Author Comments on egusphere-2026-223

Manuscript: *Deep learning for gridding climate variables from sparse station networks*

We would like to thank the reviewers for their thorough and constructive feedback. Below we provide a point-by-point response to each comment. Where applicable, we reference new or revised figures and sections in the manuscript.

AC1: Reply to RC1 (Anonymous Referee #1)

RC1-C1 — Context vs. target data distributions

> *"Does the input data (context) distribution fit the distribution of the target points? How valid are the model outside the training data 'window'? The distribution of the geographic parameters for both context and target points as well as for the grid cells should be presented and discussed."*

AC1-1:

We thank the reviewer for this important point. We have added a new figure (Figure 3) and accompanying discussion comparing the distributions of the three key static predictors — smoothed elevation, topographic position index, and distance to coast — across three groups: (a) training (context) stations ($n = 130$), (b) held-out test (target) stations ($n = 26$), and (c) all prediction grid cells within Romania's national territory ($n = 38{,}144$).

The figure shows that the training and test station distributions are well-matched across all three predictors (e.g., mean smoothed elevation: 0.22 vs. 0.21), confirming that the random 130/26 split is representative. The prediction grid falls almost entirely within the training data envelope — only 0.4% of grid cells exceed the maximum training-station elevation and 0.1% exceed the topographic position range, corresponding to the highest peaks where no stations exist. The station network covers the full range of predictor values encountered on the grid, though the grid has a stronger mode at low elevations, consistent with the fact that lowlands and hills/plateaus together cover roughly two-thirds of Romania's territory.

We have added a discussion of these findings at the end of Section 2.2 ("Spatio-temporal covariates") in the revised manuscript.

RC1-C2 — RK description too limited

> *"The RK description is too limited. A description of the deterministic trend and the semi-variogram parameters should be presented and discussed."*

AC1-2:

We agree that the RK baseline deserved a more complete description. We have expanded Section 3.1 and Appendix A to include:

1. **Deterministic trend.** The linear regression model uses all static and temporal predictor variables as covariates — specifically: smoothed elevation, topographic position index, distance to the Black Sea coast, latitude, longitude (geographic coordinates), and cyclic temporal encodings (sine/cosine of day-of-year and month, plus a linear year trend). For temperature, the model is fitted on all context stations; for precipitation amount, it is fitted only on stations reporting precipitation > 0 mm (the "rainy subset").

2. **Semi-variogram model.** The residuals from the deterministic trend are interpolated using Ordinary Kriging with a **spherical** semi-variogram model. The variogram parameters — partial sill, range, and nugget — are estimated automatically for each daily field using PyKrige's weighted least-squares fitting procedure. Because the variogram is re-fitted for every day, the parameters vary across the time series, adapting to the daily spatial structure of the residual field. We note this design choice rather than reporting fixed parameter values, since the variogram characterizes the spatial correlation of the residuals after removing the trend, and this structure changes daily.

RC1-C3 — SMACNP description too detailed

> *"The description of the SMACNP is very detailed, and these details overshadow the main message. Consider using the flow chart in figure 3 (or even a simplified version of it) as a guide to the presentations of the method. Move details that are not necessary for further reading, or not discussed further, to appendix. Be also sure to define all abbreviations, terms and symbols the first time they appear."*

AC1-3:

We agree that the level of mathematical detail in the original Section 3.2 made the description harder to follow than necessary. We have restructured the section as follows:

1. **Reorganized around the architecture figure.** The revised Section 3.2 now follows the data flow shown in Figure 4 (formerly Figure 3) as a narrative guide: Input → Encoder (three pathways) → Decoder → Output distributions. Each component is explained conceptually — what it does and why — before referencing any equations.

2. **Moved detailed equations to a new Appendix C.** The attention score formulations (Laplace attention weights, multi-head cross-attention), the complete loss function derivations (Gaussian NLL, binary cross-entropy, log-normal NLL with class weighting), and the softmax/softplus activation details have been moved to Appendix C ("SMACNP Architecture Details"). The main text retains only the key

equations necessary to understand the architectural choices (e.g., the Laplace distance kernel to explain L1 vs. L2 distance selection).

