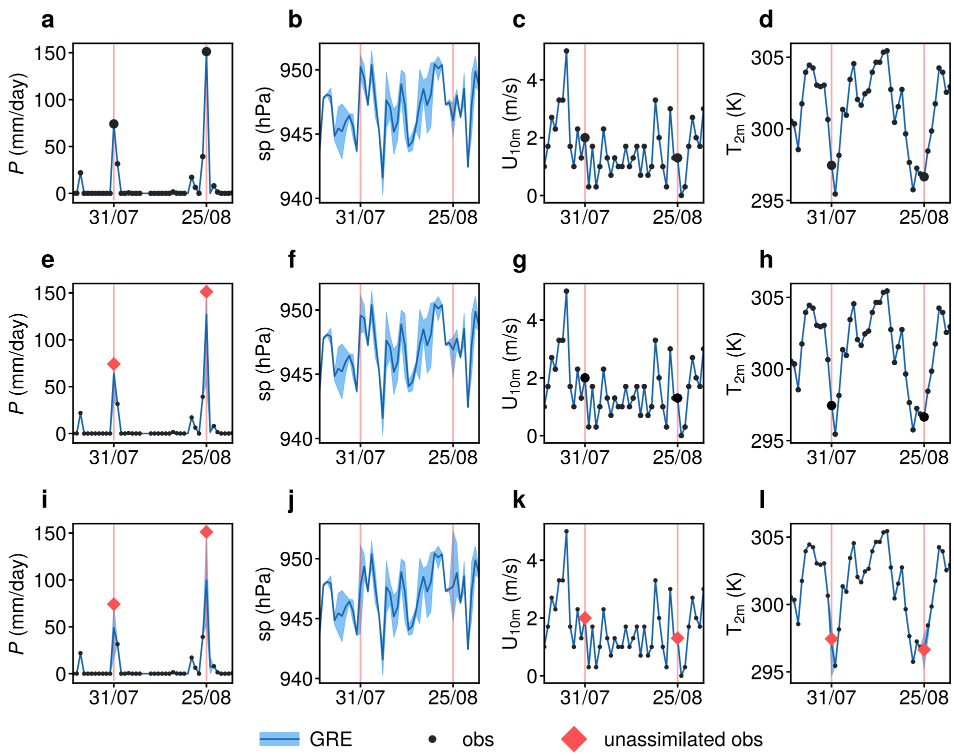
3.4 Network-Wide Impact of the Forecast Background

Most existing generative and deterministic inpainting methods treat each day independently (Ling et al., 2024; Chao et al., 2025): they learn spatial structure from gridded data and condition on the available observations, but they carry no memory from one day to the next. A key distinction of GRE is the forecast model, which propagates yesterday's analysis forward as a temporal background for today's reconstruction. The single-point experiments above showed that this background provides meaningful constraint even when no current-day observations are available. We now ask how much it contributes across the full station network and whether its contribution varies with season and station density.

Figure 5 maps the change in reconstruction skill when forecast guidance is removed, defined as the metric without forecast minus the metric with forecast, so that positive values indicate improvement. The results are calculated based on data covering the full year of 1951. Under the sparser $\mathrm{GRE}_{50\%}$ (Figure 5a), where 50% of station observations are applied to constrain estimates, most withheld stations show reductions in both wind-speed error and ensemble spread when the forecast is included, with the largest gains near the plateau–basin transition where terrain-driven variability is highest and the stations are fewest, exactly where the analysis has the least same-day observational support and therefore benefits most from temporal information carried forward. Under $\mathrm{GRE}_{80\%}$ (Figure 5b), where 80% of station observations are applied to constrain estimates, the improvements are smaller but still predominantly positive: when stations are dense, they already constrain the field well, and the forecast adds less.

The seasonal breakdown (Figure 5c,d) sharpens this picture. The forecast helps most in autumn and winter, when synoptic patterns persist from day to day and the previous day's analysis is a reliable predictor of the current state. It helps least in summer, when convective variability breaks the day-to-day connection and the forecast background becomes less informative. This dependence, with forecast information mattering most when stations are sparse and the weather is persistent, is exactly the behavior expected of a useful background field in any assimilation system. It confirms that the forecast model is not merely a generic smoother but provides physically grounded temporal context that complements the spatial prior, and it is what allows GRE to maintain both accuracy and calibrated uncertainty through the observational gaps that pervade the early instrumental record.

