

**An evolution strategy
to estimate emission
source distributions**

P. O'Brien et al.

An evolution strategy to estimate emission source distributions on a regional scale from atmospheric observations

P. O'Brien¹, D. Corcoran², and D. Lowry³

¹Martin Ryan Marine Science Institute, National University of Ireland Galway, Ireland

²Physics Department, University of Limerick, Ireland

³Department of Geology, Royal Holloway University of London, UK

Received: 20 December 2002 – Accepted: 4 March 2003 – Published: 13 March 2003

Correspondence to: D. Corcoran (david.corcoran@ul.ie)

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Print Version

Interactive Discussion

© EGU 2003

Abstract

In this paper we present an Evolution Strategy (ES) approach towards the estimation of the location and strength of surface emissions of trace gases based on atmospheric concentration measurements and back-trajectory analyses. The details of the ES developed are outlined. The ES is tested using artificial emission maps at different grid resolutions and the results compared to those obtained on the same problems using Singular Value Decomposition (SVD). In almost all cases, the ES improves on SVD at equivalent resolutions. In addition, a number of insights, which the ES approach brings to the problem of source location and emission strength, are discussed, particularly the limitations on the use of measurement and meteorological data in the determination of emission source distribution.

1. Introduction

In this paper, we present a novel approach towards constructing regional scale emission maps from the long term monitoring of atmospheric conditions at remote sites. A common problem in atmospheric monitoring is relating measurements made at particular observation stations to emission sources that have influenced the sampled air mass (Seinfeld and Pandis, 1998). The motivation here is to estimate emission source strength and location using field measurements so that these independent results may be compared with emissions inventories of anthropogenic and natural sources based on economic, production and other statistical data sources. Also it is desirable to be able to generate estimates for species and processes for which inventories are unavailable.

Taking precise measurements of the concentration of important trace gases and other species in the atmosphere is both labour and capital intensive. Furthermore a limited number of suitable locations exist where the site is sufficiently remote from local influences as to allow the collected data be interpreted as an indicator of regional

An evolution strategy to estimate emission source distributions

P. O'Brien et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Print Version

Interactive Discussion

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

© EGU 2003

activity. Therefore, one has data which is expensive to gather, restricted to a few locations (often just a single site), from which one wishes to extract as much information as possible. In this paper, we present methods by which measurements taken at two locations can be used to estimate emissions: the remote, baseline station at the Mace Head Atmospheric Research Station, on the west coast of Ireland and a suburban field site at Royal Holloway University near London (RHUL). Although the present discussion will be based on artificial measurement time series, the locations coincide with actual atmospheric monitoring stations, data from which will be incorporated into the models at a later date.

When the measurement data is combined with meteorological data, it becomes possible to ascribe source locations to the observed concentrations (also referred to as the excess concentrations or excess) at least qualitatively. For the Mace Head and RHUL data, contributions to the excess can come from a variety of sources particularly, but not exclusively, mainland Europe. Attempts to quantify the locations and strength of emissions sources from Mace Head, and similar data from other locations, have however met with varying degrees of success (Ashbaugh et al., 1985; Vasconcelos, 1996a, b; Stohl, 1996).

Recently, Ryall et al. (2001) have used a heuristic approach for determining emissions strengths and locations. The approach was based on the rational idea that those regions more or less often associated with higher detected gas concentrations must have higher or lower emission strengths. In this paper, we present preliminary results of an alternative heuristic approach. By casting the problem as an optimisation, we investigate the use of an evolution strategy in determining the spatial distribution and strength of emissions.

2. Back trajectory analysis

A popular meteorological tool to investigate emission is the transport model, including the trajectory-type model (Simmonds et al., 1993; Vasconcelos, 1996a, b; Deininger

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

© EGU 2003

and Saxena, 1997), and the dispersion-type model (Derwent et al., 1998a, b; Ryall et al., 1998, 2001; Zhang et al., 1999; Wang and Bentley, 2002). To determine emission source strength and location, the model may be scaled to match observational data (Ryall et al., 2001), alternatively data assimilation techniques such as the adjoint method (Houweling et al., 1999) or Kalman smoother (Zhang et al., 1999; Wang and Bentley, 2002) can be applied. In this work, we use back-trajectory analysis and implement source characterisation using an evolution strategy.

A back trajectory is an estimate of the recent movements of an air parcel, backwards in time from the moment when the air parcel was sampled at the monitoring station. A large number of trajectory models exist, each with their own assumptions and limitations (Kud at al., 1985; Merrill at al., 1986; Draxler, 1987; Deininger and Saxena, 1997; Stohl, 1998), however all models output essentially the same type of data, which is the estimated geographical location of the air mass at preceding time intervals. Often an estimate of the height of the air parcel above the surface at the various locations visited is also part of the trajectory model output. The model results used in this paper are from the Irish Meteorological Service Global Trajectory Model (GTM) using ECMWF analyses, details of which are given in McGrath (1989). An example of the back-trajectory data is presented in Table 1.

The injection of gas species into an air parcel depends on the parcel height and on the mixing height at the time it passed over a location. If the air parcel is above the mixing height, it is assumed ground source emissions do not influence concentration. Otherwise, it is assumed that the ground emissions are mixed uniformly within the mixing height. A constant mixing height of 1000 m was used in this study and the height of the air parcel is estimated using the pressure difference between the air parcel and the surface. A consequence of these simplifying assumptions is that the solution of “real-world” problems cannot be attempted. Instead the intention is to establish a “proof of concept” where a methodology for this notably complex problem is first developed and tested on artificial scenarios, but will in the future be modified to incorporate more sophisticated models of mixing height (Marik et al., 1995; Biraud et al., 2000), or actual

mixing height measurements.

3. Rationale

A reasonable proposition is to combine the measurement and trajectory data in order to generate quantitative estimates of sources and their location. Following Seinfeld and Pandis (1998) the excess concentration of a species within an air mass can be considered to be a result of recent influences, that is the recent emission into the air mass as it passed over source locations on its trajectory. The excess of a particular species C (as measured in ppb) arising from a trajectory j would then be the sum of emissions e (where a constant mixing height is assumed), from the geographical locations on this trajectory ϕ_t (a parametric curve sampled at discrete intervals of time t):

$$C_j = \sum_t e_j(\phi_t). \quad (1)$$

By identifying those sites, i , visited in a set of trajectories, and assuming that the emission from each location is independent of the trajectory used (also equivalent to time independence on the timescale of the trajectory set), this equation can be rewritten as

$$C_j = \sum_i n_{i,j} e_i, \quad (2)$$

where $n_{i,j}$ is a weighting factor which incorporates per trajectory the frequency with which a location has been visited and the dependence on mixing height during these visits. The weighting factor can be made as detailed as one feels necessary and can include descriptions of other physical and chemical processes influencing the injection and removal of the species into the air mass. With a suitable number of observations it should be possible to solve the resulting set of simultaneous equations for e_i using

**An evolution strategy
to estimate emission
source distributions**

P. O'Brien et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Print Version

Interactive Discussion

standard approaches to matrix inversion (see e.g. Veltkamp et al., 1995). Unfortunately, the resultant matrix is sparse, can be under or over-determined, depending on the resolution of the location sampling, and, susceptible to error and model assumptions.

