

Strategies for Estimation of Gas Mass Flux Rate Between Surface and the Atmosphere

Haroldo F. de Campos Velho¹, Débora R. Roberti²,
Eduardo F.P. da Luz¹ and Fabiana F. Paes¹

¹Laboratory for Computing and Applied Mathematics (LAC),
National Institute for Space Research
(INPE) São José dos Campos (SP)

²Department of Physics, Federal University of
Santa Maria (UFSM) Santa Maria (RS)
Brazil

1. Introduction

A relevant issue nowadays is the monitoring and identification of the concentration and rate flux of the gases from the greenhouse effect. Most of these minority gases belong to important bio-geochemical cycles between the planet surface and the atmosphere. Therefore, there is an intense research agenda on this topic. Here, we are going to describe the effort for addressing this challenging. This identification problem can be formulated as an inverse problem. The problem for identifying the minority gas emission rate for the system ground-atmosphere is an important issue for the bio-geochemical cycle, and it has been intensively investigated. This inverse problem has been solved using regularized solutions (Kasibhatla, 2000), Bayes estimation (Enting, 2002; Gimson & Uliasz, 2003), and variational methods (Elbern et al., 2007) - the latter approach coming from the data assimilation studies.

The inverse solution could be computed by calculating the regularized solutions. Regularization is a general mathematical procedure for dealing with inverse problems, looking for the smoothest (*regular*) inverse solution. For this approach, the inverse problem can be formulated as an optimization problem with constraints (*a priori* information). These constraints can be added to the objective function with the help of a regularization parameter. For adding a regularization operator, the ill-posed inverse problem becomes in a well-posed one. The maximum second order entropy principle was used as a regularization operator (Ramos et al., 1999), and the regularization parameter is found by the L-curve technique. A recent approach for solving inverse problems is provided by the use of artificial neural networks (ANN). Neural networks are non-linear mappings, and it is possible to design an ANN to be one inverse operator, with robustness to deal on noisy data.

We have applied two different strategies for addressing the inverse problem: formulated as an optimization problem, and neural network. The optimization problem has been solved

by using a deterministic method (quasi-Newton), implementation used in the E04UCF routine (NAG, 1995) – see Roberti et al. (2007), and applying a stochastic scheme (Particle Swarm Optimization: PSO) (Luz et al., 2007). The PSO is a meta-heuristic based on the collaborative behaviour of biological populations. The algorithm is based on the swarm theory, proposed by Kennedy & Eberhart (1995) and is inspired in the flying pattern of birds which can be achieved by manipulating inter-individual distances to synchronize the swarm behaviour. The swarm theory is then combined with a social-cognitive theory. The multilayer perceptron neural network (MLP-NN) with 2 hidden layers (with 15 and 30 neurons) was applied to estimate the rate of surface emission of a greenhouse gas (Paes et al., 2010). The input for the ANN is the gas concentration measured on a set of points.

2. Techniques for solving inverse problems

There is an aphorism among mathematicians that the theory of partial differential equation could be divided into two branches: an easy part, and a difficult one. The former is called the *forward* (or *direct*) *problem*, and the latter is called *inverse problem*. Of course, the anecdotal statement tries to translate the pathologies found to deal with inverse problem.

For a given mathematical problem to be characterized as a *well-posed problem*, Jacques Hadamard in his studies on differential equations (Hadamard 1902, 1952) has established the following definition:

A problem is considered well-posed if it satisfies all of three conditions:

- a. *The solution exists (existence);*
- b. *The solution is unique (uniqueness);*
- c. *The solution has continuous dependence on input data (stability).*

However, inverse problems belong to the class of ill-posed problems, where we cannot guarantee existence, uniqueness, nor stability related to the input data. A mathematical theory for dealing with inverse problems is due to the Russian mathematician Andrei Nikolaevich Tikhonov during the early sixties, introducing the regularization method as a general procedure to solve such mathematical problems. The regularization method looks for the smoothest (*regular*) inverse solution, where the data model would have the best fitting related to the observation data, subjected to the constraints. The searching by the smoothest solution is an additional information (or *a priori* information), which becomes the ill-posed inverse problem in a well-posed one (Tikhonov & Arsenin, 1977).

Statistical strategies to compute inverse solution are those based on Bayesian methods. Instead to compute a set of parameter or function as before, the goal of the statistical methods is to calculate the probability density function (PDF) (Tarantola, 1987; Kaipio & Somersalo, 2005). Bayesian methodology will not be applied here.

With some initial controversy, artificial neural networks are today a well established research field in artificial intelligence (AI). The idea was to develop an artificial device able to emulate the ability of human brain. Sometimes, this view is part of program of so called strong AI. Some authors have used the expression *computational intelligence* (CI) to express a set of methodologies with inspiration on the Nature for addressing real world problems. Initially, the CI includes fuzzy logic systems, artificial neural networks, and evolutionary computation. Nowadays, other techniques are also considered to belong to the CI, such as

swarm intelligence, artificial immune systems and others. Different from other methodologies for computing inverse solutions, neural networks do not need the knowledge on forward problem. In other words, neural networks can be used as *inversion operator* without a mathematical model to describe the direct problem. Several architectures can be adopted to implement a neural network, here will be adopted the multi-layer perceptron with back-propagation algorithm for learning process. Other approaches to implement artificial neural network for inverse problem are described by Campos Velho (2011).

2.1 Regularization theory

Regularization is recognized as the first general method for solving inverse problems, where the regularization operator is a constrain in an optimization problem. Such operator is added to the objective function with the help of a Lagrangian multiplier (also named the *regularization parameter*). The regularization represents an *a priori* knowledge from the physical problem on the unknown function to be estimated. The regularization operator moves the original ill-posed problem to a well posed one. Figure 1 gives an outline of the idea behind the regularization scheme, one of most powerful techniques for computing inverse solutions.

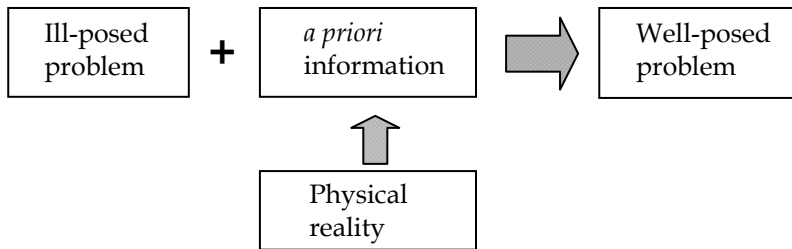


Fig. 1. Outline for regularization methods, where additional information is applied to get a well-posed problem.

The regularization procedure searches for solutions that display *global* regularity. In the mathematical formulation of the method, the inverse problem is expressed as optimization problem with constraint:

$$\min_{x \in X} \|A(x) - f^\delta\|_2^2, \text{ subject to } \Omega[x] \leq \rho \tag{1}$$

where $A(x) = f^\delta$ represents the forward problem, and $\Omega[x]$ is the regularization operator (Tikhonov & Arsenin, 1977). The problem (1) can be written as an optimization problem without constrains with the help of a Lagrange multiplier (penalty or regularization parameter):

$$\min_{x \in X} \left\{ \|A(x) - f^\delta\|_2^2 + \alpha \Omega[x] \right\} \tag{2}$$

being α the regularization parameter. The first term in the objective function (2) is the fidelity of the model with the observation data, while the second term expresses the regularity (or smoothness) required from the unknown quantity. Note that for $\alpha \rightarrow 0$, the fidelity term is overestimated; on the other hand, for $\alpha \rightarrow \infty$ all information in the mathematical model is lost. A definition of a family of regularization operator is given below.

Definition 1: A family of continuous regularization operators $R_\alpha : F \rightarrow U$ is called a regularization scheme for the inverse operation of $A(x) = f^\delta$, when

$$\lim_{\alpha \rightarrow 0} R_\alpha \{A(x)\} = x \quad (3)$$

where $x \in X$, and α is regularization parameter.

