IMPERIAL

Engineering and Physical Sciences Research Council

AI4Urban project website

Machine-learning based variational data assimilation for unknown emission source identification in urban environments

Linfeng Li, Judongyang Zhou, Jie Zheng, Kechao Lu, Fangxin Fang

Department of Earth Science and Engineering, Imperial College London

Introduction

Urban pollutants transport affected by:
- Urban boundary layer flow
- Emission sources

To identify emission sources, inversion method (data assimilation) is used
- Given observed mixing ratio at locations
- Given (or modelling) meteorological conditions
- Find emission location and rate

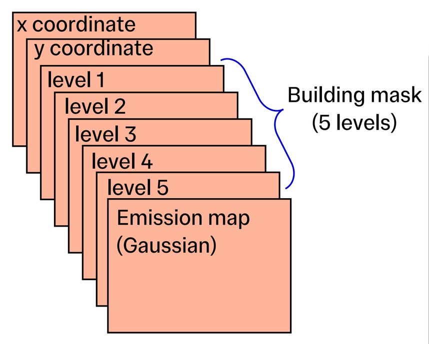
Physics-based methods:
- Based on numerical methods;
- Obeying physical conservations;
- Can be slow

Data-driven method:
- Fast
- Automatic-differentiation (no need to code adjoint)

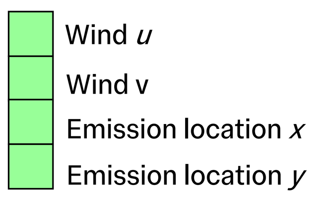
Forward model

U-Net with FiLM condition modulation

Input: (B, 8, 200, 200)



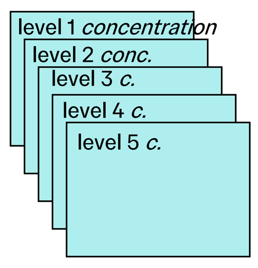
Condition: (B, 4)



U-Net Architecture

Stage	Layer	Output shape	Params
Input	–	$8 \times 200 \times 200$	–
Encoder	DoubleConv (32)	$32 \times 200 \times 200$	16,064
	MaxPool2d	$32 \times 100 \times 100$	–
	DoubleConv (64)	$64 \times 100 \times 100$	64,064
	MaxPool2d	$64 \times 50 \times 50$	–
Bottleneck	DoubleConv (128)	$128 \times 50 \times 50$	238,400
Decoder	ConvTranspose2d (64)	$64 \times 100 \times 100$	32,832
	DoubleConv (64)	$64 \times 100 \times 100$	119,360
	ConvTranspose2d (32)	$32 \times 200 \times 200$	8,224
	DoubleConv (32)	$32 \times 200 \times 200$	32,192
Output	Conv2d (1×1)	$5 \times 200 \times 200$	165

Each DoubleConv consists of:
- 1st convolution (in_ch , out_ch , kernel_size=3, padding=1)
- A FiLM modulating layer to inject condition — *FiLM: **F**eature-w**i**se **L**inear **M**odulation*
  - MLP: 4 -> 64 -> 2 * out_ch
  - Then apply to each channel: $\alpha \cdot x + \beta$
- 2nd convolution (out_ch , out_ch , kernel_size=3, padding=1)



Output: (B, 5, 200, 200)

Concentration fields on five layers

Loss function (weighted MSE loss) (weighted inversely with pixel value frequency)

Ground truth concentration fields

Training for 20,000 epochs: Training MSE error ~7e-5, Validation MSE error ~3e-4 (normalised log10 ppm)

