ML for graphs final exam

Probabilistic Weather Forecasting with Hierarchical Graph Neural Networks

Abdoul ZEBA & Amine RAZIG

ENSAE & Ecole Polytechnique, Palaiseau, France

April 9, 2025

Motivation

Challenges in Traditional Weather Forecasting

- **Traditional Forecasting**:
  - Relies on supercomputers to simulate atmospheric physics.
  - Initial 3D grid of the atmosphere is imperfect due to data gaps.
  - Uncertainty grows over time, reducing forecast accuracy.
- **Ensemble Forecasting**:
  - Generates multiple forecasts by tweaking initial conditions.
  - Computationally expensive, limiting spatial resolution and ensemble size.
- **Proposed Solution**:
  - Develop a probabilistic model using hierarchical graph neural networks (Graph-EFM).
  - Model the full distribution $p(X^{1:T}|X^{-1:0}, F^{1:T})$ of future weather states.
  - Efficiently sample ensemble forecasts to quantify uncertainty.

Methodology

Graph-based Ensemble Forecasting Model (Graph-EFM)

- **Objective**:
  - Model the distribution $p(X^{1:T}|X^{-1:0}, F^{1:T})$ of future weather states.
  - Generate ensemble forecasts to quantify uncertainty.
- **Approach**:
  - Introduce a latent variable Z^t to capture uncertainty at each time step.
  - Decompose the distribution into a product of conditional distributions:

$$p(X^{1:T}|X^{-1:0}, F^{1:T}) = \prod_{t=1}^{T} \int p(X^t|Z^t, X^{t-2:t-1}, F^t)p(Z^t|X^{t-2:t-1}, F^t)\, dZ^t$$

- **Components**:
  - Latent Map: Models $p(Z^t|X^{t-2:t-1}, F^t)$ as an isotropic Gaussian.
  - Predictor: Computes $\hat{X}^t = X^{t-1} + g(Z^t, X^{t-2:t-1}, F^t)$.

Methodology

Hierarchical Architecture

- **Hierarchical Graph Structure**:
  - Multiple graph levels $G_1, G_2, \ldots, G_L$, where $G_l = (V_l, E_l)$.
  - The number of nodes decreases with level l.
- **Information Propagation**:
  - Information flows up through $G_{1,2}, G_{2,3}, \ldots, G_{L-1,L}$ and filters back down via $G_{L,L-1}, \ldots, G_{2,1}$ using GNNs.

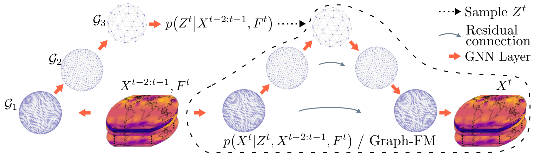


Figure: Hierarchical architecture of Graph-EFM [3].

Methodology

Model Training

- **Pre-training**:
  - Minimize the ELBO (Evidence Lower Bound) for a single time step:

$$\mathcal{L}_{\mathsf{Var}} = \lambda_{\mathsf{KL}} D_{\mathsf{KL}}\left(q(Z^t|\cdot)\middle\|p(Z^t|\cdot)\right) - \mathbb{E}_{q(Z^t|\cdot)}\left[\log p(X^t|Z^t,\cdot)\right]$$

  - Goal: Learn latent representations while maintaining spatial coherence.
- **Finetuning**:
  - Add a CRPS (Continuous Ranked Probability Score) loss to improve ensemble diversity:

$$\mathcal{L} = \mathcal{L}_{\mathsf{Var}} + \lambda_{\mathsf{CRPS}}\mathcal{L}_{\mathsf{CRPS}}$$

  - The CRPS loss penalizes differences between two independent ensemble members.

GraphCast

Learning Skillful Medium-Range Global Weather Forecasting

Predict Earth's weather at high resolution (0.25°) up to 10 days ahead, and provide deterministic forecasts using graph neural networks which have been train by minimizes a mean squared error (MSE) in an autoregressive loss setting.