The expression (2) is a practical implementation of the Definition 1. Several regularizations operators have been derived from the pioneer works. Here, only one class of these regularization operators will be described.

2.1.1 Entropic regularization

The maximum entropy (MaxEnt) principle was firstly proposed as a general inference procedure by Jaynes (1957), on the basis of Shannon's axiomatic characterization of the amount of information (Shannon & Weaver, 1949). It has emerged at the end of the 60's as a highly successful regularization technique. Similar to others regularization techniques, MaxEnt searches for solutions with *global* regularity. Employing a suitable choice of the penalty or regularization parameter, MaxEnt regularization identifies the smoothest reconstructions which are consistent with the measured data. The MaxEnt principle has successfully been applied to a variety of fields: pattern recognition (Fleisher et al., 1990), computerized tomography (Smith et al., 1991), crystallography (de Boissieu et al., 1991), non-destructive testing (Ramos & Giovannini, 1995), magnetotelluric inversion (Campos Velho & Ramos, 1997).

For the vector of parameters x_i with nonnegative components, the discrete entropy function S of vector u is defined by

$$S(x) = - \sum_{q=1}^N s_q \log(s_q), \quad \text{with} \quad \begin{cases} x = [x_1 \quad \dots \quad x_N]^T \\ s_q = x_q / \sum_{q=1}^N x_q \end{cases} \quad (4)$$

The (nonnegative) entropy function S attains its global maximum when all s_q are the same, which corresponds to a uniform distribution with a value of $S_{\max} = \log N$, while the lowest entropy level, $S_{\min} = 0$, is attained when all elements s_q but one are set to zero.

Following the Tikhonov regularization, new entropic higher order regularization techniques have been introduced: MinEnt-1 - applied to identify the 2D electric conductivity from electro-magnetic data (Campos Velho & Ramos, 1997), MaxEnt-2 - employed to the retrieval of vertical profiles of temperature in the atmosphere from remote sensing data (Ramos et al., 1999). They represent a generalization of the standard MaxEnt regularization method,

allowing a greater flexibility to introduce information about the expected shape of the true physical model, or its derivatives, for the inverse solution.

For defining higher order entropy functions, two approaches can be derived from the maximization of the entropy of the vector of *first-* and *second-differences* x . Under the assumption $x_{\min} < x_i < x_{\max}$ ($i=1, \dots, N$), the methods of *maximum entropy of order 1* (MaxEnt-1) and *maximum entropy of order 2* (MaxEnt-2) are defined below (Muniz et al., 2000):

$$p_i = \begin{cases} x_i - x_{\min} + \zeta & \text{(zeroth order)} \\ x_{i+1} - x_i + (x_{\max} - x_{\min}) + \zeta & \text{(first order)} \\ x_{i+1} - 2x_i + x_{i-1} + 2(x_{\max} - x_{\min}) + \zeta & \text{(second order)} \end{cases} \quad (5)$$

where ζ is a small parameter ($\zeta = 10^{-10}$). The argument in the Eq. (4) is to be considered with p_i , instead of x_i .

The non-extensive formulation for the entropy proposed by Tsallis (Tsallis, 1988) can be used to unify the Tikhonov and entropic regularizations (Campos Velho et al., 2006, see also Campos Velho, 2011). The Tsallis' non-extensive entropy is expressed as

$$S_q(x) = \frac{k}{q-1} \left[1 - \sum_{i=1}^M x_i^q \right] \quad (6)$$

The non-extensive parameter q plays a central role in the Tsallis' thermostatics. The Boltzmann-Gibbs-Shannon's entropy is recovered setting $q=1$. However, for $q=2$, the maximum non-extensive entropy principle is equivalent to the standard Tikhonov regularization. Of course, for $q \neq 1$ and 2 , we have a new regularization operator. In context of inverse problem, the Boltzmann constant (k) is assumed to be equal to 1.

In order to have a complete theory, the regularization operator should be known, and it is also necessary to have a scheme to compute the regularization parameter. Several methods have been developed to identify this parameter: Morosov's discrepancy criterion, Hansen's method (L-curve method), and generalized cross validation are the most used schemes.

The Morosov's criterion is based on the difference between data of the mathematical model and observations. It should have the same magnitude as measurement errors (Morosov, 1984; see also Morosov & Stessin, 1992). Therefore, if δ is the error in the measure process, α is the root of the following equation:

$$\left\{ \|A(x) - f^\delta\|_2^2 \right\}_\alpha \approx \delta \quad (7)$$

For determining N parameter in an inverse problem, and assuming that measurement errors could be modeled by a Gaussian distribution with σ^2 variance, the discrepancy criterion for independent measures could expressed as:

$$\|A(x) - f^\delta\|_2^2 \approx N\sigma^2 \quad (7.1)$$

The discrepancy principle can also be applied to the higher entropy regularization (Muniz et al., 2000). However, if the probability density function of measurement errors follows a

distribution where the second statistical moment is not defined (Lévy or Cauchy distributions, for example) the *generalized Morosov’s discrepancy principle* can be used (Shiguemori et al., 2004).

If the statistics on the observational data is not available, the generalized cross-validation method can be applied (Bertero & Bocacci, 1998; Aster et al., 2005). Considering now a linear forward problem: $A(x) \equiv Ax$, being A a matrix, the goal of the cross-validation scheme is to minimize the *generalized cross-validation function* (Aster et al., 2005):

$$V(\alpha) = \frac{N \|A(x_\alpha) - f^\delta\|_2^2}{[Tr\{I - B(\alpha)\}]^2} \tag{8}$$

being $Tr\{C\}$ the trace of matrix C , and $B(\alpha)$ is the following matrix:

$$B(\alpha) \equiv AA^*(AA^* + \alpha I)^{-1} \tag{9}$$

where A^* is the *adjoint matrix*: $(Af, g) = (f, A^*g)$ and I is the identity matrix.

Another scheme to compute the regularization parameter is the L-curve method. The L-curve criterion is a geometrical approach suggested by Hansen (1992) (see also Bertero & Bocacci, 1998; Aster et al., 2005). The idea is to find the point of maximum curvature, on the corner of the plot: $\Omega[x_\alpha] \times \|A(x) - f^\delta\|_2^2$. In general, the plot smoothness \times fidelity shows a L-shape curve type.

2.1.2 Optimization methods

This is a central issue in science and engineering, where several methods have been developed. They can be classified as deterministic and stochastic schemes. Here, only the methods used in the application treated will be described.

2.1.2.1 Quasi-Newton optimization

The minimization of the objective function $J(x)$ given by equation (2), subjected to simple bounds on x , is solved using a first-order optimization algorithm from the E04UCF routine, NAG Fortran Library (1993). This routine is designed to minimize an arbitrary smooth function subject to constraints (simple bounds, linear and nonlinear constraints), using a sequential programming method. For the n -th iteration, the calculation proceeds as follows:

1. Solve the forward problem to x and compute the objective function $J(\alpha, x)$,
2. Compute the gradient $\nabla J(\alpha, x)$ by finite difference,
3. Compute a positive-definite approximation (quasi-Newton) to the Hessian H^n :

$$H^n = H^{n-1} + \frac{b^n (b^n)^T}{(b^n)^T u^n} - \frac{H^{n-1} u^n (u^n)^T H^{n-1}}{(u^n)^T H^{n-1} u^n}$$

where

$$b^n = x^n - x^{n-1} \quad \text{and} \quad u^n = \nabla J(\alpha, x^n) - \nabla J(\alpha, x^{n-1})$$

4. Compute the search direction d^n as a solution of the following quadratic programming subproblem:

$$\text{minimize: } [\nabla J(\alpha, x^n)]^T d^n + \frac{1}{2} (d^n)^T H^n d^n$$

$$\text{under constrain: } (x_{\min})_q - x_q^n \leq d_q \leq (x_{\max})_q - x_q^n$$

5. Set: $x^{n+1} = x^n + \beta^n d^n$, where the step length β^n minimizes $\nabla J(x^n + \beta^n d^n)$.
6. Test of convergence: Stop, if x satisfies the first-order Kuhn-Tucker conditions and $\beta^n \|d\| < \varepsilon^{1/2}(1 + \|x\|)$, where ε specifies the accuracy to which one wishes to approximate the solution of the problem. Otherwise, return to step 1.

