Available online at www.sciencedirect.com
ScienceDirect
Atmospheric Environment 41 (2007) 1128–1134
ATMOSPHERIC ENVIRONMENT
www.elsevier.com/locate/atmosenv

An improved algorithm for locating a gas source using inverse methods

Laura C. Thomson[a,*], Bill Hirst[b], Graham Gibson[a], Steve Gillespie[c], Philip Jonathan[c], Kenneth D. Skeldon[a], Miles J. Padgett[a]

[a]*Department of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK*
[b]*Shell Global Solutions International BV, Amsterdam, Netherlands*
[c]*Shell Global Solutions (UK), Cheshire Innovation Park, Chester, UK*

Received 19 May 2006; received in revised form 28 September 2006; accepted 3 October 2006

Abstract

We apply an inverse problem approach to locating a known gas source in a desert setting from simultaneous measurements of gas concentration and wind data. We use a random search algorithm with simulated annealing to generate candidate distributions of source strengths and positions. These distributions are then assessed by means of a cost function, which quantifies the degree to which the postulated source distribution accounts for the measured gas concentrations. We present results from using three cost functions with differing regularisation terms. We assess the robustness of these and the differing regularisation terms by the progressive addition of random noise and systematic offsets to the concentration data. We show that for our application, the best reconstructions are obtained by using a multiplicative regularisation parameter defined to minimise the total gas emissions.

Keywords: Inverse problem; Gas dispersion; Random search; Simulated annealing; Cost function

1. Introduction

Numerous gas dispersion models exist that seek to predict gas concentrations resulting from a known source dispersing into the atmosphere. However, the inverse of this process, locating an unknown gas source from a measurement of concentration is ambiguous: a small source close to the measurement location can give the same concentration as a large source further away. This ambiguity can be reduced by making multiple measurements of gas concentration at different positions and/or under differing wind conditions. Such an approach, combined with inverse dispersion techniques, has been used to estimate gas emissions and source locations for agricultural settings (Flescha et al., 2005) and land-fills (Lehning et al., 1994).

The increasing difficulty of finding new oil and gas reserves has prompted the development of several novel prospecting tools, particularly those targeted at frontier exploration: where large areas must be screened prior to more detailed investigation. Examples of such techniques include searching

*Corresponding author.
E-mail address: L.Thomson@physics.gla.ac.uk (L.C. Thomson).

1352-2310/$ - see front matter
doi:10.1016/j.atmosenv.2006.10.003

satellite and airborne images for subtle signs of oil on the sea surface (MacDonald et al., 1996) and stimulation of fluorescence in these oil films by airborne lasers (Martin and Cawley, 1991). Another approach is based on the well-known fact that all reservoirs leak some hydrocarbons to the surface (Horvitz, 1985; Schumacher, 1999), and of these thermogenic ethane provides a useful tracer. Hirst et al. (2004) describes how measurements of atmospheric ethane concentration can be used to remotely locate ground-level sources of ethane escaping into the atmosphere. These ethane seepages are evidence of subsurface hydrocarbon systems that might be associated hydrocarbon accumulations.

Using a highly sensitive atmospheric ethane sensor based on laser diode absorption spectroscopy described fully by Gibson et al. (2002), with a 3D ultra-sonic anemometer we have carried out multiple surveys in the Middle East, each of which covered multiple hundreds of square kilometres. However, to reliably interpret these data, a verified and robust algorithm is needed to determine the positions and strengths of the sources responsible for the observed data. To that end, the reconstructions within this paper were produced using data collected from a single calibrated gas source at a known location. This survey site was a flat desert location meaning that the wind field could be assumed consistent over the area of interest. The source comprised a continuously weighed cylinder of ethane with a regulated release rate of $5\,kg\,h^{-1}$. We used a 200 m diameter network of perforated pipes to distribute the gas to simulate a diffuse area source. The gas concentrations were measured at a height of 5 m above the ground collocated with the anemometer at seven downwind locations. These locations were 2–5 km downwind of the release in a North West direction. Gas concentration and wind measurements were made for a period of at least 15 min, before moving to the next location. The whole measurement sequence was completed over 4 h from noon i.e. during the most stable wind conditions. It is normal when using a Gaussian plume model to average the short timescale data into bins, the duration of which is a significant fraction of the time of flight from source to measurement location (Gifford, 1976). However, under the stable wind conditions we experienced in the desert we found that a 1 min bin time was a reasonable compromise between representative data and data collection rate. Hence the 1 s data were averaged into 1 min bins to give approximately 100 measurements, each comprising: gas concentration, the average wind speed and direction, along with horizontal and vertical standard deviations of wind direction. These deviations in wind direction characterise the growth in the width and height of the gas plume with distance and were averaged over the whole data set at each location (15 min). A representative example of the 1 min data for three locations is given in Fig. 1. Based upon the reproducibility of the gas sensor (when observing a fixed concentration) we estimate the standard deviation of concentration measurements averaged over 1 min is less than 0.2 ppb.