4. Evolution strategy-cost function

An evolution strategy (ES) is an artificially intelligent global optimisation technique based loosely on the theory of evolution, i.e. survival of the fittest. The general operation of this technique is as follows. The vector $\mathbf{X}_{\text{opt}} = (x_1 \dots x_n)$ at which the function $f(X)$ is a minimum/maximum is to be determined. In general $f(X)$ is referred to as a cost function if the objective is minimisation and a fitness function if maximisation (here minimisation will be considered). For a vector space of many dimensions with a complicated surface geometry, the problem will be intractable analytically. The approach of an ES is to create a population of random solution vectors \mathbf{X}_j referred to in ES terminology as chromosomes. The elements of the vectors are referred to as genes in keeping with the genetic analogy. The process of solving for \mathbf{X}_{opt} involves continually establishing the cost of each chromosome and iteratively breeding new populations until the cost function is a minimum. Usually the breeding of new populations involves the operations of selection followed by recombination/crossover and mutation. The ES is closely related to the Genetic Algorithm (Mitchell, 1996) and the distinction between these approaches is often blurred.

In the methodology of an ES a trial solution is attempted, the so-called chromosome from above, and rated according to how close it is to the desired solution. Its survival into successive iterations of populations and its influence on the optimum solution depends on this rating. The method of rating requires a cost function which is now considered.

The region of interest is a geographical map whose graticule is the rectilinear coor-

An evolution strategy to estimate emission source distributions

P. O'Brien et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Print Version

Interactive Discussion

dinates x, y (see Fig. 1). We rewrite Eq. (2) in terms of the indices x, y ,

$$C_j = \sum_x \sum_y n_{x,y,j} e_{x,y}. \quad (3)$$

Sites not visited are also included in this formulation, but for these $n_{x,y,j} = 0$ and they are not included in the calculations (see also below). The objective is then given a set of observations C_j and associated back-trajectory data from which $n_{x,y,j}$ can be calculated, to determine the map $e_{x,y}$ which satisfies Eq. (3). This condition can be rewritten in terms of the function f_j

$$f_j(e_{x,y}) = C_j - \sum_x \sum_y n_{x,y,j} e_{x,y} = 0 \quad (4)$$

and the problem becomes one of optimisation by requiring a value of $e_{x,y}$ which minimises the cost function.

$$F(e_{x,y}) = \sum_j |f_j(e_{x,y})| \quad (5)$$

5. Evolution strategy-algorithm

We now describe in detail the operation of the ES which minimises the cost function of Eq. (5). The first step involved is the formation of the initial population of trial solutions $e_{x,y}$. This population is generated with as few a priori assumptions as possible. In general for inverse emission modelling methods, it is necessary to supply a priori information about the source activity, and perhaps more importantly the uncertainty of the emission strength. The a priori source locations are typically derived from inventories compiled from economic statistics and other literature sources. Whilst these methods progress to constrain and modify the initial information, the methodology here will allow the investigation of species, and sources, which are poorly, or falsely, reported in

An evolution strategy to estimate emission source distributions

P. O'Brien et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Print Version

Interactive Discussion

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

© EGU 2003

the literature, for example, emissions from activities that are neither socially nor economically important. The ES will eventually provide a means of estimating emissions from regions, e.g. the north Atlantic; for species, e.g. DMS; and for processes, e.g. methane seepage from ocean floor methane hydrate deposits, for which inventory are not currently available, or are far from reliable. Application of the ES, in the future, to such species as methane will provide a competitive testing ground for this method of source location and quantification against the positive results of the myriad of other approaches.

To formulate the trial solutions, use is made of two conditions. Firstly, it is impractical to assign an emission strength to a cell (where a cell is an individual element of the matrix $e_{x,y}$) which has not been visited by the set of back trajectories i.e. those for which $n_{x,y,j} = 0$. Secondly, the maximum emission possible from a given cell, denoted $e_{x,y}^T$, is never expected to exceed the minimum concentration detected at the sampling site for any of back trajectories which have passed over the cell. In general this is probably a generous over estimate of the possible emission from a given cell. This condition also places a heavy weighting on the validity of the individual trajectory with the minimum concentration. Other schemes for limiting the first guess emission from a cell can be devised, for example using the mean observed concentration associated with the given cell rather than the minimum.

Subject to these conditions the trial solution consists of a random number of visited cells. A random emission strength is assigned to each cell which is less than or equal to the minimum site concentration associated with any trajectory that has passed over that cell. In total a population of typically 40 trial solutions or chromosomes is generated.

The ES then progresses as follows. The population is rated according to the cost function of Eq. (5). Subsequently the lower performing 20% of the population are replaced. The methodology of replacing these solutions is now described. Firstly selection occurs. Two solutions X and Y are chosen from the surviving population, i.e. the top 80%, and combined to produce two “offspring” x and y. X is chosen at random from the top 20% of the population while Y is chosen from the entire surviving

An evolution strategy to estimate emission source distributions

P. O'Brien et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Print Version

Interactive Discussion

© EGU 2003

population. Those solutions of the lowest 20% are thus prevented from being involved in the replacement process.

Subsequently reproduction takes place. A weighted random binary mask typically 75% in favour of the chromosome X is generated which is used to perform a non-uniform crossover between the genes or elements of X and Y to form x and y. In this, 75% of the gene material of x comes from X, the remainder from Y and vice versa for y. The gene information is so far unchanged. Mutated offspring are also generated. To do this a random number of genes in both x and y are replaced (typically 10%) to produce x^m and y^m . For those cells selected to be mutated, the algorithm employed is the following

$$e_{x,y}^m = \left(\frac{\text{rand}(0 \rightarrow \pm 1) * e_{x,y}^T}{M} \right) + e_{x,y}$$

$$e_{x,y}^m = \begin{cases} e_{x,y}^m & 0 < e_{x,y}^m \leq e_{x,y}^T \\ e_{x,y}^T & e_{x,y}^m > e_{x,y}^T \\ \frac{\text{rand}(0 \rightarrow 1) * e_{x,y}^T}{M} & e_{x,y}^m \leq 0 \end{cases} \quad (6)$$

with $M = \sqrt{N_{\text{gen}}}$ for what is termed restrained mutation and $M = 1$ for unrestrained mutation. Here N_{gen} is the number of new populations which have been generated since the ES commenced. For restrained mutation the mutation rate decreases as the ES progresses.

Subsequent to mutation, the cost of all new chromosomes are evaluated and a new member of the population is taken from either x or x^m depending on which has the lowest cost. The same is true of y and y^m . The process of selection and reproduction is then repeated until the lower 20% of the population is replaced. The population is once again ordered according to minimum cost and the entire process repeated until a maximum number of iterations/generations (taken here to mean one complete processing of the population) is exceeded or the cost-value of the function falls below a desired tolerance, and an optimum solution is determined.

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

© EGU 2003

It is a matter of personal judgement what proportion of the population is discarded per generation. If a large proportion of the population is discarded, there may be limited diversity in the remaining population. In effect the gene pool is reduced sharply. A stable population of similar solutions is thus attained in a relatively small number of generations before convergence on an optimum solution is obtained. Discarding a low proportion of the population, in contrast, guarantees greater diversity in the population, at least in the earlier generations, allowing poorer solutions to progress into later generations, and allowing the possibility of their gene material to contribute to later offspring. As a result it may take a greater number of generations to tend towards an optimum solution, but that solution may be of lower cost.

6. Case studies

To investigate the ES approach for the estimation of emission source distributions a “reverse engineering” strategy was adopted (Doyle et al., 1999). Artificial emission maps were generated, then using a catalogue of back-trajectory analyses from 1996 to 2000 (McGrath, 1989) for Mace Head and RHUL (Lowry et al., 2001), an artificial time series of concentration measurements at the sample sites was constructed. In principle given the artificial series and the associated back-trajectory data, the ES approach should be able to re-generate the corresponding emission map. For comparison, Singular Value Decomposition (SVD) was also applied to this synthetic data. SVD (Press et al., 1994) is a well understood method of solving simultaneous equations which ignores inconclusive or ambiguous data in favour of generating a more robust and realistic solution.

