

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Mapping and Dilution Estimation of Wastewater Discharges Based on Geostatistics Using an Autonomous Underwater Vehicle

Patrícia Ramos¹ and Nuno Abreu²

¹*Institute for Systems and Computer Engineering of Porto (INESC Porto)
School of Accounting and Administration of Porto – Polytechnic Institute of Porto*

²*Institute for Systems and Computer Engineering of Porto (INESC Porto)
Faculty of Engineering of University of Porto
Portugal*

1. Introduction

Wastewaters are often discharged into coastal waters through outfall diffusers that efficiently dilute effluent and usually restrict any environmental impact within a small area. However, predicting this impact is difficult because of the complexity of the hydrodynamic processes that mix the wastewater and also because of the variability in oceanic conditions. Despite great improvements over the years in the understanding of these mixing processes, since models are now available that can make reasonable predictions under steady-state conditions (Hunt et al., 2010), many aspects remain unknown and unpredictable. For this reason, much effort has been recently devoted to improve ways of monitoring and characterizing sewage plumes under a variety of oceanographic conditions.

1.1 MARES AUV

Autonomous Underwater Vehicles (AUVs) have been used efficiently in a wide range of applications. They were first developed with military applications in mind, for example for mine hunting missions. Later on, scientists realized their true potential and started to use them as mobile sensors, taking measurements in difficult scenarios and at a reasonable cost (Bellingham, 1997; Bellingham et al., 1992; Fernandes et al., 2000; Nadis, 1997; Robinson et al., 1999; Yu et al., 2002). MARES (Modular Autonomous Robot for Environment Sampling) AUV has been successfully used to monitor sea outfalls discharges (Abreu et al., 2010; Abreu & Ramos, 2010; Ramos & Abreu, 2010; 2011a;b;c) (see Fig. 1). MARES is 1.5 m long, has a diameter of 8-inch and weighs about 40 kg in air. It features a plastic hull with a dry mid body (for electronics and batteries) and additional rings to accommodate sensors and actuators. Its modular structure simplifies the system's development (the case of adding sensors, for example). It is propelled by two horizontal thrusters located at the rear and two vertical thrusters, one at the front and the other at the rear. This configuration allows for small operational speeds and high maneuverability, including pure vertical motions. It is equipped with an omnidirectional acoustic transducer and an electronic system that allows for long baseline navigation. The vehicle can be programmed to follow predefined trajectories while

collecting relevant data using the onboard sensors. A Sea-Bird Electronics 49 FastCAT CTD had already been installed onboard the MARES AUV to measure conductivity, temperature and depth. MARES' missions for environmental monitoring of wastewater discharges are conducted using a GUI software that fully automates the operational procedures of the campaign (Abreu et al., 2010). By providing visual and audio information, this software guides the user through a series of steps which include: (1) real time data acquisition from CTD and ADCP sensors, (2) effluent plume parameter modeling using the CTD and ADCP data collected, (3) automatic path creation using the plume model parameters, (4) acoustic buoys and vehicle deployment, (5) automatic acoustic network setup and (6) real time tracking of the AUV mission.



Fig. 1. Autonomous Underwater Vehicle MARES.

1.2 Data processing

Data processing is the last step of a sewage outfall discharge monitoring campaign. This processing involves the ability to extrapolate from monitoring samples to unsampled locations. Although very chaotic due to turbulent diffusion, the effluent's dispersion process tends to a natural variability mode when the plume stops rising and the intensity of turbulent fluctuations approaches to zero (Hunt et al., 2010). It is likely that after this point the pollutant substances are spatially correlated. In this case, geostatistics appears to be an appropriate technique to model the spatial distribution of the effluent. In fact, geostatistics has been used with success to analyze and characterize the spatial variability of soil properties, to obtain information for assessing water and wind resources, to design sampling strategies for monitoring estuarine sediments, to study the thickness of effluent-affected sediment in the vicinity of wastewater discharges, to obtain information about the spatial distribution of sewage pollution in coastal sediments, among others. As well as giving the estimated values, geostatistics provides a measure of the accuracy of the estimate in the form of the kriging variance. This is one of the advantages of geostatistics over traditional methods of assessing pollution. In this work ordinary block kriging is used to model and map the spatial distribution of temperature and salinity measurements gathered by an AUV on a Portuguese sea outfall monitoring campaign. The aim is to distinguish the effluent plume from the receiving waters, characterize its spatial variability in the vicinity of the discharge and estimate dilution.

