Preprints of Control Conference Africa 2024
Balaclava, Mauritius, 16-17 September 2024

A RBF-FD Physics-Informed Machine Learning Approach to Air Pollution Source Estimation

Didier Georges * **Sylvain Leirens** ** **Roman Lopez-Ferber** **

* *Univ. Grenoble Alpes, CNRS, Grenoble INP (Institute of Engineering and Management), GIPSA-lab, F-38000 Grenoble, France, (e-mail: didier.georges@gipsa-lab.grenoble-inp.fr)*
** *Univ. Grenoble Alpes, CEA, Leti, F-38000 Grenoble, France, (e-mail: sylvain.leirens@cea.fr, roman.lopezferber@gmail.com)*

Abstract: In this paper, we propose a source term estimation approach for air pollution monitoring based on a physics-informed machine learning approach using radial basis function-generated finite differences (RBF-FD) approximations, rather than using neural network-based approximations. This approach looks promising for detecting a static pollution source, at a particularly low computing cost and based on a network of fixed or mobile sensors. A 3D case study demonstrates the effectiveness of the approach.

Keywords: air pollution, source term estimation, advection-diffusion PDE, physics-informed machine learning, RBF-FD approximation.

1. INTRODUCTION

Air pollution has become a pressing societal and research concern in recent years, given its significant impact on public health, environmental quality, and overall urban livability (World-Health-Organization (2020)). As urbanization continues to accelerate globally, the concentration of pollutants in urban atmospheres rises, posing serious threats to human health and well-being, as well as ecosystems and infrastructure (United-Nations-Environment-Programme (2019)).

The identification of pollution sources in urban areas is vital for several reasons. Firstly, urban populations are particularly vulnerable to the adverse effects of air pollution due to the high density of emissions sources, such as vehicles, industrial facilities, and residential heating systems. Exposure to pollutants like particulate matter (PM), nitrogen oxides (NOx), sulfur dioxide (SO2), volatile organic compounds (VOC), and ozone (O3) has been linked to a myriad of health issues, including respiratory and cardiovascular diseases, cancer, and neurological disorders. Secondly, urban air pollution not only affects human health but also has far-reaching environmental implications. Pollutants released into the atmosphere can undergo complex chemical reactions, leading to the formation of secondary pollutants and exacerbating issues like smog, acid rain, and ozone depletion (United-Nations-Environment-Programme (2019)). Additionally, airborne contaminants can deposit onto surfaces, contaminating soil, water bodies, and vegetation, and disrupting ecosystems and biodiversity. Furthermore, air pollution in urban areas contributes significantly to climate change, as certain pollutants, such as carbon dioxide (CO2) and methane (CH4), act as greenhouse gases, trapping heat in the atmosphere and exacerbating global warming. The detection and mitigation of atmospheric pollution sources in urban environments have thus emerged as crucial endeavors, garnering substantial attention from both researchers and policymakers.

In recent years, advancements in sensor technologies, data analytics, and modeling techniques have revolutionized the field of atmospheric pollution detection and monitoring in urban environments. Traditional monitoring networks, comprised of fixed-site monitoring stations, provide valuable data but are limited in spatial coverage and resolution. To overcome these limitations, researchers have increasingly turned to innovative approaches, such as mobile monitoring platforms, unmanned aerial vehicles (UAV), and low-cost sensor networks, to capture fine-scale variations in pollutant concentrations across urban areas.

Source term estimation (STE) has been extensively studied in the literature, especially in highly constrained urban areas. Many authors studied model-based estimation methods to perform urban STE such as ensemble Kalman filter Defforge et al. (2021), particle filter Septier et al. (2020), or variational calculus Kumar et al. (2015), but at the price of computationally expensive models requiring hours of computation. At the same time, the use of machine learning techniques in air pollution monitoring is growing exponentially (Li et al. (2023)).

Among numerical approaches for solving partial differential equations (PDE), the RBF-FD approach has proven to be effective in the past 10 years. Flyer et al. first developed a RBF-FD method for solving shallow water equations defined on spherical domains and showed that the RBF-FD method can be faster than discontinuous Galerkin methods with equivalent accuracy (Flyer et al. (2012)). Various ap-

⋆ This work was supported by the French Alternative Energies and Atomic Energy Commission.

plications in geosciences (Mathews et al. (2022)), biomedical engineering (Tominec et al. (2022)), and epidemiology (Oliver et al. (2022)) can be found in the recent literature.

