PARAMETER ESTIMATION IN MODEL OF ESTUARINE HYDRODYNAMICS BASED ON GENETIC ALGORITHMS.

Edgar Barbosa Lima, ebarbosa@iprj.uerj.br
Pedro Paulo Gomes Watts Rodrigues, pwatts@iprj.uerj.br
Antônio José da Silva Neto, asjneto@iprj.uerj.br
Instituto Politécnico, Universidade do Estado do Rio de Janeiro, Nova Friburgo, Riode Janeiro, Brasil

Mirtha Irizar Mesa, mirtha@electrica.cujae.edu.cu
Orestes Llanes Santiago, orestes@electrica.cujae.edu.cu
Department of Automation and Computers, Technical University of Havana (I.S.P.J.A.E.), Ciudad de La Habana, Cuba

Jader Lugon Junior, jljunior@firjan.org.br
Campus Macaé, Instituto Federal de Educação, Ciência e Tecnologia Fluminense, Macaé, Rio de Janeiro, Brasil
Centro de Tecnologia SENAI RJ Ambiental, Sistema FIRJAN, Rio de Janeiro, RJ, Brasil

Abstract. *The deterministic models solve the hydrodynamics and the mass transport of a given substance in a water body. Such models present themselves as more adequate tools to diagnose and assess fluvial environments, because once calibrated, many different scenarios can be created, being enough to modify the input given to the model. These models obey to a certain sequence of development and calibration. First, it must be developed a hydrodynamic model (HDM), which can solve the space-time distribution of velocities and water level. This model is superposed with a transport model of conservative substance (TMCS). Besides the two models already mentioned, a transport model of non-conservative substance (TMNCS) is developed. So, the HDM is based on the well established equations from fluid mechanics, and some simplifications are assumed. The TMCS is expressed by a equation where the main uncertainty is the definition of the dispersion coefficients, while for the TMNCS, much of the uncertainty involves the functional dependencies of the reactions that represent the removal and production of the substance in the riverine/estuarine environment. These last two categories of models are generically called Water Quality Models (WQM).*
Genetic algorithms are tools for searching in complex spaces which have been successfully used in many problems in differents areas. In this work a genetic algorithm is used for estimation of hydrodynamics parameters. Simulation studies show the method effectiveness and give support to decisions on water supply, pollution dispersion and also on the mixture zone studies.

Keywords*: Inverse Problems, Parameter Estimation, Genetic Algorithms, Water resources, Water Quality.*

1. INTRODUCTION

The Brazilian law establishes the monitoring and environmental diagnostic as one of the instruments for the environment management. In particular regarding water resources, the use of models with different complexity is gaining importance, to be used with the purpose of decision support to matters involving the management of these resources. Mathematical and computational models belong to this category of tools. Although they are an abstraction and simplification of the simulated real systems, such models, if properly calibrated and validated, are extremely useful, mainly due to their capacity to provide different scenarios and to manipulate a great number of variables.

Historically, the analysis of dissolved and suspended constituents in river and estuaries began with the development of the so called "black-box" models, where the removal of constituents was considered through existing in and out mass balance of a certain substance for a fraction of the domain of interest. Such approach is based on intensive monitoring, with obvious logistical and economical limitations. Alternatively, the behavior of substances in rivers can be estimated through mathematical models, either deterministic or stochastic ones. The stochastic models consist of statistical treatment of data in such a way that one can define, for example, seasonal tendencies of the pollution concentration. These models can be used to optimize monitoring programs, making possible to establish sampling and water analysis priorities, as well as the ideal sampling frequencies.

The set of equations that represents the deterministic model is usually numerically solved, creating what is called numerical simulation. The basic idea of the numerical simulation is the discretization process, which reduces the physical continuum domain with an infinity number of variables to a discrete problem, with a finite number of variables, in a way that it can be computationally solved.

