Prototype-enhanced prediction in graph neural networks for climate applications

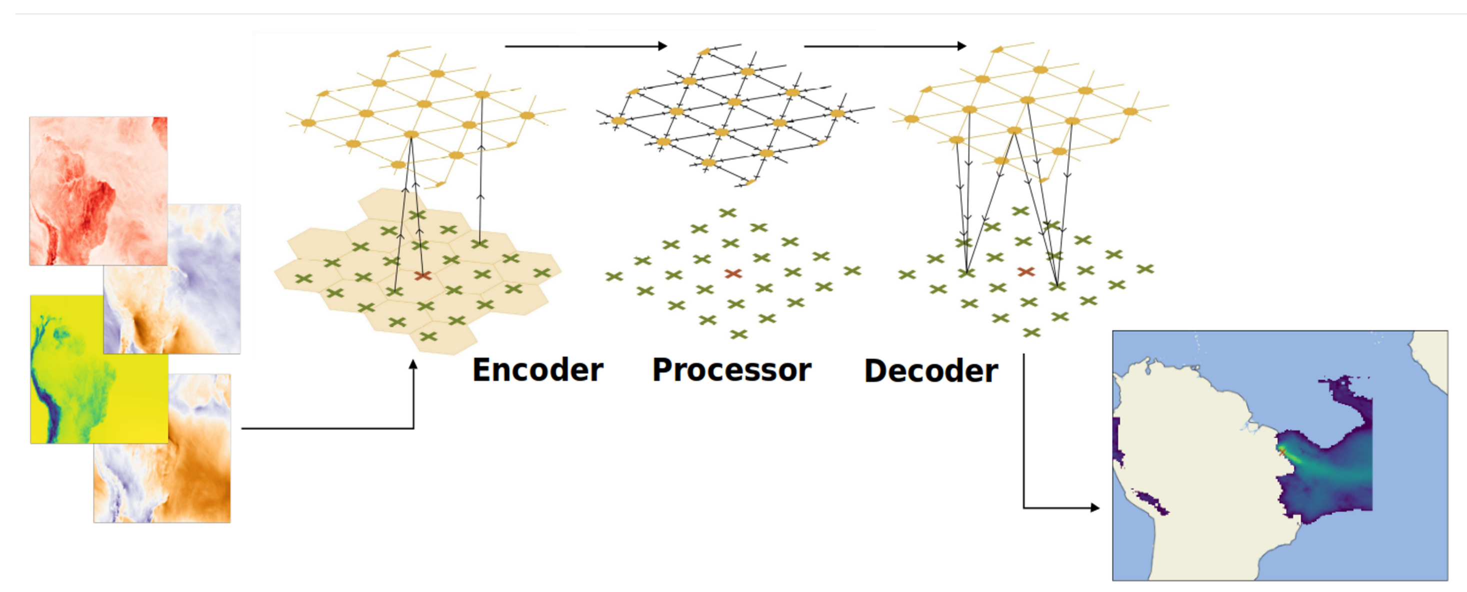
N.Keshtmand, E.Fillola, J.N.Clark, R.Santos-Rodriguez, M.Rigby
University of Bristol
yl18410@bristol.ac.uk

Motivation

- Earth system modeling often relies on complex, physics-based simulations that are computationally expensive. To address this, machine learning (ML) models are increasingly used to accelerate or replace these simulations.
- A key application is monitoring greenhouse gas (GHG) emissions using atmospheric dispersion models, particularly Lagrangian Particle Dispersion Models (LPDMs). LPDMs simulate gas particles moving backward in time from a measurement site, producing a 2D footprint that identifies upwind areas contributing to observed GHG levels (Hegarty et al., 2013) [1].
- We introduce *prototypes*—approximations of emulator outputs provided as inputs—to guide emulators toward more accurate predictions. We demonstrate the effectiveness of this approach in the context of GHG monitoring.