2.1.2.2 Particle swarm optimization

The particle swarm optimization (PSO) is a heuristic search method that uses the behavior of biological collective behavior, like birds, fishes or bees to drive the artificial particles in a search pattern that equalizes a global search with a local search, in a way that an optimum, or near to optimum, result is found (Kennedy & Eberhart, 1995, 2001). There is a socio-cognitive theory for supporting the particle swarm scheme. The cultural adaptive process encloses two components: high-level (pattern formation through the individual and the ability of problem solving), and a low-level (individual behavior), with some skills for evaluating, comparing, and emulating (Kennedy & Eberhart, 2001).

The tendency to evaluate is, maybe, the behavioral feature more present in several living beings. This is intrinsic for the learning process. The comparison among individuals from a population is a behavior process used to identify patterns to find a way to improve each one quality. The emulation behavior gives the living beings the ability to learn through the observations of others actions. This behavior is not so much common in the nature. The emulation behavior is closer to the human society, but it can be used to improve the performance of the algorithms. These three concepts can be combined, even in simplified computational rules, providing a scheme for solving optimization problems.

PSO searches for the optimum in an infinite (\mathbb{R}^∞) or \mathbb{R}^N (N -dimensional space of real numbers). Indeed, computational solutions can only obtain a projection of an infinite space into \mathbb{R}^N space. The algorithm allows the particles to fly in the search space, looking for the optimization of a predetermined function. A special position x^* of a particle in the search space represents the solution of the problem. This position can be represented by a vector of real numbers $x = (x_1, x_2, \dots, x_N)$. In the PSO, the solution x^* is searched in an iterative procedure. At each time step, or iteration of the algorithm, the position of each particle x^k is updated, by the addition of a speed vector v^k , in each dimension (space direction). The iteration could be given by

$$x^k(t) = x^k(t-1) + v^k(t) \quad (10)$$

where $x^k(t-1)$ is the previous position of the particle. The velocity that is applied to the particle is function of two basic behaviors: a cognitive component, and a social component. The velocity is up dated taken into account the behaviors mentioned, as following:

$$v^k(t) = wv^k(t-1) + c_1rand_1(p^k - x^k) + c_2rand_2(p^g - x^k) \quad (11)$$

where $c_{1,2}$ are constants (real numbers) that weights the cognitive and social components, respectively, $rand_{1,2}$ are random numbers from a uniform distribution in the interval $[0,1]$, p^k is the best position achieved by the particle and p^g is the best position of the swarm, with x^k actual position of the particle k . The parameter w is used to give more stability to the algorithm (Shi & Eberhart, 1998). The selection of the w value follows a heuristic rule:

- If $w < 0.8$: the algorithm works in a local search mode, in a exploration mode;
- If $w > 1.2$: the algorithm works in a exploitation mode, where the global is more explored;
- If $0.8 < w < 1.2$: a balance point, where exploration and exploitation are well divided.

2.2 Artificial neural networks

The artificial neuron is a simplified representation of a biological neuron, where a weighted combination of the inputs is the value for the non-linear activation function (in general, a sigmoid one). In artificial neural networks (ANN), the identification of the connection weights is called the learning (or training) phase.

Several architectures can be used for the ANN. Related to the learning process, ANNs can be described into two important groups: supervised, and unsupervised NNs. For the supervised scheme, the connection weights are selected to become the output from the ANN as close as possible to the target. For example: the weights could be found by minimizing a functional of the square differences between the ANN output and the target values. Some authors have suggested the use of a Tikhonov's functional (Poggio & Girosi, 1990). The regularization suggested by Poggio & Girosi (1990) is applied on the output values x of the NN:

$$J[w] = \sum_{i=1}^N [x_i^{\text{Target}} - x_i(w)]^2 + \alpha \|Px\| \quad (12)$$

where x is the NN output, w is the connection weight matrix, and P is a linear operator. Shiguemori et al. (2004) employed three different architectures for the NNs, with focus on determination of the initial profile for the heat conduction problem. The NNs used were: multilayer perceptron (MLP), RBF-NN, and cascade correlation. The backpropagation algorithm was used for the learning process (without regularization). Figure 2 shows a cartoon to represent a MLP-NN. Cybenko (1989) has proved that the standard multilayer NN with a single hidden layer with a sigmoid activation function, for a finite number of hidden neurons, are universal approximators on a compact subset of \mathfrak{R}^N (Cybenko theorem).

The backpropagation training is a supervised learning algorithm that requires both input and output (desired) data. Such pairs permit the calculation of the error of the network as the difference between the calculated output and the desired vector: $\varepsilon^{1/2} \equiv x^{\text{target}} - x(w)$. The

weight adjustments are conducted by backpropagating such error to the network, governed by a change rule. The weights are changed by an amount proportional to the error at that unit, times the output of the unit feeding into the weight. Equation below shows the general weight correction according to the so-called delta rule (Haykin, 1999)

$$\Delta w_{ji} = \eta \frac{\partial \varepsilon}{\partial w_{ji}} x_i + \gamma [w_{ij}(n) - w_{ij}(n-1)] = \eta \delta_j x_i + \gamma [w_{ij}(n) - w_{ij}(n-1)] \quad (13)$$

where, δ_j is the local gradient, x_i is the input signal of neuron j , η is the learning rate parameter that controls the strength of change, and γ the momentum coefficient, which decides to what degree this previous adjustment is to be considered so as to prevent any sudden changes in the direction in which corrections are made. The learning rate η and momentum γ were set to 0.1 and 0.5, respectively. Nevertheless, the activation test is an important procedure, indicating the performance of a NN. The effective test is defined using an unknown vector x that did not belong to the training set. This action is called the *generalization* of the NN.

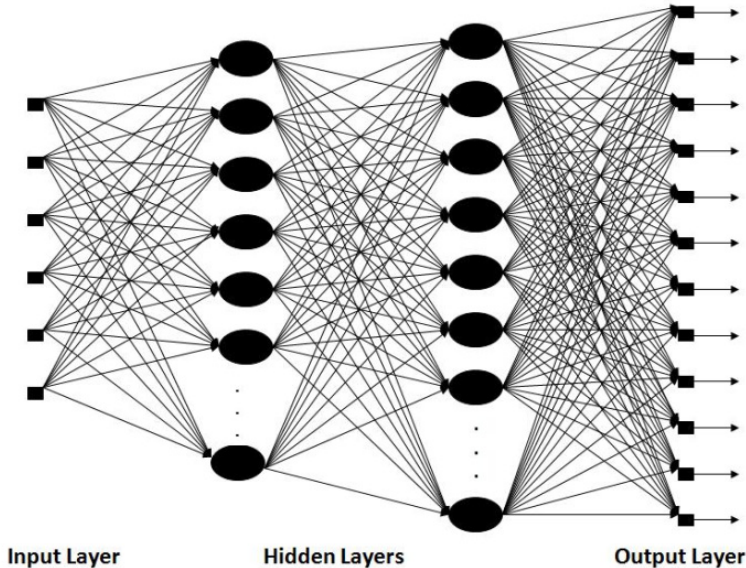


Fig. 2. Outline of fully connected MLP-NN with 2 hidden layers.

