Geostatistical Mapping of Outfall Plume Dispersion Data Gathered with an Autonomous Underwater Vehicle

M. Monego[1,a], P. Ramos[1,2,b], M. V. Neves[1,c]

[1]*Faculty of Engineer of University of Porto*

Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

Tel.: +351 22 508 1860

Fax: + 351 22 508 1443

[2]*Institute of Accountancy and Administration of Porto, Dept. of Mathematics*

R. Jaime Lopes Amorim, 4465-004 S. M. Infesta, Portugal

Email: [a]*mdmonego@fe.up.pt;* [b]*patricia@fe.up.pt*; [c]*mjneves@fe.up.pt*

Abstract - The main purpose of this study was to examine the applicability of geostatistical modeling to obtain valuable information for assessing the environmental impact of sewage outfall discharges. The data set used was obtained in a monitoring campaign to *S. Jacinto* outfall, located off the Portuguese west coast near Aveiro region, using an AUV. The Matheron's classical estimator was used the compute the experimental semivariogram which was fitted to three theoretical models: spherical, exponential and gaussian. The cross-validation procedure suggested the best semivariogram model and ordinary kriging was used to obtain the predictions of salinity at unknown locations. The generated map shows clearly the plume dispersion in the studied area, indicating that the effluent does not reach the near by beaches. Our study suggests that an optimal design for the AUV sampling trajectory from a geostatistical prediction point of view, can help to compute more precise predictions and hence to quantify more accurately dilution. Moreover, since accurate measurements of plume's dilution are rare, these studies might be very helpful in the future for validation of dispersion models.

Keywords: ordinary kriging, outfalls, plumes dispersion, autonomous underwater vehicles

1. Introduction

Outfalls are designed to promote the natural assimilative capacity of the oceans to dispose of wastewaters with minimal environmental impact. This is accomplished through the vigorous initial mixing that is followed by oceanic dispersion within spatially and temporally varying currents. Usually, those mixing processes, in conjunction to bacterial mortality, result in rapid reductions in the concentrations of contaminants and organisms present in the wastewater to near background

levels. However, coastal physical, chemical and biological processes, very dynamic and complex, and intimately coupled to the concentration and content of wastewater, are in most instances, poorly understood. Consequently, how sewage disperses and how effluent modifies and is modified by coastal environment remain in many aspects unknown and unpredictable. The impacts of discharged wastewaters on human beings may include direct contact (e.g., by swimmers, surfers, beachgoers) with chemical contaminants or pathogens, and indirect effects through the consumption of contaminated food suppliers (e.g., fish, shellfish). Much effort has been devoted recently to improve means to monitor and characterize effluent plumes under a variety of oceanographic conditions, on relevant temporal and spatial scales. However, effluent plumes dispersion is still a difficult problem to study *in situ*. The difficulties in conducting field studies arise from the rapid spatial and temporal variations in physical, chemical and biological processes and oceanographic conditions that can occur in coastal waters. Additional logistical difficulties that include variability of discharge flowrate, high costs, and large area extent to be monitored, make then reliable field measurements of coastal outfall plumes rare. Autonomous Underwater Vehicles (AUVs) already demonstrated to be very appropriate for high-resolution surveys of small features such as outfall plumes (Ramos, 2005). Some of the advantages of these platforms include: easier field logistics, low cost per deployment, good spatial coverage and sampling over repeated sections and capability of feature-based or adaptive sampling. Demands for more reliable model predictions, and predictions of quantities that have received little attention in the past are now increasing. These are driven by increasing environmental awareness, more stringent environmental standards, and application of diffusion theory in new areas. While the gross properties of the plume can be reasonably predicted by the most commonly used marine discharge models, there remain many aspects which cannot be, particularly the patchy nature of the wastefield. This patchiness, which has been observed in field studies, is not incorporated into any of those models. They implicitly assume properties to vary smoothly in space, an assumption that is true only for time-averaged plumes. If we want to calibrate these models with real data we have to be able to quantify spatial correlations and other related characteristics.

