Contents of this file

- Section S1 Details of the diffusion model
- Section S2 Details of the forecast model
- Fig. S1 Number of stations during 1951–1990 and terrain elevation
- Fig. S2 Statistical comparison of sp from CMFD and GRE
- Fig. S3 Statistical comparison of U_{10m} from CMFD and GRE
- Fig. S4 Statistical comparison of T_{2m} from CMFD and GRE
- Fig. S5 Inter-variable correlation matrix
- Fig. S6 Long-term comparison of spatially averaged daily precipitation
- Table S1 Hyperparameter configuration of the diffusion model

S1 Details of the diffusion model

The spatial prior is a score-based diffusion model 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 SDE

$$d\mathbf{x}_\tau = f_\tau(\mathbf{x}_\tau)\,d\tau + g_\tau\,d\mathbf{w}_\tau, \quad \tau\in[0,1], \tag{S1}$$

progressively transforms $\mathbf{x}_0$ into Gaussian noise $\mathbf{x}_1 \sim \mathcal{N}(\mathbf{0},\mathbf{I})$, where f_τ and g_τ are defined through a log signal-to-noise ratio (log-SNR) parameterization that yields closed-form expressions for $\alpha(\tau)$ and $\sigma(\tau)$, and $\mathbf{w}_\tau$ is a standard Wiener process. This process admits a time-reversed SDE

$$d\mathbf{x}_\tau = \left[f_\tau(\mathbf{x}_\tau) - g_\tau^2\,\nabla_{\mathbf{x}_\tau}\log p_\tau(\mathbf{x}_\tau \mid \mathbf{c})\right]d\bar{\tau} + g_\tau\,d\bar{\mathbf{w}}_\tau, \tag{S2}$$

where $d\bar{\tau}$ denotes reverse time and $\mathbf{c} = [\mathrm{DEM};\mathrm{SA}]$ encodes static topographic and solar forcing attributes. In practice, the reverse dynamics are solved using a discretized deterministic sampler equivalent to a DDIM formulation, which improves sampling efficiency while preserving consistency with the learned score field.

The score function $\nabla_{\mathbf{x}_\tau}\log p_\tau(\mathbf{x}_\tau \mid \mathbf{c})$ is approximated by a neural network s_θ, implemented as a multi-scale convolutional U-Net with residual blocks, group normalization, and skip connections. The network takes as input the noisy state $\mathbf{x}_\tau$ concatenated with conditioning variables $\mathbf{c}$, and is modulated by a sinusoidal time embedding that is injected into each residual block via scale–shift transformations. The encoder progressively reduces spatial resolution using pooling-based downsampling, while the decoder restores resolution via bilinear upsampling, allowing the model to capture spatial dependencies across scales. Rather than directly predicting noise, the model adopts a velocity parameterization (v-prediction), which has been shown to stabilize training under log-SNR schedules.

Training follows denoising score matching, where noisy samples $\mathbf{x}_\tau$ are constructed analytically as $\mathbf{x}_\tau = \alpha(\tau)\mathbf{x}_0 + \sigma(\tau)\boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I})$. The network is trained to predict the corresponding velocity target $\mathbf{v} = \alpha(\tau)\boldsymbol{\epsilon} - \sigma(\tau)\mathbf{x}_0$ by minimizing a mean squared error loss over valid grid points. The model is trained on daily four-variable fields from CMFD v2.0 (1981–2020), with precipitation and wind speed log-transformed to reduce skewness and all variables standardized using climatological statistics. Static fields (DEM and solar altitude) are concatenated as additional input channels, enabling the model to learn spatially varying climatological regimes associated with terrain and seasonal forcing. Missing values are handled through masking, and the loss is computed only over valid regions.

During inference, samples are generated by integrating the reverse diffusion process starting from Gaussian noise. To incorporate observational constraints, we employ diffusion posterior sampling (DPS), in which the reverse update is augmented with a gradient-based correction that enforces consistency between the reconstructed field and available constraints. Constraints are imposed at fixed intervals (every 5 diffusion steps), while all other steps follow the unconditional reverse dynamics. Two conditioning strategies are used: (1) constraints provided solely by sparse observations, and (2) sequential constraints in which forecast guidance is first applied and then replaced by observational constraints at the same interval. At constrained steps, the predicted clean state $\hat{\mathbf{x}}_0$ is partially replaced at known locations and further refined using the gradient of a squared-error objective with respect to $\mathbf{x}_\tau$, whereas unconstrained steps rely entirely on the learned diffusion prior. This intermittent repainting strategy enforces consistency with observations and/or forecast guidance without over-constraining the dynamics, allowing spatial information to propagate into unobserved regions. The resulting samples are spatially coherent, physically consistent, and conditioned on both learned climatological structure and sparse constraints.

