# Probabilistic downscaling of paleoclimate simulations using diffusion models

Trang Nguyen[1,2], Markella Bouchorikou[1,2], Dominik Fleitmann[3], Morteza Djamali [4], Christoph Raible[1,2]

[1] Climate and Environmental Physics, University of Bern, Bern, Switzerland

[2] Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland [3]Quaternary Geology, University of Basel, Switzerland

[4] Institut Méditerranéen de Biodiversité et d'Ecologie, French National Centre for Scientific Research (CNRS), France

## Introduction

- Traditional Earth system models are computationally intensive, making high-resolution regional downscaling for long-term paleoclimate simulations infeasible using conventional dynamical approaches.
- We applied CorrDiff, a state-of-the-art two-step diffusion downscaling framework, to paleoclimate data (CESM model). It uses a UNet for mean prediction and a diffusion model for stochastic variability and uncertainty quantification.
- The target is WRF data, the output of traditional dynamical downscaling from CESM data.

## Study area

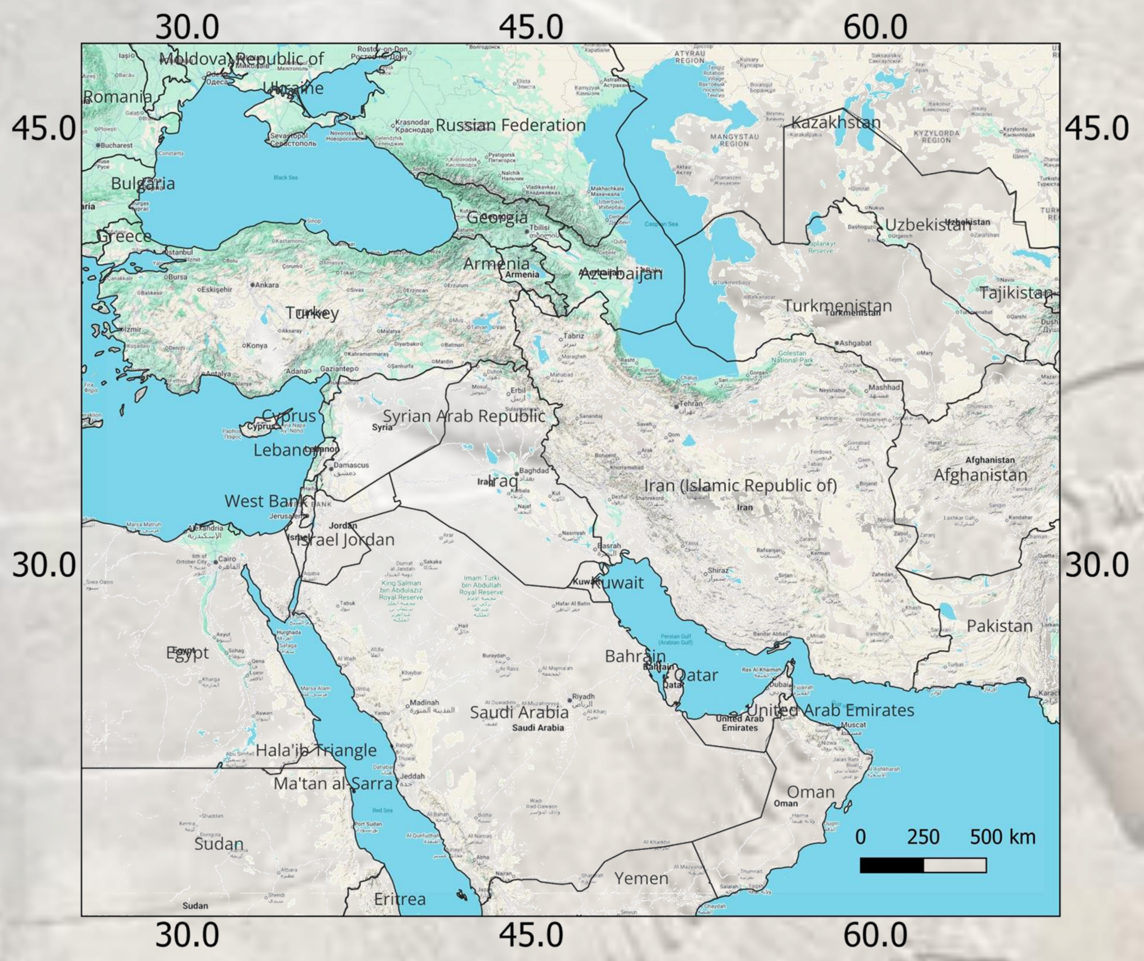


Figure 1. Southwest Asia - a climatologically dynamic region situated at the intersection of the North Atlantic, monsoonal, and Siberian systems

- Arid and Semi-Arid Climate dominates the area.
- Rainfall is highly seasonal.