In Lopez-Ferber et al. (2024), the authors propose to use the RBF-FD approach to solve direct and adjoint-based STE problems using a 3D advection-diffusion PDE in steady state in an heterogeneous domain. In this paper, we propose and evaluate on a simple case study a different STE approach based on a physics-informed machine learning (PIML) approach using RBF-FD approximations, rather than using neural network-based approximations (Raissi et al. (2019)). This approach looks promising for detecting a static pollution source, at a particularly low computing cost and based on a network of fixed or mobile sensors.

The paper is now organized as follows: Section 2 is devoted to air pollution modelling. Section 3 presents the RBF-FD PIML approach proposed in the paper. Section 4 is dedicated to a STE approach as an extension of the RBF-FD PIML formulation described in Section 3. A case study is presented in Section 5, and Section 6 sums up conclusions and perspectives.

2. AIR POLLUTION MODELLING

At the urban scale, air pollution modeling usually relies on two main assumptions:

(1) The fluid (air) is considered as incompressible.
(2) The air density is not locally changed by the pollutant.

Consequently pollutant dispersion can be well represented by Navier-Stokes equations coupled with an advection-diffusion partial differential equation (ADPDE).

In this paper, additional realistic assumptions are made:

- The mean wind velocity and mean diffusion fields are assumed to be known (for example, from meteorological data or models).
- The source term is considered stationary in time, and emits long enough to lead to a steady state.
- Pollutants are considered passive in this study (e.g. particle matter such as PM2.5 and PM10), without any further chemical reaction or photo-reaction.

Under these assumptions, the pollutant dispersion can be modelled by a 3D stationary ADPDE including a source term (see Zannetti (1990)) as follows:

$$\sum_{i=1}^{3} U_i(x)\partial_{\xi_i} u(x) = \sum_{i=1}^{3} \partial_{\xi_i}\left(K_i(x)\partial_{\xi_i} u(x)\right) + s(x), \quad (1)$$
$$\forall x \in \Omega,$$

$$\nabla u(x)\cdot\nu(x) = 0, x \in \Gamma_n,\ u(x) = 0, x \in \Gamma_d, \quad (2)$$

where $x = (\xi_1, \xi_2, \xi_3)$ denotes the vector of spatial Cartesian coordinates, $U(x) = (U_1(x), U_2(x), U_3(x))$ and $K(x) = (K_1(x), K_2(x), K_3(x))$ denote the mean wind velocity and the mean diffusion fields respectively, $u(x)$ [g m^{-3}] is the pollutant concentration, $s(x)$ is the pollutant source term [g m^{-3} s^{-1}], and $\nu(x)$ denotes the normal vector pointing towards the exterior of the boundary. ∂_{ξ_i} is the partial derivative w.r.t. ξ_i, while ∇ is the spatial gradient in $\mathbb{R}^3$. Two kinds of boundary conditions are considered: Dirichlet conditions ($u(x) = 0$) and Neumann conditions ($\nabla u(x)\cdot\nu(x) = 0$).

Equations (1)-(2) can be written in operator form as:

$$\mathcal{L}u(x) = s(x), x \in \Omega,\ \mathcal{B}u(x) = 0, x \in \partial_\Omega, \quad (3)$$

where $\mathcal{L}$ is the differential operator:

$$\mathcal{L} = \sum_{i=1}^{3} U_i(x)\partial_{\xi_i} - \sum_{i=1}^{3} \partial_{\xi_i}\left(K_i(x)\partial_{\xi_i}\right), \quad (4)$$

defined on the domain Ω, and $\mathcal{B}$ is the boundary condition operator defined on the boundary domain $\partial_\Omega = \Gamma_d \cup \Gamma_n$.

3. RBF-FD PIML FOR SOLVING ADPDE

3.1 RBF-FD approximation

Radial basis functions (RBF) are known to exhibit excellent interpolation properties in high-dimensional space (see e.g. Iske (2004); Fornberg and Flyer (2015)). A continuous function $f : y \in \Omega \subset \mathbb{R}^n \mapsto f(y) \in \mathbb{R}$ can be locally approximated in Ω by a set of N_x RBFs ϕ centered on a finite set of nodes (called center points) $X = \{x_i\}_{i=1,\cdots,N_x} \subset \Omega$. Let $\hat{f}$ denote the approximation of f at any $y \in \Omega$, based on N_x radial basis functions as:

$$\hat{f}(y) = \sum_{i=1}^{N_x} a_i \phi(\|y - x_i\|) = \Phi(y)\underline{a},\ \forall y \in \Omega, \quad (5)$$

where the a_i's are the coordinates of the function approximate in the basis of RBF. $\|\cdot\|$ denotes the Euclidean norm. $\underline{a}$ is the vector of the a_i's and $\Phi(y) = (\phi(\|y - x_1\|), \cdots, \phi(\|y - x_{N_x}\|)$.

