Geostatistical Assessment of Sewage Outfall Discharges

M. Monego[1], P. Ramos[1,2], M. V. Neves[1]
mdmonego@fe.up.pt; patricia@fe.up.pt; mjneves@fe.up.pt
[1]Faculty of Engineer of University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
[2]Institute of Accountancy and Administration of Porto, Dept. of Mathematics
R. Jaime Lopes Amorim, 4465-004 S. M. Infesta, Portugal
Phone: +351960084473, Fax: +351225081443

***Abstract*- Geostatistical techniques were used to study the spatial continuity of the salinity of a sewage outfall discharge. The data set used was obtained in a monitoring campaign to a sea outfall, located off the Portuguese west coast near Aveiro region, using an Autonomous Underwater Vehicle. The spatial analysis package geoR was used for computing variograms and ordinary kriging. The robust estimator was used to compute the experimental semivariogram, which was fitted to spherical theoretical model, after a successful cross-validation. The prediction maps of the spatial distribution of salinity show clearly the effluent plume dispersion in the field. This study suggests that geostatistical methodology can provide estimates of effluents dispersion very valuable for environmental impact and management of sea outfalls.**

I. INTRODUCTION

Outfalls are discharge points designed to promote the natural assimilative capacity of the oceans. The main advantages of marine outfalls are the natural dilution and dispersion of organic matter, pathogens and other pollutants, with minimal environmental impact. The impacts of discharged wastewaters on human beings may include direct contact and indirect effects.

Much effort has been devoted recently to improve means to monitor and characterize effluent plumes under a variety of oceanographic conditions. However, effluent plumes dispersion is still a difficult problem to study *in situ*, making reliable field measurements rare.

Autonomous Underwater Vehicles (AUVs) already demonstrated to be very appropriate for high-resolution surveys of small features such as outfall plumes [1]. In this paper we show the results of a monitoring campaign to S. Jacinto outfall, using an AUV, performed in order to study the shape and dispersion of the effluent plume.

Geostatistical modeling has been used with success to analyze, to characterize and to obtain information about the spatial distribution of the variables in study [2]-[9]. Geostatistical techniques were used to analyze the salinity measurements obtained in the vicinity of the outfall discharge. The data analysis was carried out using software R [10]. The spatial analysis package geoR was used for computing variograms and ordinary kriging [11]. In a first step the spatial structure of the observations was inspected thought a descriptive statistical analysis. The degree of spatial correlation among data in the study area as function of the distance and direction was expressed in terms of the Cressie and Hawkins robust estimator semivariogram. Cross-validation indicators and additional model parameters helped to choose the most appropriate models. Finally ordinary kriging was used to estimate salinity at unknown locations.

II. FUNDAMENTALS OF GEOSTATISTICS

Geostatistical methodology uses the semivariogram to quantify the spatial variation of the regionalized variable in study [12][13]. The semivariogram measures the mean variability between two data points as function of their distance.

The simplest and most commonly used estimator of the semivariogram is the method of moments, which is given by [14]

$$\gamma(h)=\frac{1}{2N(h)}\sum_{i=1}^{N(h)}\left[Z\left(x_i\right)-Z\left(x_i+h\right)\right]^2, \qquad (1)$$

where $\gamma(h)$ is the semivariogram, $Z\left(x_i\right)$ is the measured value at location x_i, h is the *lag* distance and $N\left(h\right)$ is the number of pairs of measurements which are $N\left(h\right)$ distance apart. The robust estimator of the semivariogram, developed by Cressie and Hawkins [15], is given by

$$\gamma(h)=\frac{\left[\frac{1}{N(h)}\sum_{i=1}^{N(h)}\left[Z\left(x_i\right)-Z\left(x_i+h\right)\right]^{\frac{1}{2}}\right]^4}{0.914+\frac{0.988}{N(h)}}, \qquad (2)$$

where $\gamma(h)$ is the semivariogram, $Z\left(x_i\right)$ is the measured value at location x_i, h is the *lag* distance and $N\left(h\right)$ is the number of pairs of measurements which are $N\left(h\right)$ distance apart. The denominator in eq. (2) is a bias correction [15].

As described in [12], this estimator should be robust to the presence of outliers and should greatly enhance the variogram continuity. It also has the advantage of not spreading the effect of outliers in computing the maps.