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026



m

Case	assi. vars	P (mm/day) spread	P (mm/day) error	sp (hPa) spread	sp (hPa) error	U_{10m} (m/s) spread	U_{10m} (m/s) error	T_{2m} (K) spread	T_{2m} (K) error
31/07	P, U_{10m}, T_{2m}	0	0	1.8	–	0	0	0	0
	U_{10m}, T_{2m}	40.9	10.5	3.0	–	0	0	0	0
	–	58.8	25.3	7.4	–	0.1	0.1	4.2	0.04
25/08	P, U_{10m}, T_{2m}	0	0	2.6	–	0	0	0	0
	U_{10m}, T_{2m}	97.8	24.5	3.1	–	0	0	0	0
	–	98.1	51.5	7.6	–	0.1	0.01	2.6	0.2

Figure 4. Reconstruction at a single grid point under progressive observation removal, for a 45-day window containing two major precipitation events (31 July, 74 mm/day; 25 August, 151 mm/day; red vertical lines). Columns show the four reconstructed variables: precipitation (P), surface pressure (sp), 10-m wind speed (U_{10m}), and 2-m temperature (T_{2m}). **(a–d)** All observed variables (P, U_{10m}, T_{2m}) assimilated. **(e–h)** Precipitation withheld on the two event days; wind speed and temperature remain assimilated. **(i–l)** All variable observations withheld on the event days; the reconstruction relies on the prior and the forecast from the previous day. Blue curves and shading show the ensemble mean and range; black dots mark assimilated observations; red diamonds mark observations that exist but are withheld. Surface pressure is not observed at this location; its spread in all three experiments is governed entirely by cross-variable coupling and the forecast background. **(m)** Ensemble spread and absolute error on the two event days for each experiment, summarizing the progressive transfer of uncertainty as observations are removed.