3. **Defined all abbreviations at first use:**
   - MLP: Multi-Layer Perceptron — a feedforward neural network that applies a sequence of linear transformations with nonlinear activations
   - Q, K, V: Query, Key, and Value — the three projections used in the attention mechanism, where the Query represents "what information is being requested," the Key represents "what information is available," and the Value represents "the actual content to be retrieved"
   - Lp: the configurable norm distance (p = 1 for Manhattan/L1, p = 2 for Euclidean/L2) used in the Laplace attention kernel
   - BCE: Binary Cross-Entropy — a loss function measuring the discrepancy between predicted probabilities and binary labels (rain/no-rain)
   - NLL: Negative Log-Likelihood
4. **Simplified dense passages.** The Global vs. Localized encoder descriptions (original L.166–169) have been condensed into a table format comparing the two configurations side-by-side (input sources for Q/K/V, attention scope, computational complexity), making the differences immediately visible.

RC1-C4 — Figures: resolution and clarity

> *"Resolution and clarity of figures need to be improved. Figure 6 and 7 are very blurry. Figure 7 intends to include a lot of information that cannot be separated. Choose colors etc. that clearly distinguish the models. (This also applies to the supplementary material)."*

AC1-4:

We appreciate the reviewer's concern regarding figure quality. We have verified that all figures were generated at 300 DPI at the source — the blurriness observed in the review copy was caused by image compression during the DOCX-to-PDF conversion pipeline used for manuscript submission. The original PNG files are high-resolution (e.g., Figure 6: 5376 × 3322 px at 300 DPI; Figure 7: 6000 × 3600 px at 300 DPI). At full resolution, all model lines, confidence intervals, and annotations are clearly distinguishable.

To prevent this issue in the revised submission:

1. **All figures are now submitted as separate high-resolution files** alongside the manuscript, ensuring no compression artifacts are introduced during document conversion.
2. **Figure 1 (station locations map) has been substantially improved.** Beyond upgrading from 200 to 300 DPI, the revised figure now includes: (a) a **DEM altitude background** (hillshaded elevation raster) that immediately conveys Romania's

topographic complexity and the relationship between station locations and terrain, and (b) a **station elevation histogram** (inset panel) showing the altitude distribution of training stations ($n = 130$, blue) and test stations ($n = 26$, red) side by side. This addition directly complements the predictor distribution analysis in Figure 3 (see AC1-1) by providing at-a-glance confirmation that the train/test split covers the full elevation range of the station network.

3. **Color choices:** The model comparison figures use distinct, well-separated colors for each model (RK, SMACNP Global, SMACNP Localized), applied consistently across all main text and supplementary figures.

RC1-C5 — Minor comments

AC1-5:

We thank the reviewer for the detailed line-by-line suggestions. All items have been addressed in the revised manuscript:

- **L.60 — Long sentence.** The sentence describing the Localized configuration hyperparameters has been split into two shorter sentences for clarity.
- **L.144 — Explain "pathway" and "encoder"; emphasize Figure 3 as guide.** The revised Section 3.2 (see our response to Comment 3) now opens with a clear definition of the encoder-decoder architecture and explicitly uses the architecture figure (now Figure 4) as the structural guide for the entire section. Each pathway is introduced with a brief conceptual explanation before any technical detail.
- **L.146 — Define MLP.** Defined at first use in the revised Section 3.2 as: "Multi-Layer Perceptron (MLP) — a feedforward neural network applying successive linear transformations with nonlinear activations."
- **L.147 — Define Lp.** Defined at first use as: "a configurable Lp norm (p = 1 for Manhattan distance, p = 2 for Euclidean distance)."
- **L.150 — Describe the functional role of Q, K, V.** The revised Section 3.2 now includes: "Query (Q) represents 'what information is being requested' (derived from target features), Key (K) represents 'what information is available' (derived from context features), and Value (V) contains 'the content to retrieve' (derived from context features and observations)."
- **L.166–169 — Simplify and define terms.** The dense Global vs. Localized comparison has been replaced by a side-by-side table (see Comment 3 response) that makes the differences immediately visible without requiring the reader to parse long prose paragraphs.
- **L.205–206 — Explain binary cross-entropy.** Defined in the revised Section 3.2 as: "Binary Cross-Entropy (BCE) — a loss function measuring the discrepancy between

predicted probabilities and binary labels (rain/no-rain).” The full loss equation is provided in the new Appendix C.