The Gaussian RBF, (inverse) multiquadric RBF, and polyharmonic splines (PHS) are among the most popular RBFs. In this work, PHS have been used, since they do not require the adjustment of a shape parameter to optimize the conditioning of the interpolation, unlike the Gaussian or multiquadric RBF (Fornberg and Flyer (2015)).

The PHS are defined as follows:

$$\phi(y) = r^{2k-1},\ k \geq 1,\ \text{with } r = \|y - x_i\|. \quad (6)$$

However, very good results have been also obtained with the inverse multiquadric:

$$\phi(y) = \frac{1}{\sqrt{1 + \epsilon^2 r^2}}, \quad (7)$$

with an appropriate choice of shape parameter ϵ.

In order to improve accuracy, polynomials may be introduced in the PHS approximation along with coordinates λ_j:

$$\hat{f}(y) = \sum_{i=1}^{N_x} a_i \phi(\|y - x_i\|) + \sum_{j=1}^{m} \lambda_j p_j(y), \quad (8)$$

associated to matching constraints $\sum_{i=1}^{N_x} a_i p_j(x_i) = 0$, for $j = 1, \cdots, m$. However, in the case study presented in Section 5, we have not observed any significant improvement in accuracy if such polynomials are introduced.

We now assume that the values of function f at each center point x_i, $f(x_i)$, $i = 1, \cdots, N_x$ are available. The vector of

the $f(x_i)$'s is denoted as $\underline{f}$. It follows that the vector $\underline{a}$ of coordinates a_i, $i = 1, \cdots, N_x$ is given by:

$$\underline{a} = A^{-1}\underline{f}, \tag{9}$$

where

$$A = \begin{pmatrix} \phi(\|x_1 - x_1\|) & \cdots & \phi(\|x_1 - x_{N_x}\|) \\ \vdots & \ddots & \vdots \\ \phi(\|x_{N_x} - x_1\|) & \cdots & \phi(\|x_{N_x} - x_{N_x}\|) \end{pmatrix}. \tag{10}$$

Using (9), approximation (5) can be directly reformulated as a function of vector $\underline{f}$:

$$\hat{f}(y) = \Phi(y)A^{-1}\underline{f}, \ \forall y \in \Omega, \tag{11}$$

More generally, this approach can be used to locally approximate any linear differential operator, and in particular the advection-diffusion and boundary condition operators $\mathcal{L}$ and $\mathcal{B}$ considered in this paper, using a simple partial derivative calculation:

$$\mathcal{L}u(y) \approx \mathcal{L}(\Phi(y))A^{-1}\underline{u}_x, \ \forall y \in \Omega, \tag{12}$$

$$\mathcal{B}u(y) \approx \mathcal{B}(\Phi(y))A^{-1}\underline{u}_x, \ \forall y \in \partial\Omega, \tag{13}$$

where $\underline{u}_x$ denotes the vector of the $u(x_i)$'s, $\mathcal{L}(\Phi(y)) = (\mathcal{L}(\phi(\|y - x_1\|)), \cdots, \mathcal{L}(\phi(\|y - x_{N_x}\|)))$ and $\mathcal{B}(\Phi(y)) = (\mathcal{B}(\phi(\|y - x_1\|)), \cdots, \mathcal{B}(\phi(\|y - x_{N_x}\|)))$.

If the set of the centers x_i, $i = 1, \cdots, N_x$ is a stencil, denoted as S_y, i.e. the set of N_x nearest neighbors of any $y \in \Omega$, chosen in a larger set of centers, the approach is called RBF-generated finite difference approximation (RBF-FD), since each operator approximation is locally defined as a function of N_x values $u(x_i)$ in $\underline{u}_x$. In order to build each stencil S_y, a set of center points, denoted as X, has to be defined on $\Omega \cup \partial\Omega$. Quasi-random (low-discrepancy) sequences such as Sobol or Halton sequences are very good candidates for the definition of X, since they are known to provide evenly filling of the spatial domain and are optimal for the approximation of integrals in the Monte-Carlo sense (Morokoff and Caflisch (1994)).

It should be pointed out that, due to their local nature, the finite-dimensional RBF approximations of the operators $\mathcal{L}$ and $\mathcal{B}$ lead to sparse vectors or matrices, since they only involve N_x $u(x_i)$, values of $u(x)$ at the nearest neighbors $x_i \in S_y \subset X$ of y. Thus, the degree of sparsity is directly governed by the number of neighbors, N_x.

