

Article

Reconstruction of Frontal Gradients Using Radial Basis Function Interpolation

Miodrag Rancic 

Lynker at NOAA/NWS/OMD, 5830 University Research Ct, College Park, MD 20740, USA;
miodrag.rancic@noaa.gov

Abstract

The accurate representation of frontal zones—characterized by sharp scalar gradients—remains a critical challenge in regional objective analysis and data assimilation, particularly when utilizing sparse or stochastically distributed observations. This study evaluates the efficacy of Multiquadric Radial Basis Functions (RBFs) as a high-order alternative to standard spatial mapping operators frequently used in machine learning atmospheric emulators. We contrast the performance of the regularized, C^∞ -continuous RBF approach against nearest neighbor and linear mesh interpolation schemes using both synthetic baroclinic wave profiles and an operational case study of the intense extratropical cyclone that impacted the East Coast of North America in mid-March 1993. To mitigate characteristic boundary artifacts and geometric clipping in bounded regional domains, we implement a targeted numerical stabilization framework combining localized boundary mirroring with four-corner domain anchoring. Our quantitative results demonstrate that the optimized RBF framework substantially improves gradient fidelity and reduces Root Mean Square Error across a wide range of observation densities. Furthermore, we evaluate the computational scalability of RBFs on high-performance computing architectures, demonstrating how Algebraic Multigrid solvers and Graphics Processing Unit acceleration mitigate the foundational $O(N^3)$ computational bottleneck. We conclude that RBF interpolation provides a physically consistent, analytically differentiable manifold that addresses the derivative discontinuities of traditional linear methods, offering a stable pre-processing framework for high-resolution meteorological analysis and machine learning optimization.

Keywords: radial basis functions (RBFs); frontal gradients; machine learning data assimilation; limited area modeling; multiquadric interpolation



Academic Editor: Stephan Havemann

Received: 10 June 2026

Revised: 7 July 2026

Accepted: 22 July 2026

Published: 24 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Atmospheric frontal zones represent primary regions of energy transformation and weather development in mid-latitudes. Defined by sharp gradients in temperature, moisture, and wind fields, these baroclinic zones drive cyclogenesis and severe weather events. Accurate capture of the morphology and magnitude of these gradients—a process fundamentally related to the calculation of the frontogenetic function [1]—is essential for both synoptic analysis and numerical weather prediction.

Despite the proliferation of remote sensing technologies, in-situ observation networks remain stochastically distributed and often sparse. Traditional objective analysis schemes, such as nearest neighbor or piecewise linear mesh interpolation—the latter frequently serving as the linear forward observation operator in operational variational frameworks—can degrade the structural representation of the front. These methods frequently introduce

unphysical artifacts, such as “staircase” discretizations or excessive numerical diffusion, which obscure the true derivative of the scalar field. Although spline-based methods offer higher-order continuity, they can be prone to Runge’s phenomenon or require complex mesh mapping logic [2].

In data assimilation, a diverse array of methodologies have been developed to map unstructured observations onto structured grids—some rely on a prior background field and others operating independently. Among the earliest non-background techniques is the ‘Kriging’ method, originally formulated for geostatistical mineral prospecting [3,4].

The problem of spatial mapping in meteorology has its foundational roots in the pioneering work of Gandin [5], who introduced Optimum Interpolation (OI) to constrain objective analysis using statistical covariance structures. In contrast to purely geometric meshless mapping, OI methods integrate a forecast background as an essential prior input, enabling the objective modeling of multivariate physical balances across the grid. While this statistical framework remains foundational to classical data assimilation, the requirement of an active background forecast field introduces significant computational friction and operational latency when preprocessing raw data for real-time machine learning emulators. Consequently, our regularized RBF approach acts as an independent, standalone observation operator—bypassing the need for prior forecast fields while explicitly preserving the sharp, local derivative structures required for downstream neural network backpropagation.

The recent development of machine learning architectures in numerical weather prediction has introduced a critical need for high-throughput observation operators that can rapidly map input data to grid points. Motivated by this requirement, this paper addresses specific geometric and structural limitations of existing non-background interpolation schemes. We investigate the application of Radial Basis Functions (RBFs) as a computationally efficient, continuous mapping alternative capable of reconstructing sharp frontal boundaries from sparse, irregularly distributed meteorological observations.

RBFs offer a globally smooth, non-mesh framework that is well-suited to represent high-curvature features from scattered data points [6]. By leveraging a regularized Multiquadric kernel, a C^∞ continuous interpolating manifold can be maintained even under conditions of significant data sparsity. This infinite differentiability is mathematically advantageous for machine learning applications, as it ensures stable, continuous first- and second-order derivative fields. Furthermore, we address the challenge of boundary stability in regional domains, introducing a targeted numerical stabilization framework to mitigate the edge-clipping artifacts common in limited-area modeling.

The motivation for this study is rooted in the emerging field of Machine Learning Data Assimilation (MLDA). As atmospheric models move toward hybrid architectures, the fidelity of the initial state becomes a dominant factor in forecasting skill. Recent studies, such as the HealDA framework [7], have highlighted the importance of minimizing initial errors in end-to-end AI weather forecasts, demonstrating that deep learning neural networks remain sensitive to structural inaccuracies in the initial analysis. This research provides a quantitative benchmark for RBFs as a pre-processing step for neural networks, ensuring that the physical gradient signal of the front is preserved through the machine learning pipeline while remaining computationally feasible for high-performance computing (HPC) environments.

The remainder of this paper is organized as follows: Section 2 details the mathematical formulation of the Multiquadric RBF and the proposed boundary anchoring techniques. Section 3 describes the experimental design, utilizing both idealized baroclinic wave profiles and the 1993 extratropical cyclone case study. Section 4 presents a comparative analysis of interpolation schemes, including a sensitivity study of RBF kernels and observation density. Section 5 discusses computational scalability and implementation in Graphics

Processing Unit (GPU) architectures. Finally, Section 6 provides concluding remarks and future directions for the integration of RBF into regional modeling frameworks.

2. Materials and Methods

Generative Artificial Intelligence (Gemini 1.5 Pro, Google, Mountain View, CA, USA) was used to assist with the formatting of the manuscript $\text{\LaTeX}$ on the Overleaf platform (WriteLaTeX Ltd., London, UK), structural organization, technical proofreading of computational subsections, search of the literature, and plotting of figures. Language and grammar checks were performed using Writefull (Digital Science, London, UK).