Once the experimental semivariogram is computed, the next step is to adjust it to a theoretical model. The most commonly used theoretical models are spherical, exponential and

Gaussian. These models give information about the structure of the spatial variation as well as the input parameters for the spatial prediction by kriging [13][16].

The prediction performances of the fitted variogram models can be evaluated and compared using the cross validation method. In the cross-validation procedure each sample is eliminated in turn and kriging is used to estimate its value with the remaining samples. The difference between the data value and the estimated value (error) is a measure of adequacy of the variogram model [17]. If the variogram model adequately describes the implicit spatial dependencies in the data set, then, the errors mean should be close to zero (indicating that the predictions are unbiased)

$$ME = \frac{1}{N}\sum_{i=1}^{N}\left[Z(x_i) - \hat{Z}(x_i)\right], \tag{3}$$

where $Z(x_i)$ is the data value at point x_i, $\hat{Z}(x_i)$ is the predicted value at point x_i, N is the number of measurements of the data set, and the standardized mean error should be close to 1

$$SME = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{Z(x_i) - \hat{Z}(x_i)}{\sigma(x_i)}\right], \tag{4}$$

where $\sigma(x_i)$ is the kriging standard deviation at point x_i.

After a variogram model is selected, kriging is applied to estimate the value of the variable in study at unsampled locations, using data from surrounded sampled points. The estimation is based on the semivariogram model and therefore takes it account the spatial variability of the variable in study.

The kriging method is a linear interpolator that belongs to the best linear unbiased estimator (BLUE) family [18]. The main purpose of kriging method is to estimate an unknown variable as a linear combination of known values:

$$\hat{Z}(x_0) = \sum_{i=1}^{M}\alpha_i Z(x_i), \tag{5}$$

where $\hat{Z}(x_0)$ is the estimated value at location x_0, M is the number of observations in the neighborhood of x_0 used in the estimative, $Z(x_i)$ is the measured value at location x_i that is in the neighborhood of x_0, and α_i are the correspondent weights.

In order to apply the ordinary kriging method, it is necessary to assume that the random function $Z(x_i)$ belongs to the stationary random functions family, which means that the mean values and the standard deviation of $Z(x_i)$ have to be independent of the location. Moreover, the unbiased condition over the weights is imposed [12]:

$$\sum_{i=1}^{M}\alpha_i = 1. \tag{6}$$

The kriging variance is minimized with the help of the Lagrange multipliers in order to impose the unbiased condition [12][16]:

$$L = E\left[\left(\hat{Z}(x_0) - Z(x_0)\right)^2\right] - 2\lambda\left(\sum_{i=1}^{M}\alpha_i - 1\right), \tag{7}$$

where λ is the Lagrange multiplier and $E\left[\left(\hat{Z}(x_0) - Z(x_0)\right)^2\right]$ is the kriging variance that can be expressed as a function of the semivariogram:

$$E\left[\left(\hat{Z}(x_0) - Z(x_0)\right)^2\right] = 2\sum_{i=1}^{M}\alpha_i\gamma(x_i, x_0) - \sum_{i=1}^{M}\sum_{j=1}^{M}\alpha_i\alpha_j\gamma(x_i, x_j). \tag{8}$$

After differencing eq. (7) with respect to α_i and λ, and equating to 0, the ordinary kriging equations are obtained [16]

$$\sum_{j=1}^{M}\alpha_j\gamma(x_i, x_j) - \lambda = \gamma(x_i, x_0), \quad i = 1, \ldots, M \tag{9}$$

The solution of this system of linear equations, with respect to α_i and λ, expressed in matrix notation is given by

$$\mathbf{X} = \mathbf{K}^{-1}\mathbf{M}, \tag{10}$$

where,

$$\mathbf{K} = \begin{bmatrix} \gamma(x_1, x_1) & \cdots & \gamma(x_1, x_M) & -1 \\ \gamma(x_2, x_1) & \cdots & \gamma(x_2, x_M) & -1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_M, x_1) & \cdots & \gamma(x_M, x_M) & -1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}$$
$$\mathbf{M} = \begin{bmatrix} \gamma(x_1, x_0) \\ \gamma(x_2, x_0) \\ \vdots \\ \gamma(x_1, x_N) \\ 1 \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_M \\ \lambda \end{bmatrix}, \tag{11}$$

being $\mathbf{K}$ the matrix of the semivariances of the observations at the neighborhood of x_0, $\mathbf{M}$ the vector of the semivariances of the neighbor observations with the location x_0, and $\mathbf{X}$ the vector of the weights α_i and the Lagrange multiplier λ.

