

SAMPLING AND RECONSTRUCTION OF TIME-VARYING ATMOSPHERIC EMISSIONS

Juri Ranieri, Ivan Dokmanić, Amina Chebira and Martin Vetterli

School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015, Switzerland
E-mail: {juri.ranieri, ivan.dokmanic, amina.chebira, martin.vetterli}@epfl.ch

ABSTRACT

We study the spatio-temporal sampling of physical fields representing the dispersion of a substance in the atmosphere. We consider the following setup: N sensors are deployed at ground level and measure the concentration of a particular substance, while M smokestacks are located in the same area and emit a time-varying amount of the substance. To recover the emission rates of the smokestacks with a limited number of spatio-temporal samples, we consider time varying emissions rates lying in two specific low-dimensional subspaces. We propose efficient algorithms and sufficient conditions to recover the emission rates of the smokestacks from the local measurements collected by the sensor network.

Index Terms— Atmospheric dispersion, source estimation, inverse problems, spatio-temporal sampling, sensors networks.

1. INTRODUCTION

Consider an industrial zone with multiple smokestacks releasing plumes. Each smokestack emits the same substance with a time-varying concentration. The transport of the substance in the atmosphere is mainly the result of three physical phenomena: advection from the wind, diffusion from turbulent eddy motion and deposition due to gravitational settling. We examine the problem of estimating the emission rates of each smokestack, without having direct access to them. We measure the substance concentration in different spatial-temporal locations using an opportunely designed sensor network. This scenario has critical importance in designing *citizen sensing* projects, such as OpenSense [1] and SafeCast [2], and in enforcing environmental laws. Even if we focus on this interesting scenario, the techniques we present here can be successfully used in other sensing scenarios. Namely, if we consider any physical field modeled by a linear partial differential equation, we can recover the emission rates of its sources from the measurements collected by a sensor network.

1.1. Previous Work

Sampling physical field to estimate the sources of the field has been already investigated since few decades. An early approach by Nehorai et al. [3] studied the localization of a single vapor-emitting source. Lu et al. investigated the problem of reconstructing the diffusion fields driven by instantaneous sources [4]. The tradeoff between spatial and temporal sampling of diffusion field to recover point sources on a discrete grid has been investigated by Ranieri et al. in [5]. A very comprehensive overview of the mathematical modeling of atmospheric dispersions is given in [6].

This work has been supported by ERC Advanced Grant - Support for Frontier Research - SPARSAM Nr : 247006.

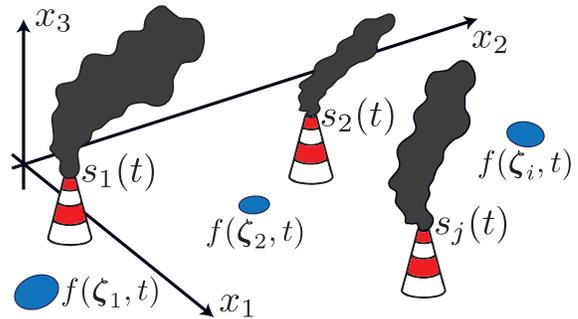


Fig. 1: A sketch of the sensing scenario. The M smokestacks are at known locations $\{\xi_j\}$. The sensors are represented as blue circles and are located at $\{\zeta_i\}$. We aim at estimate the flow of substance that is released by each smokestack $s_j(t)$.

1.2. Outline of the Paper

We state the problem and the fundamental assumptions in Section 2. In Section 3 and Section 4, we develop the reconstruction algorithms for two source models starting from the Green's function. While maintaining a practical approach, we also underline the theoretical conditions for a successful recovery of the emission rates. The theoretical results are substantiated through the numerical experiments described in Section 5.

