

Performance and Efficiency of Climate In-Situ Data Reconstruction: Why Optimized IDW Outperforms kriging and Implicit Neural Representation

Jakub Walczak

Abstract—This study evaluates three reconstruction methods for sparse climate data: the simple inverse distance weighting (IDW), the statistically grounded ordinary kriging (OK), and the advanced implicit neural representation model (MMGN architecture). All methods were optimized through hyper-parameter tuning using validation splits. An extensive set of experiments was conducted, followed by a comprehensive statistical analysis.

The results demonstrate the superiority of the simple IDW method over the other reference methods in terms of both reconstruction accuracy and computational efficiency. IDW achieved the lowest RMSE (3.00 ± 1.93), MAE (1.32 ± 0.77), and Δ_{MAX} (24.06 ± 17.15), as well as the highest R^2 (0.68 ± 0.16), across 100 randomly sampled sparse datasets from the ECA&D database. Differences in RMSE, MAE, and R^2 were statistically significant and exhibited moderate to large effect sizes. The Dunn post-hoc test further confirmed the consistent superiority of IDW across all evaluated quality measures.

The evaluation of reconstruction time and peak memory usage for different reconstruction domain sizes showed that IDW required less computational time (even when excluding the training process of MMGN). However, memory usage for both IDW and OK increased more rapidly with the number of target points, which aligns with expectations since IDW and OK store values explicitly, whereas MMGN provides an implicit representation.

The experiments highlight that methods with a small number of interpretable hyper-parameters, such as IDW, are considerably easier to optimize and more robust than complex approaches like the MMGN artificial neural model. We conclude that the use of artificial neural networks should always be critically assessed for each specific application, balancing their potential advantages against the substantial computational cost and complexity of the hyper-parameter optimization process.

Index Terms—spatial reconstruction, climate data approximation, inverse distance weighting, ordinary kriging, implicit neural representation,

I. INTRODUCTION

OBSERVATIONS play a pivotal role in advancing climate modelling, particularly in the context of climate change research [1], extreme event attribution [2], and operational oceanography [3]. Accurate observational data enable model calibration, facilitate bias correction [4], and ensure reliable quality assessment of climate models [5]–[7], forming the backbone of climate science research.

Observational data come mainly from two main sources: remotely sensed data and in-situ measurements. Remotely sensed observations, such as those obtained via satellites [8], aircraft [9], drones [10], radiosondes [11], or ground-based devices such as Doppler radar systems [12], provide extensive spatial coverage and high temporal resolution. On the other hand, in-situ observations, representing localized measurements of physical fields, offer higher accuracy and are invaluable for validation. To overcome the limitations of individual datasets, both types are often integrated [13].

Despite their value, in-situ observations face significant challenges. These include measurement biases, uneven spatial distribution, and persistent gaps, especially in regions with limited infrastructure or harsh conditions [14]. These observational sparsity hinder their direct application in climate models as to achieve spatial and temporal continuity the data needs to be spatially reconstructed [15].

In this study, we address the challenge of sparsity in in-situ observations, focusing specifically on temperature as a critical variable in climate modelling due to its central role in atmospheric processes. We carefully design experiments to reliably assess the performance of three benchmark methods of varying complexity under optimised hyper-parameters. Our findings clearly demonstrate that simple methods can, in some cases, be more accurate and easier to apply in practice, which challenges the current trend of carelessly applying artificial neural networks to a wide range of problems.

Another contribution of this work is the integration of hyper-parameter optimisation, based on *optuna* [16], into the *Climatrix* tool [17].

II. RELATED STUDIES

Artificial intelligence is already widely applied across various domains of climate sciences [18]. In particular, AI has been extensively exploited in processes such as data assimilation [19], bias reduction [20], land cover classification [21], and retrieving geophysical information from remotely sensed observations [22]. Recent advancements have also achieved impressive results in accurate weather forecasting [23] and extreme event prediction [24].

Although data-driven methods have been employed for climate data reconstruction for many years, new approaches relying on implicit neural representations (INRs) are now emerging with hope to address limitations of traditional tech-

Manuscript received April 19, 2021; revised August 16, 2021.

Jakub Walczak is affiliated to Lodz university of Technology, Lodz, Poland

✉ Jakub Walczak: <https://orcid.org/0000-0002-5632-9484>

We define the problem of *reconstruction* in this study as the task of identifying a function $\mathcal{F} : \mathbb{R}^2 \rightarrow \mathbb{R}$ that approximates a given set $\mathcal{X}$ of 3D points $\mathbf{x}_i \in \mathcal{X}$, where $\mathbf{x}_i = \{x_i^{(1)}, x_i^{(2)}, x_i^{(3)}\}$, and satisfies the following condition:

$$\forall \mathbf{x} \in \mathcal{X} \mathcal{F}(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = \mathbf{x}^{(3)} + \epsilon(\mathbf{x}), \quad (1)$$

where $x^{(j)}$ is the j -th coordinate of a point $\mathbf{x} \in \mathcal{X}$, and $\epsilon(\mathbf{x})$ represents an approximation error that is to be minimized. For interpolation, this error is strictly zero: $\epsilon(\mathbf{x}) = 0$ for reconstruction nodes.

A. Traditional Reconstruction Methods

1) *Spatial Interpolation*: The interpolation task involves finding a function that exactly matches the prescribed values of a dataset [25], i.e., $\epsilon(\mathbf{x}) = 0$ for reconstruction nodes. Depending on the data's characteristics, the underlying function may be subject to constraints or prior assumptions.

A variety types of interpolation methods exist, two of which are particularly noteworthy. The first relies on Lagrange polynomials [26], enabling the construction of an $(b - 1)$ -degree polynomial (where $b = |\mathcal{X}|$) that passes through b points. This method can also be extended to multivariate data [27].