Key hyperparameters are summarized in Table S1.

S2 Details of the forecast model

The forecast model F_ξ is implemented as a deterministic convolutional U-Net with a multi-scale encoder–decoder architecture. The network consists of a series of residual blocks with group normalization and nonlinear activation, organized across multiple spatial resolutions. The encoder progressively downsamples the input using pooling-based operations to capture large-scale

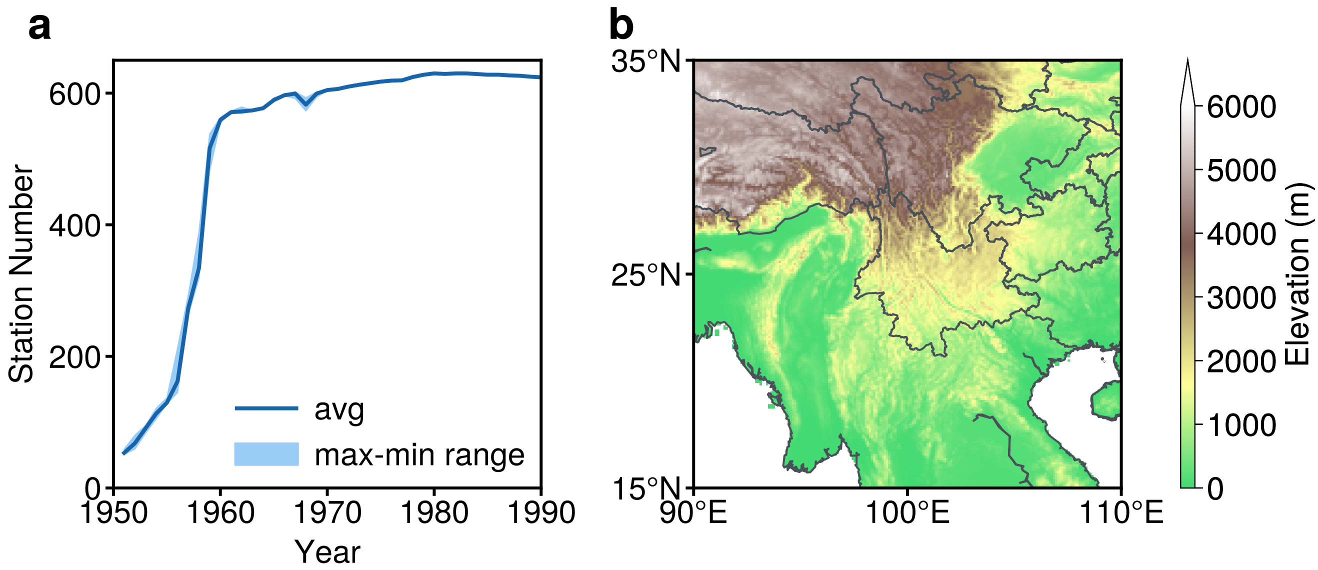

Figure S1. Number of stations during 1951–1990 (a) and terrain elevation (b).

spatial patterns, while the decoder restores resolution via bilinear upsampling. Skip connections are introduced between corresponding encoder and decoder stages to preserve fine-scale spatial information. The input to the network is the four-variable field $\mathbf{x}_t \in \mathbb{R}^{4\times H\times W}$, and the output has the same dimensionality, representing the predicted next-day state $\mathbf{x}_{t+\Delta t}$.

The model is trained on consecutive daily field pairs from CMFD v2.0, using the period 1981–2015 for training and reserving 2016–2020 as an independent validation set. Training minimizes the mean squared error:

$$\mathcal{L} = \mathbb{E}_t \left| F_\xi(\mathbf{x}_t) - \mathbf{x}_{t+\Delta t} \right|^2, \tag{S3}$$

where $F_\xi(\mathbf{x}_t)$ denotes the model-predicted next-day field from the input state $\mathbf{x}_t$, $\mathbf{x}_{t+\Delta t}$ is the corresponding target field, and the expectation $\mathbb{E}$ is taken over all consecutive daily training pairs.

