DETERMINATION OF A POLLUTANTS SOURCE IN AN ESTUARY THROUGH THE INVERSE PROBLEMS TECHNIQUE

Radael de Souza Parolin
Antônio José da Silva Neto
Pedro Paulo Gomes Watts Rodrigues
Instituto Politécnico, IPRJ/UERJ, Nova Friburgo/RJ, Brazil
radaelsp@gmail.com, ajsneto@iprj.uerj.br, pwatts@iprj.uerj.br

Abstract. A great challenge today is conciliation of water resources utilization with the expansion of cities and human activities. Considering that the water quality of a given water body is necessarily evaluated through the analysis of some biological, physical and chemical parameters, mathematical and computational models able to describe the behavior of such parameters can be an useful tool, given their ability to generate scenarios and, as a consequence, the possibility to support decisions regarding water resources management. In this work, Inverse Problems techniques are applied to estimate the source parameters (intensity and position) of a hypothetical conservative pollutant released in estuarine waters. The studied case here is the estuary of Macaé River, located in the brazilian southeast coast. The pollutant transport was modeled by the advection-diffusion equation, here solved by the Finite Element Method and the Finite Difference Method. The hydrodynamics parameters were assumed known and the mesh applied to the domain was defined according to the discretization method used to solve the direct problem. For estimation of source position here were used the Luus-Jaakola (LJ), the Particle Collision Algorithm (PCA) and the Ant Colony Optimization (ACO) Methods, and to estimate the source intensity was used the Golden Section Method. Besides, a sensibility analysis of hypothetical sampling sites position regarding the source parameters (intensity and position) was performed. In this study, synthetic pollutant concentrations with and without noise were used. For the noiseless data, all methods have successfully achieved the objective function in more than 90% of executions. Considering the number of estimates from different points on the location and also the computational cost, the PCA Method showed the highest performance. On the other hand, for the data with ±5% of noise, all methods had efficiency greater than 85%. Considering the number of estimates from different points on the location, and also the computational cost, again the PCA method showed the best performance. The results of this study demonstrated the potentiality of the Inverse Problems technique to estimate with satisfactory accuracy the location and intensity of a given pollutant source released in estuarine environments, something that can also contribute to possible environmental liabilities identification.

Keywords: pollutant transport, determination of sources, inverse problem, finite element method, finite difference method.

1. INTRODUCTION

The water environment considered in this work is is the estuary of the Macaé River, located on the north coast of the state of Rio de Janeiro, Brazil. The preservation of such environments is justified by great biological diversity. The problem proposed here is to identify the origin and magnitude of a hypothetical release of pollutant that is diluted in the waters of the estuary. Then, is implemented the solution of the direct problem, here modeled by transport equation. This model describes the behavior of a contaminant, involving hydrodynamic parameters, dispersion and a term that represents sources or sinks. The model is coupled with computational intelligence methods for parameter estimation (inverse problem) that can express the location and intensity of the pollutant source considered. Thus, the main objective is the estimation of sources through such parameters.

Concerning the determination of parameters related to pollutant sources of transport models follow some works developed. Shen and Kuo (2001) use a two-dimensional model of eutrophication laterally integrated to model eight state variables of water quality. In the model, 13 parameters from the source term are estimated, which are functions that describe the time rate of growth (or decline) mass by biochemical reactions and external addition (or removal) of the state variables. Revelli et al. (2004) and Revelli and Ridolfi (2005) estimate a function belonging to a source term of a one-dimensional transport problem of pollutants into channels. The source term is composed of two functions (one spatial and another temporal), being the temporal function calculated with boundary conditions and concentrations in a known location. Yang and Hamrick (2005) estimate open boundary conditions in a three-dimensional transport model of salinity in an estuary. Shen et al. (2006) estimate nonpoint sources of fecal coliforms in an estuary. Specifically, regarding the estimation of the location of sources or sinks in models of pollutant transport, developed studies were not found in literature.
2. TRANSPORT PROBLEM OF CONSTITUENTS