Position (No.)	Concentration (Kg/m^3)	Wind Direction (°)	Wind Speed (m/s)	SD of Horizontal Wind Direction (°)	SD of Vertical Wind Direction (°)
82	1.17E-09	163.94	3.84	10.13	5.11
83	1.30E-09	177.77	5.66	11.26	2.37
84	2.62E-09	159.05	5.33	15.07	2.58

Fig. 1. A sample of 1 min averaged data recorded at three locations.

2. Inverse problems

Given a known gas source and wind field we could calculate the expected gas concentration for any downwind location. However, given a downwind concentration measurement and knowledge of the wind field, the location of the source is ambiguous. Non-inverting problems of this type are termed inverse problems: problems that can be solved in one direction but for some physical reason cannot be solved in the opposite direction. Such problems are widely encountered in several fields, most typically with determining an enhanced reconstruction of an object from a series of low quality images (Arridge, 1999). In all such inverse problems the aim is to infer the unknown state from measured consequences of that state. In the case of gas dispersion, the unknown state is the gas source distribution of strengths and locations; and the measured consequences are the gas concentrations for the associated wind conditions and measurement locations. Our aim is to find the source distribution that will generate predicted concentrations closest to those actually measured.

Gas dispersion is a complicated phenomenon with numerous physical processes active at different length scales (Draxler, 1976) in most experimental situations access to wind field data is limited, typically comprising a very small number of measurement locations. In common with much earlier work (Hirst et al., 2004; Gifford, 1976) we employ a simple Gaussian plume dispersion model which can be implemented using a single location wind speed and direction measurement. Within our numerical model we divide the 8 km × 8 km survey area into a 16 × 16 grid, with each cell, j, containing a central diffuse source of unknown strength S_j. Each of these M sources gives rise to a downwind gas concentration approximated by a Gaussian plume model (Gifford, 1976), see Fig. 2, and the predicted concentration is the sum over all the gas sources in the grid of cells. For the ith measurement of mass concentration we have

$$C_i = \sum_{j=1}^{M} \frac{S_j}{\pi V_i \sigma_{w_i} \sigma_{h_i}} \exp\left(-\frac{1}{2}\left(\frac{\Delta_{w_i}}{\sigma_{w_i}}\right)^2\right) \times \exp\left(-\frac{1}{2}\left(\frac{\Delta_{h_i}}{\sigma_{h_i}}\right)^2\right), \quad (1)$$

where V_i is the wind velocity at the time and location of the ith measurement, Δ_{w_i} and Δ_{h_i} are the offset from the plume centre, σ_{w_i} is the width of the Gaussian plume and σ_{h_i} is the height in the plume. Both σ_{w_i} and σ_{h_i} are length measurements and are calculated using the angular standard deviations of the wind direction in both the horizontal, θ_{w_i}, and vertical, θ_{h_i}, directions, respectively,

$$\sigma_{w_i} = \theta_{w_i} z + k, \quad (2)$$

$$\sigma_{h_i} = \theta_{h_i}(z + k), \quad (3)$$

where z is the distance from the source position, j, to the measurement location, i, and we have approximated the near field concentration effects of diffuse sources by adding a constant, k, to these plume width terms, where k is in units of cell size.

