*Environmental Data Science* (2026), 5: e17, 1–12
doi:10.1017/eds.2026.10048

CAMBRIDGE UNIVERSITY PRESS

METHODS PAPER

# Emulating non-differentiable metrics via knowledge-guided learning: Introducing the Minkowski image loss

Filippo Quarenghi[1], Ryan Cotsakis[2] and Tom Beucler[3]

[1]Faculty of Geosciences and Environment, University of Lausanne, Lausanne, Switzerland
[2]University of Lausanne, Switzerland
[3]Faculty of Geosciences and Environment—FGSE Unil, University of Lausanne, Switzerland
**Corresponding author**: Filippo Quarenghi; Email: filippo.quarenghi@unil.ch

**Received:** 29 May 2026 UTC; **Revised:** 29 May 2026 UTC; **Accepted:** 04 June 2026 UTC

**Keywords:** differentiability; knowledge-guided machine learning; Minkowski functionals; radar precipitation; super-resolution

## Abstract

The "differentiability gap" presents a primary bottleneck in Earth system deep learning: Since models cannot be trained directly on non-differentiable scientific metrics and must rely on smooth proxies (e.g., MSE), they often fail to capture high-frequency details, yielding "blurry" outputs. We develop a framework that bridges this gap using two different methods to deal with non-differentiable functions: The first is to analytically approximate the original non-differentiable function into a differentiable equivalent one; the second is to learn differentiable surrogates for scientific functionals. We formulate the analytical approximation by relaxing discrete topological operations using temperature-controlled sigmoids and continuous logical operators. Conversely, our neural emulator uses Lipschitz-convolutional neural networks to stabilize gradient learning via: (1) spectral normalization to bound the Lipschitz constant and (2) hard architectural constraints enforcing geometric principles. We demonstrate this framework's utility by developing the Minkowski image loss, a differentiable equivalent for the integral-geometric measures of surface precipitation fields (area, perimeter, and connected components). Validated on the EUMETNET OPERA dataset, our constrained neural surrogate achieves high emulation accuracy, completely eliminating the geometric violations observed in unconstrained baselines, which generate physically impossible precipitation fields in up to 7.8% of cases. However, applying these differentiable surrogates to a deterministic super-resolution task reveals a fundamental trade-off: While strict Lipschitz regularization ensures optimization stability, it inherently over-smooths gradient signals, restricting the recovery of highly localized convective textures. This work highlights the necessity of coupling such topological constraints with stochastic generative architectures to achieve full morphological realism.

## Impact Statement

In scientific machine learning, trusted metrics (e.g., object counting) are often non-differentiable, breaking the gradient signal required for training. Consequently, models rely on smooth proxies (like mean-squared error) and underperform on scientific validation. We provide a solution via a "differentiable surrogate": a secondary neural network that mimics these metrics. By embedding strict rules, we ensure the surrogate provides stable, valid feedback. While demonstrated on weather radar data to enforce storm geometry, this framework allows

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/eds.2026.10048 Published online by Cambridge University Press

researchers in any field (i.e., weather downscaling, wildfire segmentation, and ocean eddy detection) to train models directly with gold-standard domain metrics.

## 1. Introduction

Structural realism in predicted precipitation fields depends on metrics that are inherently non-differentiable, creating a bottleneck for gradient-based optimization. To bridge this gap, we leverage knowledge from integral geometry, which reveals organizational biases often missed by standard statistical verification. Unlike recent work exploring surrogates in latent spaces (Patel et al., 2020), we enforce these properties directly in the optimization landscape. We introduce a framework situated at the intersection of differentiable programming (Gelbrecht et al., 2023) and knowledge-guided machine learning (Karpatne et al., 2017) (Figure 1). Within this framework, we compare an analytical approximation against a neural emulator, both designed to map the level sets of precipitation fields to key integral-geometric measures (area, perimeter, and number of connected components). By relaxing this discrete combinatorial problem into a continuous regression task (Hu et al., 2019), we enable the direct backpropagation of structural error signals. Crucially, we ensure emulation stability by adopting a knowledge-guided Lipschitz-Convolutional Neural Network (Lip-CNN) (Gouk et al., 2021). This architecture combines spectral normalization—to prevent exploding gradients—with hard geometric constraints (monotonicity and isoperimetry) embedded in the computation graph.

Concurrent efforts to integrate non-differentiable operations into automatic differentiation frameworks, such as the SoftJAX and SoftTorch libraries (Paulus et al., 2026), provide general-purpose continuous relaxations for elemental primitives (e.g., thresholding and fuzzy logic). While these drop-in replacements facilitate the formulation of analytical approximations akin to our continuous baseline, cascading numerous element-wise soft operators through complex topological algorithms frequently exacerbates gradient attenuation and mathematical entanglement. In contrast, our Lipschitz-regularized neural emulation circumvents the need for explicit operator relaxation by learning a holistic, topologically informed mapping that ensures robust structural fidelity and stable gradient flow.

Our primary contributions are: (1) proposing a general framework for injecting non-differentiable domain knowledge into training via learned surrogates, compared against standard analytical approximations; (2) introducing the Minkowski image loss, a differentiable objective that penalizes errors in rigid-motion invariants (area, perimeter, and connected components); and (3) demonstrating that hard geometric constraints paired with Lipschitz regularization prevents violating geometric principles and overfitting observed in unconstrained baselines.