The transport of constituents can be described by the advection-diffusion equation (Anderson, 1995; Miranda et al., 2002), which vertically integrated assumes the change in concentration is negligible in the vertical direction. Thus, the two-dimensional representation of constituents transport is expressed by

$$\frac{\partial C}{\partial t} + H \vec{v} \cdot \nabla C - \nabla \cdot (HD \nabla C) + HQ = 0$$

where $C(x, y, t)$ is the concentration of constituent ($kg/m^3$),
$H(x, y, t)$ is the total water depth ($m$),
$\vec{v} = (u, v)$ is the velocity vector of the fluid ($m/s$),
$Q(x, y, t)$ is the source term of $C$ ($kg/m^3s$),
$D$ is the coefficient or tensor of turbulent dispersion ($m^2/s$), given for
$$D = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix}$$
$x$ and $y$ are horizontal independent variables ($m$),
$t$ is the time ($s$),
$\vec{v} = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right)$ is the vector operator $[1/m]$.

For the boundary conditions (Neumann type) and initial condition considered respectively,

$$\frac{\partial C(x_l, y_l, t)}{\partial n} = 0$$
$$C(x, y, 0) = 0$$

where $(x_l, y_l)$ represent the boundaries and $\vec{n}$ the normal vector to them.

For the discretization of the transport model was applied the Finite Element Method (FEM) for the discretization in space (Zienkiewicz and Taylor, 1994, 2000; Domínguez and Hernández, 2007) and the Finite Difference Method (FDM) for discretization in time (Smith, 2004). Finally, we obtain a system of algebraic linear equations (SALE), being solved using the Gauss-Seidel Method (GSM) which is convergent when the system has diagonal dominant (Cunha, 2000).

The field of study of the Macaé River estuary, located in the city of Macaé in the state of Rio de Janeiro, was defined via satellite photos, spanning almost 20 km, from the headwaters to the coast. The bathymetry data (depth according to the mean sea level) of the coast region were taken from the nautical chart 1507, published by the Navy of Brazil in 1974, while the upper region was obtained from Amaral (2003). Due to the limited available bathymetry data of the upper region, the data used were linearly interpolated. For the coast, the data have referred to the nearest existing bathymetry. It is worth noting also that these data are outdated due to several changes in the estuary and nearby performed by human action in recent decades.

Figure 1a shows an example of satellite photo used for the definition of the field of study. In Figure 1b is shown the estuary geometry and the used bathymetry data. The spatial mesh consists of 600 nodes with 917 triangular finite elements, defined after a study of consistency, stability and convergence about the discretization method.

The hydrodynamic model was not solve and because does not exist information about the hydrodynamic variables of the estuary, the vectors of the velocities in the fluvial area were admitted constant with $|\vec{v}| = 1 m/s$ along the river and in the coastal region varies with time, with $\vec{v} = (\cos(\pi t/t_f), 0) m/s$, where $t_f = 43200s = 12h$. As the variation of the water level also is not available, the water column was considered simply the bathymetry. The parameters of dispersion were considered $D_{xx} = 2m^2/s$, $D_{xy} = D_{yx} = 0m^2/s$, $D_{yy} = 2m^2/s$. 
3. FORMULATION OF THE INVERSE PROBLEM

The transport problem of constituents, represented by the advection-diffusion equation (Eq. 1), with the boundary conditions (Eqs. 2 and 3), hydrodynamic variables and adopted parameters is defined as the Direct Problem (DP).

The problem of estimating the source of contaminants was formulated as an Inverse Problem (IP), which is assumed known experimental measurements of concentration and is intended to estimate the location and intensity of a source, represented by parameters of the transport problem (DP).