There are many search algorithms that could be used to identify the combination of source strengths and locations that best predict the measured concentrations. We employ an iterative direct search algorithm (Seldowitz et al., 1987) to postulate different source distributions. Within each iteration, a new postulated source distribution is generated by assigning a random source strength to a randomly selected cell, j. The predicted concentrations arising from the new set of postulated sources are then calculated from using the Gaussian plume model. The new postulated source distribution is numerically assessed by evaluating the chosen cost function, E_n, giving a single numerical value reflecting the closeness of the predicted concentrations to those actually measured. Typically, E_n would be defined such that a closer match between observed and predicted concentrations gives a lower value of E_n. The new postulated source value, S_j, is accepted or rejected depending on whether the corresponding value of E_n is reduced or not. Repeated iterations minimise the value of E_n and result in a source distribution that best accounts for the concentrations measured. An obvious candidate cost function would be that based solely on the predicted and measured concentration χ^2, for which

$$E_n = \frac{\chi^2}{N} = \frac{1}{N}\sum_{i=1}^{N} \frac{|C_{m_i} - C_{e_i}|^2}{\sigma_i^2}. \quad (4)$$

We note that an accurate determination of χ^2/N requires a correspondingly accurate knowledge of σ_i^2 which encompasses both the noise associated with the measurement technique and errors in the forward model. In Section 4 we will discuss how to estimate σ_i.

3. Simulated annealing

If we calculated the cost function over the full range of M source strengths, we would obtain an M dimensional landscape where each axis represented the range of emissions from the source in a

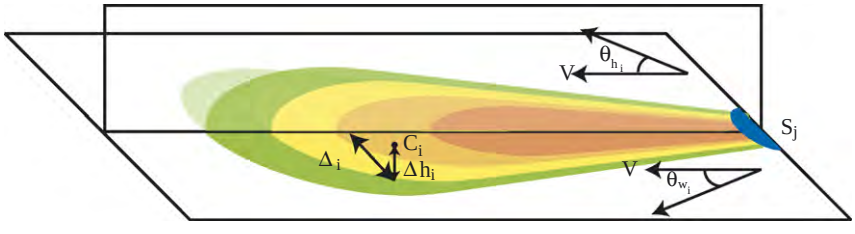

Fig. 2. The Gaussian plume model.

particular cell. Ideally, the resulting cost function landscape would contain a clearly identified global minimum, corresponding to the optimum distribution of M source strengths and locations. However, typically, this global minimum will be surrounded by local minima, to which the algorithm will converge to. We avoid this by refining the search algorithm to occasionally accept postulates that increase E_n, thereby allowing escape from a local minimum. The technique is referred to as simulated annealing and was developed by analogy with thermodynamic cooling (Press et al., 2002). The criterion for accepting an increase in the cost function is based by thermodynamic analogy on the Boltzman probability distribution,

$$R(0,1) < \exp\left(-\frac{E_n - E_{n-1}}{T_n}\right), \quad (5)$$

where $R(0,1)$ is a random number between zero and one, E_n is the value of the cost function for the new postulate, E_{n-1} is the value of the cost function at the last accepted change and T_n is the temperature or cooling parameter, which is reduced after each iteration. If $E_n < E_{n-1}$ that change is always accepted, however, if $E_n > E_{n-1}$ the change is sometimes accepted dependent on whether $\exp(E_n - E_{n-1}/T)$ is greater than the random number, R. Whether or not this holds true depends on the value of T. The value of T is reduced after each iteration according to

$$T_n = T_{n-1}(1-\varepsilon). \quad (6)$$

The initial iterations, corresponding to high values of T, accept most changes whether E_n is improved or not. As T is reduced fewer changes are accepted. Since the parameter T is described as a temperature it is usual to describe the rate of cooling in terms of thermal mass. For our reconstructions of source strengths on a 16×16 grid we typically run upwards of 4 million iterations during which time the annealing parameter, T, is progressively reduced to zero.

