Towards emulated Lagrangian particle dispersion model footprints for satellite observations

Elena Fillola[1], Raul Santos-Rodriguez[1], Matt Rigby[2]
1. Department of Engineering Mathematics, University of Bristol
2. Atmospheric Chemistry Research Group, University of Bristol
elena.fillolamayoral@bristol.ac.uk

Background: Atmospheric Transport Models

- Inverse modelling systems for **greenhouse gas (GHG) flux quantification** require an atmospheric transport model and a statistical inversion framework to produce spatial estimates of methane
- Backwards running Lagrangian particle dispersion models (LPDMs) produce ***footprints:*** sensitivity of a measurement to emissions in the area around it
- The computational cost of the LPDMs means that the growing volume of satellite GHG measurements is underutilized
- The small number of existing methods to efficiently approximate LPDM footprints use interpolation in time and space, or smoothing of low-resolution runs

Goal: Met-only Emulator

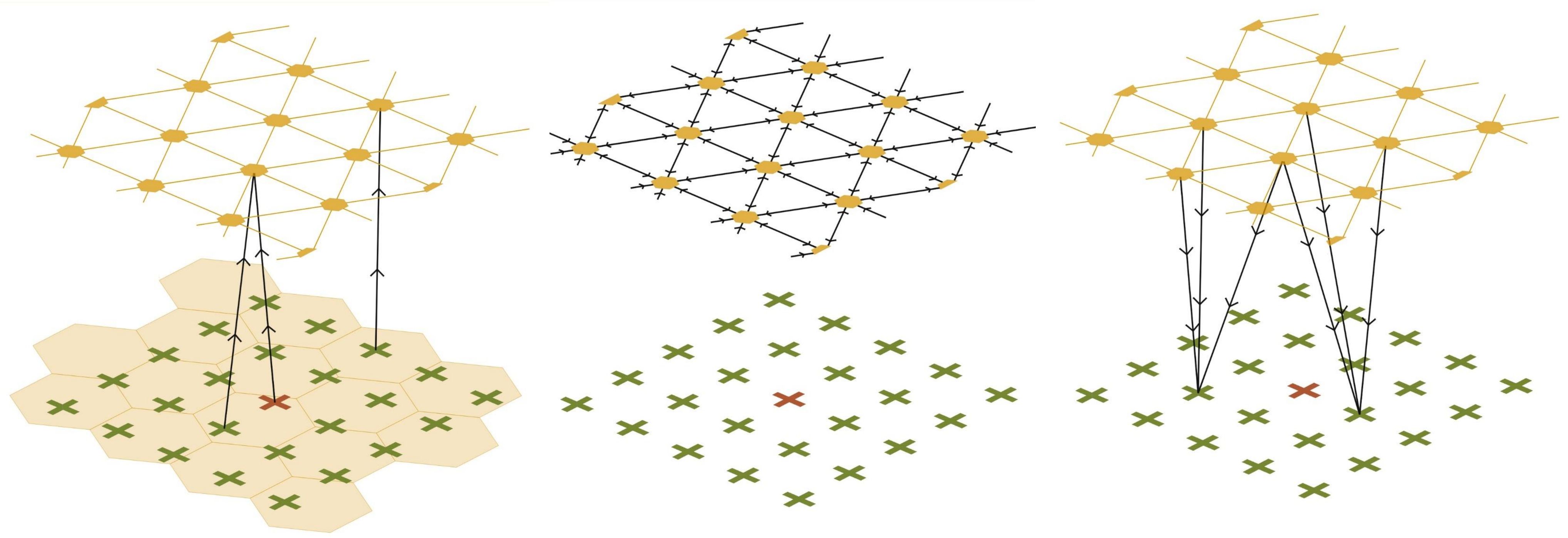
To develop a machine learning LPDM emulator, which can output full-domain footprints for satellite measurements using only meteorological and geographical inputs

Proof of Concept: Simple Emulator for In Situ Data (Fillola et al., 2023)