RC1-C6 — Precipitation-temperature dependencies and univariate DL

> *“Could you elaborate on this a bit more? Is it so that there might be dependencies between precipitation and temperature that is captured by the deep learning method? Would e.g. co-kriging of precipitation using temperature as co-variate captured some of the same? And finally, for curiosity and further understanding. Would it be possible to run the deep learning as univariate approaches for temperature and precipitation respectively?”*

AC1-6:

We thank the reviewer for this thought-provoking question. We address the three aspects below:

1. **Temperature-precipitation dependencies in the multivariate DL model.** The SMACNP model is trained jointly on temperature and precipitation — both variables are observed at the same context stations and predicted simultaneously at the target locations. The encoder pathways produce shared latent representations from the combined observations, meaning the model can implicitly learn cross-variable dependencies. For example, topographic effects that jointly influence temperature (lapse rate) and precipitation (orographic enhancement) are encoded in the same latent space. Additionally, training on both variables simultaneously has a stabilising effect: spatial patterns learned from temperature fields — which tend to be smoother and more predictable — can inform the model’s precipitation predictions, particularly in data-sparse situations.

2. **Co-kriging as an alternative.** Co-kriging is indeed a geostatistical method that enables the simultaneous estimation of two or more spatially correlated variables by modelling their cross-variograms. In principle, it could exploit the spatial cross-correlation between temperature and precipitation to improve predictions of one variable using observations of the other. However, in our experimental setup, both variables are observed at the same stations and on the same days, so the practical benefit of co-kriging would primarily lie in improving the spatial prediction of one variable by leveraging the spatial structure of the other. The SMACNP framework achieves a similar effect through shared encoder representations, but with the added capacity to capture nonlinear cross-variable relationships.

3. **Univariate DL.** It is certainly possible to train separate SMACNP models for temperature and precipitation independently. However, the multivariate approach offers two advantages: (a) training a single model is more computationally efficient than training two separate models, and (b) the framework naturally scales to more than two variables — additional climate fields (e.g., wind speed, humidity, solar radiation) can be incorporated as additional output heads without requiring

separate models for each. A formal univariate ablation study comparing single-variable models against the joint approach would be an interesting direction for future work to quantify the benefit of multi-task learning in this context.

We have added a new paragraph in Section 5 (Discussions) to elaborate on these points.

RC1-C7 — Sensitivity to training period and applicability outside it

> *"The study covers a relatively short time period. How sensitive will the method be to the training data period, and how applicable is the method for periods outside the training period?"*

AC1-7:

The SMACNP model was designed with temporal transferability in mind. Its primary predictors are time-invariant geographic features (elevation, topographic position, distance to coast), and seasonality is captured through cyclic sine/cosine encodings that do not depend on the specific training years. A linear year trend accounts for long-term climate trends. Critically, the model is conditioned on actual station observations each day — it learns a spatial interpolation function, not a climate simulator — so it generalises to any period where the topography-climate relationships and station network remain comparable.

Although 2020–2023 covers only four years, it spans ~1,460 daily fields with randomised context/target splits, providing substantial variability across seasons, weather regimes, and data sparsity conditions. Extending the training period would improve exposure to rare extremes and is a priority for future work. We have added a paragraph in Section 5 (Discussion) elaborating on temporal transferability.

AC2: Reply to RC2

We thank the reviewer for the thorough and constructive evaluation. The comments have led to meaningful improvements in both the analysis and the presentation. Below we address each point in detail.

RC2-C1 — Baseline design and DL comparisons

> *"The biggest weakness is the baseline design. RK alone is too weak to support broad claims that the proposed DL framework is superior. The paper should include stronger non-DL baselines such as random forest, xgboost, spline-based interpolation, and possibly more advanced kriging variants. More importantly, it should include DL baselines: a plain MLP using station/query covariates, deepkriging-style models, GNN-based interpolation, etc., and perhaps a simpler attention/CNP baseline. The authors mention convcnp and griddedtnp struggled, but those results should be shown, not just stated."*

AC2-1:

We appreciate the reviewer's concern about baseline breadth. We would like to clarify that the choice of SMACNP was not arbitrary — it was the outcome of a deliberate preliminary architecture screening. Before committing to the full analysis presented in the paper, we evaluated several Neural Process variants, including **ConvCNP** (Gordon et al., 2019) and **GriddedTNP** (Ashman et al., 2024). SMACNP consistently emerged as the best-performing architecture during these preliminary experiments, which is why we selected it as the framework for the complete study. In the original submission, we mentioned these preliminary results only briefly in Section 5 (Discussions); following the reviewer's recommendation, we now report them quantitatively.