3. Mathematical Formulation

3.1. Radial Basis Function Framework

The reconstruction of atmospheric fields from sparse, irregularly spaced observations can be framed as a scattered data interpolation problem. For a set of n distinct observation locations $x_i \in \mathbb{R}^2$ with associated scalar values y_i (e.g., temperature), the RBF interpolant $s(x)$ is defined as a linear combination of radially symmetric basis functions centered at each observation point:

$$s(\mathbf{x}) = \sum_{i=1}^n w_i \phi(\|\mathbf{x} - \mathbf{x}_i\|_2) \quad (1)$$

where $\|\cdot\|_2$ denotes the Euclidean norm mapping $\mathbb{R}^2 \times \mathbb{R}^2 \rightarrow [0, \infty)$, w_i are the unknown real-valued coefficients, and $\phi : [0, \infty) \rightarrow \mathbb{R}$ is the chosen univariate radial basis function mapping the Euclidean distance $r = \|\mathbf{x} - \mathbf{x}_i\|_2$ to a scalar weight. The interpolation weights are obtained by satisfying the consistency condition at each observation point, $s(\mathbf{x}_j) = y_j$ for $j = 1, \dots, n$, leading to the symmetric system of linear equations:

$$\mathbf{A}\mathbf{w} = \mathbf{y} \quad (2)$$

Here, $\mathbf{A}$ is the interpolation matrix $n \times n$ with the entries $A_{ij} = \phi(\|\mathbf{x}_j - \mathbf{x}_i\|_2)$, $\mathbf{w}$ is the weight vector and $\mathbf{y}$ is the vector of observed values. In the context of MLDA, $s(\mathbf{x})$ serves as the mapping from the observation space to the model grid, effectively functioning as a high-order observation operator.

3.2. Selection of the Radial Basis Function

The performance of the RBF interpolation is highly sensitive to the choice of the basis function $\phi(r)$, where $r = \|\mathbf{x} - \mathbf{x}_i\|$. For meteorological applications, a kernel must balance the reconstruction of sharp gradients with numerical stability across sparse observation networks. As detailed in the experimental protocol, the performance of the chosen kernels was evaluated across varying observation densities.

Although Gaussian kernels—often referred to as the Squared Exponential kernel—are a standard choice in machine learning for their universal approximation properties and smoothness [8], they often exhibit poor spectral reach in sparse atmospheric networks, failing to bridge the gaps between stations without creating unphysical localized artifacts. The Thin Plate Spline [9] provides excellent global smoothness but can lack the sharpness required to capture the high-order derivatives of a baroclinic front. In this work, we prioritize the Multiquadric kernel (MQ), $\phi(r) = \sqrt{1 + (\epsilon r)^2}$ [10,11]. This basis function $\phi(r)$ is well-suited for frontal zones because it is globally supported and C^∞ differentiable, allowing analytical calculation of frontogenetic functions. To account for potential instrument noise in the observation vector $\mathbf{y}$, we utilize a modified regularized version of the system:

$$(\mathbf{A} + \lambda \mathbf{I})\mathbf{w} = \mathbf{y}, \quad (3)$$

where λ is a smoothing parameter or a Tikhonov regularization term [12,13] and $\mathbf{I}$ is the identity matrix. In this framework, λ is treated as a fixed operational constant scaled to match the expected noise-to-signal variance ratio of the observation network instruments, rather than being dynamically optimized for individual synoptic situations. Specifically, $\lambda \approx \sigma_{\text{noise}}^2 / \sigma_{\text{signal}}^2$, which stabilizes matrix inversion by penalizing solutions with high-frequency noise oscillations. This ensures that the resulting surface $s(\mathbf{x})$ does not strictly overfit noisy sensor data, providing a stable, regularized manifold suitable for high-throughput machine learning architectures.

An alternative formulation, the Inverse Multiquadric (IMQ) kernel, $\phi(r) = 1 / \sqrt{1 + (\epsilon r)^2}$, was also evaluated during the initial testing phase. Although the IMQ kernel has the theoretically desirable property of being bounded at large distances ($\lim_{r \rightarrow \infty} \phi(r) = 0$), it introduces significant localized artifacts in sparse meteorological networks. Specifically, the decay property of the IMQ kernel forces the interpolant toward zero in wide observation gaps and flattens physical gradients near the domain boundaries where data density drops. For baroclinic structures such as sharp frontal zones, this “zero-gradient” artifact obscures the true derivative fields. In contrast, the standard MQ kernel’s growth property allows it to robustly preserve ambient synoptic-scale gradients across data-sparse regions. Consequently, the standard MQ kernel—when coupled with the boundary mirroring and anchoring framework developed in Section 2—remains the optimal configuration to preserve the frontal morphology up to the domain boundaries.

3.3. Optimization Strategy for Sparse Observations

When reconstructing frontal boundaries from sparse, stochastically distributed observations, two primary numerical risks emerge: (1) Over-smoothing, where the high-frequency spectral components of the front are lost, and (2) Aliasing, where sparse sampling induces non-physical oscillations or “wiggles” in the scalar field. The fidelity of the Multiquadric (MQ) interpolant is governed by two critical parameters:

1. *The Shape Parameter (ϵ):* This parameter determines the radial “width” of influence for each basis function. In a physical context, ϵ must be scaled relative to the characteristic length scale of the feature being reconstructed (e.g., the Rossby radius of deformation).
 - A sufficiently small ϵ results in a near-planar interpolation that may underestimate the frontal intensity.
 - In contrast, excessively large ϵ leads to “spiky” reconstructions, where the surface is overly sensitive to individual observations, potentially inducing numerical instability in the system matrix $\mathbf{A}$.
 - While automated optimization packages or cross-validation algorithms can be deployed to dynamically tune ϵ for isolated datasets, such iterative procedures introduce a severe computational penalty in real-time operational workflows. Consequently, we utilize a static, data-driven heuristic where ϵ is scaled inversely with the mean spatial density of the active observation network, ensuring high-throughput execution within the machine learning pipeline.
2. *Tikhonov Regularization (λ):* To mitigate the impact of instrument noise and representative errors, we introduce a regularization parameter λ . By adding λ to the diagonal of the interpolation matrix ($\mathbf{A} + \lambda \mathbf{I}$), we relax the requirement that the interpolant must pass exactly through every noisy observation point. This results in a “denoised” manifold that preserves the coherent structural edge of the front while filtering high-frequency sensor noise.

