2024 European Control Conference (ECC)
June 25-28, 2024. Stockholm, Sweden

# Fast Estimation of Pollution Sources in Urban Areas Using a 3D LS-RBF-FD Approach

Roman Lopez-Ferber, Didier Georges and Sylvain Leirens

***Abstract*— Source term estimation (STE) is a field of growing interest in the context of air pollution, both for people living in urban areas and for decision makers. Thus retrieving maps of sources of pollution in an urban context is a necessity. Since urban pollution mainly depends on car traffic conditions, it is important to develop fast estimation methods to quickly and enough accurately identify highly-polluting vehicles. The challenge is high since the problem requires the inversion of distributed models defined on a 3D heterogeneous domain including complex obstacles. This paper proposes an estimation method based on a flexible least squares-radial basis function-finite differences (LS-RBF-FD) reduced model of an advection-diffusion PDE on 3D heterogeneous domains representing complex urban areas. The STE problem is solved by using an adjoint-based method relying on the reduced model to effectively estimate pollutant sources given a limited number of measurements. The paper provides preliminary results demonstrating the potential of the proposed approach.**

## I. Introduction

Pollution modeling is a research field of increasing interest for public alerting and decision-making. Regulatory authorities in the European Union point out that air pollution caused by fine particulate matter (PM2.5), mainly due to car traffic, causes about 420 000 premature deaths in Western Europe [1]. These public health concerns raise the need to develop methods to accurately identify the most polluting sources in heterogeneous urban areas, within a timeframe compatible with the change of traffic conditions. In particular, source term estimation (STE) could be very useful to isolate highly pollutant cars in urban streets based on measurements provided by a crowd of low-cost sensors. STE had 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 [3], particle filter [17], or variational calculus [9], but at the price of computationally expensive models requiring hours of computation. Some authors addressed this limitation by introducing a reduced-order model of a pollutant transport PDE with the use of proper orthogonal decomposition (POD) [8] but this method is data-driven and impose to create a large database and learn a model for each use-case. Overall, current methods for urban STE are computationally expensive, since requiring a large-scale transport model to be solved several times on fine 3D grids. Therefore, 3D STE remains challenging in complex and heterogeneous urban environments characterized by the presence of various obstacles (such as buildings).

Fast and accurate computation represents the "holy grail" in computational physics. The here-proposed approach relies on the use of a meshless radial basis function-finite differences (RBF-FD) method to get a reduced-order model of the pollution dynamics defined on 3D heterogeneous domains including potentially complex obstacles and boundary conditions without the need of a structured grid (meshless approach). 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 [5]. The RBF-FD approach has proven to be effective in the past 10 years. Various applications in geosciences [12], biomedical engineering [19], hydrodynamics [4], [5], epidemiology [15] can be found in the recent literature. Advection-diffusion equations have also been investigated, e.g. Liu et al. applied a RBF-FD method to compute non-stationary pollutant dispersion, but in an urban 2D domain only [11]. To illustrate a realistic 3D configuration, Fig. 1 depicts the pollutant concentration in a abstracted 3D urban domain with three buildings. The wind velocity field is not uniform due the presence of buildings acting as obstacles, so that the pollution plume is highly dependent on the local wind conditions.

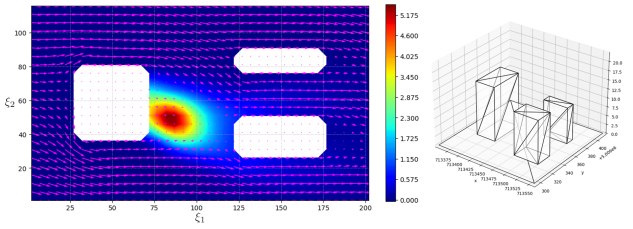


Fig. 1: Left: pollutant concentration and wind field (arrows) at a height of 2 m. Right: 3D domain composed of 3 buildings of different sizes.

The STE method proposed in this paper is a finite-dimensional variational approach based on a 3D meshless least squares-RBF-FD (LS-RBF-FD) reduced-order model. The LS-RBF-FD modelling approach is similar to that proposed in [18]. To the best of our knowledge, this is the first use of a LS-RBF-FD approach to a 3D pollution simulation

This work is supported by The French Alternative Energies and Atomic Energy Commission.

R. Lopez-Ferber and S. Leirens are with French Alternative Energies and Atomic Energy Commission, Univ. Grenoble Alpes, CEA, Leti, F-38000 Grenoble, France, roman.lopez-ferber@cea.fr, sylvain.leirens@cea.fr

D. Georges is with Univ. Grenoble Alpes, CNRS, Grenoble INP (Institute of Engineering and Management), GIPSA-lab, F-38000 Grenoble, France, didier.georges@gipsa-lab.grenoble-inp.fr

Copyright ©2024 EUCA

and estimation problem. This method appears to provide good approximate solutions even on coarse grids, which are more numerically robust and accurate in 3D domains that the RBF-FD collocation method. The approximate operators obtained with this approach exhibit high sparsity, which can be efficiently exploited to induce lower computation costs. The LS-RBF-FD approach inherently leads to reduced-order finite-dimensional models of the PDEs. Our approach benefits from both accuracy of adjoint-based methods and low computational costs of reduced-order models obtained with the LS-RBF-FD method, and facilitates STE solving with a limited number of measurements.

