Deterministic and Probabilistic Near-Earth Space Weather Forecasting with Machine Learning

Daniel Holmberg, Dec. 15, 2025
AGU New Orleans

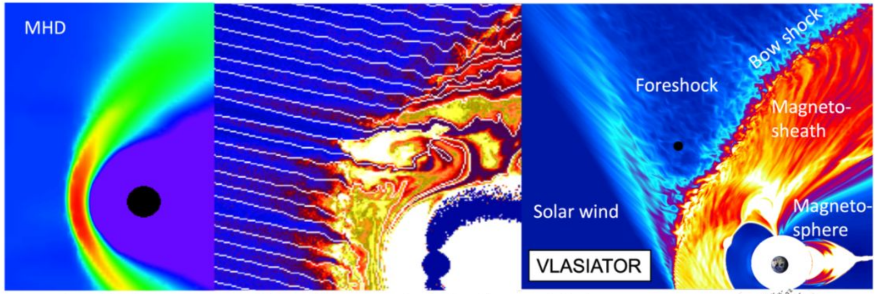
HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

FMI Space and Earth Observation Centre

UC SANTA BARBARA
Electrical & Computer Engineering

Magnetospheric simulation methods

- **Magnetohydrodynamics (MHD)** models plasma as a continuous fluid, combining fluid dynamics with Maxwell's equations to describe plasma motion under magnetic fields.
- **Particle-in-cell (PIC)** simulates many individual (or super) particles in a self-consistent electromagnetic field.
- **Hybrid-Vlasov (Vlasiator)** used in this work simulates ions through the Vlasov equation on a phase-space grid, capturing their distribution function directly, while electrons are treated as a massless, charge-neutralizing fluid.



MHD: Velocity distribution function (VDF) is not modelled. Temperature evaluated based on the assumption that the VDF is a normal distribution (Maxwellian).

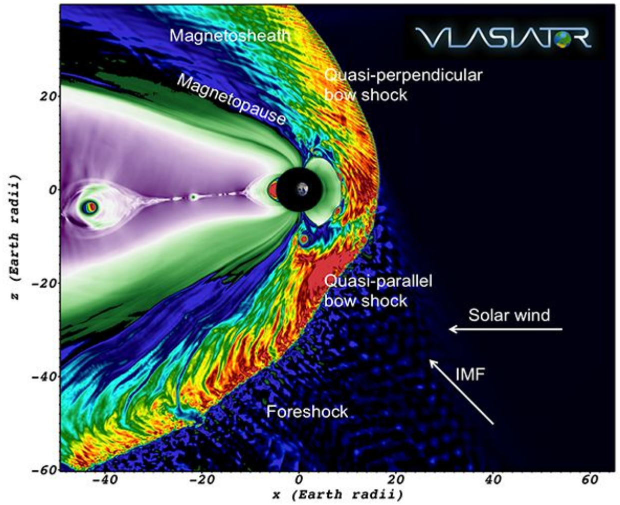
Particle-in-cell (PIC, e.g., Karimabadi et al, 2014). Distribution function is constructed from particle statistics. Since many parameters are evaluated from the VDF, the result has inherent noise.

Hybrid-Vlasov: Distribution function is modelled in a grid by phase space density. No approximations.

Hybrid-Vlasov simulations

- 2D space + 3D velocity (2D-3V) simulation on the noon–midnight (x-z) Geocentric Solar Ecliptic plane.
- Domain: x = −60 → 30 Re, z = −30 → 30 Re, spatial res 600 km.
- Inner boundary at 3.7 Re, dayside inflow with Maxwellian solar wind.
- Four runs with increasing solar-wind ion density ρ.

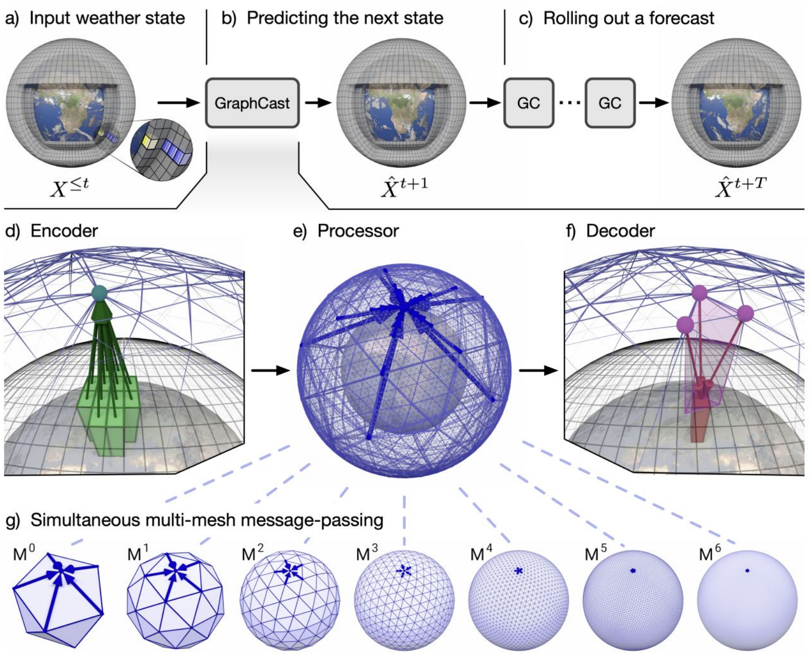
Label	ρ (cm^{-3})	$\boldsymbol{v}$ (km/s)	$\boldsymbol{B}$ (nT)	T (MK)	M_A	Δt (s)	t_{tot} (s)
Run 1	0.5	$(-750, 0, 0)$	$(0, 0, -5)$	0.5	4.9	1.0	800
Run 2	1.0	$(-750, 0, 0)$	$(0, 0, -5)$	0.5	6.9	1.0	800
Run 3	1.5	$(-750, 0, 0)$	$(0, 0, -5)$	0.5	8.4	1.0	800
Run 4	2.0	$(-750, 0, 0)$	$(0, 0, -5)$	0.5	9.8	1.0	800



