ANTITATIVE GEOLOGY AND GEOSTATISTICS

. Atkinson · C.D. Lloyd (Eds.)

oENV VII – Geostatistics for Environmental Applications

volume brings together selected contributions from geoENV 2008, the 7th International ference on Geostatistics for Environmental Applications, held in Southampton, UK, in mber 2008. This book presents the state-of-the-art in geostatistics for the environmental ces. It includes a wide range of methodological advances and applications. It offers ht and guidance for researchers, professionals, graduate students and others seeking mation on the latest perspectives in the field. The rich body of applications will enable new to geostatistics to assess the utility of the methods for their own applications. book includes 35 chapters on topics as diverse as methodological developments, applica- in the soil sciences, climatology, pollution, health, wildlife mapping, fisheries and remote ng, amongst other areas. With its focus on environmental applications of geostatistics, r than the more traditional geostatistical remit of mining and petroleum exploration, book is part of a series that presents an invaluable resource. This book will be a first port l for those who wish to apply geostatistical methods in the environmental sciences.

nce

archers, scientists, professionals and graduate students in geosciences geostatistics, spatial tics, environmental science and engineering, ecology, oceanography, climatology, logy, soil and forestry science.

N 978-90-481-2321-6

QGAG 16

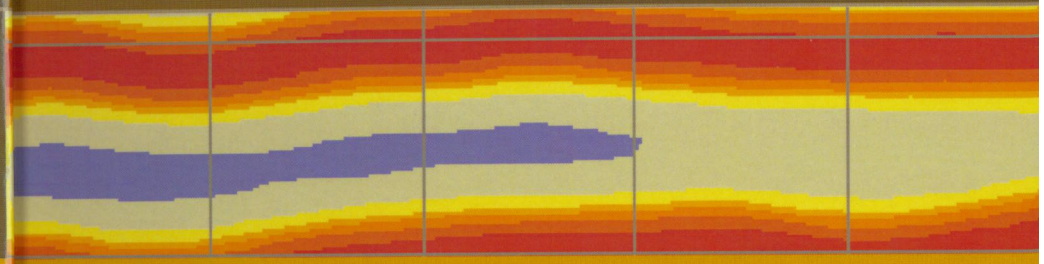
P.M. Atkinson · C.D. Lloyd (Eds.)

geoENV VII – Geostatistics for Environmental Applications

P.M. Atkinson · C.D. Lloyd (Eds.)

geoENV VII – Geostatistics for Environmental Applications

QUANTITATIVE GEOLOGY AND GEOSTATISTICS



P.M. Atkinson · C.D. Lloyd (Eds.)

geoENV VII – Geostatistics for Environmental Applications

Monestiez P et al. (eds) (2001) GeoENVIII – geostatistics for environmental applications. Kluwer, Dordrecht
Nunes C, Soares A (2004) Geostatistical space-time simulation model for characterization of air quality. In: Sanchez-Villa X et al. (eds) GeoENV IV – geostatistics for environmental applications. Kluwer, Dordrecht, pp 103–114
Pereira M, Soares A, Branquinho C (1997) Stochastic simulation of fugitive dust emissions. In: Baafi EY et al. (eds) Geostatistics Wollongong'96, vol 2. Kluwer, Dordrecht, pp 1055–1065
Pereira MJ, Almeida J, Costa C, Soares A (2001) Accounting for soft information in mapping soil contamination with TPH at an oil storage site. In: Monestiez P et al. (eds) GeoENVIII – geostatistics for environmental applications. Kluwer, Dordrecht, pp 475–486
Renard P et al. (eds) (2005) Geostatistics for environmental applications. Springer, Berlin
Rouhani S et al. (ed) (1996) Environmental and geotechnical applications. STP 1283, ASTM Pub
Russo A, Soares A, Trigo R (2008) Geostatistical model for air quality surveillance/monitoring system. GeoENV2008. Porthmouth, UK
Sanchez-Villa X et al. (eds) (2004) GeoENV IV – geostatistics for environmental applications. Kluwer, Dordrecht
Serre M, Christakos G, Lee S (2003) Soft data space-timemaping of coarse particulate matter annual arithmetic average over the U.S. In: Sanchez-Villa X et al. (eds) GeoENV IV – geostatistics for environmental applications. Kluwer, Dordrecht, pp 115–126
Soares A (ed) (1993) Geostatistics TROIA'92, vols 1 and 2. Kluwer, Dordrecht
Soares A (2001) Direct Sequential simulation and co-simulation. Math Geol 33(8):911–926
Soares A et al. (eds) (1997) GeoENVI – geostatistics for environmental applications. Kluwer, Dordrecht
Verly G et al. (eds) (1984) Geostatistics for natural resources characterization, vols 1 and 2, D. Reidel Pub., Dordrecht