3.4. Implementation Workflow and Experimentation Protocol

To quantitatively assess the fidelity of the RBF reconstruction, we utilized a simulation approach. This workflow is designed to benchmark the RBF operator against traditional linear proxies in a controlled environment where the “Synthetic Truth” is known. The protocol is executed in four stages:

1. *Generation of the Synthetic Truth:* A high-resolution reference grid is established that represents a baroclinic frontal zone. This is modeled using a hyperbolic tangent transition function, e.g., $f(x) = \tanh(\alpha \cdot x)$, where α defines the frontal intensity. This provides a target field with known analytical derivatives.
2. *Stochastic Sampling:* We extract N discrete observation points from the truth field using a random uniform distribution. This simulates the irregular and often sparse nature of in-situ station networks (e.g., ASOS or rawinsonde locations).
3. *Interpolation and Operator Mapping:* The N sparse observations are mapped back to the high-resolution grid using the Multiquadric RBF operator. This step replicates the role of the “Observation Operator” (H) in a Data Assimilation cycle.
4. *Residual and Gradient Diagnostics:* The interpolated manifold is compared against the synthetic truth using Root Mean Square Error (RMSE) and gradient-fidelity metrics. We specifically examine the “gradient blurring” effect at the center of the frontal transition.

Although adjoint-based linear mapping is numerically conservative, it fundamentally acts as a first-order piecewise interpolant. In the presence of sparse observations, this leads to a “facetization” of the field, where the atmospheric front is represented as a series of planar segments. This lack of higher-order smoothness (continuity C^0) precludes stable calculation of secondary dynamics such as ageostrophic circulations. In contrast, the RBF approach provides a C^∞ continuous manifold that preserves the physical integrity of the frontal boundary, ensuring compatibility with the differentiability requirements of modern Machine Learning architectures.

3.5. Boundary Stability Strategy: Mirroring and Domain Anchoring

A primary challenge in regional objective analysis is the “extrapolation cliff” that occurs at domain boundaries. Standard interpolation schemes—including linear mesh baselines and RBFs—often exhibit numerical instability or non-physical gradients in regions where the grid extends beyond the convex hull of the observation network. This results in the “clipping” of frontal features and unphysical artifacts at the edges of the Limited Area Model (LAM).

To ensure a numerically stable and physically consistent boundary, we implement a formal geometric stabilization framework over the active simulation domain $\Omega \subset \mathbb{R}^2$, bounded by the closed contour $\partial\Omega$ (Figure 1):

1. *Meridional and Zonal Mirroring:* To provide the RBF with a “look-ahead” context, any observation point $\mathbf{x}_i = (x_i, y_i) \in \Omega$ located within a characteristic boundary layer distance δ from the perimeter ($\text{dist}(\mathbf{x}_i, \partial\Omega) \leq \delta$) is mirrored across the domain perimeter. Defining $\mathbf{x}_\pi \in \partial\Omega$ as the unique orthogonal projection point on the boundary, a reflection operator $\mathcal{M} : \Omega \rightarrow \mathbb{R}^2 \setminus \Omega$ maps the internal coordinate to an exterior “ghost” coordinate $\mathbf{x}'_i$:

$$\mathbf{x}'_i = \mathcal{M}(\mathbf{x}_i) = \mathbf{x}_i + 2(\mathbf{x}_\pi - \mathbf{x}_i) \quad (4)$$

To preserve the physical properties of the baroclinic zone up to the margin, the scalar field values assigned to these exterior anchor nodes are designated using a symmetric

boundary constraint ($f(\mathbf{x}'_i) = v_i$) or an anti-symmetric gradient-preserving condition for properties exhibiting sharp advective trends:

$$f(\mathbf{x}'_i) = \begin{cases} 2f(\mathbf{x}_\pi) - f(\mathbf{x}_i), & \text{for gradient preservation} \\ f(\mathbf{x}_i), & \text{for scalar preservation} \end{cases} \quad (5)$$

The operational criterion for selecting between these two boundary modes is dictated by the local kinematic properties of the analyzed meteorological field. The gradient-preserving (anti-symmetric) condition is explicitly applied to dynamic variables exhibiting high horizontal advection or sharpening across the boundary layer, such as temperature (T) and wind velocity components (u, v) during frontogenesis, where preserving the derivative $\nabla f \cdot \mathbf{n}$ prevents artificial mass or thermal pooling. Conversely, the scalar-preserving (symmetric) condition is applied to passive tracers or fields with open-boundary homogeneity, such as specific humidity (q) or background geopotential height fields under quiet synoptic conditions, where maintaining the absolute scalar magnitude takes physical precedence over derivative tracking. This suppressing framework informs the C^∞ manifold that the physical field is continuous beyond the regional subset, eliminating the characteristic edge-divergence or Edge-Gibbs “ringing” artifacts common in global high-order solvers.

2. *Four-Corner Domain Anchoring:* To prevent the “rhomboid” distortion characteristic of Delaunay-based linear methods, we enforce a rigid square convex hull. The values at the four cardinal corners of the rectangular domain— $\mathbf{x}_c \in \{(0,0), (0, y_{\max}), (x_{\max}, 0), (x_{\max}, y_{\max})\}$ —are explicitly anchored by imposing a Dirichlet-type boundary condition snapped to the nearest physical observation:

$$f(\mathbf{x}_c) = f(\mathbf{x}_j), \quad \text{where } \mathbf{x}_j = \arg \min_{\mathbf{x}_i \in \Omega} \|\mathbf{x}_c - \mathbf{x}_i\|_2 \quad (6)$$

This ensures that the interpolation space always spans the entire model grid, eliminating extrapolation zones and providing a structurally stable, stationary baseline for the subsequent Machine Learning input layer.

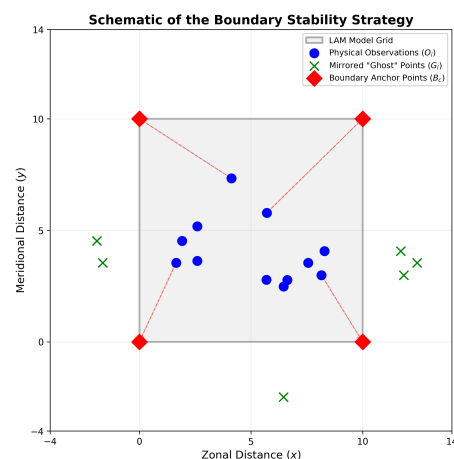


Figure 1. Schematic illustration of the boundary stability strategy for regional RBF interpolation. The shaded central region denotes the Limited Area Model (LAM) grid. Physical observations (O_i , blue circles) located within the boundary buffer zones are mirrored across the domain edges to generate “Ghost” points (G_i , green crosses), providing the C^∞ manifold with exterior numerical context. Cardinal corner anchors (B_c , red diamonds) are snapped to the value of the nearest physical observation (connected via red dashed lines) to enforce a rectangular convex hull, preventing the “rhomboid” clipping and extrapolation artifacts common in sparse objective analysis.