The second and more widely adopted approach in climate sciences is built on the intuition that closer reconstruction nodes have a greater influence than distant ones. This category includes methods such as Inverse Distance Weighting (IDW) and kriging [28]. For an arbitrary point $\bar{\mathbf{x}}$ located within the bounds of reconstruction nodes, the reconstruction function in IDW takes the following form:

$$\mathcal{F}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) = \frac{\sum_{i \in \mathcal{X}} \mathbf{x}_i^{(3)} w_i}{\sum w_i}, \quad (2)$$

where w_i is the reciprocal of the powered (usually squared) distance between a reconstructed point and a reconstruction node.

For kriging methods, which often outperform IDW [29], node weights are determined differently. They incorporate the spatial autocorrelation of the data modelled using a semivariogram. Traditional kriging assumes the data is both mean stationary and isotropic. However, advanced kriging variants allow some of these assumptions to be relaxed.

2) *Sub-grid Parametrization*: Many climate phenomena occur at spatial and temporal scales smaller than the resolution of physical models, making them challenging to capture directly. Sub-grid parametrization addresses this gap [30], [31]. This approach is widely utilized in climate science to enhance the fidelity and applicability of physical models.

Sub-grid parametrization involves tuning (and calibrating) physics-based models to better align with specific scientific objectives or observational datasets [32]. The authors of [32] identify two main aspects of this process:

- 1) **Objective tuning** — This involves optimizing parameters through well-defined mathematical approaches, such as solving Bayesian calibration problems [33].

- 2) **Subjective judgement** — This aspect refers to expert decisions made during the parametrization process, such as prioritizing specific processes or choosing parameter values based on domain knowledge, expertise, and scientific priorities.

By fusing these approaches, sub-grid parametrization enables climate models to approximate the set of observations, provided the physical model is supplied.

B. Deep Learning Methods

Deep learning models have emerged as powerful tools for data reconstruction problems in climate science [34], owing to their ability to capture hidden patterns in data through data-driven approach.

The first category of deep-learning-based methods are super-resolution techniques. They involve upscaling or enhancing the resolution of gridded datasets [35], which are typically represented as images or similarly structured data formats. In recent years, machine learning, particularly deep learning, has shown impressive capabilities in super-resolution tasks. These methods have become increasingly popular across diverse fields, with artificial neural networks being deployed to significantly improve resolution enhancement [36], [37]. These group of solutions, however, operate on gridded (image-like) structures, hence their usability for sparse in-situ observations is limited and doubtful.

For super-resolution, convolutional neural networks (CNNs) are frequently employed due to their ability to effectively capture spatial patterns and features in image-like data. In addition to CNNs, generative models such as generative adversarial networks (GANs) are also utilized to further improve super-resolution outcomes [38], [39].

A particularly promising and versatile approach in the domain of deep learning is the use of **implicit neural representations** (INRs). Despite potential confusion surrounding the term *implicit*, INR methods can be expressed both in an implicit form (3) and an explicit form (4):

$$\mathcal{F}_\theta(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}) \approx 0, \quad (3)$$

$$\mathcal{F}_\theta(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) \approx \mathbf{x}^{(3)}, \quad (4)$$

where $\mathcal{F}_\theta$ denotes the implicit neural representation, and $\mathbf{x}^{(i)}$ are components of a sample in a multidimensional space.

In this context, the term *implicit* does not refer to the mathematical form of the function itself, but rather to the way the signal is stored. Unlike methods such as IDW or kriging, which rely on discrete memory structures, INRs encode information implicitly through the parameterisation defined by the neural network's weights [40].

INRs have been successfully applied in various fields, including 3D shape reconstruction [41], where the signed distance function serves as the reconstruction objective.

One of the most recent methods proposed for sparse data reconstruction is the MMGN model introduced by [42]. This approach employs an encoder-decoder architecture enhanced with iterative Gabor-based feature extraction. The authors

report superior performance, in terms of mean squared error, compared to several established reconstruction methods, including the widely recognised SIREN architecture [40]. For this reason, we selected MMGN as the reference deep learning method in our study.

III. METHODS

A. Reference methods

As reference methods, we selected three well-known and widely applied reconstruction approaches of varying complexity:

- 1) IDW — a simple method with only three hyperparameters, easy to tune, belonging to the class of explicit representations (all reconstruction nodes are stored),
- 2) ordinary kriging — the most common variant of kriging [43], conceptually similar to IDW but with weights computed in a more sophisticated manner,
- 3) MMGN — a recent INR-based method reported to achieve superior performance in data reconstruction compared to others in its family.

B. Definitions

Sparse climate observations can be thought of as a point cloud, yet with characteristics that differ slightly from the conventional understanding of 3D point clouds typically utilized in volumetric-data studies [44]. To account for these differences, the definition of a point cloud can be extended from a simple collection of Cartesian coordinates to a representation in any metric space [45], which is not necessarily constrained to three dimensions. Hence, we defined the climate-based point cloud as expressed in (5)-(6).

$$\mathcal{P} = \{p_i \mid i \in \mathbf{I}\}, \quad (5)$$

where $\mathbf{I}$ is a set of indices, $\mathbf{I} = \{1, 2, 3, \dots, N\}$, representing a collection of points whose size is given by $|\mathbf{I}| = |\mathcal{P}| = N$.

Each point in a climate-based point cloud can therefore be defined as a tuple that combines its geographic position with the corresponding physical field values. For our study, we specifically considered a single physical field: **temperature**. This formulation is expressed in (6).

$$p_i = (\lambda, \phi, \tau), \quad (6)$$

where ϕ represents the longitude, λ denotes the latitude, and τ corresponds to the physical field, in this case, the temperature at the geographic coordinates (λ, ϕ) .

Let us confront that concept with the general notation introduced in the section Related Studies — the climate-based point cloud points $p_i \in \mathcal{P}$ are treated as the reconstruction nodes $\mathbf{x}_i \in \mathcal{X}$ ($\mathcal{P} = \mathcal{X}$). A sample climate point cloud is represented in Fig. 1.

C. Tools used

In our study, we used the *Climatrix* [17] accessor to facilitate and accelerate climate data manipulation and to enable systematic comparison of reconstruction methods.