As the number of experimental data is usually larger than the number of unknowns, the inverse problem is formulated as an optimization problem of finite dimension, which seeks to minimize the objective function that is the sum of squared residues between the calculated and the measured values of the observed variable,

\[ SSR = \left( \tilde{G}_{\text{calc}}(\tilde{P}) - \tilde{G}_{\text{med}} \right) \left( \tilde{G}_{\text{calc}}(\tilde{P}) - \tilde{G}_{\text{med}} \right)^T = \tilde{R}^T \tilde{R} \]  (4)

where \( \tilde{G}_{\text{med}} \) is the vector of measures available,
\( \tilde{G}_{\text{calc}} \) is the vector of the calculated values,
\( \tilde{P} \) is the vector of unknowns,
\( \tilde{R} \) is the vector of residuals.

The main objective is to estimate the location and intensity of a single source, punctual and constant. In this scenario, it is important to understand that the location of the source is not just a parameter of the proposed transport model, but a punctual position (node) in the middle of the two-dimensional discretized estuary. Thus, it is necessary a different treatment from any other parameter analysis. The parameter of source intensity \((\text{kg/m}^3\text{s})\) even constant, but located in separate positions, can lead to one source with different mass flow \((\text{kg/s})\), because the finite elements (triangles) have different dimensions and intensity in each finite element is the average of their nodes, and the node of the source location may belong to different elements.

A sensibility analysis was accomplished (Beck et al., 1985) to evaluate the identification possibility of the parameters front to the transport model. Then, three possible positions of sources were analyzed separately considering the intensity parameter \(F=1\text{kg/m}^3\text{s}\), with five possible positions of collect sensors of the concentration data. Figure 2 shows the locations of the sources and sensors.

Through the sensibility analysis the identification possibility was verified from the referring parameters to the source, as well as strong lineal correlation among them. In that way, we choose that to each estimate of a location of the source is estimate your intensity for this location. With that, to each point (estimate location) an one-dimensional algorithm of estimation can be used for the intensity, of easy use, that covers a considerable interval of possibilities.

The considered sampling data are synthetic data generated from resolving the transport problem (DP), represented by Eq. (1), using known parameters and a time discretization \(\Delta t = 180\text{s}\) in a period of 6\text{hours} (21600s).
The location of the source considered is the location $L_2$ (Fig. 2) with an intensity parameter $F=1\text{kg/m}^3\text{s}$. The sampling data are concentration data at locations $S_1, S_2, S_3, S_4, S_5$ (Fig. 2) for each hour (1, 2, 3, 4, 5, 6 [hours]). Therefore, sampling data comprising 30 data with 5 locations and 6 times.

All programming of the direct problem and the inverse problem was developed in C++ language as well as most numerical analyzes performed. The computational cost of each execution of the DP depends on the location and intensity of the source. For the scenario considered real, to be estimated, with the intensity and location known, the computational cost of execution of the DP is approximately 8.86s considering an accuracy of $10^{-2}$ to the solution of the SAL (PC with processor Intel Core i5, 2.67 GHz and 4 GB of RAM memory, using Windows 7 Professional operating system).

4. META-HEURISTICS

To estimate the location of the source were applied the methods Luus-Jaakola (LJ), the Particle Collision Algorithm (PCA) and Ant Colony Optimization (ACO), and to estimate the intensity the Golden Section method (GS). Regarding the estimation of the location, if the point has already been previously estimated, goes up for a new iteration of the applied methods, using their stored information if necessary. Estimated a new point, the method GS is then used to estimate the source intensity at that point. To emphasize, the objective function of the minimization problem is the sum of squared residues (SSR) between the calculated data and sampling data (Eq. 4).

The choice of the method LJ is due to its characteristic purely geometrical, by reducing the search region leading a tendency of global search to a local search with passing of the iterations (Luus e Jaakola, 1973). The PCA method has a feature that evaluates with its operators when performing a global, local, or totally random search (Sacco et al., 2006). The use of the method ACO is justified because it is a proposed method for solving discrete problems, which is the case here considered (Dorigo et al., 1996; Silva Neto e Becceneri, 2009). For estimation of intensity parameter, the Golden Section method was chosen because it has the characteristic of reducing a one-dimensional search interval, which in this case is defined by only one parameter (Bazaraa et al., 2006).