Three test geometries were considered (see Fig. 2). For the area and distributed source test geometries the grid cell size is $200 \times 200 \text{ km}^2$, for the enclosing source the grid cell size is $100 \times 100 \text{ km}^2$, and in all cases the emission rate per source cell is 100 ppb/hr (recall a constant mixing height of 1000 m is assumed). The distributed source problem consists of a 7×5 array of isolated sources with a horizontal spacing of 1000 km for column elements 2 to 5, a spacing of 1800 km for column elements 1

to 2 and a vertical spacing of 600 km. The area source problem consists of a 5×3 array of sources located to the northwest of Ireland. Finally the enclosing source has a hexagonal outline consisting of 60 connected sources.

The ES and SVD were applied to the distributed and area geometries at 3 grid resolutions, with cell sizes of $100 \times 100 \text{ km}^2$, $200 \times 200 \text{ km}^2$, and $400 \times 400 \text{ km}^2$. In Fig. 3, the number of visitations by the back-trajectories per cell is shown. As the resolution of the grid increases, on average the number of visitations per cell decreases. The benefit of improved resolution might therefore be expected to be at the cost of less information per cell upon which to base any estimation of emission.

The general parameters used for the ES and SVD are outlined in Table 2. As mentioned earlier two methods of mutation are available and to test this, adopting a grid size comparable to the test problem resolution, both methods were implemented on the area source problem. The convergence data for this test is shown in Fig. 4 which clearly demonstrates the significantly faster convergence character of unrestrained mutation. Notice also the characteristic exponential like convergence of the ES approach a recognised advantage of such techniques. Given the improved performance, unrestrained mutation was adopted for subsequent ES applications.

In Figs. 5 and 6 results of the ES (after 300 iterations) and SVD methods are shown for the distributed and area test geometries at each of the differing resolutions. Inspection of the data clearly demonstrates that in almost all cases the ES improves over the SVD results when compared at equivalent resolutions.

For both the artificial area and distributed source problems, the best ES solutions are for the grid with element size $200 \times 200 \text{ km}^2$, the resolution at which the test problems were generated. In the case of the area solution a mean area emission rate of 99.9 ppb/hr ($\sigma = 1.2$) is obtained.

The ES does not generate an adequate solution for the distributed point source array problem at the lowest resolution and misses the sources entirely. The cause of this effect is the initialisation scheme. During the initialisation of the ES, the maximum emission value for a given cell is constrained to never exceed the minimum observed

An evolution strategy to estimate emission source distributions

P. O'Brien et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Print Version

Interactive Discussion

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

© EGU 2003

concentration associated with any trajectory which has passed through the cell. The validity of this approach however depends on the grid resolution. As seen from the visitation map (Fig. 3) as the grid cell size becomes larger an increasing number of trajectories will pass through a given cell on average. Consider now a source in a cell with physical size smaller than the grid resolution. The larger the trajectory number that pass through the cell the greater the likelihood that one or more of these trajectories will not intersect the source within it. The cell strength may then be constrained by emissions from a trajectory external to the source, which is incorrect. The minimisation constraint will only be valid provided the grid cell size is smaller than the source size. The same effect is responsible for the diminished extent of the area distribution solution at low resolution. It is worth pointing out that this constraint can be relaxed or even removed and an improved solution likely obtained. The trade-off would be an increase in the number of iterations required for convergence.

For the distributed source array at the lowest resolution ($400 \times 400 \text{ km}^2$) SVD does not present a realistic solution, but at medium resolution the sources are visible. The area emission source is only clearly identifiable at a medium resolution where the mean area source emission rate is 99.3 ppb/hr ($\sigma = 0.9$) compared with the correct solution of 100 ppb/hr . Importantly, in all SVD solutions there is large area coverage of spurious emission which is both positive and unphysically negative. Without prior knowledge of the distribution it would be impossible to distinguish the true distribution of sources from artefact sources.

At the highest resolution $100 \times 100 \text{ km}^2$ for both the distributed and area source problems the correct solution is matched by the ES, though a variation in source solution emission is evident. The variation is most likely caused by the increasing number of unknowns at higher resolution and the limit imposed on the maximum number of ES iterations. In Fig. 7 the convergence trend for the high resolution distributed source is presented, and reveals continuing convergence beyond the 300 iteration limit. In stark contrast SVD fails completely at this resolution and provides no result. Even though the information available to the ES approach is limited, a strength of the technique is that

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

© EGU 2003

it manages to use this information efficiently to determine a solution approximation.

Scattergram plots at the benchmark 300 iterations for the ES solutions of the area source problem are presented in Fig. 8. The plots display the correlation between site concentrations for each back trajectory, calculated from the optimum solutions, versus the actual site concentrations generated from the corresponding artificial maps. For an exact solution a regression coefficient of unity would be obtained. From the plots it is possible to see the trend, already apparent in the solution maps above, is repeated with the best solutions obtained for the medium and high resolution grids.

One potential difficulty with the ES is the presence, particularly to the northwest, of false sources being located at the extremities of the map (see Fig. 5). This behaviour is a recurring feature, though not exclusively, of the ES technique. In Fig. 6, in addition to a background of spurious emission, similar false sources can be identified for the SVD solutions, particularly to the east. In both cases, without prior knowledge of the distribution, it could be impossible to distinguish the true sources from artefact sources. In contrast to SVD however, the spurious emission appears to be confined in the ES solutions only to specific areas and might therefore be identified by other means. For example comparison of the cell visitation map with the solution map shows clearly that the artefacts are located within a region of low trajectory visitation.

It is easy to understand why errors at the extremes of the ES solutions happen; the ES does not know when an injection occurs and therefore can assign emissions to all locations along a trajectory if it wishes. If a cell is visited by only one or relatively few air masses, the cost function which is cumulative, is less sensitive to these cells than it is to cells which are visited on many occasions. Therefore, it is efficient for the ES to assign the relatively large emission to a single, poorly sampled cell. Under the present formulation of the ES, there is no forcing or weighting factor which informs the ES that it is “wrong” to do so. Without greater a priori input, the ES cannot learn of the mistake.

In Fig. 9 the results obtained from the ES at a resolution of $100 \times 100 \text{ km}^2$, for the enclosing source geometry, is shown. The ES has only limited success, with this geometry. Although the inner limits of the source region is clearly defined, outside the

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

© EGU 2003

circle is very poorly reconstructed. It would appear that in order for the ES to clearly delineate a source, a “clean” pathway to the sampling site must exist nearby. The ES needs to be able to contrast heavily burdened air masses with “clean” air masses with a similar trajectory in order to enclose the source. The SVD failed to converge on a solution for this test.

Overall for the ES if the emission sources fully encompass the detection sites, or when there is a lack of trajectory coverage, the ES has greater difficulty in reconstructing the “true” distribution. Greatest success is had with localised sources, within well sampled regions. It can be argued therefore that high emission source regions identified by the ES which are enclosed within well sampled lower/zero emission regions on a reconstructed map are likely to be reliable indicators of actual source locations and emission strength. In contrast those cells at the extremes of trajectories, or part of a contiguous link to the extremes of the sampled region, are not conclusive indicators of the presence of source activity within the cells involved. Note that the argument often made, that the tendency of back trajectory studies to locate sources at the limits of their range indicates the real sources lie further afield, is not valid in the test case studies.

7. Conclusions

In this paper, a novel approach has been presented to estimate the distribution of emission sources on a regional scale based on the use of evolutionary optimisation and back-trajectory analysis. An ES, that is an artificially intelligent global optimiser, was developed for this purpose.

