

SCIENTIFIC REPORTS



OPEN

Source reconstruction of airborne toxics based on acute health effects information

Christos D. Argyropoulos¹, Samar Elkhalfifa¹, Eleni Fthenou², George C. Efthimiou³, Spyros Andronopoulos³, Alexandros Venetsanos³, Ivan V. Kovalets⁴ & Konstantinos E. Kakosimos¹

The intentional or accidental release of airborne toxics poses great risk to the public health. During these incidents, the greatest factor of uncertainty is related to the location and rate of released substance, therefore, an information of high importance for emergency preparedness and response plans. A novel computational algorithm is proposed to estimate, efficiently, the location and release rate of an airborne toxic substance source based on health effects observations; data that can be readily available, in a real accident, contrary to actual measurements. The algorithm is demonstrated by deploying a semi-empirical dispersion model and Monte Carlo sampling on a simplified scenario. Input data are collected at varying receptor points for toxics concentrations (*C*; standard approach) and two new types: toxic load (*TL*) and health effects (*HE*; four levels). Estimated source characteristics are compared with scenario values. The use of *TL* required the least number of receptor points to estimate the release rate, and demonstrated the highest probability (>90%). *HE* required more receptor points, than *C*, but with lesser deviations while probability was comparable, if not better. Finally, the algorithm assessed very accurately the source location when using *C* and *TL* with comparable confidence, but *HE* demonstrated significantly lower confidence.

In modern societies, where risk has a dominant role in all facets of our life^{1,2}, short-term exposure to toxic/hazardous material (HazMat), whether intentional or accidental, poses great threats to the surrounding population³⁻⁵. The Sarin gas terrorist attacks in Japan (Matsumoto⁶ and Tokyo⁷), as well as the famous chemical accidents of Bhopal, India⁸ and Seveso, Italy⁹ are characteristic examples of such incidents.

An unlikely event of a toxic/hazardous material begins with a “source” releasing the HazMat in the air, depending on the meteorological conditions, the plume is dispersed reaching, potentially, high HazMat “concentrations” that vary over space and time. A subject’s (receptor’s) “exposure” to the HazMat depends mainly on the subjects’ activity and residence time within the plume. The amount - “dose” - deposited within the body is affected by a number of biological factors. Finally, the subject’s “response” to the HazMat event will be exhibited by a series of adaptive or adverse health effects. This is the consequences analysis paradigm¹⁰ in which the final impact of a HazMat release depends on the “source-concentration-exposure-dose-response” binary relationships. However, among the required information for emergency response planning, the source characteristics present the greatest factor of uncertainty¹¹ and can differ from the true one by a factor of 10 or more¹². Thus, the identification of source characteristics is of high importance for assessing the health impact on the population and planning exposure measures for epidemiological studies¹³ covering all environmental compartments¹⁴.

The reconstruction of the source term of an airborne contaminant may be obtained by using forward (optimization/minimization) or backward (inverse) approaches, in which source characteristics are inferred from concentration or deposition measurements at different locations and time intervals by establishing source-receptor (i.e. source-concentration) relationships¹⁵. The problem itself is not new. It is driven mainly by the investigation of nuclear-energy accidents in the global¹⁶ or local scale¹⁷ and it is not limited to airborne contaminants¹⁸.

¹Department of Chemical Engineering and Mary Kay O’Connor Process Safety Center, Texas A&M University at Qatar, Education City, Doha, PO Box 23874, Qatar. ²Qatar Biobank for Medical Research, Doha, 5825, Qatar. ³Environmental Research Laboratory, INRASTES, NCSR Demokritos, Patriarchou Grigoriou & Neapoleos Str, 15310, Aghia Paraskevi, Athens, Greece. ⁴Department of Environmental Modelling, Institute of Mathematical Machine and Systems Problems, National Academy of Sciences of Ukraine, Kiev, Ukraine. Correspondence and requests for materials should be addressed to K.E.K. (email: kkakosimos@qatar.tamu.edu)

The first class of methods includes Gradient based techniques (e.g. least squares¹⁹, re-normalisation²⁰ and Broyden-Fletcher-Goldfarb-Shanno algorithm²¹), Meta-heuristics (e.g. pattern search method²², simulated annealing²³ and genetic algorithms²⁴), Bayesian inference approaches^{25,26} and Markov Chain Monte Carlo (MCMC) sampling techniques^{27,28}. More details for the methods of first category may be found in the recent review paper by Hutchinson *et al.*²⁹. The second type of methods incorporates adjoint and tangent linear models^{30–32}, Kalman filtering³³ along with Gaussian³⁴, Lagrangian^{35,36} and advanced dispersion models^{28,37}, as well as variational data assimilation techniques^{38,39}. In any case, the greater the number of available concentration measurements, the closer the guessed/estimated source rate will be to reality, resulting to more reliable and faster mitigation. For example, Say, *et al.*⁴⁰ inferred multi-annual UK emissions of Hydrofluorocarbons and Schauburger, *et al.*⁴¹ completed a fragmented set of emission rates from a wastewater treatment, both by using the respective atmospheric observations. Unfortunately, concentration measurements are rarely available on-site unless for a high-risk facility or a regional-scale accident.

The concept of information-reconstruction has also been employed on the third and fourth relationships (exposure-dose and dose-response), of the consequences paradigm i.e. the use of available dose or response data to estimate the exposure levels. However, no source reconstruction has been attempted, because the aforementioned source reconstruction models take into account only concentration data⁴². For example, Pirkle, *et al.*⁴³ used biomonitoring data, such as clinical tests and biomarkers, in order to identify lead exposures, while Hays, *et al.*⁴⁴ discussed issues inherent in using clinical tests for evaluating such data. Other researchers⁴⁵ moved one step further by using Physiologically Based Toxicokinetic (PBTK) and Biologically Based Dose-Response (BBDR) models, together with appropriate optimisation and inverse modelling techniques to reconstruct exposure to environmental chemicals, and to some extent the source itself, from biomarkers. PBTK models approximate the kinetic behaviour of chemicals and, as a result, can predict the internal dose at targeted tissues/organs⁴⁶. On the other side, BBDR models represent biological processes at the molecular and cellular level which link the target issue/organ dosage to the health outcome^{47,48}. Later, Chen, *et al.*^{49,50} employed chemical data (i.e. blood sampling and urinary data) in conjunction with Monte-Carlo sampling techniques for reconstructing past exposures. In the same direction, simplified PBTK models and clinical data were deployed to demonstrate the reconstruction of exposures, to BisPhenol A⁵¹ and Carbaryl⁵².

In all these studies, the main objective was to estimate (reconstruct) exposure to a chemical(s) by using primarily clinical tests and biomarkers. Whereas, only a few of them attempted to incorporate the observed health effects (clinical health observations e.g. asthma, gastrointestinal problems, and deaths) which are among the first type of information collected by the response personnel. Therefore, this study demonstrates the concept of reconstructing source characteristics based on health observations in the case of acute chemical, biological or radiological accidents. A simplified forward-modelling source-reconstruction algorithm was developed based on the SLAB semi-empirical dispersion model⁵³ and the Monte Carlo sampling technique. A realistic toxic gas release scenario (synthetic scenario) is presented to assess the performance and capability of the proposed concept. To the authors' best knowledge, this is the first report of this concept.

Methods

In the absence of any similar previous work, it is first necessary to identify the expectations and limitations of such an algorithm. In principle, the algorithm should be able to combine available response data to estimate the source characteristics. Typically, such a problem can be solved by either optimizing the forward solution or solving the inverse mathematical equation¹⁵ e.g. advection equation. However, there is no single equation that relates source-response directly, instead there are four different sets of equations that cover the binary relationships of source-concentration, concentration-exposure, exposure-dose, and dose-response. Often, some of these sets of expressions are based on stochastic and probabilistic approaches. Moreover, response data, collected by emergency response teams, will usually be in the form of discrete information (e.g. clinical health effects). Following the above reasoning, it is evident that the proposed concept includes a number of inherent difficulties, although promising and attractive, and the use of an advanced inverse numerical scheme could further increase the complexity. Therefore, the first version of the algorithm was designed using simplified steps and approaches, in order to be able to investigate the proposed concept and assess its performance. The overall outline of the methodology is presented (Fig. 1).