In this paper we use geostatistics in the spatial analysis of environmental data gathered with an autonomous underwater vehicle (AUV) in a monitoring campaign to a sea outfall, aiming: (i) to distinguish the effluent plume from the receiving water; (ii) to estimate the salinity value at unknown locations and map its distribution by kriging interpolation, motivated by environmental impact assessment for decision-making and (iii) to validate predictions of plume dispersion models.

The geostatistical modeling has been used with success to analyze and characterize the spatial variability of soil properties (Saby *et al*., 2006; Wei *et al*., 2007), to obtain information for assessing water and wind resources (Shoji, 2006; Shoji and Kitaura, 2006), to design a sampling strategy for estuarine sediments (Caeiro *et al*., 2003), to study the thickness of effluent-affected sediment in the vicinity of a wastewater discharge (Murray *et al*., 2002), to obtain information about the spatial distribution of sewage pollution in coastal sediments (Poon *et al*., 2000), among many others.

Although very chaotic due to turbulent diffusion, plume's dispersion process tends to a natural variability mode when the plume stops rising and the intensity of turbulent fluctuations approaches to zero (Roberts, 1996). This region is called the end of the "near field" or "initial mixing region". After the end of the near field the established wastefield spreads laterally, drifting with the ocean current diffused by oceanic turbulence. In the near field the dilution increases rapidly with downstream distance, due to the turbulent kinetic energy generated by the buoyancy and momentum of the discharge. However, after the end of the near field the rate of increase of dilution is much lower. Dilution is then usually evaluated for risk assessment purposes at the end of the near field. It is likely that after the end of the near field pollutant concentrations are spatially correlated. In this case, geostatistics appears to be an appropriate technique to estimate dilution and map the plume dispersion.

In this work we conducted a geostatistical study of salinity measurements, obtained in the vicinity of an outfall discharge, using ordinary kriging interpolation. In a first step the spatial structure of the observations was inspected thought a descriptive statistical analysis. Then, the degree of spatial correlation among data in the study area as function of the distance and direction was expressed in terms of the semivariogram. Finally, ordinary kriging was used to

estimate salinity at unknown locations, and a map of this parameter distribution in the field was generated. Cross-validation indicators and additional model parameters helped to choose the most appropriate models.

2. Geostatistical 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, 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 analyse salinity data (computed from conductivity, temperature and depth) from the horizontal section at 2 m depth, where the effluent plume was found established and dispersing horizontally. The trajectory of the AUV at that section is shown in Figure 1.
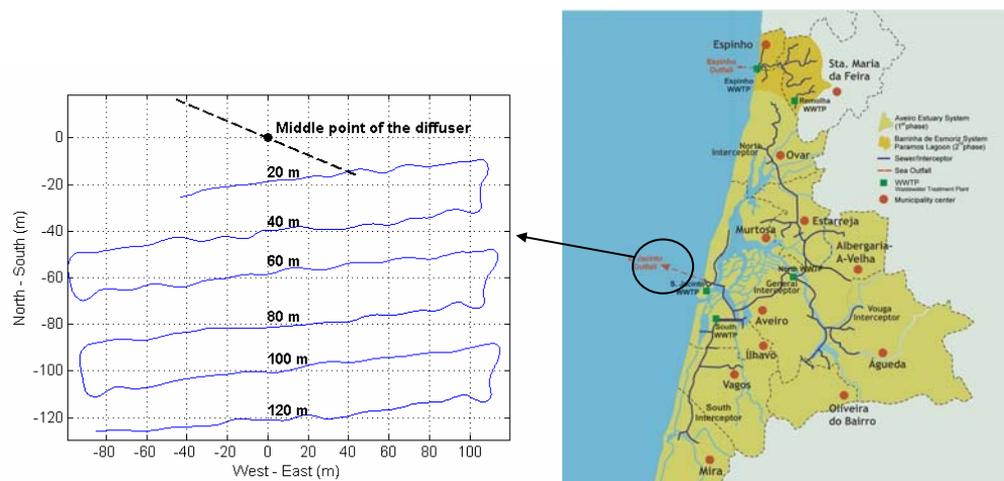

Figure 1 (a) AUV sampling trajectory at 2 meters depth; (b) Study area off the Portuguese west coast near Aveiro region.