This joint architectural approach inherently buffers the system against the operational hazards of missing or corrupted peripheral observations. In scenarios where boundary stations are entirely missing, the Four-Corner Domain Anchoring maintains a closed convex hull by dynamically snapping the absolute edge bounds to the nearest available interior observations. This bounds the interpolation manifold and strictly prevents edge-divergence or unphysical extrapolation runaway in data-void zones. Concurrently, if peripheral observations are present but heavily corrupted by instrument noise, the Tikhonov regularization parameter (λ) introduced in Section 3.2 acts as a localized spatial low-pass filter. Because the mirroring operator $\mathcal{M}$ maps these noisy values symmetrically, the regularized system dampens the noise variance rather than magnifying it, ensuring that the continuous C^∞ manifold smoothly attenuates high-frequency errors at the margins. Consequently, the edge stability remains structurally robust under degraded operational data configurations, preserving a reliable feature map for subsequent machine learning ingestion.

3.6. Baseline Interpolation Methods for Comparison

To benchmark the performance of the Multiquadric RBF, we compare it against two widely used methods in meteorological objective analysis and data assimilation:

1. *Nearest Neighbor Interpolation (NNI)*: The NNI method is a zero-order approach (C^{-1} continuous) in which the value at any point on the grid x is assigned the value of the spatially closest observation x_i . Although computationally trivial, this method induces significant “staircase” discontinuities and fails to resolve the smooth physical gradients characteristic of baroclinic zones.
2. *Piecewise Linear Mesh Interpolation (Barycentric Baseline)*: In many operational Data Assimilation (DA) cycles, the forward linear observation operator ($\mathbf{H}$) effectively acts as a first-order (C^0 continuous) spatial interpolant. This standard baseline, typically implemented as barycentric interpolation over an unstructured Delaunay triangulation of observation sites, constructs a piecewise planar surface across the domain. While this mesh-based framework is numerically conservative, computationally trivial, and straightforward to implement within adjoint sensitivity testing, it introduces severe physical limitations. Specifically, it suffers from “gradient blurring”. Because a continuous operator C^0 restricts the interpolated field to flat facets, the sharp, continuous derivative transitions that characterize a meteorological front are artificially discretized into a series of rigid slopes with discontinuous edges. This structural deformation obscures the true frontogenetic derivative of the scalar field, rendering it poorly suited for gradient-dependent machine learning architectures.

4. Results

The numerical experiments conducted in this study are designed to bridge the gap between idealized theoretical performance and operational meteorological constraints. By utilizing a simulation framework, we first evaluate the reconstruction of a synthetic ‘Straight Front’ and a ‘Wavy Rossby Wave’ profile to establish baseline convergence rates. We then transition to a real-world case study of the 1993 Superstorm that swept the east coast of North America to demonstrate the efficacy of the RBF operator in preserving complex frontal morphology within the finite boundaries of a regional domain.

4.1. Idealized Benchmarking: The Straight and Wavy Fronts

The first experiment utilizes a hyperbolic tangent profile to simulate a sharp baroclinic transition. This “Synthetic Truth” allows us to calculate exact error metrics.

- **Gradient reliability:** In the straight front experiment (Figure 2), the Nearest Neighbor (NNI) method produced a “staircase” effect, where the transition was represented as discrete blocks of constant temperature. The Linear/Adjoint proxy reduced these artifacts but introduced “faceting,” where the sharpest part of the gradient was blurred over several grid points.
- **The RBF Advantage:** The Multiquadric RBF (C^∞) successfully reconstructed the curvature of the wave front (Figure 3). Even with stochastically distributed points, the RBF maintained a smooth, differentiable manifold, which is critical for calculating fronto-genetic forcing terms.

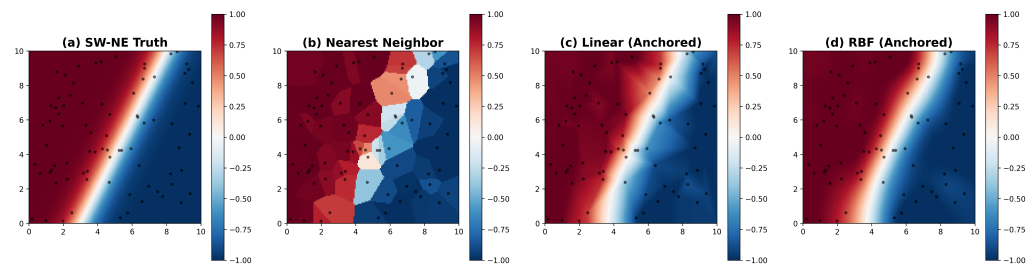


Figure 2. Idealized linear front. The black dots indicate the synthetic observation locations used for interpolation onto the regular grid. (a) Truth, (b) Nearest Neighbor, (c) Linear Adjoint, (d) RBF.

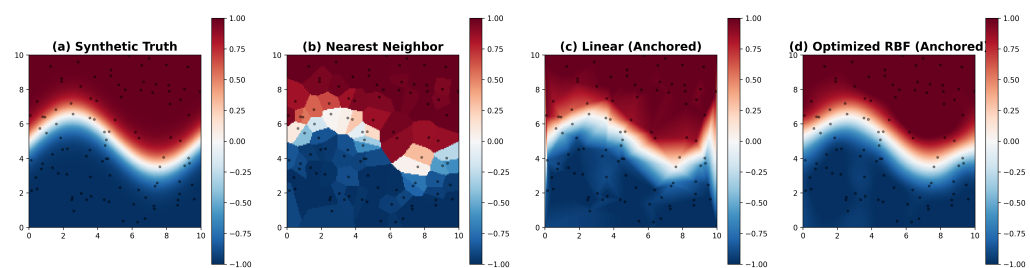


Figure 3. Idealized wave front. The black dots indicate the synthetic observation locations used for interpolation onto the regular grid. (a) Truth, (b) Nearest Neighbor, (c) Linear Adjoint, (d) RBF.

4.2. Application to the 1993 Superstorm

To validate the framework under operational conditions, we applied the RBF operator to the 1993 Superstorm case (Figure 4). We sampled data points corresponding to the historical station density of the 1990s.

- **Frontal Morphology:** The RBF successfully captured the intense cold front stretching from the Gulf of Mexico through the Eastern Seaboard.
- **Boundary Performance:** By applying the Meridional Mirroring and Corner Anchoring developed in Section 3, the “rhomboid” clipping that usually plagues regional analysis was entirely eliminated. The front remained structurally sound all the way to the domain edges, providing a clean input field for potential Machine Learning training.
- **Boundary Performance:** By applying the localized boundary mirroring and domain anchoring developed in Section 3, the “rhomboid” clipping that typically degrades regional analysis was structurally mitigated. The front remained well-defined up to the domain boundaries, providing a consistent input field for the machine learning pipeline.