First, a simplified and realistic release scenario is formulated and forward modelling is implemented at three steps (Fig. 1, white blocks). The results of the forward modelling comprise the synthetic dataset which plays the role of real data collected by the response teams and will be used as input for the source reconstruction algorithm. Note that a random numerical noise in the range of $\pm 20\%$ has been added to the concentration values (only at the forward atmospheric dispersion modelling results) to represent the expected uncertainty of real measurements and to account for the input and meteorological conditions variability. The selection of the noise range is based on the recent findings of a related EU project⁵⁴.

It should be noted that the meteorological data for the source reconstruction can be a difficult problem to address due to the lack of reliable data. In order to surpass this problem a number of researchers has proposed some interesting techniques for increasing the confidence of the selected meteorological data. For example, Allen *et al.*⁵⁵ attempted to determine directly by their proposed algorithm the surface wind direction along with the reconstruction of the source characteristics. Kolczynski *et al.*⁵⁶ performed ensemble forecasting in order to quantify the uncertainty of the meteorological data. They used multiple numerical weather prediction models adopting different initial and boundary conditions for each model configuration. Another interesting approach to treat the meteorological data uncertainty by Zajackowski *et al.*⁵⁷ is the combination of numerical weather predictions with advanced CFD techniques for the wind prediction. More specifically, the method assimilates mesoscale model data (e.g. wind profiles) into the selected CFD model presenting improved accuracy. In the present study,

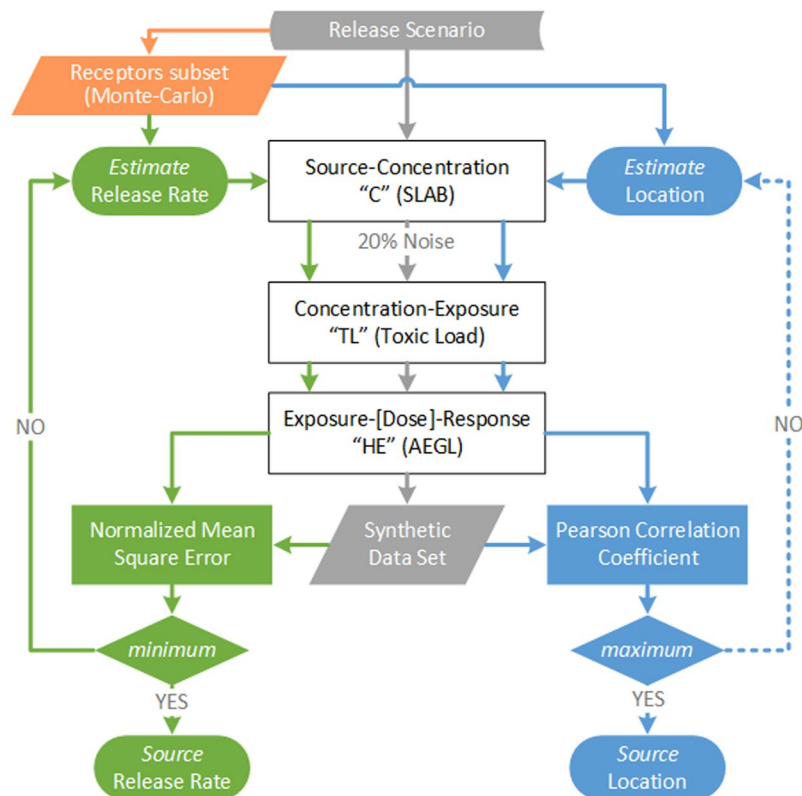


Figure 1. Outline of the developed computational algorithm and presented methodology.

we used available meteorological data from an airport location, however, in a future research we will examine the proposed algorithm with the assimilation of weather data into the ADREA CFD code^{58,59}.

Then, a sample (receptors subset) of the synthetic dataset is selected via Monte-Carlo sampling, in order to be used as input parameter into the source reconstruction algorithm. The release rate estimation and location identification take place in two different, parallel, and loop processes, each using the forward modelling scheme until the set goal is achieved (Fig. 1 green and blue blocks).

Finally, it is important to mention that the release rate and location of the source may also be calculated based on the mean square error, so it is not necessary to estimate location separately, as we shown in Fig. 1. However, we did not select this approach because there is a numerical advantage regarding the convergence of the solution, as shown in our recent work by Efthimiou *et al.*⁵⁸.

Description of the release scenario. The considered scenario involves a release from a feed pipeline of natural gas with a source rate of 100 kg s^{-1} and leak duration of 10 min. The gas contains 0.77% (w/w) of hydrogen sulphide (H_2S), a well-studied toxic agent with adverse health effects even at sub-ppm concentrations^{60,61}. The release of H_2S is directed to the non-process building, while the temperature and pressure of pipeline are assumed to be 27°C and 83 bars, respectively. The scenario is based on an own previous study⁶² and a related Quantitative Risk Assessment (not publically available). The surrounding area of the facility is characterised as flat without obstacles. The prevailing wind speed during the release is taken equal to 5 m s^{-1} , while a class *D* to “(neutral conditions) atmospheric stability has been selected.

A computational domain area of $5000 \times 1000 \text{ m}^2$ was selected for the numerical simulations. Then, the domain area was divided into a 200×40 grid, with $\Delta x = \Delta y = 25 \text{ m}$. The coordinates of the actual source location were chosen at $(x, y, z) = (500, 300, 0)$, while all the source parameters were assumed to be constant over the time. The same grid was used to identify possible source locations.

Source-concentration modelling. The SLAB atmospheric dispersion model was selected to simulate the H_2S plume and predict the concentration levels with respect to time at each grid node (x and y coordinates) and at height 2 m above the ground (selected breathing zone). The SLAB model was developed by Ermak and co-workers^{53,63} and is suitable for continuous and finite duration release scenarios⁶⁴. It has been extensively validated against lab- and field- experiments, presenting satisfactory agreement⁵³.

Concentration-exposure and exposure-dose-response modelling. The contact with a toxic agent through ingestion, inhalation and dermal route can lead to adverse health effects due to the absorption or adsorption of the material from the human body. This forms a dose and potentially turns out a disease issue⁴². Thus,

Exposure duration	10 min	30 min	1 h	4 h	8 h
H₂S (ppm)					
AEGL-1	0.75	0.60	0.51	0.36	0.33
AEGL-2	41	32	27	20	17
AEGL-3	76	59	50	37	31

Table 1. Acute exposure guideline levels (AEGLs) for H₂S (<https://www.epa.gov/aegl>).

depending on the toxic agent characteristics (i.e. nature and concentrations), the impact of the agent can vary significantly, while the duration of the exposure at low concentrations can be extended for a longer period⁶⁵.

The impact of exposure is estimated by using three acute exposure guideline levels (AEGLs)⁶⁶. The AEGL thresholds determine the exposures (i.e. constant concentrations levels) for given durations from 10 min to 8 hrs, presenting three levels of harm or the general public, namely notable discomfort (AEGL-1), disability (AEGL-2) and life threatening or death (AEGL-3)^{66,67}. However, in a real release scenario, inhalation dosages may vary significantly with durations, which pose a challenge to estimate the onset of the effects using the fixed values by AEGLs.

To surpass this problem, a toxic load algorithm, based on the AEGLs and quantification of the toxic dosage, using a non-linear equation⁶⁸, was proposed⁶⁹. When this toxic dosage exceeds a value of one, then the respective AEGL threshold has been reached. By adopting this approach, we computed the onset of each of the AEGL thresholds either as a continuous toxic load (i.e. *TL*) or as discrete health effects (i.e. *HE*, discrete values of AEGL-1, AEGL-2, AEGL-3, and “nothing”) for every available concentration profile. The *TL* predictions require the AEGL thresholds for the considered chemical (i.e. for H₂S see Table 1). A power law of 4.4 is obtained by fitting the AEGL data of Table 1 (R² better than 0.99), while we define a low and upper bound of exposure duration equal to 30 secs and 24 hrs, respectively.