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

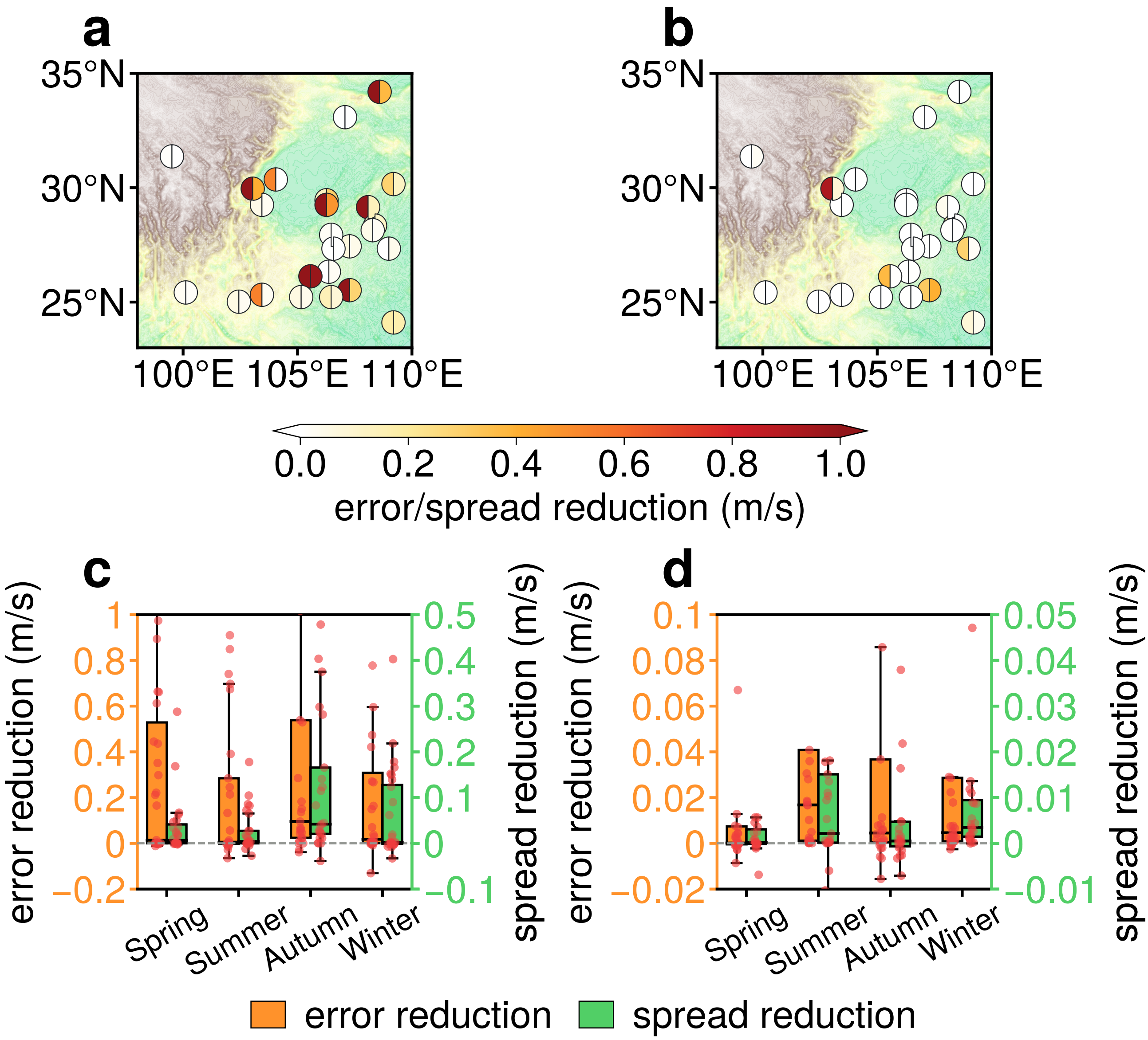


Figure 5. Change in wind-speed reconstruction skill when the forecast background is included, evaluated across the station network. Results are calculated based on data covering the full year of 1951. All panels show the difference with forecast minus without forecast, positive (red) values indicate improvement from including the forecast. **(a)** Spatial map for $GRE_{50\%}$ (conditioned on 50% of stations). At each station, the left semicircle shows the reduction in RMSE of ensemble mean estimates, and the right semicircle shows the reduction in ensemble spread; background shading indicates terrain elevation. **(b)** Same as (a) but for $GRE_{80\%}$ (conditioned on 80% of stations). **(c)** Seasonal breakdown for $GRE_{50\%}$. Orange boxplots show error reduction and green boxplots show spread reduction across stations. **(d)** Same as (c) but for $GRE_{80\%}$.

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

3.5 Calibrated Ensemble Uncertainty Through Network Evolution

The controlled experiments of Sections 3.2–3.4 established that GRE's ensemble spread responds appropriately to the removal of observations at individual stations and grid points. A historical reconstruction, however, must maintain this calibration over multi-decadal timescales as the real station network evolves, a far more demanding test, because network growth is gradual, spatially heterogeneous, and confounded with seasonal and interannual variability. We evaluate this using the 1951–1981 record over the observationally covered subregion (23°–35°N, 98°–110°E), a window that spans the most dramatic transformation of the gauge network, from fewer than fifty reporting stations in the early 1950s to roughly six hundred by the early 1980s (Figure 6d–f). Restricting the comparison to this subregion ensures that both GRE and the reference product are evaluated where the historical gauge network actually provides constraint; over the data-void western plateau, neither product is observationally anchored.

A well-calibrated ensemble must satisfy two specific properties over this period. First, *uncertainty should scale with statistical order*: the spread of a single-gridcell extreme-value statistic such as the daily spatial maximum of precipitation (P_{max}) should far exceed the spread of a spatially integrated quantity such as the domain-averaged daily precipitation (P_{mean}), because spatial averaging cancels local errors while extremes depend on resolving individual grid cells. Second, *uncertainty should decrease as observational support grows*: the ensemble spread should contract as the gauge network densifies, reflecting the tightening observational constraint rather than producing a fixed uncertainty band. We test both properties using P_{max} and P_{mean} computed daily over the subregion, two statistics that probe complementary physics: P_{max} isolates the single most intense local event each day and is governed by mesoscale and convective dynamics, whereas P_{mean} integrates over the domain and is dominated by the large-scale monsoon circulation.

The ensemble passes both tests. P_{max} carries an order of magnitude more uncertainty than P_{mean} throughout the record. In the sparse early 1950s ($\sim$40–80 gauges), the P_{max} spread frequently reaches 60–120 mm/day (Figure 6d), reflecting genuine ambiguity about where and how intense the day's heaviest rainfall is when few gauges sample the domain. By contrast, P_{mean} spread remains below 1–2 mm/day even during this period (Fig. S6 in the Supplement): even a thin network constrains the domain-integrated rainfall well, because local errors average out. This order-of-magnitude contrast (same ensemble, same network, same day, different statistic) is exactly the behavior expected of a physically consistent uncertainty estimate and would be absent from an ensemble whose spread reflected numerical noise rather than genuine observational ambiguity.