As modifications of these meta-heuristics, for the method PCA is just accomplished a modification in the function “Perturbation” of the method given by Sacco et al. (2006), presented in Fig. 3.

Already for the method ACO, the modifications in the method given by Dorigo et al. (1996) are similar modifications proposed by Silva Neto and Becceneri (2009). As the source location is defined by a point of the space mesh which is irregular, it is not possible to use two-dimensional coordinates for implementing the method ACO. Thus, we propose to represent the grid points by a vector, obtaining also a vector of probability and another of pheromone. To keep the idea of a two-dimensional geometry and also to value the neighborhood of points, since the same point already selected not recalculate the direct problem, the pheromone is updated not only to the selected point, but a neighborhood within a defined ray, decreasing with distance from the selected point. The algorithm of the ACO method used for the minimization problem is shown in Figs. 4 and 5.
Define the number of ants of the population $n_f$, the deposit rate of pheromone $f_0$, the evaporation rate of pheromone $f_e$, with $0 \leq f_e < 1$, the space of performance of pheromone $d_r$, the roulette parameter $q_0$, with $0 \leq q_0 \leq 1$, and the maximum number of generations $g_{max}$.

Compute the initial matrix $F^{(0)}$, with $f_i = 0$, and the initial matrix $PC^{(0)}$, with $pc_i = 0$, where $i = 1$ to $nn$ (number of nodes).

Generate the initial population of $n_f$ ants like random solutions and evaluate them.

Do the best ant of the population be $x_b$

for $g = 1$ to $g_{max}$ iterations

\[ f_i^{(g)} = f_i^{(g-1)}(1-f_e) + \delta f_i d_i + f_0 \]

where $d_i$ is the distance between $x_b$ and node $i$ of the generation $g-1$.

and $\delta f_i = 1$ to nodes with $d_i < d_r$, else $\delta f_i = 0$

\[ pc_i = \frac{\sum_{j=1}^{nn} f_j^{(g)}}{\sum_{j=1}^{nn} f_j^{(g)}} \]

Generate_Population ()
Evaluate Population
Do the best ant of the population be $x_b$

end
5. ESTIMATION OF THE CONTAMINANTS SOURCE

For the application of meta-heuristics in the source estimation can define different stop criteria besides a maximum number of iterations/generations. For the meta-heuristics applied in the estimation of the source location, the stop criteria used were: successful with the objective function for the estimated point, defined smaller than a desired considerable value, and a maximum number of different estimated points (also referred with a maximum number of evaluations of the objective function). The maximum number estimation of different points was regarded 60, which represents 10% of the total number of nodes of the space mesh. Regarding the estimation of the source intensity, specifically the application of the Golden Section method, the stop criteria is the size of the interval be smaller than a desired value. Moreover, each method stores and updates the best result during its iterations/generations.

The operators of the LJ method were defined as $n_{int} = 10$, $n_{ext} = 10$ and $\varepsilon = 0.30$. This way, we can estimate up to 100 points, reminding that equal points do not perform the direct problem. The reduction of side of the squares in the search region is 30% to each external iteration, which represents a reduction of 51% in area. The initial search region used is a square with side of 10000m. The operators of PCA method were defined as $N_{max} = 50$ e $n_{e} = 5$, and the random solution of the “Scattering” function was annulled setting $p_{scattering} = 0$. The operators of ACO were defined as $n_{f} = 20$, $f_{o} = 0.1$ , $f_{c} = 0.03$ , $q_{o} = 0.2$ , $dr=500$ e $g_{max}=50$. For these methods used in the estimation of the source location we tested different values for its operators, which empirically, were defined by the best results.