Both ConvCNP and GriddedTNP were trained and evaluated under the same protocol as SMACNP, using identical data splits, loss functions, and ground truth observations at the 26 held-out test stations over 2020–2023. The results are summarized below:

Temperature (daily mean, °C):

Model	MAE	RMSE	Corr	CRPS
RK	0.94	1.27	0.989	0.72
SMACNP Localized	**0.80**	**1.11**	**0.992**	**0.59**
GriddedTNP	1.05	1.42	0.988	0.81
ConvCNP	1.53	2.14	0.978	1.13

Precipitation (daily total, mm):

Model	MAE	RMSE	Corr	CRPS
RK	1.24	3.63	0.670	1.03
SMACNP Localized	**1.05**	**3.23**	**0.749**	**0.85**
ConvCNP	1.11	3.59	0.742	0.89
GriddedTNP	1.22	3.63	0.685	0.96

These results confirm the architectural selection made during the preliminary phase. SMACNP Localized is the best-performing model across both variables and all metrics. For temperature, GriddedTNP (MAE = 1.05°C) approaches the performance of Regression Kriging (0.94°C), while ConvCNP (1.53°C) falls further behind. For precipitation, both ConvCNP and GriddedTNP outperform RK and approach SMACNP Localized, demonstrating that grid-based deep learning architectures can effectively capture the large-scale spatial patterns that drive precipitation fields.

The contrasting performance across variables is informative and explains why SMACNP was the preferred architecture. Temperature prediction is dominated by local topographic relationships — particularly the elevation lapse rate — which require preserving fine-scale station-level information. Grid-based architectures (ConvCNP, GriddedTNP) discretize station observations onto a regular grid before processing, inherently smoothing these local contrasts. SMACNP's station-to-station attention mechanism preserves exact station attributes throughout the encoding and decoding process, enabling it to learn precise lapse-rate corrections. Precipitation, by contrast, is driven by larger-scale spatial patterns

(frontal systems, orographic enhancement) that are well-represented on regular grids, which explains the more competitive performance of ConvCNP and GriddedTNP for this variable. However, even for precipitation, SMACNP Localized maintains a clear advantage across all metrics.

The primary objective of our study is to identify a practical and reliable alternative to geostatistical interpolation methods — particularly Regression Kriging — that have been the standard approach for producing gridded climate datasets from national station networks for decades. Within this scope, the most relevant comparison is between the proposed SMACNP framework and the well-established RK method, which is why we structured the paper around this comparison. The SMACNP Localized model consistently outperforms all baselines across both variables, while also providing calibrated uncertainty estimates that RK does not natively produce.

Regarding the non-DL baselines (random forest, XGBoost, spline interpolation): while these are valuable tools in environmental modelling, they do not natively produce full predictive distributions — calibrated uncertainty estimates (predictive mean and variance) for continuous variables and a coherent joint model for precipitation occurrence and amount — which are essential requirements for a climate gridding framework intended for downstream applications such as drought monitoring, return-period estimation, and climate change impact assessment. For this reason, and because benchmark comparisons with machine learning regressors for spatial interpolation have been extensively reported elsewhere in the literature, we focused our evaluation on methods that produce full predictive distributions.

Following the reviewer's recommendation, we have added these quantitative results to Section 4.1 of the revised manuscript and clarified the architectural selection rationale in Section 5.

RC2-C2 — Climate-relevant diagnostics

> *"The reported metrics are a good start. Pointwise MAE/RMSE/correlation plus probabilistic CRPS/coverage/hit rate/FAR are useful, but the paper should add more climate-relevant diagnostics like wet-day only errors, dry/wet frequency bias, precipitation intensity bias, extreme-event metrics such as R95p/R99p/Rx1day, etc. For precipitation especially, MAE can hide failures on extremes."*

AC2-2:

We agree that standard point-wise metrics can mask important aspects of precipitation model performance, particularly regarding the frequency–intensity decomposition and the representation of extremes. We note, however, that the submitted manuscript already includes a substantial stratified evaluation of precipitation in this direction.

Specifically, Figure 6 of the manuscript presents a stratified uncertainty evaluation that separates model performance into four distinct regimes: (a) temperature z-scores, (b) wet-day precipitation z-scores (amount calibration conditional on rain occurrence), (c) dry-day

prediction magnitude (how close predictions are to zero on observed dry days), and (d) precipitation occurrence diagnostics including the hit rate and false alarm rate. This analysis already addresses the reviewer's concern about wet/dry frequency assessment and conditional accuracy on rainy days.