The gauge-density dependence is equally clear. As the network grows through the late 1950s, the P_{max} spread contracts sharply and then stabilizes into a seasonal rhythm once the network reaches its modern density (Figure 6f). The long-term anticorrelation between spread and gauge count persists across all three decades, with residual summer maxima reflecting the irreducible difficulty of constraining convective extremes even under dense coverage. This contraction occurs over real, gradual network evolution rather than the synthetic withholding of Sections 3.2–3.4, and it demonstrates that the ensemble is responding to actual changes in observational support at the multi-decadal scale.

Despite the wide uncertainty in extremes during the sparse early record, the GRE ensemble envelope encloses the CMFD P_{max} series throughout the three decades (Figure 6a), reproducing both the pronounced annual cycle (summer-monsoon max-

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

ima routinely exceeding 120 mm/day, winter minima below 30 mm/day) and the interannual modulation of peak intensity, without drift or systematic amplitude bias as the network changes. The zoomed June–August views for a sparse year (1952, Figure 6b) and a dense year (1980, Figure 6c) illustrate the mechanism at the event scale: in both years the ensemble follows individual synoptic episodes, but in 1952 the envelope is visibly wider. For P_{mean}, agreement is tight throughout (Fig. S6 in the Supplement): the reconstruction conserves the regional water budget across the full record, confirming that the wide early-record spread in P_{max} reflects localized extreme-value uncertainty rather than a bulk wet or dry bias. GRE and CMFD share the same station inputs and training-era climatology, so this consistency is expected where the network is dense; the distinguishing property is that where the network is sparse, GRE provides an explicit uncertainty envelope that CMFD does not.

Together, these results extend the single-point and single-year calibration of Sections 3.2–3.4 to the multi-decadal, network-scale regime. An ensemble whose spread tracks observational support, discriminates between statistics of different order, and remains bounded throughout gives downstream users, whether analyzing multi-decadal trends in precipitation extremes, evaluating regional climate model hindcasts, or assessing hydrological hazards during the pre-satellite era, a principled basis for weighting the reconstruction: wide spread signals genuine ambiguity that should temper interpretation, while narrow spread signals a well-constrained estimate.

4 Conclusions

Long-term, high-resolution climate records are fundamental to diagnosing regional extremes, evaluating climate models, and informing hydrological design. However, the station networks that underpin them thin and shift as one moves back through the instrumental record. Conventional gridded products cope with this thinning through prescribed background-error covariances, which inevitably smooth localized extremes and lose amplitude fidelity precisely when accurate intensity estimates matter most.

This study demonstrates that a learned generative prior, trained jointly on multiple surface variables from satellite-era data, can replace these prescribed covariances within a sequential assimilation framework (Lorenc, 1986; Kalnay, 2003; Evensen, 2003). The resulting daily 0.1° ensemble analyses over Southwest China (1951–2024) preserve localized precipitation extremes that the baseline product attenuates, reconstruct withheld station observations more faithfully than conventional interpolation, and maintain physically coherent multivariate fields, including surface pressure, which is not directly observed but is constrained through its learned co-variability with the observed variables. The forecast background contributes most where the early record is most vulnerable: under sparse coverage, in complex terrain, and during synoptically persistent seasons. Across all experiments, ensemble spread tracks the strength of the available observational constraint, widening when observations are removed and contracting when they are restored, without diverging. Extended over three decades of real network evolution (1951–1981), the reconstructed daily precipitation extremes over the gauge-covered subregion remain statistically consistent with the reference product while the ensemble spread contracts in step with the densifying network, and the spatially averaged precipitation stays tightly constrained throughout (Section 3.5; Figure 6 and Fig. S6 in the Supplement), demonstrating that the calibration established in controlled withholding experiments carries over to the historical record itself.