Inversion model

Bayesian formulation and MAP estimation

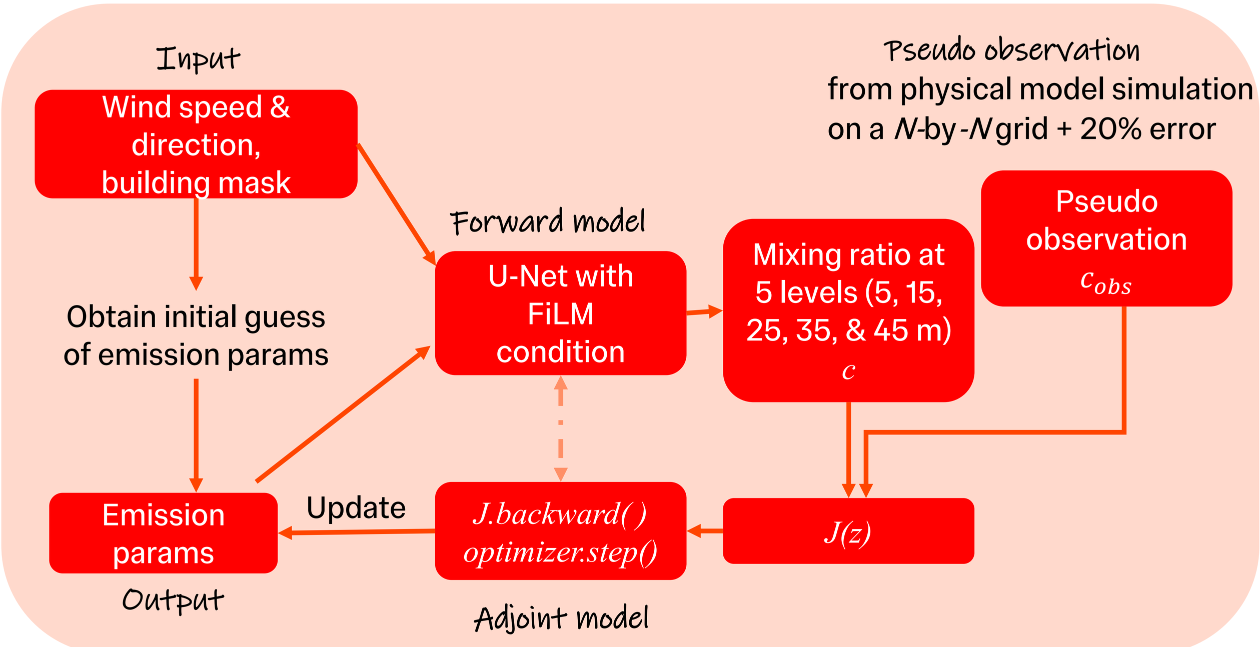
- Background error: uniform distribution
  - Assuming no prior information
- Observation error: Gaussian distribution
  - Assuming all obs points' errors are independent and at same level (20% ppm)
- Using max a posterior estimation, **cost function**

$$J(z) = \frac{1}{2}\left(\log_{10}(c_{obs}) - \log_{10}\left(\boldsymbol{H}(\boldsymbol{M}(z))\right)\right)^T \mathbf{R}^{-1}\left(\log_{10}(c_{obs}) - \boldsymbol{H}(\boldsymbol{M}(z))\right)$$

where $z = [x_e, y_e, r_e]^T$ $\boldsymbol{M}$: forward model $\boldsymbol{H}$: obs operator

Inversion process

- Selection of pseudo-observations:
  - A grid of N-by-N obs points on 35-m level
  - 20% random perturbation (error) added to pseudo-obs.
- Initial guess of z : (x_e, y_e, r_e)
  - Trial of a series of candidates
  - select the one with minimal J as initial guess
- Optimisation method:
  - `torch.optim.Adam`
  - Learning rate = 1
  - Call `J.backward()` for gradient
  - `optimizer.step()` for updating parameters