2. ASSUMPTIONS AND PROBLEM STATEMENT

Mathematically, we consider the following inverse problem: estimate the source emission rates $s_j(t)$ of M smokestacks using S sensors capable of measuring the local concentration of a substance. The substance concentration is a spatio-temporal physical field that can be mathematically modeled by partial differential equations (PDEs) [6]. We assume we know the locations of the smokestacks $\{\xi_j\}_{j=1}^M$ and the locations of the sensors $\{\zeta_i\}_{i=1}^S$. Each sensor collects a set of T samples at different time instants $\{t_l\}_{l=1}^T$. These samples are local measurements of the field, $f(\zeta_i, t_l)$. An illustration of the sensing scenario is given in Fig. 1. We assume that the PDE is linear; thus, we can derive the spatio-temporal concentration $f(\mathbf{x}, t)$ of the substance using the convolution of the field's sources with a Green's function $g(\mathbf{x}, t)$, [7]. Note that turbulent diffusion is usually modeled using Navier-Stokes equations, that are non-linear and therefore prevent the use of the Green's function. However, for the specific case of atmospheric dispersion, the advection-diffusion equation and the Gaussian plume approximation are considered sufficiently accurate by environmental engineers [6]. Throughout the paper, we do not specify the particular Green's function, to keep the

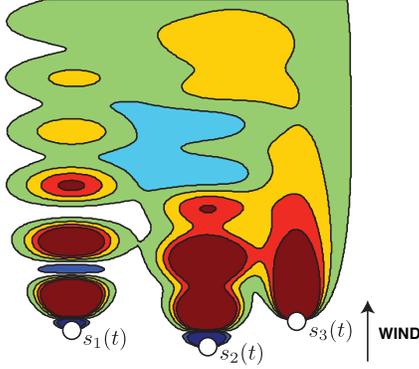


Fig. 2: A simulation of atmospheric dispersion. Three sources, the white circles, are emitting the same substance with emission rates $s_j(t) \in \mathcal{H}_K$. The value of the field is color coded: red represents the highest concentration, while dark blue represents the lowest one.

description as general as possible. In fact, $g(\mathbf{x}, t)$ changes according to the examined scenario, the boundary conditions, the presence of an inversion layer and several other factors.

Our second important assumption regards the characteristics of the emission rates. We assume that the j th smokestack is a point source in the spatial domain and the respective emission rate is a time-varying function $s_j(t)$. We consider two distinct models for the emission rates. Both models have a limited number of degrees of freedom:

1. **Model 1:** The waveform $s_j(t)$ belongs to the subspace \mathcal{H}_K defined by the span of K known functions $\phi_{k,j}(t)$:

$$s_j(t) = \sum_{k=1}^K \alpha_{k,j} \phi_{k,j}(t), \quad (1)$$

where the $\alpha_{k,j}$ are the unknowns of the j th source.

2. **Model 2:** The waveform $s_j(t)$ is defined using the concept of a signal with a finite rate of innovation (FRI) [8]. Namely, each smokestack produces only K innovations over the considered period of time. Using this abstraction, we can model many types of signals, including streams of Diracs, piecewise constant and piecewise polynomials.

The assumed source models are important for the following reasons:

- They are sufficiently flexible to deal with many types of sources, and provide an elegant way to solve the problem of estimating their appearance times [9], [3], [4].
- They effectively regularize the otherwise ill-conditioned inverse problem.

An example of the field generated by a dispersive phenomenon driven by a constant wind and time-varying emission rates is given in Fig. 2.

To summarize, given the set of ST samples of the field collected in space and time by the sensor network, we aim at the recovery of the emission rate waveforms $s_j(t)$, knowing the Green's function $g(\mathbf{x}, t)$ of the dispersive phenomenon. The main contributions of this paper are: first, an efficient algorithm to reconstruct the concentration field generated by sources lying in \mathcal{H}_K . This reconstruction is stable to noise corrupting the measured samples; second, an efficient algorithm to recover a concentration field generated by an FRI source. This algorithm is optimal in the sense of number of necessary sensors and stable to noise; and third, an efficient way to de-

termine the initial time instant of the release; a problem with direct practical implications.

3. RECOVERING EMISSION RATES LYING IN A LINEAR SUBSPACE

In this section, we analyze the recovery of the emission rates living in the linear subspace \mathcal{H}_K . Namely, each emission rate is defined according to (1).