M. Palmroth, et al. Vlasov methods in space physics and astrophysics. Living reviews in computational astrophysics (2018)

Proposed method

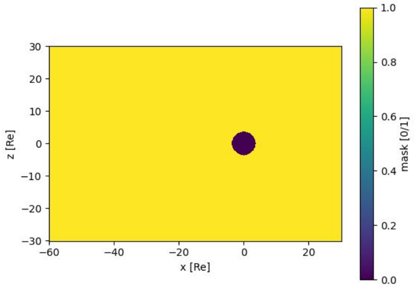
- Train a graph neural network (GNN) to *autoregressively* predict next simulation frame.
- Meaning, predicted state is used as input to predict the following after that.
- Large advantage in terms of speed with respect to numerical simulations.
- Modern generative models open the doors for fast ensemble forecasts and uncertainty quantification.



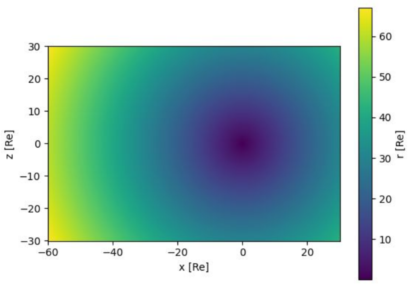
R. Lam et al., Learning skillful medium-range global weather forecasting. Science (2023)

Propagated features

- Magnetic and electric fields
- Plasma moments: velocity, density, pressure, temperature
- Encode also static coordinates (x, z, radial distance)
- Released openly in Zarr format to enable ML studies on highly resolved plasma.



binary grid mask

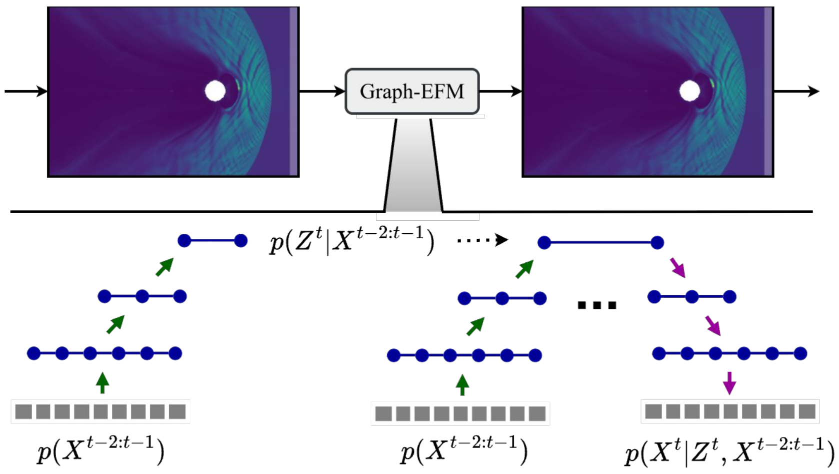


radial distance

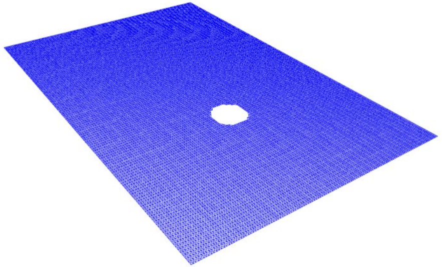
Variable	Label	Unit
Magnetic field x-component	B_x	nT
Magnetic field y-component	B_y	nT
Magnetic field z-component	B_z	nT
Electric field x-component	E_x	mV/m
Electric field y-component	E_y	mV/m
Electric field z-component	E_z	mV/m
Velocity field x-component	v_x	km/s
Velocity field y-component	v_y	km/s
Velocity field z-component	v_z	km/s
Particle number density	ρ	$1/\text{cm}^3$
Plasma pressure	P	nPa
Plasma temperature	T	MK

Model architecture

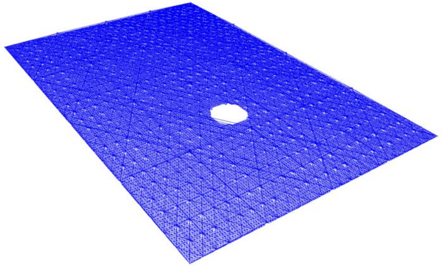
- Encode from high-resolution data *grid* on to a coarser *mesh*.
- Process node and edge representations using *interaction networks*, learning a latent mesh representation.
- Decode from mesh back to data grid yielding the predicted next state of the simulator.
- Probabilistic model injects noise into coarsest mesh level. Learns the full distribution.
- Can sample arbitrarily many ensemble members → forecast uncertainty.