The solution of the set equations depends on parameters that are mostly difficult to obtain. Such parameters like roughness height and turbulent dispersion coefficients are used for model calibration, usually throughout a manual method (Vaz *et al.*, 2007).

In this work an automatic method is proposed in order to obtain the roughness height and turbulent dispersion coefficients of a estuary, by means of the solution of a inverse problem bases in the Genetic Algorithm Method, using the Macaé river estuary model developed with the MOHID system (Water Modeling System, Instituto Superior Técnico - Universidade de Lisboa). Results obtained with this method are presented and discussed.

2. MODELATION

The mathematical formulation adopted to simulate the mass transport in estuaries and rivers usually takes into account the time variable. Regarding the spatial coordinates, such formulation can be three, two or one-dimensional, depending on the physical and morphological characteristics of the system. The estuarine flow is mainly three-dimensional, as a function of the stratification that takes place in the water column, with the salt water (denser) flowing under the freshwater. However, estuaries subjected to large tidal amplitude (more than 4 meters) and with a relatively small river flow, can be simulated with two-dimensional hydrodynamic equations (Dyer, 1997; Miranda et al, 2002; Rosman, 1989), where it is considered that the flow variables have small changes in the depth direction. These equations are

$$\frac{\partial h}{\partial t}+\frac{\partial (uh)}{\partial x}+\frac{\partial (vh)}{\partial y}=0 \quad (1)$$

$$\frac{\partial u}{\partial t}+\frac{\partial (uu)}{\partial x}+\frac{\partial (uv)}{\partial y}=K_x\frac{\partial^2 u}{\partial x^2}+K_y\frac{\partial^2 u}{\partial y^2}+F \quad (2)$$

$$\frac{\partial v}{\partial t}+\frac{\partial (uv)}{\partial x}+\frac{\partial (vv)}{\partial y}=K_x\frac{\partial^2 v}{\partial x^2}+K_y\frac{\partial^2 v}{\partial y^2}+F \quad (3)$$

where h is the free surface level, u and v are respectively the longitudinal and transversal velocity components, K_x and K_y are the turbulent dispersion coefficients and F is the source/sink term, that can include parametrization of the momentum generation/dissipation due to wind forces or the friction in the estuary bed. Assuming source and drain only due to friction of the bed, can parameter the tension (T) for

$$T=\frac{\rho g u(u^2+v^2)}{{C_h}^2} \quad (4)$$

were C_h is the coefficient of Chézy that consider the roughness height in your formulation.

The transport of solute and suspended constituent can be described by a 2D-horizontal equation of advection-dispersion, expressed by

$$\frac{\partial c}{\partial t}+\frac{\partial (uc)}{\partial x}+\frac{\partial (vc)}{\partial y}=D_x\frac{\partial^2 c}{\partial x^2}+D_y\frac{\partial^2 c}{\partial y^2}+R \quad (5)$$

where c is the solute or suspended constituent concentration, Dx and Dy are turbulent dispersion coefficients and R takes into account all internal reactions and external sources/sinks that can change the concentration of the simulated constituent. Many numerical solutions have been proposed for this set of 4 equations with 4 unknowns (h, u, v and c). The MOHID simulator solve this equations with the Finite Volume Method (Versteeg and Malalasekera, 1995), using the UPWIND scheme for the advective terms, central differences for the dispersive terms and Crank-Nicolson in time.

This model was applied to the Macaé estuary, located at the southeast of Brazilian coast (Fig. 1). The Macaé basin, which has a population of around 141.000 people, has been submitted to some severe environmental impacts, especially its estuarine zone. This has motivated the authorities to implement some actions, which include the computational modeling of estuarine waters.

The modeled domain included an extension of approximately 20 km, from the head to the outer region of the estuary, at the coast. It was adopted a spatial discretization of 40 m, based on quadratic cells. The bathymetry data of the coast region was taken from the nautical chart 1507, edited by Brazilian Navy in 1974, while the upper region bathymetry was obtained from Amaral (2003). The Figure 2 shows the discretized domain, and the adopted bathymetry.