The rest of the paper is organized as follows: Section II introduces the LS-RBF-FD method. Section III describes the formulation and solution of the adjoint-based STE problem. A case study is presented in section IV, and Section V sums up conclusions and perspectives.

## II. LS-RBF-FD DIRECT MODEL

### A. Assumptions and air pollution model

At the urban scale, air pollution modeling usually relies on two main assumptions: The fluid is considered as incrompressible, and the pollutant does not change the air density locally. Under these assumptions, pollutant dispersion can be represented by Navier-Stockes equations coupled with the advection-diffusion partial differential equation (ADPDE). In this paper, the Navier-Stockes equations providing the mean wind velocity and mean diffusion fields are obtained from a computational fluid dynamics simulation software. The velocity and diffusion fields are assumed to be constant or slowing time-varying. Additionally, the source term is considered stationary both in time and space, and emits long enough to reach a steady state. Finally, 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 realistic assumptions in the context of air pollution in urban area [20], the 3D stationary ADPDE including source term writes:

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

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

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, $u(x)$ [$\mu g\,m^{-3}$] denotes the pollutant concentration, $s(x)$ denotes the pollutant source term [$\mu g\,m^{-3}\,s^{-1}$], and $\nu(x)$ denotes the normal vector pointing towards the exterior of the boundary. $\partial_{\xi_i}$ is the partial derivative w.r.t. $\xi_i$, while $\nabla$ is the spatial gradient in $\mathbb{R}^3$. Two kinds of boundary conditions are considered: Dirichlet conditions ($u(x) = 0$) and Neumann conditions ($\nabla u(x)\cdot\nu(x) = 0$). (1)-(2) can be written in operator form as:

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

where $\mathcal{L}$ is the differential operator $\sum_{i=1}^{3} U_i(x)\partial_{\xi_i}u(x) - \sum_{i=1}^{3}\partial_{\xi_i}\left(K_i(x)\partial_{\xi_i}u(x)\right)$ defined on the domain $\Omega$, and $\mathcal{B}$ is the boundary condition operator defined on the boundary domain $\partial_\Omega = \Gamma_d \cup \Gamma_n$, respectively.

### B. RBF-based methods to solve PDEs

Radial basis functions (RBFs) are known to exhibit excellent interpolation properties (see e.g. [6]) so that a continuous function $f : y \in \Omega \mapsto f(y) \in f(\Omega)$ can be accurately approximated by a set of $N_x$ RBFs $\phi$ centered on a finite set of nodes (collocation points) $X = \{x_i\}_{i=1,\cdots,N_x} \subset \Omega$. Let $\hat{f}$ denote the interpolation function of $f$:

$$\hat{f}(y) = \sum_{i=1}^{N_x} a_i\phi(\|y - x_i\|),\ \forall y \in \Omega, \tag{4}$$

where the $a_i$'s are the coordinates of the function approximate in the basis of RBFs. $\|\cdot\|$ denotes the Euclidean norm. Among the most popular RBFs are the Gaussian RBFs, multiquadric RBFs, and the polyHarmonic splines (PHS), the latter being used in this work, since it does not require a shape parameter, unlike the Gaussian or multiquadric RBF [6], to be tuned to optimize conditioning:

$$\phi(y) = r^{2k-1},\ k \geq 1,\ \text{with } r = \|y - x_i\|. \tag{5}$$

In order to improve accuracy, a polynomial basis is introduced in the PHS-RBF approximation along with coordinates $\lambda_j$:

$$\hat{f}(y) = \sum_{i=1}^{N_x} a_i\phi(\|y - x_i\|) + \sum_{j=1}^{m}\lambda_j p_j(y), \tag{6}$$

associated to matching constraints $\sum_{i=1}^{N_x} a_i p_j(x_i) = 0$, for $j = 1, \cdots, m$. For the sake of simplicity and without restriction, the RBF formulations will not explicitly integrate polynomials in what follows, despite they are actually used in this work.

In the 90's, Kansa [7] introduced a RBF approximation of a PDE solution. Unlike the mesh-based methods, RBF-based methods do not require any complex mesh and are therefore well adapted to deal with complex and heterogeneous domains. However, Kansa's approach suffers from a severe computational limitation when solving large systems, since the RBF operator matrix is not sparse. To overcome this limitation, the RBF-FD methods has been introduced [5]. The idea is to locally approximate the differential and boundary condition operators by using a RBF approximation only based on the $n$ closest neighbors (a stencil) of each node. RBF-FD formulas have been theoretically studied in [2] and exhibit nice convergence properties.

### C. RBF-FD approximation of operators

Consider now two sets of nodes defined in the domain $\Omega$ and on its boundary $\partial\Omega$.