## Data & Model

| | | |
|---|---|---|
| Inputs (from CESM) | t | 3D var - 500 and 850 hPa |
| | z | |
| | u | |
| | v | |
| | rh | |
| | TREFHT (surface air temperature) | 2D variables |
| | U10 | |
| | V10 | |
| | PSL (sea level pressure) | |
| | DOY_sin | Trigonometric encodings of the Day of Year (DOY) |
| | DOY_cos | |
| Inputs (from WRF) | Landmask | Static variables |
| | DEM | |
| Outputs (from WRF) | T2M | 2D variables |
| | Precip | |

Table 1. The model uses 18 input channels from CESM coarse-resolution data (1*2 deg spatial resolution, grid size 11 × 12 over the domain) and predicts 2 output channels at WRF high-resolution (10 km spatial resolution, grid size 245 × 295). All data are at daily timesteps.

### 1. Regression

| Model Architecture | Training Duration | Optimizer | Batch Settings | Performance | Regularization |
|---|---|---|---|---|---|
| Base channels: 128 | 500,000 | Type: Adam | Total batch size: 64 | Dataloader workers: 4 | Gradient clip: 1.0 |
| Channel multipliers: [1,2,2,2,2] | | LR: 0.0001 | Batch per GPU: 8 | | |
| Attention resolutions: [28] | | Ramp-up: 10,000 | Grad accum: Auto | | |
| | | Decay factor: 0.5 | | | |
| | | Decay rate: 500,000 | | | |



## Results

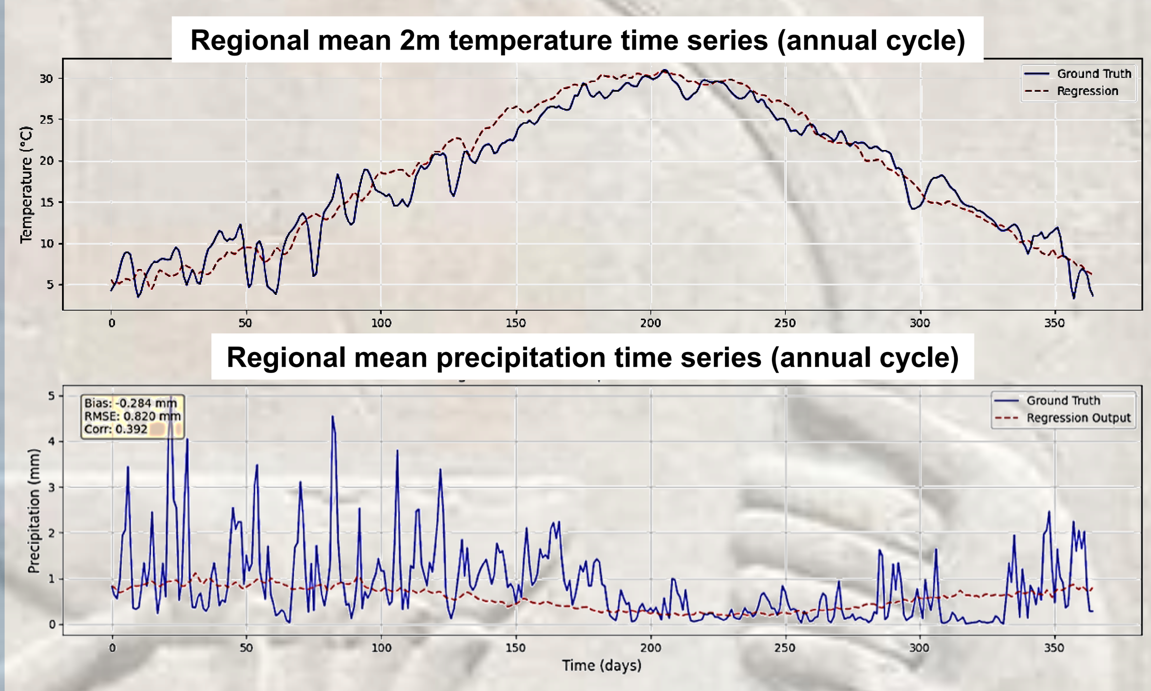


Figure 2. Time series of spatially averaged 2m temperature and precipitation from regression output and target (WRF data) over one year of testing data.