- **Architecture**: Three main components:
  1. **Encoder**: Maps input data to a multi-mesh graph representation.
  2. **Processor**: Updates graph nodes using learned message-passing.
  3. **Decoder**: Maps processed graph features back to the grid.
- **Key Features**:
  - State-of-the-art performance for deterministic forecasts.
  - Efficient and scalable for global weather prediction.

Graph-EFM Experiments

Local and Global Forecasting

Global Results (ERA5) [1]

Variable	Model	Lead time 5 days		
		RMSE	CRPS	SpSkR
z500	GraphCast*	**387**	236	-
	Graph-EFM	399	**169**	**1.18**
2t	GraphCast*	1.65	1.00	-
	Graph-EFM	**1.64**	**0.71**	**0.98**

Variable	Model	Lead time 10 days		
		RMSE	CRPS	SpSkR
z500	GraphCast*	808	498	-
	Graph-EFM	**695**	**299**	**1.15**
2t	GraphCast*	2.82	1.69	-
	Graph-EFM	**2.32**	**1.00**	**0.99**

Regional Results (MEPS) [2]

Variable	Model	Lead time 24h		
		RMSE	CRPS	SpSkR
z500	GraphCast*	**153**	108	-
	Graph-EFM	172	**91**	**0.84**
wvint	GraphCast*	**1.51**	1.01	-
	Graph-EFM	1.61	**0.79**	**0.57**

Variable	Model	Lead time 57h		
		RMSE	CRPS	SpSkR
z500	GraphCast*	**201**	138	-
	Graph-EFM	219	**115**	**0.75**
wvint	GraphCast*	**2.82**	1.32	-
	Graph-EFM	2.08	**1.00**	**0.53**

Our Experimental Methodology

Data Presentation and Graph Modelisation

Dataset

- ERA5

Variables

- `10m` `u`-component of wind
- `10m` `v`-component of wind
- `2m` temperature

Geographical Area

- North: 50.85°
- West: -1.46°
- South: 48.36°
- East: 3.37°

Modelisation of a Deterministic and a Probabilistic Graph Model

- Definition of both models
- Comparison of deterministic vs. probabilistic approach
- Application and interpretation

The experiments we conducted are available on our *GitHub Repository*

Our Experimental Methodology

Deterministic Model: Encoder-Decoder GCN

- Basic Graph Convolutional Network (GCN) architecture
- Encoder-decoder structure for representation learning

Probabilistic Model: Variational GCN see

- **Variational GCN Encoder:**
  - Low-level: in_channels $\rightarrow$ hidden $\rightarrow$ hidden/2
  - Mid-level: hidden/2 $\rightarrow$ hidden/4
  - High-level: hidden/4 $\rightarrow$ latent (μ, logvar)
- **GCN Decoder:**
  - 3-layer architecture transforming latent space to predictions
- **Training Approach:**
  - Loss: Combination of reconstruction loss (MSE) and KL divergence
  - Strategy: Sequence-to-sequence prediction on graph time series

Experiments: Results

Deterministic vs Probabilistic Model Performance

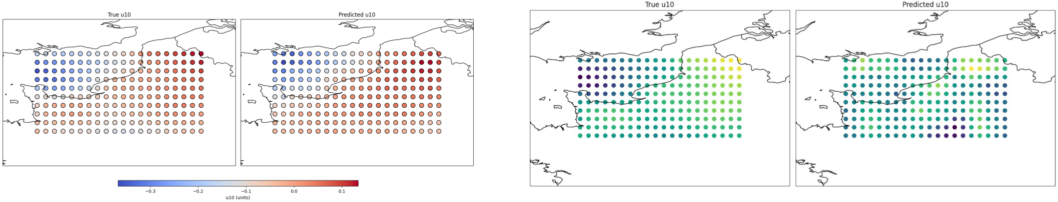


Figure: Deterministic model (left) / Probabilistic model (right) on U-wind component

	Deterministic			**Probabilistic**		
	`t2m`	`u10`	`v10`	`t2m`	`u10`	`v10`
MAE	27.498	**2.088**	**0.506**	**27.487**	2.947	0.860
RMSE	32.536	**2.173**	**0.624**	**32.522**	3.031	1.004

Key Observation: *The deterministic model yields lower errors for wind, while both models perform similarly for t2m.*

Conclusion

THANK YOU FOR YOUR ATTENTION

QUESTION ?