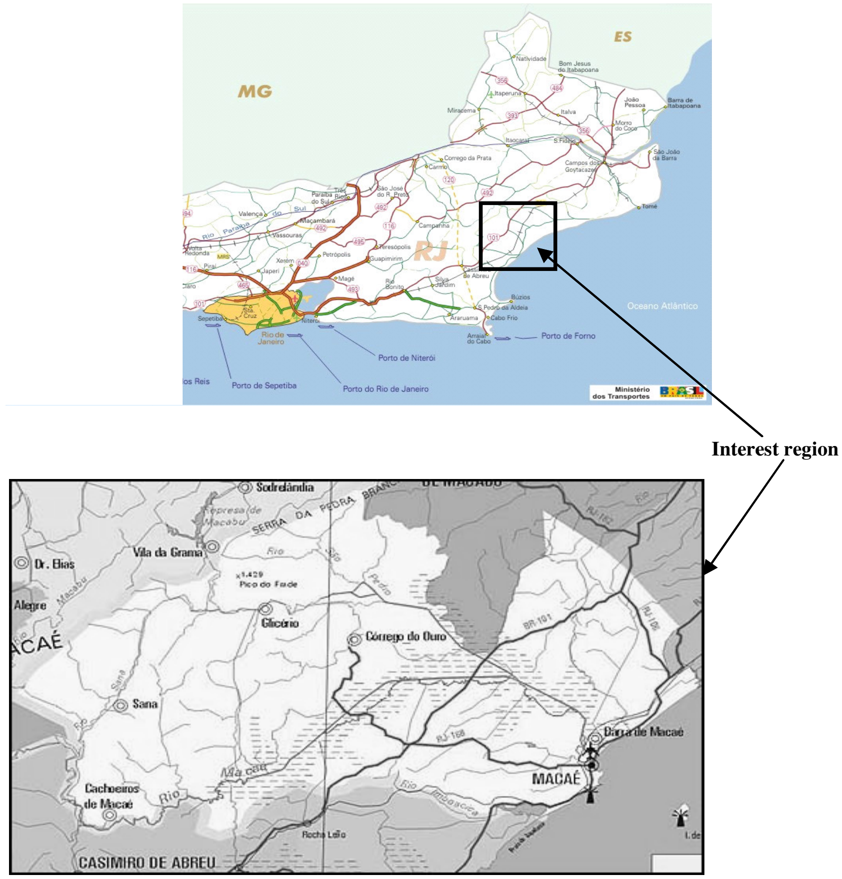

Figure 1: Estuary localization and detail of simulated region in this study

(http://www.macaetour.com.br/mapas.html)

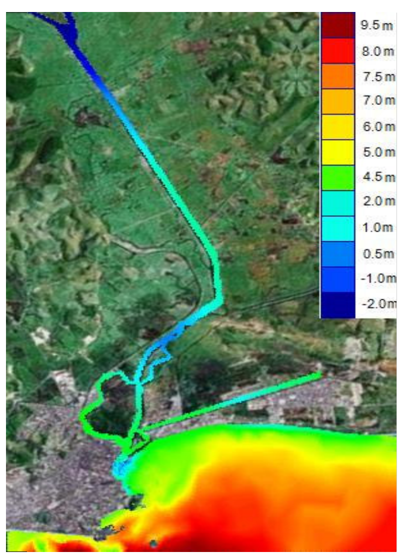

Figure 2: Simulated domain and model bathymetry used.

Two boundary conditions were prescribed for the hydrodynamic model: (i) in the riverine boundary it was set a river discharge of 7.8 m^3 s^{-1}, typical of the dry season for the Macaé river close to estuarine region (Amaral, 2003); (ii) in the marine boundary it was simulated an astronomic tide with 17 components, of which the amplitude and phase are shown in Table 1.