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

Maurici Monego, Patrícia Ramos, and Mário V. Neves

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 nearby 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.

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,

M. Monego (✉), P. Ramos, and M.V. Neves
Faculty of Engineering of University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
e-mail: mdmonego@fe.up.pt; patricia@fe.up.pt; mjneves@fe.up.pt

P. Ramos
Institute of Accountancy and Administration of Porto, Department of Mathematics,
R. Jaime Lopes Amorim, 4465-004 S. M. Infesta, Portugal
e-mail: patricia@fe.up.pt

P.M. Atkinson and C.D. Lloyd (eds.), *geoENV VII – Geostatistics for Environmental*

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 environments 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 the means to monitor and characterize effluent plumes under a variety of oceanographic conditions, on relevant temporal and spatial scales. However, effluent plume 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 reliable field measurements of coastal outfall plumes rare.

Autonomous Underwater Vehicles (AUVs) have already been demonstrated to be 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, 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 targeted 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.

Geostatistical modeling has been used 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 sampling strategies for estuarine sediment collection (Caeiro et al., 2003),

to study the thickness of effluent-affected sediment in the vicinity of wastewater discharges (Murray et al., 2002), and 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 "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 sense, geostatistics appears to be an appropriate technique to estimate dilution and map the plume dispersion.

In this work we conduct 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 model.

2 Geostatistical Analysis

The data set used in this analysis was obtained in a monitoring campaign of *S. Jacinto* outfall, located off the Portuguese west coast near *Aveiro* region, using the AUV of the Underwater Systems and Technology Laboratory of University of Porto. A rectangular area of $200 \times 100\ m^2$ starting 20 m downstream from the middle point of the outfall diffuser was covered. As planned, the vehicle performed six horizontal trajectories at 2, 4, 6, 8, 10 and 12 m depth. In each horizontal section the vehicle described six parallel transects, perpendicular to the current direction, of 200 m length and spaced at 20 m. While navigating at a constant velocity of approximately 1 m/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 at 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 this section is shown in Fig. 1.

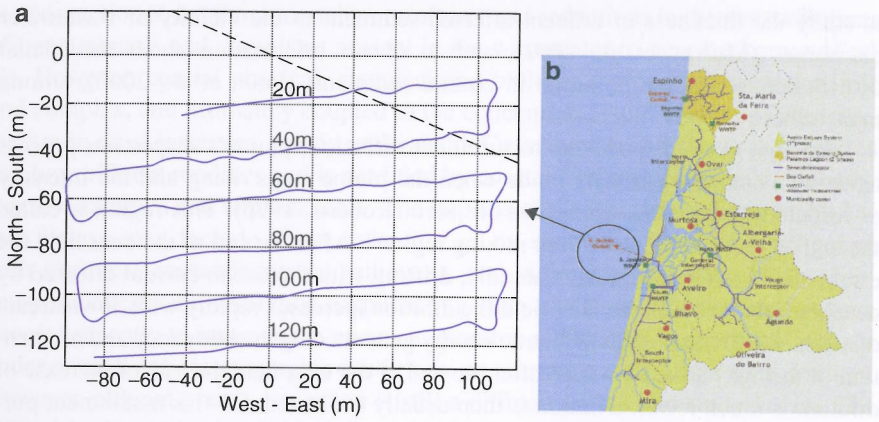


Fig. 1 (**a**) AUV sampling trajectory at 2 m depth. (**b**) Study area off the Portuguese west coast near *Aveiro* region

