IMPERIAL

A Systematic Comparison of Machine Learning Models for Reconstructing Atmospheric Fields from Sparse Observations

Masters Thesis Presentation

Supervisor: Dr. Sibo Cheng

Author: Vincent Lefeuve

Introduction
Motivation

Increasing frequency of climate disasters

- Floods, storms, droughts, extreme temperatures, wildfires
- Sharp rise since the 1970s

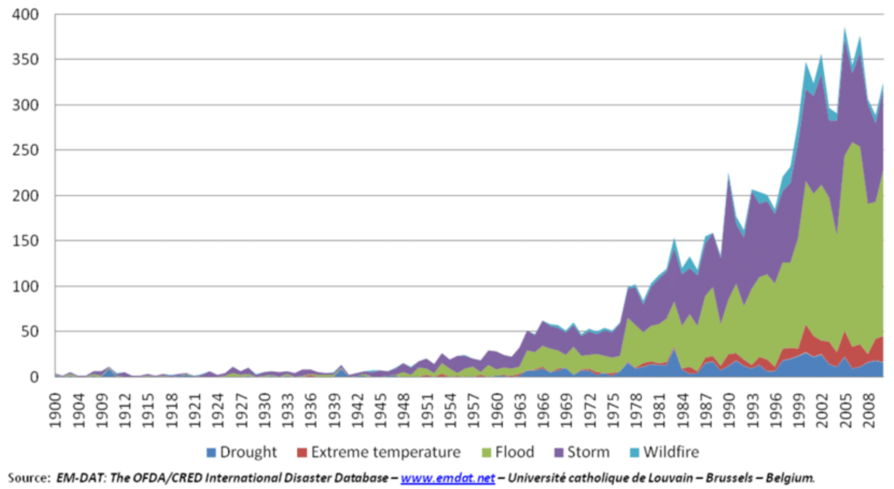


Fig. 1. Numbers of extreme weather events globally, by year (Simpson and Burpee 2014)

Reliance on the Global Observing System

- Satellites, aircraft, ships, buoys, radars, automatic stations
- Data remains sparse and unevenly distributed
- Need for improved reconstruction methods

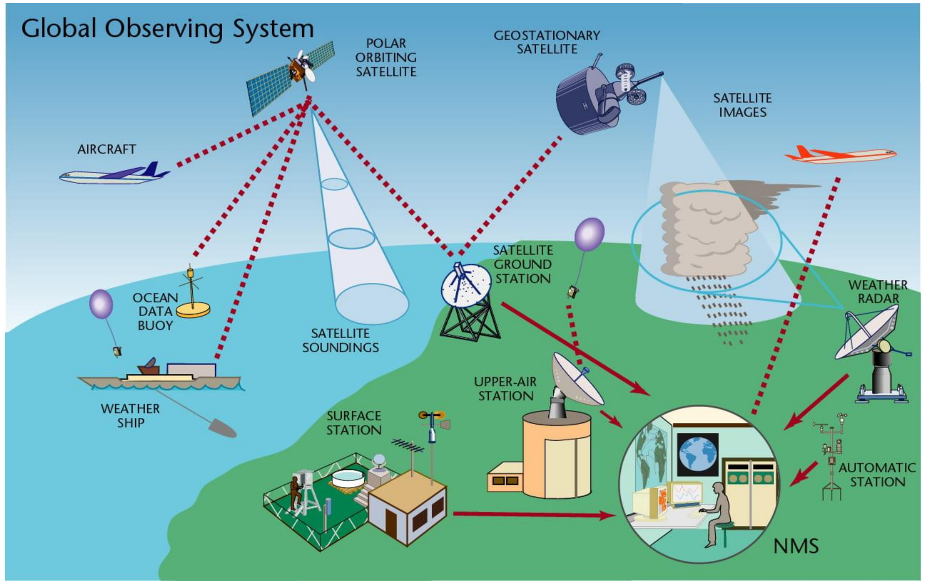


Fig. 2. Observation components of the global observing system (WMO, 2020)

Motivation
Current Methods

- Current practice: **Data assimilation** (accurate but costly, multiple steps)
- Alternatives: **Kriging** ($O(n^3)$, expensive) or **Cubic interpolation** (fast but inaccurate)
- Both only blend observations, limited physical insight
- Yet, we now have **massive observational & reanalysis datasets**

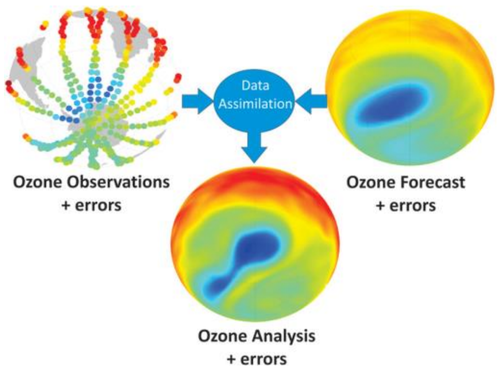


Fig. 3. Schematic of how data assimilation adds value to observational and model information (Lahoz et. Al, 2014)

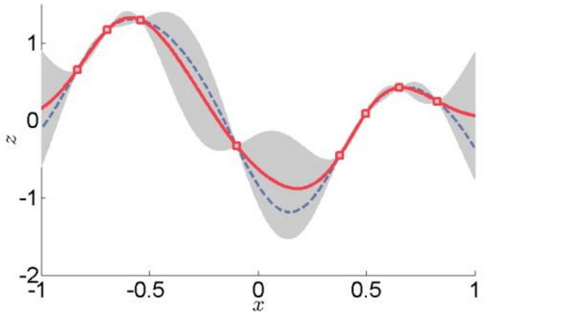


Fig. 4. Example of one-dimensional data interpolation by kriging (Wikipedia, 2016)

Machine Learning could be a scalable and powerful new approach

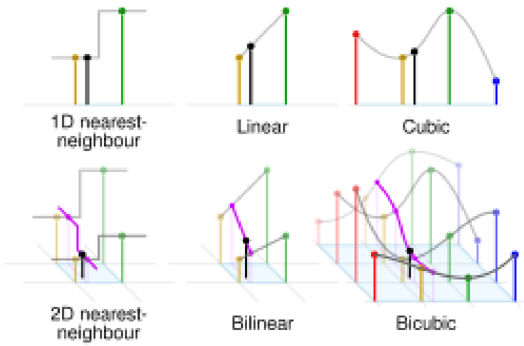


Fig. 5. Comparison of Bicubic interpolation with some 1- and 2-dimensional interpolations. (Wikipedia, 2020)