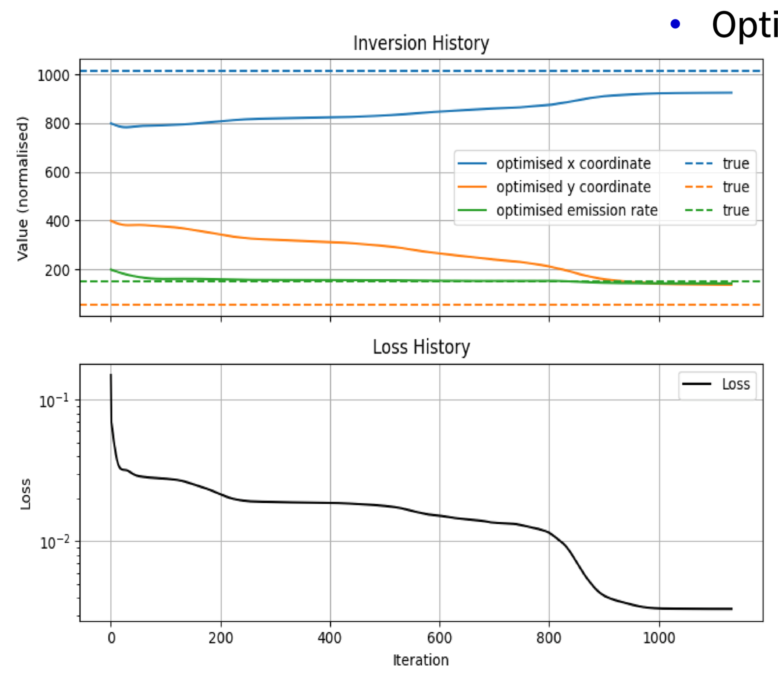
Results

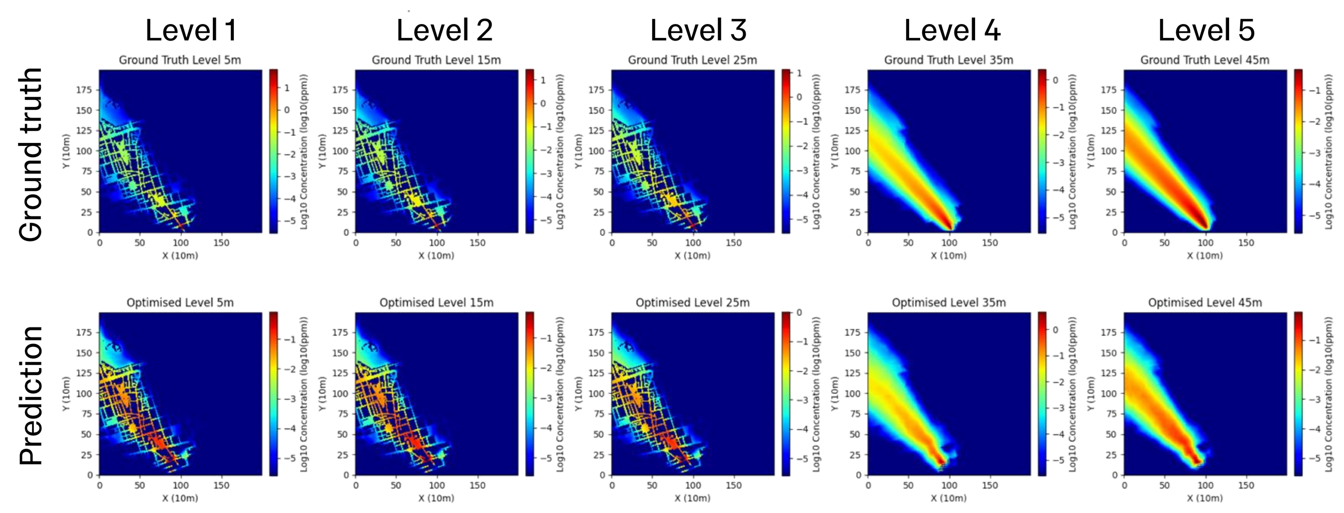
Inversion history of a particular case

Single emission source

- 10 m/s SE wind, single emission location
- Ground truth: location (1015, 55), 150 ton/yr
- Optimised: location (925.1, 137.3), 142.0 ton/yr

Using 16 ob pnts (on a 4x4 grid)



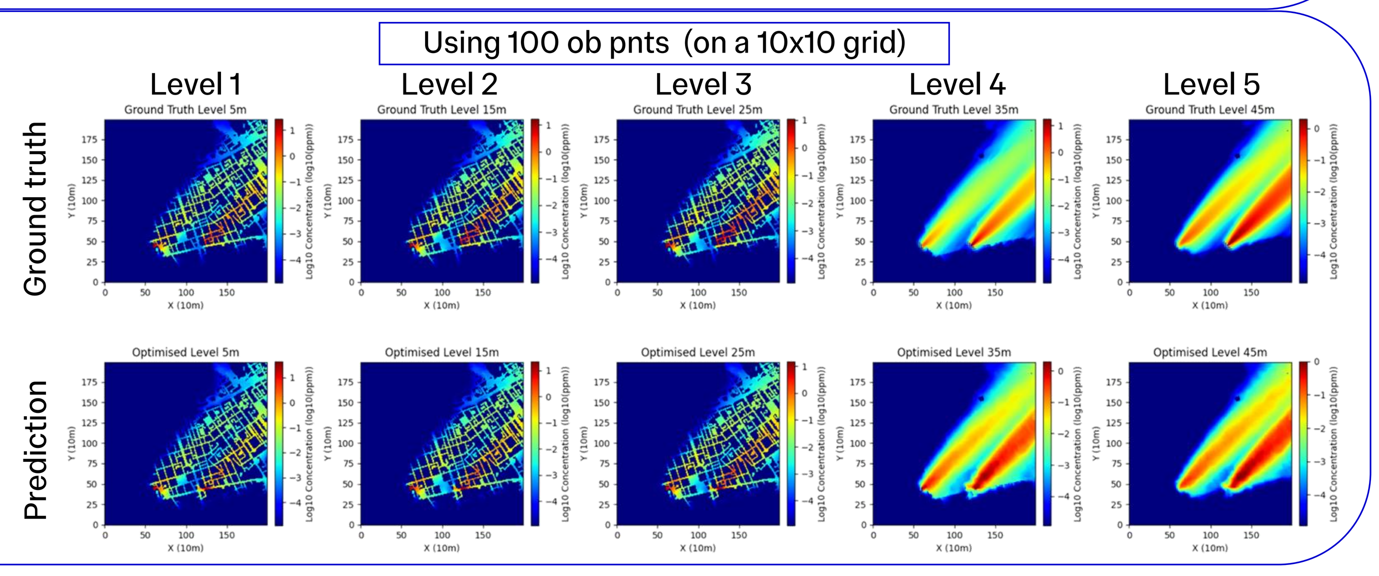


- Predicted concentration using optimised emission parameters
- All colour-bar using log10 (ppm) scale.