4. Choice of cost function

Although $E_n = \chi^2/N$ is a useful general indicator of the closeness of two data sets; for our application a simple minimisation of χ^2/N generally leads to an unrealistic set of source strengths. This arises because it is possible to over minimise χ^2/N: in effect one is fitting artefacts and noise in the data, (Press et al., 2002). To counter this it is usual to incorporate a regularisation term into the cost function. This is chosen to favour combinations of source strengths that correspond to physically realistic values and/or behaviours. The most obvious restriction is that the gas sources should be positive. This constraint is inherent in our approach because when postulating a new source strength, only positive values are considered. Another constraint is that in keeping with geophysical expectations we restrict the maximum source strength, thereby defining the limit of the parameter space. Our main goal within this paper is to compare the effects of different regularisation terms within the cost function. We seek a regularisation scheme that gives the most accurate reconstruction and is robust to likely sources of noise or error in the concentration data. The accuracy of a reconstruction can be judged both in terms of the predicted source location and source strength. Within our application it is the source position which is of most significance.

It is important to appreciate that the choice of cost function is independent of the method of minimisation. Providing we successfully minimise E_n then it is the choice of cost function alone that determines the distribution of source strengths and locations the algorithm identifies.

Initially we used $E_n = \chi^2/N$ as the cost function. Fig. 3 shows the corresponding $E_n = \chi^2/N$ reconstruction compared to the known release rate and position. Although the cylinder is approximately located this minimisation gives rise to strong sources at positions distant from the source location. Not only is this unlikely, but in our specific case of a known source, it is simply wrong. This illustrates the consequence of minimising χ^2/N, when the system is under-determined.

Many different regularisation terms are possible, ranging from those that have a statistical rationale

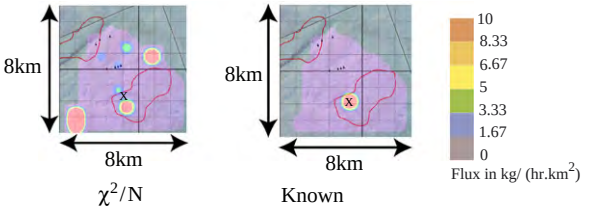

Fig. 3. A χ^2/N reconstruction compared to the known source distribution (the expected reconstruction for the area), the cylinder release area is shown with colour indicating ethane flux, with the scale shown.

to those that are more arbitrary and depend on the specific nature of the problem. Of the former perhaps the best known are those based on entropy: a maximum entropy cost function aims to minimise the disorder of the source distribution favouring distributions that are smooth. An alternative approach is based solely on the smoothness of the source strength distributions. A third, and distinctly different approach, is based on minimising the total emission rate. The maximum entropy cost function (maxent) is given by (Press et al., 2002)

$$E(\text{maxent}) = \frac{\chi^2}{N} + \lambda \sum_{j=1}^{M} S_j \ln\left(\frac{S_j}{\sum_{j=1}^{M} S_j}\right), \tag{7}$$

where λ is a factor that determines the relative importance of χ^2/N and the regularisation term. However, this requires an accurate knowledge of the noise distribution, σ, otherwise χ^2 itself is subject to an error, discussed in Section 2. We estimate the appropriate value of σ by simply running an initial χ^2/N reconstruction for a given data set and from this reconstruction calculating the standard deviation of the residuals, effectively normalising χ^2/N to unity. For the entropy regulariser there is a strong statistical argument for setting λ to bring χ^2/N to within a specified statistical range, $1 \pm \sqrt{2/N}$ (Press et al., 2002).

The maxent regularisation term is insensitive to the interchange of two or more sources. An alternative regularisation function is based on the smoothness of the resulting reconstruction and is sensitive to interchange of sources,

$$E(\text{smooth}) = \frac{\chi^2}{N} + \lambda \sum_{j=1}^{M} \frac{(S_j - \langle S_j \rangle_8)^2}{\alpha + (S_j - \langle S_j \rangle_8)^2}, \tag{8}$$

where each source strength is compared to the average of the eight sources in the grid cells immediately adjacent to j denoted by $\langle S_j \rangle_8$. This favours reconstructions that have strong correlations between neighbouring sources. Again λ dictates the weighting between χ^2/N and the regulariser. α has a similar effect on the function to that of λ but also prevents singularities in the cost function. Amongst reconstructions with the same value of χ^2/N, regularisation by either entropy or smoothness favours smooth reconstructions.