Motivation
Machine Learning

- **Current ML approaches are fragmented**
  - Deterministic models (e.g. U-Nets, trained with MSE)
  - Probabilistic models (e.g. diffusion models like DDPM)

- **Lack of systematic evaluation**
  - No unified benchmark for field reconstruction
  - Unlike other domains (e.g. ImageNet for image classification, GLUE for NLP, ClimateNet for extreme weather detection)

- **Need for this study**
  - Provide comparative analysis of deterministic vs. generative ML methods
  - Establish benchmarks for weather field reconstruction

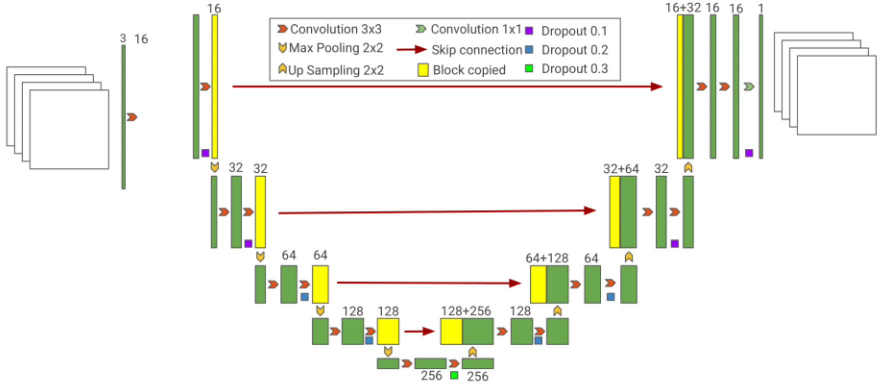


Fig. 6. Example of a U-Net Architecture (T. Alves, 2020)

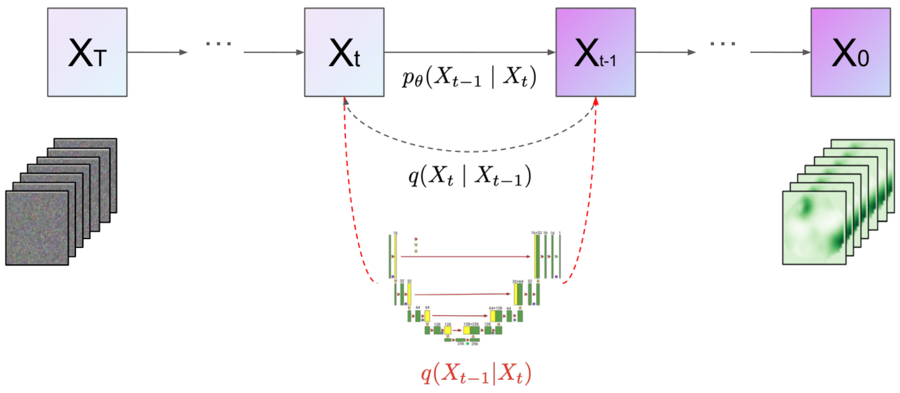
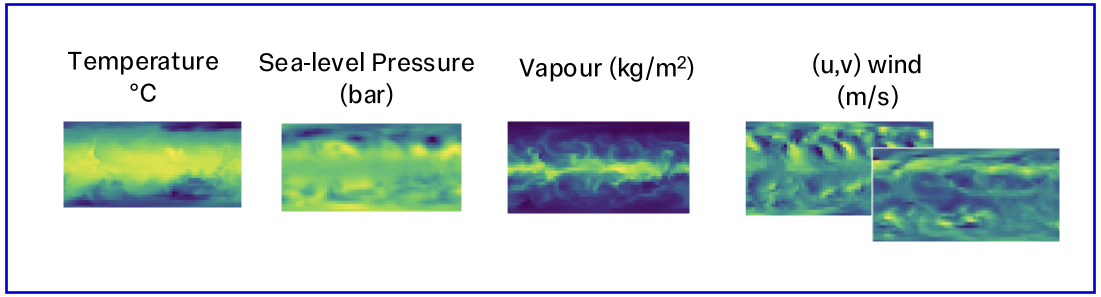


Fig. 7. Schematic of a DDPM

Data Preparation
Dataset

WeatherBench 2



Random selection of 10% of the grid points
+
Train/Test split

Dataset
- Weatherbench2
- Variables:
  - Temperature
  - Sea-level pressure
  - Total Column Water Vapour
  - Wind

Subsampling Strategy
- Random selection of K% pixels
- Selected pixels = observations
- Remaining pixels = ignored

Methodology
Voronoi Tessellation

Most of this work uses Voronoi Tessellation as input to the models

- Transform sparse sensor measurements into a structured image
- Construct **Voronoi cells** around each observation point
- Each cell assigned the value of its nearest observation
- Produces a continuous input field for the neural network
- Enables deep learning–based field reconstruction from irregular data

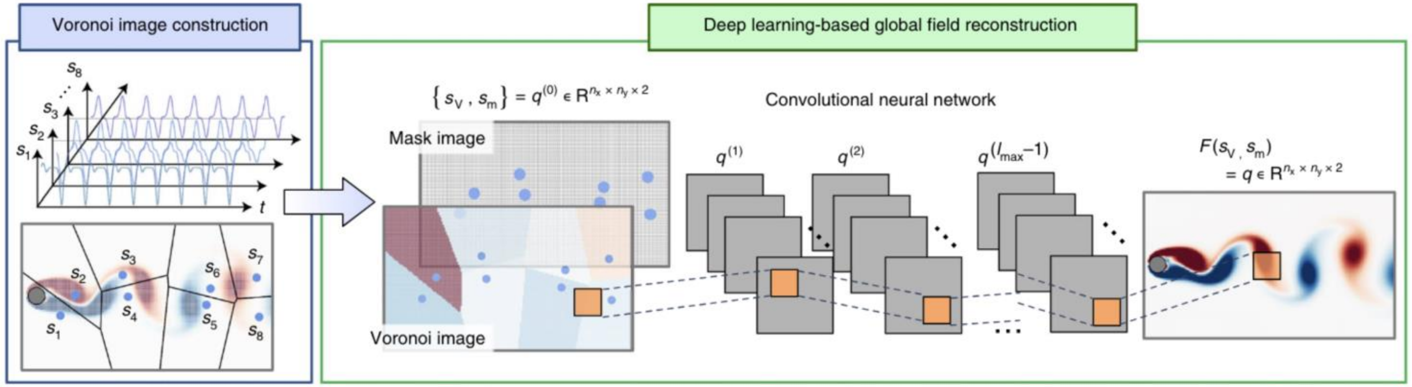


Fig. 8. Voronoi tessellation-assisted global data recovery from discrete sensor locations (Fukami et. Al. 2021)

$$s_{c,i} = \{x \in \Omega \mid d(x, \mathrm{Pos}(y_{c,i})) \leq d(x - \mathrm{Pos}(y_{c,j})), \forall j \neq i\}, \quad \text{with } s_{c,i}(x) = y_{c,i}, \ \forall x \in s_{c,i}.$$

Methodology

VT-UNet

Architecture

- Encoder–decoder with conv. down/upsampling.
- Skip connections → preserve scales & fine details.