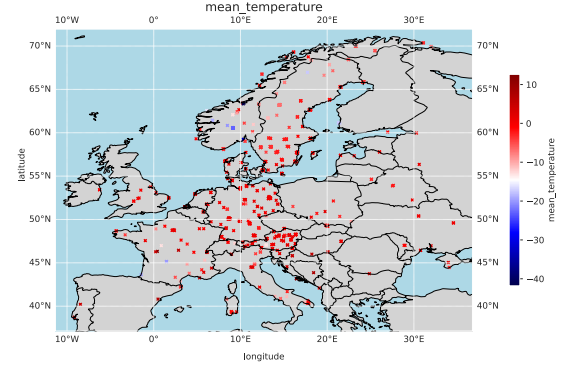


Fig. 1: A sample climate point cloud representing temperature field

D. Quality Metrics

In our study, we adopted a robust, evidence-based methodological framework for evaluation, as proposed in [46].

To evaluate the quality of the reference climate reconstruction methods, we employed three complementary quality measures: root mean square error (*RMSE*), mean absolute error (*MAE*), and the determination coefficient (R^2). These metrics offer distinct insights into the method's performance.

- 1) **Root mean squared error** (7) — this metric measures error in the same units as the observations, providing an interpretable indication of overall predictive accuracy.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\tau_i - \hat{\tau}_i)^2}, \quad (7)$$

where N represents the cardinality of the climate point cloud, τ_i denotes the observed temperature at point p_i , and $\hat{\tau}_i$ is the predicted temperature derived from the approximation function.

- 2) **Mean absolute error** (8) — this metric calculates the average absolute difference between observed and predicted values, making it more robust against outliers than *RMSE*.

$$MAE = \frac{1}{N} \sum_{i=1}^N |\tau_i - \hat{\tau}_i|, \quad (8)$$

- 3) **Determination coefficient** (9) according to [47] is more informative than *RMSE* and *MAE* due to its upper bound set to 1.

$$R^2 = 1 - \frac{\sum (\tau_i - \hat{\tau}_i)^2}{\sum (\tau_i - \bar{\tau})^2}, \quad (9)$$

where $\bar{\tau}$ is the mean value of the observed temperature.

Additionally, we report the maximum absolute difference Δ_{MAX} (10) between observed and predicted temperature values. This metric highlights the worst-case prediction error, providing insight into each method's behaviour under extreme conditions.

$$\Delta_{MAX} = \max_{i \in \mathbf{I}} |\tau_i - \hat{\tau}_i| \quad (10)$$

To evaluate the efficiency of the reconstruction methods, we further measured for simulated target data both the elapsed time $t_r(M)$ and the maximum memory occupancy $m_r(M)$ required to reconstruct $M \in \{10^2, 5 \cdot 10^2, 5 \cdot 10^3, 5 \cdot 10^4, 5 \cdot 10^5, 5 \cdot 10^6\}$ points.

E. Dataset

At the beginning, it is important to emphasize a fundamental limitation: **a ground-truth dataset of dense, in-situ observations does not exist**. Therefore, to evaluate the effectiveness of the proposed method, we rely on blended in-situ observations provided by the European Climate Assessment & Dataset (ECA&D) project [48], which serves as the basis for the gridded E-OBS dataset [49].

The ECA&D project, a continuation of the original ECA initiative coordinated by the Royal Netherlands Meteorological Institute (KNMI), focuses on analysis of **daily** precipitation and temperature (particularly trends in extreme values [50]) records from meteorological stations across Region VI (Europe), as defined by the World Meteorological Organization (WMO) [51]. The dataset is provided in two formats: non-blended, which contains raw observations from partner institutions, and blended, where missing values are filled using observations from nearby meteorological stations (within 12.5 km and with a relative elevation difference no greater than 25 m) [50].

Due to its collaborative nature, the ECA&D dataset exhibits spatial and temporal heterogeneity in coverage.

In this study, we use the blended version, which underpins the widely adopted E-OBS gridded climate product.

To ensure seasonal and spatial representativeness, and to enable robust statistical analysis, we randomly selected 100 ($|\mathcal{S}| = 100$) single dates across the entire ECA&D archive to form a dataset $\mathcal{S}$. We considered only those dates where number of valid observations with the quality flag set to `valid` exceeds 500. The selection was performed using the NumPy [52] pseudo-random number generator with a fixed seed of 0 to ensure reproducibility.

We chose this relatively large number of dates to capture variability in spatial station density and to mitigate potential biases introduced by anomalous weather events or extreme conditions.

For each selected date, all available in-situ observations were randomly divided into three mutually exclusive subsets: training, validation, and testing.

By focusing on single-day snapshots, we eliminate temporal dependencies, thereby reducing the risk of data leakage — particularly the ‘neighbourhood bias’ described in [53]. Each single-date sample serves as an independent input to the instantaneous reconstruction task.

Across all dates, 60% of the available observations were assigned to the training set, with the remaining 40% equally and randomly split between the validation and test sets, all derived from the same day’s data. Cardinalities of splits may vary across dates due to time-dependent differences in spatial coverage in ECA&D dataset. Summary statistics for the training and validation splits are shown in Table I. To

TABLE I: Summary statistics for training and validation splits across 100 randomly selected dates.

Split	Min [$^{\circ}C$]	Mean [$^{\circ}C$]	St. dev. [$^{\circ}C$]	Max [$^{\circ}C$]
Train	−57.00	6.65	10.41	37.60
Validation	−53.40	6.68	10.44	37.60

prevent data leakage and unintentional methodological bias, we did not compute statistics for the test split.

F. Experiment Procedure

1) *Reproducibility*: To ensure full reproducibility of the experiments described in this paper, we manually set seeds for pseudo-random number generators when processing each individual date.

Information regarding library versions and specific implementation details is available in the `experiments` directory of the *Climatrix* repository we contributed to¹.

2) *Dataset preprocessing*: Based on exploratory data analysis, we found no need for additional data cleaning or imputation.