Another cost function we have used is based on our expectation that the sum of the sources be small (Hirst et al., 2004). Rather than implementing this as an additive term to the cost function, we used the total flux as a multiplicative factor,

$$E(\text{flux}) = \frac{\chi^2}{N} \lambda \sum_{j=1}^{M} \left(S_j + \frac{1}{\lambda}\right). \tag{9}$$

Again λ sets the degree of regularisation and also ensures the multiplicative factor cannot fall below unity.

We will now compare the performance of these three cost functions when applied to invert data for the known release. In all the reconstructions shown in this paper we have subtracted 0.7 ppb from all the concentration measurement data; this is to remove the effect of the background atmospheric ethane concentration at this latitude and time of year (Rudolph, 1995). We must next determine the optimum value of λ for each of the cost functions. This will be the value of λ that gives the reconstruction closest to the known result for that

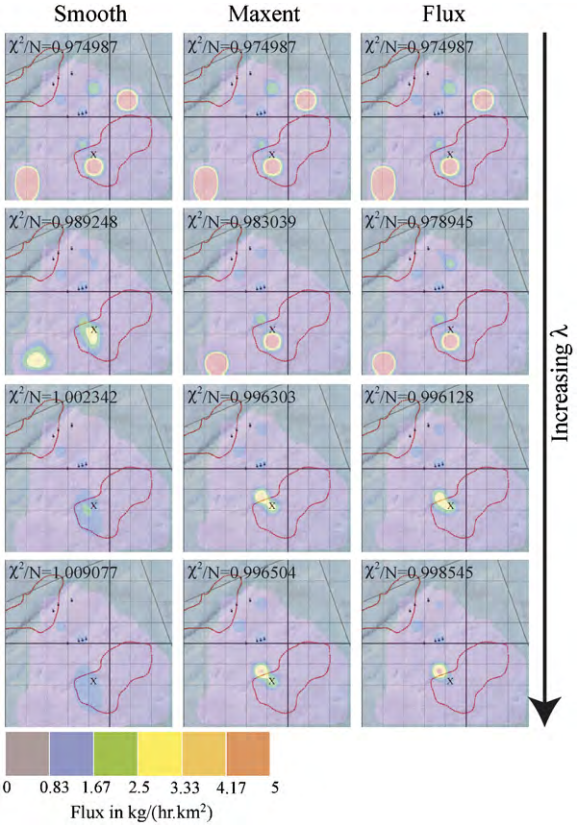

Fig. 4. Reconstructions shown for each cost function for a range of λ from $\lambda = 0$ (the χ^2/N solution) to a value which increases χ^2/N by a factor $1 + \sqrt{2/N}$. The location of the release is marked with an ×.

cost function. Fig. 4 shows reconstructions obtained using each of our three cost functions for a range of values of λ, λ is increased from 0 (i.e. the χ^2/N solution) to the value of λ which results in an increase of χ^2/N by a factor $1+\sqrt{2/N}$.

When using the optimised λ values, all three cost functions approximately locate the cylinder release. However, the smoothing cost function results in the source being, incorrectly, spread over adjacent cells. Next we shall establish which of the cost functions is most robust to likely sources of experimental uncertainty: namely added noise and baseline offsets in the concentration data.

5. Noise sensitivity

We are able to test the robustness of each cost function by modifying our controlled release data to include Gaussian noise added to the measured concentrations. In Fig. 5 we see that even adding up to 1.2 ppb of noise to the intrinsic sensor noise of 0.2 ppb does not prevent reasonable reconstruction of the gas source. This significant level of noise immunity arises because multiple measurements under similar wind directions are taken at each measurement location. As might be expected, we see that using the smoothing regularisation spreads the