III. GEOSTATISTICAL METHODS

A. Exploratory Analysis

The data set used in this analysis was obtained in a monitoring campaign to S. Jacinto outfall, located off the Portuguese west coast near Aveiro region (see Fig. 1), using the AUV of Underwater Systems and Technology Laboratory of University of Porto. A rectangular area of 200x100m^2 starting 20 m downstream from the middle point of the outfall diffuser was covered. As planned, the vehicle performed 6

horizontal trajectories at 2, 4, 6, 8, 10 and 12 m depth. In each horizontal section the vehicle described 6 parallel transects, perpendicular to the current direction, of 200 m long spaced of 20 m. While navigating at a constant velocity of approximately 2 knots (~1m/s), CTD (conductivity, temperature, depth) data were collected and recorded at a rate of 2.4 Hz. Consecutive measurements at horizontal sections were then distanced of about 0.4 m. In this study we analyze salinity data (computed from conductivity, temperature and depth) from the horizontal section at 2 m and at 4 m depth (see Fig. 2). Only about 5% of the raw data of salinity (computed from conductivity, temperature and depth), collected by the AUV at 2 m and 4 m depth, were considered due to interpolation conditions (see Fig. 3 and 4).

Fig. 1 Study area off the Portuguese west coast near Aveiro region.

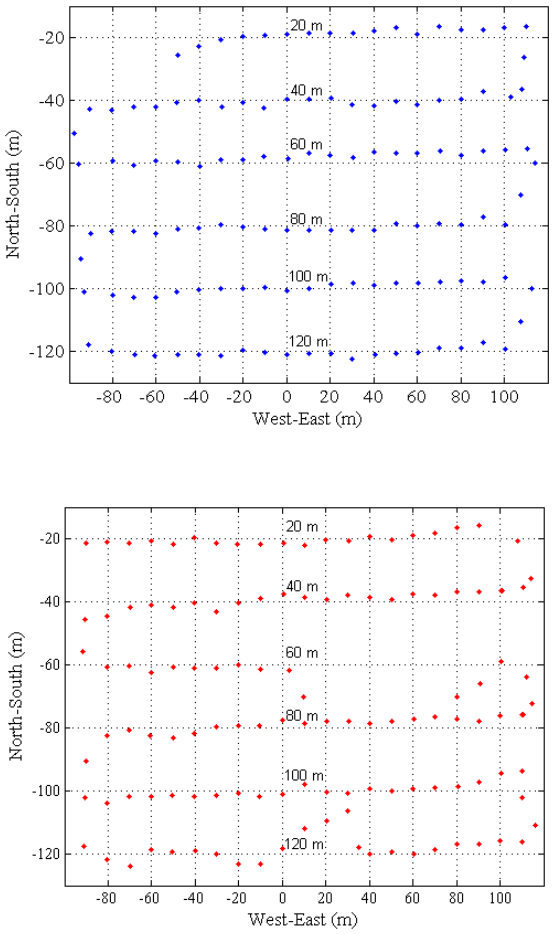

Fig. 2 Position of data sampled at 2 m (blue) and 4m (red) depth.