Let $X$ be a set of nodes that covers the domain $\Omega$, consisting of two subsets $X_\Omega$ corresponding to nodes chosen in the domain interior $\Omega$, and $X_{\partial\Omega}$, corresponding to nodes on the boundary $\partial\Omega$. These nodes will be used as centers

for the local RBF approximation of the operators. Quasi-random low-discrepancy sequences such as Sobol or Halton sequences are recommended, 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. Let $N_{x_\Omega} = \text{card}(X_\Omega)$ and $N_{x_{\partial\Omega}} = \text{card}(X_{\partial\Omega})$, with $N_x = \text{card}(X) = N_{x_\Omega} + N_{x_{\partial\Omega}}$, where $\text{card}(\cdot)$ denotes the cardinality of a set.

In a similar manner, let $Y$ be a set of evaluation nodes consisting of two subsets $Y_\Omega$ and $Y_{\partial\Omega}$. $Y$ can also be generated using a quasi-random sequence. Let $N_{y_\Omega} = \text{card}(Y_\Omega)$ and $N_{y_{\partial\Omega}} = \text{card}(Y_{\partial\Omega})$, with $N_y = \text{card}(Y) = N_{y_\Omega} + N_{y_{\partial\Omega}}$.

The RBF-FD approximation is based on the local RBF approximation of $u(x)$ on each stencil $S_{y_k}$, where $S_{y_k}$ contains the $n$ closest neighbors denoted as $x_i^{y_k} \in X, i = 1, ..., n$, of each $y_k \in Y$ (including $y_k$). For each stencil $S_{y_k}$ of cardinality $n$, it is possible to express the approximate solution $u_\text{a}$ at each collocation point $x_i^{y_k}$ as the local linear combination of $n$ RBFs given by:

$$u_\text{a}(x_i^{y_k}) = \sum_{j=1}^{n} a_j \phi(\|x_i^{y_k} - x_j^{y_k}\|),\ x_i^{y_k} \in S_{y_k}. \tag{7}$$

Let $\underline{u}_\text{a}(y_k)$ denote the vector of the $u_\text{a}(x_i^{y_k})$'s, for all $x_i^{y_k} \in S_{y_k}$, then $\underline{a}$, the vector of the coordinates $a_j$'s, is given by:

$$\underline{a} = A(y_k)^{-1} \underline{u}_\text{a}(y_k), \tag{8}$$

where

$$A(y_k) = \begin{pmatrix} \phi\left(\|x_1^{y_k} - x_1^{y_k}\|\right) & \dots & \phi\left(\|x_1^{y_k} - x_n^{y_k}\|\right) \\ \vdots & \ddots & \vdots \\ \phi\left(\|x_n^{y_k} - x_1^{y_k}\|\right) & \dots & \phi\left(\|x_n^{y_k} - x_n^{y_k}\|\right) \end{pmatrix}. \tag{9}$$

Apart from (7), and by using (8), local approximations of the operators $\mathcal{L}$ and $\mathcal{B}$ are easily derived at each $y_k \in Y$ as:

$$\begin{aligned} \mathcal{L}u_\text{a}(y_k) &= \sum_{j=1}^{n} a_j \mathcal{L}\phi(\|y_k - x_j^{y_k}\|) = \mathcal{L}\Phi(y_k)\underline{a} \\ &= \mathcal{L}\Phi(y_k)A(y_k)^{-1}\underline{u}_\text{a}(y_k),\ y_k \in Y_\Omega, \quad (10) \\ \mathcal{B}u_\text{a}(y_k) &= \sum_{j=1}^{n} a_j \mathcal{B}\phi(\|y_k - x_j^{y_k}\|) = \mathcal{B}\Phi(y_k)\underline{a} \\ &= \mathcal{B}\Phi(y_k)A(y_k)^{-1}\underline{u}_\text{a}(y^k),\ y_k \in Y_{\partial\Omega}. \quad (11) \end{aligned}$$

It clearly appears that $\mathcal{L}u_\text{a}(y_k)$ and $\mathcal{B}u_\text{a}(y_k)$ can be expressed as linear combinations of $\underline{u}_\text{a}(y_k)$ of the form:

$$\mathcal{L}u_\text{a}(y_k) = \sum_{j=1}^{n} w_j^{\mathcal{L}}(y_k)\underline{u}_\text{a}(y_k) = \underline{w}^{\mathcal{L}}(y_k)\underline{u}_\text{a}(y_k), \tag{12}$$

$$\mathcal{B}u_\text{a}(y_k) = \sum_{j=1}^{n} w_j^{\mathcal{B}}(y_k)\underline{u}_\text{a}(y_k) = \underline{w}^{\mathcal{B}}(y_k)\underline{u}_\text{a}(y_k), \tag{13}$$

where the $w_j^{\mathcal{L}}(y_k)$'s and $w_j^{\mathcal{B}}(y_k)$'s are the RBF-FD coefficients of the finite-difference operators.

Note also that this approach allows the local interpolation of $u_\text{a}$ at any point $y$ in the domain $\Omega$ by using the interpolation formula:

$$u_\text{a}(y) = \Phi(y)A(y)^{-1}\underline{u}_\text{a}(y),\ y \in \Omega, \tag{14}$$

where $\underline{u}_\text{a}(y)$ denotes the vector of $u_\text{a}(x_i^y)$'s, for all $x_i^y \in S_y$.