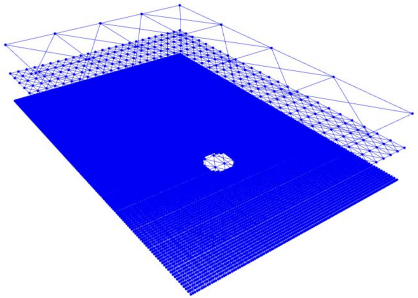
Mesh variations compared for the GNN



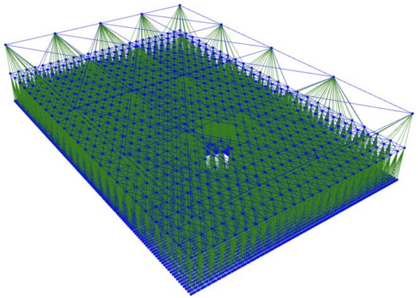
(a) Simple graph



(b) Multiscale graph



(c) Graph layers



(d) Hierarchical graph

Training objectives

Minimize composite loss over many autoregressive steps.

- Graph-FM: Sum of weighted Mean Squared Error (MSE) and magnetic divergence loss with derivatives discretized using second–order central differences.
- Graph-EFM: Variational autoencoder that maximizes Evidence Lower Bound (ELBO) with weighted Continuous Ranked Probability Score (CRPS) loss + divergence penalty.

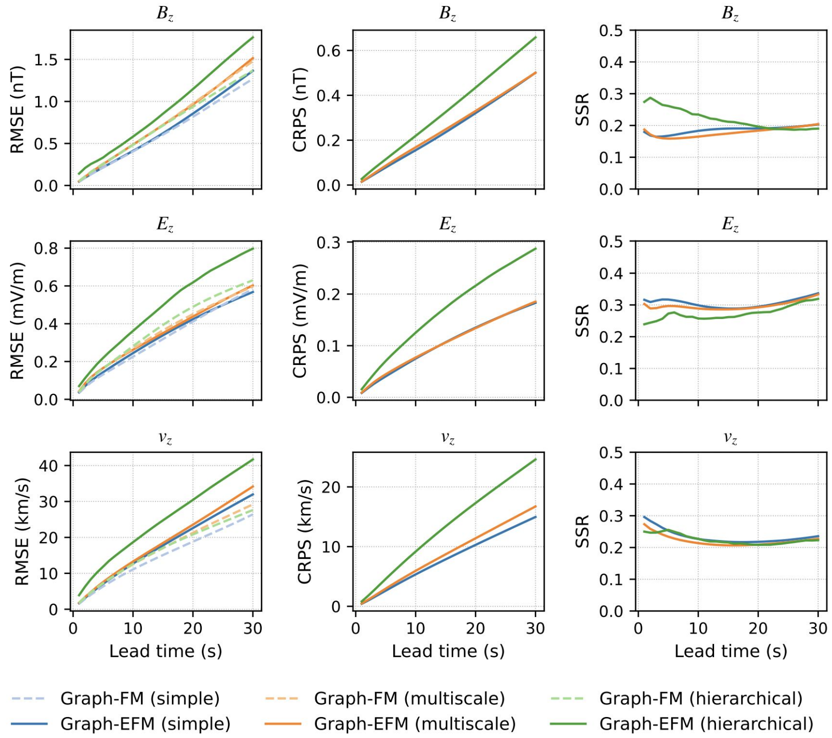
$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{TN} \sum_{t=1}^{T} \sum_{n=1}^{N} \sum_{i=1}^{d_x} \omega_i \lambda_i \left(\hat{X}_{n,i}^{t} - X_{n,i}^{t} \right)^2$$

$$\mathcal{L}_{\mathrm{Div}} = \frac{1}{TN} \sum_{t=1}^{T} \sum_{n=1}^{N} \left(\frac{\partial \hat{B}_x^t}{\partial x} + \frac{\partial \hat{B}_z^t}{\partial z} \right)_n^2$$

$$\mathcal{L} = \mathcal{L}_{\mathrm{MSE}} + \lambda_{\mathrm{Div}} \mathcal{L}_{\mathrm{Div}}$$