The results of the ES have been compared to the Singular Value Decomposition matrix method of solving systems of linear equations. The methods have been applied to the geometric test problems of an area source, a distributed source and an enclosing source, where constant emission for the test source elements was chosen. To generate both the test problems and solutions, a constant mixing height was assumed for the back-trajectory analyses. The use of artificial test problems also excludes both

model and measurement errors which would be present in a real world scenario. Given these limitations, comparison of the ES to SVD is favourable at equivalent resolutions, with the ES approach being more successful in both the location and quantification of source emissions. At high resolution the SVD method fails to converge on any viable solution.

For the operation of the ES no a priori emission data was assumed. The generalised approach allows, in principle, for the determination of emission locations and strengths in cases where inventories are either unavailable or possibly incorrect. Nevertheless, if reliable inventories are available this data could be used to initialise the ES solutions. For example, a potential difficulty with the ES and also, for that matter SVD, is a tendency to locate false emissions at the extremes of the trajectories. For the ES, the cost function formulated above is insensitive to the emission values associated with infrequently visited cells. With certain species it may be possible to exclude falsely identified regions by virtue of a priori knowledge of the nature of the emission sources, and so further refine the cost function. As a specific example, HCFC's are strictly anthropogenic, and significant sources must be land based. Therefore, one might exclude ocean regions from the ES from the beginning. Nevertheless, it is more often not possible to predetermine possible source regions in this manner.

Comparing the operation of the ES at different resolutions has revealed that, as might be intuitively expected, the best solution is obtained at the resolution of the original test problem. In a real-world application the size of sources to be identified may not be known in advance. It has been shown, however, that at higher resolution, the test sources are located, though at computational expense. Ultimately higher grid resolutions will be limited by low trajectory number per cell, which has been identified as a likely cause of spurious source emission. At low resolution, it is expected that removing the minimisation constraint, imposed on the initialisation of threshold emissions, will lead to successful source identification, though its removal will also be computationally expensive. Thus regardless of resolution, it is expected that the ES will converge on a solution provided that the trajectory visitation per cell is sufficiently high. In this con-

An evolution strategy to estimate emission source distributions

P. O'Brien et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Print Version

Interactive Discussion

text, it would be worth exploring an adaptive grid approach, where the grid element size distribution might be adjusted to generate an optimum trajectory visitation number per cell.

The most exacting test piece investigated was the enclosing source, for which the ES established the inside boundary only. Such a test may be representative of any sampling site where the arriving air mass always passes over an emission source, as might be the case for a continental site, for example. SVD failed definitively for this test, and without the contrast of a proper background or clean concentration level, it is likely that any inverse modelling approach would have difficulty with source characterisation. Nevertheless, the use of a more sophisticated physical model, such as a dispersion instead of a trajectory model, might lead to better characterisation, given its more refined description of potential areas which have an influence on the sampled air mass. It should be pointed out that the use of an ES approach need not be confined to back-trajectory analyses. The problem need only be cast in terms of a cost function where the model relating emissions to observation can be as complex as required. Regardless of the model adopted, consideration will need to be given in the future to measurement and modelling errors for the solutions to real measurement data.

Finally, it is recognised that the presence of spurious sources undermines the results obtained from the ES approach. This could be an indication that the information which we seek may not be discernible from the data. Certainly the appearance of localised false sources in the SVD solutions would support this reasoning. The ES and indeed other methods may yield results which reconstruct observed time series of measurements for example, but there is no guarantee that the emissions maps constructed represent reality. The constructed map is just one which is a solution to the mathematical problem posed. In this debate the authors are of the opinion that the ES represents an interesting new approach to the problem, which although may not (as yet) yield an exact solution, presents important insight into the complexity of problems often hidden in the more analytical approaches.

Acknowledgement. The authors wish to thank Met Eireann and the ECMWF for ongoing provi-

An evolution strategy to estimate emission source distributions

P. O'Brien et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Print Version

Interactive Discussion

sions of back trajectory data to the Atmospheric Physics Group at National University of Ireland, Galway.

References

- Ashbaugh, L. L., Malm, W. C., and Sadeh, W. Z.: A residence time probability analysis of sulfur concentrations at Grand Canyon National park, *Atmos. Environ.*, 19, 1263–1270, 1985.
- Biraud, S., Ciais, P., Ramonet, M., Simmonds, P., Kazan, V., Monfray, P., and O'Doherty, S.: European greenhouse gas emissions estimated from continuous atmospheric measurements and radon 222 at Mace Head, Ireland, *J. Geophys. Res.*, 105, 1351–1366, 2000.
- Deininger, C. K. and Saxena, V. K.: A Validation of back trajectories of air masses by principal component analysis of ion concentration in cloud water, *Atmos. Environ.*, 31, 295–300, 1997.
- Derwent, R. G., Simmonds, P. G., Doherty, S., and Ryall, D. B.: The Impact of the Montreal Protocol on Halocarbon concentrations in Northern Hemisphere Baseline and European Air Masses at Mace Head, Ireland over a ten year period from 1987–1996, *Atmos. Environ.*, 32, 3689–3702, 1998a.
- Derwent, R. G., Simmonds, P. G., Doherty, S., Ciais, P., and Ryall, D. B.: European source strengths and Northern Baseline Concentrations of Radiatively Active Trace Gases at Mace Head, Ireland, *Atmos. Environ.*, 32, 3703–3715, 1998b.
- Doyle, S., Corcoran, D., and Connell, J.: Automated mirror design using an evolution strategy, *Optical Engineering* 38, 323–333, 1999.
- Draxler, R. R.: Sensitivity of a Trajectory Model to the Spatial and Temporal Resolution of the Meteorological Data during CAPTEX, *J. of Climate & Applied Meteorology*, 26, 1577–1588, 1987.
- Houweling, S., Kaminski, T., Dentener, F., Lelieveld, J., and Heimann, M.: Inverse modelling of methane sources and sinks using the adjoint of a global transport model, *J. Geophys. Res.*, 104, 26137–26160, 1999.
- Kud, Y. H., Skumanich, M., Haagenson, P. L., and Chang, J. S.: The Accuracy of Trajectory Models as Revealed by the Observing System Simulation Experiments, *Monthly Weather Review*, 113, 1852–1876, 1985.
- Lowry, D., Holmes, C. W., Rata, N. D., O'Brien, P., and Nisbet, E. G.: London methane emis-

An evolution strategy to estimate emission source distributions

P. O'Brien et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Print Version

Interactive Discussion

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

© EGU 2003

sions: Use of diurnal changes in concentration and $d^{13}C$ to identify urban sources and verify inventories, *J. Geophys. Res. Atmos.*, 106, 7427–7448, 2001.

Marik, F. J., Fischer, K. W., McDonald T. D., and Samson, P. J.: Comparison of Methods for Estimating Mixing Height Used during the 1992 Atlanta Field Intensive, *J. of Applied Meteorology*, 34, 1802–1814, 1995.

McGrath, R.: Global Trajectory Model, Internal Memorandum Met Eireann, 1989.

Merril, J. T., Bleck, R., and Boudra, D.: Techniques of Lagrangian Trajectory Analysis in Isentropic Coordinates, *Monthly Weather Review*, 114, 571–581, 1986.

Mitchell, M.: An introduction to Genetic Algorithms, MIT Press, Cambridge Massachusetts, 1996.

Press, W. H., Teukolsky, S. A., Vettering, W. T., and Flannery, B. P.: Numerical Recipes in C, Cambridge University Press, New York, 1994.

Ryall, D. B., Maryon, R. H., Derwent, R. G., and Simmonds, P. G.: Modelling long range transport of CFCs to Mace Head, Ireland, *Quarterly Journal of the Royal Meteorological Society*, 124, 417–446, 1998.

Ryall, D. B., Derwent, R. G., Manning, A. J., Simmonds, P. G., and O'Doherty, S.: Estimating source regions of European emissions of trace gases from observations at Mace Head, *Atmos. Environ.*, 35, 2507–2523, 2001.