2. Geostatistical analysis

2.1 Stationary random function models

The most widely used geostatistical estimation procedures use stationary random function models. A random function is a set of random variables that have some spatial locations and whose dependence on each other is specified by some probabilistic mechanism. A random function is stationary if all the random variables have the same probability distribution and if any pair of random variables has a joint probability distribution that depends only on the separation between the two points and not on their locations. If the random function is stationary, then the expected value and the variance can be used to summarize the univariate behavior of the set of random variables. The parameter that is commonly used to summarize the bivariate behavior of a stationary random function is its covariance function, its correlogram, and its variogram. The complete definition of the probabilistic generating mechanism of a random function is usually difficult even in one dimension. Fortunately, for many of the problems we typically encounter, we do not need to know the probabilistic generating mechanism. We usually adopt a stationary random function as our model and specify only its covariance or variogram (Isaaks & Srivastava, 1989; Kitanidis, 1997; Wackernagel, 2003).

2.2 Ordinary kriging

Ordinary kriging method is often referred with the acronym BLUE which stands for “Best Linear Unbiased Estimator”. “Linear” because its estimates are weighted linear combinations of the available data; “Unbiased” since it tries to have the mean error equal to 0; and “Best” because it aims at minimizing the variance of the errors. Let us then see how the concept of a random function model can be used to decide how to weight the nearby samples so that our estimates are unbiased. For any point at which we want to estimate the unknown value, our model is a stationary random function that consists of n random variables, one for the value at each of the n sample locations, $Z(\mathbf{x}_1), Z(\mathbf{x}_2), \dots, Z(\mathbf{x}_n)$, and one for the unknown value at the point we are trying to estimate $Z(\mathbf{x}_0)$. Each of these random variables has the same probability law; at all locations, the expected value of the random variable is m and the variance is σ^2 . Every value in this model is seen as an outcome (or realization) of the random variable. Our estimate is also a random variable since it is a weighted linear combination of the random variables at the n sampled locations (Cressie, 1993; Goovaerts, 1997; Isaaks & Srivastava, 1989; Kitanidis, 1997; Stein, 1999; Wackernagel, 2003; Webster & Oliver, 2007):

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^n w_i \cdot Z(\mathbf{x}_i). \quad (1)$$

The estimation error is defined as the difference between the random variable modeling the true value and the estimate:

$$\varepsilon(\mathbf{x}_0) = Z(\mathbf{x}_0) - \hat{Z}(\mathbf{x}_0). \quad (2)$$

The estimation error is also a random variable. Its expected value, often referred to as the bias, is

$$E[\varepsilon(\mathbf{x}_0)] = m \left(1 - \sum_{i=1}^n w_i \right). \quad (3)$$

Setting this expected value to 0, to ensure an unbiasedness estimate results in:

$$\sum_{i=1}^n w_i = 1. \quad (4)$$

This is known as the condition of unbiasedness (Isaaks & Srivastava, 1989; Kitanidis, 1997; Wackernagel, 2003). The expression of the variance of the modeled error is

$$\sigma_{\varepsilon(\mathbf{x}_0)}^2 = \text{var} [Z(\mathbf{x}_0)] - 2 \text{cov} [\hat{Z}(\mathbf{x}_0), Z(\mathbf{x}_0)] + \text{var} [\hat{Z}(\mathbf{x}_0)]. \quad (5)$$

Since we have already assumed that all of the random variables have the same variance σ^2 , then $\text{var} [Z(\mathbf{x}_0)] = \sigma^2$. The second term in Equation 5 can be written as

$$\begin{aligned} -2 \text{cov} [\hat{Z}(\mathbf{x}_0), Z(\mathbf{x}_0)] &= -2 \sum_{i=1}^n w_i \cdot \text{cov} [Z(\mathbf{x}_i), Z(\mathbf{x}_0)] \\ &= -2 \sum_{i=1}^n w_i C_{i0}. \end{aligned} \quad (6)$$

The third term of Equation 5 can be expressed as

$$\begin{aligned} \text{var} [\hat{Z}(\mathbf{x}_0)] &= \sum_{i=1}^n \sum_{j=1}^n w_i \cdot w_j \cdot \text{cov} [Z(\mathbf{x}_i), Z(\mathbf{x}_j)] \\ &= \sum_{i=1}^n \sum_{j=1}^n w_i w_j C_{ij}. \end{aligned} \quad (7)$$

Then, the expression of the error variance comes in the following way:

$$\sigma_{\varepsilon(\mathbf{x}_0)}^2 = \sigma^2 - 2 \sum_{i=1}^n w_i C_{i0} + \sum_{i=1}^n \sum_{j=1}^n w_i w_j C_{ij}. \quad (8)$$