Table 1: Period and amplitude of the components adopted in the tidal simulation model

Components	Period (s)	Amplitude(m)	Components	Period (s)	Amplitude(m)
M2	44714.16	0.369	MN4	22569.03	0.012
S2	43200	0.191	MS4	21972.02	0.011
O1	92949.63	0.100	2N2	46459.35	0.010
K1	86164.09	0.059	M1	89399.69	0.009
K2	43082.05	0.054	M3	29809.44	0.004
N2	45570.05	0.046	MO3	30190.69	0.003
Q1	96726.08	0.026	MK3	29437.7	0.002
M4	22357.08	0.024	SN4	22176.69	0.002
P1	86637.21	0.021			

The main objective of this study was to perform a sensitivity analysis and to simulate the transport of conservative constituents in the estuarine region, more specifically salt. So, it was necessary to specify the value of this property at the riverine, adopted as 0.037 psu (unit of salinity), and at the sea, adopted as 36 psu. As initial condition a salinity of 20 psu was set for the whole dominium.

Results obtained with this model are presented at Lima *et al* (2008) and Rodrigues *et al* (2009).

3. INVERSE PROBLEM FORMULATION

3.1. Inverse solution for a system of partial differential equations

In most of the scientific disciplines and particularly in engineering there are problems characterized by differential equations with associated initial and boundary conditions. When these problems are solved in a direct way, the result is generally a functional relationship or a system of equations, which can be used to calculate values of the dependent variable for given values of the independent variable. The inverse solution of systems of partial differential equations constitutes a complex problem, for which there are no universally accepted methods. Given an applicable direct solution to a system of partial differential equations, it is possible to propose an inverse problem as a problem of optimization. An algorithm to achieve this is (Karr *et al*, 2000):

- Suppose a solution to the inverse problem. This can include the supposition of an initial or boundary condition, or a typical parameter for a given problem.
- Feed the supposed condition to the direct solution of the partial differential equation system, calculating in this way values of the dependent variable y. Here the output of the direct solution is a vector of values corresponding to the times in which the values of y are measured. This vector of solutions will be denoted as calculated and it will be represented as $\hat{y}$.
- Compare the calculated values $\hat{y}$ with the values of the dependent variable y measured in consistent times with those for which $\hat{y}$ was calculated.

The success of this approach is the mechanism for which the supposed condition is improved in the subsequent invocations of the first step. Optimization is the procedure to upgrade the suppositions of the conditions and in this case a genetic algorithm, whose characteristics are explained in the next section, will be used.

The most applied function in the measure of prediction error is the sum of the square error (SSE).

$$SSE(\hat{y},\hat{\theta}) = \sum_{t_i=1}^{N_T} (y(t_i) - \hat{y}(t_i,\hat{\theta}))^2 \quad (5)$$

where θ represents the parameters to be estimated and N_T is the total number of experimental data.

3.2 Genetic Algorithm

One method of Evolutionary Computation with practical applications is the Genetic Algorithm (GA) Fig. 3 shows the general structure of an Evolutionary Algorithm.

```
t = 0
Create Initial Population P(t)
While no termination criterion is satisfied
    Evaluate fitness P(t)
    Select reproducers from P(t)
    Generate a new population P(t+1)
    t = t + 1
End
```

Figure 3. General structure of an Evolutionary Algorithm.

GA is based in three basic operators selection, crossover or recombination and mutation:

Selection operator selects chromosomes in the population for reproduction. The fitter the chromosome, the more times it is likely to be selected to reproduce;

Crossover operator randomly chooses a locus and exchanges the subsequences before and after that locus between two chromosomes to create two offspring. For example, the strings 10000100 and 11111111 could be crossed over after the third locus in each to produce the two offspring 10011111 and 11100100. The crossover operator roughly mimics biological recombination between two single−chromosome (haploid) organisms;

Mutation operator randomly flips some of the bits in a chromosome. For example, the string 00000100 might be mutated in its second position to yield 01000100. Mutation can occur at each bit position in a string with some probability, usually very small (e.g., 0.001) (Melanie, 1999).