4

2.1. Exploratory analysis

Table 1 gives the summary statistics of the salinity data set (2470 measurements). The salinity ranged from 35.152 to 35.607 psu. The mean value of the data set was 35.451 psu, being close to the median valued that was 35.463 psu. As in conventional statistics, a normal distribution for the variable under study is desirable in linear geostatistics (Wackernagel, 2003).

It can be seen from Table 1 that both skewness and kurtosis values are low indicating an approximated normal distribution of the raw data.

Table 1 Summary statistics of the salinity data set

Summary statistics of the salinity data set	
Number of data	2470
Minimum	35.152 psu
Mean	35.451 psu
Median	35.463 psu
Maximum	35.607 psu
Variance	0.004
Standard Deviation	0.067
Skewness	-0.52
Kurtosis	0.006

Figure 2 shows the frequency distribution of the salinity data set. The 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 1. This can be justified by the sampling strategy adopted, which was not completely succeeded. Since transects were all perpendicular to the current direction (and not parallel to the outfall diffuser), the ones closer to the diffuser still caught the plume ascending giving lower values of salinity.
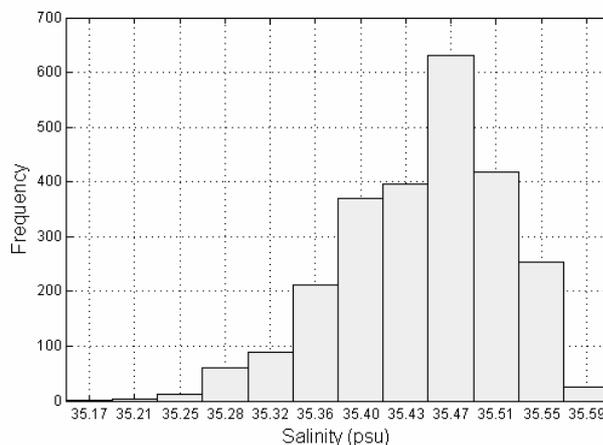

Figure 2 Frequency distribution of the salinity data set.

2.2. Semivariogram

Geostatistical methodology uses the semivariogram to quantify the spatial variation of the variable in study (Cressie, 1993; Isaaks and Srivastava, 1989). The semivariogram measures the mean variability between two data points as function of their distance. Matheron's classical estimator of the semivariogram was used in this study, whose computing equation is (Matheron, 1965):

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

where $\gamma(h)$ is the semivariogram, $Z(x_i)$ is the salinity value measured at location x_i, h is the lag distance and $N(h)$ is the number of pairs of measurements which are h distance apart. The experimental semivariogram is calculated for several lag distances.

Once the experimental semivariogram is computed, the next step is to adjust it to a theoretical model. This model gives information about the structure of the spatial variation as well as the input parameters for the spatial prediction by kriging. The most commonly used theoretical models are circular, spherical, exponential and gaussian (Kitanidis, 1997).

Figure 3 shows the omnidirectional experimental semivariogram and the fitted models.
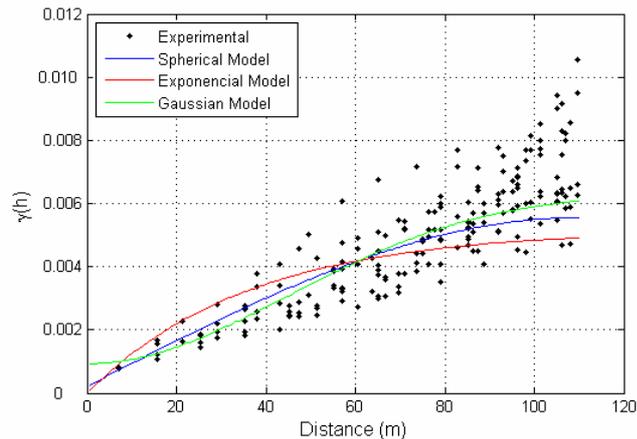

Figure 3 Omnidirectional experimental semivariogram and fitted models.