However, we applied method-specific preprocessing. For ordinary kriging, we normalised latitude and longitude to the range $[-1, 1]$ and standardised the physical field values by subtracting the mean and dividing by the standard deviation.

IDW does not require any preprocessing.

For the MMGN method, we applied min–max scaling to the range $[-1, 1]$, following the original setup [42] so that the numerical problem with float overflowing can be avoided.

3) *Experimental trial configuration and split setup*: Our investigation focused on analysing the quality and efficiency of the three optimised reference reconstruction methods.

We sampled 100 time instants from the ECA&D dataset and trained the selected methods using predefined training splits. Validation splits were used to compute the optimisation criterion for hyper-parameter tuning (see Section III-G). Test splits were held out and used exclusively to report performance for the tuned methods at each sampled time instant.

After applying the optimised hyper-parameters, we reconstructed values for all spatial points in the test split. Each reconstruction method produced a set of reconstructed values $\hat{\mathcal{T}}$ over the same spatial domain as the test split. The elements of this set are tuples $r_i = (\lambda_i, \phi_i, \hat{\tau}_i)$, where λ_i, ϕ_i are the geographical coordinates of the i -th test point.

To investigate reconstruction times for the three reference methods, we used the optimised hyper-parameters obtained for each sample throughout the experiment. The training split of each sample served as the set of reconstruction nodes, while synthetically generated datasets were used as targets to be reconstructed. These target datasets domains were sampled from a uniform distribution (taking as limits the minimum and maximum values of latitude and longitude or a training dataset) with increasing cardinalities of $\{10, 100, 500, 1000, 2000, 5000, 10000\}$ points. For MMGN, we neglected time necessary for training a model (we assumed the weights are preloaded in advance).

¹*Climatrix* project repository: <https://github.com/jamesWalczak/climatrix>.

G. Hyper-parameter optimisation

To facilitate hyper-parameter optimisation, we extended the *Climatrix* tool with the *HParamFinder* component, which relies on the *optuna* library.

For optimisation, we adopted Gaussian process priors for Bayesian optimisation [54]. Although Bayesian optimisation does not guarantee convergence to the global optimum due to its probabilistic nature, brute-force optimisation is often impractical for multidimensional space of numerical hyper-parameters.

Table II presents the optimisation setup for all three reference methods.

TABLE II: Bayesian optimisation setup for reference methods

Method	# initial samples	# iterations
IDW	50	100
OK	50	100
MMGN	50	200

Because MMGN has many more hyper-parameters to optimise, we allocated twice as many trials as for IDW or OK.

For each method, we empirically set constraints on the hyper-parameters to ensure that most optimised values fall within, but not at the boundaries of, the imposed ranges. To verify this, we tracked histograms of continuous hyper-parameters and frequencies of categorical choices.

As the optimisation criterion, we minimised the Mean Absolute Error (MAE) (8) on the validation split of each sample.

1) *Bounds for IDW*: IDW requires only two hyper-parameters: the number of nearest neighbours and the power coefficient, which controls how rapidly influence decreases with distance. Their bounds are given in Table III.

TABLE III: Optimisation bounds for IDW

Parameter	Type	Values/Range
# nearest neighbours	integer	[1, 50]
Distance power	real number	$[10^{-7}, 5.0]$

2) *Bounds for OK*: Ordinary kriging involves several hyper-parameters with bounds listed in Table IV.

TABLE IV: Optimisation bounds for ordinary kriging

Parameter	Type	Values/Range
# averaging bins for semivariogram	integer	[2, 50]
Anisotropy scaling factor	real number	$[10^{-5}, 5]$
Coordinate type	categorical	{euclidean, geographic}
Semivariogram model	categorical	{linear, power, gaussian, spherical, exponential, hole-effect}

3) *Bounds for MMGN*: MMGN is a neural network model, and therefore involves both generic and architecture-specific hyper-parameters. Their bounds are shown in Table V. We fixed the number of epochs at 500, during which the optimisation criterion was minimised.

TABLE V: Optimisation bounds for MMGN

Parameter	Type	Values/Range
learning rate	real number	$[10^{-5}, 10^{-2}]$
L2 regularisation	real number	[0, 0.1]
mini-batch size	integer	[32, 1024]
# hidden dimensions	categorical	{32, 64, 128, 256, 512, 1024}
# latent dimensions	categorical	{32, 64, 128, 256, 512, 1024}
# layers	integer	[1, 10]
input scale	real number	[2, 1024]
alpha	real number	[0.0, 100.0]

Here, the learning rate controls the step size during gradient descent, while L2 regularisation mitigates overfitting. Hidden and latent dimensions govern the internal feature representations. The input scale is a parameter of the Gabor filter, and α is the concentration parameter of the gamma distribution used in the filter. For details, see the code accompanying MMGN [42].

4) *Statistical considerations*: To evaluate the outcomes of our experiments, we performed a series of statistical analyses. In particular, we applied the non-parametric Kruskal–Wallis test [55], [56] to assess the impact of reconstruction method on each reported quality measure. When the Kruskal–Wallis test indicated significance, we performed post hoc analyses using Dunn’s test [57] with Holm–Bonferroni correction for multiple comparisons [58].

We also reported effect sizes. For the Kruskal–Wallis test, we computed η^2 following [59]:

$$\eta^2 = \frac{H - k + 1}{k \cdot |\mathcal{S}| - k}, \quad (11)$$

where H is the Kruskal–Wallis statistic, $|\mathcal{S}| = 100$, and k is the number of groups — in our case set to three (as we have three methods: IDW, OK, and MMGN).

For pairwise Dunn’s tests, we calculated the biserial rank correlation coefficient r [59]:

$$r = \frac{2 \cdot (\bar{R}_1 - \bar{R}_2)}{n_1 + n_2}, \quad (12)$$

where $\bar{R}_1$ and $\bar{R}_2$ are the mean ranks of groups 1 and 2, and n_1 and n_2 are their respective sizes. In our case, $n_1 = n_2 = |\mathcal{S}| = 100$, as we sampled 100 instances from the ECA&D dataset.