In what follows, we adopt the following notations for the operator approximations obtained from (12) and (13):

$$\mathcal{L}u(y) \approx \Psi_L(y)\underline{u}, \ \forall y \in \Omega, \tag{14}$$

$$\mathcal{B}u(y) \approx \Psi_B(y)\underline{u}, \ \forall y \in \partial\Omega, \tag{15}$$

where $\Psi_L(y)$ and $\Psi_B(y)$ are the sparse row-vectors defined as the finite-dimensional approximation of $\mathcal{L}$ and $\mathcal{B}$ respectively, and $\underline{u}$ denotes the vector of the $u(x_i), \forall x_i \in X$. The cardinality of the set of non-zero elements of each of the matrices $\Psi_L(y)$ and $\Psi_B(y)$ is exactly $N_x = \text{card}(S_y)$, with $N_x \ll \text{card}(X)$.

In the same way, it is also possible to define a sparse interpolation operator of $u(x)$:

$$u(y) \approx \Psi(y)\underline{u}, \ \forall y \in \Omega, \tag{16}$$

where $\Psi(y)$ is the sparse row-vector obtained from the local approximations (11) obtained from the stencil S_y.

3.2 Formulation of the RBF-FD PIML solver

Following the approach proposed by Tominec and Breznik (2021) (also denoted as LS-RBF-FD) and applied to the ADPDE problem (1)-(2), we search for an approximate solution of the ADPDE problem for a given source distribution $s(x)$, based on the RBF-FD approach, and using residual minimization:

$$\min_{\underline{u}} \frac{1}{N_\Omega}\sum_{k=1}^{N_\Omega} |\Psi_L(y_k)\underline{u} - s(y_k)|^2 + \frac{\alpha}{N_{\partial\Omega}}\sum_{k=1}^{N_{\partial\Omega}} |\Psi_B(y_k)\underline{u}|^2, \ \alpha > 0, \tag{17}$$

where the y_k's belongs to two subsets Y_Ω and $Y_{\partial\Omega}$ generated using quasi-random sequences defined on Ω and $\partial\Omega$, respectively, with $N_\Omega = \text{card}(Y_\Omega)$ and $N_{\partial\Omega} = \text{card}(Y_{\partial\Omega})$. α is a weighting coefficient used to give more or less weight to the satisfaction of the boundary conditions.

If $\text{card}(Y_\Omega) + \text{card}(Y_{\partial\Omega}) > \text{card}(X)$, (17) is an over-determined least-squares optimization problem that can be compactly reformulated as follows:

$$\underline{u} = \arg\min_v \|Dv - f\|^2, \tag{18}$$

with

$$D = \begin{pmatrix} \frac{1}{\sqrt{N_\Omega}}D_\Omega \\ \sqrt{\frac{\alpha}{N_{\partial\Omega}}}D_{\partial\Omega} \end{pmatrix}, \ f = \begin{pmatrix} \frac{1}{\sqrt{N_\Omega}}\underline{s}(Y_\Omega) \\ 0 \end{pmatrix}, \tag{19}$$

where D_Ω and $D_{\partial\Omega}$ denotes the sparse matrices built with the $\Psi_L(y_k)$'s for all $y_k \in Y_\Omega$ and the $\Psi_B(y_k)$'s for all $y_k \in Y_{\partial\Omega}$, respectively. $\underline{s}(Y_\Omega)$ denotes the vector of the $s(y_k)$, for all $y_k \in Y_\Omega$.

Since the problem (18) is linear-quadratic, provided D has full rank, a unique explicit solution is given by:

$$\underline{u} = D^+\underline{s}, \tag{20}$$

where $D^+ = (D^{\mathrm{T}}D)^{-1}D^{\mathrm{T}}$ is the left pseudo-inverse of D. In practice, the solution $\underline{u}$ can be effectively computed by using an iterative least square solver optimized for sparse problems, such as the algorithm proposed in Paige and Saunders (1982).

Remarks. We call this approach RBF-FD PIML because it is closely related to the physics-informed neural network (PINN) approach initially proposed in Raissi et al. (2019). However, using a neural network requires the solution of a (often under-determined) nonlinear regression problem with a potentially large set of parameters (weights and biases) and without a clear knowledge of the most appropriate network architecture needed. Instead, the RBF-FD approach is based on a limited set of parameters, which are here the solution of the ADPDE at the center points in X. Furthermore, the formulation of the PIML problem (17) is easy and the precision can be adjusted by increasing the size of the stencils N_x, and the number of points defined in X and Y (hyper-parameters). When dealing with linear PDEs, the key point is not to introduce the center locations as part of the parameter set, which would otherwise have led to a non-linear regression problem instead of a linear one. Consequently, an explicit solution (20) can be

obtained using the RBF-FD PIML approach. In fact, this approach can be seen as based on locally defined single-layer RBF networks, in which the bias parameters are not degrees of freedom to be determined but are fixed a priori by quasi-random sequences. Of course, a fair comparison between the two approaches in terms of solution accuracy and complexity (computation time, storage, parallelization capabilities etc.) would be welcome, but this is beyond the scope of this paper and will be the subject of further work.