The up date for the weight matrix of neural connections is obtained with the innovation described in Equation (13):

$$w_{ij}(n) = w_{ij}(n-1) + \Delta w_{ji}$$

where the iterative process follows up to reach a maximum number of epochs (iterations) or the required precision to be caught. The delta rule can be justified if the diagonal matrix ηI (the learning parameter multiply by identity matrix) could be understood as an approximate inverse of Hessian matrix (a matrix C is defined as an approximate inverse if: $\|I - CA\| <$

1, see Conte and de Boor (1980)). Therefore, the free parameter η should be determined to be an approximation for the inverse of the Hessian matrix, producing similar results as the Newton method. However, the Newton method, or even the steepest descent scheme, is not good to deal if a shallow local minimum. The momentum term ($\gamma \neq 0$, in Eq. (13)) is added to circumvent some instabilities (Rumelhart et al., 1986) and avoid pathological behaviour with shallow error surface (Haykin, 1999). However, any other optimization method can be used instead of delta rule. Indeed, the PSO scheme described in Section 2.1.2.2 can be employed to determine the weights of the neural network (Li and Chen, 2006).

3. The Lagrangian stochastic model LAMBDA

The Lagrangian particle model LAMBDA was developed to study the transport process and pollutants diffusion, starting from the Brownian random walk modeling (Anfossi & Ferrero, 1997). In the LAMBDA code, full-uncoupled particle movements are assumed. Therefore, each particle trajectory can be described by the generalized three dimensional form of the Langevin equation for velocity (Thomson, 1987):

$$du_i = a_i(\mathbf{x}, \mathbf{u}, t)dt + b_{ij}(\mathbf{x}, \mathbf{u}, t)dW_j(t) \quad (14a)$$

$$d\mathbf{x} = (\mathbf{U} + \mathbf{u})dt \quad (14b)$$

where $i, j = 1, 2, 3$, and \mathbf{x} is the displacement vector, \mathbf{U} is the mean wind velocity vector, \mathbf{u} is the Lagrangian velocity vector, $a_i(\mathbf{x}, \mathbf{u}, t)$ is a deterministic term, and $b_{ij}(\mathbf{x}, \mathbf{u}, t)dW_j(t)$ is a stochastic term and the quantity $dW_j(t)$ is the incremental Wiener process.

The deterministic (drift) coefficient $a_i(\mathbf{x}, \mathbf{u}, t)$ is computed using a particular solution of the Fokker-Planck equation associated to the Langevin equation. The diffusion coefficient $b_{ij}(\mathbf{x}, \mathbf{u}, t)$ is obtained from the Lagrangian structure function in the inertial subrange, ($\tau_K \ll \Delta t \ll \tau_L$), where τ_K is the Kolmogorov time scale and τ_L is the Lagrangian decorrelation time scale. These parameters can be obtained employing the Taylor statistical theory on turbulence (Degrazia et al., 2000).

The drift coefficient, $a_i(\bar{\mathbf{x}}, \bar{\mathbf{u}}, t)$, for forward and backward integration is given by

$$a_i P_E = -\frac{\partial}{\partial u_j} (B_{i,j} P_E) + \varphi_i(\mathbf{x}, \mathbf{u}, t) \quad (15a)$$

with,

$$\frac{\partial \varphi_i}{\partial u_i} = -\frac{\partial P_E}{\partial t} - \frac{\partial}{\partial x_i} (u_i P_E) \quad \text{and} \quad \varphi_i \rightarrow 0 \quad \text{when} \quad \mathbf{u} \rightarrow \infty \quad (15b)$$

where $c_v = -1$ for forward integration and $c_v = 1$ for backward integration, $P_E = P(\mathbf{x}, \mathbf{u}, t)$ is the non-conditional PDF of the Eulerian velocity fluctuations, and $B_{i,j} = (1/2)b_{i,k}b_{j,k}$. Of course, for backward integration, the time considered is $t' = -t$, and velocity $\mathbf{U}' = -\mathbf{U}$, being \mathbf{U} the mean wind speed. The horizontal PDFs are considered Gaussians, and for the vertical coordinate the truncated Gram-Charlier type-C of third order is employed (Anfossi

& Ferrero, 1997). The diffusion coefficients, $b_{ij}(\mathbf{x}, \mathbf{u}, t)$, for both forward and backward integration is given by

$$b_{ij} = \delta_{ij} \left[2 \frac{\sigma_i^2}{\tau_{Li}} \right]^{1/2} \quad (16)$$

where δ_{ij} is the Kronecker delta, σ_i^2 and τ_{Li} are velocity variance at each component and the Lagrangian time scale (Degrazia et al., 2000), respectively. With the coordinates and the mass of each particle, the concentration is computed (see below).

The inverse problem here is to identify the source term $S(t)$. A source-receptor approach can be employed for reducing the computer time, instead of running the direct model, Eq. (14), for each iteration. This approach displays an explicit relation between the pollutant concentration of the i -th receptor related the j -th sources:

$$C_i = \sum_{j=1}^{N_s} M_{ij} S_j \quad (17)$$

where the matrix M_{ij} is the *transition matrix*, and each matrix entry given by

$$M_{ij} = \begin{cases} (V_{S,j}/V_{R,i})(\Delta t/N_{S,j})N_{R,i,j} & \text{- forward model;} \\ (\Delta t/N_{R,i})N_{S,i,j} & \text{- backward model.} \end{cases} \quad (18)$$

where $V_{R,i}$ and $V_{S,j}$ are the volume for the i -th receptor and j -th source, respectively; $N_{S,j}$ and $N_{R,i}$ are the number of particle realised by the j -th source and i -th sensor, respectively; $N_{R,i,j}$ and $N_{S,i,j}$ are the number of particle released by the j -th source and detected by the i -th receptor.

4. Determining gas flux between the ground and the atmosphere

For testing the procedure to estimate the emission rate procedure, it is considered the area pollutant sources placed in a box volume, where the horizontal domain and vertical height are given by: (1500 m × 1000 m) × 1000 m. There are two embedded regions R_1 and R_2 into computational domain, with following horizontal domain (600 m × 600 m) for each region, and 1 m of height, and they are realising contaminants with two different emission rates. Figure 1 shows the computational scenario in a two-dimensional projection (x, y): the six sensors are placed at 10 m height and spread horizontally with the coordinates presented in Table 1. The domain is divided into sub-domains with 200 m × 200 m × 1 m. The emission rates for each sub-domain (S_{A_k}) is as following:

$$R_1 = S_{A_2} = S_{A_3} = S_{A_4} = S_{A_7} = S_{A_8} = S_{A_9} = 10 \text{ gm}^{-3}\text{s}^{-1} \text{ (Region-1)}$$

$$R_2 = S_{A_{12}} = S_{A_{13}} = S_{A_{14}} = S_{A_{17}} = S_{A_{18}} = S_{A_{19}} = 20 \text{ gm}^{-3}\text{s}^{-1} \text{ (Region-2)}$$

$$S_{A_1} = S_{A_5} = S_{A_6} = S_{A_{10}} = S_{A_{11}} = S_{A_{15}} = S_{A_{16}} = S_{A_{20}} = S_{A_{21}} = S_{A_{22}} = S_{A_{23}} = S_{A_{24}} = S_{A_{25}} = 0$$

The dispersion problem is modelled by the LAMBDA backward-time code, where data from the Copenhagen experiment are used, for the period (12:33 h - 12:53 h, 19/October/1978) (Gryning & Lyck, 1984, 1998). It is assumed a logarithmic vertical profile for the wind field

$$U(z) = \frac{u^*}{\kappa} \ln \left(\frac{z}{z_0} \right) \quad (19)$$

being $U(z)$ the main stream, u^* the fiction velocity, κ the von Kármán constant, z the height above the ground, and $z_0=0.06$ m the roughness. The wind speed was measured at 10 m, 120 m, and 200 m. The numerical value for u^* was obtained from the best fitting with the measured wind speed - see Table 2 - and the equation (19).