For both Kruskal–Wallis and Dunn’s tests, we set the significance threshold at $\alpha = 0.05$.

IV. RESULTS

Table VI presents the aggregated quality measures for the three examined reference reconstruction methods across all test splits of $\mathcal{S} = 100$ samples. Reported values correspond to the median followed by the interquartile range (as distributions are strongly abnormal).

Following the established methodology, we performed an omnibus Kruskal–Wallis test for each metric separately. The results are reported in Table VII.

TABLE VI: Quality measure values for the three evaluated methods across all test splits of $|\mathcal{S}| = 100$ samples. Each entry shows the median followed by the interquartile range. Green cells indicate the best value of a given metric, while red cells indicate the worst.

Metric	IDW	OK	MMGN
RMSE	3.00 $\pm$ 1.93	5.40 $\pm$ 3.30	4.11 $\pm$ 2.46
MAE	1.31 $\pm$ 0.77	3.81 $\pm$ 2.33	1.93 $\pm$ 1.02
R^2	0.67 $\pm$ 0.16	-0.76 $\pm$ 1.92	0.46 $\pm$ 0.21
Δ_{MAX}	24.06 $\pm$ 17.15	24.85 $\pm$ 17.40	35.83 $\pm$ 21.46

TABLE VII: Results of the omnibus Kruskal–Wallis test for all quality measures.

Metric	H statistic	p -value	Effect size
RMSE	70.67	$4.50 \cdot 10^{-16}$	0.23
MAE	169.66	$1.44 \cdot 10^{-37}$	0.56
R^2	288.95	$4.3 \cdot 10^{-46}$	0.70
Δ_{MAX}	28.24	$7.39 \cdot 10^{-7}$	0.01

For all quality measures, the p -values are far below the presumed significance threshold (0.05), indicating strong evidence of statistically significant differences between the methods. The effect sizes are substantial for all metrics except Δ_{MAX} , where the value is negligible.

The complete results of the Dunn post-hoc tests are provided in the associated experiment notebook². In brief, p -values for RMSE, MAE, and R^2 are well below 0.05, confirming statistical significance. For Δ_{MAX} , significant differences were also observed, except for the IDW vs. OK comparison, where $p = 0.32$. Considering the median values of all reported quality measures across the three methods (IDW, OK, MMGN), the results clearly demonstrate the superiority of IDW over both OK and MMGN in terms of RMSE, MAE, and R^2 when reconstruction proceeds the hyper-parameters optimisation.

In Table VIII, the r biserial rank correlation coefficients are presented for the Dunn post-hoc tests.

TABLE VIII: Effect size for Dunn post-hoc tests for all reported measures

Metric	Pair	r
RMSE	IDW vs OK	-0.65
	IDW vs MMGN	-0.41
	OK vs MMGN	0.35
MAE	IDW vs OK	-0.95
	IDW vs MMGN	-0.48
	OK vs MMGN	0.78
R^2	IDW vs OK	0.99
	IDW vs MMGN	0.64
	OK vs MMGN	-0.91
Δ_{max}	IDW vs OK	-0.08
	IDW vs MMGN	-0.41
	OK vs MMGN	-0.33

As shown in Table VIII, the effect sizes for RMSE, MAE, and R^2 are large across all pairwise comparisons, favouring IDW over both MMGN and OK (negative values for error measures and positive for R^2). For Δ_{max} , the reported effect

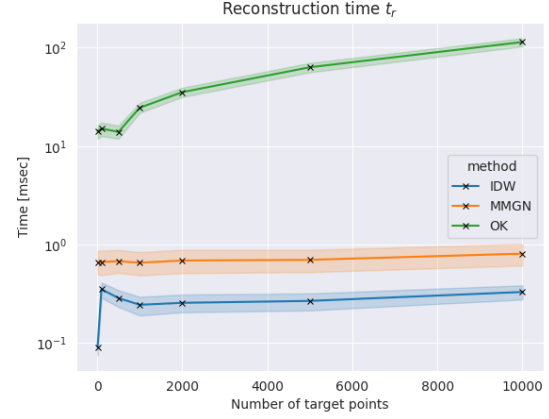


Fig. 2: Reconstruction time for optimized hyper-parameters with an increasing number of points to be reconstructed.

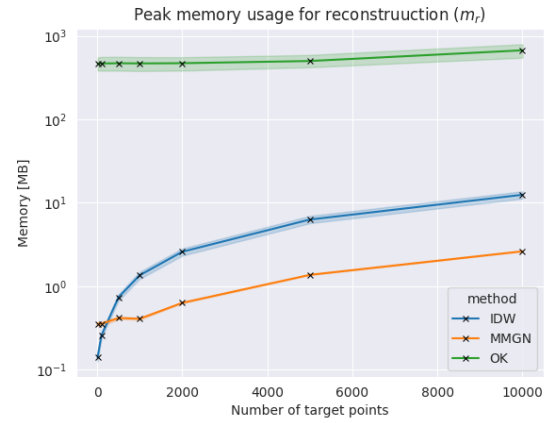


Fig. 3: Peak memory usage during reconstruction for optimized hyper-parameters with an increasing number of points to be reconstructed.

sizes are moderate (for IDW vs. MMGN and OK vs. MMGN) or small (for IDW vs. OK).

We now proceed to compare the computational efficiency of the methods in terms of reconstruction time, t_r (Fig. 2), and peak memory usage, m_r (Fig. 3). Due to the large differences in the required time and memory consumption for ordinary kriging, a logarithmic scale was applied to the y-axis. For all three reference methods, we depicted the 95th confidence intervals for both time and memory.

Although the IDW method belongs to the class of explicit representations and therefore retains all training points in memory, there is no noticeable difference in the order of reconstruction time for up to 10,000 points. Interestingly, the IDW method requires less time than MMGN across all examined target domain sizes and significantly less than OK, whose time and memory demands are far beyond comparison. The short reconstruction time of the IDW method is particularly evident for sparse yet moderately sized training sets, where efficient access is provided by the kd-tree structure used for nearest-neighbour retrieval. For noticeably larger training sets,

²The notebook with a detailed analysis is available at https://github.com/jamesWalczak/climatrix/blob/main/experiments/jwalczak/01_Apr_02_compare_recon_method/notebooks/analysis.ipynb

however, the difference between IDW and MMGN is expected to shift in favour of the implicit neural representation.