Estimation of semivariances was carried out using a lag distance of 10 m. As described in Soares (2000), the range value used for the computation of the experimental semivariogram should be less than the half major distance between the points of the data set. In this study a range value of 110 m was considered.

Anisotropy was investigated by calculation semivariogram for several directions. However, no effect of anisotropy could be shown. The semivariogram was fitted to three theoretical models: spherical, exponential and gaussian. The nugget, sill and range parameters of the fitted models are shown in Table 2.

Table 2 Parameters of the fitted semivariogram models.

Models	Nugget	Sill	Range	Nugget/Sill (%)
Spherical	0.00021	0.00555	109.772	3.9
Exponential	0	0.00492	109.772	0
Gaussian	0.00093	0.00608	109.772	15.3

The degree of spatial dependence of the variable in study can be evaluated through the nugget/sill ratio. According to Wei *et al*. (2007) the spatial dependence can be verified through of the nugget/sill ratio. Nugget/sill ratios less than 25% suggest that the variable has a strong spatial dependence; nugget/sill ratios between 25% and 75% suggest that the variable has a moderate spatial dependence; and nugget/sill ratios above 75% suggest that the variable has low spatial dependence.

As can be observed in Table 2, the nugget/sill ratios of salinity for all the semivariogram models are low and less then 25%, suggesting that this variable has a strong spatial dependence and that local variations could be captured as expected.

2.3. Cross-Validation

Cross-validation was used to compare the prediction performances of the three fitted semivariogram models. In the cross-validation procedure each sample is eliminated in turn and the remaining samples are used by kriging to predict the eliminated observation. The differences between the observations and the predictions are evaluated using the mean error (ME), the root mean squared error (RMSE), the root mean kriging variance (RMKV), the mean standardized error (MSE), and the root mean squared standardized error (RMSSE), computed respectively according to the following equations:

$$\mathrm{ME} = \frac{1}{N} \sum_{i=1}^{N} \left[\hat{Z}(x_i) - Z(x_i) \right] \qquad (2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left[\hat{Z}(x_i) - Z(x_i) \right]^2} \qquad (3)$$

$$RMKV = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \sigma^2(x_i)} \qquad (4)$$

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{\hat{Z}(x_i) - Z(x_i)}{\sigma^2(x_i)} \right] \qquad (5)$$

$$RMSSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left[\frac{\hat{Z}(x_i) - Z(x_i)}{\sigma^2(x_i)} \right]^2} \qquad (6)$$

where $\hat{Z}(x_i)$ is the predicted value at cross-validation point x_i, $Z(x_i)$ is the actual (measured) value at point x_i, N is the number of measurements of the data set, and $\sigma^2(x_i)$ is the kriging variance at cross-validation point x_i.

Table 3 shows these indicators for the spherical, exponential and gaussian models that helped to choose the best semivariogram model among these candidates.

Table 3 Cross-validation parameters for the fitted semivariogram models.

Models	ME	RMSE	RMKV	MSE	RMSSE
Spherical	-0.000038	0.01476	0.0183	-0.00242	0.8077
Exponential	0.0000029	0.01409	0.0086	0.000185	1.6310
Gaussian	-0.0002985	0.02495	0.0336	-0.010370	0.7461

For a model that provides accurate predictions, the ME should be close to zero indicating that the predictions are unbiased. The RMSE should be as small as possible, indicating that the predictions are close to the measured values. The RMKV should be naturally as small as possible. If the kriging variances are accurate then the RMSSE should be close to 1 (Wackernagel, 2003). If it is higher, the kriging prediction is too optimistic about the variability of the estimatives. The results given by Table 2 and Table 3 suggest that the spherical model should be use to estimate salinity over the studied area.

2.4. Ordinary Kriging

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

8

spatial variability of the variable in study. The kriging method belongs to the best linear unbiased estimators (BLUE) family.