For the use of the GS method was used the initial interval $[0.1; 3.913]$ with the stop criteria $l_{c} = 0.5$. This interval was chosen to allow estimation of a value close to the real intensity parameter of the source which is 1 with a smaller number of iterations. In this conditions, seven different intensities are calculated for the selected location, namely the direct problem is calculated seven times every new estimated location.

20 executions were accomplished for each meta-heuristic, which are then compared by a nonparametric test, the Wilcoxon signed rank test (Kanji, 2006), considering a confidence level of 95%. In Table 1 are presented the averages of the results and in Tab. 2 the comparison of the efficiency of the methods in terms of objective function (SSR), the number of different estimated points for the location (NP) and the computational cost (time).

<table>
<thead>
<tr>
<th>Methods</th>
<th>Success (%)</th>
<th>NP</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LJ-SA</td>
<td>100</td>
<td>40</td>
<td>2094,15</td>
</tr>
<tr>
<td>PCA-SA</td>
<td>95</td>
<td>24</td>
<td>1362,55</td>
</tr>
<tr>
<td>ACO-SA</td>
<td>90</td>
<td>36</td>
<td>1906,30</td>
</tr>
</tbody>
</table>

Table 2. Wilcoxon test for comparison of meta-heuristics

<table>
<thead>
<tr>
<th>Compared Criterion</th>
<th>Compared Methods</th>
<th>$R^+$</th>
<th>$R^-$</th>
<th>Critical Value</th>
<th>$p$-value</th>
<th>Significant Difference?</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQR</td>
<td>LJ-SA vs. PCA-SA</td>
<td>95</td>
<td>115</td>
<td>52</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>LJ-SA vs. ACO-SA</td>
<td>85,5</td>
<td>124,5</td>
<td>52</td>
<td>0,5</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>PCA-SA vs. ACO-SA</td>
<td>94,5</td>
<td>115,5</td>
<td>52</td>
<td>0,5</td>
<td>No</td>
</tr>
<tr>
<td>NP</td>
<td>LJ-SA vs. PCA-SA</td>
<td>185,5</td>
<td>24,5</td>
<td>52</td>
<td>0,0026</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>LJ-SA vs. ACO-SA</td>
<td>125</td>
<td>85</td>
<td>52</td>
<td>0,4552</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>PCA-SA vs. ACO-SA</td>
<td>40,5</td>
<td>169,5</td>
<td>52</td>
<td>0,016</td>
<td>Yes</td>
</tr>
<tr>
<td>Time</td>
<td>LJ-SA vs. PCA-SA</td>
<td>178</td>
<td>32</td>
<td>52</td>
<td>0,0064</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>LJ-SA vs. ACO-SA</td>
<td>134</td>
<td>76</td>
<td>52</td>
<td>0,279</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>PCA-SA vs. ACO-SA</td>
<td>52</td>
<td>158</td>
<td>52</td>
<td>0,0479</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Note: The best meta-heuristics are highlighted in bold.

As can be seen in Tab. 2, the objective function achieved for the three meta-heuristics has no significant difference, evidencing the success of over 90% (Tab. 1). In terms of efficiency, the meta-heuristic PCA was superior, achieving success with the estimation of a smaller number of points (NP), besides a lower computational cost (time).

In real situations with sampling data collected experimentally, there is always the existence of errors inherent in the equipment used, the numerical accuracy adopted, design of experiments, the experimental conditions and simplifications, and human errors in the experimental handling.

Because of the use of synthetic data sampling and the impossibility of escape of experimental errors, it is proposed to introduce noise in the data to bring them closer to the reality of the experimental field. Furthermore, it is important to evaluate the methods of solving the inverse problem facing this adversity, with more realistic data.

The noise introduced in the sampling data was at most 5% at around the same. The operators of the meta-heuristics used were the same as in previous applications.

Then, 20 executions were accomplished for each meta-heuristic, which are then compared by Wilcoxon test, considering a confidence level of 95%. In Table 3 are presented the averages of the results and in Tab. 4 the comparison...
of the efficiency of the methods in terms of objective function (SSR), the number of different estimated points for the location (NP) and the computational cost (time).