For MMGN, which is an implicit neural representation method, the poorer-than-expected performance results from the overhead associated with artificial neural network processing. The most critical factors include memory transfers between CPU and GPU, data batching, and data denormalisation required to match the expected input range. It should also be emphasized that for MMGN the reported time corresponds solely to the reconstruction phase; the model training time is not included and may further impact the overall computational efficiency of the pipeline.

V. DISCUSSION AND CONCLUSIONS

In this study, we addressed the problem of reconstructing continuous physical fields from sparse *in-situ* observations using three reference methods: inverse distance weighting (IDW), ordinary kriging (OK), and an implicit neural representation model, MMGN. Based on careful and exhaustive experiments, we conclude that, after hyper-parameter optimization, the simplest method (IDW) yields the best results among all three reference methods across all evaluated quality measures. This superiority is statistically significant and exhibits a large effect size for RMSE, MAE, and R^2 .

We also found that IDW, though reliant on distance computations and an explicit in-memory representation of the training set, requires considerably less time to perform reconstruction. This advantage comes at the cost of a somewhat higher, though roughly constant-factor, peak memory consumption.

Our findings challenge the widespread belief in the universal applicability of artificial neural networks for all types of problems. We wish to emphasize that we do not deny the high capabilities of ANNs; however, they are tools whose selection should be guided by the nature of the problem at hand. In some cases, their complexity and flexibility may be unnecessary or even problematic. We do not claim that ANNs (and in this particular case, the MMGN method) are incapable of efficiently solving the stated problem. In fact, ANNs possess remarkable potential, which is often difficult to fully exploit due to the high dimensionality of their hyper-parameter space and the complexity of optimization.

While IDW relies on only two easily interpretable parameters, MMGN involves a significantly larger number of hyper-parameters, some of which could not be thoroughly explored due to computational constraints and time limitations. The key conclusion from our experiments is that simpler methods can often outperform more advanced ones, not necessarily because of their limited capabilities, but because the optimization of complex models is constrained by the vastness of their hyper-parameter space.

Furthermore, it is worth noting the computational efficiency of the compared methods. In our case, where the task involves reconstructing sparse spatial points rather than processing massive datasets, the in-memory (explicit) representation inherent to IDW requires considerably fewer computational resources than the implicit neural representation used in MMGN.

Finally, it should be noted that hyper-parameters were optimized using validation splits, which may not always ensure

smooth reconstruction across the entire European domain. That is the problem that should be challenged by future research.

A. Future Work

The present study primarily focuses on fields distributed across a geographic coordinate system. An important direction for future research would be to incorporate topographic factors such as elevation, aspect, and slope, as numerous studies have demonstrated their significant influence on spatial and climatic field patterns [60]–[62]. Including these topographic variables could enhance both the accuracy and the physical relevance of the reconstructed fields.

IMPACT STATEMENT

By addressing the critical challenge of data sparsity, our findings demonstrate the effectiveness of the simple inverse distance weighting (IDW) method in reconstructing observational datasets. We believe this study represents a meaningful step toward more efficient and reliable modelling of climate and meteorological systems, paving the way for further advancements in climate science and its practical applications.

Moreover, we show that simple and easily interpretable models can often outperform more advanced approaches due to their computational efficiency and the smaller, more manageable hyper-parameter search space.

ACKNOWLEDGEMENT

We are grateful to Dr. Stefania A. Ciliberti[©] and Dr. Marco Mancini[©] for their invaluable advice on methodological issues and remarks regarding the first draft of the paper.

The calculations mentioned in this paper were performed using the BlueOcean computational cluster which is part of TUL Computing & Information Services Center infrastructure.

APPENDIX

SUPPLEMENTARY MATERIALS

All scripts, environment setup, reproducibility containers definitions, and results were published under MIT license and are available as contribution to *Climatrix* project³.

REFERENCES

- [1] H.-D. Guo, L. Zhang, and L.-W. Zhu, “Earth observation big data for climate change research,” *Advances in Climate Change Research*, vol. 6, no. 2, pp. 108–117, 2015.
- [2] N. Christidis, P. A. Stott, A. A. Scaife, A. Arribas, G. S. Jones, D. Copey, J. R. Knight, and W. J. Tennant, “A new hadgem3-a-based system for attribution of weather and climate-related extreme events,” *Journal of Climate*, vol. 26, no. 9, pp. 2756–2783, 2013.
- [3] F. Davidson, A. Alvera-Azcarate, A. Barth, G. B. Brassington, E. P. Chassignet, E. Clementi, P. De Mey-Frémaux, P. Divakaran, C. Harris, F. Hernandez, *et al.*, “Synergies in operational oceanography: the intrinsic need for sustained ocean observations,” *Frontiers in Marine Science*, vol. 6, p. 450, 2019.
- [4] J. R. Eyre, “Observation bias correction schemes in data assimilation systems: A theoretical study of some of their properties,” *Quarterly Journal of the Royal Meteorological Society*, vol. 142, no. 699, pp. 2284–2291, 2016.

³The project remote repository is available under the following address: <https://github.com/jamesWalczak/climatrix>.