These algorithms should work in a wide interval of their parameters, but with differences in the efficiency, what indicates the importance of the designer's approach. Another of the aspects to consider in a GA is the fitness function, which offers information about the quality from the possible solutions to a problem. Execution parameters and fitness function define the GA completely. Selection, recombination and mutation processes form a generation in the execution of a GA, and are executed until a satisfactory solution or a specified number of generations is reached.

Another of the aspects to consider in a GA is the fitness function, which offers information about the quality from the possible solutions to a problem. Execution parameters and fitness function define the GA completely. Selection, recombination and mutation processes form a generation in the execution of a GA, and are executed until a satisfactory solution or a specified number of generations is reached.

3.3 Parameters estimation in model of estuarine hydrodynamics

To estimate the roughness height and the turbulent dispersion coefficients of the Macaé river estuary, a short period of time was adopted, in order to reduce the computational time required in this calculation. There are two variables that can be easily obtained in the field, the water free level (using a ruler) and salinity (using a condutivimeter). A study was performed to choose the variable that is more sensitive for the parameters of interest.

According to Abbot and Basco (1980), the roughness height at the Macaé river estuary, that presents a great amount of sediment transport, is somewhat between 0.007 m and 0.05 m. Considering the model configuration adopted here (time and space discretization, etc) MOHID hydrodynamic module would not accept values outside the range between 0.01 m^2s^{-1} and 5.0 m^2s^{-1} for the turbulent diffusion coefficient, which represents a quite large interval within it a number must be chosen. In Fig. 4 it is presented the results for different values of roughness height and turbulent dispersion coefficients inside the above intervals. One can see that the water free level is not sensitive for the turbulent dispersion variation, while the salinity is sensitive to both.

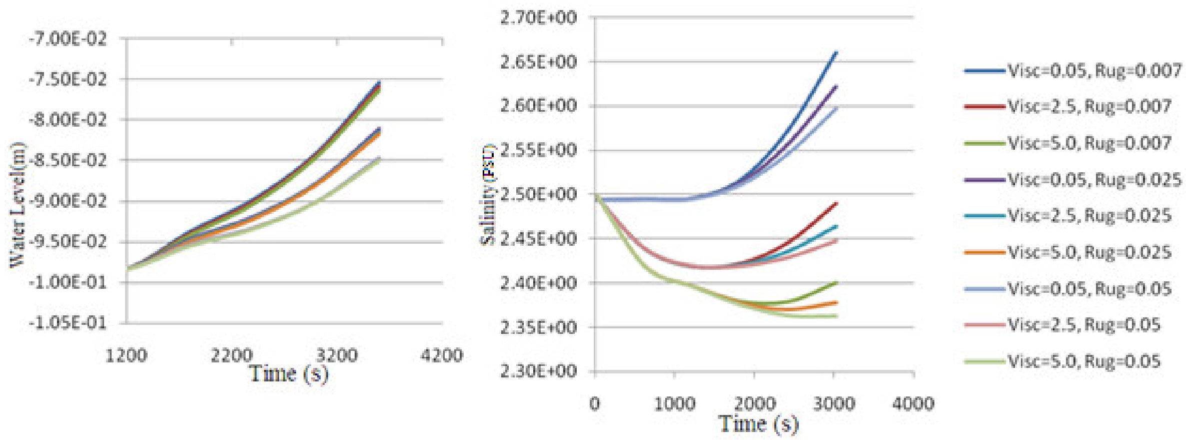

Figure 4. Results obtained with different parameter values for water free level and salinity.

Considering this observed aspects salinity results where choose as the state variable for comparison. Besides, the first results were performed considering a period of one hour of simulation, after the lowest point of tide, in order to capture the salinity penetration at the estuary with minimum influence of previous results.