### D. The RBF-FD method

The RBF-FD collocation method will consist in choosing $Y = X$ (collocation property) and solving the following linear system:

$$D\underline{u}_\text{a} = \underline{f}, \tag{15}$$

where $\underline{u}_\text{a}$ is the vector of the approximate solution evaluated at each collocation point of $Y$ and $\underline{f}$ is the vector containing $\underline{s}$ which is the vector of source term values evaluated at nodes of $X_\Omega$ concatenated with the vector of null boundary conditions evaluated at the collocation points of $X_{\partial\Omega}$ . $D$ is the matrix containing the $w_i^{\mathcal{L}}(y_k)$'s and $w_i^{\mathcal{B}}(y_k)$'s properly located at the indices of the points of each stencil $S_{y_k}$, for all $y_k$ in $Y$. Since the stencil size $n$ is usually chosen much smaller that $N_x$, the finite-dimensional operator matrix $D$ is sparse.

### E. The LS-RBF-FD method

The LS-RBF-FD method has been recently investigated in [10], [18] as an approach capable of overcoming the limitations of the RBF-FD collocation method in terms of numerical stability and accuracy. Indeed, the RBF-FD collocation approach showed poor accuracy and stability for the computation of 3D ADPDE solutions, according to our tests. The proposed approach can be interpreted as a least squared Galerkin method based on a discretized version of the following squared $L_2$-norm of the residual $r(v), v \in V \subset$ Sobolev space $W^{2,2}(\Omega)$ of the PDE problem (3):

$$\begin{aligned} u_\text{a} &= \arg\min_{v \in V} \|r(v)\|^2_{L_2(\Omega)} \quad (16) \\ &= \arg\min_{v \in V} \int_\Omega \left(\mathcal{L}v(y) - s(y)\right)^2 dy + \int_{\partial\Omega} \left(\mathcal{B}v(y)\right)^2 dy. \end{aligned}$$

By using a Monte-Carlo method for the integral approximations and (10)-(11), $\|r(v)\|^2_{L_2(\Omega)}$ can be approximated as:

$$\begin{aligned} \|r(v)\|^2_{L_2(\Omega)} \approx \frac{|\Omega|}{N_{y_\Omega}} \sum_{k=1}^{N_{y_\Omega}} (\mathcal{L}\Phi(y_k)A(y_k)^{-1}\underline{v}(y_k) - s(y_k))^2 \\ + \frac{|\partial\Omega|}{N_{y_{\partial\Omega}}} \sum_{k=1}^{N_{y_{\partial\Omega}}} (\mathcal{B}\Phi(y_k)A(y_k)^{-1}\underline{v}(y_k))^2, \quad (17) \end{aligned}$$

where $|\Omega| = \int_\Omega 1dy$ and $|\partial\Omega| = \int_{\partial\Omega} 1dy$.

Equation (17) can be rewritten in compact form as :

$$\|r(v)\|^2_{L_2(\Omega)} \approx \|D_\Omega \underline{v} - \underline{s}\|^2 + \|D_{\partial\Omega}\underline{v}\|^2. \tag{18}$$

The LS-RBF-FD method is not a collocation method since the evaluation set $Y$ now differs from the node set $X$. $Y$ is chosen to oversample $\Omega$ compared to $X$, i.e. $\text{card}(Y) \gg \text{card}(X)$. Therefore, $\underline{u}_\text{a}$ is the solution of an overdetermined least squares optimization problem defined by:

$$\underline{u}_\text{a} = \arg\min_{\underline{v}} \|D\underline{v} - \underline{f}\|^2, \tag{19}$$

where $\underline{u}_{\text{a}}$ and $\underline{v} \in \mathbb{R}^{N_x}$, $D \in \mathbb{R}^{N_y \times N_x}$ and $\underline{f} \in \mathbb{R}^{N_y}$. By using (18), $D$ and $\underline{f}$ can be expressed as:

$$D = \begin{pmatrix} D_\Omega \\ D_{\partial\Omega} \end{pmatrix}, \ \underline{f} = \begin{pmatrix} \underline{s}(Y_\Omega) \\ 0 \end{pmatrix}, \tag{20}$$

where $\underline{s}(Y_\Omega)$ denotes the vector of the source term evaluated at the nodes of $Y_\Omega$.

Unlike the RBF-FD collocation method, the LS-RBF-FD method helps to relax hard constraints induced by the boundary conditions. The stability of $u_{\text{a}}$ is also reinforced by the fact that the evaluation points of $Y$ are more numerous than the unknowns $u_{\text{a}}$ which helps to improve the well-posedness of the problem. The approximate differential operator $D$ in (20) remains sparse due to the use of local approximations on stencils of limited size. The LS-RBF-FD approach could also be adapted to solve a time-dependent PDE.