The concentration field generated by one smokestack located at ξ is a convolution with the Green's function $g(\mathbf{x}, t)$,

$$f(\mathbf{x}, t) = g(\mathbf{x}, t) * [\delta(\mathbf{x} - \xi)s(t)], \quad (2)$$

where $*$ is the convolution, δ is a Dirac delta that defines the spatial location of the smokestack and $s(t)$ is the emission rate. Given the linearity of the PDE, if we have M smokestacks we rewrite (2) as

$$f(\mathbf{x}, t) = \sum_{j=1}^M g(\mathbf{x}, t) * [\delta(\mathbf{x} - \xi_j)s_j(t)]. \quad (3)$$

Plugging the source model (1) into the field equation (3) and using the definition of the Dirac delta we obtain

$$f(\mathbf{x}, t) = \sum_{k=1}^K \sum_{j=1}^M \alpha_{k,j} g(\mathbf{x} - \xi_j, t) * \phi_{k,j}(t). \quad (4)$$

Let $\theta_{k,j}(\mathbf{x}, t) \stackrel{\text{def}}{=} g(\mathbf{x} - \xi_j, t) * \phi_{k,j}(t)$. We can rewrite (4) as

$$f(\mathbf{x}, t) = \sum_{k=1}^K \sum_{j=1}^M \alpha_{k,j} \theta_{k,j}(\mathbf{x}, t) = \boldsymbol{\theta}(\mathbf{x}, t)^T \boldsymbol{\alpha}, \quad (5)$$

where in the last step we vectorize the double sum and obtain an inner product between two vectors of length KM . Note that the vector $\boldsymbol{\theta}(\mathbf{x}, t)$ is known, since we know the functions $\phi_{k,j}(t)$. On the other hand, the KM elements of $\boldsymbol{\alpha}$ are the unknown parameters that we want to estimate using local concentration measurements.

Now, let us consider having S sensors located at $\{\zeta_i\}_{i=1}^S$, each taking T samples of the concentration at times $\{t_l\}_{l=1}^T$. We obtain ST measurements that can be represented as the product between a known matrix Θ and the vector of unknowns $\boldsymbol{\alpha}$,

$$\begin{bmatrix} f(\zeta_1, t_1) \\ \vdots \\ f(\zeta_S, t_T) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}(\zeta_1, t_1)^T \\ \vdots \\ \boldsymbol{\theta}(\zeta_S, t_T)^T \end{bmatrix} \boldsymbol{\alpha} = \Theta \boldsymbol{\alpha}, \quad (6)$$

where Θ is a $ST \times MK$ known matrix. We can now state the first result—a sufficient condition for a numerically stable recovery of a signal lying in a subspace spanned by $\{\phi_{k,j}(t)\}_{k=1, j=1}^{K, M}$.

Proposition 1. Consider a concentration of a substance released by M smokestacks, whose dispersion is described by a PDE having the Green's function $g(\mathbf{x}, t)$. Assume that the emission rate waveforms $\{s_j\}_{j=1}^M$ belong to a subspace \mathcal{H}_K and that the concentration is measured in space and time by a sensor network composed of S sensors. If we collect enough measurements so that $ST \geq MK$ and $\text{rank}(\Theta) = MK$; then we can uniquely estimate the parameters of the sources $\boldsymbol{\alpha}$ using least square estimation. Moreover, the reconstruction error is upper bounded by the condition number $\kappa(\Theta)$

as

$$\frac{\|\tilde{\alpha} - \alpha\|}{\|\alpha\|} = \mathcal{O}(\kappa^2(\Theta)), \quad (7)$$

where $\tilde{\alpha}$ is the least square estimate of α .

Note that the rank of Θ and its conditioning strictly depend on the subspace \mathcal{H}_K and on the spatio-temporal sampling grid. We leave a detailed discussion of this fundamental aspect to [10].

4. RECOVERING EMISSION RATES WITH A FRI

In this section, we derive a reconstruction algorithm for emission rates $s_j(t)$ that are τ -periodic and have at most K innovations per period [8]. First, we show how to recover the Fourier series of the emission rates $s_j(t)$ from the Fourier series of the sensors measurements, considering the physical field as a communication channel. Second, we use the annihilation filter to determine the location and the amplitude of the K innovations of each signal $s_j(t)$. We further specialize the result for singular *Gaussian puffs* and for piecewise constant emission rates, two interesting signal models for monitoring smokestacks [6]. Note that, using these signal models we can easily represent sources appearing at unknown time instants.

