Project: **1529**

Project title: **SKOP. Beyond the Limits of Climate Intelligence**

Principal investigator: **Ivanka Pelivan**

Report period: **2025-11-01 to 2026-10-31**

1. Objectives and work completed

At this project we investigate whether probabilistic machine learning can move beyond conventional bias correction of CMIP6 projections and improve station-scale climate information for renewable-energy. The completed work focuses on surface wind: approximately 1,000 globally distributed stations, hourly NOAA observations (GHCNh weather stations) and multiple CMIP6 models and scenarios obtained from DKRZ Levante. A reusable processing and evaluation workflow now separates marginal distribution fidelity, conditional probabilistic skill and annual resource uncertainty. This implements the large-scale validation objective of the ESA Philab project.

Quantile Delta Mapping (QDM) was designated as the classical reference baseline in the project design. The completed memberwise quantile-mapping experiments provide the conventional bias-correction benchmark. Histogram-based Gradient Boosting (HGB) was introduced as the first machine-learning baseline, before the more flexible Mixture Density Network (MDN) stage.

2. Machine-learning baseline and quantitative results

HGB uses training-derived month-by-hour anomalies and nine ensemble summary features to estimate 99 conditional wind-speed quantiles. Model selection uses probabilistic loss, followed by quantile rearrangement and reconstruction in physical units. Expected wind is obtained by integrating the quantile function. We also tested empirical quantile mapping (EQM) after HGB to correct the pooled marginal distribution.

Method	Wasserstein (m/s)	Mean resource error (%)	CRPS (m/s)
Raw CMIP6	1.01426	27.859	1.40969
Classical bias correction	0.12226	2.746	1.20045
HGB	0.13617	2.671	1.14977
HGB + EQM	0.12368	2.745	1.14948

Table 1. Mean resource error is the absolute period-mean deviation for bias correction and the expected-wind mean error for HGB; it is not annual P50. CRPS denotes the Continuous Ranked Probability Score.

Classical correction reduced Wasserstein distance by approximately 88% relative to raw CMIP6. HGB achieved about 4.2% lower CRPS than classical correction and slightly lower mean-resource error, but initially poorer marginal fidelity. Post-HGB EQM reduced HGB Wasserstein distance by 9.17%, improving about 67% of stations and bringing the marginal result close to classical correction. It left CRPS almost unchanged. These results identify a calibration vs. conditional-skill trade-off, rather than uniform machine-learning superiority.

An important methodological correction replaced averages of hourly conditional medians with integrated conditional expectations. In the matched diagnostic, mean-resource error fell from 7.58% to 2.70%. This corrected the annual estimator rather than demonstrating an equivalent increase in predictive skill, and prevents misleading interpretation of resource estimates.

3. Scientific findings and research decisions

The experimental programme tested loss-based model selection, broader hyperparameter searches, separate configurations for low/central/high quantiles, percentile-based predictors and affine/EQM post-calibration. Larger searches, three-band models and percentile inputs produced negligible gains; short-period affine calibration worsened transfer. These negative results eliminate unnecessary complexity and focus further computing on better-supported approaches.

Stronger reference tests revealed the remaining limitation. HGB achieved approximately +6.72% CRPS skill against unconditional history, but -0.74% against month-by-hour probabilistic climatology. **Monthly correlation was approximately 0.61**, while climatology was around 0.56. Historical distributions also remained competitive in marginal accuracy (ie. Wasserstein Distance). The work therefore establishes effective calibration, while showing that additional climate-conditioned skill must be demonstrated separately.

Annual uncertainty is a distinct research target: accurate hourly quantiles do not define dependence between hours, and therefore do not by themselves yield a validated predictive annual P90. Plain HGB achieved 2.89% error in the annual expected-wind P50 diagnostic, approximately tied with history; post-EQM slightly worsened it. These findings direct the next stage towards conditional density representation and temporal dependence, not further routine HGB tuning.

4. Levante use and ongoing MDN experiments

Levante has supported CMIP6 data access, CPU-parallel BC and HGB training and large-scale benchmarking. **With 99 quantile models at approximately 1,000 stations**, a full HGB configuration involves roughly **100,000 final quantile fits, before hyperparameter screening and repeat experiments**.

MDN experiments are currently running on the separate HPC6 system because GPUs were quickly available there, enabling faster iteration on deep-learning architectures and training stability. This complements Levante's CPU and CMIP6 data-access role. MDNs learn mixture weights, means and variances to represent conditional distributions; their comparative skill remains under evaluation.

5. Results in Germany

Eighteen weather stations located in Germany have been used in these experiments, results in Germany are much better than the global average:

	Wasserstein Distance	Mean resource error	CRPS	Mean resource error over the year
Bias correction	0.149	**0.0202**	0.595	-
HGB + EQM	0.154	0.0232	**0.577**	0.020
Historical	**0.143**	0.0215	0.599	**0.018**

Table 2. Mean resource error is the absolute period-mean deviation for bias correction and the expected-wind mean error for HGB; it is not annual P50. CRPS denotes the Continuous Ranked Probability Score.

Significantly, **HGB + EQM is better in 13 stations out of 18 in terms of reconstruction of the mean resource error on a yearly basis** (over 10 years) in Germany.

6. Outputs, recognition and next steps

The project has delivered a station-scale benchmark corpus, a first machine-learning reference, comparative result archives and evidence-based architecture decisions. The project has been published on Destination Earth (**5th User Exchange Workshop**):

Destination_Earth_Geoskop-Beyond-the-Limits-of-Climate-Intelligence