



*Supplement of*

## **A machine learning emulator for Lagrangian particle dispersion model footprints: a case study using NAME**

**Elena Fillola et al.**

*Correspondence to:* Elena Fillola ([elena.fillolamayoral@bristol.ac.uk](mailto:elena.fillolamayoral@bristol.ac.uk))

The copyright of individual parts of the supplement might differ from the article licence.

## SECTION S1

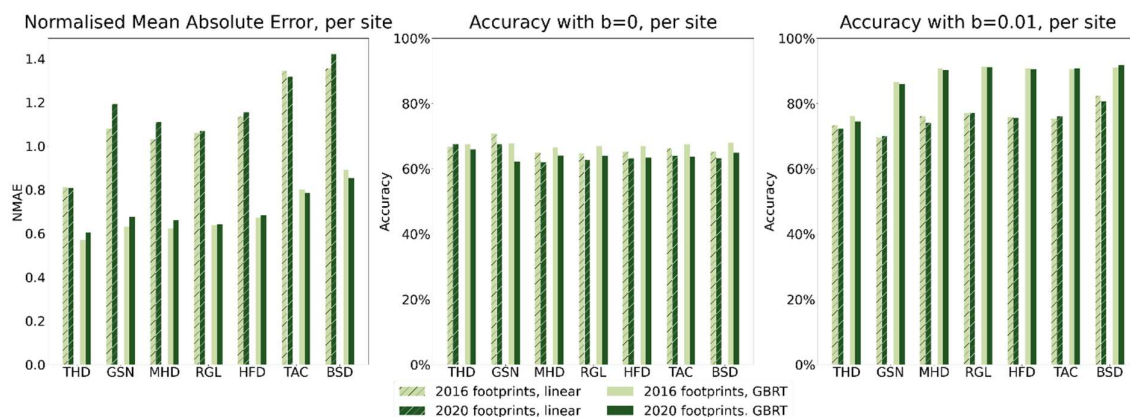
We compare the performance of the emulator against a baseline ridge linear model, using the same training and testing data and the same cell-by-cell approach. Ridge regression is a type of regularised linear regression, where an L2 penalty term is added to the loss function to shrink the regression coefficients, aiming to reduce the impact of collinearity in the data. The amount of shrinkage is controlled by a parameter lambda. See Hastie et al. (2001, Section 3.4.1) for more on ridge regression.

The ridge regression model is tuned in a similar fashion to the GBRT model (see section 3.5), and we find that a lambda value of 1 is most fitting. We evaluate the footprints output by the linear emulation in the same way as the GBRT-generated ones. The results for the whole dataset are summarised in table S1, and figure S2a and S2b shows the evaluation disaggregated by site (the GBRT results are shown as well in figures 3 and 5 respectively). These results demonstrate, in particular through the NMAE, that the GBRT model has far higher predictive skills than a linear model.