First, let us also assume that the emission rates are absolutely integrable over the periods, $s_j(t) \in L^1([0, \tau])$. Then, we can define the coefficients of the Fourier series $C_n(s_j)$ with $n \in \mathbb{Z}$. Let $f^i(t)$ be the field $f(\zeta_i, t)$ measured by the i th sensor. If the Green's function is absolutely integrable, $g \in L^1(\mathbb{R})$, then we have

$$f^i(t) = \sum_{j=1}^M g(\zeta_i - \xi_j, t) * s_j(t), \quad (8)$$

and $f^i(t) \in L^1([0, \tau])$. Under these assumptions, the Fourier series $C_n(f^i)$ exists and is given as a filtering of $C_n(s_j)$, ($1 \leq j \leq S$) in the Fourier domain as

$$C_n(f^i) = \sum_{j=1}^M \hat{g}\left(\zeta_i - \xi_j, \frac{n}{\tau}\right) C_n(s_j), \quad (9)$$

where \hat{g} is the Fourier transform of the Green's function. First, we compute the coefficients $C_n(f^i)$ by sampling the signal $f^i(t)$ after an anti-aliasing filter. Then, for each n , we compute the Fourier coefficients of each source's emission rate by solving the following linear system,

$$\begin{bmatrix} C_n(f_1) \\ \vdots \\ C_n(f_S) \end{bmatrix} = \begin{bmatrix} \hat{g}\left(\zeta_1 - \xi_1, \frac{n}{\tau}\right) & \dots & \hat{g}\left(\zeta_1 - \xi_M, \frac{n}{\tau}\right) \\ \vdots & & \vdots \\ \hat{g}\left(\zeta_S - \xi_1, \frac{n}{\tau}\right) & \dots & \hat{g}\left(\zeta_S - \xi_M, \frac{n}{\tau}\right) \end{bmatrix} \begin{bmatrix} C_n(s_1) \\ \vdots \\ C_n(s_M) \end{bmatrix}. \quad (10)$$

For convenience, we define the matrix $\mathbf{G}^{(n)}$ by $(\mathbf{G}^{(n)})_{i,j} \stackrel{\text{def}}{=} \hat{g}\left(\zeta_i - \xi_j, \frac{n}{\tau}\right)$. Again, the number and the locations of the sensors are fundamental to determine the rank and the condition number of the matrices $\mathbf{G}^{(n)}$. Therefore, they also determine the uniqueness and the stability of their solutions—the coefficients $C_n(s_j)$.

This result is particularly interesting when applied to emission rates with FRI. Let us assume that each smokestack emits K puffs of a substance with unknown concentrations $\{a_{k,j}\}$ and at unknown

time instants $\{t_{k,j}\}$,

$$s_j(t) = \sum_{k=1}^K a_{k,j} \delta(t - t_{k,j}). \quad (11)$$

Using the annihilation filter on the Fourier coefficients $C_n(s_j)$ as described in [8], we can stably recover the amplitudes and the release times of the puffs once we have $2K + 1$ contiguous Fourier coefficients for each source's emission rate.

We can further extend this result to a piecewise constant source model with jumps of height $\Delta_{k,j}$ and at time instant $t_{k,j}$. We define one period of the j th emission rate as

$$\tilde{s}_j(t) = \sum_{k=1}^K \int_0^t \Delta_{k,j} \delta(s - t_{k,j}) ds, \quad t \in [0, \tau] \quad (12)$$

and its Fourier transform as

$$\hat{\tilde{s}}_j(f) = \sum_{k=1}^K \frac{\Delta_{k,j}}{i2\pi f} \exp\{-i2\pi f t_{k,j}\}. \quad (13)$$

Emission rates are now the periodic version of (12), that is

$$s_j(t) = \sum_{n \in \mathbb{Z}} \hat{\tilde{s}}_j(t - n\tau). \quad (14)$$

We use the Poisson summation formula to obtain

$$s_j(t) = \frac{1}{\tau} \sum_{n \in \mathbb{Z}} \hat{\tilde{s}}_j\left(\frac{n}{\tau}\right) \exp\left\{i2\pi \frac{nt}{\tau}\right\}, \quad (15)$$

where $\hat{\tilde{s}}_j(n/\tau)$ are the samples of (13) and the Fourier series coefficients $C_n(s_j)$. Starting from this expansion, the reconstruction algorithm recovers the piecewise constant emission rates using an annihilation filter. These steps are summarized in Algorithm 1.