On the validation set, the deterministic forecast model achieves MAEs of 1.9 mm/day for P, 0.8 hPa for sp, 0.2 m/s for U_{10m}, and 0.7 K for T_{2m}. These errors summarize the one-day predictive skill of F_ξ and support its use as a temporal constraint in the conditional diffusion sampling process.

During inference, the forecast $\tilde{\mathbf{x}}_{t+\Delta t} = F_\xi(\mathbf{x}_t)$ is incorporated into the diffusion sampling procedure as a temporal constraint at selected steps, providing temporal guidance while allowing the diffusion prior to recover spatial variability.

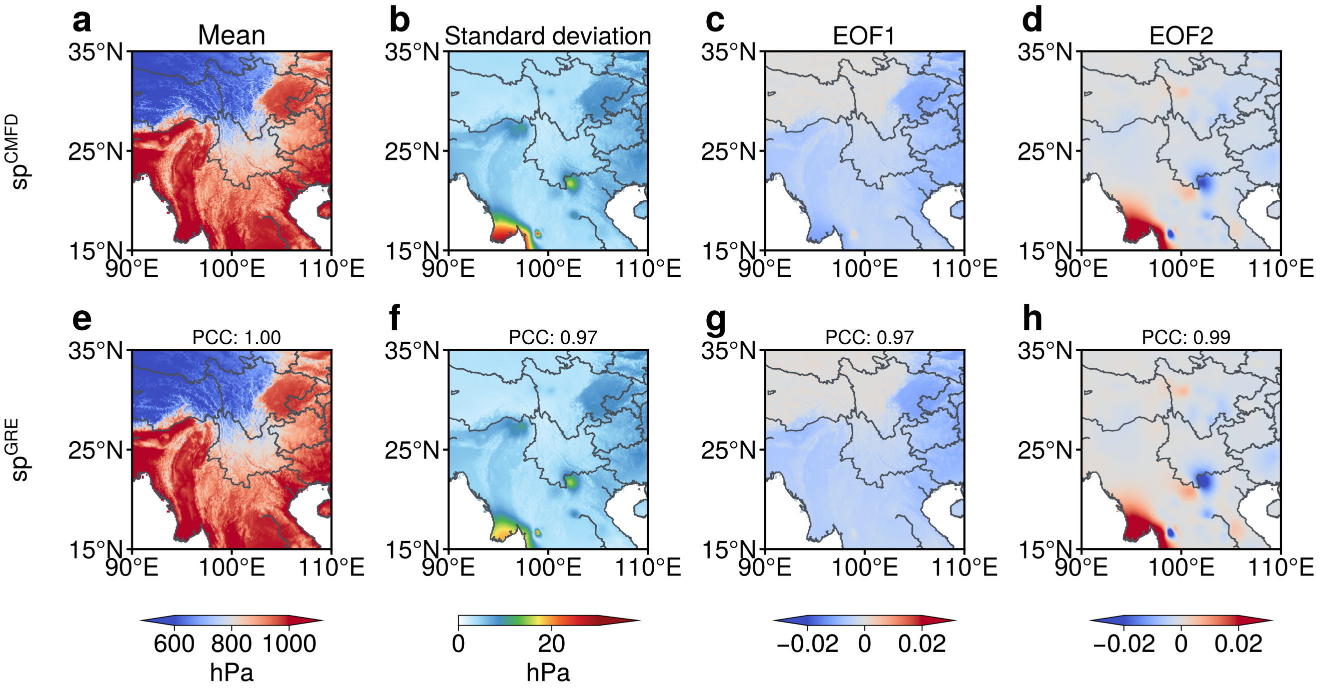

Figure S2. Statistical comparison of sp from CMFD (a–d) and GRE (e–h), including mean, variance, and the leading two EOF modes (EOF1 and EOF2).

Table S1. Hyperparameter configuration of the diffusion model.