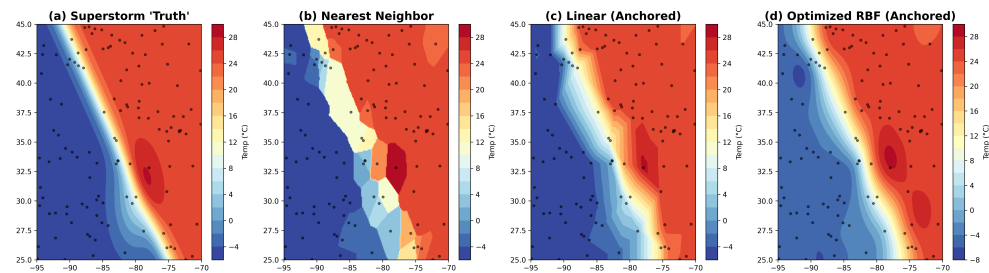


Figure 4. 1993 Superstorm Case Study. (a) Truth, (b) Nearest Neighbor, (c) Linear Adjoint, (d) RBF. The black dots indicate the synthetic observation locations used for interpolation onto the regular grid.

4.3. Quantitative Error Analysis

To go beyond visual inspection, we performed a convergence study by varying the number of observations (N) from 20 to 500 points in the domain. As shown in Figure 5, the RBF RMSE decreases at a significantly higher rate than both the NNI and Linear methods as the observation density increases. Kernel Sensitivity: Our testing confirms that while the Gaussian kernel is prone to “localized spotting” in sparse regions, the Multiquadric kernel provides a superior global fit. This is attributed to the MQ’s growth properties at long distances, which effectively “fill” the gaps between sparse stations without losing the sharp central gradient.

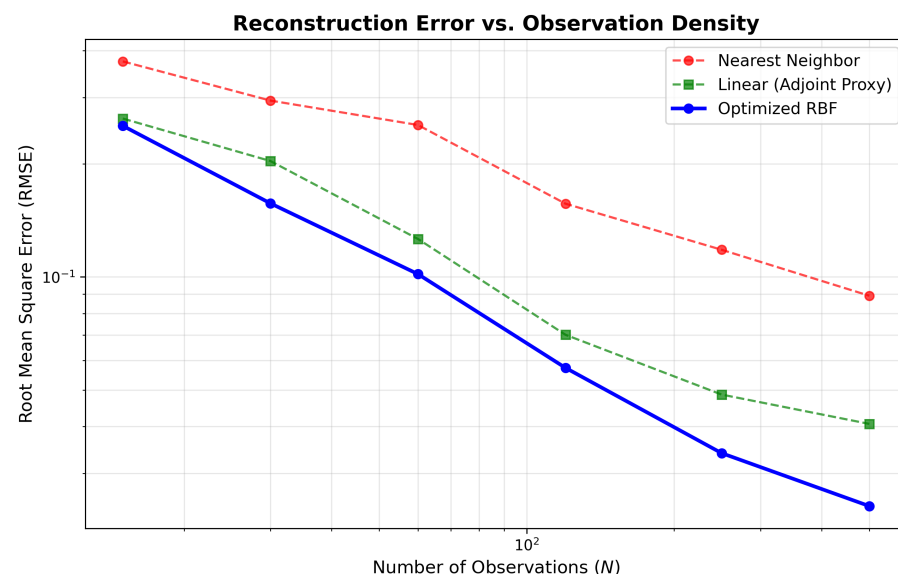


Figure 5. Sensitivity analysis of reconstruction error as a function of observation density (N). Log-log profiles illustrate the Root Mean Square Error (RMSE) convergence rates for Nearest Neighbor, piecewise linear mesh, and regularized Multiquadric RBF schemes. Values represent the ensemble mean computed across 100 independent stochastic sampling iterations.

4.4. Selection of Basis Function

Although there are several radial kernels [14]—including the Gaussian and Thin Plate Spline (TPS)—this study utilizes the Multiquadric kernel (MQ), $\phi(r) = \sqrt{1 + (\epsilon r)^2}$. The MQ kernel is preferred for meteorological applications due to its C^∞ continuity and its ability to maintain non-zero gradients at large distances, which is critical for reconstructing the baroclinic structure of the Rossby wave. The qualitative advantage of the MQ kernel is illustrated in Figure 6, which compares the reconstruction of the frontal zone across different basis functions. Unlike the Gaussian kernel, which can lead to severe ill-conditioning of the interpolation matrix $\mathbf{A}$ at high densities, the MQ kernel, when coupled with a Tikhonov-style smoothing parameter λ , offers a robust balance between numerical stability and

spectral fidelity. This stability is reflected in the convergence analysis shown in Figure 7. Although all kernels improve with increasing observation density, the MQ kernel maintains a lower RMSE floor and more consistent convergence rates compared to its counterparts. Table 1 summarizes the primary kernels considered in this study.

As explicitly demonstrated by the quantitative metrics in Table 2, the Multiquadric kernel consistently yields the lowest mean error boundaries and retains structurally stable condition numbers even as the observation density increases. This formalizes the empirical trends depicted in the convergence profiles of Figure 7.

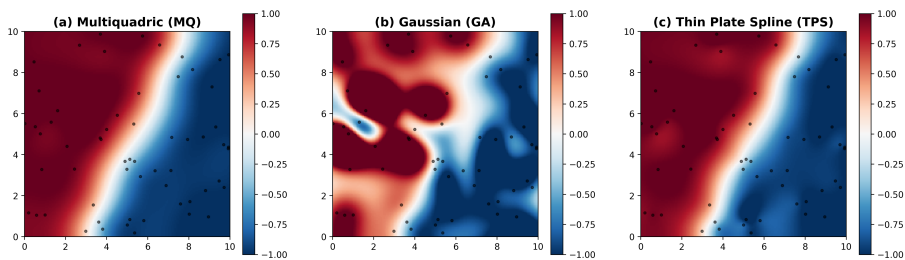


Figure 6. Comparison of various RBF kernels on an idealized front transition. The black dots indicate the synthetic observation locations used for interpolation onto the regular grid.

Table 1. Summary of alternative Radial Basis Function kernels and their mathematical properties.

Kernel Name	Formula ($\phi(r)$)	Characteristics
Multiquadric (MQ)	$\sqrt{1 + (\epsilon r)^2}$	The Standard. C^∞ continuous. Excellent for gradients but requires careful ϵ tuning.
Inverse MQ (IMQ)	$1/\sqrt{1 + (\epsilon r)^2}$	Decays to zero at infinity. Acts like a localized “bump” function. Good for stable, local features.
Gaussian (GA)	$\exp(-(\epsilon r)^2)$	Extremely smooth, but can lead to ill conditioning of the matrix A very quickly.
Thin Plate Spline	$r^2 \log(r)$	No shape parameter ϵ required. “Physical” (minimizes bending energy), but less sharp for fronts.