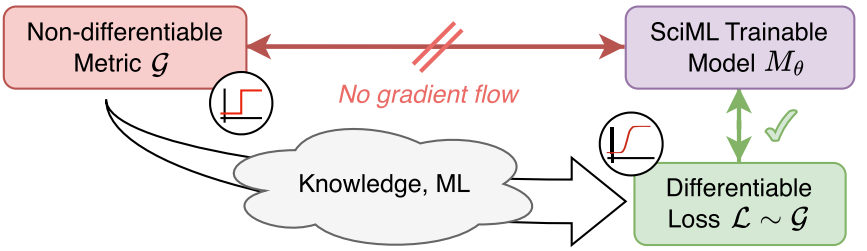



***Figure 1.*** *Conceptual overview of the differentiable surrogate training framework. The lack of gradient flow (red connection) prevents the direct optimization of SciML models ($M_\theta$) using non-differentiable scientific metrics ($\mathcal{G}$), which typically involve discrete operations like thresholding (step function plot). We overcome this limitation by using knowledge and machine learning techniques to create a differentiable surrogate loss ($\mathcal{L}$) that approximates the target metric. This differentiable loss, with a well-behaved optimization landscape, can then be successfully integrated into the model's training pipeline (green connection), enabling the direct backpropagation of structurally and physically relevant error signals.*

https://doi.org/10.1017/eds.2026.10048 Published online by Cambridge University Press

## 2. Methodology

### *2.1. Approaches to non-differentiable metrics*

Scientific evaluation often relies on functionals $f : \mathcal{X} \to \mathcal{Y}$ involving discrete operations such as counting, thresholding, or binning. These operations are inherently non-differentiable; they manifest as piecewise constant functions whose gradient $\nabla_{\mathbf{x}} f$ is zero almost everywhere or undefined at jump discontinuities (Dirac deltas), producing a loss surface of flat plateaus that prevents the backpropagation of error signals required to update an upstream model. Two primary methodologies exist to circumvent this limitation. The first relies on *analytical approximation*, wherein discrete operations are substituted with continuous, smooth relaxations, yielding a closed-form differentiable surrogate. The second leverages *neural emulation*, employing an artificial neural network to map the input space directly to the target metric; because neural networks are differentiable by construction, this learned surrogate integrates seamlessly into the computational graph.

#### *2.1.1. Analytical approximation of non-differentiable metrics*

The analytical approximation strategy aims to replace non-differentiable operations with smooth variants that permit stable gradient flow. For example, binary thresholding—mathematically represented by the Heavisine step function $H(x)$—is commonly relaxed using a logistic sigmoid function $\sigma(x/\tau)$. Here, $\tau > 0$ acts as a temperature parameter controlling the steepness of the approximation. As $\tau \to 0$, the continuous function converges to the exact discrete behavior, but its gradients vanish almost everywhere; conversely, larger values of $\tau$ provide stable gradients at the cost of approximation fidelity. Consequently, $\tau$ is often annealed dynamically during optimization. Similar continuous relaxations are employed for discrete selection operations, such as replacing $\arg\max$ with a differentiable softmax approximation (Jang et al., 2017). While these analytical substitutions preserve the mechanistic interpretability of the original metric, deriving stable, closed-form relaxations for complex, multistep spatial metrics is not always mathematically possible.

#### *2.1.2. Emulation of non-differentiable metrics: Surrogate modeling and Lipschitz regularization*

To restore differentiability, we approximate the discrete map $f$ with a parameterized neural surrogate $f_\theta : \mathcal{X} \to \mathcal{Y}$. Minimizing the regression error on static data is insufficient: For $f_\theta$ to serve as a stable loss function, it must provide "well-behaved" gradients that are informative across the input domain $\mathcal{D}$.

*Gradient propagation*: A critical requirement is nonvanishing input sensitivity. To avoid the "dead zones" associated with rectifying activations (e.g., ReLU, where $\nabla = 0$ for $x < 0$), we employ smooth, $C^\infty$ activation functions—specifically the Gaussian Error Linear Unit (GELU; Hendrycks and Gimpel, 2016). Furthermore, to facilitate gradient flow through deep architectures, we utilize residual connections $\mathbf{y} = \mathbf{x} + F(\mathbf{x})$, allowing error signals to propagate unimpeded through the identity path.

*Manifold smoothing*: Early in training, the upstream model may yield a distribution $P_g$ far from the data distribution $P_{\text{data}}$; if trained only on $P_{\text{data}}$'s support, the emulator can yield arbitrary gradients for $\mathbf{x} \sim P_g$. We hence inject Gaussian noise $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ during emulator training, effectively convolving the target functional with a smoothing kernel and improving robustness to off-manifold queries. (In practice, we found this step unnecessary for the OPERA dataset, given the Mix-Up augmentation described in Section 3.1, and set $\sigma^2 = 0$ in all reported experiments.)

*Lipschitz constraints*: Finally, we bound the gradient norm to prevent optimization instability. Unconstrained neural networks can learn arbitrarily sharp gradients, leading to exploding loss terms. We enforce stability by constraining the Lipschitz constant $K$ of the network mapping. While soft constraints like the gradient penalty (Gulrajani et al., 2017) are common in adversarial training, they do not guarantee bounds on unseen data. Instead, we enforce a hard constraint via spectral normalization (Miyato et al., 2018) on all weight matrices $W_l$. This ensures that the global Lipschitz constant is bounded by the product of the layer-wise spectral norms: $\|f_\theta(\mathbf{x}) - f_\theta(\mathbf{y})\|_2 \leq K \|\mathbf{x} - \mathbf{y}\|_2$, implying $\sup_{\mathbf{x} \in \mathcal{D}} \|\nabla_{\mathbf{x}} f_\theta(\mathbf{x})\|_2 \leq K$, almost everywhere in $\mathcal{D}$, where typically $K \approx 1$ and $\|.\|_2$ is the $L^2$ norm. While fixing $K$ before training limits the network's expressivity if the true functional varies rapidly

https://doi.org/10.1017/eds.2026.10048 Published online by Cambridge University Press

(Ducotterd et al., 2022), for the integral-geometric measures considered in Section 2.2, empirical results suggest that the stability benefits of a hard $K = 1$ constraint outweighs the theoretical expressivity loss.