Table 3. Summary of mean of the applications of different meta-heuristics using data sampling with noise

<table>
<thead>
<tr>
<th>Methods</th>
<th>Success (%)</th>
<th>NP</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LJ-SA</td>
<td>95</td>
<td>39</td>
<td>2124.8</td>
</tr>
<tr>
<td>PCA-SA</td>
<td>90</td>
<td>26</td>
<td>1475.45</td>
</tr>
<tr>
<td>ACO-SA</td>
<td>85</td>
<td>43</td>
<td>2088.15</td>
</tr>
</tbody>
</table>

Table 4. Wilcoxon test for comparison of meta-heuristics using sampling data with noise

<table>
<thead>
<tr>
<th>Compared Criterion</th>
<th>Compared Methods</th>
<th>R+</th>
<th>R−</th>
<th>Critical Value</th>
<th>p-value</th>
<th>Significant Difference?</th>
</tr>
</thead>
<tbody>
<tr>
<td>LJ-SA vs. PCA-SA</td>
<td>96.5</td>
<td>113.5</td>
<td>52</td>
<td>1</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>SQR</td>
<td>LJ-SA vs. ACO-SA</td>
<td>88</td>
<td>122</td>
<td>52</td>
<td>0.875</td>
<td>No</td>
</tr>
<tr>
<td>PCA-SA vs. ACO-SA</td>
<td>94</td>
<td>116</td>
<td>52</td>
<td>0.5</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>NP</td>
<td>LJ-SA vs. PCA-SA</td>
<td>170.5</td>
<td>39.5</td>
<td>52</td>
<td>0.0145</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>LJ-SA vs. ACO-SA</td>
<td>80</td>
<td>130</td>
<td>52</td>
<td>0.3505</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>PCA-SA vs. ACO-SA</td>
<td>21.5</td>
<td>190.5</td>
<td>52</td>
<td>0.0018</td>
<td>Yes</td>
</tr>
<tr>
<td>Time</td>
<td>LJ-SA vs. ACO-SA</td>
<td>158</td>
<td>52</td>
<td>52</td>
<td>0.0479</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>PCA-SA vs. ACO-SA</td>
<td>39</td>
<td>171</td>
<td>52</td>
<td>0.0137</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Note: The best meta-heuristics are highlighted in bold.

As in previous applications, even when using sampling data with noise, the objective function achieved for three meta-heuristic has no significant difference which can be seen in Tab. 4, evidencing the success of over 85% (Tab. 3). In terms of efficiency, the meta-heuristic PCA was superior again, achieving success with the estimation of a smaller number of points (NP), besides a lower computational cost (time).

6. CONCLUSIONS

To estimate the location and intensity of the source in the proposed transport problem of the idealized Macaé River estuary, it is perceived through sensibility analysis that both parameters are correlated, which makes the estimation procedure hard through the inverse problem. According to this aspect, alternatively was chosen to estimate the location of the source, and during this process, for each specific location, estimating the intensity parameter.

Using synthetic sampling data without noise the meta-heuristics applied were comparably efficient to estimate the location and intensity of the source with respect to the objective function (SSr) succeeding in more than 90% of executions. Considering the number of different estimated points for the location and also the computational cost, the PCA method was superior.

With the use of noise on the order of 5% in the sampling data, the meta-heuristics applied were again efficient as the objective function, successfully greater than 85%, without showing significant difference between them. When evaluating the number of different estimated points to the source location and the computational cost, once again the PCA method overcomes the implementation of the methods LJ and ACO.

A new perspective of improvement of the results would be the application of hybrid methods, engaging different methods in a single goal, with different characteristics, using what there is best individually. This task of subtle experimentation is part of the work continuity of sources estimation of pollutants in rivers and estuaries.

7. ACKNOWLEDGEMENTS

The authors acknowledge the support provided by CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, and FAPERJ, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro.

8. REFERENCES


9. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.