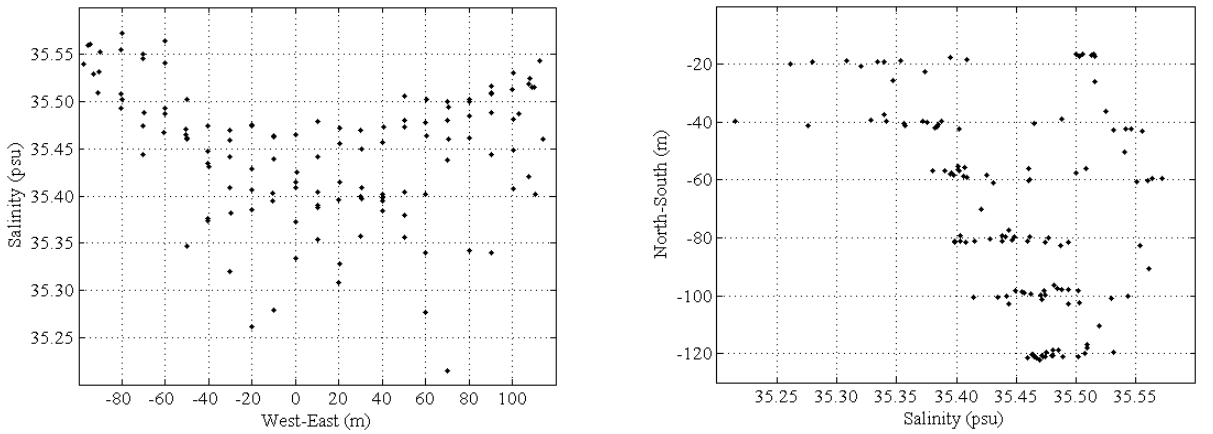

Fig. 3 Raw data of salinity measured by the AUV at 2 m depth.

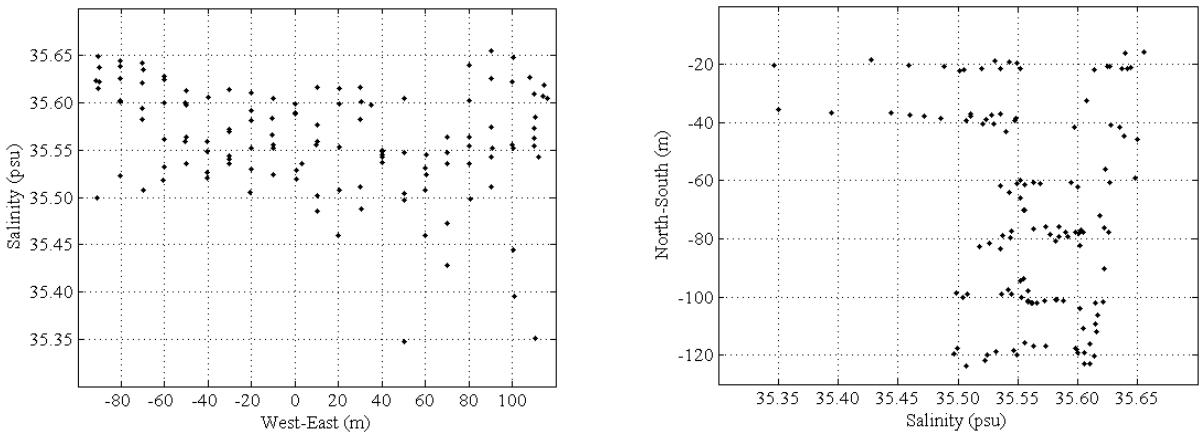

Fig. 4 Raw data of salinity measured by the AUV at 4 m depth.

For analyzing the spatial variability of the salinity data set, the classic statistical results were computed. The summary statistics of the salinity data set at 2 m and at 4 m depth is listed Table I.

The salinity at 2 m depth ranged from 35.2149 to 35.5723 psu. The mean value of the data set was 35.4469 psu and the median value was 35.4609 psu. The skewness and kurtosis values were respectively -0.6502 and 3.3742 indicating a lightly negatively skewed distribution.

The salinity at 4 m depth ranged from 35.3469 to 35.6554 psu. The mean value of the data set was 35.5608 psu and the median value was 35.5588 psu. The skewness and kurtosis values were high and respectively -1.0522 and 5.0757 indicating a negatively skewed distribution.

TABLE I
SUMMARY STATISTICS OF THE SALINITY DATA SET

Statistics	Section at 2 m depth	Section at 4 m depth
Number of data	126	126
Minimum	35.2149 psu	35.3469 psu
Mean	35.4469 psu	35.5608 psu
Median	35.4609 psu	35.5588 psu
Maximum	35.5723 psu	35.6554 psu
Variance	0.0048	0.0032
Stan. Dev.	0.0692	0.0566
Skewness	-0.6502	-1.0522
Kurtosis	3.3742	5.0757

Fig. 5 shows the histogram of the salinity data set of 2 m depth. The small left tail of the histogram shows a lightly negatively skewed distribution, which is in accordance with the negative value of the skewness parameter of Table I.