### 2.2. Minkowski image loss: Injecting knowledge from integral geometry

#### 2.2.1. Integral-geometric characterization via Minkowski functionals

To capture the structural properties of physical fields beyond pixel-wise intensity, we adopt a framework based on the *Minkowski functionals* from integral geometry (Schneider and Weil, 2008). For a compact set $K \subset \mathbb{R}^2$ with a sufficiently smooth boundary (see Supplementary Appendix A), there exist three Minkowski functionals: the area, half the perimeter, and the Euler characteristic. These functionals are denoted $A_K$, $P_{K/2}$, and $\chi_K$, respectively. The Euler characteristic describes the connectivity of the set and is defined in 2D as $\chi_K = \mathrm{CC}_K - H_K$, where $\mathrm{CC}_K$ is the number of connected components and $H_K$ is the number of holes. The suitability of these measures for a loss function is rigorously grounded in Hadwiger's Characterization Theorem (Hadwiger, 1957), which establishes that any functional on $K$ satisfying rigid-motion invariance, continuity, and additivity can be expressed as a linear combination of these three measures. These mathematical properties map directly to critical advantages in neural network training: (i) *Rigid-motion invariance:* Decouples the loss from the exact spatial positioning of extreme events, mitigating the "double penalty" problem (Ebert, 2008) inherent in pixel-wise metrics where a slight displacement is penalized twice (as a miss and a false alarm). (ii) *Continuity:* Ensures gradient stability, as minor shape deformations result in commensurate, small changes in the loss value, unlike discontinuous metrics (e.g., pure object counting), which exhibit step functions.

#### 2.2.2. The γ-vector and Minkowski image loss formulation

A continuous physical field is transformed into $N$ binary images (excursion sets) by thresholding at $N$ intensity levels, $u^{(1)}, \ldots, u^{(N)}$. For each excursion set, we compute the area ($A$), perimeter ($P$), and the number of connected components (CC). The values of these functionals for the $i$-th threshold are denoted $A^{(i)}$, $P^{(i)}$, and $\mathrm{CC}^{(i)}$. We concatenate these statistics into a single nonnegative vector $\boldsymbol{\gamma}$ that provides a multilevel integral-geometric summary of the field:

$$\boldsymbol{\gamma} = \left[A^{(1)}, P^{(1)}, \mathrm{CC}^{(1)}, \ldots, A^{(N)}, P^{(N)}, \mathrm{CC}^{(N)}\right]^{\top} \in \mathbb{R}^{3N}. \tag{1}$$

The distance between two $\boldsymbol{\gamma}$-vectors serves as a proxy for the structural divergence between the predicted and target fields. Integrating this metric into training ensures the network minimizes geometric error alongside pixel-wise error. The reader is referred to Supplementary Appendix B.2 for further details regarding the computation of the $\gamma$-vector from data.

A significant challenge is the order-of-magnitude disparity between vector components; $A^{(i)}$ (pixel summation) can be orders of magnitude larger than $\mathrm{CC}^{(i)}$ (sparse count). Without normalization, gradients would be dominated by area terms, ignoring topological structure. To address this, we define the *Minkowski image loss* $\mathcal{M}$ using a log-transformed $L^1$ distance. This transformation effectively normalizes the scale differences, operating as a relative error metric that balances the contribution of large-magnitude features (area) and small-magnitude features (connectivity):

$$\mathcal{M}(\hat{\boldsymbol{\gamma}}, \boldsymbol{\gamma}) = \| \log(\hat{\boldsymbol{\gamma}} + \mathbf{1}) - \log(\boldsymbol{\gamma} + \mathbf{1}) \|_1 \tag{2}$$

where $\mathbf{1}$ is a vector of ones to ensure numerical stability and $\|.\|_1$ is the $L^1$ norm. This Minkowski image loss functions as a regularizer alongside standard pixel-wise metrics (e.g., MSE). While pixel loss enforces local intensity accuracy, $\mathcal{M}$ acts as a global structural constraint. Since the direct computation of $\boldsymbol{\gamma}$ involves non-differentiable operations (thresholding and counting) that break the computation graph, we employ the knowledge-guided differentiable surrogate described in Section 2.1. This surrogate allows the error signal $\nabla_{\hat{\gamma}}\mathcal{M}$ to be backpropagated through the emulator to update the trainable model.

https://doi.org/10.1017/eds.2026.10048 Published online by Cambridge University Press

### *2.3. Application to integral-geometric measures*

While analytical approximations directly relax discrete metrics, formulating closed-form solutions for complex topologies proves computationally restrictive and prone to vanishing gradients. As a robust alternative, we introduce a neural emulator as a learned differentiable surrogate, approximating the non-differentiable mapping $F : \mathcal{X} \to \mathbb{R}^{3N}$ from physical fields to $\boldsymbol{\gamma}$-vectors. Standard pixel-wise metrics (e.g., MSE) inherently smooth the multiscale spatial organization critical to precipitation modeling. By embedding this emulator into the loss function, we directly penalize geometric errors. We train the model to regress exact ground-truth descriptors (area, perimeter and connected components) generated via GUDHI (Maria et al., 2014). The bipartite architecture pairs a Lipschitz-regularized feature extractor with geometrically constrained heads to decode valid $\boldsymbol{\gamma}$-vector components.

#### *2.3.1. Analytical approximation of integral-geometric measures*

Our analytical approximation substitutes the non-differentiable Heaviside step function with a temperature-controlled sigmoid, mapping the continuous physical field into a pseudo-probability space $P \in (0,1)$ to enable differentiable tensor operations. We calculate the area as the weighted sum of $P$ and derive perimeter from finite-difference approximations of spatial derivatives. To compute the Euler characteristic (a proxy for connected components), we evaluate the alternating sum of grid vertices, edges, and faces, replacing discrete boolean logic with the Gödel t-norm (min operator) to ensure continuous gradient flow. To stabilize these topological metrics against minor fluctuations, we preprocess the input tensor using differentiable morphological filters and a local persistence threshold mask, ultimately aggregating and scaling the measures via a symmetric logarithm.

#### *2.3.2. Emulation of integral-geometric measures via knowledge-guided learning*

***Feature extractor with Lip-CNN.*** To map the physical field $\mathbf{x} \in \mathbb{R}^{H \times W}$ to the latent geometric manifold $\mathbf{z}$, we employ a CNN encoder designed to respect the mathematical constraints of the $\boldsymbol{\gamma}$-vector. The input field is normalized to $[0,1]$ via a fixed scaling factor. The backbone processes the field through four hierarchical stages (Figure 2, left), with spectral normalization (Miyato et al., 2018) applied to all convolutional kernels except the first layer, which is left unconstrained to allow unrestricted input scaling. Deep feature extraction is handled by "robust blocks" containing residual connections and group normalization, ensuring invariance to batch statistics. We use global sum pooling rather than average pooling, since area and connected components are extensive quantities whose additivity must be preserved as the latent representation scales with domain size.