This PIML approach was compared to the analytical solution of the 2D ADPDE with Dirichlet boundary conditions, based on Fourier series (see Chapter 3, Section 3.4 in Lopez-Ferber (2024)) and shows excellent accuracy results. This RBF-FD PIML solver will be used to provide a reference solution in the case study of Section 5. Now, we extend this approach to source term estimation.

4. SOURCE TERM ESTIMATION BASED ON RBF-FD PIML

The source term estimation (STE) problem consists in reconstructing the unknown distribution of a source $s(x)$ and its associated pollutant concentration $u(x)$ over the domain Ω from M measurements u_k^s made by a set of a fixed or mobile sensors in domain Ω. We formulate the STE problem as an extension of the RBF-FD PIML problem (17) by adding the least-square error between the measurements and the output predictions made by a model of the sensors, where $C(y_k^s)$ denotes the output operator at location y_k^s of the k-th measurement u_k^s:

$$\begin{aligned}\min_{\underline{u},\underline{s}} \frac{1}{M}\sum_{k=1}^{M}\frac{1}{r_k}|C(y_k^s)\underline{u}-u_k^s|^2 \\ +\frac{\alpha_1}{N_\Omega}\sum_{k=1}^{N_\Omega}|\Psi_L(y_k)\underline{u}-\Psi(y_k)\underline{s}|^2 \\ +\frac{\alpha_2}{N_{\partial\Omega}}\sum_{k=1}^{N_{\partial\Omega}}|\Psi_B(y_k)\underline{u}|^2,\ \alpha_1,\alpha_2>0, \end{aligned} \tag{21}$$

where $\underline{s}$ is the vector of the approximate $s(x_i)$'s, $\forall x_i \in X_\Omega$, α_1 and α_2 are weighting coefficients, and the r_k's are the variance of the noise of measurement k (this makes it possible to adjust the confidence placed in the measurements according to the quality of different sensors).

The operators $C(y_k^s)$ are here defined, using the interpolation operator (16), as follows:

$$\mathcal{C}(y_k^s) = \Psi(y_k^s). \tag{22}$$

However, an alternative definition is possible if we consider an integral output operator for problem (3) of the form:

$$\mathcal{C}(y_k^s)u = \int_\Omega \Delta(x-y_k^s)u(x)dx, \tag{23}$$

where Δ is the characteristic function of the sensor (for example, a Gaussian function). Then, using (16) in (23), $C(y_k^s)$ is given by:

$$C(y_k^s) = \int_\Omega \Delta(x-y_k^s)\Psi(x)dx,\ k=1,\cdots,M. \tag{24}$$

The integrals can be approximated using a quasi-Monte-Carlo method based on the low discrepancy sequence used for X_Ω.

Once again, problem (21) can be expressed in the same way as problem (18)-(19) and an explicit solution can be easily derived using a left pseudo-inverse calculus or an iterative least-square solver optimized for sparse problems. It should be emphasized that this approach can be used to estimate any continuous source distribution without any restriction on the shape of the source.

5. CASE STUDY

In order to evaluate the performance of the proposed STE approach, we consider in this paper a 3D case study involving the estimation of a single source in a 3D domain without obstacles and a mix of Dirichlet and Neumann boundary conditions. The "true" measurements are obtained from the high-resolution solution of the ADPDE using the RBF-FD PIML solver described in Section 3. Table 1 provides the physical description of the case study.

Table 1. Physical parameters of the case study in international units.

Parameters	Values
Cubic domain $(L_{\xi_1}, L_{\xi_2}, L_{\xi_3})$	$(250, 250, 50)$
Wind velocity (U_1, U_2, U_3)	$(3, 5, 1)$
Diffusion (K_1, K_2, K_3)	$(20, 20, 10)$
4 Neumann BC for:	$(\xi_1 = L_{\xi_1},\ \xi_2 = L_{\xi_2},$ $\xi_3 = 0,\ \xi_3 = L_{\xi_3})$
2 Dirichlet BC for:	$(\xi_1 = 0,\ \xi_2 = 0)$

The simulated "true" source term used to generate the reference concentration field is a parameterized Gaussian function defined over Ω:

$$s : x \mapsto a_s \exp\left(-\frac{1}{2\mu^2}\|x-x_s)\|^2\right),\ x\in\Omega, \tag{25}$$

where $a_s = 10\,\mathrm{g\,m^{-3}\,s^{-1}}$, the source position $x_s = [0.25L_{\xi_1}\,\mathrm{m},\ 0.25L_{\xi_2}\,\mathrm{m},\ 1\,\mathrm{m}]$ and $\mu = 10\,\mathrm{m}$. The source term is shown in a 2D layer at an altitude of 1 m in Fig. 1. The resulting pollutant concentration computed by a high-resolution RBF-FD PIML solver is shown in Fig. 2. The PHS $\phi(y) = r^5,\ r = \|y-x_i\|$ were used for the PIML solver and STE. The stencil size N_x is chosen equal to 100. Here, N_x must not be less than 80 to obtain sufficiently accurate results. The parameters of the reference PIML solver are shown in Table 2.