Component	Value
Grid size	200×200
Variables	P, sp, U_{10m}, T_{2m}
Input channels	4 + 2 (static attributes)
Output channels	4
Base channel dimension	64
Channel multipliers	(1, 2, 4, 4)
Diffusion steps (training)	1000
Noise parameter range	$\lambda = 20$
Batch size	4
Learning rate	1×10^{-5}
Optimizer	AdamW
Gradient clipping	1.0
Sampler steps (inference)	200
Observation-guidance weight	$\lambda_o = 1$
Forecast-guidance weight	$\lambda_f = 1$

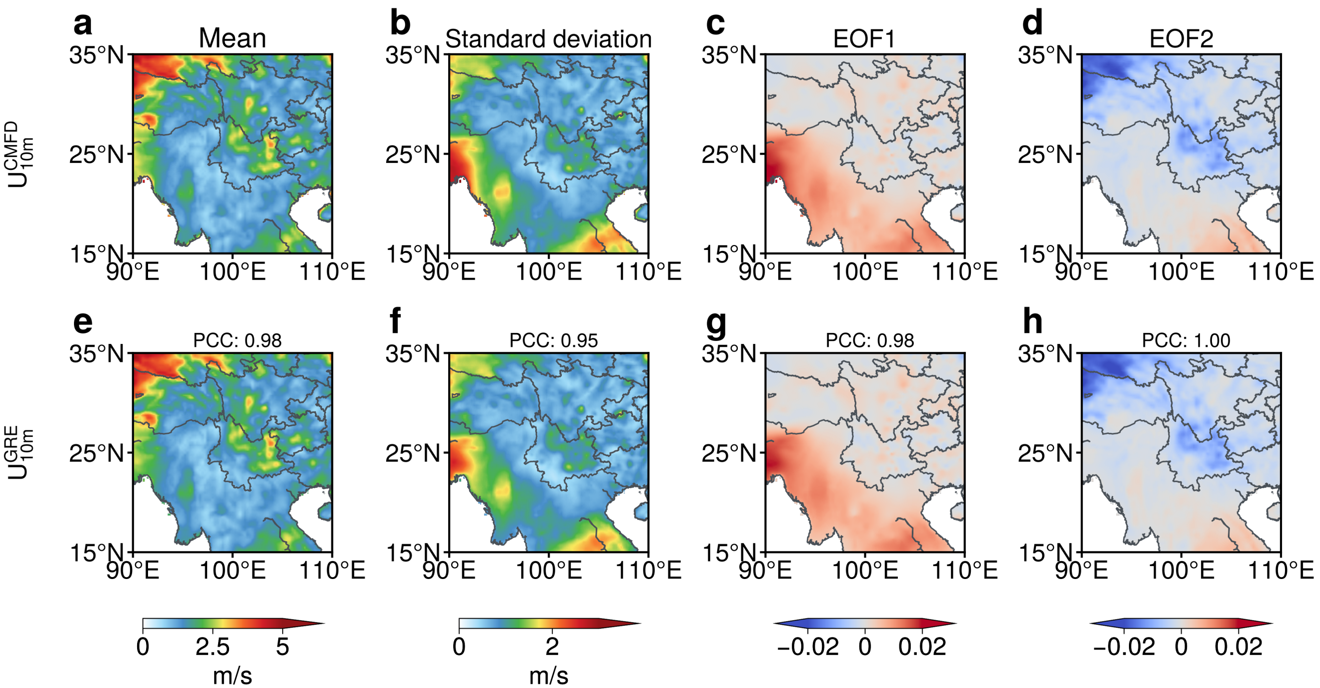

Figure S3. Statistical comparison of U_{10m} from CMFD (a–d) and GRE (e–h), including mean, variance, and the leading two EOF modes (EOF1 and EOF2).