***Knowledge-guided geometric constraints.*** Unlike standard multi-task regression, where outputs are independent, our architecture couples the output heads to strictly enforce geometric consistency.

*Area head (monotonicity via integration).* Directly regressing the area metric $A^{(i)}$ for each threshold fails to guarantee the monotonicity property $A^{(i)} \geq A^{(i+1)}$ (as threshold increases, excursion set area decreases). To resolve this, we use the fact that the area function is effectively the cumulative distribution function of the field intensities. The network predicts a discrete probability density using a softmax layer and a scalar total area $A_{\text{total}}$. The final area vector $\widehat{\mathbf{A}}$ is computed via the cumulative sum of these probabilities. This inductive bias ensures that the predicted area is strictly monotonic by construction.

*Perimeter head (isoperimetric inequality).* We strictly condition the perimeter on the area by predicting a dimensionless "roughness" coefficient $\widehat{\mathbf{r}} \geq 0$ (via a Softplus activation) rather than the absolute perimeter. The perimeter is then reconstructed as $\widehat{\mathbf{P}} = \sqrt{4\pi\widehat{\mathbf{A}}} \odot (\mathbf{1} + \widehat{\mathbf{r}})$. This hard-codes the isoperimetric inequality ($P^2 \geq 4\pi A$) into the forward pass. It prevents the model from generating a perimeter smaller than that of a perfect circle with equivalent area, ensuring 0% geometrical violations.

*Topological head (positivity).* The number of connected components CC is a counting measure. To reflect this, the output is rectified using Softplus ($\log(1 + e^x)$), ensuring $\widehat{\mathrm{CC}} \in \mathbb{R}_+$ while maintaining differentiable nonzero gradients, unlike a hard ReLU clip, which may lead to "dead neurons".

https://doi.org/10.1017/eds.2026.10048 Published online by Cambridge University Press

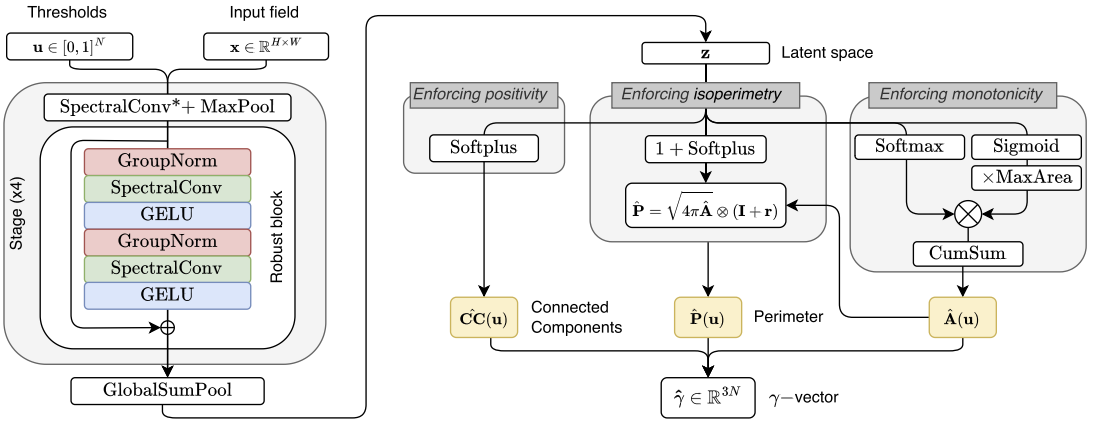



***Figure 2.*** *(Left) Lipschitz-bound feature extractor: The backbone acts as a hierarchical encoder. Stability is enforced via spectral normalization and residual connections. We use global sum pooling to preserve the extensivity of geometric features. (Right) Geometric constraint heads: The area $\widehat{A}(u)$ is constructed via integration of a predicted probability density to enforce monotonicity. The perimeter $\widehat{P}(u)$ is conditioned on the area via the isoperimetric inequality, preventing geometric principles' violations.*

***Training setup.*** Motivated by the target data's statistical properties, we optimize $\theta$ using the Minkowski image loss: $\mathcal{L}_{\text{train}} = \mathcal{M}(\hat{\boldsymbol{\gamma}}_\theta(\mathbf{x}), \boldsymbol{\gamma}_{\text{true}})$. Logarithmic scaling prevents large disparities between area and connected components, while the $L^1$ norm ensures stability against outliers from complex storm structures.

## 3. Application: Observed instantaneous surface precipitation fields

Emulating Minkowski functionals is particularly relevant for precipitation, which—unlike smooth fields—exhibits high intermittency, non-Gaussian statistics, and complex multiscale geometries. Standard pixel-wise losses often fail to capture these morphological properties, leading to structurally amorphous ("blurred") predictions. By applying our Minkowski image loss to radar-based precipitation estimates, we aim to encourage realistic structures in generative tasks, specifically km-scale downscaling.