Since the MOHID use text files, with text format, for data input, it was opted to change the parameters values inside these files and run MOHID in an automatic way, so that it was not necessary to change anything inside its code. In this way, using a FORTRAN code for GA, the MOHID is executed and returns its results GA, and so on. This scheme is presented in Fig. 5.

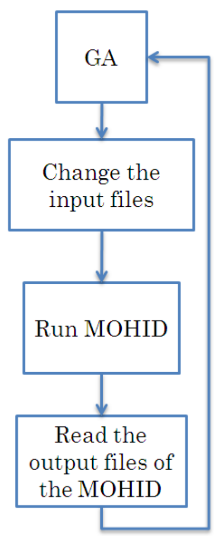

Figure 5. GA scheme for the estimation of parameters using MOHID system.

4. RESULTS

As experimental data were not available right now, it was tested using simulated data obtained using the MOHID model using the values of 2.5 m^2s^{-1} for the turbulent dispersion coefficients and 0.025 m for the roughness height. It was used 60 values of salinity obtained in a period of 1 hour of simulation. The GA used 30 generations with 5 individuals each, with a percentage of 0.5 of crossover and 0.04 of mutation.

In Fig. 6 it is shown the results error using values obtained for 1 hour simulation that resulted in the following values:

Roughness height = $2.596063037058265 \times 10^{-2}$ m
Turbulent dispersion coefficients = 2.49212598501463 m^2s^{-1}

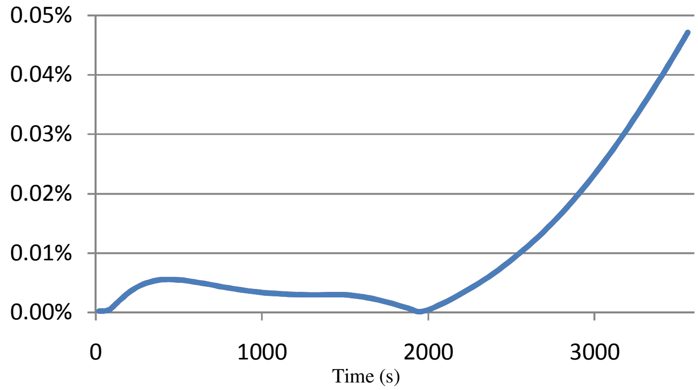

Figure 6. Results error with estimated parameter for 1 hour simulation.

Since the error tent to increase with time, it was used the same estimated parameters to compare with the original ones after 12 hours of simulation (see Fig. 7). One can see that even when using a much larger time period for the parameter estimation, the error does not reach 1%, showing that this approach is in fact, efficient.

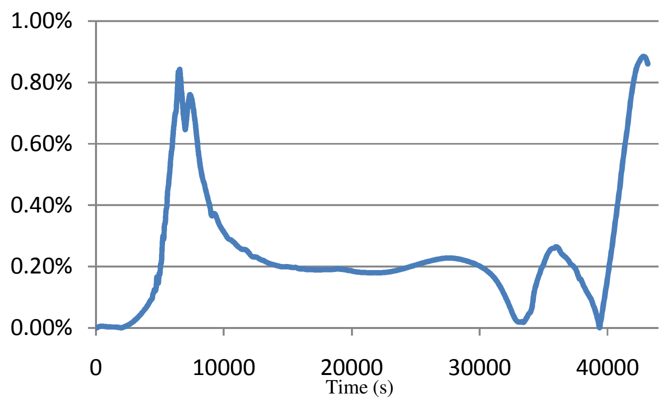

Figure 7 Results of 12 hours simulation using the parameters estimated with 1 hour simulation data.

In order to evaluate the robustness of the method, noise was added to the simulated data. A 4% noise is considered usual for this kind of measurements. In Fig. 8 is shown the comparison between the original result and the 4% noisy data.