Algorithm 1 Recovery of Finite Rate of Innovation Emission Rates

Input: Sensor locations $\{\zeta_i\}_{i=1}^S$, smokestack locations $\{\xi_j\}_{j=1}^M$

Output: Smokestacks' emission rates $s_j(t)$

1. Collect samples of the field from each sensor, opportunely low-pass them, and recover $S(2K + 1)$ coefficients $C_n(f_j)$,
 2. Build $2K + 1$ linear systems as in (10) and retrieve $2K + 1$ Fourier coefficients for every emission rate $s_j(t)$,
 3. Use the annihilation filter method to recover the K innovations from the Fourier series coefficients (13).
-

When designing the sensor network, certain tradeoffs must be considered. The number of time measurements per period τ determines the cutoff frequency of the low-pass filter and therefore the number of Fourier coefficients we can recover. This defines a clear upper bound on the number of innovations that we can recover. The number of sensors must be at least equal to the number of smokestacks: the larger it is, the better the conditioning of the linear systems (10). If we sample the field more frequently, we gain stability in estimating the innovation amplitudes and locations. Note that it is possible to use effective denoising techniques to reduce errors due to ill conditioning in (10), such as Cadzow's method [11]. Furthermore, the maximum number of innovations that we can re-

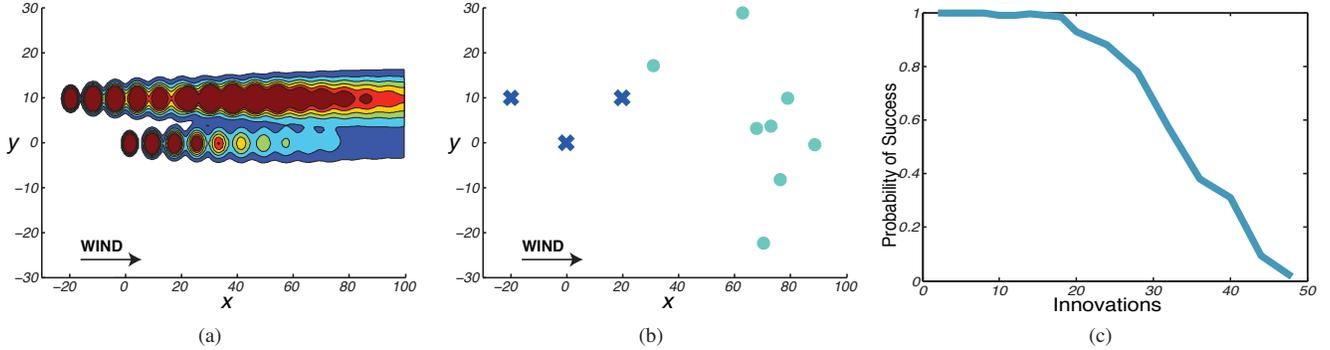


Fig. 3: (a) Concentration of the substance at $t = 2000s$ with a constant wind speed $\mu = 20m/s$, when the emission rates are random piecewise constant functions with $K = 10$ innovations. (b) Location of the sources (blue crosses) and of the sensors (light blue circles). (c) Probability of success of Algorithm 1 as a function of the number of innovations K . The sensor network showed in (b) collected 100 Fourier coefficients $C_n(f^i)$ from each sensor.

trieve depends also on the bandwidth of the Green's function. In fact, it might happen that the system matrix in (10) is rank-deficient, thus limiting the number of Fourier coefficients we can consider. We leave a more detailed discussion of the bandwidth of the Green's function and of diffusion fields in general to [10].

We conclude this section with the proposition that summarizes the derived results for the recovery of FRI emission rates.