I/O

- Input: Sparse fields (Voronoï-tessellated) + sensor locs (6 ch).
- Output: Reconstructed fields (5 ch).

Training

- MSE loss.

Advantages

- Learns multi-scale features.
- Captures global + local context.
- Lightweight, proven baseline.

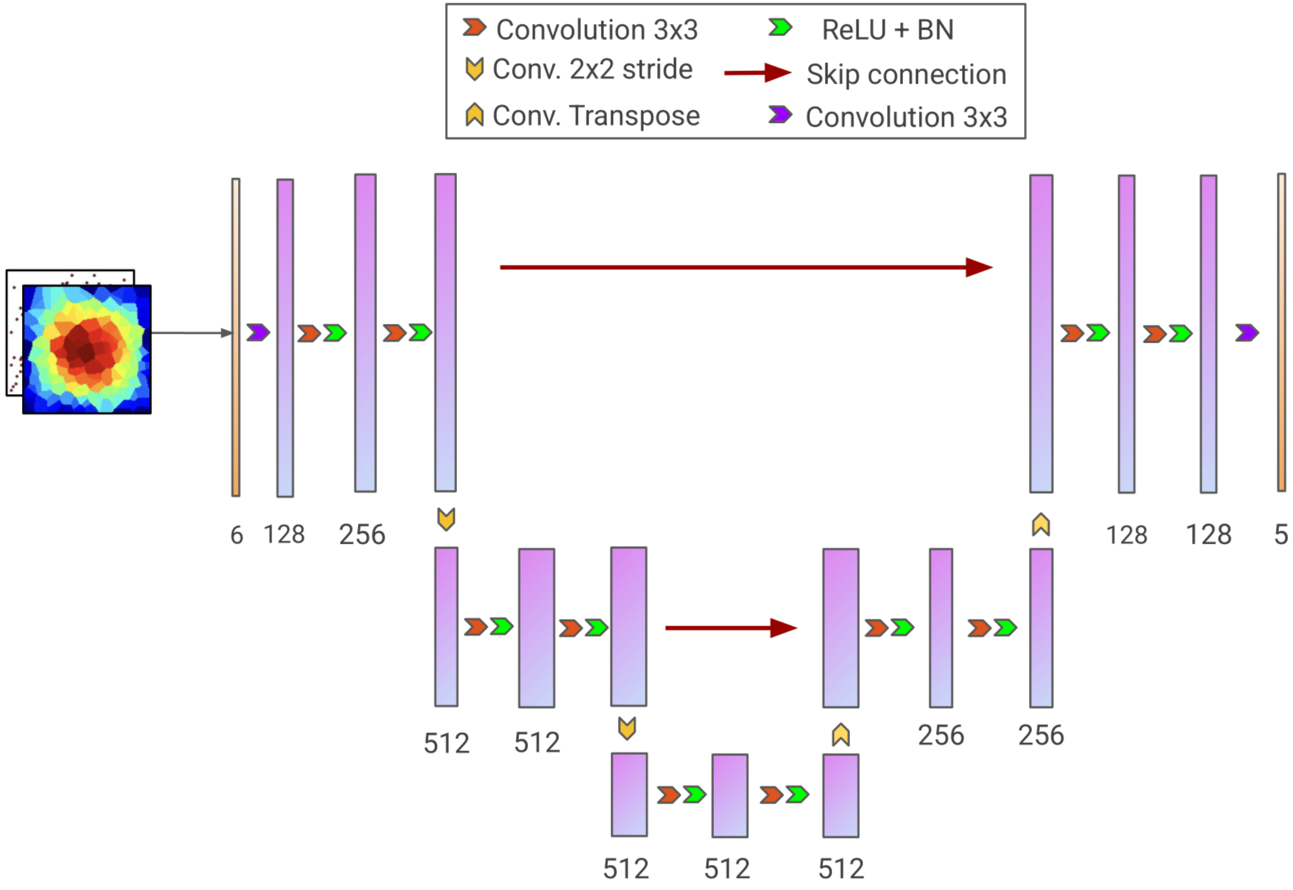


Fig. 9. Architecture of the U-Net Model

Methodology
VT-ResNet

Description
- Residual learning with skip connections → deeper, more stable training.

Advantages
- Captures local + global dependencies.
- More accurate than plain CNNs.

Architecture
- Input: Sparse fields (Voronoï-tessellated) + sensor locs (6 ch).
- Core: 6 × ResBlocks.
- Exit: Conv → output field.
- Trained with MSE loss.

Limitations
- Limited receptive field → weak on long-range deps (earth curvature).
- Deterministic