Sensor-number	x (m)	y (m)
1	400	500
2	600	300
3	800	700
4	1000	500
5	1200	300
6	1400	700

Table 1. Position of sensors in physical domain.

For this experiment the wind direction had an angle $\theta=180^\circ$, and the boundary layer height is $h=1120$ m (Gryning & Lyck, 1998). The turbulence parameterization follows Degrazia et al. (2000), for computing the wind variance (σ_i^2) and the Lagrangian decorrelation time scale (T_{L_i}). These two turbulent parameters are considered constant for the whole boundary layer, and their numerical values were calculate at $z=10$ m level. This characterises a stationary and vertically homogeneous atmosphere.

The sensors dimension, where the fictitious particles have arrived, was de $0,1 \text{ m} \times 0,1 \text{ m} \times 0,1 \text{ m}$, centred in the computational cells presented in Table 1. Next, the reverse trajectories are calculated for 1000 particles per sensor. The parameters for numerical simulations were 1800 time-steps with $\Delta t = 1$ s, meaning 30 min for the whole simulation. After 10 min, the concentration was computed at each 2 min, for the remaining 20 min of simulation. The mean concentration is found using the following expression:

$$C^{Mod}(x_j, y_j) = \frac{1}{10} \sum_{n=1}^{10} \left[\sum_{i=1}^{25} S_i \frac{\Delta t}{N_{PES,j}} N_{PVF,i,j,n} \right] \quad (20)$$

where (x_j, y_j) is position the each sensor (the number of sensors is 6), $N_{PES,j}=1000$ is the number of particles arriving at each sensor. $N_{PVF,i,j,n}$ is the quantity of fictitious particles in the i -th volume source, at the n -th instant.

Time (h:min)	Speed U (ms ⁻¹)			Wind direction (°)		
	10 m	120 m	200 m	10 m	20 m	200 m
12:05	2.6	5.7	5.7	290	310	310
12:15	2.6	5.1	5.7	300	310	310
12:25	2.1	4.6	5.1	280	310	320
12:35	2.1	4.6	5.1	280	310	320
12:45	2.6	5.1	5.7	290	310	310

Table 2. Measured wind speed for the Copenhagen experiment (Gryning & Lyck, 1998).

The inverse approach is tested using synthetic measured concentration data. The synthetic observations are emulated as:

$$C^{Exp}(x_j, y_j) = C^{Mod}(x_j, y_j)[1 + \lambda\mu] \quad (21)$$

where $C^{Mod}(x, y)$ is the gas concentration computed from the forward model (14), μ is a random number associated to the Gaussian distribution with zero mean and unitary standard deviation, and λ is the level of the noise (Gaussian white noise). In our tests, $\lambda=0.05$ (Experiment-1) and $\lambda=0.1$ (experiment-2) were used. In other words, two numerical experiments are performed, using two different maximum noise levels: 5% and 10%.

The optimization method is used to find the minimum for the functional:

$$F(\alpha, \mathbf{S}) = \sum_{j=1}^6 [C_j^{Exp} - C_j^{Mod}(\mathbf{S})]^2 + \alpha \Omega(\mathbf{S}) \quad (22)$$

where the vector $\mathbf{S} = [S_{A_1}, S_{A_2}, \dots, S_{A_{25}}]^T$ represent the emission rate from 25 sub-domains, presented in Figure 3. However, some sub-domains are considered releasing nothing. Therefore, the number of the unknown parameters to be estimated by inverse method decreases to 12 cells: $\mathbf{S} = [S_{A_2}, S_{A_3}, S_{A_4}, S_{A_7}, S_{A_8}, S_{A_9}, S_{A_{12}}, S_{A_{13}}, S_{A_{14}}, S_{A_{17}}, S_{A_{18}}, S_{A_{19}}]$. Another *a priori* information is smoothness of the inverse solution (regularization operator). As mentioned before, the second order maximum entropy is applied, Eq. (13c). The use of entropic regularization for this type of inverse problem was independently proposed by Roberti (2005) and Bocquet (2005a, 2005b) – see also Roberti et al. (2007) and Davoine & Bocquet (2007).

There are several methods to compute the regularization parameter. Here, a numerical experimentation was employed to determine this parameter. Roberti (2005) had found the values $\alpha = 10^{-6}$ for Experiment-1, and $\alpha = 10^{-5}$ for Experiment-2.

The first inverse solution was obtained by applying the quasi-Newton optimization method (described in Section 2.1.1). In the Experiment-1, a global error around 2% is found for the inverse solution. For the region-1 (lower emission rate), the error in the estimation is 10%, and for region-2 the error is 3%. The biggest errors for estimation are verified for sub-domains A_9 and A_{14} , with error of 55% and 42%, respectively.

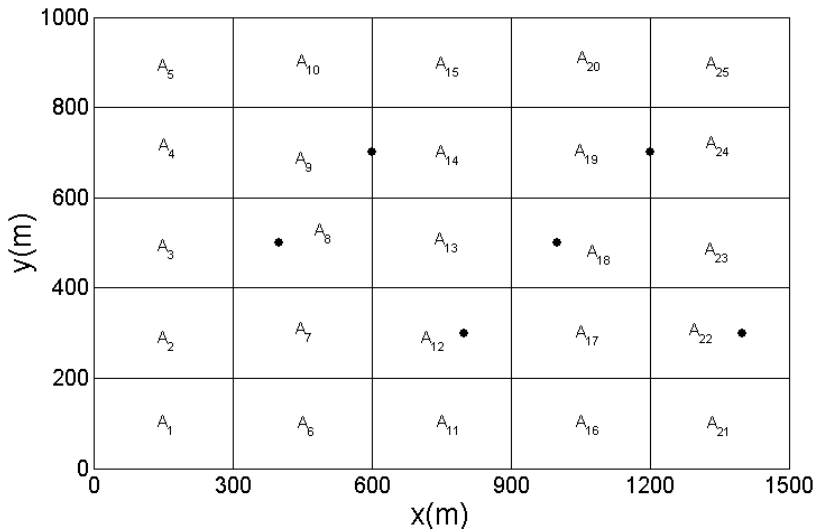


Fig. 3. Sub-domains representation, in a two-dimensional projection (x,y), with different emission rates, and black points (•) represent the sensor position at 10 m height.

In order to try some improvement for the inverse solution, the minimum value for the functional (22) is computed using the PSO scheme. This method is a stochastic technique and it deals with a population (candidate inverse solutions). Therefore, the numerical results represent the average from 25 seeds used by the PSO algorithm, working with 12 particles, with the following parameters: $w = 0.2$, $c_1 = 0.1$, and $c_2 = 0.2$. The total error obtained with the PSO was less than inverse solution computed from the quasi-Newton method. It is also important to point out that for the PSO approach no regularization operator was used ($\alpha = 0$ in Eq. (22)).

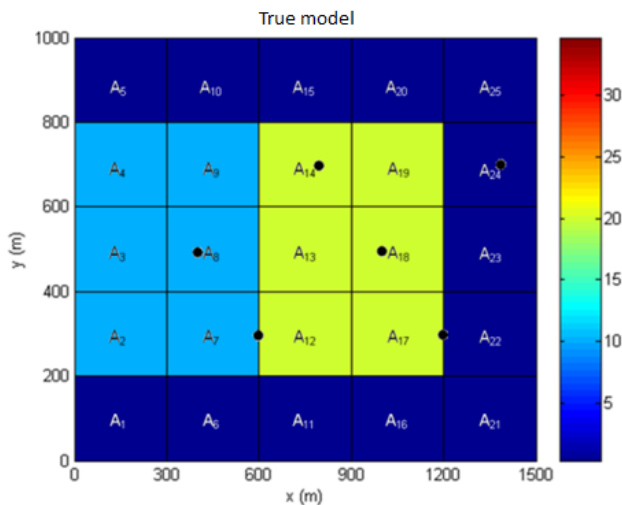
The inverse problem was also addressed by neural network. The MLP-NN was designed and trained to find the gas flux between the soil and the atmosphere. However, there are two previous steps before the application of MLP-NN. Firstly, it is necessary to determine the appropriated activation function, number of hidden layers for the NN, and the number of neurons for each layer. Secondly, the values of the connection weights (learning process) must be identified.