Fig. 6 shows the histogram of the salinity data set of 4 m depth. The left tail of the histogram shows a negatively skewed distribution, which is in accordance with the negative value of the skewness parameter of Table I.

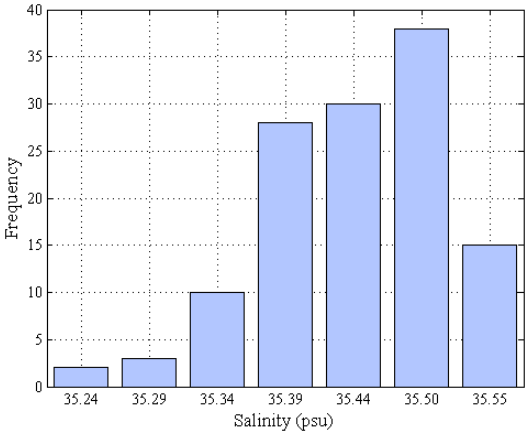

Fig. 5 Histogram of salinity at 2 m depth.

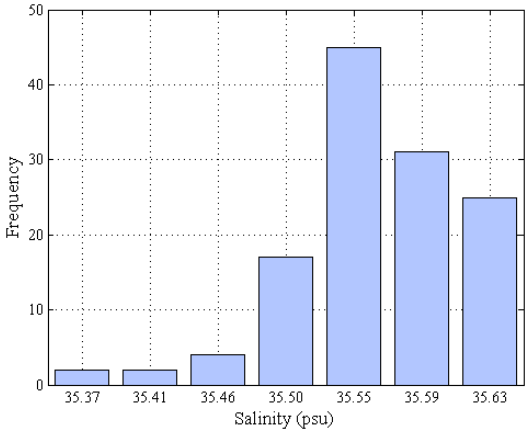

Fig. 6 Histogram of salinity at 4 m depth.

B. *Semivariogram*

Fig. 7 shows the omnidirectional experimental semivariogram of salinity of 2 m depth, computed using eq. 2, and the spherical model fitted. Fig. 8 shows the omnidirectional experimental semivariogram of salinity of 4 m depth using eq. 2, and the spherical model fitted.

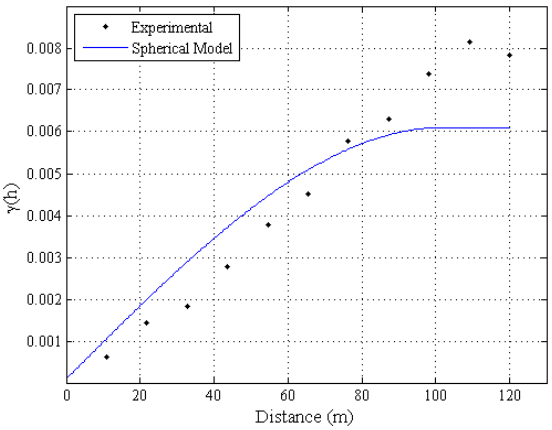

Fig. 7 Omnidirectional experimental semivariogram of 2 m depth and spherical model.

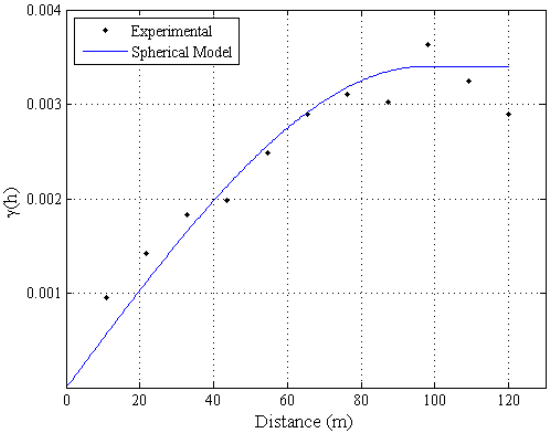

Fig. 8 Omnidirectional experimental semivariogram of 4 m depth and spherical model.

Estimation of semivariances was carried out using a lag distance of 12 m. In this study a range value of 120 m was considered. Anisotropy was investigated by calculating the semivariogram for several directions. However, no effect of anisotropy could be shown. The nugget, sill and range parameters of the fitted spherical models for the two depths are shown in Table II.