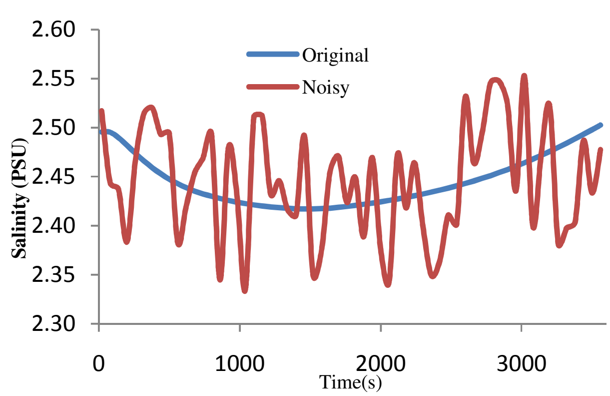

Figure 8. Original results and 4% noisy data for 1 hour simulation.

Due to the small variation of the data in the 1 hour period, the noise represents a great function variation, for this reason the parameter estimation in this example was proven to be low efficient, resulting in:

Roughness height = $4.390551248110655 \times 10^{-2}$ m

Turbulent dispersion = 1.79763779624945 m^2s^{-1}

For this reason, the simulation period was increased for 2 hours and is shown in Fig. 9.

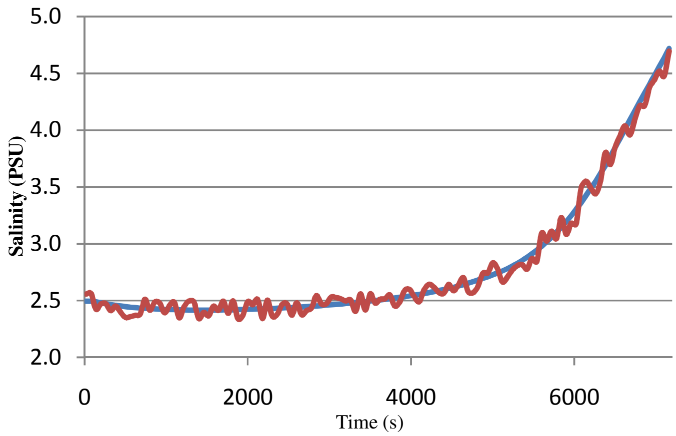

Figure 9. Original results and 4% noisy data for 2 hour simulation.

Using the 2 hours period of simulation, the results were much closer to the original ones:

Roughness height = $2.629921305191329 \times 10^{-2}$ m

Turbulent dispersion = 2.49212598501463 m^2s^{-1}

In Fig. 10 is presented the absolute error for a 12 hours simulation using the estimated values. One can see that the noise error is not significantly increased, being below 1.2%.

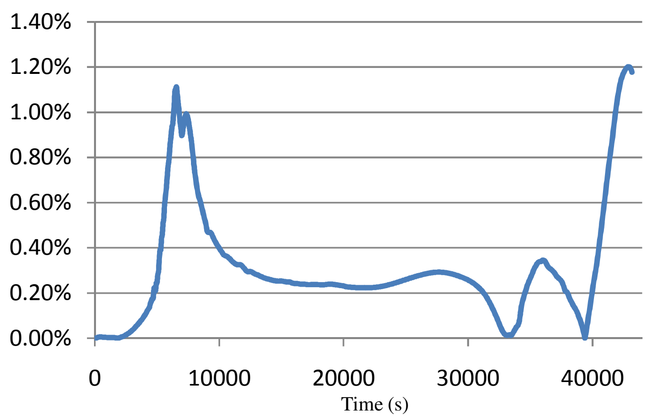

Figure 10. Error for 12 hours simulation considering parameters estimated with 2 hours and 4% salinity noise data.

5. CONCLUSIONS AND FUTURE WORKS

A better parameter estimation has a crucial importance for the modeling of pollutant dispersion in water bodies, especially in studies focused on environmental impact in mixing zones.