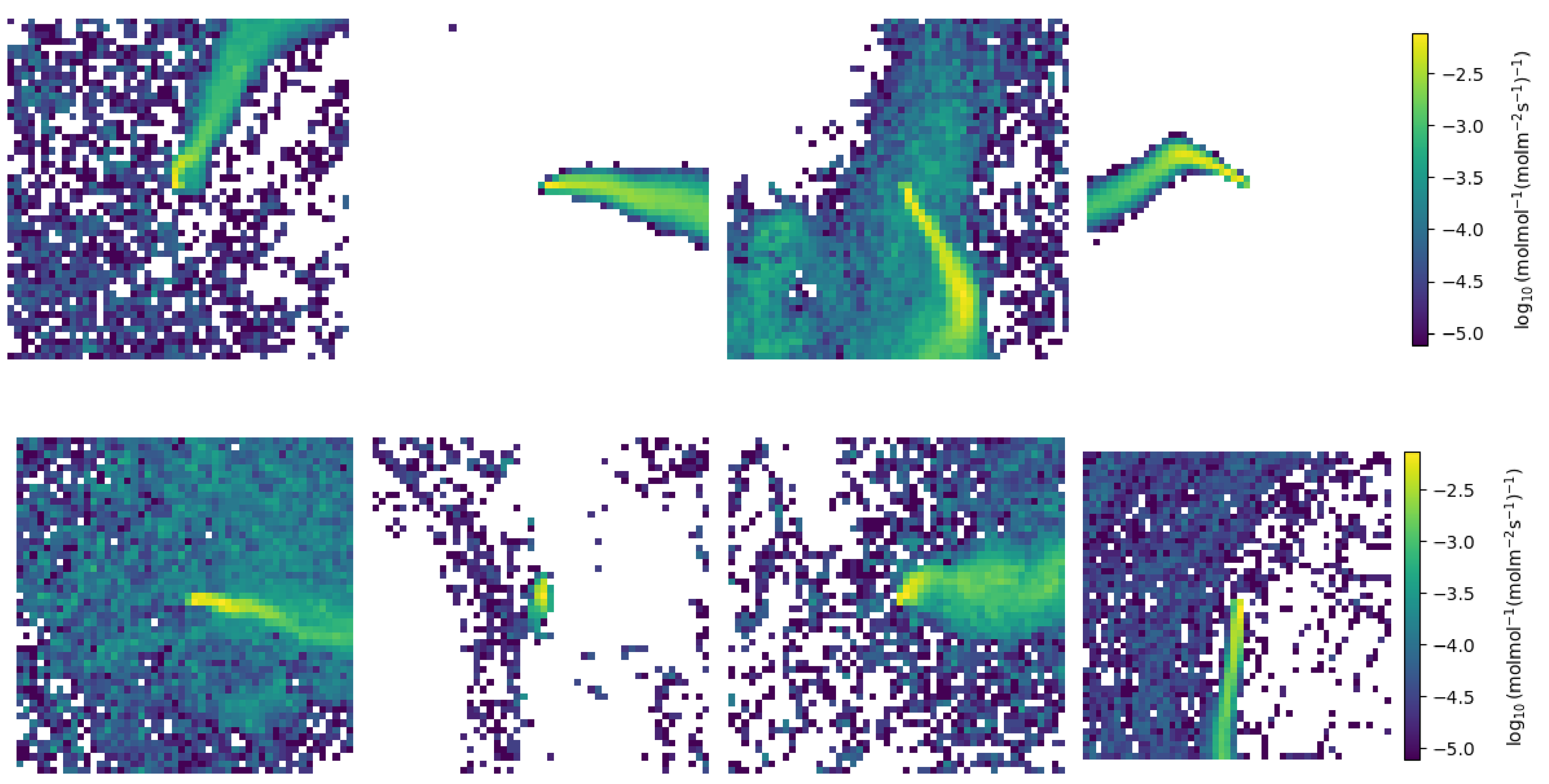
Method - Architecture

- Data- We use footprints produced by NAME, the UK's Met Office LPDM for GOSAT measurements (Greenhouse gases Observing SATellite (Parker et al., 2020)) over Brazil for 2014-16 [2].
- We adopt the Encoder-Processor-Decoder architecture from Fillola et al. (2023) to predict footprints, using meteorological and topographical inputs on a latitude-longitude grid [3]. A prototype footprint is added as an additional input.