Double emission sources

Using 100 ob pnts (on a 10x10 grid)

- 10 m/s SW wind, two emission locations
- Ground truth:
  - location #1: (615, 455), 150 ton/yr
  - location #2: (1215, 455), 564 ton/yr
- Optimised:
  - location #1: (615.5, 453.8), 159.8 ton/yr
  - location #2: (1207.6, 431.5), 497.3 ton/yr



Statistics of all inversion test cases

Single source inversion

Statistics	Location error (m)	Emission rate error (%)
mean	242.94	4.53
std	283.23	12.17
min	0.12	0
25%	48.14	0.02
50%	160.26	0.82
75%	335.11	4.33
max	2401.35	145.88

Total number of test cases: 824

Double sources inversion

Statistics	Source 1 Location error (m)	Source 1 Location error (m)	Source 2 Location error (m)	Source 2 Location error (m)
mean	137.28	31.05	147.39	30.63
std	260.86	38.04	278.35	37.61
min	0.26	0	0.33	0.04
25%	25.95	7.48	26.27	6.90
50%	49.43	18.09	51.35	17.81
75%	144.16	41.23	146.97	41.23
max	2196.48	458.69	2360.80	364.14

Total number of test cases: 2000

Data used in this study

Simulated data is used in this study
- Study region: 2 km by 2 km by 500 m around BT tower in Camden, London
- Using urban LES model PALM
- Varying building height based on real OpenStreetMap data
- Releasing 1 single point source at each case
- 4-hr simulation (steady turbulence)
- a total of 4000 cases.

Wind direction
S, SW, W, NW, N, NE, E, SE

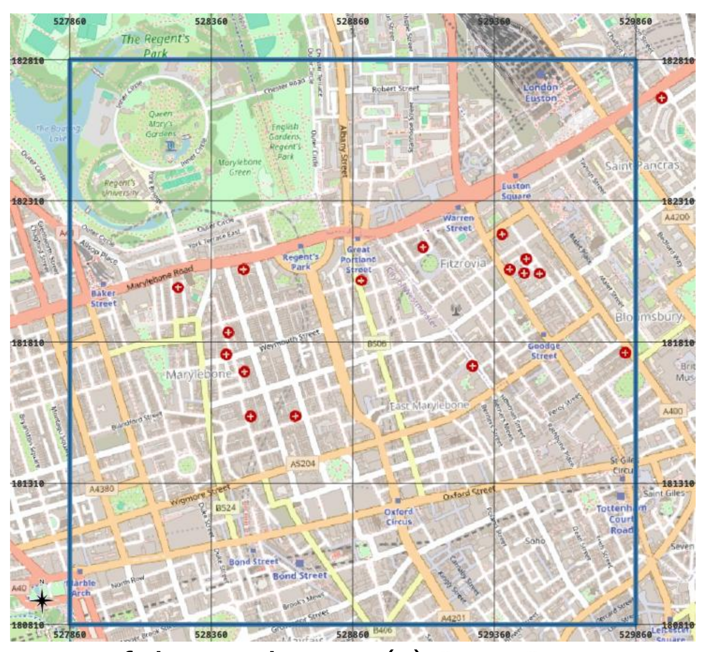
Wind speed
4, 6, 8, 10, 12 m/s

(Log profile is used, numbers here is wind speed at 190 m height)

Emission location
x = 15, 215, ..., 1815
y = 55, 255, ..., 1855
A total of 100 locations

Emission rate
Fixed at 564 ton/yr

(For other emission rates, linearity of transport eq. can be used to derive concentration)



Map of the study area (c) OpenStreetMap

Reference:
- (C) OpenStreetMap Distributed under the Open Data Commons Open Database License (ODbL) v1.0 (for further information, please see https://www.openstreetmap.org/copyright/en)]
- PALM: Maronga, B., et al. (2020). "Overview of the PALM model system 6.0." Geosci. Model Dev. 13(3): 1335-1372.
- U-Net: Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Cham: Springer international publishing, 2015.
- FiLM: Perez, E., et al. "Visual reasoning with a general conditioning layer, Courville." In Proceedings of the AAAI Conference on Artificial Intelligence. 2017.

Imperial College London

Email: l.li20@imperial.ac.uk (L. Li)