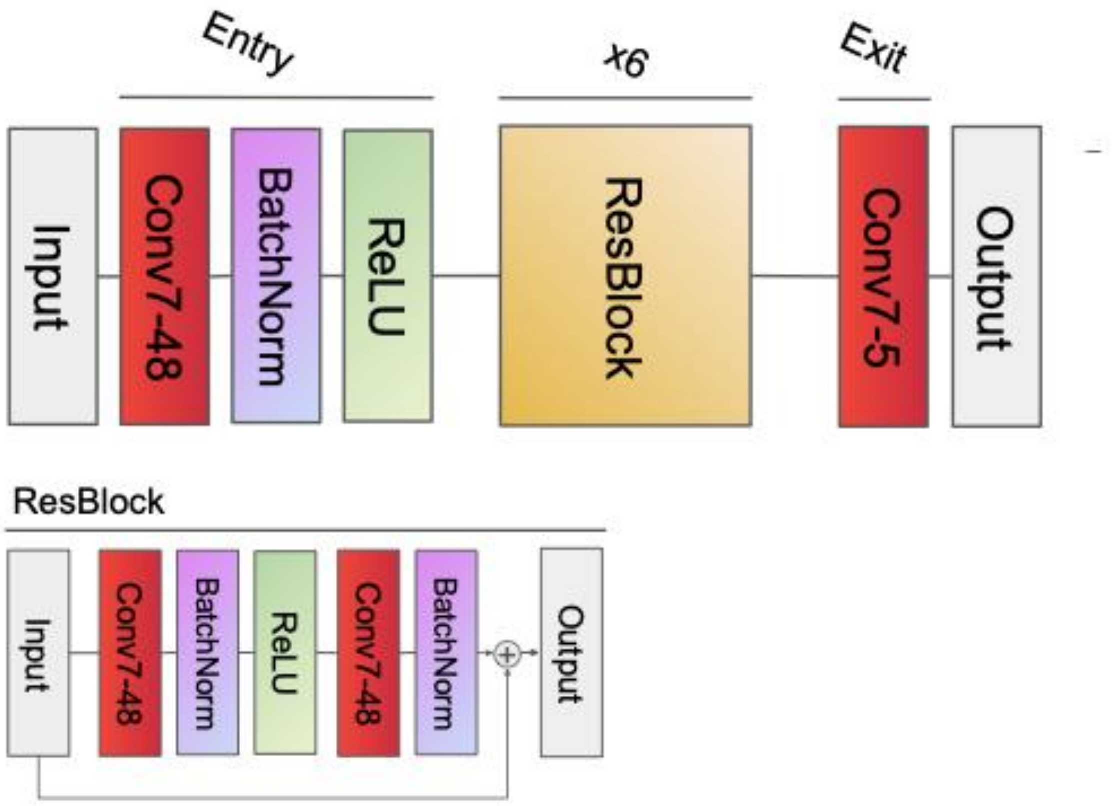


Fig. 10. Architecture of the ResNet Model

Methodology

ViTAE-SL

Description

- Vision Transformer–based AutoEncoder.
- Inputs split into patches → linear projection + positional embeddings.
- Transformer encoder models spatial relationships and long-range dependencies.

Advantages

- Captures **global context** beyond CNN receptive fields.
- Learns **interactions across distant regions**.
- Hybrid design (transformer + conv) balances **global** and **local detail**.

Architecture

- Linear projection of patches + positional embeddings.
- Transformer encoder → feature representations.
- Linear decoder reshapes to grid; conv decoder refines local structure.
- Trained with MSE loss.

Output

- Reconstructed atmospheric fields (5 channels).

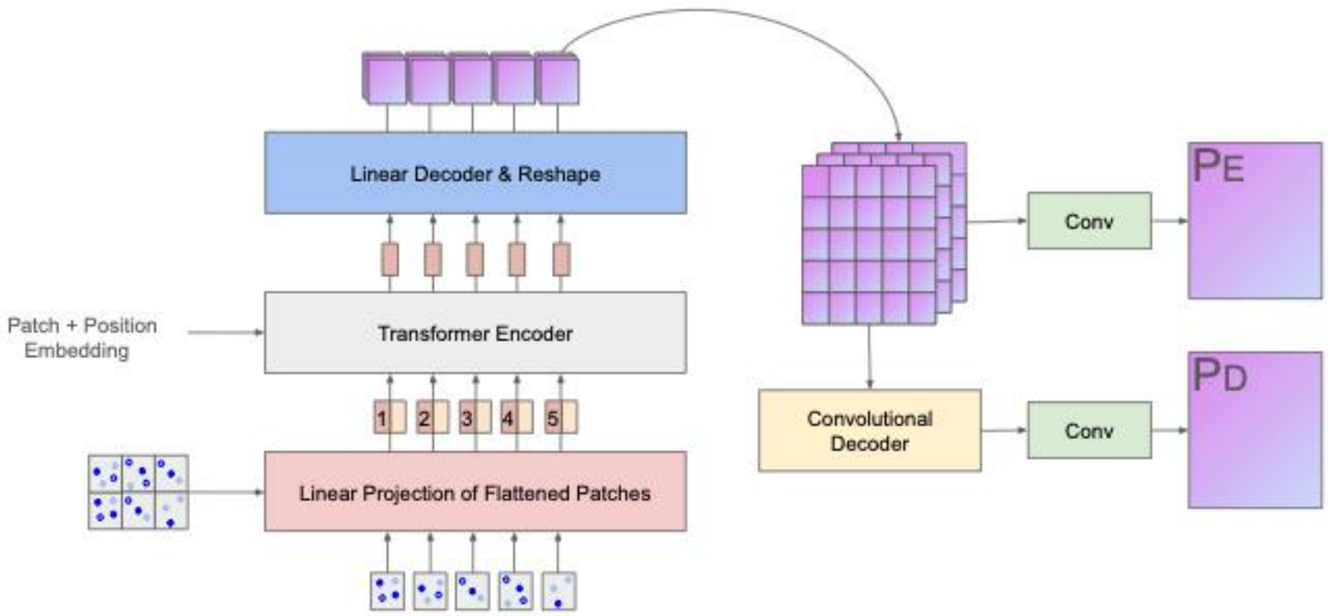


Fig. 11. Architecture of the ViTAE-SL model@x

Methodology
CWGAN-GP

Conditional Wasserstein GAN with Gradient Penalty

- Generator learns to reconstruct fields from sparse observations
- Discriminator distinguishes between real and generated fields
- U-net Generator, Small Discriminator

Training objective

- Wasserstein loss for stable training
- Gradient penalty to enforce Lipschitz constraint
- Additional L1 loss for reconstruction accuracy

Advantage

- Captures data distribution
- Produces realistic and diverse reconstructions

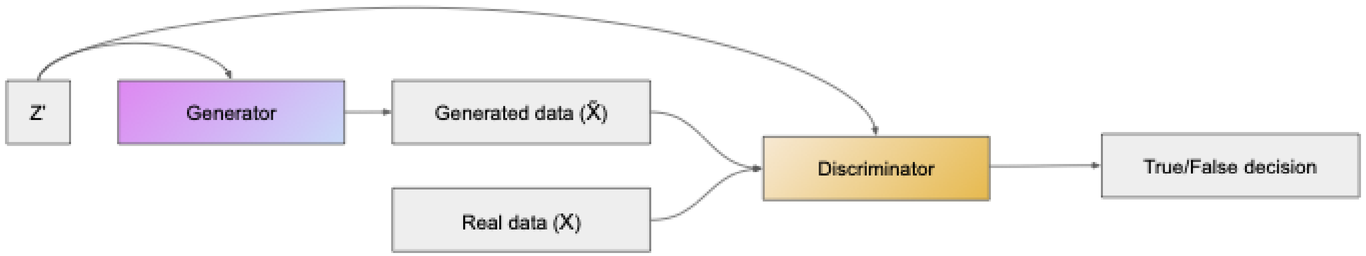


Fig. 12. Schematic of the Conditional GAN

$$\mathcal{L}_G^{\mathrm{CWGAN}} = -\mathbb{E}_{\tilde{X}\sim P_g}\left[D(\tilde{X} \mid Z')\right] + \lambda_2 \mathbb{E}_{X,\tilde{X}}\left[\|X - \tilde{X}\|_1\right]$$