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

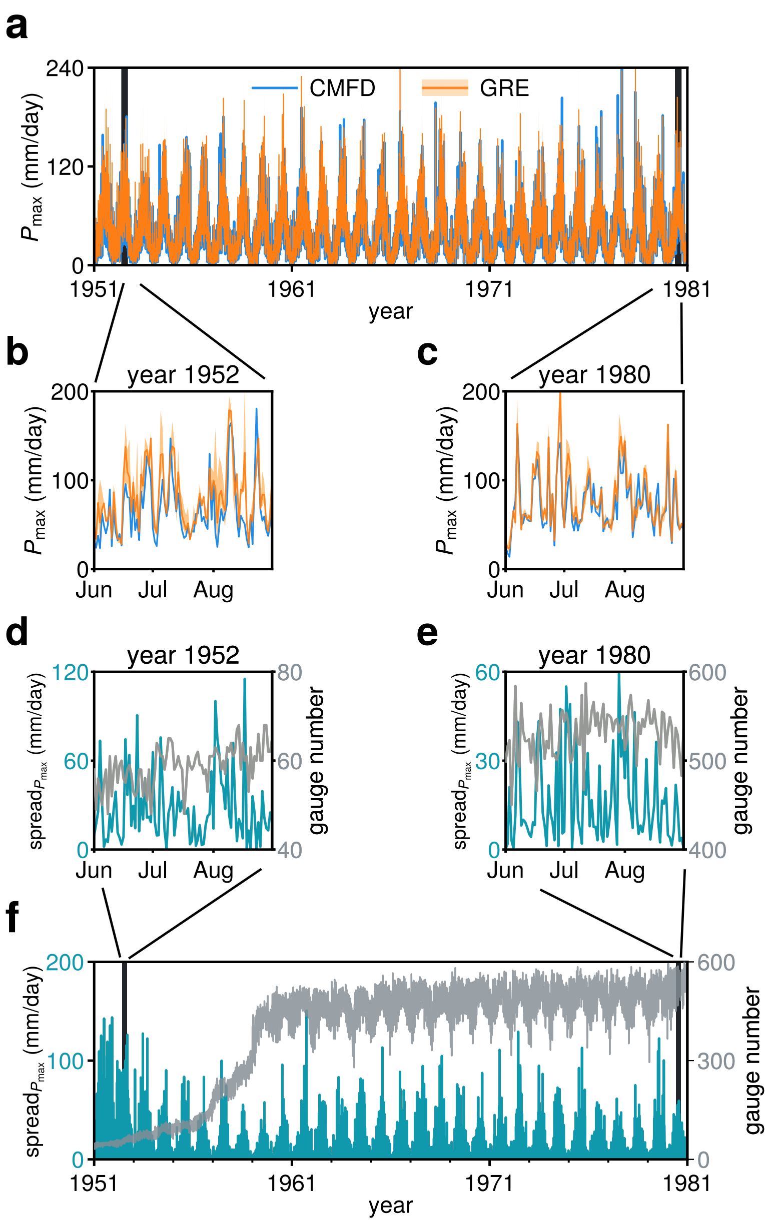


Figure 6. Calibrated ensemble uncertainty over three decades of network evolution. All statistics are computed daily over the observationally covered subregion (23°–35°N, 98°–110°E) for 1951–1981. **(a)** Daily spatial maximum of precipitation ($P_{\max}$, mm/day) from CMFD (blue) and the GRE ensemble (orange shading spans the ensemble range). Vertical black bars mark the two years selected for detailed comparison. **(b, c)** June–August $P_{\max}$ for a sparse-network year (1952) and a dense-network year (1980). **(d, e)** GRE ensemble spread of $P_{\max}$ (cyan) and number of reporting gauges (gray) for 1952 and 1980; note the difference in spread magnitude between the two years. **(f)** Long-term ensemble spread of $P_{\max}$ (cyan) and gauge count (gray), showing the contraction of uncertainty as the network densifies through the late 1950s and the persistent anticorrelation over the full three decades. The companion analysis for the spatially averaged daily precipitation (P_{mean}), which shows an order-of-magnitude smaller spread throughout, is shown in Fig. S6 in the Supplement.

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

These properties make the framework well suited to constructing long-term, high-resolution climate records for regions where station density has changed substantially over time. The resulting ensemble analyses could support studies of multi-decadal trends in precipitation extremes (Zhai et al., 2005; Fowler et al., 2021), evaluation of regional climate model hindcasts against observational benchmarks, cross-validation against independent gauge-based precipitation datasets such as CHM_PRE V2 (Hu et al., 2025), and assessment of hydrological hazards during periods that predate satellite coverage.

Two aspects of the current framework warrant further development. First, the diffusion model is trained on the satellite-era climatology (1981–2020) and treats the spatial prior as stationary. In reality, large-scale boundary forcing evolves on interannual to decadal timescales: ENSO modulates monsoon precipitation across Southwest China, and secular warming shifts the temperature and moisture fields. The prior therefore cannot represent spatial patterns that fall outside the satellite-era training distribution, for instance, teleconnection-driven rainfall anomalies during an ENSO phase that was rare in the training period. A natural extension is to condition the prior on slowly varying covariates such as sea-surface temperature indices or large-scale circulation fields, which are available from observational records and reanalyses extending back well over a century. This would allow the prior to adapt to the prevailing boundary forcing rather than defaulting to a fixed climatological envelope.