2.1 *Exploratory Analysis*

Table 1 gives the summary statistics of the salinity data set (2,470 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 value 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.

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 in Table 1. This can be justified by the sampling strategy adopted. 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 much lower values of salinity.

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 a function of their distance. Matheron's classical estimator of the semivariogram was used in this study, whose computing equation is (Matheron, 1965):

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \mathbf{h})]^2 \qquad (1)$$

Table 1 Summary statistics of the salinity data set

Summary statistics of the salinity data set	
Number of data	2,470
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

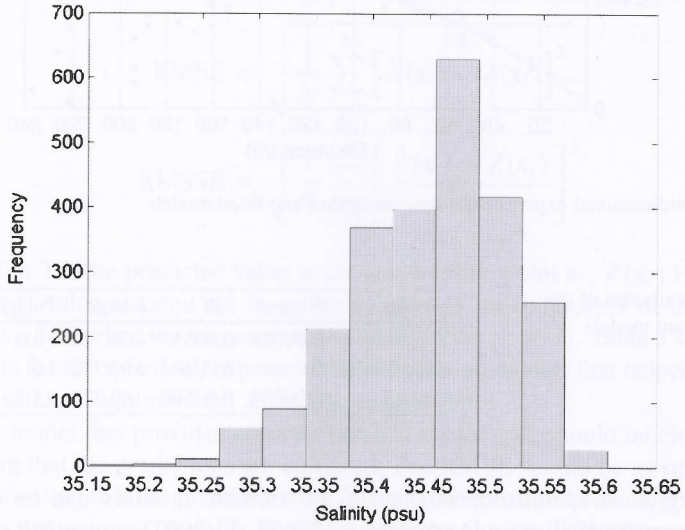


Fig. 2 Frequency distribution of the salinity data set

where $\gamma(\mathbf{h})$ is the semivariogram, $Z(\mathbf{x}_i)$ is the salinity value measured at location $\mathbf{x}_i$, $\mathbf{h}$ is the lag distance and $N(\mathbf{h})$ is the number of pairs of measurements which are $\mathbf{h}$ distance apart. The experimental semivariogram is calculated for several lag distances. Once the experimental semivariogram is computed, the next step is to fit it into a theoretical model. This model gives information about the structure of the spatial variation being also used 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 of the salinity data set and the models spherical, exponential and Gaussian fitted.

Estimation of semivariances was carried out using a lag distance of 10 m. Anisotropy was investigated by calculation of semivariogram for several directions. However, no effect of anisotropy could be shown. The nugget, sill, and range parameters of the three fitted models are shown in Table 2.