The proposed algorithm “EAGLE”⁶⁹ is based on an application of the “Induction Parameter Model”⁷⁰ which employs a non-linear equation⁶⁸ in order to quantify the toxic dosage⁷¹. EAGLE is also capable of coupling with a dispersion model in order to enable the prediction of the onset of AEGLs thresholds for time-varying plume involving chemical agents with tabulated AEGLs. The approach can be determined by the following expressions:

$$TL(t) = \int_0^t TL_{rate}(t') dt' \quad (1)$$

$$TL_{rate}(t) := \frac{dTL}{dt} = \frac{1}{t_b} \left[\frac{C(t)}{C_{t_b}} \right]^n \quad (2)$$

where *TL* (*t*) is the integrated toxic load, *t* is the time, *t_b* is the reference AEGL time band exposure step, *C_{t_b}* is the concentration corresponds to the reference AEGL time band exposure time (Table 1) and *n* (=−0.23 for H₂S) is the power exponent. In this work, the EAGLE algorithm was programmed to evaluate the toxic load (exposure) and the corresponding health effects (response) resulting from the estimated H₂S concentration levels

Source reconstruction and Monte Carlo sampling. In order to obtain an optimal estimate of the source rate (*Q_s*), we adopted an optimisation routine that is based on minimising a cost function, namely normalised mean square error:

$$NMSE = \frac{\langle (C_m - C_o) \rangle}{(\langle C_m \rangle - \langle C_o \rangle)} \rightarrow \min \quad (3)$$

where triangle brackets denote arithmetic averaging, while *m* and *o* subscripts stand for model and observations, respectively. The iterative calculation was performed by the “fminbnd” built-in function of MATLAB R2016a; an algorithm based on golden section search and parabolic interpolation^{72,73}.

The identification of the release location can be obtained from another optimisation routine that is based on maximizing a cost function based on the Pearson correlation coefficient, *J*⁷⁴:

$$J = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{C_{o,i} - \langle C_o \rangle}{\sigma_o} \right) \left(\frac{C_{m,i} - \langle C_m \rangle}{\sigma_m} \right) \rightarrow 1 \quad (4)$$

where *C* is the concentration and *σ* is the standard deviation. *J* approaches a value of unit for two perfectly correlated sets of observations and modelling results. Since the source rate does not directly participate in the correlation coefficient, the obtained solution is independent of the choice of the source rate^{58,75}. The source location could be calculated with a multidimensional optimisation function, but in our case for illustrative reasons we compute *J* (Eq. (4)) at all possible grid points (receptor points) because of the small computational demand.

Monte Carlo simulations are also performed to generate multiple samples of varying size of observations. Hereafter, we refer to the size of the observations sample as number of receptors (ranging from 2 to 100) and the number of samples as the number of iterations (10, 100, 1000, and 10000). In other words, each iteration is comprised by a predefined number of randomly selected receptors. The receptors’ network or distribution is an important aspect of source reconstruction and this is captured by studying multiple and varying number of

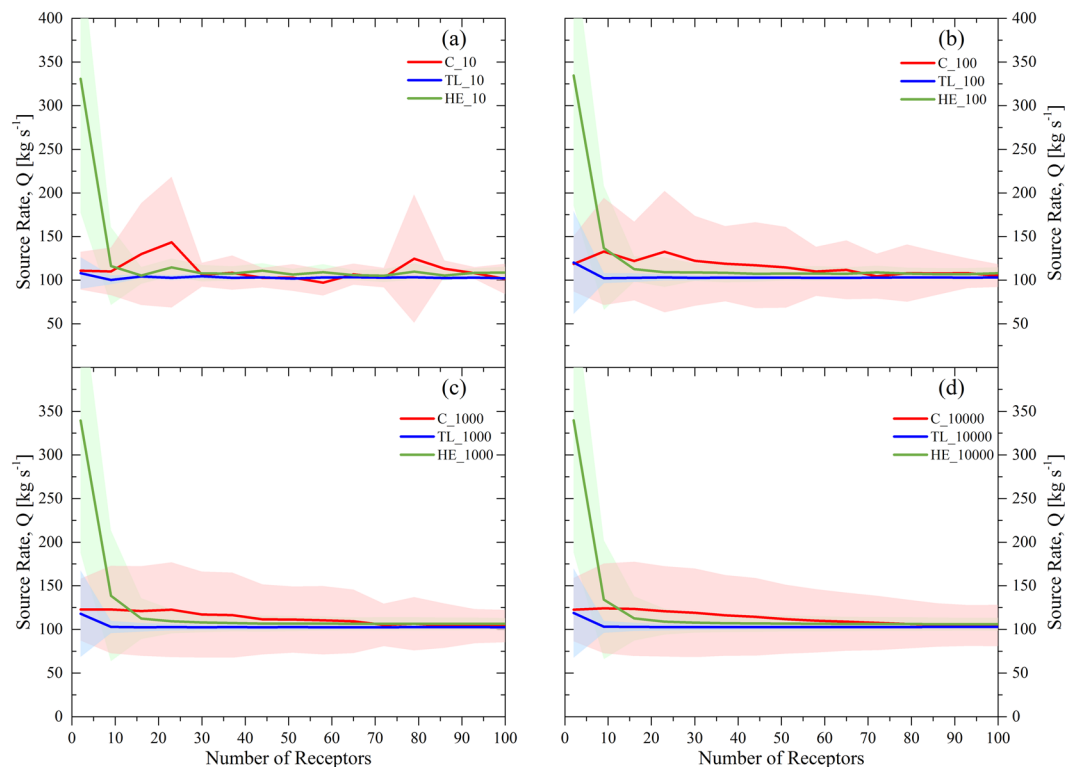


Figure 2. Estimation of source rate value according to the selected number of receptors: (a) 10 iterations, (b) 100 iterations, (c) 1000 iteration and (d) 10000 iterations, for Concentration (C ; $\color{red}{\blacksquare}$), Toxic Load (TL ; $\color{blue}{\blacksquare}$) and Health Effects (HE ; $\color{green}{\blacksquare}$) considered parameters. The three shaded areas represent the 2σ range for C (red shaded area), TL (blue shaded area) and HE (green shaded area), respectively.

iterations. However, the effect of the receptors' position and other specific characteristics have not been studied explicitly. Finally, the sampled results (subset of the synthetic dataset) are then imported into the reconstruction algorithm.

Results

This section presents the results of the source reconstruction algorithm i.e. performance towards estimating the correct source release rate and location. It also includes results of the algorithm sensitivity against the two most important parameters, the number of receptors and number of iterations. The time period at which data are collected, although critical in a real incident, did not affect the behavior of the algorithm. Therefore, here only results for the first 600 s after the incident are presented.

SLAB computation times are in the scale of < 1 s per simulation. Our inverse modelling approach scales linearly with the number of available receptor points, in practice less than 20–40 receptors. Overall computation time was less than 1 min per iteration (i.e. specific group of receptors). We used a multi-CPU system in order to conduct the multiple-iterations studies.

Prediction of source rate (Q_s). Figure 2(a–d) exhibit the estimated source rate based on the different number of receptors ranging from 2 to 100 sample points. The proposed model was run for four number of iterations (i.e. 10, 100, 1000 and 10000) in order to investigate the impact of increasing the magnitude on the convergence of the results. We observe that in any case of chosen number of iterations the estimated source rate approaches sooner/faster the value of the real one ($\sim 100 \text{ kg s}^{-1}$) when the dataset is obtained from the TL (solid blue line). Accordingly, the 2σ range of TL (blue shaded area) is also smaller than C (red shaded area) and HE (green shaded area) methods.

By using the HE method (green solid line), we notice that there is a delay to approach the real value and a larger scattering area (green shaded area) compared to the other two methods. The use of C values as input data in the source reconstruction algorithm gives better convergence than HE dataset but worse than TL .

Figure 3 illustrates the number of occurrences versus the estimated values of source rate for 100 receptors and two number of iterations (1000 and 10000). The frequency values for the estimated values of source rate were calculated by using input data of C , TL and HE . The estimated results for the source rate after 1000 iterations lie between 62.8 and 130 kg s^{-1} for C method, from 90.26 to 115.74 kg s^{-1} for TL method, and 87.5 – 113.5 kg s^{-1} for HE method. Note that the smaller the obtained surface from the curves at each selected number of iterations, the higher the portion of estimates being close to the real solution. Similar results, we can observe for 10000 iterations, however, the number of occurrences is much higher than in the case of 1000 iterations.