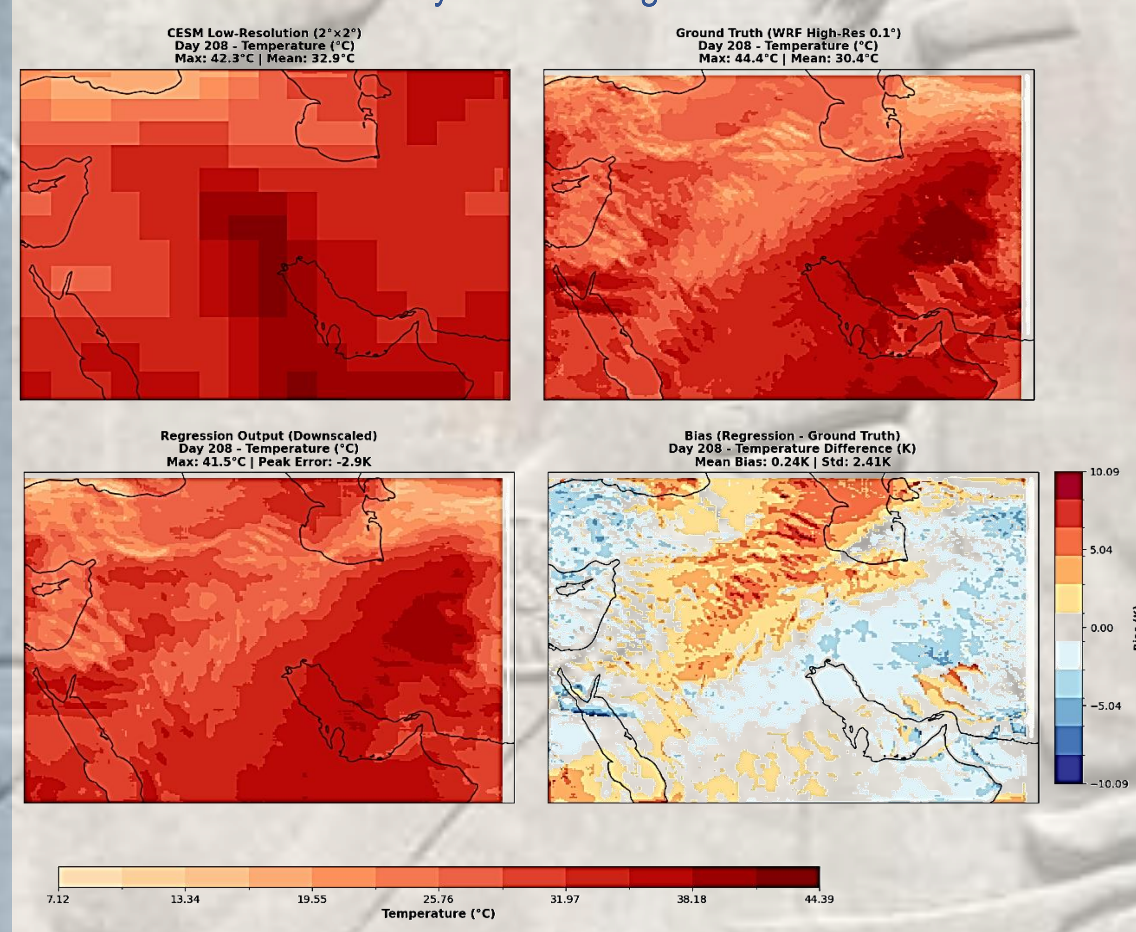


Figure 3. Comparison of 2m temperature maps on July 27, 1937: interpolated low-resolution CESM data, WRF target data, and the regression and the bias.

## Model

- Training period: 1934 to 1936, testing period: 1937.
- Preprocessing: Interpolate CESM to grid of WRF (bicubic), fill NaN by 0, Z score normalization, structure to patches (32*32 cells) with 50% overlapped
- Loss function:
  - Regression used standard MSE (Mean Squared Error) with no time-dependent weighting
  - Diffusion used denoising score matching loss (MSE-based)

### 2. Diffusion

| Category | Parameter | Value |
|---|---|---|
| Model Architecture | Base channels | 128 |
| | Channel multipliers | [1, 2, 2, 2, 2] |
| Number of ensembles | | 4 |
| | Attention resolutions | [28] |
| Training Duration | Total training images | 500,000 |
| Optimizer | Type | Adam |
| | Learning rate (lr) | 0.0001 |
| | LR ramp-up steps | 10,000 |
| | LR decay factor | 0.5 |
| | LR decay rate | 500,000 steps |
| Batch Settings | Total batch size | 8 |
| | Batch size per GPU | 4 |
| | Gradient accumulation | Automatic (8/4/GPUs) |
| Diffusion Schedule | Number of timesteps | 1,000 |
| | Beta schedule | linear |
| | Beta start | 0.0001 |
| | Beta end | 0.02 |
| | P_mean (noise mean) | -1.2 |
| | P_std (noise std) | 1.2 |
| Performance | Data loader workers | 8 |
| Regularization | Gradient clip threshold | 1.0 |