Second, classical sequential filters weight the forecast background by a flow-dependent error covariance, so that the analysis trusts the background more when it agrees with the observations and less when it does not. The current framework uses fixed guidance weights: the background receives the same influence regardless of how well the forecast matches the stations reporting on that day. A poor forecast, for example following a rapid regime transition that the simple next-day model cannot capture, would therefore pull the analysis away from the observations rather than being down-weighted. A practical remedy is to compare the forecast with the available station data before each analysis step and scale the background guidance accordingly: strong agreement would increase the weight, while large discrepancies would reduce it, mimicking the adaptive gain of an ensemble Kalman filter. Implementing such flow-dependent guidance would further improve calibration during regime transitions and convective episodes where day-to-day predictability is low.

Beyond these two directions, broader extensions include scaling the framework to continental and global domains, assimilating satellite retrievals and paleoclimate proxies alongside station data, and pushing reconstructions into the late-19th century, where sparse pressure and temperature records could anchor the analysis in combination with the prior and long-record boundary forcing.

Code and data availability. The exact version of the GRE source code used for this study is archived on Zenodo at https://doi.org/10.5281/zenodo.21525755 (Chao, 2026) and is released under the Creative Commons Attribution 4.0 International licence.

The China Meteorological Forcing Dataset (CMFD) version 2.0 used to train both the diffusion prior and the forecast model is available from the National Tibetan Plateau / Third Pole Environment Data Center at https://doi.org/10.11888/Atmos.tpdc.302088 (He et al., 2024); its CSTR persistent identifier is https://cstr.cn/18406.11.Atmos.tpdc.302088. The GEBCO_2026 Grid used as the digital elevation model is available from the NERC EDS British Oceanographic Data Centre NOC at https://doi.org/10.5285/4f68d5c7-45eb-f999-e063-7086abc036fa (GEBCO Bathymetric Compilation Group 2026, 2026). Daily station observations of precipitation, 10-m wind speed and 2-m temperature are provided by the China Meteorological Administration through the China Meteorological Data Service Centre (China Meteorological

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

Data Service Centre, 2026); the data are accessed via https://data.cma.cn/en/ under the data provider's terms of use and no DOI is assigned. Solar angle (SA) is computed from longitude, latitude and date using the open-source `Pysolar` Python library.

Author contributions. **BP, QL, AS:** conceptualization, methodology, funding acquisition. **JC:** formal analysis, software, data curation, writing – original draft. **HY:** data curation. **ZL, JX, SG:** validation, investigation. **ZH, HS:** resources, computational infrastructure. **ZX:** investigation, visualization. **BP, QL, AS:** supervision, writing – review and editing. All authors discussed the results and commented on the manuscript.

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. This work is supported by the Yunnan Fundamental Research Program (Major Project): Spatiotemporal Evolution, Mechanisms, and Risk Prediction of Winter–Spring Compound Drought over Yunnan (Grant No. 202501BC070001). We thank Sencan Sun (Tsinghua University), Shangshang Yang (Technical University of Munich), and Peili Wu (UK Met Office) for valuable discussions and feedback on this work.

References

Anderson, B. D.: Reverse-time diffusion equation models, Stochastic Processes and their Applications, 12, 313–326, 1982.

Chao, J.: GRE: Official PyTorch code of "Generative reconstruction of high-resolution historical climate fields from station observations", Zenodo, https://doi.org/10.5281/zenodo.21525755, version "code", archived 24 July 2026; concept DOI https://doi.org/10.5281/zenodo.21525754, 2026.

Chao, J., Yang, S., Chen, G., Huang, C., Ling, F., and Luo, J.-J.: Learning to infer weather states using partial observations, Journal of Geophysical Research: Machine Learning and Computation, 2, e2024JH000 492, https://doi.org/10.1029/2024JH000492, 2025.

China Meteorological Data Service Centre: China Meteorological Data Service Centre, https://data.cma.cn/en/, last access: 24 July 2026, 2026.

Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., and Ye, J. C.: Diffusion posterior sampling for general noisy inverse problems, in: International Conference on Learning Representations, 2023.

Efron, B.: Tweedie's formula and selection bias, Journal of the American Statistical Association, 106, 1602–1614, https://doi.org/10.1198/jasa.2011.tm11181, 2011.

Evensen, G.: The ensemble Kalman filter: theoretical formulation and practical implementation, Ocean Dynamics, 53, 343–367, https://doi.org/10.1007/s10236-003-0036-9, 2003.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., et al.: Evaluation of climate models, in: Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 741–866, Cambridge University Press, 2014.