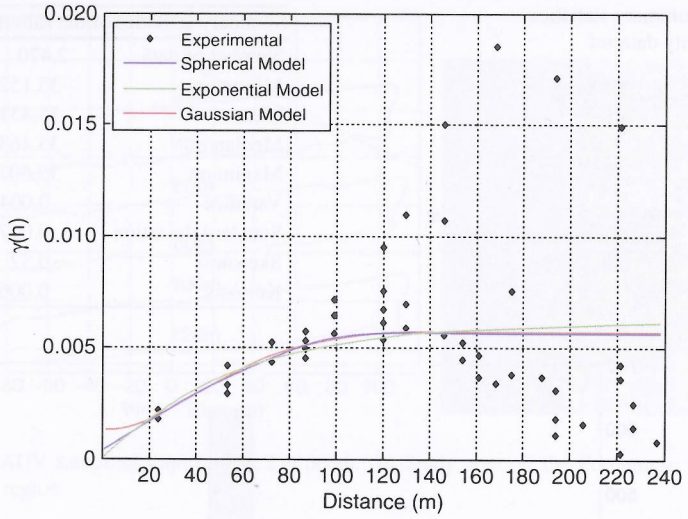


Fig. 3 Omnidirectional experimental semivariogram and fitted models

Table 2 Parameters of the 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), 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 the variable has a strong spatial dependence and that probably local variations could be captured, as expected.

2.3 *Cross-Validation*

Cross-validation was used to compare the prediction performances of the three semivariogram models. In this 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 then evaluated using the mean error (ME), the root mean squared error (RMSE), and the root mean

Table 3 Cross-validation parameters for the semivariogram models

Models	ME	RMSE	RMSSE
Spherical	-3.8×10^{-5}	0.01476	0.8077
Exponential	0.29×10^{-5}	0.01409	1.6310
Gaussian	-29.9×10^{-5}	0.02495	0.7461

squared standardized error (RMSSE), computed respectively according to the following equations:

$$\mathrm{ME} = \frac{1}{N}\sum_{i=1}^{N}\left[\hat{Z}(\mathbf{x}_i) - Z(\mathbf{x}_i)\right] \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\hat{Z}(\mathbf{x}_i) - Z(\mathbf{x}_i)\right]^2} \tag{3}$$

$$\mathrm{RMSSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\hat{Z}(\mathbf{x}_i) - Z(\mathbf{x}_i)}{\sigma^2(\mathbf{x}_i)}\right]^2} \tag{4}$$

where $\hat{Z}(\mathbf{x}_i)$ is the predicted value at cross-validation point $\mathbf{x}_i$, $Z(\mathbf{x}_i)$ is the actual (measured) value at point $\mathbf{x}_i$, N is the number of measurements of the data set, and $\sigma^2(\mathbf{x}_i)$ is the kriging variance at cross-validation point $\mathbf{x}_i$. Table 3 shows these indicators for the spherical, exponential and Gaussian models that helped to choose the best semivariogram model among these candidates.

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. If the kriging variances are accurate, then the RMSSE should be close to 1 (Wackernagel, 2003). If it is higher, the kriging predictions are too optimistic about the variability of the estimates. The results given by Table 2 and Table 3 suggest that the spherical model should be used to estimate salinity over the studied area.

2.4 *Ordinary Kriging*

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

The kriging method belongs to the best linear unbiased estimators (BLUE) family. It is said to be linear because the estimated value is a linear combination of the measurements, being written in the form of:

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^{M}\alpha_i Z(\mathbf{x}_i) \tag{5}$$

where $\hat{Z}(\mathbf{x}_0)$ is the estimated value for location $\mathbf{x}_0$, M is the number of observations in the neighborhood of $\mathbf{x}_0$ 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 of the horizontal section at 2 m depth using the spherical, exponential and Gaussian models are shown in Fig. 4. All 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 current direction. It appears as a region of lower salinity compared to 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 some small local variations were not captured. Figure 5 shows the prediction error map when using the spherical model. It can be seen, as expected, that the prediction error is smaller the closer the prediction to the trajectory of the vehicle.

Salinity differences compared to 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.

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 increase in horizontal resolution and also possibly due to the kriging results.