$$\mathcal{L}_D^{\mathrm{CWGAN}} = \mathbb{E}_{\tilde{X}\sim P_g}\left[D(\tilde{X} \mid Z')\right] - \mathbb{E}_{X\sim P_r}\left[D(X \mid Z')\right] + \lambda_1 \mathbb{E}_{\hat{X}\sim P_{\hat{X}}}\left[\left(\left\|\nabla_{\hat{X}} D(\hat{X} \mid Z')\right\|_2 - 1\right)^2\right]$$

Methodology
DDIM

Denoising Diffusion Implicit Model

Forward process
- Gradually adds Gaussian noise to data
- Transforms input into pure noise

Reverse process
- Neural network learns to denoise step by step
- Recovers clean data distribution from noise

Advantages
- Stable training compared to GANs
- High-quality and realistic reconstructions

Inconvenients
- Inference is Expensive
- Can be sensitive to hyperparameters

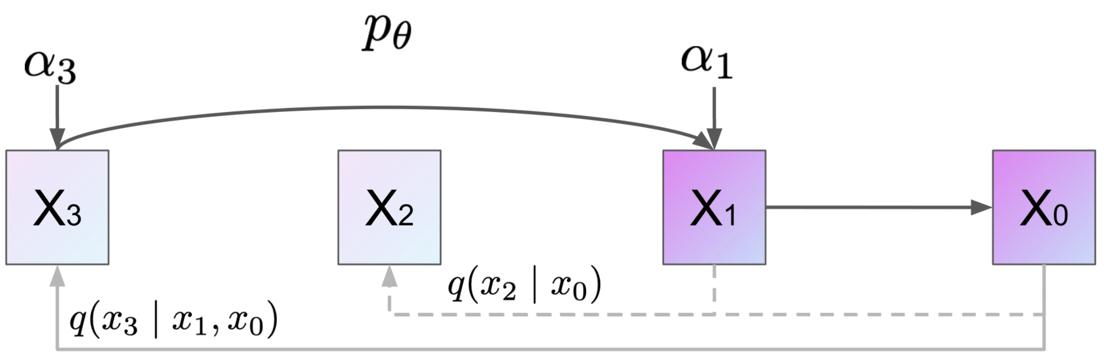


Fig. 13. Schematic of the DDIM reverse sampling process

Parameters:
- T=1000, 50 steps DDIM
- Average of 5 predictions

Methodology
Model and Conditioning

Architecture
- We use the same U-net as the VT-Unet
- Time embedding is added with cosine embedding

Spatial conditioning
- Encodes sensor observations into feature maps
- Convolutional layers extract spatial structure

Integration with diffusion model
- Conditioning maps resized to match feature maps
- Added to intermediate network layers (feature fusion)

Advantage
- Preserves spatial correlations from observations
- Guides diffusion process toward physically consistent reconstructions

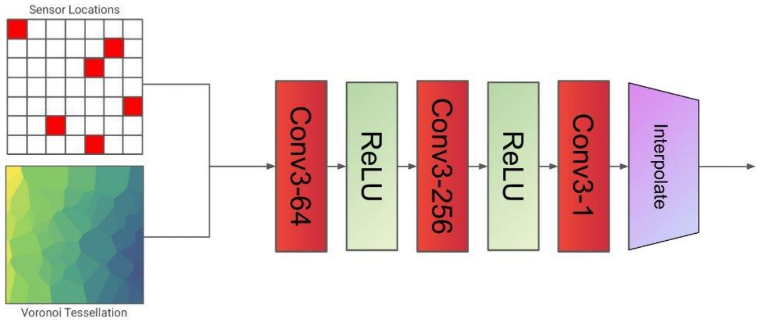


Fig. 14. Schematic of the spatial conditioning method

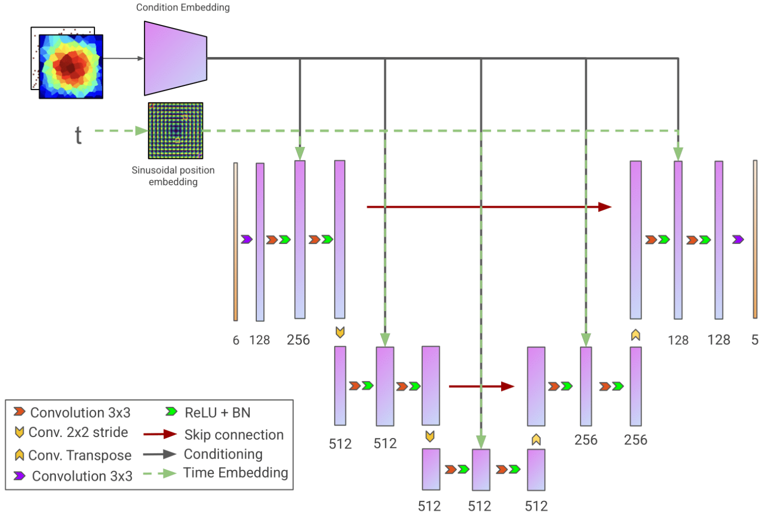


Fig. 15. Schematic of the U-Net architecture with time and condition embedding

Results
Benchmark

Table 1. performance comparison of the Field Reconstruction Methods

Model	↓ RRMSE	↓ MAE	↑ SSIM[1]
Kriging	0.4338	0.6757	0.5325
Cubic Interp.	0.5037	0.7265	0.3750
VT-UNet	0.284	0.191	0.835
VT-ResNet	0.299	0.207	0.810
ViTAE-lite	0.3108	0.2202	0.7785
ViTAE-base	0.2718	0.1878	0.8292
ViTAE-large	**0.2494**	**0.1692**	0.8555
CWGAN-GP	0.2664	0.1805	0.8407
DDIM (spatial cond.)	0.2578	0.1764	**0.8621**

- **All ML models** outperform the baselines (Kriging, Cubic Interpolation).
- **ViTAE-Large** shows the best raw reconstruction performance (↓RRMSE, ↓MAE).
- **DDIM** achieves the highest structural similarity (↑SSIM), with more spatially consistent fields.
- **Generative approaches (DDIM, CWGAN-GP)** consistently outperform CNNs (U-Net, ResNet).
- Results are **statistically significant** due to the large evaluation sample size.