Method - Prototype Generation

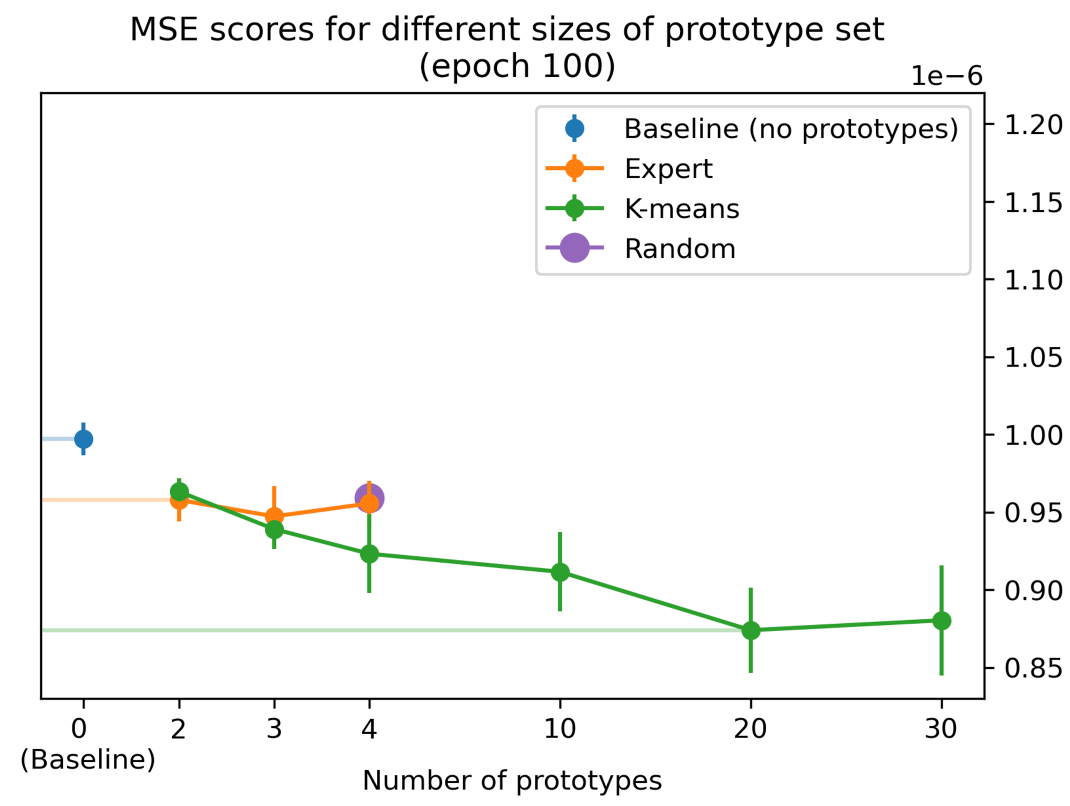
- The prototype—an inexpensive initial approximation of the output—can be generated using low-cost input-based calculations, artificial samples, or selected training examples.
- We select n footprints from the training data as prototypes. Each training sample is assigned one prototype as an additional input to support prediction.
- We demonstrate two methods for selecting prototypes: (1) a human expert manually chooses n footprints to represent diverse conditions, such as upwind areas aligned with the four cardinal directions (top row); (2) a data-driven approach using k-means clustering, where the footprint nearest each cluster center is selected as a prototype (bottom row).