The hyperbolic tangent was used as activation function. Different topologies for MLP-NN were tested with 6 inputs (measured points for 6 different sensors) and 12 outputs (emission for each subdomain). Good results were achieved for two neural networks with 2 hidden layers: NN-1 with 6 and 12 (6:6:12:12), and NN-2 with 15 and 30 (6:15:30:12) neurons for the hidden layers, respectively.

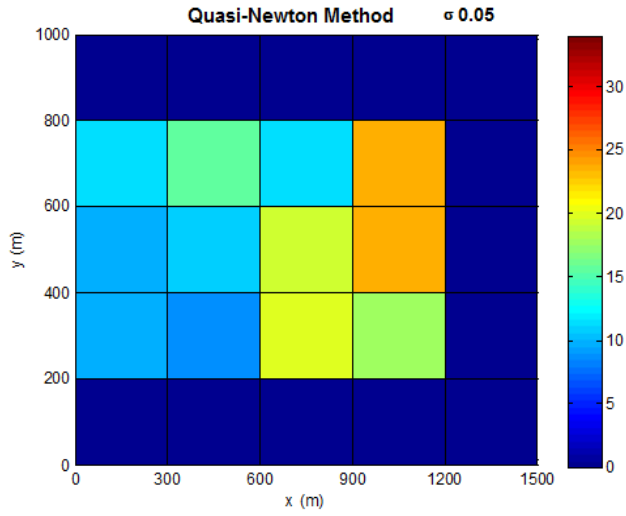
Table 3 shows the numerical values computed for the emission rate for each marked cell. Good results were obtained for all methodologies. The cells 12 and 14 had a worst estimation using optimization solution. Inversions with MLP-NN were better, in particular for the emission of the cells 12 and 14. Figures 4 and 5 show the emission estimation for the experiment-1 (5% of noise level) and experiment-2 (10% of noise level), only results obtained with NN-2 (6:15:30:12) are shown (the best results with MLP-NN).

Sub-domain	Exact	Q-N	PSO	MLP-NN-1	MLP-NN-2
A ₂	10	8.97	9.83	10.33	10.11
A ₃	10	9.97	10.40	10.11	10.11
A ₄	10	12.53	10.79	10.12	10.20
A ₇	10	7.99	10.50	10.27	10.20
A ₈	10	10.15	12.06	10.24	10.06
A ₉	10	11.56	11.28	10.21	10.17
A ₁₂	20	13.85	14.56	21.28	20.97
A ₁₃	20	22.65	22.67	21.32	21.00
A ₁₄	20	14.14	15.85	21.47	20.95
A ₁₇	20	19.99	21.56	21.47	20.99
A ₁₈	20	21.17	20.05	21.38	20.99
A ₁₉	20	24.90	21.74	20.96	20.96

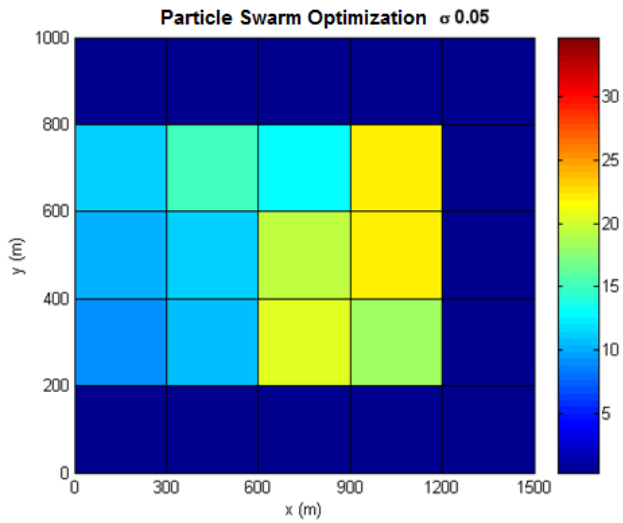
Table 3. Estimation of gas flux emission by the methodologies: quasi-Newton (Q-N), PSO, MLP-NN-1 (6:6:12:12) MLP-NN-2 (6:15:30:12).



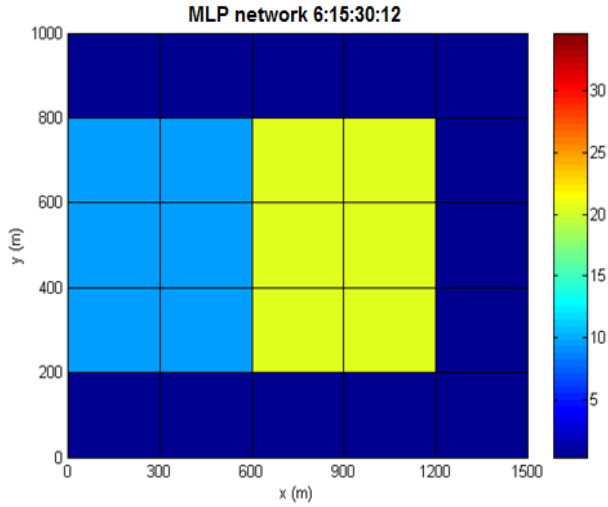
(a)



(b)

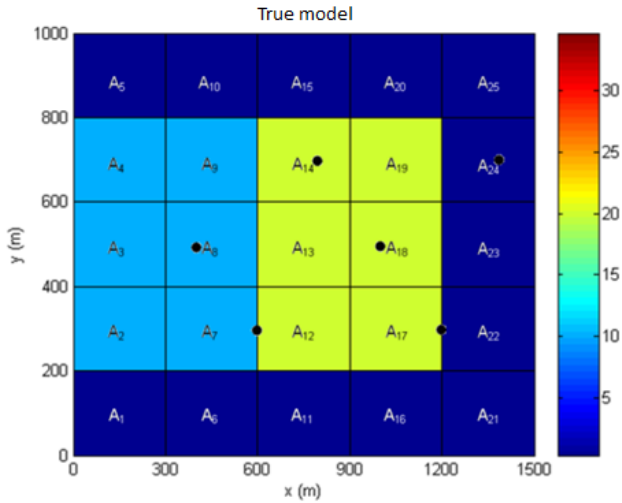


(c)

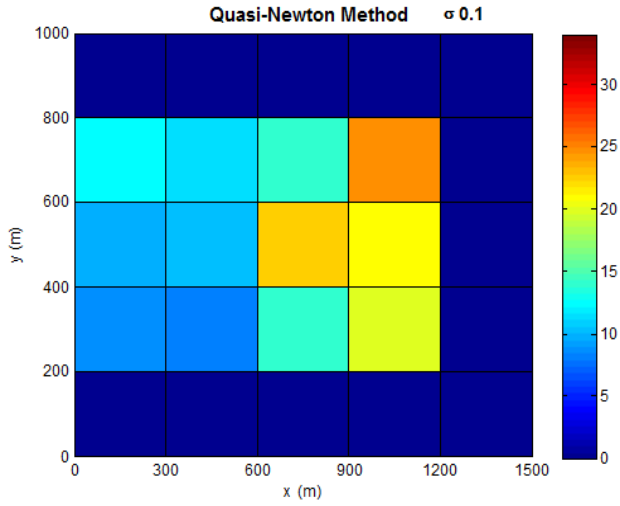


(d)