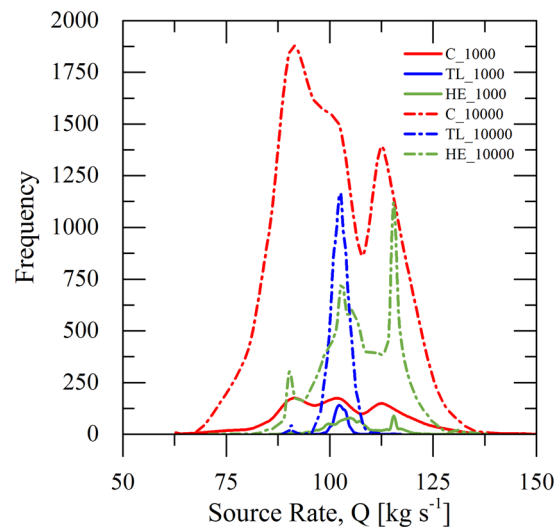


Figure 3. Values of number of occurrences versus the source rate estimation of 1000 (solid line) and 10000 (short dash dot line) iterations, respectively, for Concentration (*C*; red colour), Toxic Load (*TL*; blue colour) and Health Effects (*HE*; green colour) considered parameters. The number of selected receptors is taken equal to 100.

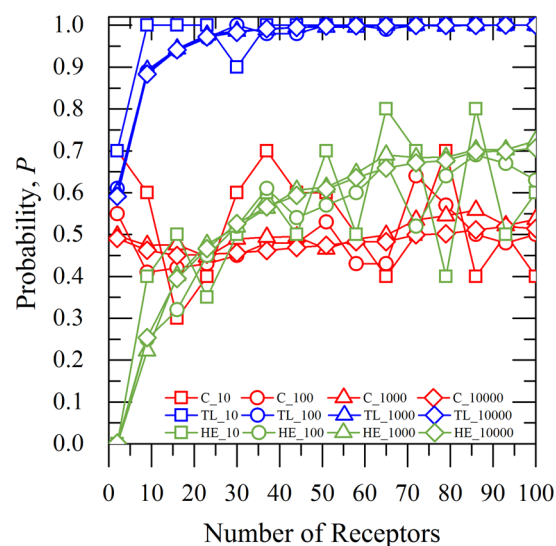


Figure 4. Values of Probability factor depending on the selected number of receptors. Number of 10 (\square), 100 (\circ), 1000 (Δ) and 10000 (\diamond) iterations, respectively, for Concentration (*C*; \color{red}), Toxic Load (*TL*; \color{blue}) and Health Effects (*HE*; \color{green}) considered parameters.

Another important factor of assessing our results is to compute the probability of the estimated values of source rate to approach the real value versus the selected number of receptors (sample points)³², as shown in Fig. 4. As the probability values approach one, the estimated value of source rate is close to real source rate. From Fig. 4, it is also seen that the highest probability values are obtained from the *TL* method, which after a certain number of receptors (~ 40) approaches one, regardless of the selected number of iterations. For *C* dataset, we observe that the probability values are close to 0.5 for more than 50 receptors and large number of iterations (i.e. 1000 (red triangle symbol) and 10000 (red diamond symbol)). For *HE*, we notice that the probability of the solution approaches 0.7 for more than 60 receptors and in some cases (green square symbol) overpasses it even with relatively small number of iterations (i.e. 10). It is also indicated that the probability of the solution for *HE* is better than *C* for more than 25 receptors. It is worth mentioning that the increased number of iterations also helps to achieve faster convergence of the results.

On the other hand, the probability values for the estimated source rate are gradually decreasing for less than 50 receptors, however, in the case of *C* the decrease of the obtained probability is almost negligible for 100 (red circle symbol), 1000 (red triangle symbol) and 10000 (red diamond symbol) iterations, excluding the case of 10 (red square symbol) iterations.

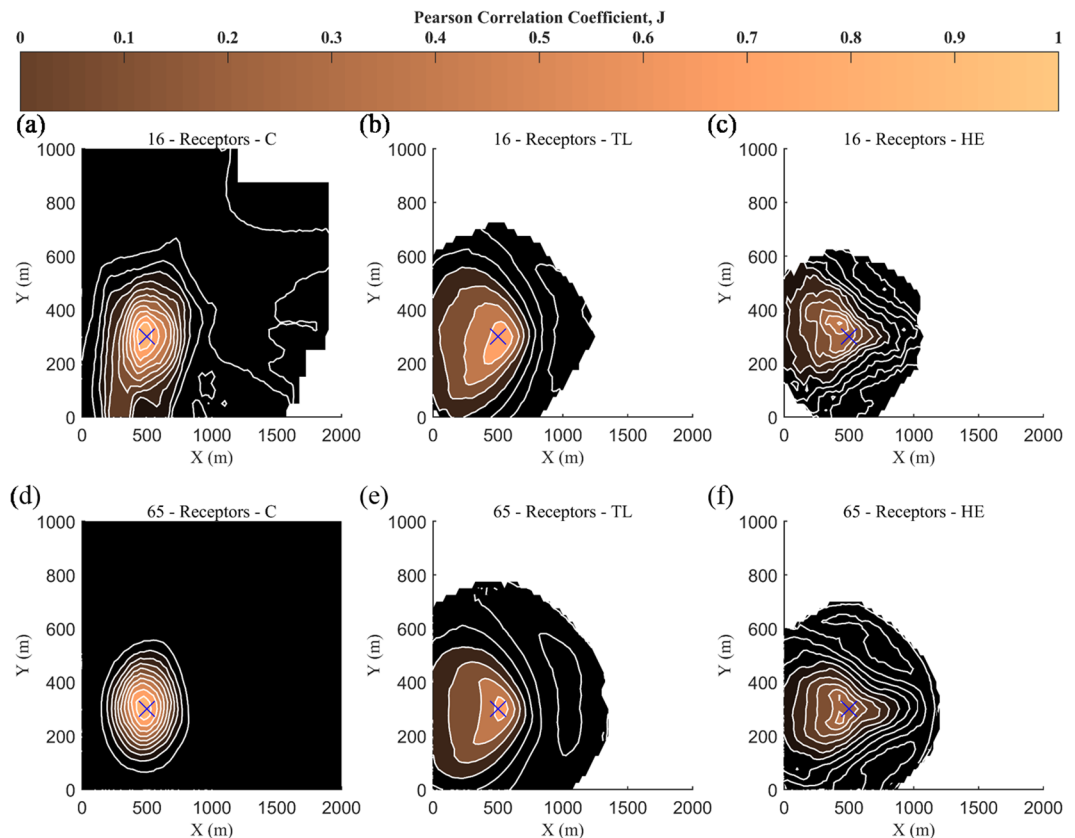


Figure 5. Spatial distribution of the correlation coefficient, J , for 16 (a–c) and 65 (d–f) receptors used for values of Concentration (a,d), Toxic Load (b,e) and Health Effects (c,f). The blue cross symbol denotes the real source location and the selected number of random iterations is taken equal to 100.

Prediction of source location. The location of the source release was predicted using three different input datasets (C , TL and HE), and each time the number of receptors was varied between 2 and 100, for four different number of iterations (10, 100, 1000 and 10000). The present results assess the performance of the proposed algorithm based on the different types of input data and number of receptors required in order to obtain the desired results (source location).

Figure 5 presents the spatial distribution of the correlation coefficient (J) corresponding to each guessed source location coordinates, for 16 and 65 receptors with input data from C , TL and HE and for 100 iterations. The colour bar depicts the colours that correspond to the values of Pearson Correlation Coefficient (PCC). Values close to one indicate that the solution is approaching the real release location (blue cross symbol). The prediction of source location is well identified, in all cases, regardless of the low J values for HE . This significant outcome indicates the advantages of the J compared to the NMSE for the source location estimation. As a result, we can claim that the source location is estimated with high accuracy when C or TL are available, and high probability when HE is adopted.

Figure 6 depicts the results for the prediction of source location using as input data health effect observations and 100 iterations. It is seen that the level of the complexity associated with the use of health effects as input parameter has effect on the minimum required number of receptors for the source location reconstruction. It is also noticed that after 58 receptors the reconstruction of source location is well predicted, but the required number of receptors for the desired results is larger than C and TL scenarios.