Seinfeld, J. H. and Pandis, S. N.: Atmospheric Chemistry and Physics, Wiley Interscience, New York, 1998.

Simmonds, P. G., Cunnold, D. M., Dollard, G. J., Davies, T. J., McCulloch, A., and Derwent, R. G.: Evidence of the phase out of CFC use in Europe over the period 1987–1990, *Atmos. Environ.*, 27, 1397–1407, 1993.

Stohl, A.: Trajectory statistics – A new method to establish source-receptor relationships of air pollutants and its application to the transport of particulate sulfate in Europe, *Atmos. Environ.*, 30, 579–587, 1996.

Stohl, A.: Computation, accuracy and applications of trajectories – A review and bibliography, *Atmos. Environ.*, 32, 6, 947–966, 1998.

Vasconcelos, L. A. de P.: A tracer calibration of back trajectory analysis at the Grand Canyon, *J. Geophys. Res.*, 101, 19329–19336, 1996a.

Vasconcelos, L.A. de P.: Spatial resolution of a transport technique, *J. Geophys. Res.*, 101, 19337–19342, 1996b.

Veltkamp, A. C., Eisma, R., Vermuelen, A. T., Kieskamp, W. M., Van Den Bulk, W. C. M.,

Beemsterboer, B., Zwaagstra, O., Mols, J. J., Hensen, A., and Wyers, G. P.: Validation of Methane Source Strengths. The Netherlands Energy Research Foundation (ECN) Report ECN-C95-034, 1995.

5 Wang, Y.-P. and Bentley, S. T.: Development of a spatially explicit inventory of methane emissions from Australia and its verification using atmospheric concentration data, Atmos. Environ., 36, 4965–4975, 2002.

Zhang, X. F., Heemink, A. W., Janssen, L. H. J. M., Janssen, P. H. M., and Sauter, F. J.: A computationally efficient Kalman smoother for the evaluation of the CH₄ budget in Europe, Applied Mathematical Modelling, 23, 109–129, 1999.

10

**An evolution strategy
to estimate emission
source distributions**

P. O'Brien et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Print Version

Interactive Discussion

Table 1. Sample back trajectory data

Trajectory number 1							
Start: Lat 53.3° N				Pressure level 1000 hpa			
Lon 0.0° W				Time 12:00 UTC on 12.12.0			
End: Time 12:00 UTC on 8.12.0							
Forecast fields from				Time 18:00 UTC on 12.12.0			
Hours	Lat	Lon	Level	U-wind	V-wind	W-wind	Ps
0	53.3° N	9.9° W	1000	-1.1	-5.3	-0.06	981
1	53.47° N	9.84° W	980.4	-5.5	0.9	-0.08	981.9
2	53.44° N	9.6° W	981.1	-3.5	1.7	-0.07	982.5
3	53.38° N	9.46° W	981.9	-1.8	2.4	-0.06	983.4
4	53.3° N	9.4° W	983	-0.2	3	-0.06	984.7
5	53.2° N	9.43° W	985	1.3	3.2	-0.05	986.4
6	53.09° N	9.54° W	986.8	2.9	3.6	-0.05	988.5
7	52.96° N	9.7° W	988.7	3.1	4.5	-0.06	989.8
8	52.8° N	9.87° W	989.6	3.9	6.1	-0.08	991.4
9	52.59° N	10.09° W	992.4	4.5	6.9	-0.08	993.5
10	52.35° N	10.35° W	993.6	5.9	8.7	-0.09	995.7
11	52.07° N	10.68° W	996.6	6.7	9.1	-0.08	997.7

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

An evolution strategy to estimate emission source distributions

P. O'Brien et al.

Table 2. Parameters used for the ES and SVD

Analysis ES	Population	Generations	Mutation	Crossover	Replacement	X/Y
Enclosing source	100	300	Unrestrained	0.25	0.2	0.2/0.8
Area Source, long run	100	1000	Unrestrained	0.25	0.2	0.2/0.8
All other analyses	40	300	Unrestrained	0.25	0.2	0.2/0.8
SVD	W_{\min} threshold (Press et al., 1994)					
All analysis	$1.0 \times 10^{-4} * W_{\max}$					

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Print Version

Interactive Discussion

© EGU 2003

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

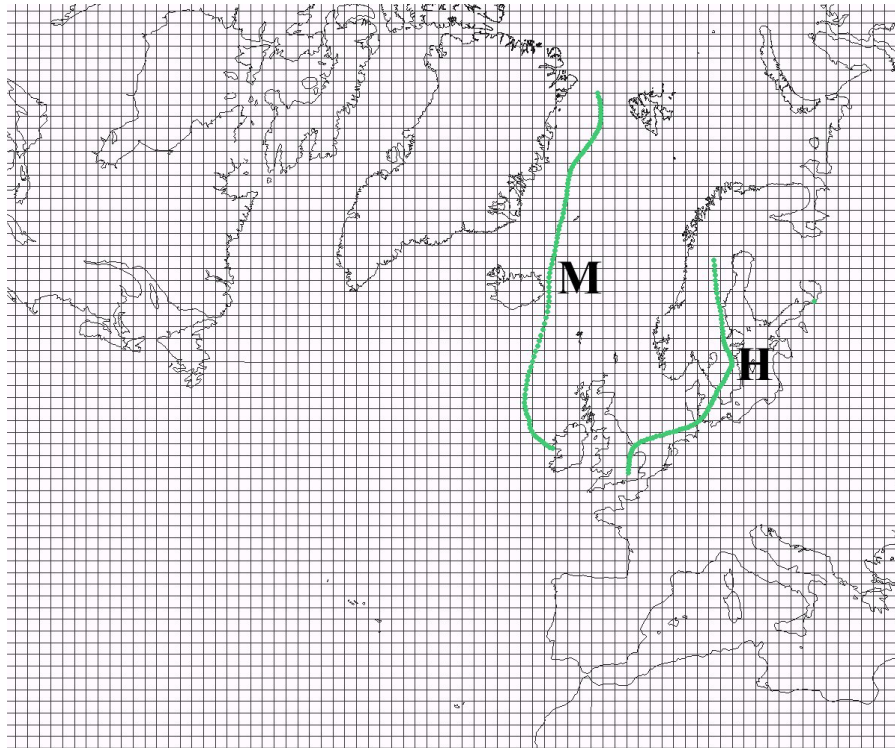


Fig. 1. The map and an example of the superimposed rectilinear coordinate graticule with cell size $100 \times 100 \text{ km}^2$ used in subsequent simulations. Sample back-trajectories arriving at Mace Head (M) and Royal Holloway, University of London (H) are shown.