To complement this existing analysis and address the reviewer's request for climate-specific extreme indices, we have computed additional diagnostics at the 26 held-out test stations over 2020–2023 (using a standard WMO wet-day threshold of 1.0 mm):

Diagnostic	RK	SMACNP Global	SMACNP Localized
Wet-day frequency bias (pred/obs)	1.200	1.235	**1.193**
Mean intensity bias (pred/obs, wet days)	0.633	0.746	**0.762**
Wet-day MAE (mm)	3.83	3.46	**3.43**
Wet-day RMSE (mm)	7.03	6.52	**6.36**
R95p bias (%)	−58.2	−53.5	**−50.9**
R99p bias (%)	−66.1	−63.3	**−60.5**
Rx1day bias (%)	−31.0	−33.5	**−27.2**

SMACNP Localized outperforms both RK and SMACNP Global on all 7 diagnostics. All models slightly overpredict wet-day frequency (~20%), a common tendency of spatial interpolation in convective-dominated regions. All models also underestimate wet-day intensity due to spatial smoothing — SMACNP Localized recovers 76% of the observed intensity vs. only 63% for RK. For aggregate extreme indices (R95p, R99p), all models underestimate totals from the most intense days, but SMACNP Localized shows the smallest biases (−51% and −61%, respectively). For Rx1day (mean annual maximum one-day precipitation, computed per station per year following the ETCCDI standard), SMACNP Localized also shows the smallest bias (−27%) compared to RK (−31%) and SMACNP Global (−34%). This is because kriging, while exact at observation points, tends to smooth spatial gradients and underestimate peak intensities at locations between stations, whereas the probabilistic model better captures the station-specific extreme behaviour through its learned attention weights. For all 7 diagnostics — including all aggregate extreme indices (R95p, R99p, Rx1day), frequency, intensity, and conditional accuracy — SMACNP Localized outperforms RK.

Extreme precipitation underestimation is a known limitation of all spatial interpolation approaches over sparse networks — extreme convective events are highly localised and may not be captured by the nearest stations. SMACNP mitigates this better than the alternatives, but eliminating the bias entirely would require denser observations or auxiliary data sources such as radar. We have added these diagnostics as Table 3 in Section 4.1, with an accompanying summary paragraph.

RC2-C3 — Regional scope

"The regional scope is too narrow. Test in an additional region or clearly reframe as a Romania-focused case study, acknowledging limitations of single-country evaluation."

AC2-3:

We agree and have reframed the work as a Romania-focused case study. Romania was chosen deliberately because it concentrates several key challenges: complex Carpathian topography (0–2,544 m), a sparse and irregular station network (156 stations, ~40 km average spacing), mixed precipitation regimes (Mediterranean, continental, oceanic influences), and the absence of an existing high-resolution gridded daily dataset — the type of product that many European countries have produced using geostatistical methods.

The methodology itself is general and transferable to other regions with comparable data availability. We now state this explicitly in the manuscript and acknowledge that demonstrating transferability across different climates and station configurations is a priority for future work.

RC2-C4 — Validation protocol clarity

"The paper says 130 stations are used for model development/validation and 26 stations are held out for final testing, which is good, but it should explicitly state how validation targets are sampled and whether the target station/day is ever included among the context observations during training or validation. This is important because if the model is allowed to see the target station's own observation as part of the context set, even indirectly, the reported skill would be inflated. The paper should also clarify that scalers, hyperparameter tuning, and model selection use only training/validation stations, never the held-out test stations."

AC2-4:

We agree it was not described clearly enough. We have revised Section 2.1 (where the train/test split is introduced) and the Experimental Setup (Section 3.2) to make the following explicit:

The 26 test stations were selected randomly once before any model development and are never used for training, validation, tuning, or model selection. The training protocol involves two nested levels of data partitioning. First, the available days (2020–2023) are split into 80% for training and 20% for validation; the validation days are used exclusively for early stopping and model selection. Second, within each training day, the 130 training-pool stations are randomly partitioned into a context set (~50%, model input) and the loss is evaluated at all 130 stations ("on-the-grid" training). This means that the model must learn both to reconstruct observations at context stations and to interpolate at the remaining non-context stations, which constitutes the genuine test of generalization within each training step. The context partition is re-randomized at every step, so every station regularly appears in both roles. At evaluation time, all 130 training stations serve as

context and predictions are made at the 26 held-out locations. All data normalization is fitted on the training pool only, and all hyperparameter tuning (Optuna) uses only training-pool stations. There is no data leakage at any stage.