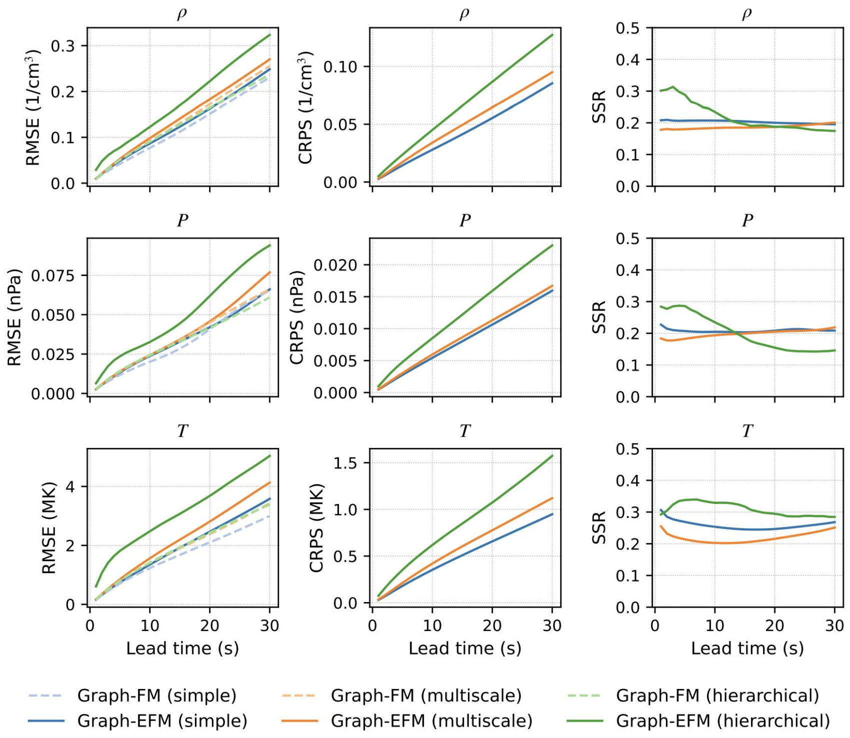
Forecast RMSE, CRPS and Spread-Skill-Ratio

- Metrics shown per lead time for *E*, *B*, and *v* z-components.
- RMSE: pointwise forecast error; lower is better.
- CRPS: measures how well the predicted distribution matches the true value.
- Spread-Skill-Ratio (SSR): Ratio of ensemble spread to RMSE (ideal ≈ 1). Here SSR ≈ 0.2–0.3, indicating underdispersive ensembles.
- The model captures epistemic uncertainty from model limitations, with no aleatoric uncertainty since the data contain no observational noise.

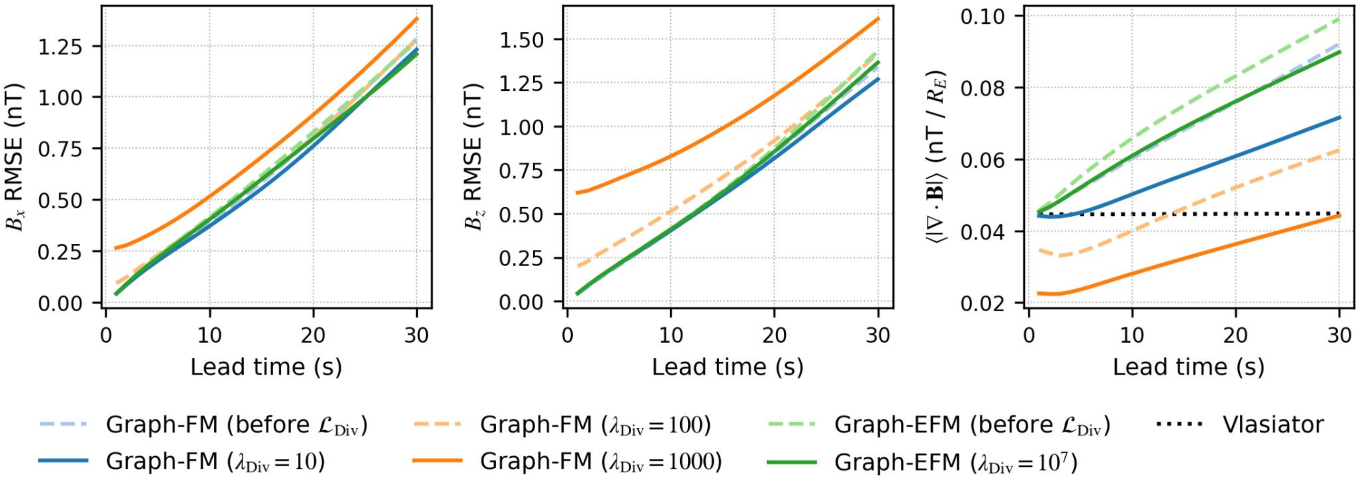


Forecast RMSE, CRPS and Spread-Skill-Ratio

- Metrics shown per lead time for ρ, P, and T z-components.
- RMSE: pointwise forecast error; lower is better.
- CRPS: measures how well the predicted distribution matches the true value.
- Spread-Skill-Ratio (SSR): Ratio of ensemble spread to RMSE (ideal ≈ 1). Here SSR ≈ 0.2–0.3, indicating underdispersive ensembles.
- The model captures epistemic uncertainty from model limitations, with no aleatoric uncertainty since the data contain no observational noise.