Since the problem (19) is linear-quadratic, provided $D$ has full rank, a unique explicit solution is given by:

$$\underline{u}_{\text{a}} = D^{+}\underline{f}, \tag{21}$$

where $D^{+} = (D^{\text{T}}D)^{-1}D^{\text{T}}$ is the left pseudo-inverse of $D$. In practice, the solution $\underline{u}_{\text{a}}$ can be effectively computed by using an iterative least square solver optimized for sparse problems, such as the algorithm proposed in [16].

However, it is interesting to derive the necessary condition for optimality of the problem (19) given by:

$$\begin{aligned} & D^{\text{T}}D\underline{u}_{\text{a}} - D^{\text{T}}\underline{f} = 0 \\ \Leftrightarrow\ & E\underline{u}_{\text{a}} - D_\Omega^{\text{T}}\underline{s}(Y_\Omega) = 0, \end{aligned} \tag{22}$$

where $E = D_\Omega^{\text{T}}D_\Omega + D_{\partial\Omega}^{\text{T}}D_{\partial\Omega}$.

Equation (22) will be used as *direct model* of the adjoint-based STE method proposed in the next section.

## III. Source term estimation

The problem is to simultaneously estimate the unknown source term vector $\underline{s} \in \mathbb{R}^{N_x}$ and the pollutant concentration $\underline{u} \in \mathbb{R}^{N_x}$ both defined on $X$, from a limited number of measurements $\underline{y}_{\text{mes}} \in \mathbb{R}^{N_s}$, where $N_s$ is the number of measurements properly located in the domain $\Omega$. STE is performed using a batch of the $N_s$ measurements and relies on an adjoint-based method. Here the source term is parametrized by using the RBF interpolation formula (14):

$$s(z) = \Phi(z)A(z)^{-1}\underline{s}(z), \tag{23}$$

where $\underline{s}(z)$ denotes the vector of the source term evaluated on the stencil $S_z$. It follows that the approximate source term can be evaluated on $Y_\Omega$ by using (23). The resulting operator, which provides a source approximation defined in $Y_\Omega$ from a source approximation defined in $X$, will be denoted as $\Psi_s$ such as $\underline{s}(Y_\Omega) = \Psi_s\underline{s}$.

The measurement operator, here denoted as $C$, is defined as an interpolation operator, very similar to that of (23):

$$y = C\underline{u} = \Psi(Y_{\text{mes}})\underline{u} \tag{24}$$

where $y$ denotes the model output vector and $Y_{\text{mes}}$ is the set of the measurement locations in $\Omega$ (which do not necessarily coincide with the nodes in $Y_\Omega$).

### A. An adjoint-based approach based on the LS-RBF-FD reduced model

The STE problem consists in solving the following least square optimization problem (a LS-RBF-FD model-based regularized minimization of the least square error of output prediction):

$$\begin{aligned} \min_{\underline{s},\underline{u}} \mathcal{J}(\underline{s},\underline{u}) = \frac{1}{2}\|C\underline{u} - \underline{y}_{\text{mes}}\|^2_{R^{-1}} + \frac{1}{2}\|\underline{s}\|^2_{B^{-1}} \\ \text{s.t. } E\underline{u} - D_\Omega^{\text{T}}\Psi_s\underline{s} = 0, \end{aligned} \tag{25}$$

where $R$ and $B$ are measurement noise covariance and regularization matrices respectively.

We can state the following result:

*Theorem 3.1:* The optimal solution $(\underline{s}^*, \underline{u}^*)$ of the STE problem (25) is solution of the following linear system:

$$\begin{pmatrix} K & E \\ E & -Q \end{pmatrix} \begin{pmatrix} \underline{u} \\ \lambda \end{pmatrix} = \begin{pmatrix} C^{\text{T}}R^{-1}\underline{y}_{\text{mes}} \\ 0 \end{pmatrix}, \tag{26}$$

where $K = C^{\text{T}}R^{-1}C$ and $Q = D_\Omega^{\text{T}}\Psi_s B \Psi_s^{\text{T}} D_\Omega$. Furthermore, $\underline{s}^* = B\Psi_s^{\text{T}}D_\Omega\lambda^*$, where $\lambda^*$ is the adjoint of $\underline{u}^*$. Provided that the symmetric matrix $E$ is positive definite, $\underline{u}^* = [E + QE^{-1}K]^{-1}QE^{-1}C^{\text{T}}R^{-1}\underline{y}_{\text{mes}}$, and $\lambda^*$ is given by $\lambda^* = -E^{-1}C^{\text{T}}R^{-1}(C\underline{u}^* - \underline{y}_{\text{mes}})$.