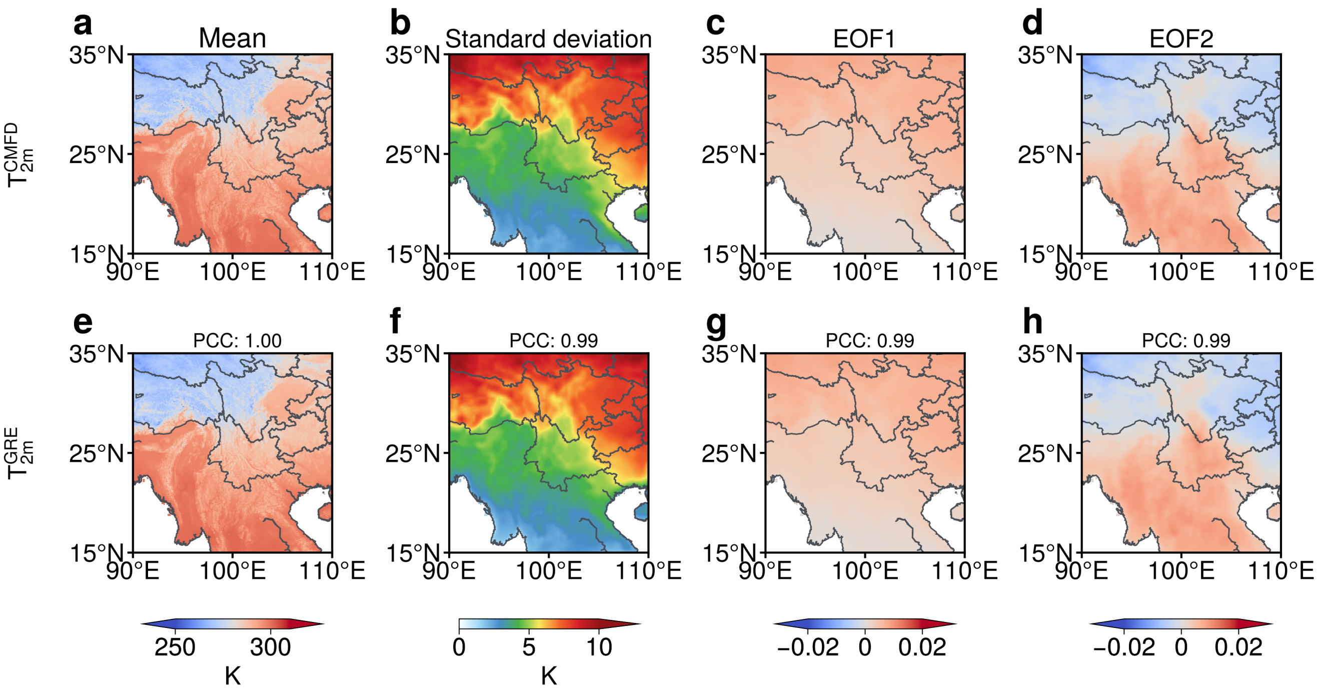

Figure S4. Statistical comparison of T_{2m} from CMFD (a–d) and GRE (e–h), including mean, variance, and the leading two EOF modes (EOF1 and EOF2).

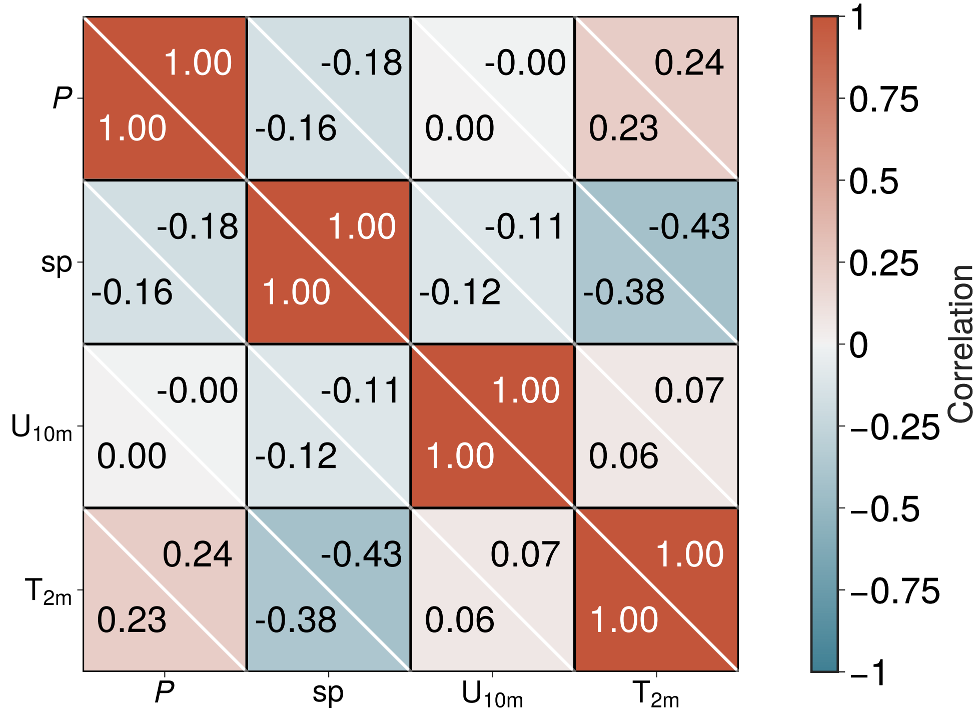

Figure S5. Inter-variable correlation matrix among precipitation (P), surface pressure (sp), 10-m wind speed (U_{10m}), and 2-m temperature (T_{2m}). In each cell the upper value is computed from the CMFD training data and the lower value from the 10,000 unconditionally generated GRE fields, showing that the prior reproduces the multivariate (cross-variable) correlation structure.