According to Fig. 7, it is observed that with increasing the number of iterations the results are converged faster and the obtained source location approaches the real one (blue cross symbol). It is also worth mentioning that even with small number of iterations and relatively small number of receptors (i.e. 37) the source location is well retrieved, in particular for the cases of C and TL input data.

Figure 8 depicts the dependence of the correlation coefficient (J) over the selected number of receptors. The obtained results are derived from 10, 100, 1000 and 10000 iterations for the corresponding C , TL and HE input. The correlation coefficient approaches the highest values around 0.95 for C (red solid line), after a certain number of receptors (less than 10) regardless of the set of iterations. Similar behaviour is observed for TL but the values of J are lower—around 0.8. On the other hand, the lowest values of J (~0.65) are obtained from the HE (green solid line), as shown in Fig. 8. In any case, uncertainty of the source location proves to be independent of the number of receptors and iterations for any test past the 10–20 receptors. In other words, a small number of receptors yields satisfactory values for J , as less as 10 receptors. It is also worth mentioning that no value for the source rate is necessary.

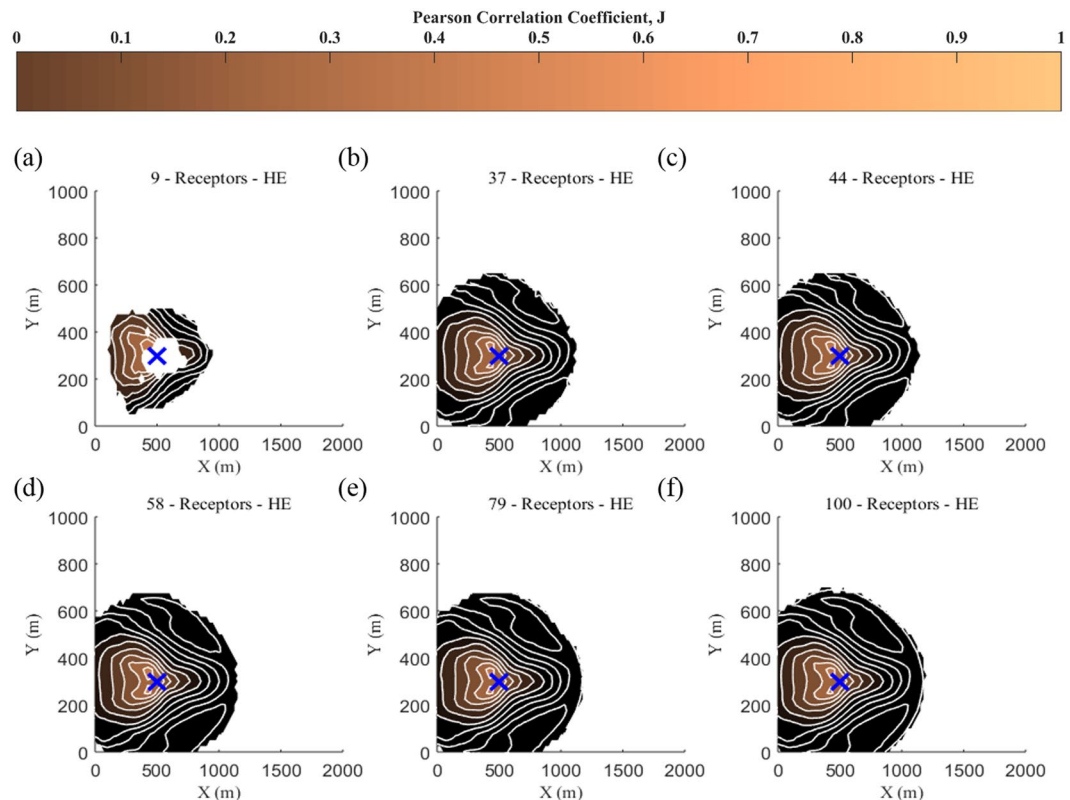


Figure 6. Spatial distribution of the correlation coefficient, J , for health effects observations and 100 iterations using (a) 16, (b) 37, (c) 44, (d) 58, (e) 79 and (f) 100 receptors. The blue cross symbol denotes the real source location.

Discussion

The sensitivity of the numerical results to the selected input data was gauged by varying the number of receptor points used in each scenario individually. With respect to the performance of the proposed scheme in reconstructing the source rate, the obtained information from the three considered methods (C , TL and HE) allowed the prediction of “averaged” source term rate in the range of $45\text{--}450\text{ kg s}^{-1}$. Therefore, a wide variation in the quality of the algorithm’s prediction was observed, in particular for the use of HE dataset as input parameter.

Our results show that the use of C data along with the increase of sample points used in reconstructing the release source rate leads to better convergence and agreement of the obtained results with the actual release source rate. The necessity to have large number of input data for the source reconstruction algorithm is due to the inherent variability and uncertainty of predicting the real release rate. Herein, this is reflected mostly with the introduced noise to the synthetic data, as discussed in Section 2. With the absence of artificial noise, this is a simple well-posed problem at which the obtained solution algorithm converges in few iterations together with a small number of receptors. However, in reality, uncertainty is high and only scarce number of data are typically available for use in the source rate estimation, thus the examined scenarios are only for exploratory purposes aimed primarily at assessing the limitations associated with the proposed computational framework.

By using the averaged source rate predictions obtained from the TL approach, it is seen that less variability and good agreement is achieved, when utilizing 10 receptor points and above (Fig. 2). The results are also independent for a selected number of 51 receptors and above. The different performance of the proposed reconstruction algorithm between C and TL input data can be attributed to the fact that the TL is estimated using an integral of time and hence it contains prior information (history) of the specified period at which the data are collected. With the use of HE as input parameter the performance of the method also improves with significantly reduced variability compared to the C input data. On the other hand, the prediction of the source location, using HE data, requires twice as much information (receptors and iterations) in order to achieve a comparable estimate as in the case of C and TL data (Fig. 7).

Another parameter that influences the performance of the algorithm and quality of the results is the number of iterations used in generating the source rate predictions. By increasing the number of iterations is shown to yield better convergence and less spread of the results around the average. This can be attributed to the use of large number of iterations, which accounts for different possible combinations of a certain number of receptors rather than only a single dataset of a given number of receptors. During realistic conditions, however, the errors of the obtained solutions are likely to appear within the confidence limits, which were estimated using the large number of iterations. It should also be noted that in realistic conditions the errors of model predictions might be larger than those generated in simulation exercises. Therefore, the confidence intervals may be also larger. The use of