*Proof.* We first introduce the Lagrangian of the problem (25): $L(\underline{s},\underline{u},\lambda) = \frac{1}{2}\|C\underline{u} - \underline{y}_{\text{mes}}\|^2_{R^{-1}} + \frac{1}{2}\|\underline{s}\|^2_{B^{-1}} + \lambda^{\text{T}}(E\underline{u} - D_\Omega^{\text{T}}\Psi_s\underline{s})$. Lagrange's theory provides the following necessary conditions for optimality:

$$\nabla_{\underline{u}} L = 0 \Leftrightarrow C^{\text{T}}R^{-1}(C\underline{u} - \underline{y}_{\text{mes}}) + E\lambda = 0, \tag{27}$$

$$\nabla_{\underline{s}} L = 0 \Leftrightarrow B^{-1}\underline{s} - \Psi_s^{\text{T}}D_\Omega\lambda = 0, \tag{28}$$

$$\nabla_{\lambda} L = 0 \Leftrightarrow E\underline{u} - D_\Omega^{\text{T}}\Psi_s B\underline{s} = 0, \tag{29}$$

where (27) is the adjoint model equation. Under the assumption that $E$ is positive definite (thus invertible), $\lambda^*$ can be explicitly computed as $\lambda^* = -E^{-1}C^{\text{T}}R^{-1}(C\underline{u}^* - \underline{y}_{\text{mes}})$. From (28), we get $\underline{s}^* = B\Psi_s^{\text{T}}D_\Omega\lambda^*$. From (29) and by using the expression of $\underline{s}^*$, we get $E\underline{u}^* - Q\lambda^* = 0$. Finally, $\underline{u}^* = [E + QE^{-1}K]^{-1}QE^{-1}C^{\text{T}}R^{-1}\underline{y}_{\text{mes}}$, by using the expression of $\lambda^*$. Notice that $E + QE^{-1}K$ is always invertible if $E$ is positive definite, since the symmetric matrix $QE^{-1}K$ is at least positive semi-definite.

The invertibility of the matrix $E$ is guaranteed if and only if the matrix $D$ in (20) has full rank. This condition is easily met by using the PHS basis functions and a sufficiently large number of nodes in $X$ and $Y$. However, obtaining analytical conditions seems to be out of reach. In practice, the $2N_x \times 2N_x$ linear system (26) can be effectively solved by using a least square or a Krylov method, well suited for large-scale problems, without explicit inversion of $E$.

## IV. Case study: simulated source term estimation in a urban area

### A. Description of the case study

The problem is to estimate the Gaussian source inducing the pollution field in Fig. 1 (left subplot). The studied 3D domain mimics an urban area with 3 buildings of different

heights of 10 m, 15 m and 12 m, respectively. The dimensions of the domain are $L_{\xi_1} = 205\,\mathrm{m}$, $L_{\xi_2} = 120\,\mathrm{m}$ and $L_{\xi_3} = 105\,\mathrm{m}$. The 3D domain is displayed in Fig. 1 (right subplot). In this work, the PHS5 basis functions are used since they appear to be here a good choice in terms of accuracy and stability:

$$\phi(y) = r^5, \text{ with } r = \|y - x_i\|. \tag{30}$$

Additionally, second-order regularization polynomials are used to improve accuracy.

The LS-RBF-FD direct model (22) used in (25) has the following additional characteristics: $N_y = 15\,000$ evaluation nodes are used to ensure a sufficiently high resolution. For the set $X$, the node distribution is much coarser with $N_x = 2\,060$ nodes which naturally leads to a reduction of the variables to estimate $\underline{s}$ and $\underline{u}$. Null flux (Neumann) boundary conditions are assigned to the ground, the edges of the buildings and to the top of the domain, while null Dirichlet conditions are assigned to the edges of the domain. The size of the stencils is equal to 90, which here appears to be a good compromise in terms of both approximation accuracy and sparsity preservation.

The matrices $R$ and $B$ in (25) are here chosen as $R = rI_{N_s}$ and $B = bI_{N_x}$, where $I_{N.}$ denotes the $N. \times N.$ identity matrix. $r$ is related to the variance of the measurement noise and $b$ is a regularization coefficient. With this choice of the matrices and the studied domain, the best results have been obtained with $r = 1$ and $b = 0.15$.

*1) Wind and turbulence fields computation:* The 3D wind-field and diffusion fields are computed by using the Micro-SWIFT-M code [14] which is part of the PMSS software suite [13]. The model produces both the mean wind field $U$ ($\mathrm{m\,s^{-1}}$) and the mean diffusion field $K$ ($\mathrm{m^2\,s^{-1}}$).

*2) Measurements:* The forward problem was simulated with addition of white noise to produce measurements $\underline{y}_{\text{mes}}$ such that:

$$\underline{y}_{\text{mes}} = \underline{y}^0 + \underline{\epsilon}^{\text{obs}} \text{ with } \underline{\epsilon}^{\text{obs}} \sim \mathcal{N}(0, \sigma^2), \tag{31}$$

where $\underline{y}^0$ is the "true" pollutant concentration computed from a high-resolution LS-RBF-FD simulation with $N_y = 50\,000$, $N_x = 15\,648$ (see Fig. 1), and $\underline{\epsilon}^{\text{obs}}$ is a Gaussian white noise vector of variance $\sigma^2$ with $\sigma = 0.002$. The simulated "true" source term used to generate the reference concentration field is a parameterized Gaussian function defined over $\Omega$:

$$s : x \mapsto a_s \exp\left(-\frac{1}{2}(x - x_s)^{\mathrm{T}}\Sigma^{-1}(x - x_s)\right), \; x \in \Omega, \tag{32}$$

where the amplitude $a_s = 1\,\mu\mathrm{g\,m^{-3}\,s^{-1}}$, the source position $x_s = [0.42L_{\xi_1}\,\mathrm{m},\, 0.39L_{\xi_2}\,\mathrm{m},\, 0.02L_{\xi_3}\,\mathrm{m}]$ and $\Sigma$ is a diagonal matrix whose non-zero elements are the squared source spread value chosen as $8\,\mathrm{m}$ in each of the three spatial dimensions. The source term is depicted in a 2D layer at an altitude of 2 m on Fig. 3.