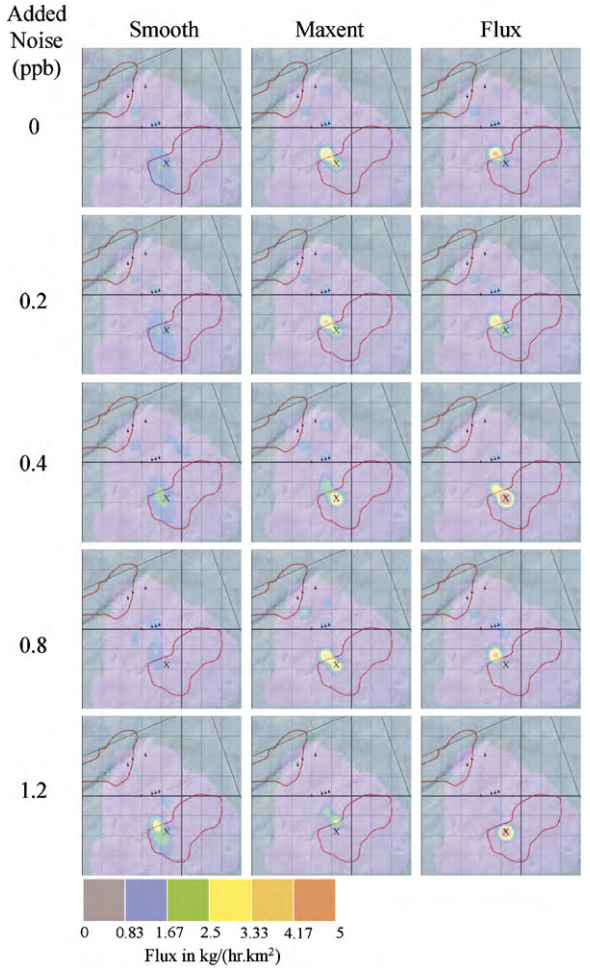

Fig. 5. Reconstructions for each cost function with increasing added noise, colour indicates ethane flux at source. The first row shows the reconstructions based on the original data. Subsequent rows show reconstructions for each of the cost functions as increasing amounts of random noise are added to the concentration data. The location of the release is marked with an ×.

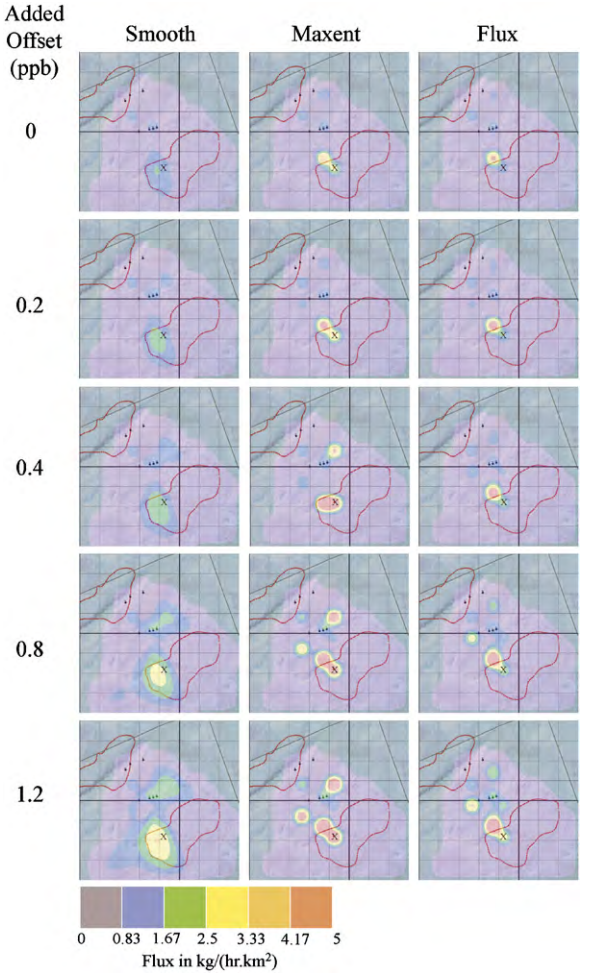

Fig. 6. Reconstructions for each cost function with increasing added concentration offsets, colour indicates ethane flux at source. The first row shows the reconstructions based on the original data. Subsequent rows show reconstructions for each of the cost functions as increasing amounts of added offset are added to the concentration data. The location of the release is marked with an ×.

reconstructed source over the neighbouring grid cells. The maxent and flux minimiser regularisation functions perform similarly with the latter finding the best overall reconstruction that achieved by adding 1.2 ppb of noise, see Fig. 5.