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \qquad \mathrm{RRMSE}^* = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}}{\bar{y}}, \quad \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

1. SSIM: Structural Similarity Index Measure

* In this case the standard deviation was used because data is normalised and can take negative values.

Evaluation
Error distribution

- **ML models**: Relative L2 error concentrated at small values.
- **Narrow histograms** → stable & reliable reconstruction.
- **Wind fields**: consistently harder to reconstruct for all models.
- **Baselines**: worse accuracy **and** more unstable (wider error distributions).

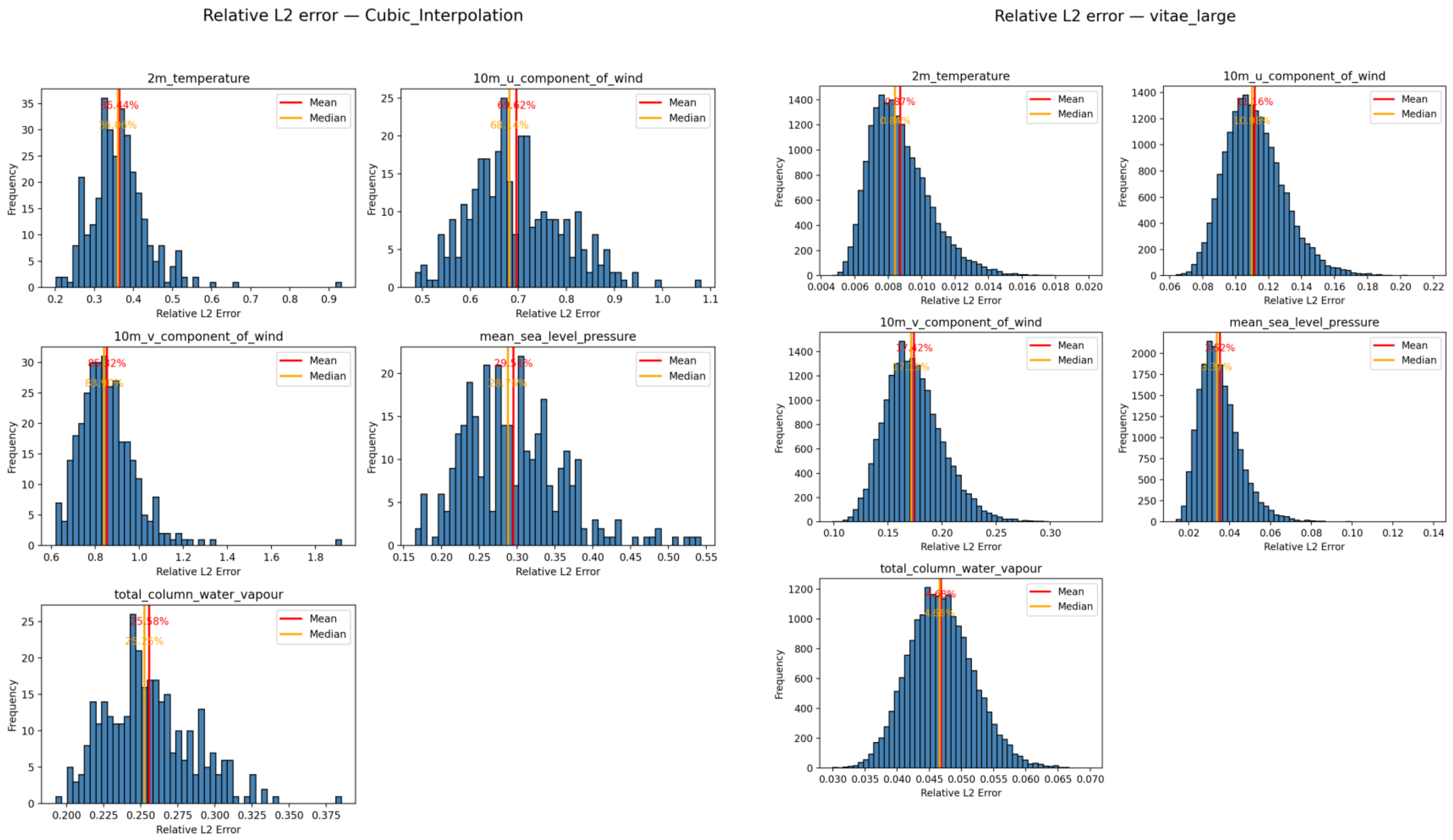


Fig. 16. Relative L2 error distribution across all fields for Cubic Interpolation and ViTAE-Large

Evaluation
Visual Assessment

- **ML reconstructions**: visually very close to ground truth, hard to distinguish between models.
- **DDIM**: slightly sharper fields → thanks to multi-step generative process.
- **ViTAE-Lite**: oversmoothed, lacks small-scale detail.
- **Model size trend**: more parameters → richer small-scale variations.
- ML models **go beyond interpolation** → produce realistic, physically consistent fields.