Table 2. Parameters of the RBF-FD PIML solver.

N_x	card(X)	card(Y_Ω)	card$(Y_{\partial\Omega})$	α
100	4096	16384	1024×6	1

Two scenarios are presented in this paper for the STE: both are based on a reduced RBF-FD model with 2 times fewer center and evaluation points than those used in the reference solver. Table 3 shows the parameters of the two scenarios. Scenario 1 corresponds to 1500 measurements, while scenario 2 relies on 150 measurements only. We simulated noise measurement by adding a Gaussian noise with zero mean and a standard deviation σ of 0.1 to the measurements provided by the reference RBF-FD PIML solver. The location of the measurements is here chosen to coincide with the points in a Sobol sequence defined on the 3D domain. The weighting coefficients α_1 and α_2 are

chosen equal to 1 and 20, respectively. The coefficients r_k are all equal to σ^2 (variance of the measurement noise). All the simulations are performed in MATLAB.

Table 3. Parameters of the STE.

M	N_x	card(X)	card(Y_Ω)	card$(Y_{\partial\Omega})$
		Scenario 1		
1500	100	2048	8192	512×6
		Scenario 2		
150	100	2048	8192	512×6

Unsurprisingly, scenario 1 produces the best results in terms of source location (with L_2 location error less than 1 m), and amplitude estimation (see Fig. 4 and Fig. 3), while providing a good-precision estimate of the pollutant field. However, scenario 2 (with only 150 measurements) is still able to provide an accurate estimate of the source location (with L_2 location error less than 2 m), despite the appearance of artefacts visible in Fig. 5, and a reasonable estimate of the pollutant field (see Fig. 6).

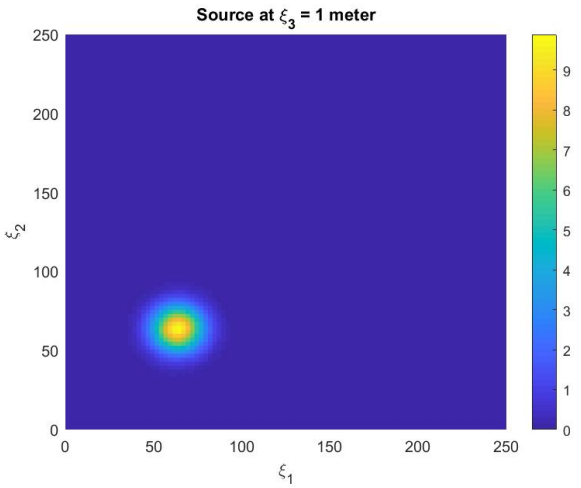

Fig. 1. True source obtained from the reference PIML solver.

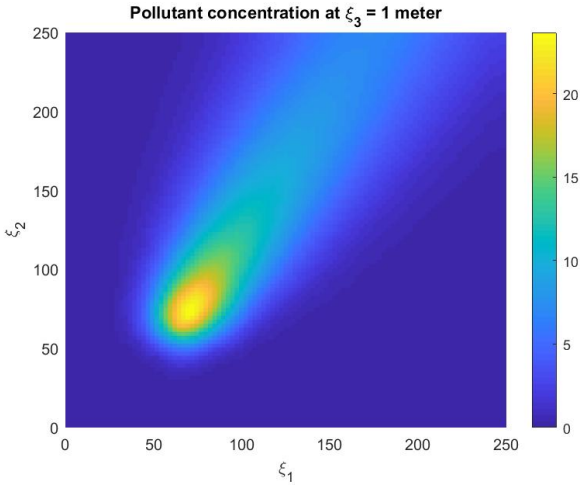

Fig. 2. True pollutant concentration obtained from the reference PIML solver.