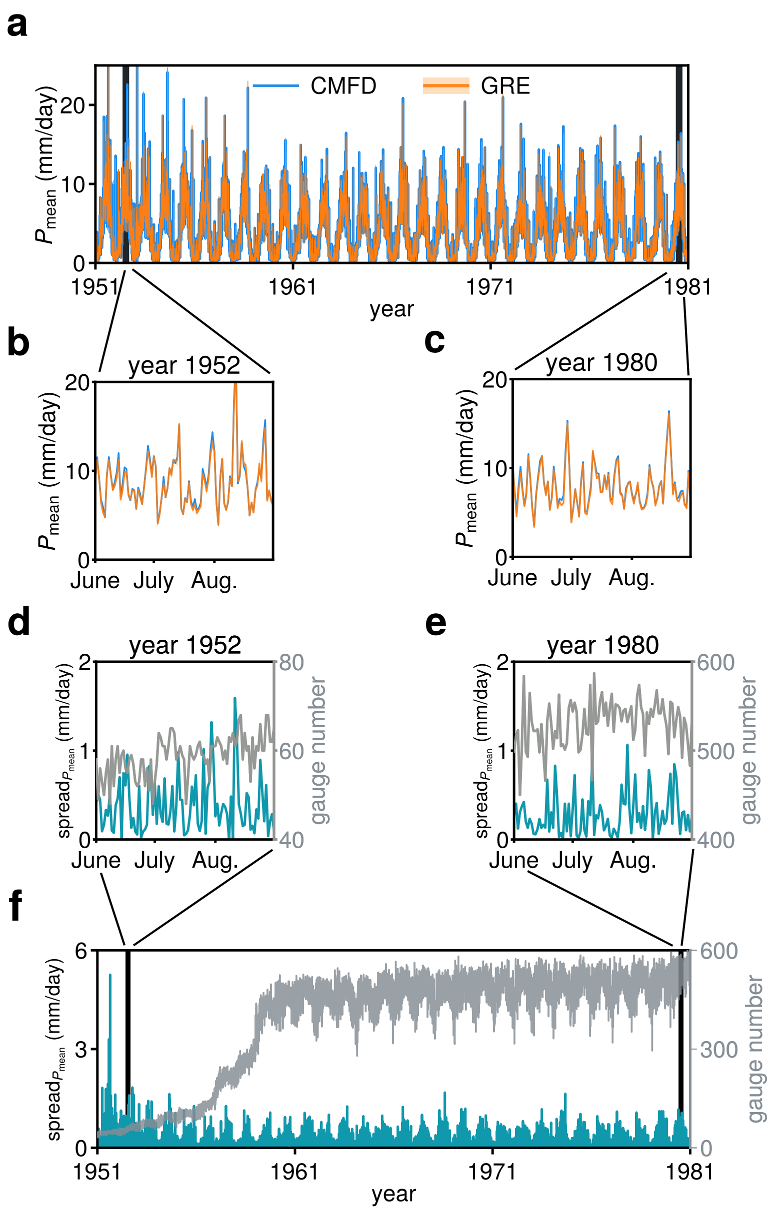

Figure S6. Long-term comparison of spatially averaged daily precipitation between CMFD and the GRE ensemble over the observationally covered subregion (23°–35°N, 98°–110°E) for 1951–1981. (a) Time series of the daily spatial mean of precipitation (P_{mean}, mm/day), from CMFD (blue) and the GRE ensemble (orange; shading spans the ensemble range). Vertical black bars mark the two years selected for detailed comparison (1952, sparse network; 1980, dense network). (b, c) Zoomed-in views of June–August P_{mean} for 1952 and 1980, respectively. (d, e) Corresponding GRE ensemble spread of P_{mean} (spread$_{P_{mean}}$, mm/day, cyan) and the number of reporting gauges in the subregion (gray) for 1952 and 1980. (f) Long-term time series of the GRE ensemble spread (cyan) and gauge number (gray).