- Proof-of-concept model with simpler task and simpler design – **footprints for tower site data**
- Footprint value at each cell modelled with independent gradient-boosted regression trees, for a grid centered around the measurement point.
- One model trained per site, using 2-hourly meteorological fields from 2014-2015
- Prediction speed-up of five orders of magnitude
- R^2 scores for CH_4 concentration prediction across sites of 0.69 for 2016
- Limitations: not scalable with domain size

Graph Networks to Emulate Satellite Data

The emulator consists of Graph Networks in a **Encode-Process-Decode** architecture that combines the native data space, a latitude-longitude grid, with an abstract icosahedron mesh. The two grids are always centered around the measurement point and therefore the physical location of the lat-lon grid changes for every sample. Black arrows indicate the flow of information between nodes.



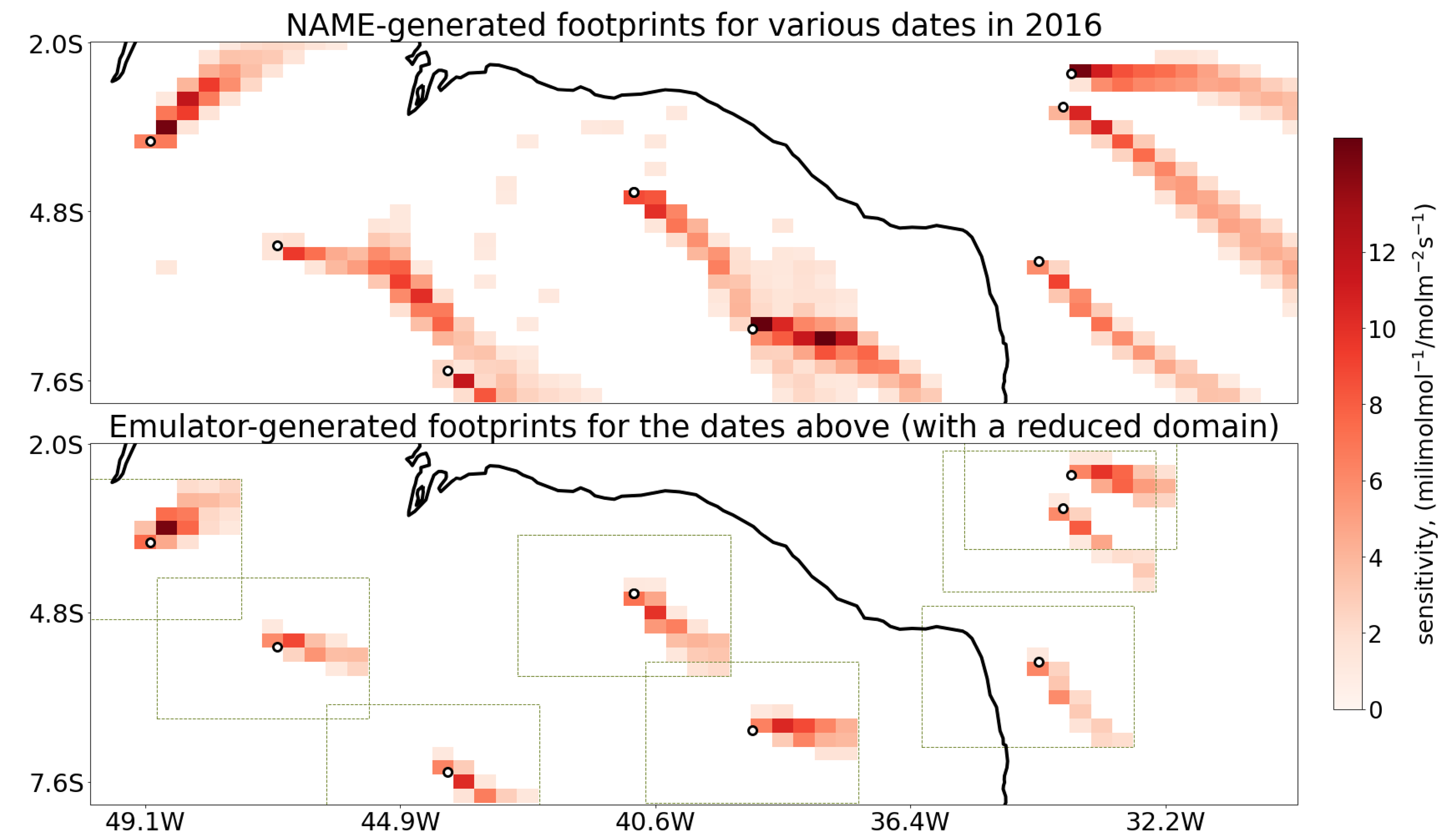
Encoder: meteorology and geographical data at each node in the lat-lon grid is encoded in the closest node of the icosahedron mesh