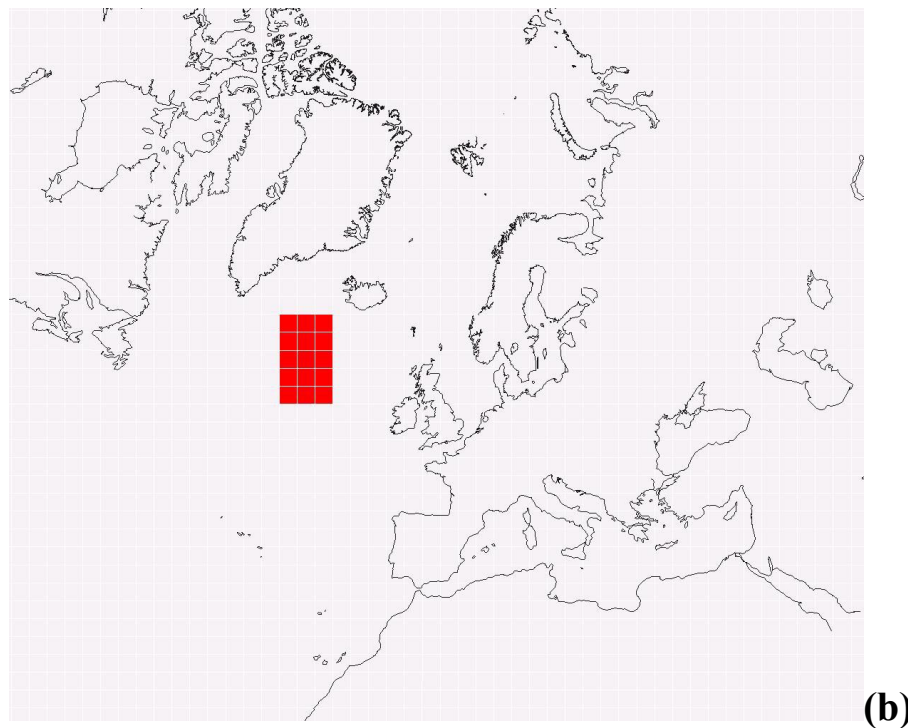
[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

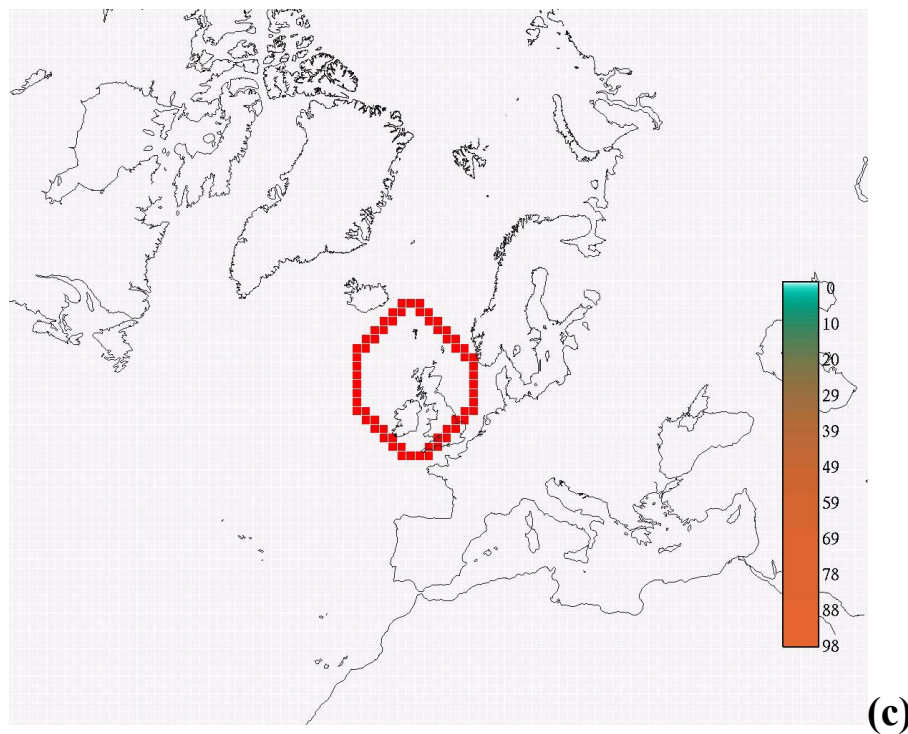
[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

Fig. 2. Test geometries for the artificial emission maps: **(a)** distributed source **(b)** area source and **(c)** enclosing source.

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

**Fig. 2.** Continued.[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

**Fig. 2.** Continued.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Print Version	
Interactive Discussion	

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

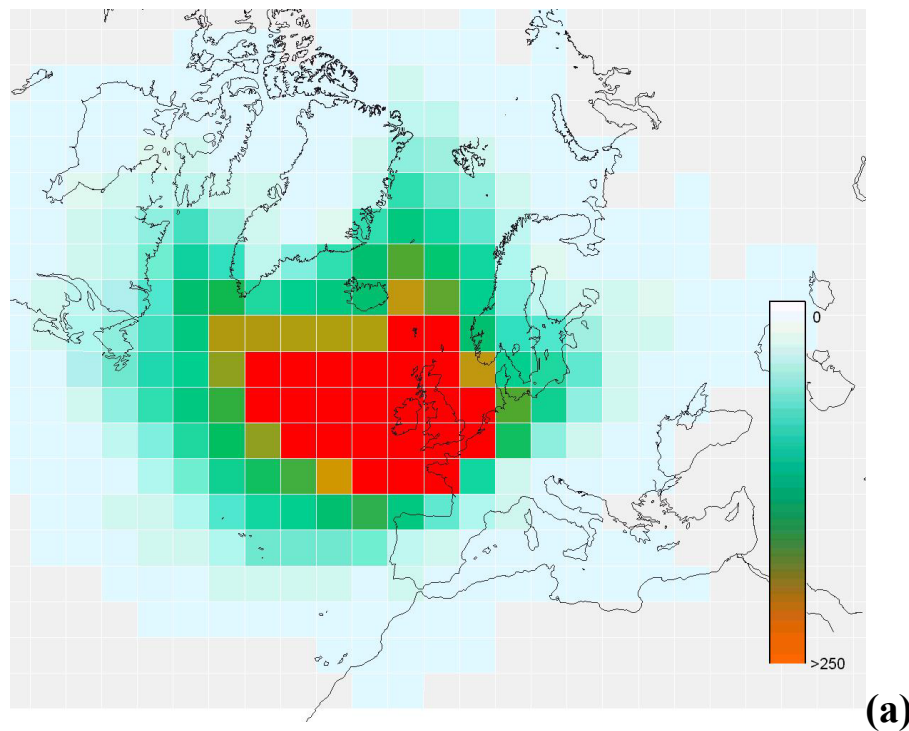
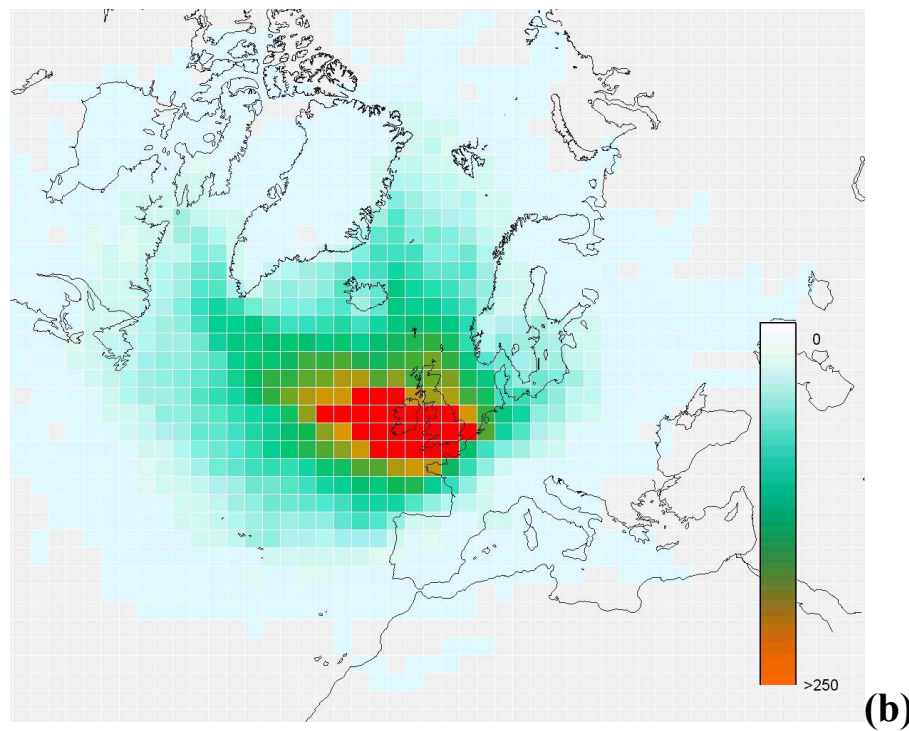