Fowler, H. J., Lenderink, G., Prein, A. F., Westra, S., Allan, R. P., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Do, H. X., Guerreiro, S., Haerter, J. O., Kendon, E. J., Lewis, E., Schaer, C., Sharma, A., Villarini, G., Wasko, C., and Zhang, X.: Anthropogenic intensification of short-duration rainfall extremes, Nature Reviews Earth & Environment, 2, 107–122, https://doi.org/10.1038/s43017-020-00128-6, 2021.

GEBCO Bathymetric Compilation Group 2026: The GEBCO_2026 Grid — a continuous terrain model for oceans and land at 15 arc-second intervals, NERC EDS British Oceanographic Data Centre NOC, https://doi.org/10.5285/4f68d5c7-45eb-f999-e063-7086abc036fa, 2026.

He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., and Li, X.: The first high-resolution meteorological forcing dataset for land process studies over China, Scientific Data, 7, 25, https://doi.org/10.1038/s41597-020-0369-y, 2020.

He, J., Yang, K., Li, X., Tang, W., Shao, C., Jiang, Y., and Ding, B.: China meteorological forcing dataset v2.0 (1951–2024), National Tibetan Plateau / Third Pole Environment Data Center, https://doi.org/10.11888/Atmos.tpdc.302088, cSTR: https://cstr.cn/18406.11.Atmos.tpdc.302088, 2024.

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, R. J., 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.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

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

Hu, J., Miao, C., Su, J., Zhang, Q., Gou, J., and Sun, Q.: An upgraded high-precision gridded precipitation dataset for the Chinese mainland considering spatial autocorrelation and covariates, Earth System Science Data, 17, 3987–4004, https://doi.org/10.5194/essd-17-3987-2025, 2025.

Huang, Z., Huang, Y., Zhang, Y., Ren, K., and Chen, L.: Diffusion-based data assimilation from sparse station observations, arXiv preprint, 2024.

Kadow, C., Hall, D. M., and Ulbrich, U.: Artificial intelligence reconstructs missing climate information, Nature Geoscience, 13, 408–413, https://doi.org/10.1038/s41561-020-0582-5, 2020.

Kadow, C., Illing, S., Lucio-Eceiza, E. E., Bergemann, M., Kunst, O., Schartner, T., Pankatz, K., Hall, D. M., and Ulbrich, U.: Artificial intelligence reveals past climate extremes by reconstructing historical records, Nature Communications, 15, 9191, https://doi.org/10.1038/s41467-024-53389-y, 2024.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, 2003.

Kingma, D., Salimans, T., Poole, B., and Ho, J.: Variational diffusion models, Advances in neural information processing systems, 34, 21 696–21 707, 2021.

Ling, F., Luo, J.-J., Li, Y., Tang, T., Bai, L., Ouyang, W., and Yamazaki, T.: Diffusion model-based probabilistic downscaling for 180-year East Asian climate reconstruction, npj Climate and Atmospheric Science, 7, 131, https://doi.org/10.1038/s41612-024-00638-0, 2024.

Liu, B., Tang, Q., Zhao, G., Gao, L., Shen, C., and Pan, B.: Physics-guided long short-term memory network for streamflow and flood simulations in the Lancang–Mekong river basin, Water, 14, 1429, 2022.

Lorenc, A. C.: Analysis methods for numerical weather prediction, Quarterly Journal of the Royal Meteorological Society, 112, 1177–1194, https://doi.org/10.1002/qj.49711247414, 1986.

Luo, J.-J., Xia, J., Pan, B., Ham, Y.-G., Li, X., Shangguan, W., Xue, W., Wang, Y., Mu, B., Hong, Y., et al.: AI for atmosphere–ocean sciences: advancements, challenges and ways forward, National Science Review, 13, nwag063, 2026.

Manshausen, P., Cohen, Y., Watson-Parris, D., and Stier, P.: Generative data assimilation of sparse weather station observations at kilometer scales, arXiv preprint, 2025.

Mardani, M., Brenowitz, N., Cohen, Y., Pathak, J., Chen, C.-Y., Liu, C.-C., Vahdat, A., Kashinath, K., Kautz, J., and Pritchard, M.: Residual corrective diffusion modeling for km-scale atmospheric downscaling, Communications Earth & Environment, 6, 58, https://doi.org/10.1038/s43247-025-02040-1, 2025.