- [5] J. Gómez-Navarro, J. Montávez, S. Jerez, P. Jiménez-Guerrero, and E. Zorita, "What is the role of the observational dataset in the evaluation and scoring of climate models?," *Geophysical Research Letters*, vol. 39, no. 24, 2012.
- [6] F. Tapiador, A. Navarro, V. Levizzani, E. García-Ortega, G. Huffman, C. Kidd, P. Kucera, C. Kummerow, H. Masunaga, W. Petersen, *et al.*, "Global precipitation measurements for validating climate models," *Atmospheric Research*, vol. 197, pp. 1–20, 2017.
- [7] V. Eyring, P. M. Cox, G. M. Flato, P. J. Gleckler, G. Abramowitz, P. Caldwell, W. D. Collins, B. K. Gier, A. D. Hall, F. M. Hoffman, *et al.*, "Taking climate model evaluation to the next level," *Nature Climate Change*, vol. 9, no. 2, pp. 102–110, 2019.
- [8] D. J. Crichton, C. A. Matmann, L. Cinquini, A. Braverman, D. Waliser, M. Gunson, A. F. Hart, C. E. Goodale, P. Lean, and J. Kim, "Sharing satellite observations with the climate-modeling community: software and architecture," *IEEE software*, vol. 29, no. 5, pp. 73–81, 2012.
- [9] X. Zhang, J. Zhou, S. Liang, L. Chai, D. Wang, and J. Liu, "Estimation of 1-km all-weather remotely sensed land surface temperature based on reconstructed spatial-seamless satellite passive microwave brightness temperature and thermal infrared data," *ISPRS journal of photogrammetry and Remote Sensing*, vol. 167, pp. 321–344, 2020.
- [10] K. B. Bärfuss, H. Schmithüsen, and A. Lampert, "Drone-based meteorological observations up to the tropopause—a concept study," *Atmospheric Measurement Techniques*, vol. 16, no. 15, pp. 3739–3765, 2023.
- [11] B. Ingleby, P. Pauley, A. Kats, J. Ator, D. Keyser, A. Doerenbecher, E. Fucile, J. Hasegawa, E. Toyoda, T. Kleinert, *et al.*, "Progress toward high-resolution, real-time radiosonde reports," *Bulletin of the American Meteorological Society*, vol. 97, no. 11, pp. 2149–2161, 2016.
- [12] M. Lindsog, K. Salonen, H. Järvinen, and D. Michelson, "Doppler radar wind data assimilation with hirlam 3dvar," *Monthly weather review*, vol. 132, no. 5, pp. 1081–1092, 2004.
- [13] B. Arabi, M. S. Salama, J. Pitarch, and W. Verhoef, "Integration of in-situ and multi-sensor satellite observations for long-term water quality monitoring in coastal areas," *Remote Sensing of Environment*, vol. 239, p. 111632, 2020.
- [14] V. Bessenbacher, L. Gudmundsson, and S. I. Seneviratne, "Optimizing soil moisture station networks for future climates," *Geophysical Research Letters*, vol. 50, no. 4, p. e2022GL101667, 2023.
- [15] V. Bessenbacher, D. Schumacher, M. Hirschi, S. I. Seneviratne, and L. Gudmundsson, "Gap-filled multivariate observations of global land-climate interactions," *Journal of Geophysical Research: Atmospheres*, vol. 128, no. 24, p. e2023JD039099, 2023.
- [16] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2623–2631, 2019.
- [17] J. Walczak and W. Żyndul, "climatrix: Climate data reconstruction made simple." <https://github.com/jamesWalczak/climatrix>, 2025.
- [18] S. Dewitte, J. P. Cornelis, R. Müller, and A. Munteanu, "Artificial intelligence revolutionises weather forecast, climate monitoring and decadal prediction," *Remote Sensing*, vol. 13, no. 16, p. 3209, 2021.
- [19] S. Boukabara, E. Maddy, N. Shahroudi, R. N. Hoffman, T. Connor, S. Upton, J. E. Ten Hoeve, V. Krasnopolsky, and K. Garrett, "Artificial intelligence (ai) techniques to enhance satellite data use for nowcasting and nwp/data assimilation," in *100th American Meteorological Society Annual Meeting*, AMS, 2020.
- [20] T. Schneider, S. Lan, A. Stuart, and J. Teixeira, "Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations," *Geophysical Research Letters*, vol. 44, no. 24, pp. 12–396, 2017.
- [21] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2217–2226, 2019.
- [22] V. M. Krasnopolsky and H. Schiller, "Some neural network applications in environmental sciences. part i: forward and inverse problems in geophysical remote measurements," *Neural Networks*, vol. 16, no. 3–4, pp. 321–334, 2003.
- [23] R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, *et al.*, "Learning skillful medium-range global weather forecasting," *Science*, vol. 382, no. 6677, pp. 1416–1421, 2023.
- [24] G. Nearing, D. Cohen, V. Dube, M. Gauch, O. Gilon, S. Harrigan, A. Hassidim, D. Klotz, F. Kratzert, A. Metzger, *et al.*, "Global prediction of extreme floods in ungauged watersheds," *Nature*, vol. 627, no. 8004, pp. 559–563, 2024.
- [25] R. Franke, "Scattered data interpolation: tests of some methods," *Mathematics of computation*, vol. 38, no. 157, pp. 181–200, 1982.
- [26] P. O. J. Scherer, *Interpolation*, pp. 15–35. Heidelberg: Springer International Publishing, 2013.
- [27] T. Sauer and Y. Xu, "On multivariate lagrange interpolation," *Mathematics of computation*, vol. 64, no. 211, pp. 1147–1170, 1995.
- [28] E. Ozelkan, S. Bagis, E. C. Ozelkan, B. B. Ustundag, M. Yucel, and C. Ormeci, "Spatial interpolation of climatic variables using land surface temperature and modified inverse distance weighting," *International Journal of Remote Sensing*, vol. 36, no. 4, pp. 1000–1025, 2015.
- [29] D. Zimmerman, C. Pavlik, A. Ruggles, and M. P. Armstrong, "An experimental comparison of ordinary and universal kriging and inverse distance weighting," *Mathematical Geology*, vol. 31, pp. 375–390, 1999.
- [30] M. D. Fielding, S. A. Schäfer, R. J. Hogan, and R. M. Forbes, "Parametrizing cloud geometry and its application in a subgrid cloud-edge erosion scheme," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 729, pp. 1651–1667, 2020.
- [31] J. Yuval and P. A. O’Gorman, "Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions," *Nature communications*, vol. 11, no. 1, p. 3295, 2020.
- [32] F. Hourdin, T. Mauritsen, A. Gettelman, J.-C. Golaz, V. Balaji, Q. Duan, D. Folini, D. Ji, D. Klocke, Y. Qian, *et al.*, "The art and science of climate model tuning," *Bulletin of the American Meteorological Society*, vol. 98, no. 3, pp. 589–602, 2017.
- [33] M. C. Kennedy and A. O’Hagan, "Bayesian calibration of computer models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, 2001.
- [34] Y. Wang, H. A. Karimi, and X. Jia, "Reconstruction of continuous high-resolution sea surface temperature data using time-aware implicit neural representation," *Remote Sensing*, vol. 15, no. 24, p. 5646, 2023.
- [35] N. Nguyen, P. Milanfar, and G. Golub, "A computationally efficient superresolution image reconstruction algorithm," *IEEE transactions on image processing*, vol. 10, no. 4, pp. 573–583, 2001.
- [36] J. Ma, X. Wang, and J. Jiang, "Image superresolution via dense discriminative network," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 7, pp. 5687–5695, 2019.
- [37] Q. Wang, Q. Li, and X. Li, "Hyperspectral image superresolution using spectrum and feature context," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 11, pp. 11276–11285, 2020.
- [38] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE signal processing magazine*, vol. 35, no. 1, pp. 53–65, 2018.
- [39] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced gan for remote sensing image superresolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5799–5812, 2019.
- [40] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," *Advances in neural information processing systems*, vol. 33, pp. 7462–7473, 2020.
- [41] J. Ye, Y. Chen, N. Wang, and X. Wang, "Gifs: Neural implicit function for general shape representation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12829–12839, 2022.
- [42] X. Luo, W. Xu, Y. Ren, S. Yoo, and B. Nadiga, "Continuous field reconstruction from sparse observations with implicit neural networks," *arXiv preprint arXiv:2401.11611*, 2024.
- [43] H. Wackernagel, "Ordinary kriging," in *Multivariate geostatistics: an introduction with applications*, pp. 79–88, Springer, 2003.
- [44] B. Fei, W. Yang, W.-M. Chen, Z. Li, Y. Li, T. Ma, X. Hu, and L. Ma, "Comprehensive review of deep learning-based 3d point cloud completion processing and analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 22862–22883, 2022.
- [45] S. A. Bello, S. Yu, C. Wang, J. M. Adam, and J. Li, "Deep learning on 3d point clouds," *Remote Sensing*, vol. 12, no. 11, p. 1729, 2020.
- [46] J. Walczak, "Evidence-based methodological framework for machine learning studies," *Authorea Preprints*, 2025.
- [47] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation," *PeerJ computer science*, vol. 7, p. e623, 2021.
- [48] A. M. Klein Tank, J. Wijngaard, G. Können, R. Böhm, G. Demarée, A. Gocheva, M. Mileta, S. Pashiardis, L. Hejkrlik, C. Kern-Hansen, *et al.*, "Daily dataset of 20th-century surface air temperature and precipitation series for the european climate assessment," *International Journal of Climatology: A Journal of the Royal Meteorological Society*, vol. 22, no. 12, pp. 1441–1453, 2002.