### B. Results

Two estimations were performed with 1 500 and 200 measurements randomly distributed on the whole 3D computational domain. The resulting estimated source terms, and the estimated concentration field derived from STE are displayed in Fig. 2. The estimated source position $\hat{x}_s$ is rather well computed in the two inversion test cases. 1 500 measurements are sufficient to obtain an estimation of the source position with errors inferior to $5\,\mathrm{m}$ in the $(\xi_1, \xi_2)$-plane (Fig 2-A), while $\hat{x}_s$ is shifted by $25\,\mathrm{m}$ along $\xi_1$ and $10\,\mathrm{m}$ along $\xi_2$ by using 200 measurements (Fig 2-C). Some inversion artifacts caused by the inversion process are visible in Fig 2-C where the source reaches 0.5 units instead of 0 close to the bottom right building. Nevertheless sources are retrieved well enough to produce pollutant field patterns (Fig 2-B and D) close to the field of reference returned by our full model shown in Fig 1. It is not surprising that the field resulting from an STE with 1 500 measurements (Fig 2-B) is closer to the reference field, in terms of configuration and order of magnitude, than that resulting from only 200 measurements shown in Fig 2-D.

The computation of our STE approach is reasonably fast and appears to be compatible with realistic real-time constraints in an embedded configuration. Once the operator matrix $D$ in (20) is computed, the estimated source term and pollutant field, solution of the problem (25), are obtained in about 4 minutes by using an Intel® Xeon® E-2276M 6-core processor and 32 GB of RAM running Python code, without exploiting parallel computing capabilities.

## V. Conclusions and perspectives

In this paper, promising preliminary results of a fast reduced model-based atmospheric source term estimation method well suited for complex urban areas are presented. To the best of our knowledge, this paper proposes the first application of the LS-RBF-FD method for the model reduction and simulation of a 3D ADPDE, together with the first use of LS-RBF-FD in an adjoint-based estimation method. This methodology appears to be effective in quickly producing good estimates of isolated sources and meets the objective of monitoring pollution in a complex 3D urban environment. By choosing the number of approximation and evaluation nodes, along with the stencil size, the trade-off between computational complexity and accuracy for model-based STE purpose is easily tunable. The estimation of distributed sources is also effective, although not presented in the paper due to the length limitation, thanks to the flexible RBF parameterization of the source term.

Our work in progress is currently dedicated to the evaluation of the approach in real urban districts, including the extension to the non-stationary case, and determining the number of measurements required for ensuring a valid estimation. Furthermore, the integration of LS-RBF-FD approach will be investigated in the framework of Bayesian methods such as ensemble Kalman filtering, together with the optimal navigation of mobile sensors to ensure the best possible estimate given meteorological situation and urban configuration. Future work will also be devoted to a comparison in terms of accuracy and computational speed with physics-informed machine learning methods.