Nai, C., Pan, B., Chen, X., Tang, Q., Ni, G., Duan, Q., Lu, B., Xiao, Z., and Liu, X.: Reliable precipitation nowcasting using probabilistic diffusion models, Environmental Research Letters, 19, 034 039, 2024.

Nai, C., Chen, X., Yang, S., Xiao, Z., and Pan, B.: Boosting weather forecast via generative superensemble, npj Climate and Atmospheric Science, 8, 377, 2025.

Ou, Z., Nai, C., Pan, B., Zheng, Y., Shen, C., Jiang, P., Liu, X., Tang, Q., Li, W., and Pan, M.: Probabilistic diffusion models advance extreme flood forecasting, Geophysical Research Letters, 52, e2025GL115 705, 2025.

Pan, B., Hsu, K., AghaKouchak, A., and Sorooshian, S.: Improving precipitation estimation using convolutional neural network, Water Resources Research, 55, 2301–2321, 2019.

Pan, B., Wang, L.-Y., Zhang, F., Duan, Q., Li, X., Pan, X., Chen, X., Ling, F., Wang, S., Pan, M., et al.: Probabilistic diffusion model for stochastic parameterization–a case example of numerical precipitation estimation, Authorea Preprints, 2023.

Price, I., Sanchez-Gonzalez, A., Alet, F., Andersson, T. R., El-Kadi, A., Masters, D., Ewalds, T., Stott, J., Mohamed, S., Battaglia, P., et al.: Probabilistic weather forecasting with machine learning, Nature, 637, 84–90, 2025.

https://doi.org/10.5194/egusphere-2026-4341
Preprint. Discussion started: 17 August 2026

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, https://doi.org/10.1038/s41586-019-0912-1, 2019.

Schneider, T.: Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values, Journal of Climate, 14, 853–871, https://doi.org/10.1175/1520-0442(2001)014<0853:AOICDE>2.0.CO;2, 2001.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics, in: International conference on machine learning, pp. 2256–2265, PMLR, 2015.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B.: Score-based generative modeling through stochastic differential equations, in: International Conference on Learning Representations, 2021.

Song, Y., Shen, L., Xing, L., and Ermon, S.: Solving inverse problems in medical imaging with score-based generative models, in: International Conference on Learning Representations, 2022.

Vincent, P.: A connection between score matching and denoising autoencoders, Neural Computation, 23, 1661–1674, https://doi.org/10.1162/NECO_a_00142, 2011.

Wan, J., Yang, T., Yu, Q., Liu, R., Li, W., Song, H., Wang, X., Wang, J., Xue, F., Xiao, Z., et al.: Generative machine learning for skilful 3D radar nowcasting, npj Climate and Atmospheric Science, 2026.

Wang, J., Chao, J., Yang, S., Nai, C., Ren, K., Deng, K., Chen, X., Liu, Y., Han Wen, Q., Xiao, Z., et al.: Generative high-resolution ensemble weather forecast supports renewable energy planning and operation, Geophysical Research Letters, 53, e2025GL119 044, 2026.

Wu, J. and Gao, X. J.: A gridded daily observation dataset over China region and comparison with the other datasets, Chinese Journal of Geophysics, 56, 1102–1111, https://doi.org/10.6038/cjg20130406, 2013.

Xu, Z. X., Gong, T. L., and Li, J. Y.: Decadal trend of climate in the Tibetan Plateau — regional temperature and precipitation, Hydrological Processes, 22, 3056–3065, https://doi.org/10.1002/hyp.6892, 2008.

Yang, S., Chao, J., Chen, G., Huang, C., Ling, F., and Luo, J.-J.: Generative assimilation and prediction for weather and climate, arXiv preprint, 2025.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N., and Kitoh, A.: APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges, Bulletin of the American Meteorological Society, 93, 1401–1415, https://doi.org/10.1175/BAMS-D-11-00122.1, 2012.

Zhai, P., Zhang, X., Wan, H., and Pan, X.: Trends in total precipitation and frequency of daily precipitation extremes over China, Journal of Climate, 18, 1096–1108, https://doi.org/10.1175/JCLI-3318.1, 2005.

Zhang, Q., Xu, C.-Y., Zhang, Z., Chen, Y. D., Liu, C.-L., and Lin, H.: Spatial and temporal variability of precipitation maxima during 1960–2005 in the Yangtze River basin and possible association with large-scale circulation, Journal of Hydrology, 353, 215–227, https://doi.org/10.1016/j.jhydrol.2007.11.023, 2008.