### *3.1. Dataset and preprocessing*

We validate our approach using the pan-European EUMETNET radar composites provided by the European Operational Program for Exchange of Weather Radar Information (OPERA) (Saltikoff et al., 2019). We utilize the instantaneous surface rain rate product, leveraging high-definition streams (ODISSEY/NIMBUS) with 15-minute updates. Spanning August 1, 2023, to October 30, 2024, this archive yields >1M patches of $128 \times 128$ pixels ($256 \times 256$ km) across diverse regimes, from localized summer convection to large-scale winter synoptic storms. We partition the dataset chronologically to preclude temporal autocorrelation: data from August 2023 through June 2024 is utilized for training, July 2024 for validation, and August through October 2024 is reserved for independent testing. We apply a drizzle threshold $\delta = 0.1\,\text{mm}\,\text{h}^{-1}$; intensity values below $\delta$ are zeroed out to remove sensor noise. For training, data undergoes a log-linear transformation to compress the dynamic range: $x_{\text{norm}} = \log(1+x)/S$, where $S$ is the global maximum of the log-transformed training set. To enforce robustness, we apply an offline "Mix-Up" strategy in physical space. We generate synthetic convex combinations of ground truth fields and noise-injected interpolated fields to ensure off-manifold robustness: $\mathbf{x}_{\text{mix}} = \lambda \mathbf{x}_{\text{real}} + (1-\lambda)\left(\mathbf{x}_{\text{interp}} + \boldsymbol{\epsilon}\right)$ where $\lambda \sim \text{Beta}(\alpha, \alpha)$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$.

https://doi.org/10.1017/eds.2026.10048 Published online by Cambridge University Press

### *3.2. Evaluated architectures and baselines*

#### *3.2.1. Statistical baseline: bicubic interpolation*

As a naive deterministic baseline that preserves total domain mass without introducing high-frequency artifacts, we employ bicubic interpolation. This non-parametric approach acts as a lower bound for structural fidelity, isolating the performance gains achievable purely through the spatial redistribution of the low-resolution input.

#### *3.2.2. Analytical approximation*

To evaluate the efficacy of the proposed continuous relaxation, we integrate the analytical surrogate (detailed in Section 2.3.1) into the optimization loop. This model operates on identical inputs but replaces the learned emulation heads with differentiable temperature-controlled sigmoid and Gödel t-norm operations.

#### *3.2.3. Surrogate CNN emulators ablation study*

To quantify the impact of embedding domain knowledge, we perform an ablation study comparing three distinct architectures. This hierarchy disentangles the benefits of gradient stabilization (Lipschitz continuity) from the benefits of hard geometric constraints.

*Unconstrained CNN*: A standard convolutional neural network employing ReLU activations, batch normalization, and global average pooling. This represents a naive deep learning baseline. It treats the components of the $\boldsymbol{\gamma}$-vector $(A, P, \mathrm{CC})$ as independent regression targets.

*Unconstrained Lip-CNNs* use our specialized backbone (spectral normalization, group normalization, residual layers, and global sum pooling) to ensure gradient stability but retain standard, independent regression heads. This isolates the benefit of Lipschitz regularization but lacks structural constraints; it remains capable of predicting geometrically impossible states (e.g., non-monotonic area curves).

*Constrained Lip-CNN*: The full architecture described in Section 2.3.2. This model combines the Lipschitz-regularized backbone with the geometrically constrained output heads, structurally enforcing monotonicity, nonnegativity, and isoperimetric consistency.

### *3.3. Verification via diagnostic input optimization*

We assess the utility of the emulator's gradients using a diagnostic feature inversion test. This procedure involves freezing the emulator weights $\theta$ and iteratively updating a noise vector $\mathbf{x}$ to minimize the distance to a target metric value $\mathbf{y}_{\text{target}}$: $\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t - \alpha \nabla_{\mathbf{x}} \|f_\theta(\mathbf{x}_t) - \mathbf{y}_{\text{target}}\|_2$. This acts as a unit test for the gradient field. If $f_\theta$ has learned structurally robust features, the optimization recovers an input $\mathbf{x}^*$ consistent with domain physics (e.g., connected precipitation blobs). If the optimization produces high-frequency adversarial noise, it indicates the gradients are ill-conditioned, requiring stricter regularization.

### *3.4. Application of Minkowski image loss to super-resolution of instantaneous surface rain rate*

To evaluate our framework, we formulate a perfect-model super-resolution task for instantaneous surface rain rates from the OPERA dataset, emulating a transition from coarse global circulation models (~25 km) to high-resolution radar (2 km). We generate low-resolution inputs via 12.5 × block-averaging, concatenating them with a normalized digital elevation model to condition the downscaling on orographic forcing. Target fields undergo a normalized log-space transformation to stabilize optimization across highly skewed precipitation distributions. We benchmark a deterministic log-space residual U-Net against a conditional denoising diffusion probabilistic model (DDPM), evaluated exclusively via deterministic DDIM sampling to ensure rigorous comparison. We test whether augmenting the mean squared error (MSE) with the Minkowski image loss (via analytical approximation or neural emulation) imparts generative-level structural realism to deterministic architectures, thereby circumventing the substantial computational sampling costs of diffusion models. Supplementary Appendix E details the implementation.

https://doi.org/10.1017/eds.2026.10048 Published online by Cambridge University Press

***Table 1.*** *Comparison of analytical approximation and emulation*

| Architecture | $\mathcal{M}$ (↓) | $R^2$ (↑) | $R^2_A$ | $R^2_P$ | $R^2_{CC}$ | $v_{Iso}$ (%) |
|---|---|---|---|---|---|---|
| PCR | 5.36 | 0.35 | 0.46 | 0.39 | 0.25 | 27.0 |
| Analytical approx. | **1.03** | 0.67 | **0.99** | 0.92 | 0.14 | **0.3** |
| CNN (unconstrained) | 5.77 ± 0.02 | 0.95 ± 0.02 | 0.94 ± 0.02 | 0.94 ± 0.01 | 0.96 ± 0.02 | 7.8 ± 0.3 |
| Lip-CNN (unconstrained) | 5.74 ± 0.01 | 0.88 ± 0.03 | 0.79 ± 0.03 | 0.91 ± 0.01 | 0.95 ± 0.03 | 5.8 ± 0.2 |
| **Lip-CNN (constrained)** | 5.73 ± 0.01 | **0.99** ± 0.01 | **0.99** ± 0.01 | **0.99** ± 0.01 | **0.97** ± 0.02 | **0.00** |

*Note.* Bold values represent the best significant results (column wise). Performance is measured via the Minkowski image loss $\mathcal{M}$ (lower is better), the coefficient of determination $R^2$ for the full vector and individual components (area [A], perimeter [P], connected components [CC]), and the rate of isoperimetric violations ($v_{Iso}$).