Equation 8 expresses the error variance as function of the n weights, once chosen the random model function parameters, namely the variance σ^2 and all the covariances C_{ij} . The minimization of $\sigma_{\varepsilon(\mathbf{x}_0)}^2$ is constrained by the unbiasedness condition imposed earlier, which can be solved using the method of Lagrange multipliers. We start by introducing a new parameter μ , called the Lagrange multiplier, in Equation 8 in the following way:

$$\sigma_{\varepsilon(\mathbf{x}_0)}^2 = \sigma^2 - 2 \sum_{i=1}^n w_i C_{i0} + \sum_{i=1}^n \sum_{j=1}^n w_i w_j C_{ij} + 2 \underbrace{\mu \left(\sum_{i=1}^n w_i - 1 \right)}_{=0}. \quad (9)$$

Then we minimize $\sigma_{\varepsilon(\mathbf{x}_0)}^2$ by calculating the $n + 1$ partial first derivatives of Equation 9 with respect to the n weights and the Lagrange multiplier, and setting each one to 0, which

produces the following system of equations:

$$\begin{aligned}
 \frac{\partial(\sigma_{\varepsilon(\mathbf{x}_0)}^2)}{\partial(w_1)} &= -2C_{10} + 2 \sum_{j=1}^n w_j C_{1j} + 2\mu = 0 \Rightarrow \sum_{j=1}^n w_j C_{1j} + \mu = C_{10} \\
 &\vdots \\
 \frac{\partial(\sigma_{\varepsilon(\mathbf{x}_0)}^2)}{\partial(w_i)} &= -2C_{i0} + 2 \sum_{j=1}^n w_j C_{ij} + 2\mu = 0 \Rightarrow \sum_{j=1}^n w_j C_{ij} + \mu = C_{i0} \\
 &\vdots \\
 \frac{\partial(\sigma_{\varepsilon(\mathbf{x}_0)}^2)}{\partial(w_n)} &= -2C_{n0} + 2 \sum_{j=1}^n w_j C_{nj} + 2\mu = 0 \Rightarrow \sum_{j=1}^n w_j C_{nj} + \mu = C_{n0} \\
 \frac{\partial(\sigma_{\varepsilon(\mathbf{x}_0)}^2)}{\partial(\mu)} &= 2 \left(\sum_{i=1}^n w_i - 1 \right) = 0 \Rightarrow \sum_{i=1}^n w_i = 1
 \end{aligned}$$

which can also be written in a compact way as

$$\sum_{j=1}^n w_j C_{ij} + \mu = C_{i0}, \quad \forall i = 1, 2, \dots, n; \quad \sum_{i=1}^n w_i = 1. \quad (10)$$

This system of equations, often referred to as the *ordinary kriging system*, can be written in matrix notation as

$$\begin{aligned}
 \mathbf{C} \cdot \mathbf{W} &= \mathbf{D} \\
 \underbrace{\begin{bmatrix} C_{11} & \dots & C_{1n} & 1 \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & \dots & C_{nn} & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}}_{(n+1) \times (n+1)} \cdot \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{bmatrix}}_{(n+1) \times 1} &= \underbrace{\begin{bmatrix} C_{10} \\ \vdots \\ C_{n0} \\ 1 \end{bmatrix}}_{(n+1) \times 1}
 \end{aligned} \quad (11)$$

The set of weights and the Lagrange multiplier that will produce an unbiased estimate of $Z(\mathbf{x}_0)$ with the minimum error variance are then given by

$$\mathbf{W} = \mathbf{C}^{-1} \cdot \mathbf{D}. \quad (12)$$

The value of $\sigma_{\varepsilon(\mathbf{x}_0)}^2$ can be obtained in a quicker way using an alternative expression to Eq. 8. Multiplying each of the n equations given in Eq. 10 by w_i and summing these n equations leads to the following:

$$\sum_{i=1}^n \sum_{j=1}^n w_i w_j C_{ij} = \sum_{i=1}^n w_i C_{i0} - \mu.$$

Substituting this into Equation 8 the minimized error variance comes as follows:

$$\sigma_{\varepsilon(\mathbf{x}_0)}^2 = \sigma^2 - \left(\sum_{i=1}^n w_i C_{i0} + \mu \right), \quad (13)$$

or, in terms of matrices as

$$\sigma_{\varepsilon(\mathbf{x}_0)}^2 = \sigma^2 - \mathbf{W}^T \cdot \mathbf{D}. \quad (14)$$

The minimized error variance is usually called the *ordinary kriging variance*.