Processor: a message-passing GNN updates the information in the icosahedron mesh iteratively, spreading the information from each node to its neighbours

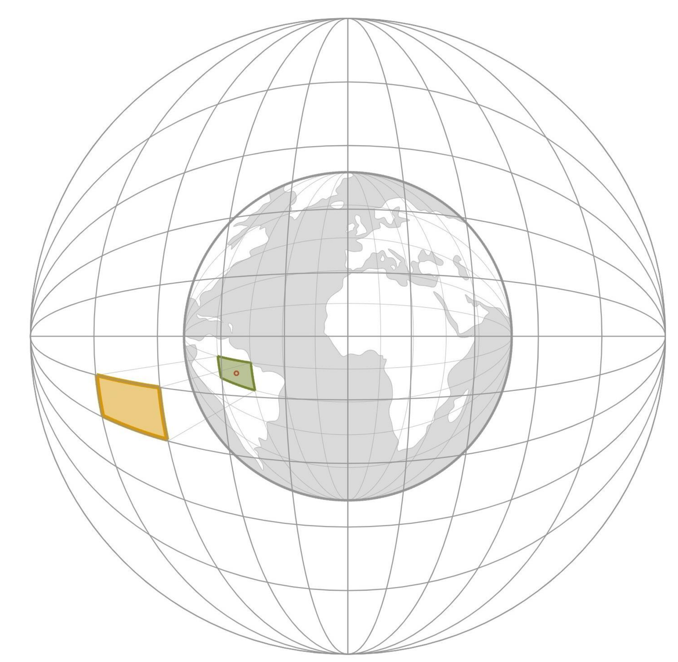
Decoder: the value of the footprint at each point in the lat-lon grid is predicted with the three nearest nodes in the icosahedron mesh

Example Application: GOSAT over Brazil

- Demonstrated for GOSAT measurements over Brazil
- Inputs: Met Office Unified Model fields and geographical information for 2014-2015 to predict NAME satellite footprints
- Reduced domain – 10x10 latitude-longitude grid
- R^2 score for CH_4 concentration prediction over Brazil of 0.60 for 2016



Sample Results: NAME-generated and emulator-generated footprints for GOSAT measurement locations over North-East Brazil at various dates in 2016. Dotted line shows the emulated area around each point



Left - **Diagram of the two data spaces,** centered around the measurement point

Next Steps

- Increase prediction domain
- Analysis of error distribution
- Transfer learning across world regions
- Integration with inverse modelling system to calculate methane fluxes

References and Acknowledgements

Fillola, Elena et al. "A machine learning emulator for Lagrangian particle dispersion model footprints: a case study using NAME." *Geoscientific Model Development* (2023)
Battaglia, Peter et al. "Relational inductive biases, deep learning, and graph networks" arXiv:1806.01261 (2018)

This research has been supported by the Google PhD Fellowship and was carried out using the computational facilities of the Advanced Computing Research Centre at the University of Bristol.
Figures co-created with Daniel Martos Herrer.