Fig. 3. A map revealing the number of cell visitations by back-trajectories at the **(a)** low $400 \times 400 \text{ km}^2$, **(b)** medium $200 \times 200 \text{ km}^2$ and **(c)** high $100 \times 100 \text{ km}^2$ resolutions.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

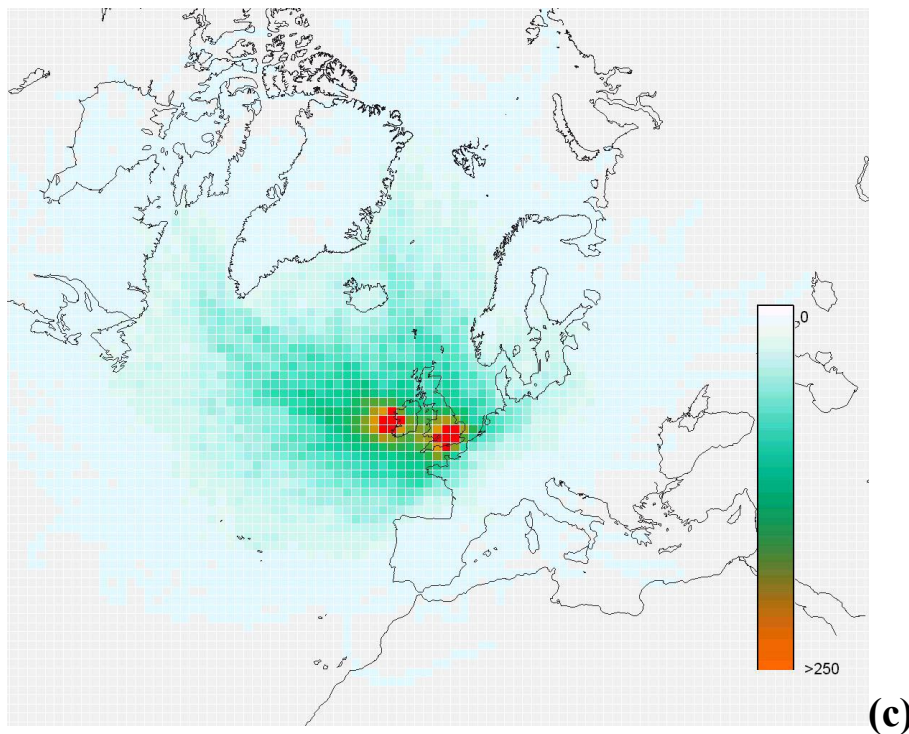
© EGU 2003

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

**Fig. 3.** Continued.[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

© EGU 2003

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

**Fig. 3.** Continued.[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

© EGU 2003

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

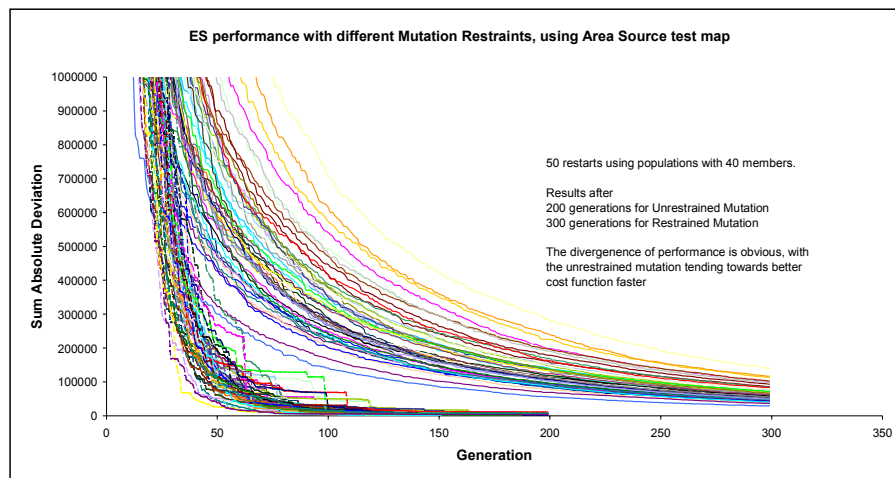


Fig. 4. The convergence trends of the ES for various start conditions for restrained and unrestrained mutation applied to the area source problem for a $200 \times 200 \text{ km}^2$ cell size.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

© EGU 2003

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

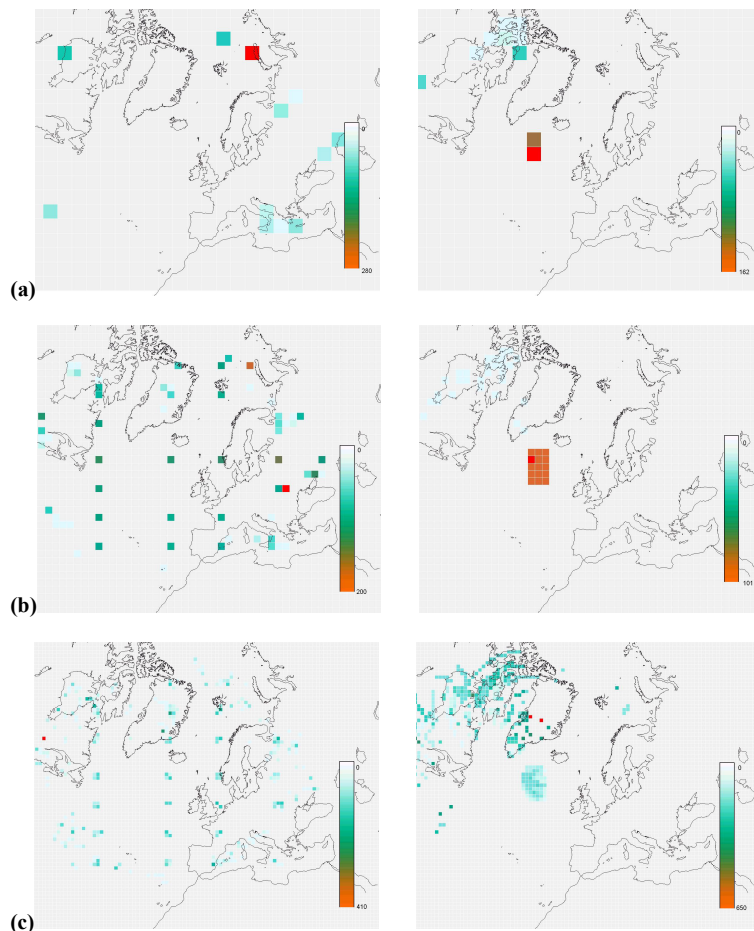


Fig. 5. Solution emission maps generated by ES for the distributed source (left) and area source (right) at **(a)** low, **(b)** medium and **(c)** high resolutions.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