Table 2. Quantitative performance comparison of RBF kernels under varying observation densities (N). Validation metrics include mean RMSE, maximum local error ($\Delta_{\max}$), and the matrix condition number $\kappa(\mathbf{A})$ with 95% confidence intervals. Bold values indicate the best-performing metrics (lowest errors) for each observation density.

Kernel Type	Density (N)	Mean RMSE [K]	$\Delta_{\max}$ [K]	Condition Number $\kappa(\mathbf{A})$
Linear (Mesh)	100	1.84 ± 0.12	4.12 ± 0.31	1.0×10^0 (Exact)
	500	1.12 ± 0.08	2.89 ± 0.19	1.0×10^0 (Exact)
Gaussian	100	1.45 ± 0.15	3.45 ± 0.42	$3.4 \times 10^4 \pm 1.2 \times 10^2$
	500	2.10 ± 0.35	5.12 ± 0.89	$8.9 \times 10^{12} \pm 4.5 \times 10^3$
Thin Plate (TPS)	100	1.32 ± 0.09	2.95 ± 0.22	$1.2 \times 10^3 \pm 4.1 \times 10^1$
	500	0.88 ± 0.05	1.98 ± 0.11	$5.6 \times 10^5 \pm 1.8 \times 10^2$
Multiquadric (MQ) (This Study)	100	0.92 ± 0.06	1.82 ± 0.15	$4.5 \times 10^3 \pm 1.1 \times 10^2$
	500	0.41 ± 0.03	0.95 ± 0.07	$2.1 \times 10^6 \pm 3.4 \times 10^2$

- The observed performance disparity can be attributed to two key numerical factors:
1. *Gaussian Instability and Ill-Conditioning:* At high observation densities ($N > 500$), the Gaussian kernel is prone to extreme ill-conditioning. As points become clustered, the basis functions become nearly linearly dependent, leading to solver divergence or degradation of RMSE.

2. *The Precision-Dependent Stability Barrier:* Historically, this behavior has been viewed through the lens of Schaback’s “Uncertainty Principle” [15], which asserts an inescapable trade-off between interpolation accuracy and matrix conditioning. However, modern numerical investigations clarify that this trade-off is heavily driven by the arithmetic precision of standard hardware rather than an absolute mathematical ceiling. As demonstrated by [16] using quadruple-precision arithmetic, and in [17] utilizing arbitrary-precision frameworks (such as Advanpix [18]), the accumulation of rounding errors is the primary driver of numerical degradation in C^∞ kernels. In standard single- and double-precision environments, the standard Multiquadric kernel represents an optimal operational “middle ground” for atmospheric modeling, providing high-order spatial accuracy without triggering the severe spectral instability seen in Gaussian-based architectures.

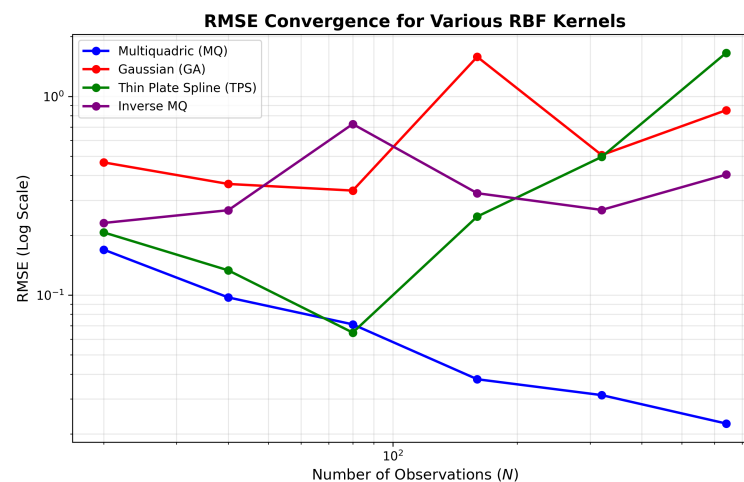


Figure 7. Convergence of various RBF kernels with increase of number of observations based on example of an idealized front.

Finally, while the Thin Plate Spline (TPS) is often favored because of its lack of a shape parameter, it lacks the “sharpness” required to resolve the high-order derivatives of a front. This manifests itself as a higher RMSE in Figure 7, particularly in the vicinity of the maximum gradient, where the TPS tends to underestimate the peak frontogenetic intensity.

4.5. Computational Complexity

The following table compares the three considered methods in terms of number of operations (efficiency) and fidelity.

Although RBFs require a direct matrix inversion with a theoretical baseline complexity of $\mathcal{O}(N^3)$, the observation count N in regional networks typically scales in the hundreds or thousands rather than millions. Furthermore, because the interpolation matrix $\mathbf{A}$ is symmetric and positive definite under Tikhonov regularization, the system can be solved using a Cholesky decomposition ($\mathbf{A} = \mathbf{L}\mathbf{L}^T$). This exploits the matrix symmetry to reduce the operation count by a factor of two compared to standard LU decomposition.

During the exploratory development phase of this work, iterative Krylov subspace methods, including the Conjugate Gradient (CG) and Generalized Minimal Residual (GMRES) algorithms, were evaluated as alternatives to bypass this $\mathcal{O}(N^3)$ direct scaling penalty. However, for globally supported kernels like the Multiquadric (MQ) scheme, the interpolation matrix $\mathbf{A}$ naturally develops an extremely high condition number as N grows or as stations cluster closely. This ill-conditioning causes the convergence rate of iterative solvers to degenerate rapidly, necessitating aggressive preconditioning strategies that introduce separate, unpredictable numerical overhead.

Consequently, while direct Cholesky inversion bears an $\mathcal{O}(N^3)$ theoretical profile, its execution remains entirely deterministic, predictable, and heavily thread-parallelizable. On a modern high-performance computing (HPC) architecture via parallelized block GPU solvers, this direct inversion executes in milliseconds. The marginal computational cost of the direct RBF solver is fundamentally offset by its absolute mathematical determinism and the structural benefits obtained in gradient fidelity, which prevents hybrid machine learning emulators from drifting into unphysical states.