- [49] R. C. Cornes, G. Van Der Schrier, E. J. Van Den Besselaar, and P. D. Jones, "An ensemble version of the e-obs temperature and precipitation data sets," *Journal of Geophysical Research: Atmospheres*, vol. 123, no. 17, pp. 9391–9409, 2018.
- [50] ECA&D, "European climate assessment & dataset (eca&d)'algorithm theoretical basis document (atbd), version 10.7.," 2013.
- [51] World Meteorological Organization, "Wmo regions." <https://wmo.int/about-wmo/regions>. Accessed June 3, 2025.
- [52] C. R. Harris, K. J. Millman, S. J. Van Der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, *et al.*, "Array programming with numpy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.
- [53] S. Kapoor, E. M. Cantrell, K. Peng, T. H. Pham, C. A. Bail, O. E. Gundersen, J. M. Hofman, J. Hullman, M. A. Lones, M. M. Malik, *et al.*, "Reforms: Consensus-based recommendations for machine-learning-based science," *Science Advances*, vol. 10, no. 18, p. eadk3452, 2024.
- [54] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *Advances in neural information processing systems*, vol. 25, 2012.
- [55] W. H. Kruskal, "A nonparametric test for the several sample problem," *The Annals of Mathematical Statistics*, pp. 525–540, 1952.
- [56] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.
- [57] O. J. Dunn, "Multiple comparisons using rank sums," *Technometrics*, vol. 6, no. 3, pp. 241–252, 1964.
- [58] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian journal of statistics*, pp. 65–70, 1979.
- [59] M. Tomczak and E. Tomczak, "The need to report effect size estimates revisited. an overview of some recommended measures of effect size," 2014.
- [60] H. F. Diaz, M. Grosjean, and L. Graumlich, "Climate variability and change in high elevation regions: past, present and future," *Climatic change*, vol. 59, no. 1, pp. 1–4, 2003.
- [61] J. He, W. Zhao, A. Li, F. Wen, and D. Yu, "The impact of the terrain effect on land surface temperature variation based on landsat-8 observations in mountainous areas," *International Journal of Remote Sensing*, vol. 40, no. 5-6, pp. 1808–1827, 2019.
- [62] N. Pepin, E. Arnone, A. Gobiet, K. Haslinger, S. Kotlarski, C. Notarnicola, E. Palazzi, P. Seibert, S. Serafin, W. Schöner, *et al.*, "Climate changes and their elevational patterns in the mountains of the world," *Reviews of Geophysics*, vol. 60, no. 1, p. e2020RG000730, 2022.



Jakub Walczak defended his PhD in 2022, he is an AI researcher with both academic and industrial experience across a wide range of domains, including climate and ICT industry. He specializes in the ethical and multidisciplinary applications of artificial intelligence, bio-inspired advancements, and the development of novel AI paradigms.