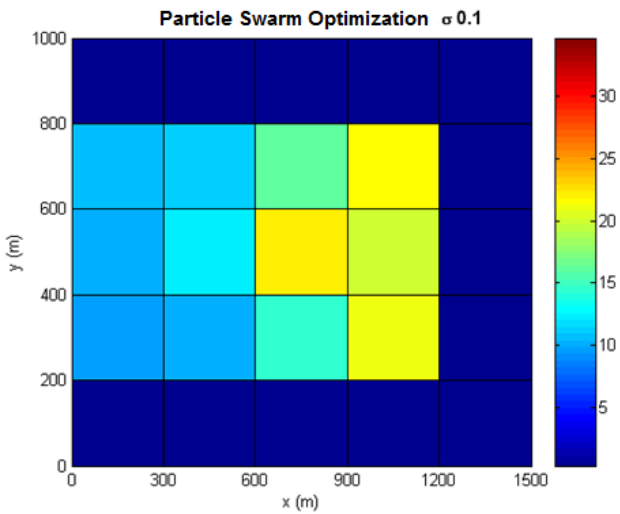
Fig. 4. Gas flux rate for Experiment-1: (a) true values, (b) quasi-Newton (Q-N), (c) PSO, and (d) MLP-NN.



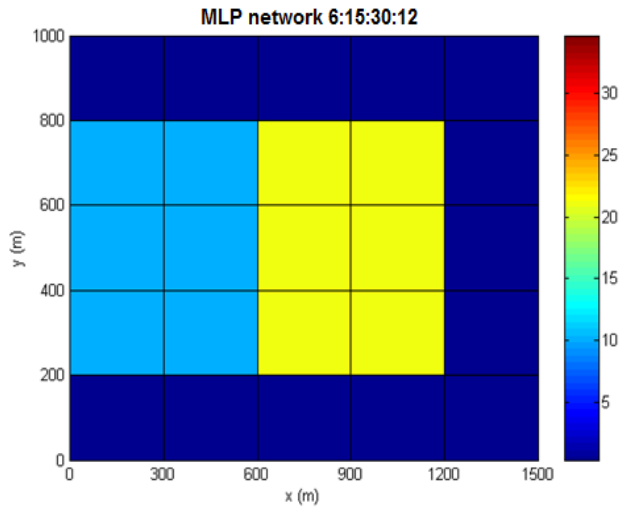
(a)



(b)



(c)



(d)

Fig. 5. Gas flux rate for Experiment-1: (a) true values, (b) quasi-Newton (Q-N), (c) PSO, and (d) MLP-NN.

5. Conclusions

The estimation of gas flux between the surface and the atmosphere is a key issue for several objectives: air pollution quality, climate change monitoring, short and medium range weather forecasting, and ecology. This lecture presents two formulations to address this important inverse problem: optimization problem approach, and artificial neural networks.

When the inverse problem is formulated as an optimization problem, a regularization operator is, in general, required. The optimal solution could be computed using a deterministic or a stochastic searching methods. Deterministic methods has a faster convergence, but stochastic schemes can overcome the local minima. The regularization strategy requires the estimation of the regularization parameter, responsible by a fine balance between adhesion term (square difference between measurements and model data) and of the regularization function. For the present estimation, the PSO inverse strategy does not use the regularization. This is an advantage, but there are much more calculations of the objective function than the deterministic method.

The artificial neural networks produced the better results for the worked example. More investigations are needed to confirm this good performance. We are generating different weather conditions, for all seasons of the year and for different latitudes, to provide more robust conclusions. Any way, the results are very encouraging. As a final remark, two topologies for MLP neural network were identified. Of course, the best result is an obvious criterion to select the topology, but an additional criterion could be based on the complexity: less complex NN (less neurons) is better. Using these two criteria, the NN-1 (6:6:12:12) could be better than NN-2 (6:15:30:12).

The methodology for computing inverse solution was described for *in situ* measurements. The new wave of environmental satellites already started (Clerbaux et al., 1999; Buchwitz et al., 2004; Crevoisier et al., 2004; Carvalho & Ramos, 2010). The next step is to use a cascade of inversions to estimate the surface flux. The first inversion is to determine the gas vertical profile concentration from the radiances measured by the satellite. The second inversion would be to identify the gas surface flux from the gas concentration found by the first inversion. We are going to adapt the methodology described here for this new observation data.

6. Acknowledgments

The authors want to thank Drs. Domenico Anfossi (retired) and Silvia Trini Castelli from the Istituto di Scienze dell'Atmosfera e del Clima (ISAC), Turim (Italy), to allow us to use the LAMBDA code, and for scientific cooperation. Author H. F. Campos Velho acknowledges the CNPq, Brazilian agency for research support – Proc. 311147/2010-0.

7. References

- Anfossi, D. & Ferrero, E. (1997). Comparison among Empirical Probability Density Functions of the Vertical Velocity in the Surface Layer Based on Higher Order Correlations, *Boundary-Layer Meteorology*, Vol. 82, pp. 193-218. ISSN 0006-8314.
- Aster, R. C., Borchers, B. & Thurber, C. H. (2005). *Parameter Estimation and Inverse Problems*, Academic Press, ISBN 0-12-065604-3, USA.
- Bertero, M. & Boccacci, P. (1998). *Introduction to Inverse Problems in Imaging*, Taylor & Francis, ISBN 0-7503-0435-9.
- Bocquet M. (2005a). Reconstruction of an atmospheric tracer source using the principle of maximum entropy, I: Theory. *Quarterly Journal of the Royal Meteorological Society*, Vol. 131, No. 610, pp. 2191-2208. ISSN 0035-9009
- Boissieu, J. De, Papoular, R. J. & Janot C. (1991). Maximum entropy method as applied in quasi-crystallography. *Europhysics Letters*, Vol. 16, pp. 343-347.
- Buchwitz, M., Noel, S., Bramstedt, K., Rozanov, V. V., Eisinger, M., Bovensmann, H., Tsvetkova, S. & Burrows, J. P. (2004). Retrieval of trace gas vertical columns from SCIAMACHY/ENVISAT near-infrared nadir spectra: first preliminary results. *Advances in Space Research*, Vol. 34, No. 809-814.
- Carvalho, A. R. & Ramos, F. M. (2010). Retrieval of carbon dioxide vertical concentration profiles from satellite data using artificial neural networks. *TEMA: Tendências em Matemática Aplicada e Computacional*, Vol. 11, pp. 205-216.
- Clerbaux, C., Hadji-Lazaro, J., Payan, S., Camy-Peyret, C. & Megie G. (1999). Re-trieval of CO columns from IMG/ADEOS spectra. *IEEE Transactions on Geo-science and Remote Sensing*, Vol. 37, pp. 1657-1661. ISSN: 0196-2892
- Crevoisier, C., Chedin, A. & Scott, N. A. (2004). AIRS channel selection for CO₂ and other trace-gas retrievals, *Quarterly Journal of the Royal Meteorological Society*, Vol. 129, pp. 2719-2740.
- Campos Velho, H. F. & Ramos, F. M. (1997). Numerical Inversion of Two-Dimensional Geoelectric Conductivity Distributions from Eletromagnetic Ground Data. *Brazilian Journal of Geophysics*, Vol. 15, pp. 133-143. ISSN: 0102-261X.
- Campos Velho, H. F., Ramos, F. M., Shiguemori, E. H., & Carvalho J. C. (2006). A Unified Regularization Theory: The Maximum Non-extensive Entropy Principle, *Computational and Applied Mathematics*, Vol. 25, pp. 307-330. ISSN: 1807-0302.