The GA approach was proven to be rather efficient for the estimation of the parameters roughness height and turbulent dispersion coefficients, resulting to a calibration with error below 1.2%. Besides, one can see that even using a short period of measurements time it was capable of estimating the parameters with good precision and calibrating the model for a larger time. Also it was observed that, when the data noise is larger, the period of measurements should be increased to obtain a good estimation.

Throughout the study of Fig. 4, one can see that the salinity is more sensitive to the turbulent dispersion coefficients, while the water free level is more sensitive to the roughness height variation. So it was proposed that in future works one could use both measurements to estimate those parameters, and also using other inverse problems methods.

6. ACKNOWLEDGEMENTS

The authors Acknowledge the financial support provided by CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico, FAPERJ, Fundação Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro, and CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

7. REFERENCES

Abbot, M. B. and Basco, D. R., 1989, “Computational Fluid Dynamics, an Introduction for Engineers.”, Logan Group, UK Limited.

Amaral, K., 2003, “Macaé Estuary: Modelling as a Tool for an Integrated Management of Water Resources”, M.Sc. Thesis, COPPE/UFRJ, Rio de Janeiro, RJ, Brazil, pp. 150. (in Portuguese).

Baird, J.I. and Whitelaw, K., 1992, “Water Quality Aspects of Estuary Modelling. In Roger A. Falconer (Ed.) Water Quality Modelling.”, Ashgate Publishing Limited, Aldershot, pp. 139.

CONSÓRCIO INTERMUNICIPAL MRA-5., 2004, “The Design of Macaé River (RJ) Water Basin Plan, Rio de Janeiro”, pp. 52 (in Portuguese).

Dowding, K. J., Blackwell, B. F. and Cochran, R. J. , 1999, “Applications of Sensitivity Coefficients for Heat Conduction Problems”, Numerical Heat Transfer, Part B, v. 36, pp. 33-55.

Dyer, K. , 1997, “Estuaries. A Physical Introduction”, John Wiley & Sons Ltd, London, pp. 195.

Karr, C.L., Yakushin, I. and Nicolosi, K., 2000, “Solving inverse initial-value, boundary-value problems via genetic algorithm”, Engineering Applications of Artificial Intelligence 13, pp. 625-633.

Lima, Edgar B., Silva Neto, Antônio J. and Rodrigues, Pedro P. G. W., 2008, “Simulação bidimensional do escoamento e transporte no estuário do rio Macaé”.Encontro de Modelagem computacional, Volta Redonda, Brazil.

Melanie, Mitchell, 1999, “An Introduction to Genetic Algorithms”, A Bradford Book The MIT Press Cambridge, Massachusetts, London, England, Fifth printing.

Miranda, L. B., Castro, B.M. and Kjerfve, B., 2002, “Principles of Estuarine Physical Oceanography”. Editora Universidade de São Paulo, São Paulo. (In Portuguese).

Vaz, N., Diasa, J. M., Leitão,P. C. Rita Nolasco., 2007, “Application of the Mohid-2D model to a mesotidal temperate coastal lagoon”. Computers & Geosciences 33, pp.1204–1209.

Rosman, P.C. , 1989, “Circulation Models in Shallow Wa-ter Bodies, In R.C. Vieira da Silva (Ed.)”, Numerical Methods in Water Resources, ABRH, Rio de Janeiro, pp. 381 (In Portuguese).

Rodrigues, Pedro P. G. W., Lima, Edgar B., Silva Neto, Antônio J. and Lugon Junior, Jader, 2008, “Modelling and sensitivity analysis for estuarine hydrodynamics”, IX Simposio Internacional de Automatización, Havana, Cuba.

Versteeg, H.K. and Malalasekera.,W., 1995, “Computational Fluid Dynamics, An Introduction to Finite Volume Method”. Logman Group, London, pp. 257.

8. RESPONSIBILITY NOTICE

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