Method - Prototype Selection

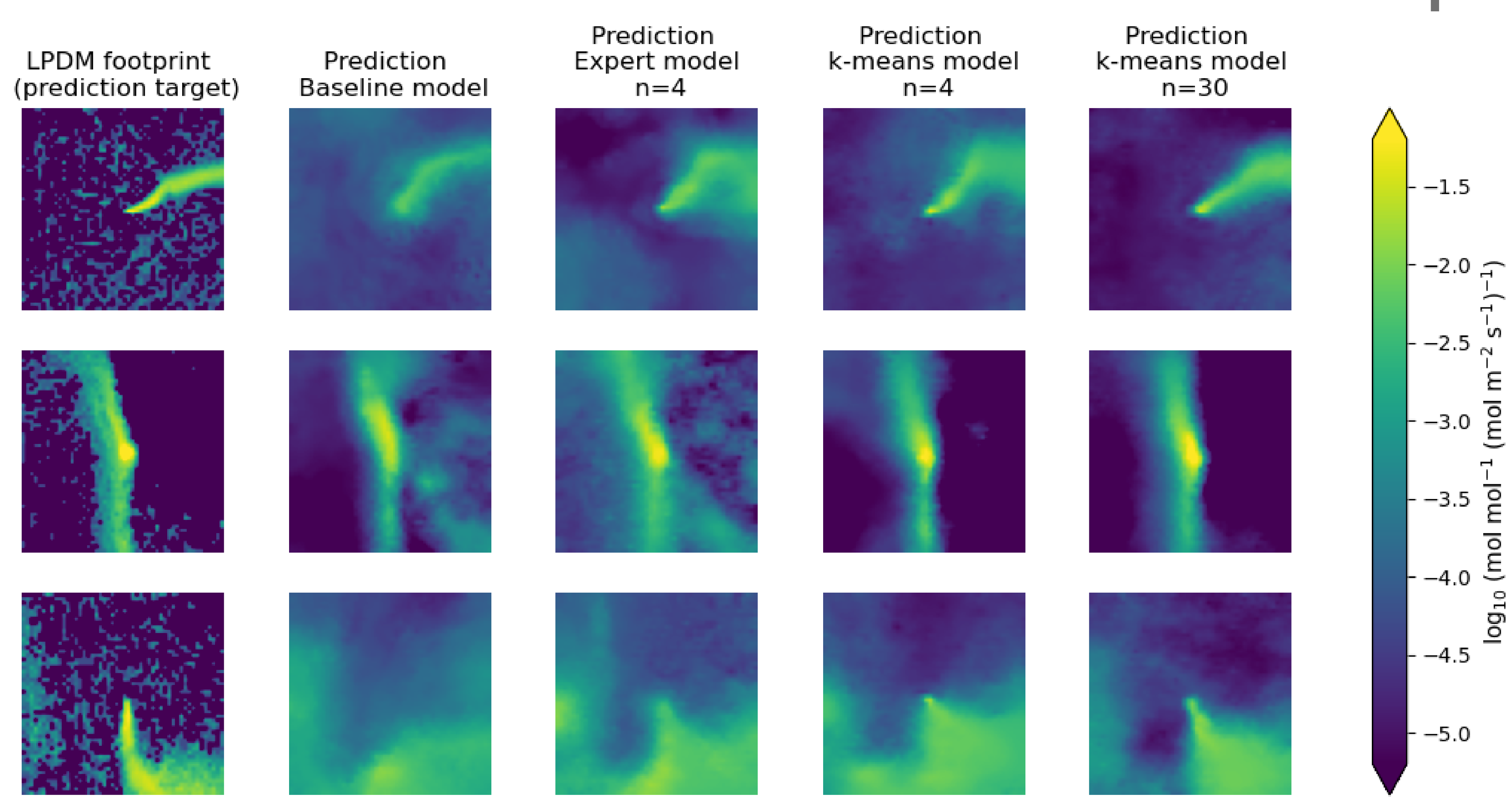
- To isolate the impact of prototypes, we present an oracle case where true footprints are available for prototype assignment.
- Each footprint is matched with the closest prototype based on L_2 distance in a 64-dimensional PCA space.
- In practice, since the true footprint is unknown, prototypes must be selected from inputs (e.g., via a classifier).

Quantitative Results



- We compare the baseline model (0 prototypes, as in Fillola et al. (2023)) [3] with our prototype-based approach using varying numbers of prototypes (n), selected either by an expert or via k-means. Performance is measured by mean squared error between predicted and true footprints (lower is better).
- Using at least two prototypes—regardless of selection method—improves over the baseline. For k-means, performance generally improves as n increases from 2 to 20.

Qualitatitve Results



- The baseline model without prototypes captures the main upwind direction of the true footprint but produces overly smoothed outputs that miss higher values. In contrast, models using prototypes generate sharper footprints that better capture wind direction and high-intensity regions.

Future work

- Developing a deployment pipeline including prototype classifiers and exploring how they impact the performance.
- Exploring applications of prototypes in other physical emulators/domains.
- Improve performance.

References

1. J. Hegarty *u. a.*, *Journal of Applied Meteorology and Climatology* **52**, 2623–2637 (2013).
2. R. J. Parker *u. a.*, *Earth System Science Data* **12**, 3383–3412 (2020).
3. E. Fillola *u. a.*, **presentedat**NeurIPS 2023 Workshop on Tackling Climate Change with Machine Learning, (`https://www.climatechange.ai/papers/neurips2023/74`).