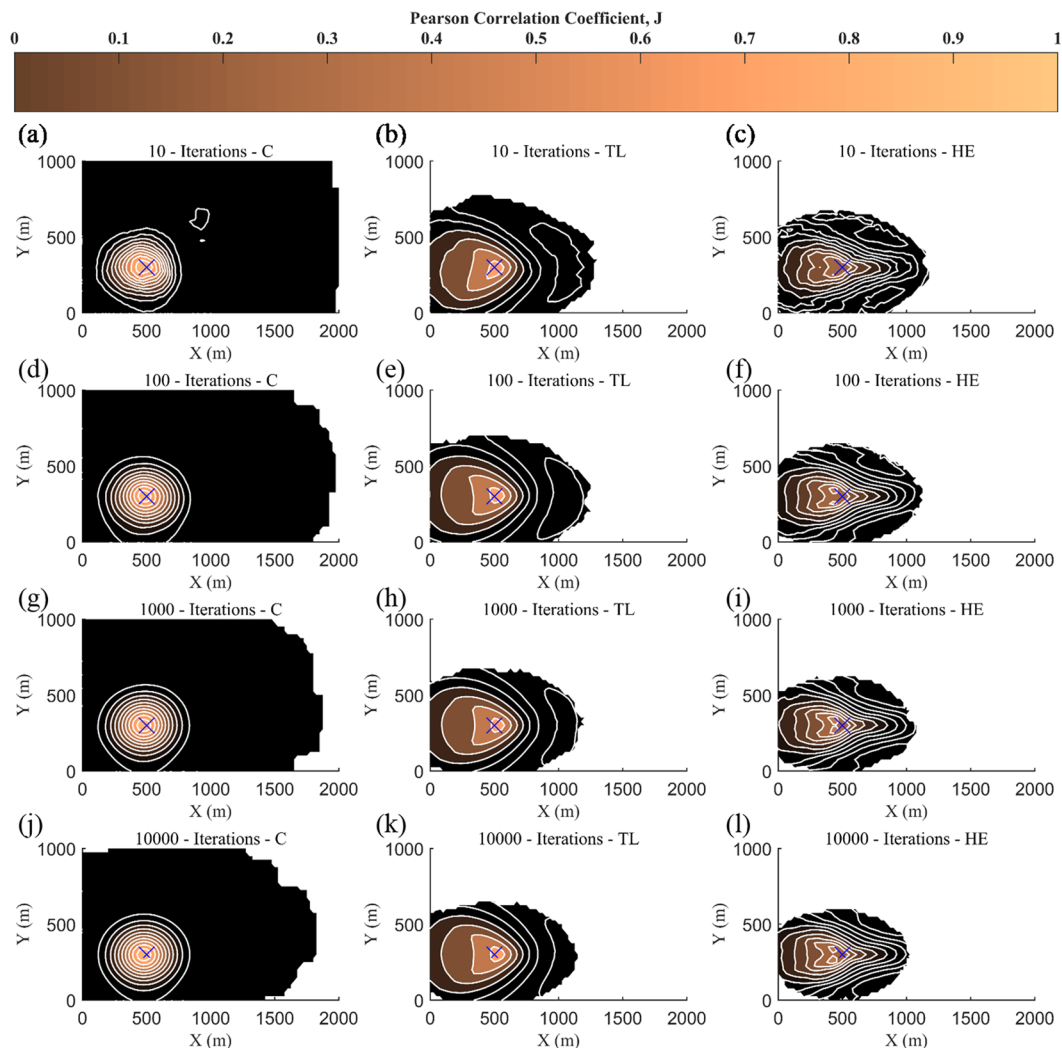


Figure 7. Spatial distribution of the correlation coefficient, J , for values of Concentration (**a,d,g,j**), Toxic Load (**b,e,h,k**) and Health Effects (**c,f,i,l**) using 10 iterations (**a,b,c**), 100 iterations (**d,e,f**), 1000 iterations (**g,h,i**) and 10000 iterations (**j,k,l**). The used number of receptors is 37. The blue cross symbol denotes the real source location.

random combination herein is for exploratory purposes to assess the performance of the algorithm against the uncertainty related to the use of sets of points with the same size but collected at different locations.

Nonetheless, there is room for future work in order to increase the accuracy of the results. Such as the use of more advanced dose-response models (e.g. PBTK), inclusion of variable meteorological data and site topography (e.g. CFD models) and more complex information related to the toxic release (e.g. used in asymmetric attacks or transport accidents, degradation of agents⁷⁶, and multiple agents/sources⁷⁷). Overall uncertainty could be reduced by the utilization of other statistical tools like Markov chain Monte Carlo, Bayesian inference, and clustering to determine the probability of the solution²². Alternatively, the implementation of more advanced numerical algorithms could partially overcome the limitations of the simple dispersion model. Ma and Zhang⁷⁸ improved the source parameters estimation, of a simple Gaussian model, by combining an integrated Gaussian-Machine Learning model with the particle swarm optimization approach and later⁷⁹ introducing the Tikhonov regularization approach. Further improvement and increasing of the accuracy of the results may be gained by including in the formulation observations from multiple time periods, instead of a single time-integration (dose) per location (sampling point) as it was demonstrated in other cases⁸⁰ and herein. Finally, pre-calculated response (health impact) databases could be combined with stochastic methods such as artificial neural networks to develop response-source correlations which then could be used with real-time observations to reconstruct the source-parameters, similarly to dispersion pattern recreation without the source knowledge⁸¹.

Conclusions

The work described herein focuses on the prediction of the released amount of a hazardous substance and release location using health observations instead of concentration levels (standard approach) or dosages in some cases. The proposed methodology is based on the coupling of multiple methods to resolve the

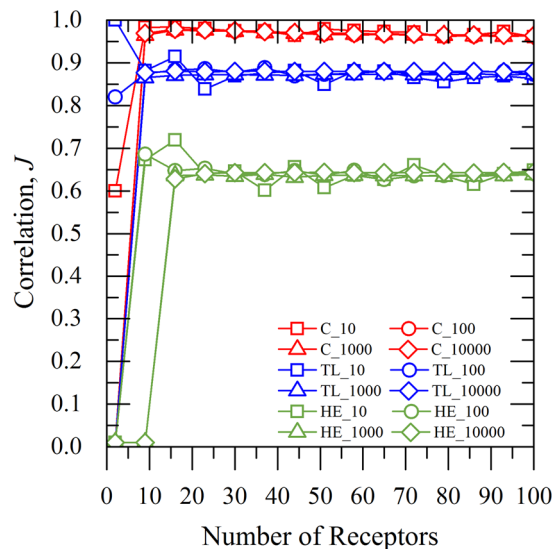


Figure 8. Values of correlation factor depending on the selected number of receptors. Number of 10 (\square), 100 (\circ), 1000 (\triangle) and 10000 (\diamond) iterations, respectively, for Concentration (C; \color{red}), Toxic Load (TL; \color{blue}) and Health Effects (HE; \color{green}) considered parameters.

source-concentration-exposure-dose-response binary relationships and fully elaborate the consequences analysis paradigm.

The results present good predictions of the source characteristics using the health observations as input for the proposed source reconstruction algorithm. The accuracy of the method, in reconstructing the source parameters decreases when a small set of observations or input data are available. Nonetheless, it is concluded that the current methodology is an appropriate tool for advance emergency preparedness and response during a HazMat release in urban environments, and that this is done within practical computer resources.

In a real life situation, the application of source-reconstruction approaches is challenging because of the lack of multiple monitoring stations and of the dependence, amongst others, on the meteorological conditions, emissions, and the released agent characteristics. It is also difficult to collect such information a priori. On the contrary, the proposed methodology is one step closer to the practical applications since it employs symptom-observations as input, instead of concentrations, as in most earlier works. There is room for much future work such as the use of the proposed algorithm with ADREA CFD code for addressing the same problem in urban environments^{58,82}. The further investigation of human behaviour during an evacuation from a threatened area to a safe place⁸³ and the calculation of toxic load versus the infrastructure effects (e.g. buildings, obstacles, trees, etc) could also be beneficial for the safety engineers.

References

1. Beck, U. *Risk Society: Towards a new modernity*. (Sage, 1992).
2. Giddens A. Risk and Responsibility. *Modern Law Review* **62**, 1–10 (1999).
3. Ashraf, A. M., Argyropoulos, C. D., Olewski, T., Vechot, L. & Kakosimos, K. E. Comparative study on toxic gas infiltration in a non-process area using CFD and multi-zone models. In *Hazards 26*: Edinburgh International Conference Center, Institution of Chemical Engineers Symposium Series No 161, Edinburgh, UK, 2016.
4. Argyropoulos, C. D., Sideris, G. M., Christolis, M. N., Nivolianitou, Z. & Markatos, N. C. Modelling pollutants dispersion and plume rise from large hydrocarbon tank fires in neutrally stratified atmosphere. *Atmos. Environ.* **44**, 803–813 (2010).
5. Argyropoulos, C. D., Ashraf, A. M., Vechot, L. & Kakosimos, K. E. Coupling multi-zone and CFD models for investigating indoor air quality. In *Second International Conference on Energy and Indoor Environment for Hot Climates*, ASHRAE: Doha, Qatar, 2017.
6. Morita, H. *et al.* Sarin poisoning in Matsumoto, Japan. *The Lancet* **346**, 290–293 (1995).
7. Okumura, T. *et al.* Report on 640 Victims of the Tokyo Subway Sarin Attack. *Ann. Emerg. Med.* **28**, 129–135 (1996).
8. Chouhan, T. R. The unfolding of Bhopal disaster. *J. Loss Prevent. Proc. Ind.* **18**, 205–208 (2005).
9. Pocchiari, F., Silano, V. & Zampieri, A. Human health effects from accidental release of tetrachlorodibenzo-p-dioxin at Seveso, Italy. *Ann. N.Y. Acad. Sci.* **320**, 311–320 (1979).
10. Assael, M. J. & Kakosimos, K. E. *Fires, Explosions, and Toxic Gas Dispersions*. (CRC Press, 2010).
11. Korsakissok, L., Mathieu, A. & Didier, D. Atmospheric dispersion and ground deposition induced by the Fukushima Nuclear Power Plant accident: A local-scale simulation and sensitivity study. *Atmos. Environ.* **70**, 267–279 (2013).
12. Wheeler, T. A., Hora, S. C., Cramond, W. R. & Unwin, S. D. Analysis of core damage frequency from internal events: Expert judgment elicitation. Report No. No. NUREG/CR-4550, SAND86-2084, (Sandia National Laboratories, Albuquerque, NM 87185, New Mexico, USA, 1989).
13. Bongers, S. *et al.* Challenges of exposure assessment for health studies in the aftermath of chemical incidents and disasters. *J. Expos. Sci. Environ. Epidemiol.* **18**, 341–359 (2008).
14. Zhan, Y. *et al.* Estimating Emissions and Environmental Fate of Di-(2-ethylhexyl) Phthalate in Yangtze River Delta, China: Application of Inverse Modeling. *Environ. Sci. Technol.* **50**, 2450–2458, (2016).
15. Rao, S. K. Source estimation methods for atmospheric dispersion. *Atmos. Environ.* **41**, 6964–6973 (2007).
16. Zhang, X. L. *et al.* Iterative ensemble Kalman filter for atmospheric dispersion in nuclear accidents: An application to Kincaid tracer experiment. *J. Hazard. Mater.* **297**, 329–339 (2015).