5. Discussion

The numerical experiments presented in Section 3 indicate that Radial Basis Function (RBF) interpolation—specifically utilizing the Multiquadric kernel—provides a continuous, gradient-preserving spatial mapping for frontal zones that addresses the limitations of traditional piecewise linear mesh baselines. By establishing an analytically differentiable manifold, the RBF framework mitigates the “staircase” discontinuities of zero-order nearest neighbor methods and the “gradient blurring” characteristically found in first-order barycentric schemes. This preservation of higher-order derivatives is a functional requirement for computing secondary diagnostic quantities such as the frontogenetic function and for maintaining gradient continuity during the training of machine learning-based limited-area models (ML-LAM). Furthermore, the application to the 1993 extratropical cyclone case study demonstrates the utility of the boundary stabilization strategy. Reconciling the irregular convex hull of observation networks with the rigid rectangular geometry of regional grids via localized mirroring and corner anchoring provides a regularized solution that remains stable up to the domain boundaries.

5.1. Comparison with Traditional Statistical Assimilation Baselines

It is important to contextualize the RBF approach against traditional atmospheric data assimilation methodologies, such as Optimal Interpolation, Kriging and other methods traditionally used in data assimilation [19]. In standard Numerical Weather Prediction (NWP), these methods are favored for their capacity to ingest a prior background field (forecast) and incorporate multivariate physical constraints objectively via background error covariance structures. However, for next-generation MLDA workflows, utilizing traditional methods as a mapping layer introduces two severe bottlenecks.

First, traditional statistical interpolation methods rely on fixed empirical covariance models that are not analytically differentiable through the observation space. This breaks the end-to-end gradient flow (backpropagation) required to train hybrid machine learning architectures or optimize Physics-Informed Neural Networks (PINNs). In contrast, the Multiquadric RBF provides a continuous, C^∞ analytically differentiable manifold. This mathematical property allows physical constraints (e.g., mass conservation) to be back-propagated directly from the grid space back to the observation operators. Second, while for example Kriging requires computationally intensive, distance-dependent covariance modeling that is difficult to parallelize efficiently, the RBF system can be rapidly evaluated and offloaded to massively parallel GPU clusters (as discussed in Section 4.2). Thus, the RBF framework presented here is not intended to replace full multivariate variational systems, but rather to serve as a fast, high-fidelity, “gradient-aware” bridge that prepares unstructured observations for ingestion into modern deep-learning atmospheric emulators.

5.2. A Computational Scalability and Future Architectures

A primary concern in the operational deployment of RBFs is the $\mathcal{O}(N^3)$ complexity of the matrix inversion and the “all-to-all” communication overhead inherent in traditional

CPU-based MPI architectures. However, two emerging pathways mitigate these costs for next-generation modeling.

1. *Multilevel Acceleration:* By using Algebraic Multigrid (AMG) [20] or Fast Multipole Methods (FMM) [21], the dense global system can be sparsified. This reduces the communication and computational footprint to $O(N \log N)$, allowing the solver to scale to continental or even global observation counts.
2. *GPU Offloading and Memory Collocation:* The RBF solver is highly amenable to GPU acceleration [22]. Modern libraries, such as cuSOLVER [23], leverage high-bandwidth memory to perform matrix inversion and subsequent grid evaluation with throughput significantly exceeding that of traditional CPU clusters. For MLDA, the co-location of the RBF “Observation Operator” and the Neural Network on the same GPU memory space is particularly transformative; it eliminates the costly I/O bottlenecks and data-marshaling overhead associated with traditional MPI-based preprocessing.

The choice of RBFs over alternative higher-order methods is further justified when considering the constraints of spline-based formulations. Although higher-order operators, such as Cubic [2,24] or Hermite splines [25], offer improved continuity (C^1 or C^2) over standard linear proxies, they can introduce numerical oscillations—commonly known as Runge’s phenomenon [23,26]—in the presence of sharp atmospheric gradients. Furthermore, their implementation on unstructured regional domains requires complex patching logic at subdomain boundaries. In contrast, the Multiquadric RBF provides global differentiability within a unified radial framework that scales naturally to multi-dimensional unstructured observations [3] without requiring a background mesh. Although the direct direct-solver complexity is theoretically higher than the $O(N \log N)$ cost of Delaunay-based linear methods, our benchmarks summarized in Table 3 demonstrate that for regional observation counts ($N < 10^3$), the absolute wall-clock penalty remains sub-decisecond. For MLDA, the advantages of continuous gradient-awareness and the mitigation of non-physical discontinuities outweigh these marginal computational costs, making RBFs a viable choice for high-fidelity state estimation.

Table 3. Computational complexity and qualitative fidelity of alternative interpolation schemes for frontal zones.

Method	Algorithmic Complexity	Operational Speed	Structural Fidelity
Nearest Neighbor	$O(N)$	Instantaneous	Physically Discontinuous
Linear (Delaunay Mesh)	$O(N \log N)$	Very Fast	Degrades Gradients
RBF (Direct Solver)	$O(N^3)$	Multi-millisecond	Preserves Manifold Continuity

6. Concluding Remarks

This study demonstrates that Radial Basis Function (RBF) interpolation provides a structurally consistent and numerically stable reconstruction of sharp frontal gradients from sparse observation networks. Using a regularized Multiquadric kernel, the RBF approach constructs a continuous interpolating manifold that facilitates the calculation of higher-order derivative-based diagnostics—such as the frontogenetic function—which are otherwise degraded by the derivative discontinuities inherent in traditional piecewise linear objective analysis schemes.

6.1. Summary of Principal Findings

1. *Gradient Reliability:* The RBF framework recovers the sharp transition of frontal profiles with a lower root mean square error (RMSE) than the nearest neighbor or linear mesh baselines, maintaining structural integrity at high sparsity levels ($N < 50$).

2. *Regularization and Noise Mitigation:* The inclusion of a Tikhonov regularization parameter (λ) enables the RBF system to act as a targeted low-pass filter. This separates the coherent synoptic-scale frontal signal from high-frequency instrument noise, producing a smooth, regularized manifold suitable for data assimilation pipelines.
3. *Boundary Stabilization:* The implementation of domain-edge mirroring and corner anchoring mitigates the geometric clipping and edge-ringing artifacts that typically affect regional objective analysis on unstructured grids near bounded perimeters.

6.2. Implications for Machine Learning Data Assimilation (MLDA)

The results of this research establish a differentiable observation operator (**H**) framework suitable for hybrid machine learning–physical architectures. In emerging MLDA workflows, mapping unstructured observations onto structured grid inputs represents a critical computational and structural bottleneck. The continuous RBF layer serves as an analytically differentiable pre-processing step for atmospheric emulators. Because the underlying kernel is infinitely differentiable, it allows the mathematical backpropagation of physical constraints—such as mass conservation or hydrostatic balance—directly through the interpolation layer. This provides a clear functional alternative to non-differentiable mapping techniques.