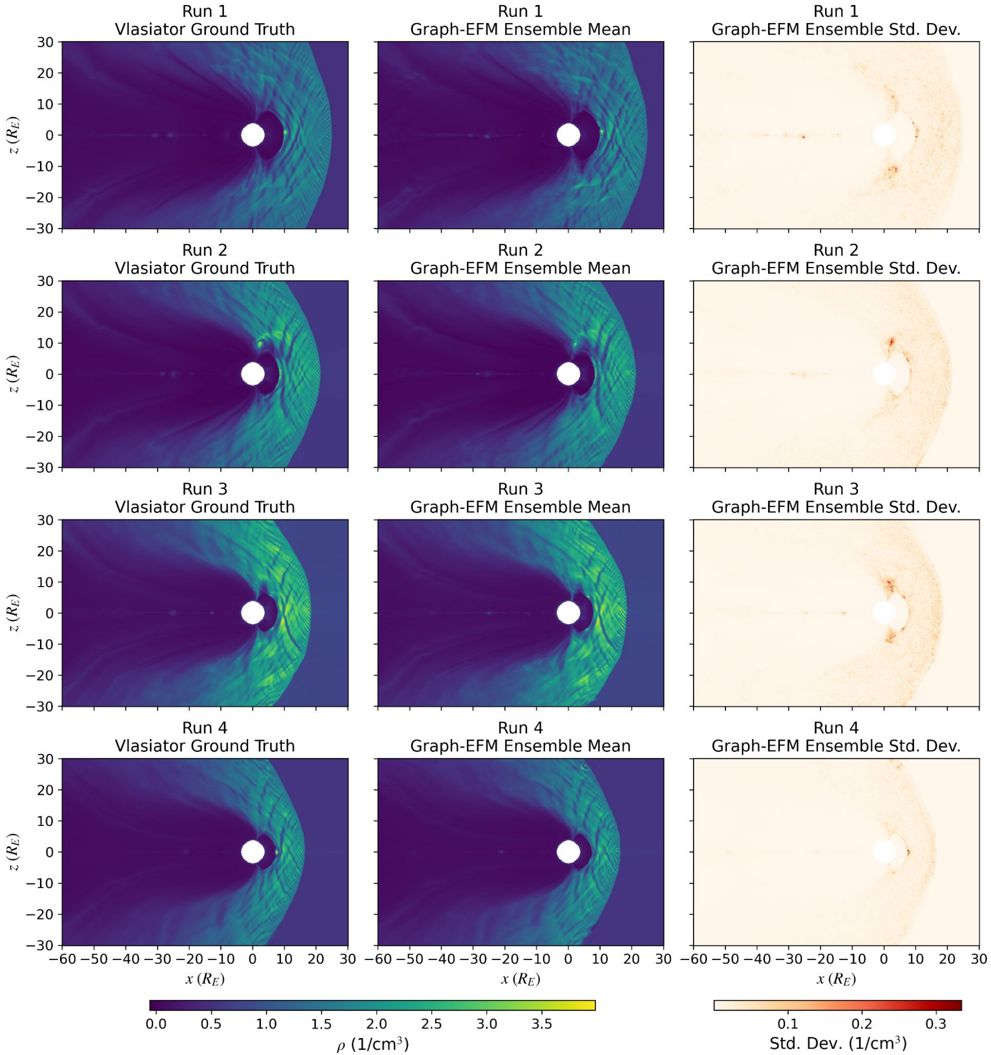


Effect of divergence penalty



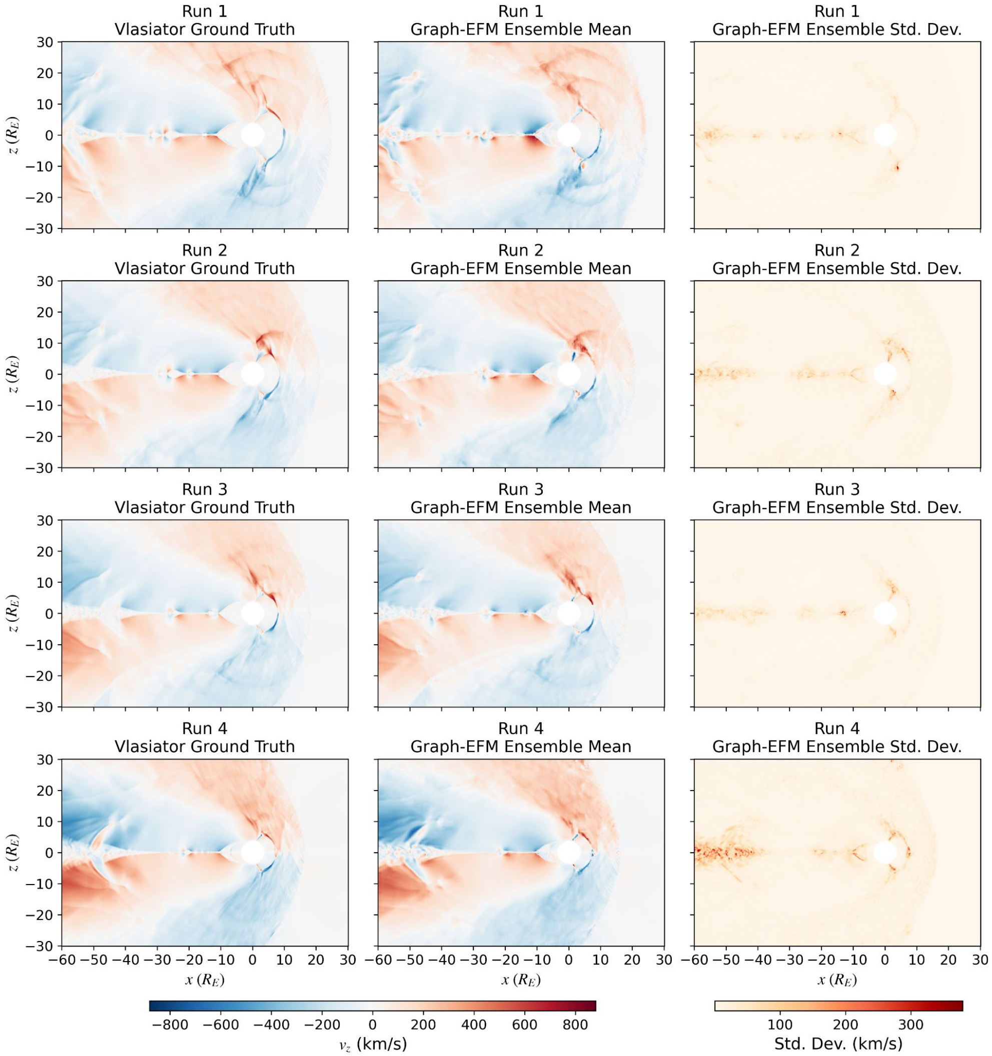
Example forecast

Predicted ρ and their uncertainty 