TABLE II
PARAMETERS OF THE SPHERICAL MODEL FITTED TO THE SEMIVARIOGRAM

Parameters	Section at 2 m depth	Section at 4 m depth
Nugget	0.0001	0
Sill	0.0061	0.0034
Range	101.89	97.30

C. *Cross-Validation*

Kriging cross-validation was used to evaluate the performance of the semivariogram model. The Table III shows the cross-validation parameters (computed according to eq. 3 and 4) relative to the two spherical models fitted. For a model that provides accurate predictions, the errors mean, ME should be close to zero and the standardized mean error, SME should be close to 1.

TABLE III
CROSS VALIDATION PARAMETERS OF THE SEMIVARIOGRAM MODELS

Parameters	Section at 2 m depth	Section at 4 m depth
ME	-0.000095	0.000584
SME	1.096	1.348

D. *Ordinary kriging*

Ordinary kriging was applied to estimate the value of salinity at unsampled locations using data from surrounded sampled points. The estimation was based on the spherical semivariogram models using the package geoR.

IV. RESULTS

The prediction map of salinity from the horizontal section of the 2 m depth, using the respective spherical semivariogram model, is shown in Fig. 9.

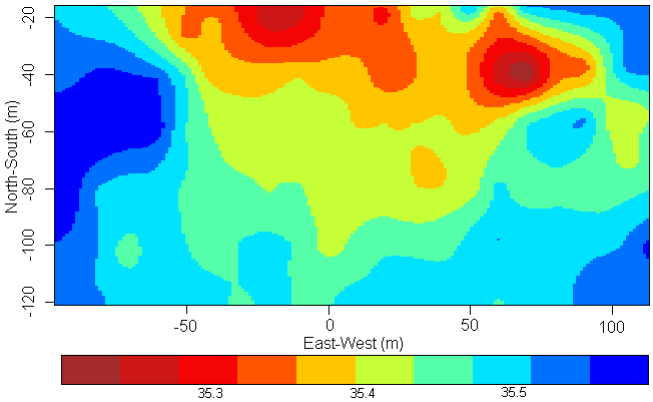

Fig. 9 Prediction map of salinity distribution at 2 m depth.

This map shows clearly the spatial variation of salinity in the studied area. From this map it is possible to identify unambiguously the effluent plume and its dispersion downstream in the north-south direction. The plume appears as a region of lower salinity compared with the surrounding ocean waters at the same depth. It is also possible to observe the plume edges since the wastefield width is shorter than the survey width. The plume exhibits a considerably more complex structure than the compact shape of the classical picture of the buoyant plume, but not so patchy as in previous studies, maybe due to the improvements in

sampling and also possibly due to the kriging successful results.

The prediction map of salinity from the horizontal section of the 4 m depth, using the respective spherical semivariogram model, is shown in Fig. 10. In this map, the patches that correspondent to the plume are smaller than those at the 2 m section, indicating that the plume was surfacing, spreading latterly.

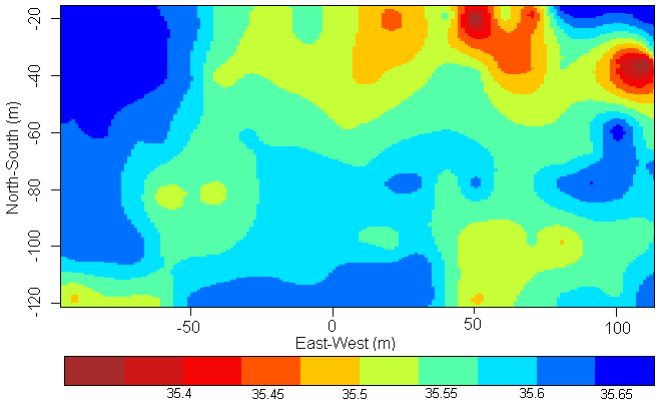

Fig. 10 Prediction map of salinity distribution at 4 m depth.