## 4. Results

### *4.1. Accuracy and geometric fidelity of analytical approximation and emulation*

To establish performance bounds, we evaluate a linear principal component regression (PCR) baseline. PCR demonstrates poor predictive capability (overall $R^2 = 0.35$) and a high isoperimetric violation rate ($v_{Iso} = 27.0\%$), confirming the highly nonlinear mapping from physical fields to integral-geometric measures (Table 1). Conversely, the analytical approximation achieves the lowest Minkowski image loss ($\mathcal{M} = 1.03$) and near-zero isoperimetric violations (0.3%). While perfectly capturing the area ($R^2_A = 0.99$) and approximating perimeter well ($R^2_P = 0.92$), its overall $R^2$ (0.67) is hindered by a near-complete failure to predict connected components ($R^2_{CC} = 0.14$; see Supplementary Appendix C).

While all neural architectures converged, the constrained Lip-CNN achieves the best performance across all metrics, definitively eliminating isoperimetric violations ($v_{Iso} = 0\%$). The unconstrained Lip-CNN underperforms the vanilla CNN baseline; while $\mathcal{M}$ remains comparable, $R^2$ scores degrade significantly (e.g., area $0.94 \rightarrow 0.79$). Imposing spectral normalization bounds the Lipschitz constant, limiting representation power. Without geometric constraint heads, this "stiffened" backbone struggles to fit high-variance data. Pairing the Lipschitz backbone with geometric constraints (Lip-CNN Constr.) recovers and surpasses baseline performance. Architectural inductive biases (monotonicity and isoperimetry) effectively reduce hypothesis space complexity, acting as a scaffold that allows the stable backbone to converge to a physically valid solution despite reduced capacity. Supplementary Appendix D.1 provides further evaluation and examples of the emulators' capabilities.

### *4.2. Diagnostic inversion and gradient stability*

Figure 3 demonstrates that both the *vanilla CNN* and *constrained Lip-CNN* reconstruct coherent precipitation structures from geometric signatures, confirming Minkowski functionals encode sufficient morphological information to guide optimization from Gaussian noise to a localized storm object. However, textural quality diverges: the *vanilla CNN* produces disjointed, "blob-like" structures with sharp, artificial boundaries, whereas Lipschitz-regularized models generate continuous intensity gradients that naturally reproduce the smooth decay from convective cores to stratiform peripheries. In Supplementary Appendices D.2 and D.3, we report further evaluation of the constrained Lip-CNN emulator's gradient quality.

### *4.3. Deterministic super-resolution*

Table 2 details super-resolution performance metrics. The bicubic interpolation baseline exhibits seemingly strong spectral fidelity (RAPS = 1.04). However, this metric is an artifact of spatial smoothing; interpolation merely redistributes low-frequency kinetic energy without generating physical high-frequency features. This failure is captured by its high topological error ($\mathcal{M} = 3.44$) and blocky visual structure (Figure 4).

https://doi.org/10.1017/eds.2026.10048 Published online by Cambridge University Press

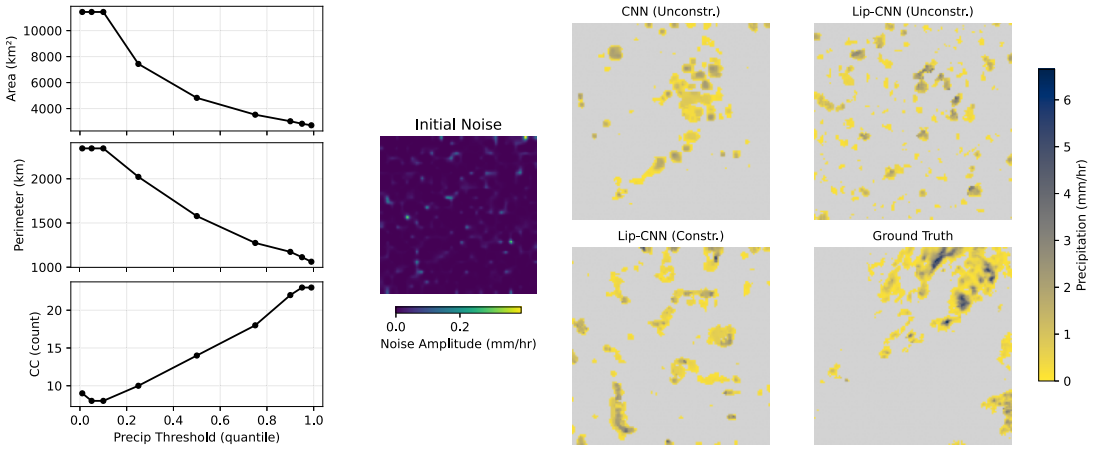



***Figure 3.*** *Feature inversion results. Qualitative comparison of precipitation fields reconstructed by inverting the target Minkowski vector* $\boldsymbol{\gamma}_{\mathrm{gt}}$ *for a test sample (left). All architectures recover the storm's magnitude, while only constrained models generate coherent, smooth intensity gradients characteristic of convective cells (bottom left).*

***Table 2.*** *Quantitative comparison of super-resolution deterministic architectures*

| Architecture | $\mathcal{M}$ (↓) | *MAE* (↓) | SAL | RAPS (↓) |
|---|---|---|---|---|
| Interp. | 3.44 | 0.06 | 0.37, **–0.01**, 0.06 | 1.04 |
| UNet | 3.21 | **0.05** | 0.32, 0.09, 0.15 | 2.88 |
| UNet + $\mathcal{M}_{\mathrm{Ana}}$ | 2.74 | **0.05** | 0.41, –0.20, **0.03** | 1.37 |
| UNet + $\mathcal{M}_{\mathrm{Emu,Lip\text{-}CNN(Constr.)}}$ | 3.25 | **0.05** | 0.31, –0.29, **0.07** | 1.78 |
| DDIM | 2.96 | 0.09 | **–0.08**, 0.26, 0.17 | **0.97** |

*Note.* Bold values represent the best significant results (column wise). Performance is measured via the Minkowski image loss ($\mathcal{M}$), the mean absolute error (MAE), the structure-amplitude-location (SAL) metric, and the radially averaged power spectral density (RAPS) error.