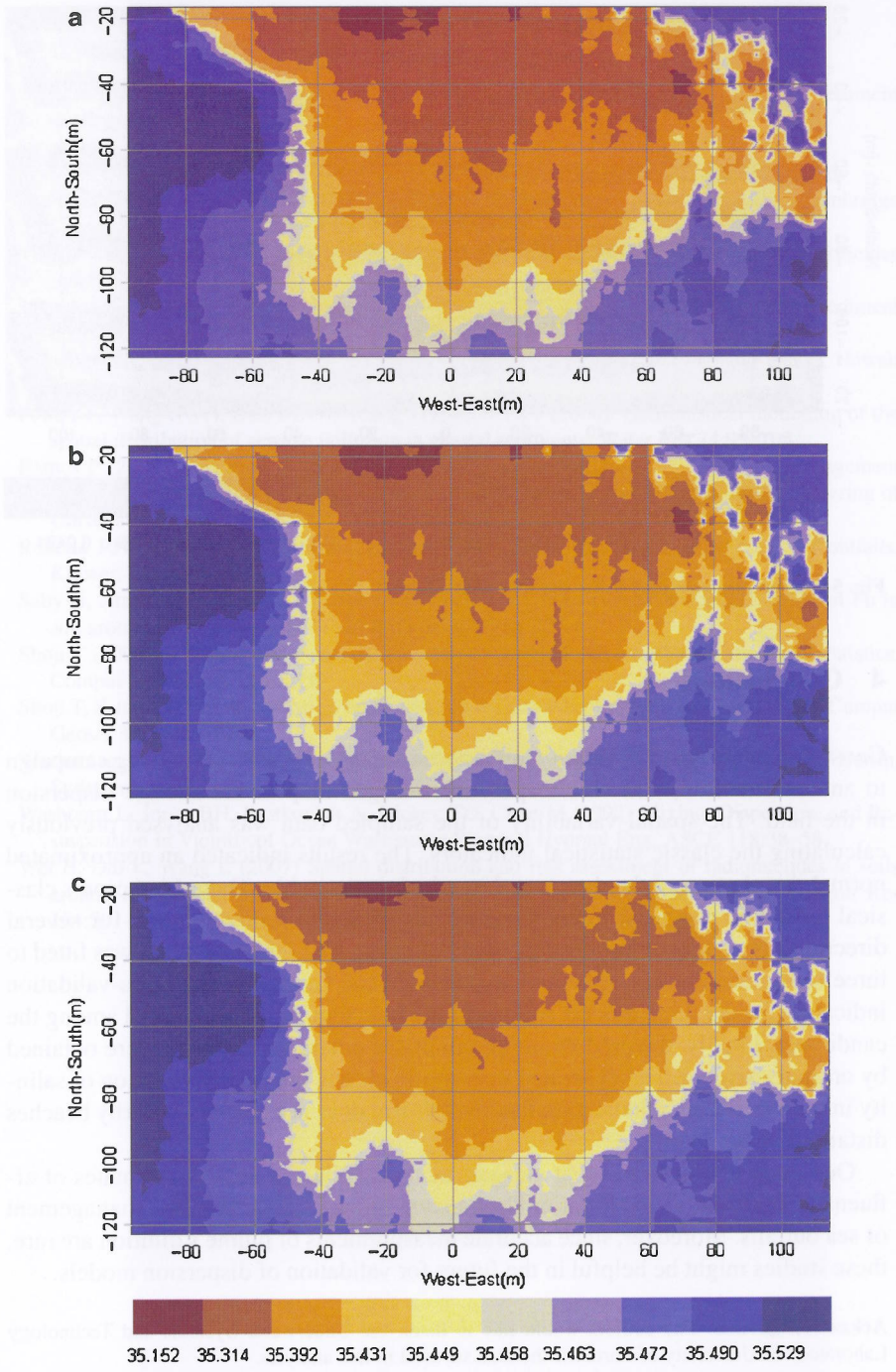


Fig. 4 Prediction maps of salinity distribution using the: (**a**) spherical model. (**b**) Exponential model. (**c**) Gaussian model