17. Zhang, X. L., Su, G. F., Yuan, H. Y., Chen, J. G. & Huang, Q. Y. Modified ensemble Kalman filter for nuclear accident atmospheric dispersion: Prediction improved and source estimated. *J. Hazard. Mater.* **280**, 143–155 (2014).
18. Zhang, X. & Huang, M. Ensemble-based release estimation for accidental river pollution with known source position. *J. Hazard. Mater.* **333**, 99–108 (2017).
19. Sharan, M., Singh, S. K. & Issartel, J. P. Least square data assimilation for identification of the point source emissions. *Pure Appl. Geophys.* **169**, 483–497 (2012).
20. Issartel, J. P. Emergence of a tracer source from air concentration measurements, a new strategy for linear assimilation. *Atmos. Chem. Phys.* **5**, 249–273 (2005).
21. Broyden, C. G. The convergence of a class of double-rank minimization algorithms: 2. The New Algorithm. *J. Inst. Maths Appl.* **6**, 222–231 (1970).
22. Zheng, X. & Chen, Z. Back-calculation of the strength and location of hazardous materials releases using the pattern search method. *J. Hazard. Mater.* **183**, 474–481 (2010).
23. Kirkpatrick, S. Optimization by simulated annealing: Quantitative studies. *J. Stat. Phys.* **34**, 975–986 (1984).
24. Haupt, S. E., Young, G. S. & Allen, C. T. Validation of a receptor–dispersion model coupled with a genetic algorithm using synthetic data. *J. Appl. Meteorol. Climatol.* **45**, 476–490 (2006).
25. Keats, A., Yee, E. & Lien, F.-S. Bayesian inference for source determination with applications to a complex urban environment. *Atmos. Environ.* **41**, 465–479 (2007).
26. Senocak, I., Hengartner, N. W., Short, M. B. & Daniel, W. B. Stochastic event reconstruction of atmospheric contaminant dispersion using Bayesian inference. *Atmos. Environ.* **42**, 7718–7727 (2008).
27. Wade, D. & Senocak, I. Stochastic reconstruction of multiple source atmospheric contaminant dispersion events. *Atmos. Environ.* **74**, 45–51 (2013).
28. Chow, F. K., Kosović, B. & Chan, S. Source inversion for contaminant plume dispersion in urban environments using building-resolving simulations. *J. Appl. Meteor. Climatol.* **47**, 1553–1572 (2008).
29. Hutchinson, M., Oh, H. & Chen, W.-H. A review of source term estimation methods for atmospheric dispersion events using static or mobile sensors. *Information Fusion* **36**, 130–148 (2017).
30. Courtier, P., Derber, J., Errico, R. O. N., Louis, J.-F. & Vukićević, T. Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology. *Tellus A* **45**, 342–357 (1993).
31. Marchuk, G. I. V. 295 *Mathematics and Its Applications* 468 (Springer Netherlands, 1995).
32. Kovalets, I. V., Andronopoulos, S., Venetsanos, A. G. & Bartzis, J. G. Identification of strength and location of stationary point source of atmospheric pollutant in urban conditions using computational fluid dynamics model. *Math Comput Simulat* **82**, 244–257 (2011).
33. Bocquet, M. *et al.* Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models. *Atmos. Chem. Phys.* **15**, 5325–5358 (2015).
34. Drews, M., Lauritzen, B. & Madsen, H. Analysis of a Kalman filter based method for on-line estimation of atmospheric dispersion parameters using radiation monitoring data. *Radiat. Prot. Dosim.* **113**, 75–89 (2005).
35. Zheng, D. Q., Leung, J. K. C., Lee, B. Y. & Lam, H. Y. Data assimilation in the atmospheric dispersion model for nuclear accident assessments. *Atmos. Environ.* **41**, 2438–2446 (2007).
36. Pudykiewicz, J. A. Application of adjoint tracer transport equations for evaluating source parameters. *Atmos. Environ.* **32**, 3039–3050 (1998).
37. Kumar, P., Feiz, A.-A., Singh, S. K., Ngai, P. & Turbelin, G. Reconstruction of an atmospheric tracer source in an urban-like environment. *Journal of Geophysical Research: Atmospheres* **120**, 12589–12604 (2015).
38. Elbern, H., Schmidt, H., Talagrand, O. & Ebel, A. 4D-variational data assimilation with an adjoint air quality model for emission analysis. *Environ Modell Softw* **15**, 539–548 (2000).
39. Robertson, L. & Langner, J. Source function estimate by means of variational data assimilation applied to the ETEX-I tracer experiment. *Atmos. Environ.* **32**, 4219–4225 (1998).
40. Say, D. *et al.* Re-Evaluation of the UK's HFC-134a emissions inventory based on atmospheric observations. *Environ. Sci. Technol.* **50**, 11129–11136 (2016).
41. Schaubberger, G., Piringer, M., Baumann-Stanzer, K., Knauder, W. & Petz, E. Use of a Monte Carlo technique to complete a fragmented set of H₂S emission rates from a wastewater treatment plant. *J. Hazard. Mater.* **263**, 694–701 (2013).
42. Lioy, P. J. Exposure science: A view of the past and milestones for the future. *Environ. Health Perspect.* **118**, 1081–1090 (2010).
43. Pirkle, J. L., Sampson, E. J., Needham, L. L., Patterson, D. G. & Ashley, D. L. Using biological monitoring to assess human exposure to priority toxicants. *Environ. Health Perspect.* **103**, 45–48 (1995).
44. Hays, S. M., Becker, R. A., Leung, H. W., Aylward, L. L. & Pyatt, D. W. Biomonitoring equivalents: A screening approach for interpreting biomonitoring results from a public health risk perspective. *Regul. Toxicol. Pharm.* **47**, 96–109 (2007).
45. Georgopoulos, P. G. *et al.* Reconstructing population exposures to environmental chemicals from biomarkers: Challenges and opportunities. *J Expos Sci Environ Epidemiol* **19**, 149–171 (2009).
46. Mumtaz, M., Fisher, J., Blount, B. & Ruiz, P. Application of physiologically based pharmacokinetic models in chemical risk assessment. *Journal of Toxicology* **2012**, 904603 (2012).
47. US-EPA. Approaches for the application of Physiologically Based Pharmacokinetic (PBPK) models and Supporting data in risk assessment. (U.S. Environmental Protection Agency, Washington, DC, 2006).
48. Boris, J. P. In *Dust in the wind: challenges for urban aerodynamics*, 35th AIAA Fluid Dynamics Conference and Exhibit. (American Institute of Aeronautics and Astronautics, 2005).
49. Chen, C.-C., Shih, M.-C. & Wu, K.-Y. Exposure estimation using repeated blood concentration measurements. *Stochastic Environmental Research and Risk Assessment* **24**, 445–454 (2010).
50. Chen, C.-C., Shih, M.-C. & Wu, K.-Y. Exposure reconstruction using a physiologically based toxicokinetic model with cumulative amount of metabolite in urine: a case study of trichloroethylene inhalation. *Stochastic Environmental Research and Risk Assessment* **26**, 21–31 (2012).
51. Christensen, K. L. Y., Lorber, M., Ye, X. & Calafat, A. M. Reconstruction of bisphenol A intake using a simple pharmacokinetic model. *J Expos Sci Environ Epidemiol* **25**, 240–248 (2015).
52. Brown, K. *et al.* Reconstructing exposures from biomarkers using exposure-pharmacokinetic modeling – A case study with carbaryl. *Regul. Toxicol. Pharm.* **73**, 689–698 (2015).
53. Ermak, D. L. *User's manual for SLAB: An atmospheric dispersion model for denser than air release.* (Lawrence Livermore National Laboratory, 1990).
54. COST Action ES1006. Evaluation, improvement and guidance for the use of local-scale emergency prediction and response tools for airborne hazards in built environments (2015).
55. Allen, C. T., Young, G. S. & Haupt, S. E. Improving pollutant source characterization by better estimating wind direction with a genetic algorithm. *Atmos. Environ.* **41**, 2283–2289 (2007).
56. Kolczynski, W. C., Stauffer, D. R., Haupt, S. E. & Deng, A. Ensemble variance calibration for representing meteorological uncertainty for atmospheric transport and dispersion modeling. *J. Appl. Meteorol. Climatol.* **48**, 2001–2021 (2009).
57. Zajackowski, F. J., Haupt, S. E. & Schmehl, K. J. A preliminary study of assimilating numerical weather prediction data into computational fluid dynamics models for wind prediction. *J. Wind Eng. Ind. Aerodyn.* **99**, 320–329 (2011).