Comparing the baseline UNet with topologically constrained architectures illustrates deterministic optimization challenges. The unconstrained UNet minimizes spatial distortion (MAE = 0.05) but actively dampens spatial variance. Figure 4 confirms this: the UNet predicts an amorphous precipitation shield, failing to resolve localized convective cores, yielding the highest spectral error (RAPS = 2.88). Augmenting the UNet with the analytical Minkowski loss ($\mathcal{M}_{\mathrm{Ana}}$) quantitatively improves RAPS (1.37) and topological error ($\mathcal{M} = 2.74$) without compromising MAE (0.05). However, visually, both analytical and emulator-constrained ($\mathcal{M}_{\mathrm{Emu,Lip-CNN(Constr.)}}$) outputs only marginally improve upon the baseline. The fields remain artificially smooth, indicating that the current surrogate loss implementation insufficiently enforces morphological realism.

Benchmarked against the stochastic DDIM, a distinct trade-off emerges. The DDIM achieves the most realistic energy spectrum (RAPS = 0.97) and structure amplitude ($S = -0.08$), successfully synthesizing high-frequency speckle and intense peaks. However, it incurs higher pixel-wise error (MAE = 0.09). This elevated MAE manifests the double-penalty effect typical of stochastic downscaling: synthesized high-frequency details, while physically realistic in texture, misalign with ground-truth spatial coordinates. Although Minkowski-constrained UNets aim to bridge this gap by preserving exact spatial alignment, their inability to visually reproduce structural fidelity highlights a limitation. Future work must refine the weighting and optimization of integral-geometric constraints to match diffusion model quality without the associated computational sampling costs.

https://doi.org/10.1017/eds.2026.10048 Published online by Cambridge University Press

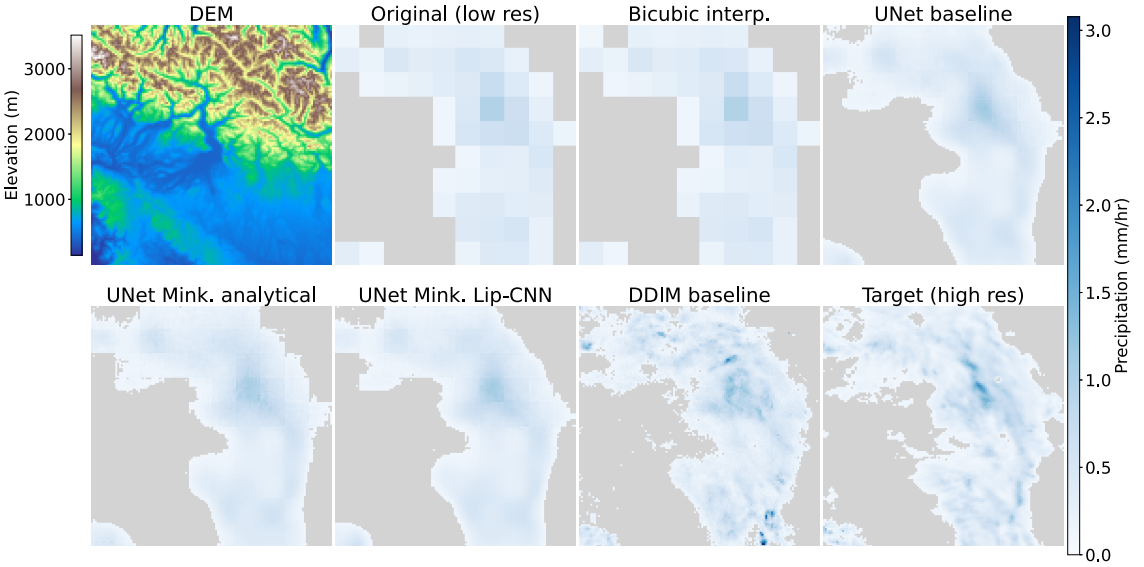



***Figure 4.*** *Structural fidelity in precipitation downscaling. The unconstrained UNet produces an amorphous shield, missing localized convective peaks. Integral-geometric constraints (analytical and Lip-CNN) yield only marginal improvement. The stochastic DDIM baseline recovers textural realism and multiscale variance, confirming the limitations of deterministic optimization.*

## 5. Discussion and conclusions

This work establishes a rigorous framework for integrating non-differentiable integral-geometric domain knowledge into gradient-based training loops via learned surrogates, and demonstrates it on the EUMETNET OPERA radar dataset. We successfully trained a Lipschitz-regularized neural emulator to accurately map physical precipitation fields to their corresponding Minkowski functionals, structurally enforcing critical inductive biases such as monotonicity and the isoperimetric inequality. While incorporating differentiable surrogates into deep learning is a rapidly evolving frontier (Rehmann et al., 2025; Nguyen et al., 2026), our empirical results highlight fundamental challenges in transitioning from accurate metric emulation to effective generator optimization.

The constrained Lip-CNN achieved a 0% isoperimetric violation rate, definitively eliminating the physically impossible geometries produced by unconstrained baselines. However, our super-resolution experiments reveal a fundamental trade-off: hard Lipschitz constraints, while essential for preventing optimization collapse, over-smooth the latent representation. The backpropagated error signals consequently lack the localized precision required to steer fine-scale spatial features; the surrogate recognizes macroscopic geometry but cannot provide the sharp, pixel-level corrective gradients needed to define distinct convective boundaries. Emerging libraries for generalized automatic differentiation of discrete operations (Paulus et al., 2026) offer complementary analytical tools, but sequential element-wise relaxations for complex topological metrics remain prone to gradient attenuation—reinforcing the case for a top-down learned surrogate once the gradient transfer bottleneck is resolved.

Ultimately, this framework addresses a ubiquitous requirement in Earth system science: the need to model fields where geometrical accuracy is encoded in coherent structures rather than independent grid points. For applications like statistical downscaling, where the spatial organization of extreme events directly dictates localized climate impacts, our findings underscore that formulating mathematically consistent loss functions is only half the solution. Future work must couple these rigorous topological

https://doi.org/10.1017/eds.2026.10048 Published online by Cambridge University Press

constraints with stochastic generative architectures—such as denoising diffusion probabilistic models—to combine the strict integral-geometric guarantees of the Minkowski image loss with the textural synthesis capabilities required for full morphological realism.