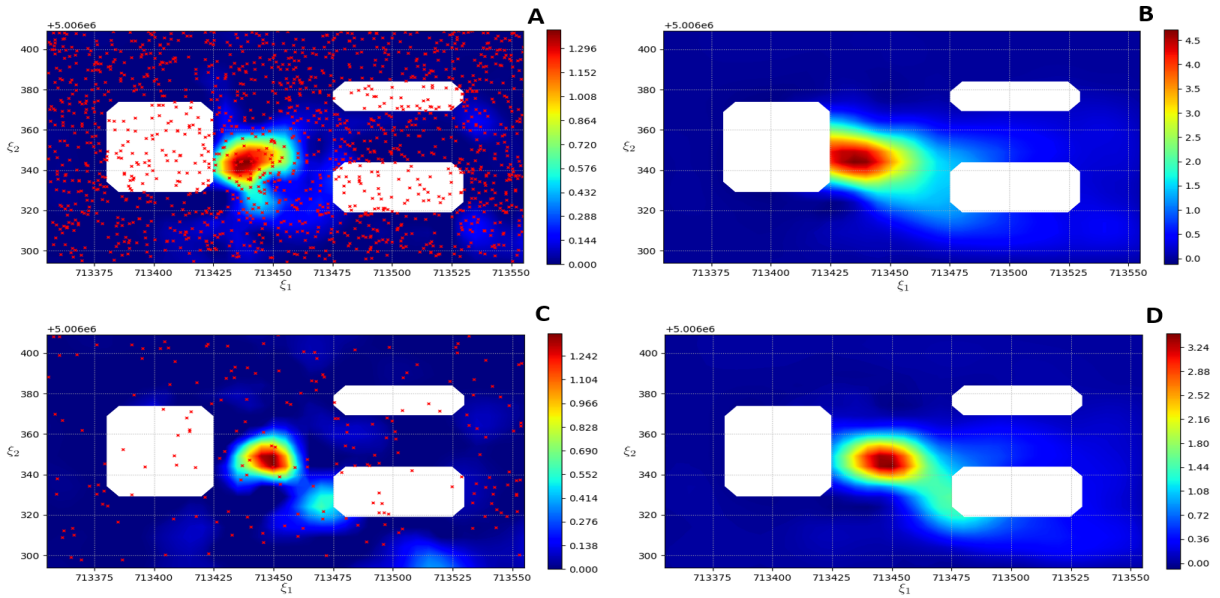


Fig. 2: **A** : Source term estimate with 1 500 noisy measurements. **B**: Corresponding reconstructed concentration field. **C**: Source term estimate with 200 noisy measurements. **D**: Corresponding reconstructed concentration field. Red crosses correspond to 3D measurement locations projected onto the 2D layer at $\xi_3 = 2\,\mathrm{m}$, where the results are plotted.

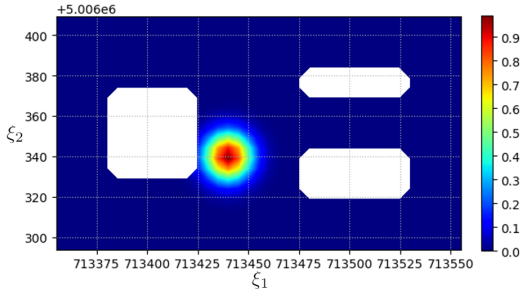


Fig. 3: True source at $\xi_3 = 2\,\mathrm{m}$.

## References


[1] Air quality in Europe - 2018 — European Environment Agency.

[2] V. Bayona, M. Moscoso, M. Carretero, and M. Kindelan. RBF-FD formulas and convergence properties. *Journal of Computational Physics*, 229(22):8281–8295, 2010.

[3] C. L. Defforge, B. Carissimo, M. Bocquet, R. Bresson, and P. Armand. Improving Numerical Dispersion Modelling in Built Environments with Data Assimilation Using the Iterative Ensemble Kalman Smoother. *Boundary-Layer Meteorol*, 179(2):209–240, May 2021.

[4] M. Dehghan and M. Abbaszadeh. The use of proper orthogonal decomposition (POD) meshless RBF-FD technique to simulate the shallow water equations. *Journal of Computational Physics*, 351:478–510, Dec. 2017.

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

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

[7] E. Kansa. Multiquadrics—A scattered data approximation scheme with applications to computational fluid-dynamics—I surface approximations and partial derivative estimates. *Computers & Mathematics with Applications*, 19(8-9):127–145, 1990.

[8] R. Khodayi-mehr, W. Aquino, and M. M. Zavlanos. Model-Based Active Source Identification in Complex Environments. *IEEE Trans. Robot.*, 35(3):633–652, June 2019. arXiv: 1706.01603.

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

[10] S. Li, L. Ling, and K. C. Cheung. Discrete least-squares radial basis functions approximations. *Applied Mathematics and Computation*, 355:542–552, 2019.

[11] J. Liu, X. Li, and X. Hu. A RBF-based differential quadrature method for solving two-dimensional variable-order time fractional advection-diffusion equation. *Journal of Computational Physics*, 384:222–238, May 2019.

[12] N. H. Mathews, N. Flyer, and S. E. Gibson. Solving 3D Magnetohydrostatics with RBF-FD: Applications to the Solar Corona. *Journal of Computational Physics*, page 111214, Apr. 2022.

[13] J. Moussafir, O. Oldrini, G. Tinarelli, and J. Sontowski. A new operational approach to deal with dispersion around obstacles : the MSS (Micro Swiftt Spray) software suite. page 5, 2004.

[14] O. Oldrini, M. Nibart, P. Armand, C. Olry, J. Moussafir, and A. Albergel. Introduction of Momentum Equations in MICRO-SWIFT. page 5, 2013.

[15] M. Oliver, D. Georges, and C. Prieur. Spatialized epidemiological forecasting applied to Covid-19 pandemic at departmental scale in France. *Systems & Control Letters*, 164:105240, June 2022.

[16] C. C. Paige and M. A. Saunders. LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares. *ACM Trans. Math. Softw.*, 8(1):43–71, Mar. 1982.

[17] F. Septier, P. Armand, and C. Duchenne. A bayesian inference procedure based on inverse dispersion modelling for source term estimation in built-up environments. *Atmospheric Environment*, 242:117733, Dec. 2020.

[18] I. Tominec and E. Breznik. An unfitted RBF-FD method in a least-squares setting for elliptic PDEs on complex geometries. *Journal of Computational Physics*, 436:110283, July 2021.

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

[20] P. Zannetti, editor. *Air Pollution Modeling: Theories, Computational Methods and Available Software*. Springer US, 1990.