30 steps ahead for all four runs.

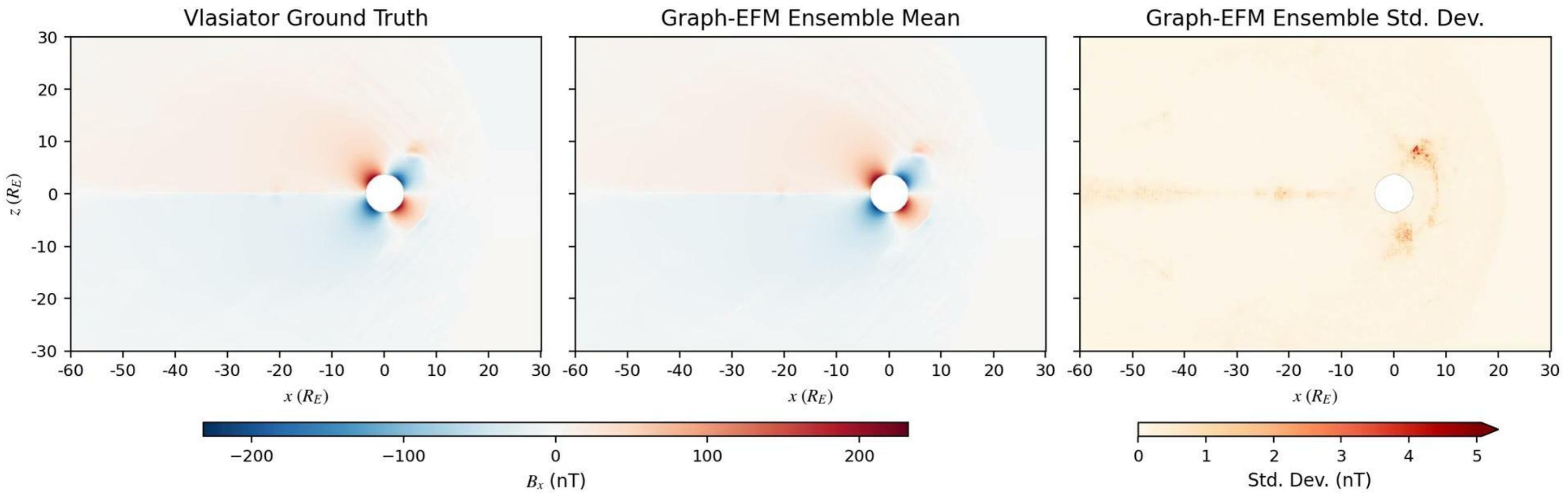


Example forecast

Predicted v_z and their uncertainty 30 steps ahead for all four runs.