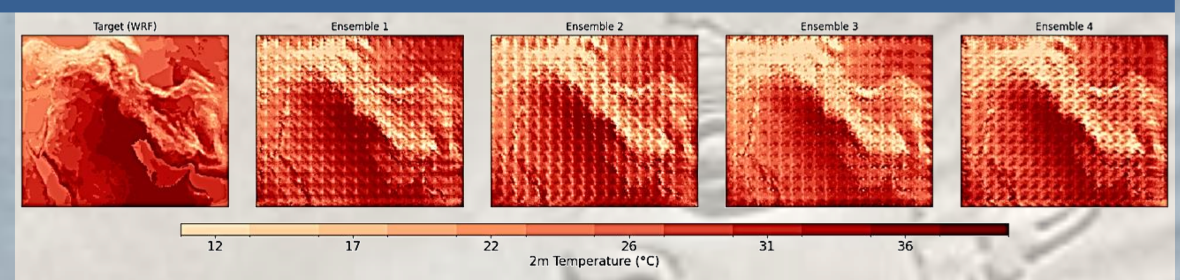


Figure 4. 2m temperature maps of WRF data (target) and 4 ensemble members of the diffusion model on July 27, 1937.

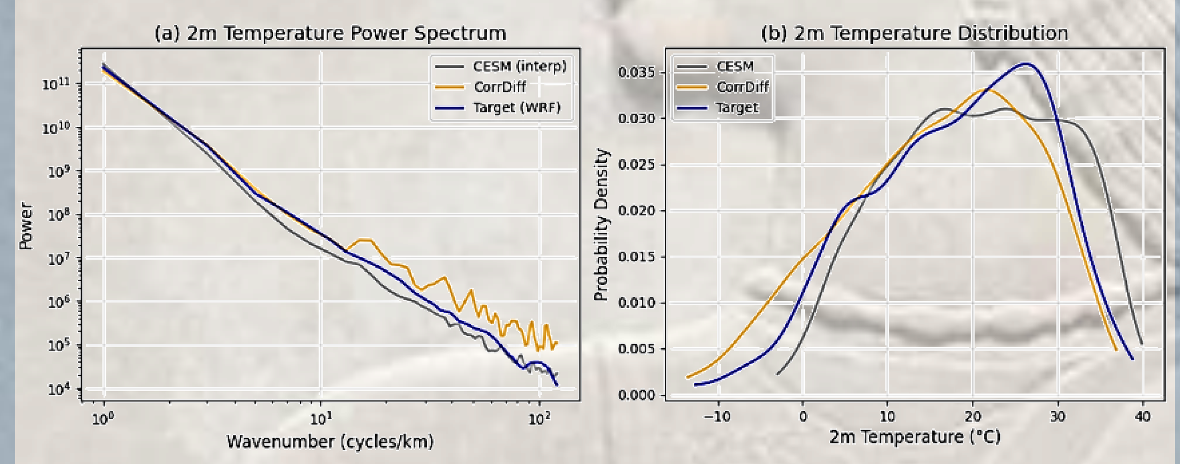


Figure 5. The power spectra and distributions are compared for CESM (after interpolation), WRF – target and CorrDiff prediction for 2m temperature.

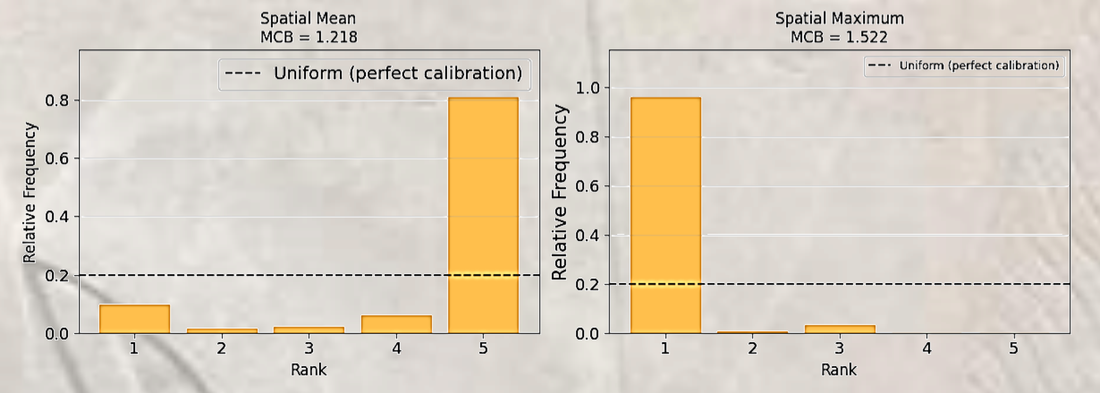


Figure 6. Rank histogram for 2m temperature for 4 ensemble members in CorrDiff prediction