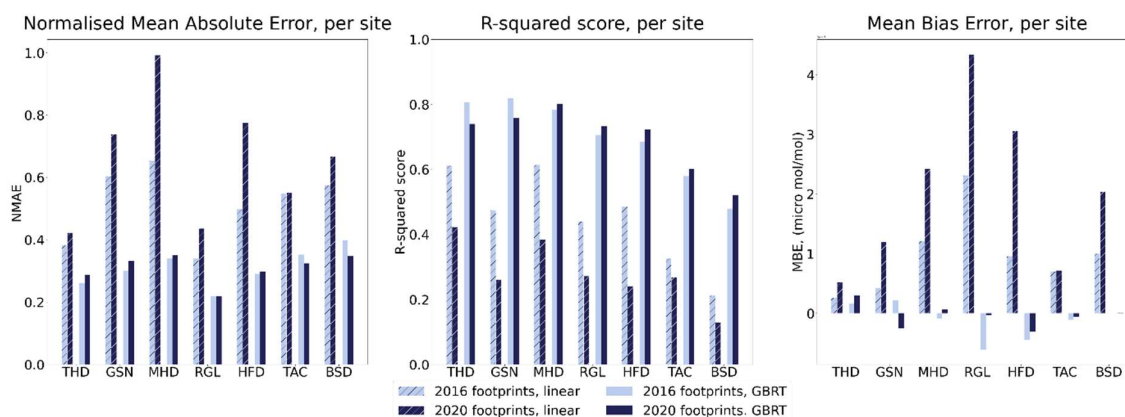
Metric	GBRT model		Ridge regression model	
	2016	2020	2016	2020
NMAE (footprints)	0.689	0.701	1.12	1.15
Accuracy with b=0	67.3%	64.0%	66.3%	64.4%
Accuracy with b=0.01	88.1%	87.8%	75.7%	75.1%
NMAE (mole fraction)	0.308	0.308	0.514	0.655
R-squared score	0.694	0.697	0.451	0.283
Mean Bias Error (micro mol/mol)	-0.0125	-0.0043	0.0978	0.204

**Table S1.** Average metric results for 2016 and 2020 across all seven sites tested, for the GBRT model (the emulator described in the manuscript) and a ridge regression model (a baseline linear model).

a)



b)



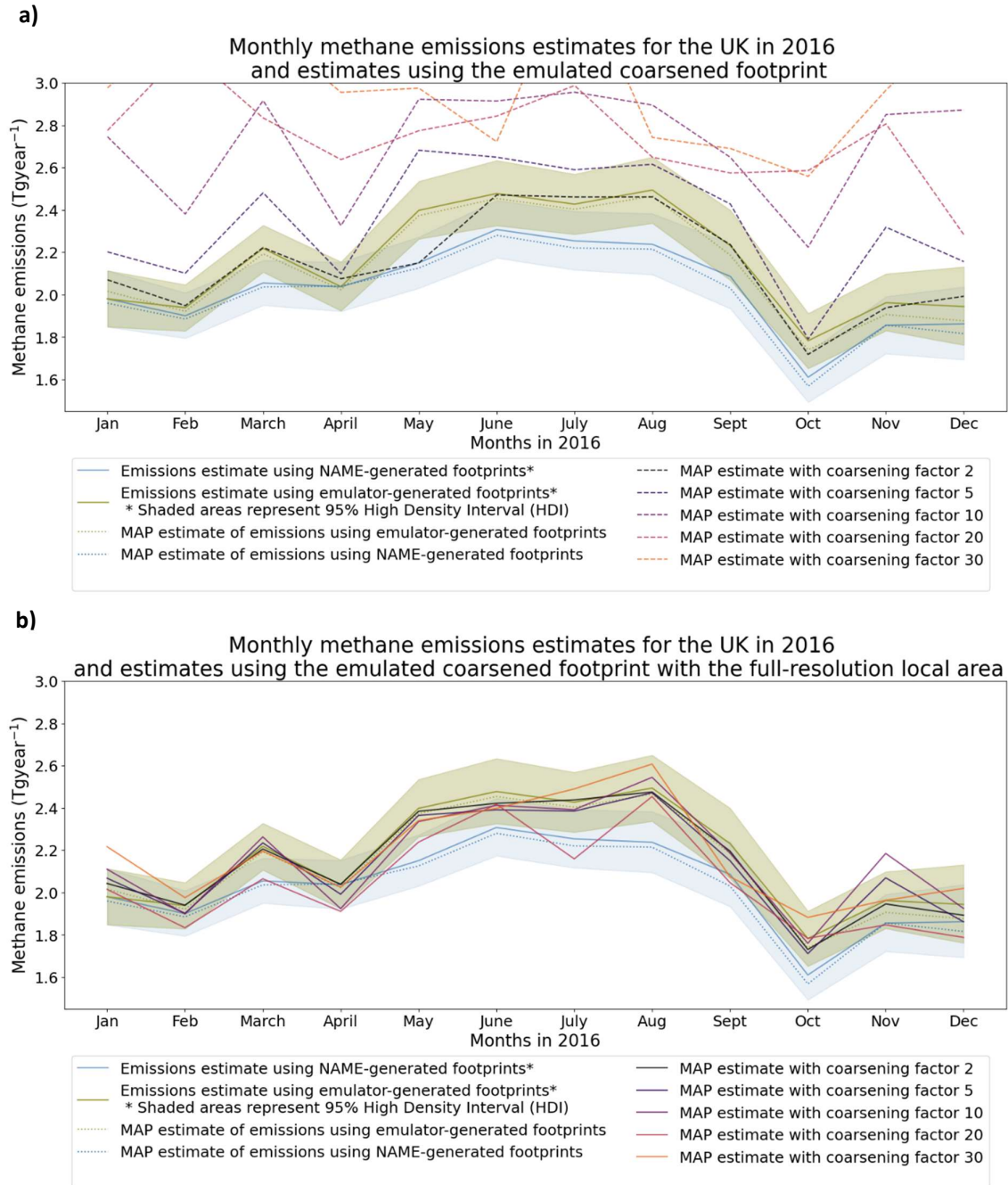
**Figure S2.** Evaluation of emulators, per site and per year, for the GBRT model and a baseline ridge linear model. a) shows footprint-to-footprint comparison, using metrics NMAE and accuracy with  $b=0$  (all footprint values) and  $b=0.01$  (high values), and b) shows mole fraction comparison, using metrics NMAE, R-squared score and MBE.