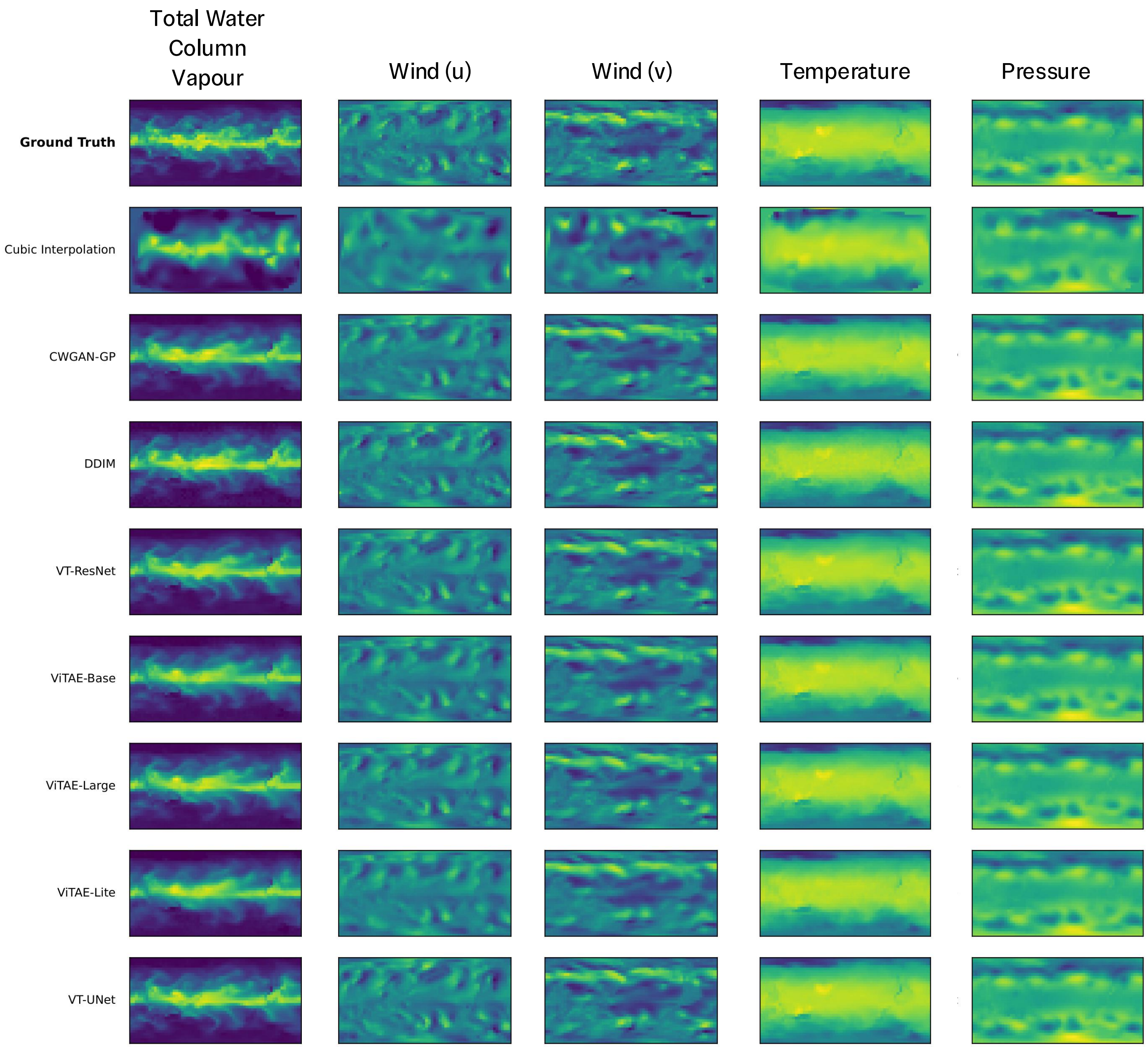


Fig. 17. An example of field reconstructions across all 5 fields

Evaluation
Sparsity robustness

- Evaluated models at **0.5, 1, 2, 5, 7.5, 10, 20, 50%** observations in order to benchmark the models' robustness to sparsity

- **ViTAE family**: most robust under high sparsity
- **Probabilistic models (DDIM, CWGAN)** outperform standard CNNs (U-Net) by a significant margin

- **Suggests:**
  - Probabilistic approaches are more robust:
    - Adversarial training
    - Multi-step generation process

  - Transformer architectures bring clear benefits

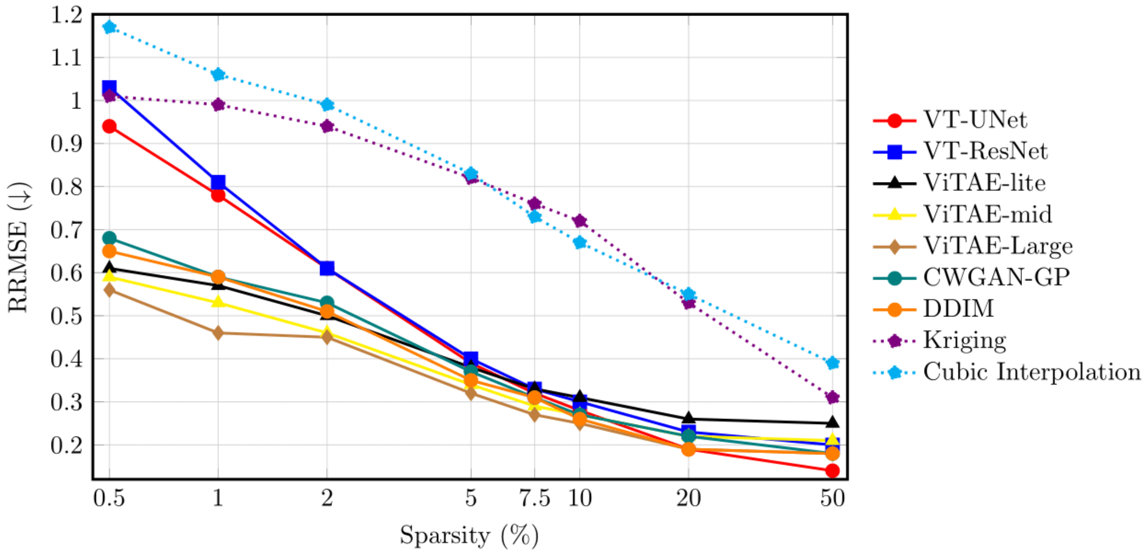


Fig. 19. RRMSE VS sensor coverage on WeatherBench 2 (log-% on x)

Results

Extreme Weather Events

- Selected **extreme events** (temperature extremes, Tropical Cyclones, Extratropical Cyclones, Atmospheric Rivers) from test set using statistical procedure.
- **ML models** maintain performance close to full-dataset results.
- Small event sample sizes → differences **not statistically significant**. (Between ML models)
- **Baselines** (Kriging, Cubic Interp.) degrade **substantially** under extremes.