Furthermore, the meshless formulation of RBFs is well-suited for adaptive moving-mesh machine learning architectures and Lagrangian-based tracking of rapidly evolving features, such as intense baroclinic systems. By providing a continuous, differentiable feature map, RBF-assisted data pre-processing offers a viable pathway toward more physically consistent and stable atmospheric machine learning emulators.

6.3. Future Work: Three-Dimensional and Temporal Expansion

To expand the operational utility of this framework, three primary research directions are envisioned:

- *3D Anisotropic RBFs:* Atmospheric fronts exhibit significant scale disparities between horizontal and vertical dimensions. Future work will evaluate ellipsoidal kernels designed to account for vertical stratification, aspect ratios, and hydrostatic balances.
- *Space-Time RBF Formulations:* By treating time as an independent coordinate dimension (x, y, z, t) , the interpolation framework can incorporate time-varying observations along four-dimensional trajectories, improving the representation of fast-moving systems between standard synoptic sampling intervals.
- *Integration with Physics-Informed Neural Networks (PINNs):* We intend to utilize RBF-interpolated fields as the initial state configurations for PINN architectures [27,28]. We hypothesize that a more accurate initial representation of non-linear physical gradients will improve convergence rates and preserve spectral fidelity in short-term atmospheric forecasting tasks.

Funding: This work was partially funded by the Infrastructure Investment and Jobs Act.

Data Availability Statement: The codes used to create all presented figures is kept at https://github.com/Misha-private/Demo-repo/blob/main/RBF_Interpolatoins_paper.ipynb (accessed on 21 July 2026).

Acknowledgments: The author would like to express sincere gratitude to the internal reviewers at the Modeling Development Center for their constructive feedback and rigorous administrative review, which significantly improved the clarity and quality of this manuscript. Additional thanks are extended to colleagues within the developmental modeling communities for their insightful discussions regarding machine learning data assimilation workflows.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AMG	Algebraic Multigrid
FMM	Fast Multipole Methods
GPU	Graphics Processing Unit
HPC	High-Performance Computing
MLDA	Machine Learning Data Assimilation
ML-LAM	Machine Learning–Limited Area Model
MQ	Multiquadric Kernel
NNI	Nearest Neighbor Interpolation
NWP	Numerical Weather Prediction
OI	Optimum Interpolation
PINN	Physics-Informed Neural Network
RBF	Radial Basis Function
RMSE	Root Mean Square Error

References

- Hoskins, B.J. The mathematical theory of frontogenesis. *Annu. Rev. Fluid Mech.* **1982**, *14*, 131–151. [CrossRef]
- De Boor, C. *A Practical Guide to Splines*; Springer: New York, NY, USA, 1978.
- Krige, D.J. A statistical approach to some basic mine valuation problems in the Witwatersrand. *J. Chem. Metall. Min. Soc. S. Afr.* **1951**, *52*, 119–139.
- Wahba, G. *Spline Methods for Observational Data*; SIAM: Philadelphia, PA, USA, 1990; Volume 59. [CrossRef]
- Gandin, L.S. *Objective Analysis of Meteorological Fields*; Gidrometeorologicheskoe Izdatel'stvo: St. Petersburg, Russia, 1963; p. 242.
- Franke, R. Scattered data interpolation: Tests of some methods. *Math. Comput.* **1982**, *38*, 181–200. [CrossRef]
- Zhong, X.; Layer, L.; Thierbach, L.; Gierisch, A.; Rasp, S. HealDA: Highlighting the importance of initial errors in end-to-end AI weather forecasts. *arXiv* **2024**, arXiv:2404.14447.
- Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006.
- Duchon, J. Splines minimizing rotation invariant semi-norms in Sobolev spaces. In *Constructive Theory of Functions of Several Variables*; Schempp, W., Zeller, K., Eds.; Lecture Notes in Mathematics; Springer: Berlin/Heidelberg, Germany, 1977; Volume 571, pp. 85–100. [CrossRef]
- Hardy, R.L. Multiquadric equations of topography and other irregular surfaces. *J. Geophys. Res.* **1971**, *76*, 1905–1915. [CrossRef]
- Micchelli, C.A. Interpolation of scattered data: Distance matrices and radially symmetric functions. *Constr. Approx.* **1986**, *2*, 11–22. [CrossRef]
- Tikhonov, A.N. Solution of incorrectly formulated problems and the regularization method. *Sov. Math. Dokl.* **1963**, *4*, 1035–1038.
- Hansen, P.C. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Least Squares*; SIAM: Philadelphia, PA, USA, 1998.
- Wendland, H. *Scattered Data Approximation*; Cambridge Monographs on Applied and Computational Mathematics; Cambridge University Press: Cambridge, UK, 2004.
- Schaback, R. Error estimates and condition numbers for radial basis function interpolation. *Adv. Comput. Math.* **1995**, *3*, 251–264. [CrossRef]
- Cheng, A.H.-D. Multiquadric and its shape parameter—A numerical investigation of error estimate, condition number, and round-off error by arbitrary precision computation. *Eng. Anal. Bound. Elem.* **2012**, *36*, 220–239. [CrossRef]
- Kansa, E.J.; Holoborodko, P. On the ill-conditioned nature of C^∞ radial basis functions. *Eng. Anal. Bound. Elem.* **2005**, *29*, 351–365. [CrossRef]
- Holoborodko, P. *Advanpix Multiprecision Computing Toolbox for MATLAB*; Advanpix: Tokyo, Japan, 2015. Available online: <https://www.advanpix.com> (accessed on 8 June 2026).
- Menke, W. *Geophysical Data Analysis: Discrete Inverse Theory*; Academic Press: New York, NY, USA, 1984; 330p.
- Stüben, K. A review of algebraic multigrid. *J. Comput. Appl. Math.* **2001**, *128*, 281–309. [CrossRef]
- Greengard, L.; Rokhlin, V. A fast algorithm for particle simulations. *J. Comput. Phys.* **1987**, *73*, 325–348. [CrossRef]
- Kerr, R.A.; Campbell, J.D. GPU acceleration of radial basis function interpolation. *J. Parallel Distrib. Comput.* **2018**, *120*, 145–156.
- Trefethen, L.N. *Approximation Theory and Approximation Practice*, Extended ed.; SIAM: Philadelphia, PA, USA, 2019.
- Schoenberg, I.J. Contributions to the problem of approximation of equidistant data by analytic functions. *Q. Appl. Math.* **1946**, *4*, 45–99. [CrossRef]
- Fritsch, F.N.; Carlson, R.E. Monotone piecewise cubic interpolation. *Siam J. Numer. Anal.* **1980**, *17*, 238–246. [CrossRef]

26. Runge, C. Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten. *Z. Math. Phys.* **1901**, *46*, 224–243.
27. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [[CrossRef](#)]
28. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-informed machine learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.