## SECTION S2

We conduct a sensitivity test to demonstrate the importance of the emulated area (the local region around the measurement point) in comparison to the rest of the domain by calculating emission estimates with coarsened footprints, simulating lower resolution runs of the LPDM. We coarsen the emulated footprints (which consist of the local emulated area of size 10x10 and the NAME-generated data in the rest of the domain) by dividing the image into independent windows of size  $F \times F$  cells where  $F$  is the coarsening factor, and in each window we replace the value of all cells with the mean.

Figure S3a shows the results of running the inversion with emulated footprints coarsened to different factors throughout the entire domain. Figure S3b shows the inversion estimates with the coarsened emulated footprints, but the local area preserved at full resolution. For efficiency, we obtain the emission estimates using the maximum a posteriori (MAP) probability rather than running the full Markov chain Monte Carlo (MCMC) sampling, but we show that the two are very similar compared to the a posteriori uncertainty by generating the MAP estimates for the full resolution footprints (shown as dotted lines in figure S3) and comparing them to the MCMC estimates.

The mean percentage absolute error between the emissions estimate inferred using full-resolution emulator-generated footprints and the estimate using the coarsened footprint is around 10% for a coarsening factor of 5, and over 40% for a coarsening factor of 30. In comparison, the error for the coarsened footprint with the emulated area at full resolution present errors of under 5% for all coarsening factors. This indicates that the inversion is highly sensitive to a loss of fidelity in the footprints within our emulated region, but substantially less sensitive outside of this region. We propose this test provides an indication that our inversion results should be relatively insensitive to substantial uncertainties in footprint magnitude outside of the emulated regions.



**Figure S3.** Monthly methane emission estimates (as shown in Fig. 7 in the manuscript) compared to emission estimates from coarsened footprints, without the full-resolution emulated area (a) and with it (b), calculated with Maximum A Posteriori (MAP). In both graphs, the MAP emissions estimates for the NAME-generated and the emulator-generated footprints are shown as dotted lines, to demonstrate that the MAP approach and the MCMC approach are equivalent. The inversion is performed for the emulated footprints coarsened to five different coarsening factors (where coarsening means dividing the image into independent windows of size  $F \times F$  cells where  $F$  is the coarsening factor, and in each window replacing the value of all cells with the mean).