The kriging estimator is said to be linear because the estimated value is a linear combination of the measurements, and is written in the form of:

$$\hat{Z}(x_0) = \sum_{i=1}^{M} \alpha_i Z(x_i) \qquad (7)$$

where $\hat{Z}(x_0)$ is the estimated value for location x_0, M is the number of observations in the neighborhood of x_0 used in the estimative, and α_i are the correspondent weights. Ordinary kriging is used when the mean value of the variable in study is unknown. For this estimator to be unbiased, for any value of the mean, it is required that $\sum_{i=1}^{M} \alpha_i = 1$. The estimated value is obtained by minimizing the kriging variance with the help of the Lagrange multipliers in order to impose the unbiased condition (Cressie, 1993; Kitanidis, 1997).

3. Results

The kriged maps of salinity from the horizontal section at 2 m depth using the spherical, exponential and gaussian semivariogram models are shown in Figure 4. All the maps show clearly the spatial variation of salinity in the studied area. From these maps it is possible to identify unambiguously the effluent plume and its dispersion downstream in the north-south direction. It 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.

We may say that the results obtained with the three semivariogram models are quite similar. However, in the prediction using the gaussian model same local variations were not capturated. Salinity differences compared with the surrounding waters at 2 m depth started to be about 0.455 psu in the first two transects (20 and 40 m), decreasing to about 0.293 psu in the third transect (60 m), to about 0.215 psu in the forth transect (80 m), to about 0.176 psu in the fifth transect (100 m), ending almost equally to background waters at 120 m distance with a difference of about 0.071 psu.

Washburn *et al.* (1992) observed salinity differences compared with the surrounding waters of the order of 0.1 psu, while Petrenko *et al.* (1998) found differences of the order of 0.2 psu.
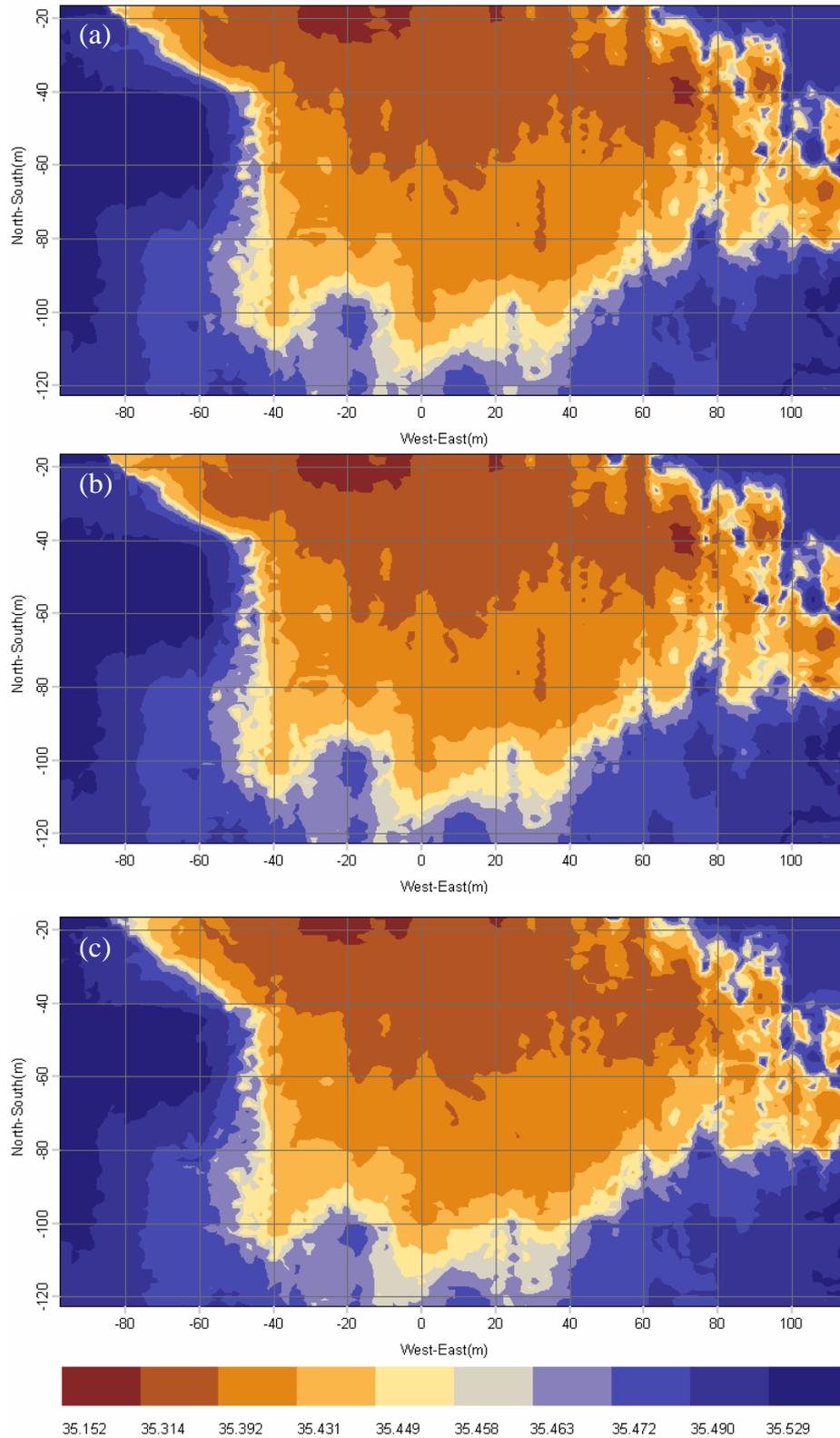