6. Offset sensitivity

We also want the reconstruction results to be robust to offsets in measured concentrations. This is particularly relevant because the exact background ethane concentration in the atmosphere is uncertain and subject to daily variation. In order to test robustness to offsets, we add a concentration offset to the data collected from the controlled release and calculate the reconstructions using each of the cost functions. Fig. 6 shows reconstructions for each of the three cost functions with progressively increasing offsets added to the concentration measurements. From the reconstructions shown in Fig. 6 we draw similar conclusions to those made in relation to the addition of Gaussian noise, namely the flux minimzer cost function is the most robust.

7. Conclusion

We have shown that we can locate (and estimate the release rate of) a source of ethane dispersing in a desert environment by inverting the gas dispersion process. We have used a random search algorithm in conjunction with simulated annealing to evaluate the performance of three cost functions. We have shown that despite the addition of Gaussian noise or concentration offsets to the data our algorithm continues to locate the gas source with each of our three cost functions. For our application the cost function minimising the total source emissions is the most robust to the experimentally relevant effects of increased noise or concentration offsets in the measured data.

References

Arridge, S.R., 1999. Optical tomography in medical imaging. Inverse Problems 15 (2).

Draxler, R.R., 1976. Determination of atmospheric diffusion parameters. Atmospheric Environment 10, 99–105.

Flescha, T., Wilsona, J.D., Harperb, L.A., Crenna, B.P., 2005. Estimating gas emissions from a farm with an inverse-dispersion technique. Atmospheric Environment 39 (27), 4863–4874.

Gibson, G., Monk, S.D., Padgett, M.J., 2002. A field-portable, laser-diode spectrometer for the ultra-sensitive detection of hydrocarbon gases. Journal of Modern Optics 49 (5–6), 769–776.

Gifford, F.A., 1976. Lectures on Air Pollution and Environmental Impact Analyses. American Meterological Society.

Hirst, B., Gibson, G., Gillespie, S., Archibald, I., Podlaha, O., Skeldon, K.D., Courtial, J., Monk, S., Padgett, M., 2004. Oil and gas prospecting by ultra-sensitive optical gas detection with inverse gas dispersion modelling. Geophysical Research Letters 31.

Horvitz, L., 1985. Geochemical-exploration for petroleum. Science 229, 821–827.

Lehning, M., Chang, D.P.Y., Shonnard, D.R., Bell, R.L., 1994. An inversion algorithm for determining area-source emissions from downwind concentration measurements. Journal of Air and Waste Management Association 44 (10), 1204–1213.

MacDonald, I.R., Reilly, Jr., J.R., Best, S.E., Venkataramaiah, R., Guinasso, Jr., N.L., Amos, J., 1996. A remote sensing inventory of active oil seeps and chemosynthetic communities in the northern Gulf of Mexico. In: Schumacher, D., Abrams, M.A. (Eds.), Hydrocarbon Migration and its near-surface expression, AAPG Mem., vol.66. pp. 27–37.

Martin, B.A., Cawley, S.J., 1991. Onshore and offshore petroleum seepage: contrasting a conventional study in papua new guinea and airborne laser fluorescing over the arafura sea. The APEA Journal 31, 333–353.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2002. Numerical Recipes in C++. Cambridge University Press, Cambridge.

Rudolph, J., 1995. The tropospheric distribution and budget of ethane. Journal of Geophysical Research 100 (D6), 11369–11382.

Schumacher, D., 1999. Exploring for Oil and Gas Traps, American Association of Petroleum Geologists.

Seldowitz, M.A., Allebach, J.P., Sweeney, D.W., 1987. Synthesis of digital holograms by direct binary search. Applied Optics 26 (14), 2788–2798.