RC2-C5 — Figures and presentation

> *"The figures and presentation need work. Some figures are visually useful, but the maps and uncertainty plots should be made cleaner, larger, and easier to compare. Figure 3 is not even centered! The paper should avoid saying that sharper precipitation maps are 'more realistic' unless this is validated against independent observations or robust spatial diagnostics; sharper can also mean artifact."*

AC2-5:

We agree with both points and have made the following changes:

1. **Figure quality.** As noted in our response to Reviewer 1 (Comment 4), all figures were originally generated at 300 DPI, but image compression during DOCX-to-PDF conversion degraded their quality in the review copy. In the revised submission, all figures are submitted as separate high-resolution PNG files (300 DPI) to avoid compression artefacts. Figure 4 (the architecture diagram) has been re-centred and reformatted for clarity.

2. **"Sharper = more realistic" claim.** We have revised the relevant passage in Section 4 to avoid this unsupported assertion. The revised text describes the spatial structure of the predicted fields in objective terms (e.g., "SMACNP fields exhibit finer spatial detail than RK, particularly in mountainous areas where topographic gradients are steepest") without claiming that finer detail is inherently more realistic. We acknowledge that sharper predictions could reflect either genuine skill in resolving local gradients or spurious spatial artefacts, and that distinguishing between these possibilities would require validation against independent high-resolution data sources (e.g., radar estimates or dense temporary station networks).

RC2-C6 — Contribution framing

> *"The contribution should be framed more modestly as an application and evaluation of an existing architecture rather than a new deep learning framework."*

AC2-6:

We accept this suggestion. The SMACNP architecture builds on the well-established Conditional Neural Process framework (Garnelo et al., 2018) and its attentive variants (Kim et al., 2019). Our contribution is not a new architecture per se, but rather the adaptation, configuration, and systematic evaluation of this family of models for the specific task of daily climate gridding from sparse national station networks — a domain where geostatistical methods have been the standard approach for decades.

We have revised the manuscript to frame the contribution accordingly: as an applied evaluation demonstrating that a localized attention-based Neural Process can outperform conventional Regression Kriging for climate gridding over complex terrain, while also providing calibrated uncertainty estimates. The architectural choices (localized vs. global attention, hurdle model for precipitation, static-only predictors) are presented as design decisions motivated by the specific requirements of the climate gridding task, rather than as novel methodological contributions. The Introduction has been partially revised as part of RC2-C3 (case-study reframing); here we add an explicit contribution list and revise the Discussion opening accordingly.

References

Ashman, M., Diaconu, C., Langezaal, E., Weller, A., and Turner, R. E.: Gridded Transformer Neural Processes for Large Unstructured Spatio-Temporal Data, https://doi.org/10.48550/arXiv.2410.06731, 2024.

Garnelo, M., Rosenbaum, D., Maddison, C., Ramalho, T., Saxton, D., Shanahan, M., Teh, Y. W., Rezende, D., and Eslami, S. A.: Conditional neural processes, in: International conference on machine learning, 1704–1713, 2018.

Gordon, J., Bruinsma, W. P., Foong, A. Y. K., Requeima, J., Dubois, Y., and Turner, R. E.: Convolutional Conditional Neural Processes, https://doi.org/10.48550/ARXIV.1910.13556, 2019.

Kim, H., Mnih, A., Schwarz, J., Garnelo, M., Eslami, A., Rosenbaum, D., Vinyals, O., and Teh, Y. W.: Attentive Neural Processes, https://doi.org/10.48550/ARXIV.1901.05761, 2019.

Klein Tank, A. M. G., Zwiers, F. W., and Zhang, X.: Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation, Climate Data and Monitoring WCDMP-No. 72, WMO-TD No. 1500, World Meteorological Organization, Geneva, 56 pp., 2009.

Benjamin Murphy, Roman Yurchak, and Sebastian Müller: GeoStat-Framework/PyKrige: v1.7.3, https://doi.org/10.5281/ZENODO.17372225, 2025.