Proposition 2. *Let us assume that the emission rates of each of the M smokestacks have K innovation per period. Assume further that we collect enough measurements in space and time so that $ST \geq M(2K+1)$ and $\text{rank}(\mathbf{G}^{(n)}) = MK$. Then, we can uniquely recover the locations and the amplitudes of the K innovations using the annihilation filter method.*

5. NUMERICAL RESULTS

For the recovery of emission rates belonging to a subspace \mathcal{H}_K we run multiple numerical simulations. As expected, the reconstruction performance largely depends on the locations of the sensors and on the wind speed. The wind has a positive impact on the recovery. Indeed, a higher wind speed is equivalent to having the sensors closer to the smokestacks.

Given the limited available space, we focus on the second method, given in Section 4. We placed three sources at $\xi_1 = (0, 0)$, $\xi_1 = (-20, 10)$ and $\xi_1 = (20, 10)$. Each source has K random innovations. We measure the field with $S = 8$ sensors placed at random. In Fig. 3, we show the percentage of successful reconstructions as a function of the number of innovations, together with two plots representing the field at a particular time instant and the locations of sources and sensors. We underline the following point: the number of recoverable innovations depends on the bandwidth of the Green's function $g(\mathbf{x}, t)$ and not on the number of sensors, as soon as $S \leq M$. In fact, the performance starts to decrease when the low-pass effect of the diffusive phenomena prevents us to recover a sufficient amount of Fourier coefficients $C_n(s_j)$.

6. CONCLUSION

We considered the problem of estimating the emission rates of smokestacks, releasing a substance that is subsequently dispersed in the atmosphere.

We proposed two reconstruction techniques for two different signal models, based on local measurements collected by a sensor network. We defined sufficient conditions such that the reconstruction is successful and stable. Our experiments align with the theory and show that exact reconstruction is feasible when the proposed conditions are satisfied.

These techniques can be applied to other sensing scenarios, where a linear field driven by time-varying sources is sampled by a sensor network in space and time. In the future, we will study two important aspects of the reconstruction stability: the optimal sensor location and the bandwidth of the Green's function.

7. REFERENCES

- [1] K. Aberer, S. Sathé, D. Chakraborty, A. Martinoli, G. Barrenetxea, B. Faltings, and L. Thiele, "OpenSense: open community driven sensing of environment," in *ACM SIGSPATIAL Intern. Workshop on Geostreaming*, New York, 2010, ACM Press.
- [2] "Safecast," <http://www.safecast.org/>.
- [3] A. Nehorai, B. Porat, and E. Paldi, "Detection and localization of vapor-emitting sources," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 243–253, 1995.
- [4] Y. M. Lu, P. L. Dragotti, and M. Vetterli, "Localizing point sources in diffusion fields from spatiotemporal samples," in *Proc. 9th Intern. Conf. on Sampling Theory and Applications*, Singapore, Feb. 2011.
- [5] Juri Ranieri, Amina Chebira, Yue Lu, and M. Vetterli, "Sampling and reconstructing diffusion fields with localized sources," in *IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, Prague, May 2011.
- [6] J. M. Stockie, "The mathematics of atmospheric dispersion modeling," *SIAM Review*, vol. 53, no. 2, pp. 349, 2011.
- [7] D.G. Duffy, *Green's functions with applications*, Chapman and Hall/CRC, 2001.
- [8] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," *IEEE Signal. Proc. Mag.*, vol. 25, no. 2, pp. 31–40, 2008.
- [9] I. Dokmanic, J. Ranieri, A. Chebira, and M. Vetterli, "Sensor networks for diffusion fields: detection of sources in space and time," in *Proc. Allerton Conf. on Comm., Control and Comp.*, Monticello, Sept. 2011.
- [10] J. Ranieri, I. Dokmanic, A. Chebira, and M. Vetterli, "Signal processing of diffusive fields: bandwidth, sampling and reconstruction," *to be submitted, IEEE Trans. Signal Process.*
- [11] J. A. Cadzow, "Signal enhancement—a composite property mapping algorithm," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 36, no. 1, pp. 49–62, 1988.