Table 2. performance comparison of the Field Reconstruction Methods on Extreme Weather Events

Model	↓ RRMSE	↓ Difference
Kriging	0.87	+0.15
Cubic Interp.	1.21	+0.54
VT-UNet	0.28	0
VT-ResNet	0.32	+0.02
ViTAE-lite	0.34	+0.03
ViTAE-base	0.288	+0.01
ViTAE-large	**0.26**	+0.01
CWGAN-GP	0.28	+0.01
DDIM (spatial cond.)	**0.26**	0

Results

Computational cost and efficiency

- **ViTAE family**: best FLOPs ↔ performance trade-off (benefits from transformer design + small field size).
- **DDIM**: worse ratio, though performance could improve with better model design.
- **GANs**: outperform CNNs without extra inference cost (unless sampling multiple passes).
- **Overall:**
  - At **10% obs.**, probabilistic ≈ deterministic (small gains).
  - At **high sparsity**, probabilistic models perform **much better**.
  - Probabilistic models are a lot **harder to train** (instability, hyperparameters…)

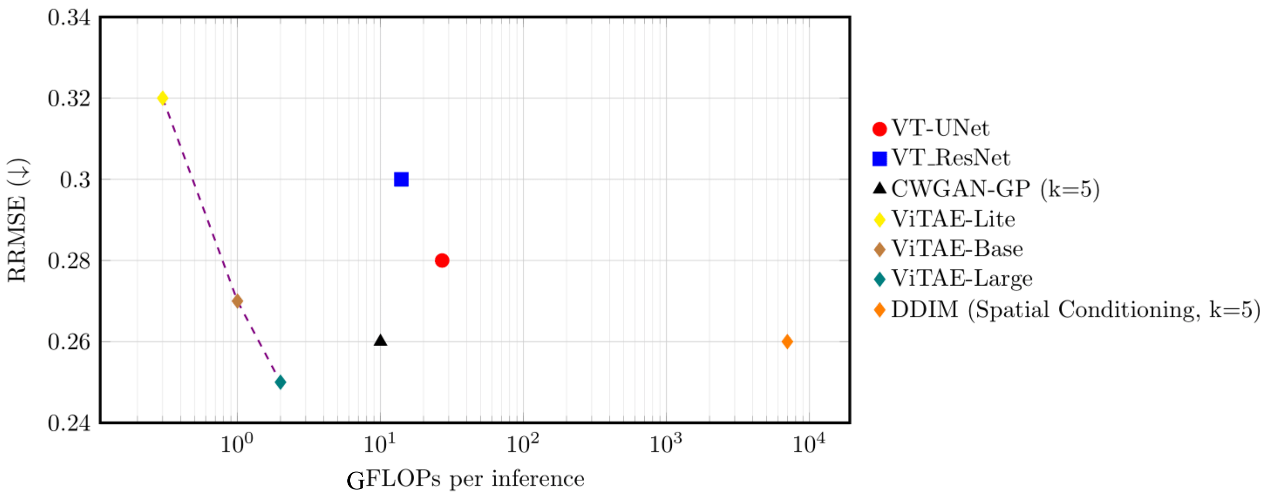


Fig. 20. RRMSE VS GFLOPs required per inference across all models

Conclusion

Contributions and Limitations

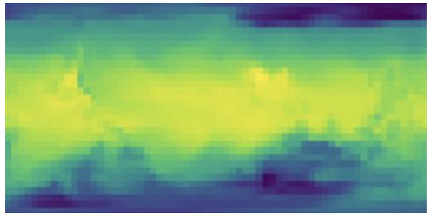
- **Problem Recap**: Field reconstruction from sparse observations is crucial, current methods (DA, kriging, cubic interpolation) are costly or limited.

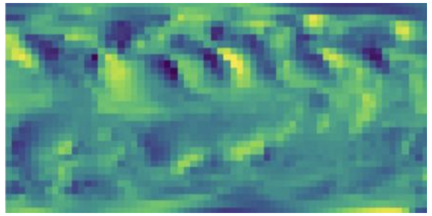
- **Our Contribution**: Comparative study of deterministic (U-Nets, ViTAE) vs probabilistic (DDIM, CWGAN) ML models on WeatherBench2 under varying sparsity.

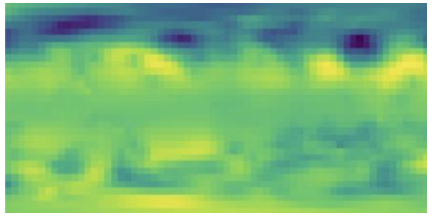
- **Key Findings**:
  - Probabilistic approaches show higher robustness under extreme sparsity.
  - Transformer-based architectures (ViTAE family) consistently more stable.
  - CNN baselines degrade quickly with fewer observations.

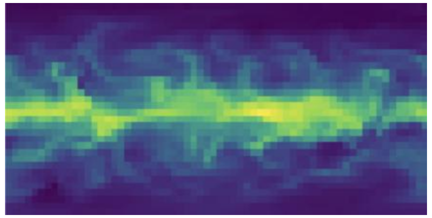
- **Implication**: Evidence that generative + transformer-based models are promising directions for robust field reconstruction.

- **Next Steps**: Larger-scale evaluation, integration with data assimilation pipelines, exploration of hybrid methods.









IMPERIAL

Thank you

Sources
Bibliography*

[1]. Fukami K, Romit Maulik, Ramachandra N, Koji Fukagata, Taira K. Global field reconstruction from sparse sensors with Voronoi tessellation-assisted deep learning. Nature Machine Intelligence. 2021 Oct 28;3(11):945–51.

[2]. Fan H, Cheng S, de Nazelle AJ, Arcucci R. ViTAE-SL: A vision transformer-based autoencoder and spatial interpolation learner for field reconstruction. Computer Physics Communications [Internet]. 2024 Dec 16;308:109464. Available from: https://www.sciencedirect.com/science/article/pii/S0010465524003874?via%3Dihub

[3]. Zhuang Y, Cheng S, Duraisamy K. Spatially-aware diffusion models with cross-attention for global field reconstruction with sparse observations. Computer Methods in Applied Mechanics and Engineering [Internet]. 2024 Dec 6;435:117623. Available from: https://www.sciencedirect.com/science/article/pii/S0045782524008776`

[4]. Wikipedia Contributors. Kriging [Internet]. Wikipedia. Wikimedia Foundation; 2018. Available from: https://en.wikipedia.org/wiki/Kriging

[5]. Wikipedia Contributors. Bicubic interpolation [Internet]. Wikipedia. Wikimedia Foundation; 2019. Available from: https://en.wikipedia.org/wiki/Bicubic_ interpolation

[6]. Hu C, Kikumoto H, Zhang B, Jia H. Fast estimation of airflow distribution in an urban model using generative adversarial networks with limited sensing data☆. Building and Environment [Internet]. 2023 Dec 19;249:111120. Available from: https://www.sciencedirect.com/science/article/pii/S0360132323011472

[7]. Carrassi A, Bocquet M, Bertino L, Evensen G. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. WIREs Climate Change. 2018 Jul 9;9(5).

[8]. Rasp S, Hoyer S, Merose A, Langmore I, Battaglia P, Russel T, et al. WeatherBench 2: A benchmark for the next generation of data-driven global weather models. arXiv (Cornell University). 2023 Aug 29;

* All the sources are available in the report