It should be pointed out that all the RBF-FD operators (14), (15), and (16) in the PIML solver and the STE are computed using MATLAB code parallelization. The

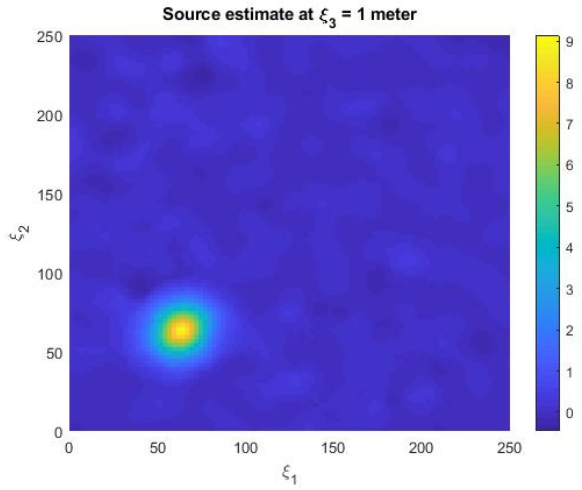

Fig. 3. Source estimate with 1500 measurements.

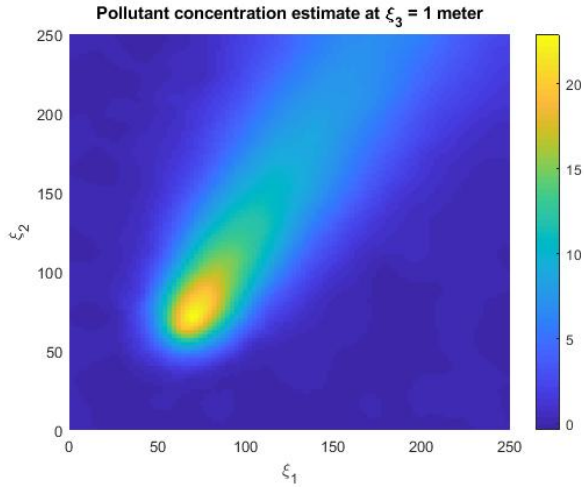

Fig. 4. Pollutant concentration estimate with 1500 measurements.

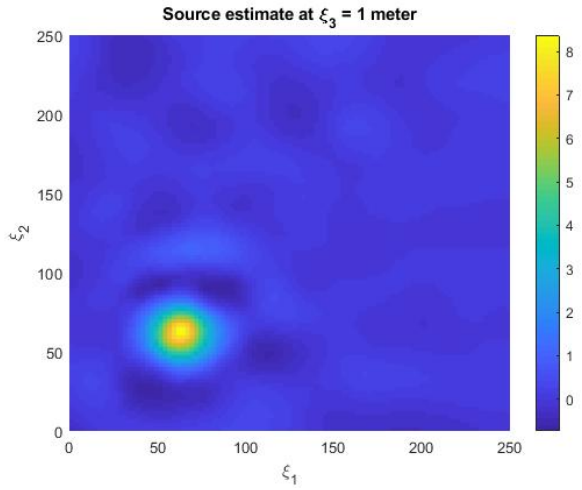

Fig. 5. Source estimate with 150 measurements.

ability to parallelize code is a very interesting feature of this approach. The solution for the reference PIML solver and the STE are obtained from a left pseudo-inverse calculation. Table 4 provides the average computation

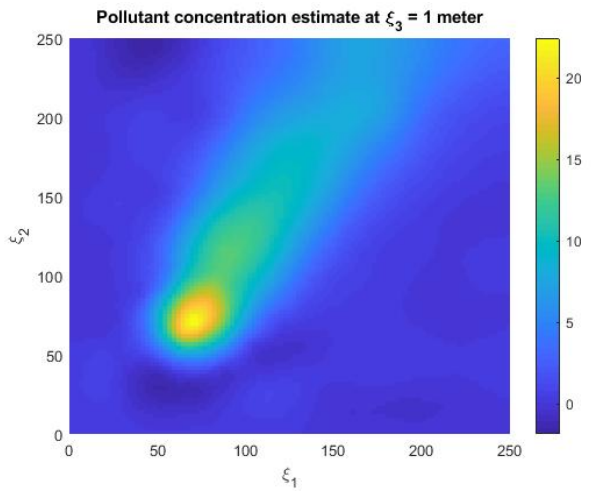

Fig. 6. Pollutant concentration estimate with 150 measurements.

times needed on a PC equipped with a 8-Core CPU (AMD Ryzen 7 3700X, 3600 Mhz) used for parallelization, with computation costs compatible with real-time applications.

Table 4. Computation times with the parameters in Tables 2 and 3.