58. Efthimiou, G. C. *et al.* An optimized inverse modelling method for determining the location and strength of a point source releasing airborne material in urban environment. *Atmos. Environ.* **170**, 118–129 (2017).
59. Kovalets, I. V. *et al.* Inverse identification of an unknown finite-duration air pollutant release from a point source in urban environment. *Atmos. Environ.* **181**, 82–96 (2018).
60. Hessel, P. A., Herbert, F. A., Melenka, L. S., Yoshida, K. & Nakaza, M. Lung health in relation to hydrogen sulfide exposure in oil and gas workers in Alberta, Canada. *Am. J. Ind. Med.* **31**, 554–557 (1997).
61. Lambert, T. W., Goodwin, V. M., Stefani, D. & Stroscher, L. Hydrogen sulfide (H₂S) and sour gas effects on the eye. A historical perspective. *Sci. Total Environ.* **367**, 1–22 (2006).
62. Argyropoulos, C. D., Ashraf, A. M., Markatos, N. C. & Kakosimos, K. E. Mathematical modelling and computer simulation of toxic gas building infiltration. *Process Safety and Environmental Protection* **111**, 687–700 (2017).
63. Ermak, D. L., Chan, S. T., Morgan, D. L. & Morris, L. K. A comparison of dense gas dispersion model simulations with burro series LNG spill test results. *J. Hazard. Mater.* **6**, 129–160 (1982).
64. Mannan, M. S. *Lees' loss prevention in the process industries: Hazard identification, assessment and control*. 4 edn, Vol. 1 (Butterworth-Heinemann, 2012).
65. Lioy, P. J., Laskin, J. D. & Georgopoulos, P. G. Preparedness and response to chemical and biological threats: the role of exposure science. *Ann. N.Y. Acad. Sci.* 1–10 (2016).
66. Crump, K. S. *et al.* What role for Biologically Based Dose–Response models in estimating low-dose Risk? *Environ. Health Perspect.* **118**, 585–588 (2010).
67. HSE. Indicative human vulnerability to the hazardous agents present offshore for application in risk assessment of major accidents. (Health and Safety Executive, SPC/Tech/OSD/30, England, 2013).
68. Ten Berge, W. F., Zwart, A. & Appelman, L. M. Concentration–time mortality response relationship of irritant and systemically acting vapours and gases. *J. Hazard. Mater.* **13**, 301–309 (1986).
69. Boris, J. P. & Patnaik, G. Acute Exposure Guideline Levels (AEGs) for Time Varying Toxic Plumes. Report No. NRL/MR/6003–14-9493, 1–29 (Laboratories for Computational Physics and Fluid Dynamics, Naval Research Laboratory, Washington, DC, 2014).
70. Oran, E. S. & Boris, J. P. *Numerical simulation of reactive flow: 5-5.1, The Induction Parameter Model*. 150–151 (2nd edition, Cambridge University Press, New York, N.Y., 2001).
71. Gunatilaka, A., Skvortsov, A. & Gailis, R. A review of toxicity models for realistic atmospheric applications. *Atmos. Environ.* **84**, 230–243 (2014).
72. Brent, R. P. *Algorithms for minimization without derivatives*. 195 (Prentice-Hall, Englewood Cliffs, 1973).
73. Forsythe, G. E., Malcolm, M. A. & Moler, C. B. *Computer methods for mathematical computations*. 1 edn, 270 (Prentice Hall, 1977).
74. Pearson, K. Note on Regression and Inheritance in the Case of Two Parents. *Proceedings of the Royal Society of London* **58**, 240–242 (1895).
75. Efthimiou, G. C. *et al.* Modification and validation of a method for estimating the location of a point stationary source of passive non-reactive pollutant in an urban environment. In *17th International Conference of Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes*, Budapest, Hungary, 2016.
76. Stuart, A. L. & Wilkening, D. A. Degradation of biological weapons agents in the environment: Implications for terrorism response. *Environ. Sci. Technol.* **39**, 2736–2743 (2005).
77. Zhang, X., Raskob, W., Landman, C., Trybushnyi, D. & Li, Y. Sequential multi-nuclide emission rate estimation method based on gamma dose rate measurement for nuclear emergency management. *J. Hazard. Mater.* **325**, 288–300 (2017).
78. Ma, D. & Zhang, Z. Contaminant dispersion prediction and source estimation with integrated Gaussian-machine learning network model for point source emission in atmosphere. *J. Hazard. Mater.* **311**, 237–245 (2016).
79. Ma, D., Tan, W., Zhang, Z. & Hu, J. Parameter identification for continuous point emission source based on Tikhonov regularization method coupled with particle swarm optimization algorithm. *J. Hazard. Mater.* **325**, 239–250 (2017).
80. Sung, L. Y., Shie, R. H. & Lu, C. J. Locating sources of hazardous gas emissions using dual pollution rose plots and open path Fourier transform infrared spectroscopy. *J. Hazard. Mater.* **265**, 30–40 (2014).
81. Wang, B., Chen, B. & Zhao, J. The real-time estimation of hazardous gas dispersion by the integration of gas detectors, neural network and gas dispersion models. *J. Hazard. Mater.* **300**, 433–442 (2015).
82. Efthimiou, G. C. *et al.* In *Validation of an inverse method for the source determination of a hazardous airborne material released from a point source in an urban environment*. *Air Pollution Modeling and its Application XXV* 329–332 (Springer International Publishing, Cham, 2018).
83. Argyropoulos, C. D., Ashraf, A. M. & Kakosimos, K. E. Multi-scale modelling pollutant dispersion and exposure against an accidental toxic release in an administration building using large eddy simulation and evacuation models. In *18th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory purposes*, Bologna, Italy, 2017.

Acknowledgements

This publication was made possible by a NPRP award [NPRP 7-674-2-252] from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors. The High Performance Computing resources and services used in this work were provided by the IT Research Computing group in Texas A&M University at Qatar.

Author Contributions

C.A. conducted the calculations and analysed the results. He prepared the main text of the manuscript along with the post-processing, and also contributed on the literature review. K.K. conceived the concept presented here and developed the algorithm used for the calculations. He also contributed on the manuscript preparation. S.E. prepared the first state-of-the-art review and helped developing the algorithm and preliminary results. S.A. and A.V. contributed on the refining of the concept and the results interpretation. E.F. assisted on the development of section 2.3. G.E. and I.K. developed the basis of section 2.4 and assisted on the results interpretation. All authors reviewed the manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018