Figure 4 Prediction maps of salinity distribution using the: (a) spherical model, (b) exponential model, (c) gaussian model.

A sharp difference in salinity at the effluent plume lateral edges is clearly visible, being the wastefield spreading almost centered in the survey area. This indicates that the sampling strategy designed was successful even for a surfacing plume which can be considered as the most complicated case in terms of natural tracer tracking.

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 horizontal resolution and also possibly to the kriging successful results.

3. Conclusions

Geostatistical analysis of salinity obtained with an autonomous underwater vehicle in a monitoring campaign to an ocean outfall was able to produce a kriged map of the sewage dispersion in the field. The spatial variability of the sampled data was analysed previously calculating the classic statistical indicators. The results indicated an approximated normal distribution of the data samples which is desirable. Then, the Matheron's classical estimator was used the compute the experimental semivariogram for several directions. No effect of anisotropy could be shown. The semivariogram was fitted to three theoretical models: spherical, exponential and gaussian. The cross-validation indicators for the spherical, exponential and gaussian models suggested the best semivariogram model among these candidates. Finally, the predictions of salinity at unknown locations were obtained by ordinary kriging. The generated map shows clearly the spatial variation of salinity in the studied area, indicating that the effluent does not reach the near by beaches distanced of about 3km.

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.

Acknowledgments

References

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

Cressie, N. (1993). Statistics for spatial data. A Wiley Interscience Publication, New York, 900p.

Isaaks, E.H., Srivastava, R.M., (1989). Applied Geostatistics. Oxford University Press, New York, 561p.

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

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

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

Petrenko, A. A., Jones, B. H., Dickey, T. D. (1998). Shape and Initial Dilution of Sand Island, Hawaii Sewage Plume. Journal of Hydraulic Engineering, ASCE, 124, pp. 565-571.

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

Ramos, P. (2005). Advanced Mathematical Modeling for Outfall Plume Tracking and Management using Autonomous Underwater Vehicles based Systems. PhD Thesis, Faculty of Engineer, University of Porto.

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

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

Soares, A. (2000). Geoestatística para as Ciências da Terra e do Ambiente, IST Press, Lisboa, 206. ISBN: 972-8469-10-1.

Shoji, T. (2006). Statistical and geostatistical analysis of wind. A case study of direction statistics. Computers & Geosciences, 32, pp 1025-1039.

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

Wackernagel, H. (2003). Multivariate geostatistics: An introduction with applications. Berlin, Springer, 291p.

Washburn, L., Jones, B. H., Bratkovich, A., Dickey, T. D., Chen, M., (1992). Mixing, Dispersion, and Resuspension in Vicinity of Ocean Wastewater Plume. Journal of Hydraulic Engineering, ASCE, 118, pp. 38-58.

Wei, H., Dai, L., Wang, L. (2007). 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, 104, pp. 201-208.