Nature of the computation	Average time
Solver (RBF-FD operators + inversion)	52 s
Reduced model for STE (RBF-FD operators)	16 s
RBF-FD PIML STE (inversion)	10 s

6. CONCLUSIONS AND PERSPECTIVES

In this paper, we have proposed a physics-informed machine learning approach based on RBF-FD approximations. This approach seems promising for detecting and locating a static pollution source with good accuracy, at a particularly low computing cost and based on a network of fixed or mobile sensors. Future work will focus on evaluating this approach in a heterogeneous domain as encountered in an real urban environment involving buildings, the use of real data, and the estimation of non-stationary sources with mobile sensors. A comparative study with the PINN approach will also be carried out.

REFERENCES

Defforge, C.L., Carissimo, B., Bocquet, M., Bresson, R., and Armand, P. (2021). Improving Numerical Dispersion Modelling in Built Environments with Data Assimilation Using the Iterative Ensemble Kalman Smoother. *Boundary-Layer Meteorol*, 179(2), 209–240. doi:10.1007/s10546-020-00588-9. URL `https://doi.org/10.1007/s10546-020-00588-9`.

Flyer, N., Lehto, E., Blaise, S., Wright, G.B., and St-Cyr, A. (2012). A guide to RBF-generated finite differences for nonlinear transport: Shallow water simulations on a sphere. *Journal of Computational Physics*, 231(11), 4078–4095. doi:10.1016/j.jcp.2012.01.028.

Fornberg, B. and Flyer, N. (2015). *A Primer on Radial Basis Functions with Applications to the Geosciences*. Society for industrial and applied mathematics edition.

Iske, A. (2004). Multiresolution methods in scattered data modelling. *Lecture Notes in Computational Science and Engineering*, 37.

Kumar, P., Feiz, A.A., Singh, S.K., Ngae, P., and Turbelin, G. (2015). Reconstruction of an atmospheric tracer source in an urban-like environment. *Journal of Geophysical Research: Atmospheres*, 120(24), 12589–12604. doi:10.1002/2015JD024110. Publisher: John Wiley & Sons, Ltd.

Li, Y., Sha, Z., Tang, A., Goulding, K., and Liu, X. (2023). The application of machine learning to air pollution research: A bibliometric analysis. *Ecotoxicology and Environmental Safety*, 257.

Lopez-Ferber, R. (2024). *RBF-FD approaches for modelling urban air pollution and source estimation*. Ph.D. thesis, Univ. Grenoble Alpes, Grenoble.

Lopez-Ferber, R., Georges, D., and Leirens, S. (2024). Fast estimation of pollution sources in urban areas using a 3d ls-rbf-fd approach. In *Proceedings of the European Control Conference 2024, Stockholm, Sweden*.

Mathews, N.H., Flyer, N., and Gibson, S.E. (2022). Solving 3D Magnetohydrostatics with RBF-FD: Applications to the Solar Corona. *Journal of Computational Physics*, 111214. doi:10.1016/j.jcp.2022.111214.

Morokoff, W.J. and Caflisch, R.E. (1994). Quasi-random sequences and their discrepancies. *Journal of Scientifc Computing*, 16(6).

Oliver, M., Georges, D., and Prieur, C. (2022). Spatialized epidemiological forecasting applied to Covid-19 pandemic at departmental scale in France. *Systems & Control Letters*, 164, 105240. doi: 10.1016/j.sysconle.2022.105240.

Paige, C.C. and Saunders, M.A. (1982). LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares. *ACM Trans. Math. Softw.*, 8(1), 43–71. doi:10.1145/355984.355989. URL `https://doi.org/10.1145/355984.355989`.

Raissi, M., Perdikaris, P., and Karniadakis, G. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378, 686–707. doi: 10.1016/j.jcp.2018.10.045.

Septier, F., Armand, P., and Duchenne, C. (2020). A bayesian inference procedure based on inverse dispersion modelling for source term estimation in built-up environments. *Atmospheric Environment*, 242, 117733. doi:10.1016/j.atmosenv.2020.117733.

Tominec, I. and Breznik, E. (2021). An unfitted RBF-FD method in a least-squares setting for elliptic PDEs on complex geometries. *Journal of Computational Physics*, 436, 110283. doi:10.1016/j.jcp.2021.110283.

Tominec, I., Villard, P.F., Larsson, E., Bayona, V., and Cacciani, N. (2022). An unfitted radial basis function generated finite difference method applied to thoracic diaphragm simulations. *Journal of Computational Physics*, 469, 111496.

United-Nations-Environment-Programme (2019). Air pollution in asia and the pacific: Science-based solutions.

World-Health-Organization (2020). Ambient (outdoor) air pollution.

Zannetti, P. (ed.) (1990). *Air Pollution Modeling: Theories, Computational Methods and Available Software*. Springer US. doi:10.1007/978-1-4757-4465-1. URL `https://www.springer.com/gp/book/9781475744675`.