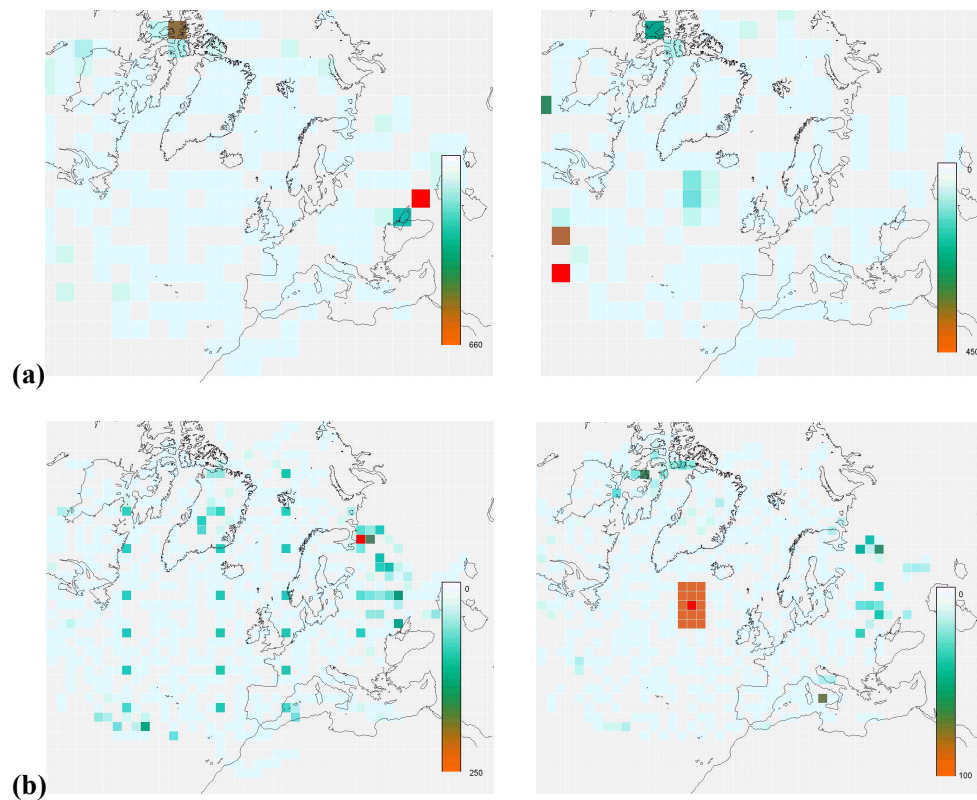


Fig. 6. Solution emission maps generated by SVD for the distributed source (left) and area source (right) problems at (a) low and (b) medium resolutions.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

An evolution strategy to estimate emission source distributions

P. O'Brien et al.

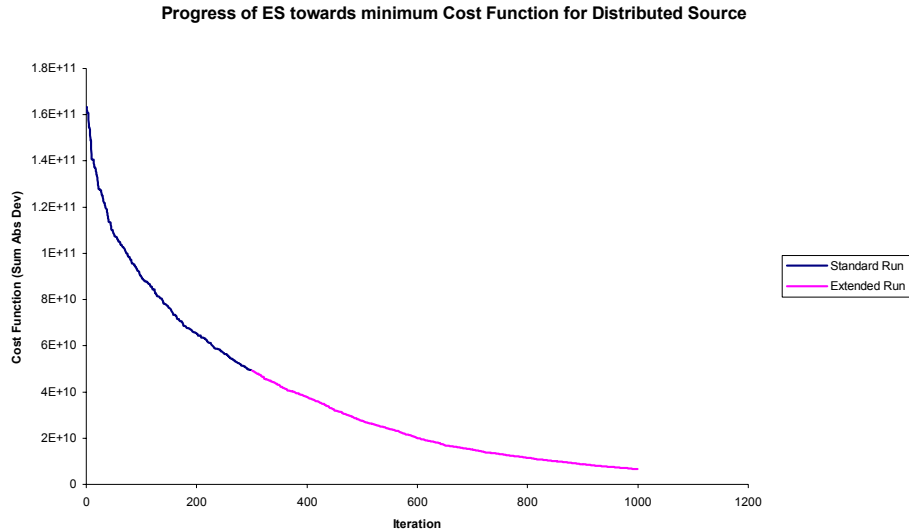


Fig. 7. The extended convergence trend for the ES solution to the distributed source problem at high resolution.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Print Version

Interactive Discussion

An evolution strategy to estimate emission source distributionsP. O'Brien et al.

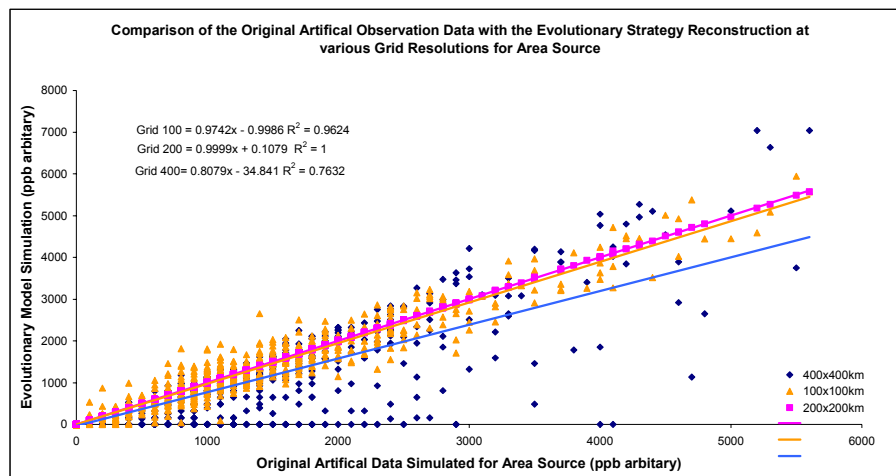


Fig. 8. Scattergram plots for the ES solutions of the area source problem at low, medium and high resolutions.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)

© EGU 2003

**An evolution strategy
to estimate emission
source distributions**P. O'Brien et al.

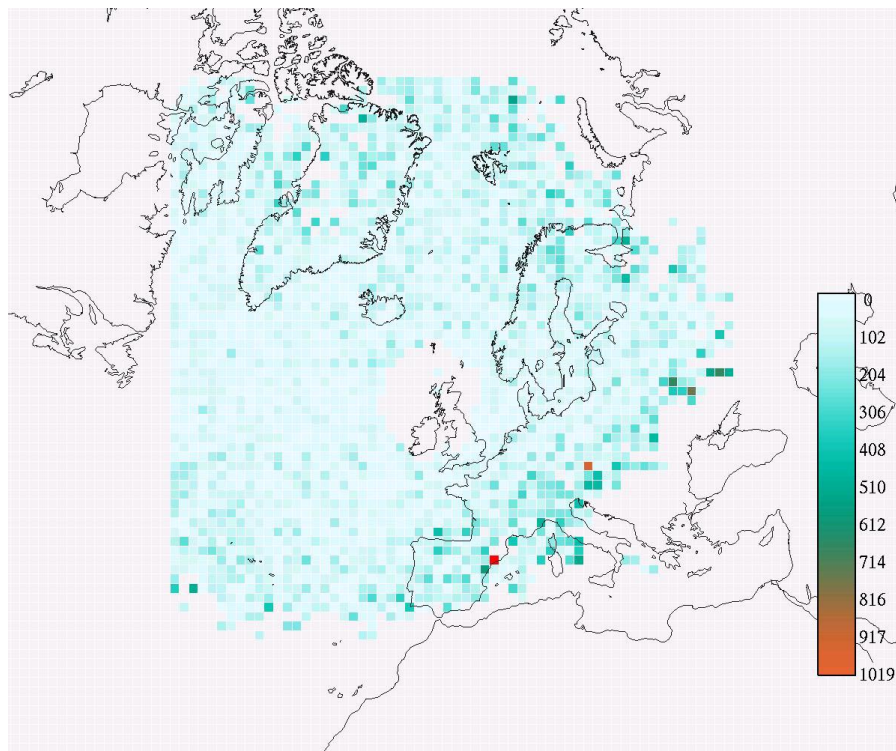


Fig. 9. Solution emission map generated by ES for the enclosing source problem.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)