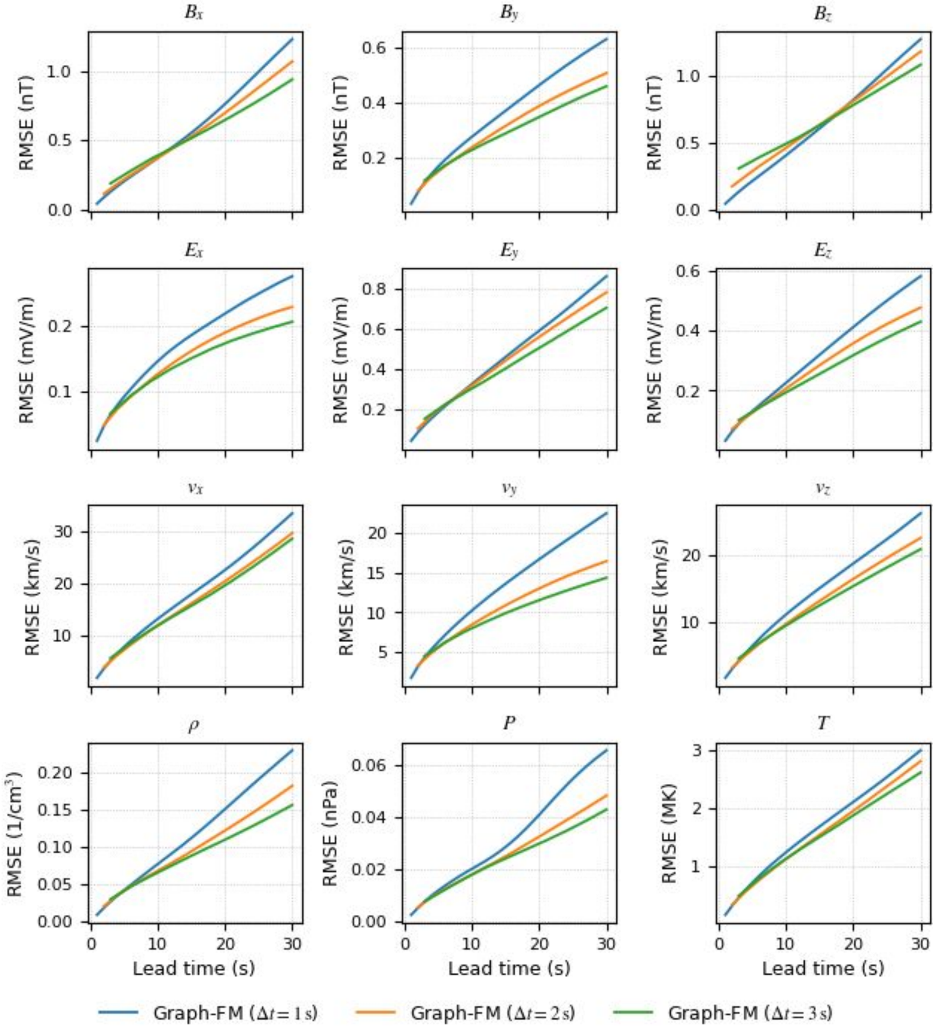


Example forecast



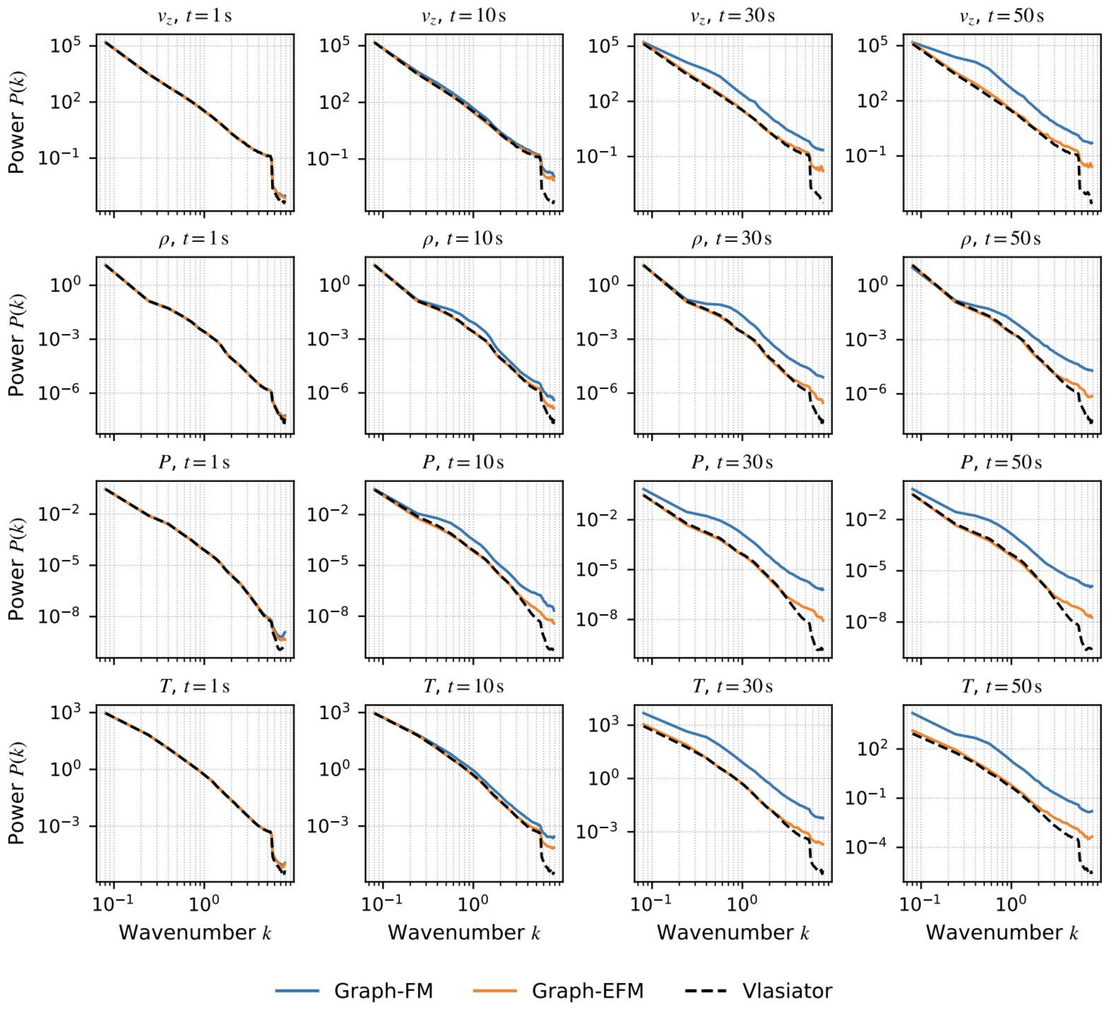
Step size comparison

- Trained Graph-FM with timesteps 1s, 2s, 3s by subsampling → larger timestep is trained on less data.
- Still larger step sizes incur less cumulative error.
- Would argue that much larger timesteps can be used as long as the temporal extent of the training data is there to match it.



Power spectra

- Power spectra reveal whether models preserve the correct scale-dependent structure of the system.
- At higher lead times and higher wavenumbers (smaller spatial scales), ML models tend to drift from the reference spectra.