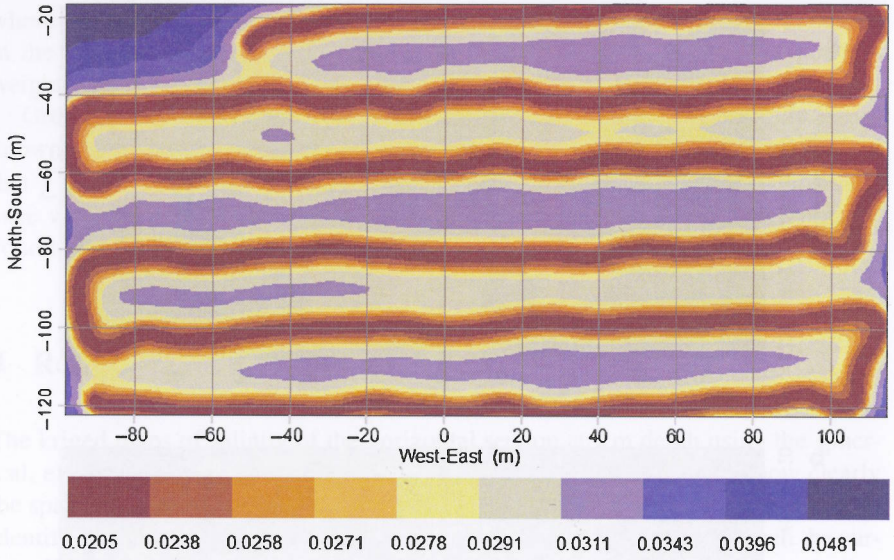


Fig. 5 Prediction error map using the spherical model

4 Conclusions

Geostatistical analysis of salinity, obtained with an AUV 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, Matheron's classical estimator was used to 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 three models suggested the best semivariogram model among the 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 nearby beaches distanced about 3 km.

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

Acknowledgments 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

Caeiro S, Painho M, Goovaerts P, Costa H, Sousa S (2003) Spatial sampling design for sediment quality assessment in estuaries. Environ Model Softw 18:853–859

Cressie N (1993) Statistics for spatial data. Wiley Interscience, New York

Isaaks EH, Srivastava RM (1989) Applied geostatistics. Oxford University Press, New York

Kitanidis P (1997) Introduction to geostatistics: applications in hydrogeology. Cambridge University Press, New York

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

Murray CJ, Leeb HJ, Hampton MA (2002) Geostatistical mapping of effluent-affected sediment distribution on the Palos Verdes shelf. Cont Shelf Res 22:881–897

Petrenko AA, Jones BH, Dickey TD (1998) Shape and Initial Dilution of Sand Island, Hawaii Sewage Plume. J Hydraul Eng, ASCE 124:565–571

Poon KF, Wong RWH, Lam MHW, Yeung HY, Chiu TKT (2000). Geostatistical modelling of the spatial distribution of sewage pollution in coastal sediments. Water Res 34:99–108

Ramos P (2005) Advanced mathematical modeling for outfall plume tracking and management using autonomous underwater vehicles based systems. Ph.D. thesis, Faculty of Engineering of University of Porto

Roberts PJW (1996) Environmental hydraulics. In: Singh VP, Hager WH (eds) Sea outfalls. Kluwer, The Netherlands

Saby N, Arrouays D, Boulonne L, Jolivet C, Pochot A (2006) Geostatistical assessment of Pb in soil around Paris, France. Sci Total Environ 367:212–221

Shoji T (2006) Statistical and geostatistical analysis of wind. A case study of direction statistics. Comput Geosci 32:1025–1039

Shoji T, Kitaura H (2006) Statistical and geostatistical analysis of rainfall in central Japan. Comput Geosci 32:1007–1024

Wackernagel H (2003) Multivariate geostatistics: an introduction with applications. Berlin, Springer

Washburn L, Jones BH, Bratkovich A, Dickey TD, Chen M (1992) Mixing, Dispersion, and Resuspension in Vicinity of Ocean Wastewater Plume. J Hydraul Eng, ASCE 118:38–58

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. Environ Res 104:201–208