References

H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thépaut.
The era5 global reanalysis.
Quarterly Journal of the Royal Meteorological Society, 146(730):1999–2049, 2020.

M. Müller, M. Homleid, K.-I. Ivarsson, M. A. Køltzow, M. Lindskog, K. H. Midtbø, U. Andrae, T. Aspelien, L. Berggren, D. Bjørge, P. Dahlgren, J. Kristiansen, R. Randriamampianina, M. Ridal, and O. Vignes.
Arome-metcoop: A nordic convective-scale operational weather prediction model.
Weather and Forecasting, 32(2):609 – 627, 2017.

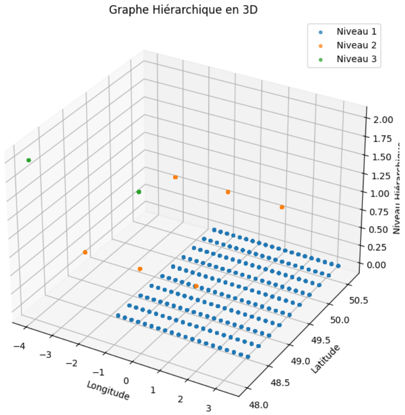
J. Oskarsson, T. Landelius, M. P. Deisenroth, and F. Lindsten.
Probabilistic weather forecasting with hierarchical graph neural networks, 2024.

Appendix: Our Experimental Methodology

Manual Construction of a Hierarchical Graph

Manual Construction of a Hierarchical Graph

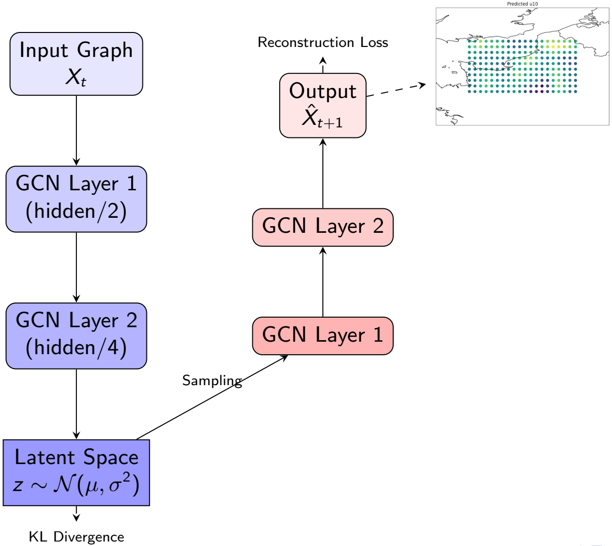
- Definition of the hierarchical structure
- Manual node and edge creation
- Validation of the hierarchy



Appendix: Our Experimental Methodology

Our variational Graph Convolutional Network Architecture.

▸ Go back



Appendix : Graph Cast

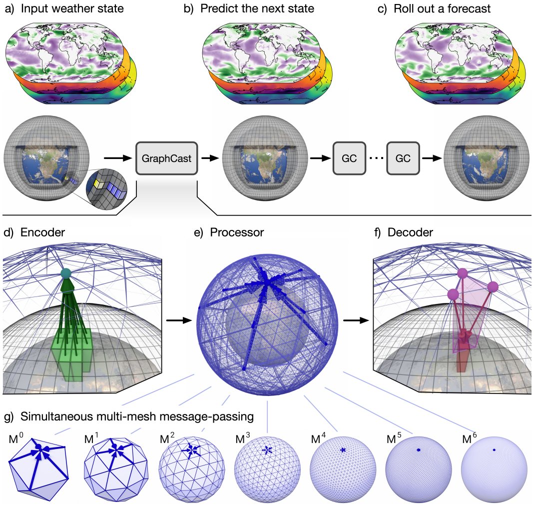


Figure: Model schematic GaphCast (lam2023)