**Supplementary material.** The supplementary material for this article can be found at http://doi.org/10.1017/eds.2026.10048.

**Data availability statement.** The OPERA radar data used in this study are openly available under the CC BY license. Data can be accessed via the EUMETNET Open Radar Data (ORD) API. Access requires IP whitelisting; requests should be directed to the OPERA support team (support.opera@eumetnet.eu). Comprehensive documentation for data discovery and retrieval is available at https://github.com/EUMETNET/openradardata-documentation. The code implementing the full pipeline, including the Minkowski functional emulator, the analytical approximation, and the super-resolution training and evaluation framework, is publicly available through Zenodo and GitHub at https://doi.org/10.5281/zenodo.20448779.

**Acknowledgments.** The authors would like to thank L. Räss for technical assistance. The authors would also like to acknowledge D. Nerini, D. Domeisen, V. Chavez, L. Moret, and O. Miralles for their contributions to the initial conceptualization of this work.

**Author contribution.** Conceptualization: F.Q., R.C. and T.B. Methodology: F.Q. and R.C. Software: F.Q. Formal analysis: F.Q. and R.C. Investigation: F.Q. Data curation: F.Q. Writing—original draft: F.Q. Writing—review and editing: F.Q. and T.B. Visualization: F.Q. Supervision: T.B. All authors approved the final submitted draft.

**Funding statements.** F.Q. and T.B. acknowledge support from the Swiss National Science Foundation (SNSF) under Grant No. 10001754 (“RobustSR” project).

**Competing interest.** The authors declare none.

**Ethical statement.** The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

**Provenance statement.** This article is part of the Climate Informatics 2026 proceedings and was accepted in Environmental Data Science based on the Climate Informatics peer review process.

## References

**Ducotterd S**, **Goujon A**, **Bohra P**, **Campos J and Unser M** (2022) Improving lipschitz-constrained neural networks by learning activation functions. arXiv preprint. *arXiv:2210.16222*.

**Ebert EE** (2008) Fuzzy verification of high-resolution gridded forecasts: A review and proposed framework. *Meteorological Applications 15*(1), 51–64.

**Gelbrecht M**, **White A**, **Bathiany S and Boers N** (2023) Differentiable programming for earth system modeling. *Geoscientific Model Development 16*(11), 3123–3135.

**Gouk H**, **Frank E**, **Pfahringer B and Cree MJ** (2021) Regularisation of neural networks of lipschitz continuity. *International Journal of Computer Vision 129*, 444–457.

**Gulrajani I**, **Ahmed F**, **Arjovsky M**, **Dumoulin V and Courville A** (2017) Improved training of wasserstein gans. *Advances in Neural Information Processing Systems 30*, pp. 4 and 5.

**Hadwiger H** (1957) *Vorlesungen über Inhalt, Oberfläche Und Isoperimetrie*. Springer.

**Hendrycks D and Gimpel K** (2016) Gaussian error linear units (gelus). arXiv preprint. *arXiv:1606.08415*.

**Hu X**, **Li F**, **Samaras D and Chen C** (2019) Topology-preserving deep image segmentation. *Advances in Neural Information Processing Systems 32*, pp. 4–6.

**Jang E**, **Gu S and Poole B** (2017) *Categorical Reparameterization with Gumbel-Softmax.* https://arxiv.org/abs/1611.01144 (accessed November 2025-March 2026).

**Karpatne A**, **Atluri G**, **Faghmous JH**, **Kumar V**, et al. (2017) Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering 29*(10), 2318–2331.

**Maria C**, **Boissonnat J-D**, **Glisse M and Yvinec M** (2014) *The Gudhi Project: Simplicial Topology for Data Analysis*. GUDHI Editorial Board.

**Miyato T**, **Kataoka T**, **Koyama M and Yoshida Y** (2018) Spectral normalization for generative adversarial networks. In *6th International Conference on Learning Representations (ICLR 2018), Vancouver, BC.*.

**Nguyen XT**, **Chen L**, **Dorigo T**, **Gauger NR**, **Vischia P**, **Nardi F**, **Awais M**, **Hanif H**, **Abbas S and Kapoor R** (2026) *Differentiable surrogate for detector simulation and design with diffusion models.* https://arxiv.org/abs/2601.07859v1 (accessed 9 January 2026)

**Patel Y**, **Hodan T and Matas J** (2020) Learning surrogates via deep embedding. In *European Conference on Computer Vision*. Springer, pp. 205–221. arXiv:2603.08824, 2026.

**Paulus A**, **Geist AR**, **Musil V**, **Hoffmann S**, **Beker O, and Martius G** (2026). *Softjax & softtorch: Empowering automatic differentiation libraries with informative gradients.* https://arxiv.org/abs/2603.08824 (accessed November 2025-March 2026)

**Rehmann A**, **Black N**, **Bjorgaard J**, **Angioi A**, **Paleyes A**, **Heim N**, **Häfner D and Lavin A** (2025) *Surrogate-based differentiable pipeline for shape optimization.* https://arxiv.org/abs/2511.10761 (accessed November 2025-March 2026)

**Saltikoff E**, **Haase G**, **Delobbe L**, **Gaussiat N**, **Martet M**, **Idziorek D**, **Leijnse H**, **Novák P**, **Lukach M and Stephan K** (2019) Opera: The radar project. *Atmosphere 10*(6), 320.

**Schneider R and Weil W** (2008) *Stochastic and Integral Geometry.* Berlin, Heidelberg: Springer.

**Cite this article:** Quarenghi F, Cotsakis R and Beucler T (2026). Emulating non-differentiable metrics via knowledge-guided learning: Introducing the Minkowski image loss. *Environmental Data Science*, 5: e17. doi:10.1017/eds.2026.10048

https://doi.org/10.1017/eds.2026.10048 Published online by Cambridge University Press