V. CONCLUSION

The salinity data used in this study were collected with an autonomous underwater vehicle in a monitoring campaign to an ocean outfall. After analyzing spatial distributions of salinity at 2 m and 4 m depth, geostatistical analysis was used to produce a prediction map of the sewage dispersion in the field. The spatial variability of the sampled data was analyzed previously, calculating the classic statistical parameters. The results indicated a negatively skewed distribution on both depths. Then, the robust estimator was used to compute the experimental semivariogram for several directions. No effect of anisotropy could be shown. The semivariogram was fitted to the spherical theoretical model, after a successful cross-validation. Finally, the predictions of salinity at unknown locations were obtained by ordinary kriging. The generated maps show clearly the spatial variation of salinity in the studied area, indicating that the effluent was surfacing, and far away from the beaches at 3 km distance.

Our study demonstrates that geostatistical analysis can provide good estimates of effluents dispersion, very valuable for environmental impact assessment and management of sea outfalls. Moreover, since accurate measurements of plume's dilution are rare, these studies might be very helpful in the future for validation of dispersion models.

ACKNOWLEDGMENT

The authors would like to thank the Underwater Systems and Technology Laboratory of University of Porto for the data set used in this analysis.

REFERENCES

[1] P.J.W. Roberts, "Environmental Hydraulics", V. P. Singh and W. H. Hager Edn., Kluwer Academic Press Publishers, Printed in Netherlands, Chapter Sea Outfalls, 1996, pp. 63-110.

[2] N. Saby, D. Arrouays, L. Boulonne, C. Jolivet, A. Pochot, "Geostatistical assessment of Pb in soil around Paris, France", Science of the Total Environment, 367, 2006, pp. 212-221.

[3] H. Wei, L. Dai, L. Wang, "Spatial distribution and risk assessment of radionuclides in soils around a coal-fired power plant: A case study from the city of Baoji, China", Environmental Research, 2007, 104, pp. 201-208.

[4] T. Shoji, "Statistical and geostatistical analysis of wind. A case study of direction statistics", Computers & Geocienses, 32, 2006, pp 1025-1039.

[5] T. Shoji, H. Kitaura, "Statistical and geostatistical analysis of rainfall in central Japan", Computers & Geosciences, 32, 2006, pp. 1007-1024.

[6] S. Caeiro, M. Painho, P. Goovaerts, H. Costa, S. Sousa, "Spatial sampling design for sediment quality assessment in estuaries", Environmental Modelling & Software, 18, 2003, pp. 853–859.

[7] C.J. Murray, H.J. Leeb, M.A. Hampton, "Geostatistical mapping of effluent-affected sediment distribution on the Palos Verdes shelf", Continental Shelf Research, 22, 2002, pp. 881–897.

[8] K.F. Poon, R.W. H. Wong, M.H.W. Lam, H.Y. Yeung, T.K.T. Chiu, "Geostatistical modelling of the spatial distribution of sewage pollution in coastal sediments", Water Research, 34, 2000, pp. 99-108.

[9] R. Orús, M. Hernández-Pajares, J.M. Juan,J. Sanz, "Improved of global ionospheric VTEC maps by using kriging interpolation technique", Journal of Atmospheric and Solar-Terrestrial Physics, 67, 2005, pp. 1598-1609.

[10] R Development Core Team, "R: a language and environment for statistical computing", http://www.R-project.org, 2004, R Foundation for Statistical Computing.

[11] P.J. Ribeiro Jr, P.J. Diggle, "geoR: a package for geostatistical analysis", R-News, 1, 2001, pp. 15–8.

[12] N. Cressie, "Statistics for spatial data", A Wiley Interscience Publication, New York, 1993, 900p.

[13] E.H. Isaaks, R.M. Srivastava, "Applied Geostatistics", Oxford University Press, New York, 1989, 561p.

[14] G. Matheron, "Les variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature", Paris. France: Masson; 1965, 305p.

[15] N. Cressie, D. Hawkins, "Robust estimation of the semivariogram", Mathematical Geology, 12, 1980, pp. 115-125.

[16] P. Kitanidis, "Introduction to geostatistics: Applications in hydrogeology", New York (USA), Cambridge University Press, 1997, 249p.

[17] H. Wackernagel, "Multivariate geostatistics: An introduction with applications", Berlin, Springer, 2003, 291p.

[18] J.P. Chilès, D. Pierre, "Geostatistics: Modeling Spacial Uncertainty", New York (USA), A Wiley Interscience Publication, 1999, 695p.