- Graph-FM shows significant drift, whereas Graph-EFM mainly lose structures at the finest-scales.



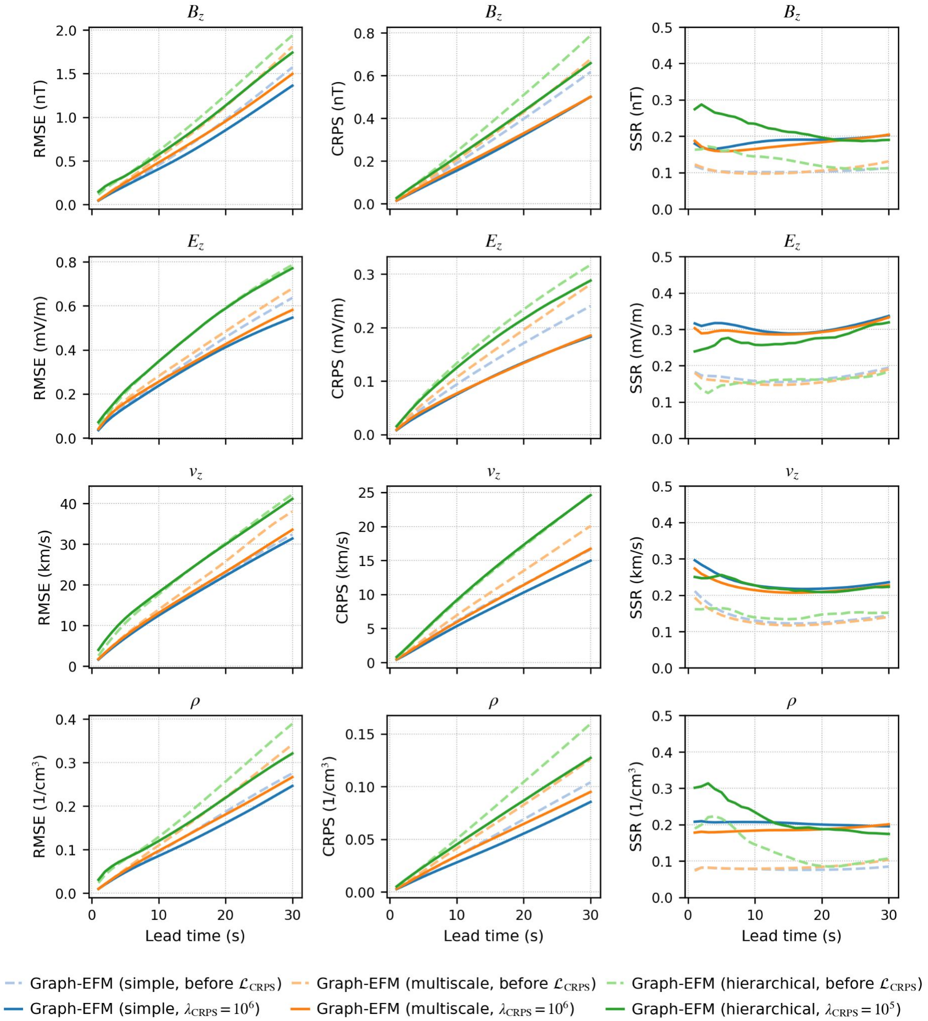
Outlook

- Autoregressive models produces a cumulative error, and smoothening for long rollouts.
- Long sequence of training data (≥ solar cycle) and larger step sizes could circumvent that issue.
- Terrestrial weather progress benefit from decades of openly available reanalysis, i.e. simulation with assimilated observations. Similar setup for space weather to enable data-driven forecasting?
- Graph-based models well-suited also for refined grids such as Vlasiator in 3D.
- Adapt to heliospheric model like WSA-ENLIL or EUHFORIA?

Preprint

Code + data

Effect of CRPS Finetuning



Ensemble size comparison