- Campos Velho H. F. (2011). Inverse Problems and Regularization (Chapter 8, Part II). In *Thermal Measurements and Inverse Techniques* (H. R. B. Orland, O. Fudym, D. Maillat, R. M. Cotta. (Org.)), pp. 283-313, CRC Press, New York. ISBN: 10 1436845557
- Conte, S. D., C. de Boor (1980). *Elementary Numerical Analysis, An Algorithmic Approach*. McGraw Hill, New York, ISBN 07001224477.
- Cybenko G. (1989). Approximations by superpositions of sigmoidal functions. *Mathematics of Control, Signals, and Systems*, Vol. 2, No. 4, pp. 303-314.
- Degrazia, G. A., Anfossi, D., Carvalho, J. C., Mangia, C., Tirabassi, T. & Campos Velho H. F. (2000). Turbulence parameterization for PBL dispersion models in all stability conditions. *Atmospheric Environment*, Vol. 34, pp. 3575-3583.
- E04UCF routine (1995). NAG Fortran Library. Oxford: Mark 17.
- Enting I. G. (2002). *Inverse Problems in Atmospheric Constituent Transport*, University Press, Cambridge, ISBN 0521812100.
- Elbern, H., Strunk, A., Schmidt, H., & Talagrand O. (2007). Emission rate and chemical state estimation by 4-dimensional variational inversion. *Atmospheric Chemistry and Physics*, Vol. 7, pp. 1725-1783. ISSN: 16807367
- Ferrero, E., Anfossi, D. & Brusasca G. (1995). Lagrangian particle model Lambda: evaluation against tracer data, *International Journal of Environment and Pollution*, Vol. 5, pp. 360-374.
- Fleisher, M., Mahlab, U. & Shamir J. (1990). Entropy optimized filter for pattern recognition. *Applied Optics*, Vol. 29, pp.2091-2098.
- Gimson, N. R. & Uliasz, M. (2003). The determination of agricultural methane emissions in New Zealand using inverse modeling techniques. *Atmospheric Environment*, Vol. 37, pp. 3903- 3912.
- Gryning, S. E. & Lyck, E. (1984). Atmospheric dispersion from elevated source in an urban area: comparison between tracer experiments and model calculations. *Journal of Climate Applied Meteorology*, Vol. 23, pp. 651-654.
- Hadarmard J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique, *Bull. Princeton Univ.*, Vol. 13, pp. 49-52.
- Hadarmard, J. (1952). *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. Dover Publications, N. Y.
- Hansen P.C. (1992). Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, Vol. 34, pp. 561-580.
- Haykin, S. (1999). *Neural Networks, a Comprehensive Foundation*. 2nd Edition, Prentice Hall, New Jersey, ISBN 0132733501.
- Jaynes E. T. (1957). Information theory and statistical mechanics. *Physical Review*, Vol. 106, 620-630.
- Kaipio, J. & Somersalo, E. (2005). *Statistical and Computational Inverse Problems*, Springer, ISBN 0387220739, USA.
- Kennedy, J., & Eberhart, R. C. (1995). Particle swarm optimization. *Proc. IEEE Int'l. Conf. on Neural Networks*, IV, 1942-1948. Piscataway, NJ: IEEE Service Center.
- Luz, E. F. P. (2007). *Estimation of atmospheric pollutant source by particle swarm*. M.Sc. Thesis, National Institute for Space Research (INPE), Applied Computing, (in Portuguese).
- Luz, E. F. P., Campos Velho, H.F. de, Beceneri, J.C. & Roberti, D. R., (2007). Estimating Atmospheric Area Source Strength Through Particle Swarm Optimization, *Inverse Problems, Design and Optimization Symposium*, vol. 1, pp. 354-359. ISBN: 978-1-59916-279-9.
- Li, Y. & Chen, X. (2006). A New Stochastic PSO Technique for Neural Network Training. *International Symposium on Neural Networks (ISNN)*, Chengdu (China), pp. 564-569.

- Morozov, V.A. (1984). *Methods for Solving Incorrectly Posed Problems*. Springer Verlag. ISBN 0387960597.
- Morozov, V. A., & Stessin M. (1992). *Regularization Methods for Ill-Posed Problems*. CRC Press.
- Muniz, W. B., Campos Velho, H. F. & Ramos, F. M. (1999). A comparison of some inverse methods for estimating the initial condition of the heat equation. *Journal of Computational and Applied Mathematics*, Vol. 103, pp.145-163. ISSN: 0377-0427
- Muniz, W. B., Ramos, F. M. & Campos Velho H. F., (2000). Entropy- and Tikhonov-based Regularization Techniques Applied to the Backwards Heat Equation, *Computres & Mathematics with Applications*, Vol. 40, pp. 1071-1084. ISSN: 08981221.
- Paes, F. F., Campos Velho, H.F. de, & Ramos, F. M. (2010). Land fluxes for the biogeochemical cycles estimated by artificial neural networks, *Inverse Problems, Design and Optimization Symposium*. Proceedings in CD-Rom.
- Ramos, F. M. & Giovannini A. (1995). Résolution d'un problème inverse multidimensionnel de diffusion de la chaleur par la méthode des éléments analytiques et par le principe de l'entropie maximale. *International Journal of Heat and Mass Transfer*, Vol. 38, pp. 101-111. ISSN: 0017-9310.
- Ramos, F. M., Campos Velho, H.F. de, Carvalho, J.C. & Ferreira, N.J. (1999). Novel approaches on entropic regularization. *Inverse Problems*, Vol 15, No. 5, pp. 1139-1148. ISSN: 0266-5611.
- Roberti, D.R., Anfossi, D., Campos Velho, H.F. de, & Degrazia, G.A. (2005). Estimation of Location and Strength of the Pollutant Sources, *Ciencia e Natura*, pp. 131-134. ISSN: 0100-8307.
- Roberti D. R. (2005). *Inverse Problems in Atmospheric Physics*. D.Sc. Thesis, Federal University of Santa Maria.
- Roberti, D.R., Anfossi, D., Campos Velho, H.F. de, & Degrazia G.A. (2007). Estimation of Pollutant Source Emission Rate from Experimental Data, *Il Nuovo Cimento*, Vol. 30 C, No. 02, pp. 177-186. ISSN: 0369-3554.
- Rumelhart, D. E., G. E. Hinton, R. J. Williams (1986). Learning representations of back-propagation erros. *Nature*, Vol. 323, pp. 533-536.
- Shannon, C.E. & Weaver W. (1949). *The Matemathical Theory of Communication*. University of Illinois Press.
- Shiguemori, E. H., Campos Velho, H. F., & Silva J. D. S. Da, (2004). Generalized Morozov's principle. *Inverse Problems, Design and Optimization Symposium*, Vol. 2, pp. 290-298.
- Shi, Y. & Eberhart R. C. (1998). A modified particle swarm optimizer. *Proceedings of the IEEE International Conference on Evolutionary Computation*. pp. 69-73, IEEE Press, Piscataway, NJ (USA).
- Smith, R. T., Zoltani, C. K., Klem, G. J. & Coleman, M. W. (1991). Reconstruction of tomographic images from sparse data sets by a new finite element maximum entropy approach. *Applied Optics*, Vol. 30, pp. 573-582. ISSN: 0003-6935.
- Tarantola A. (1987). *Inverse Problem Theory*. Elsevier Science Publishers, Amsterdam.
- Tikhonov, A.N. & Arsenin V.I. (1977). *Solutions of Ill-posed Problems*, John Wiley & Sons.
- Thomson, D. J. (1987). Criteria for the selection of stochastic models of particle trajectories in turbulent flows, *Journal of Fluid Mechanics*, Vol. 180, pp. 529-556.
- Tsallis C. (1988). Possible generalization of Boltzmann-Gibbs statistics, *Journal of Statistical Physics*, Vol. 52, p. 479-487. ISSN: 0022-4715.