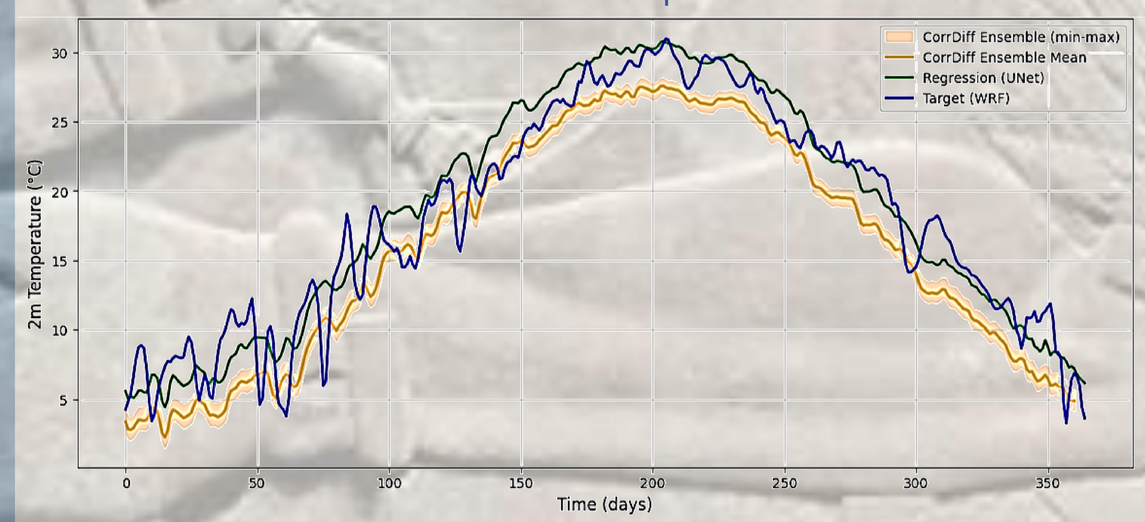


Figure 7. Time series of spatially averaged 2m temperature: comparison of WRF target data, regression output, and CorrDiff ensemble mean (diffusion output).

| Deterministic | Energy Score | CRPS | Spectral | Range |
|---|---|---|---|---|
| MSE: 65.64 (°C)$^2$ | ES: 6.63 | Pointwise: 5.13°C | RALSD: 3.86 dB | GT min: -28.9°C |
| RMSE: 8.10°C | ESpred: 8.33°C | CRPS-mp: 4.51°C | | GT max: 44.4°C |
| | ESvar: 1.71°C | Spatial loss: -0.61°C | | Diff min: -32.6°C |
| | Balance: 6.63 | | | Diff max: 48.6°C |

Table 2. Evaluation for 2m temperature CorrDiff prediction regarding mean squared error (MSE), energy score loss (ES), prediction error (ESpred), variability (ESvar), CRPS averaged over all grid points (CRPS) and after max-pooling across patches of size 10 (CRPS-mp), as well as Radially-averaged Log Spectral Distance (RALSD). Metrics are averaged across all days to evaluate the overall performance on the entire dataset.

## Conclusion & Outlooks

- CorrDiff reproduces the mean pattern of 2m temperature but not precipitation (Fig. 2, Fig. 3).
- The diffusion output exhibits patch edge artifacts and a cold bias in 2m temperature predictions (Fig. 5, Fig. 6, Fig. 7).
- The CorrDiff prediction of 2m temperature is overconfident (ensemble spread is too narrow), the ensemble mean is far from the truth, the temperature range is too wide, and the model predicts both more extreme cold and more extreme heat than actually occur (Tab. 2 and Fig.6).

**Outlooks:**

- Revise the loss function to improve the performance of the diffusion model.
- For precipitation: Apply normalization based on the ECDF of precipitation to preserve dry-day frequency at each grid point.
- Train the model with conditioning on the previous time step to improve temporal sequence learning.

## Reference

Mardani, M., Brenowitz, N., Cohen, Y. *et al.* Residual corrective diffusion modeling for km-scale atmospheric downscaling. *Commun Earth Environ* **6**, 124 (2025). https://doi.org/10.1038/s43247-025-02042-5