2.3 Block kriging

A consideration in many environmental applications has been that ordinary kriging usually exhibits large prediction errors (Bivand et al., 2008). This is due to the larger variability in the observations. When predicting averages over larger areas, i.e. within blocks, much of the variability averages out and consequently block mean values have lower prediction errors. If the blocks are not too large the spatial patterns do not disappear. The block kriging system is similar to the point kriging system given by Equation 11. The matrix \mathbf{C} is the same since it is independent of the location at which the block estimate is required. The covariances for the vector \mathbf{D} are point-to-block covariances. Supposing that the mean value over a block V is approximated by the arithmetic average of the N point variables contained within that block (Goovaerts, 1997; Isaaks & Srivastava, 1989), i.e.

$$Z_V \approx \frac{1}{N} \sum_{j=1}^N Z(\mathbf{x}_j), \quad (15)$$

the point-to-block covariances required for vector \mathbf{D} are

$$\bar{C}_{iV} = \text{cov}[Z(\mathbf{x}_i), Z_V] = \frac{1}{N} \sum_{j=1}^N C_{ij}, \quad \forall i = 1, 2, \dots, n. \quad (16)$$

The block kriging variance is

$$\sigma_V^2 = \bar{C}_{VV} - \left(\sum_{i=1}^n w_i \bar{C}_{iV} + \mu \right), \quad (17)$$

where \bar{C}_{VV} is the average covariance between pairs of points within V :

$$\bar{C}_{VV} = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N C_{ij}. \quad (18)$$

An equivalent procedure, that can be computationally more expensive than block kriging, is to obtain the block estimate by averaging the N kriged point estimates within the block (Goovaerts, 1997; Isaaks & Srivastava, 1989).

2.4 Spatial continuity

Spatial continuity exists in most earth science data sets. When we look at a contour map, or anything similar, the values do not appear to be randomly located, but rather, low values tend to be near other low values and high values tend to be near other high values. I.e. two measurements close to each other are most likely to have similar values than two measurements far apart (Isaaks & Srivastava, 1989). To compute the set of weights and the Lagrange multiplier, that will produce each estimate and the resulting minimized error variance, we need to know the covariances of \mathbf{C} and \mathbf{D} matrices. As we said before, since our random function is stationary, all pairs of random variables separated by a

distance and direction \mathbf{h} (known as lag) have the same joint probability distribution. The covariance function, $C(\mathbf{h})$ is the covariance between random variables separated by a lag \mathbf{h} (Isaaks & Srivastava, 1989; Kitanidis, 1997; Wackernagel, 2003). For a stationary random function, the covariance function $C(\mathbf{h})$ is:

$$C(\mathbf{h}) = E [Z(\mathbf{x}) Z(\mathbf{x} + \mathbf{h})] - \{E [Z(\mathbf{x})]\}^2. \quad (19)$$

The covariance between random variables at identical locations is the variance of the random function:

$$C(\mathbf{0}) = E [\{Z(\mathbf{x})\}^2] - \{E [Z(\mathbf{x})]\}^2 = var [Z(\mathbf{x})] = \sigma^2. \quad (20)$$

The semivariogram, or simply variogram, is half the expected squared difference between random variables separated by a lag \mathbf{h} :

$$\gamma(\mathbf{h}) = \frac{1}{2} E [\{Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})\}^2] = \frac{1}{2} var [Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})]. \quad (21)$$

The quantity $\gamma(\mathbf{h})$ is known as the semivariance at lag \mathbf{h} . The “semi” refers to the fact that it is half of a variance. The variogram between random variables at identical locations is zero, i.e. $\gamma(\mathbf{0}) = 0$. Using Equations 19, 20 and 21, we can relate the variogram with the covariance function as:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}) = \sigma^2 - C(\mathbf{h}). \quad (22)$$

In practice, the pattern of spatial continuity chosen for the random function is usually taken from the spatial continuity evident in the sample data set. Geostatisticians usually define the spatial continuity of the sample data set through the variogram and solve the ordinary kriging system using covariance (Isaaks & Srivastava, 1989). The maximum value reached by the variogram is called the sill. The distance at which the sill is reached is called the range. The vertical jump from zero at the origin to the value of semivariance at extremely small separation distances is called the nugget effect. The estimator of the variogram usually used, known as Matheron’s method-of-moments estimator (MME) is (Matheron, 1965; Webster & Oliver, 2007)

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \mathbf{h})]^2, \quad (23)$$

where $z(\mathbf{x}_i)$ is the value of the variable of interest at location \mathbf{x}_i and $N(\mathbf{h})$ is the number of pairs of points separated by the particular lag vector \mathbf{h} . Cressie and Hawkins (Cressie & Hawkins, 1980) developed an estimator of the variogram that should be robust to the presence of outliers and enhance the variogram spatial continuity, having also the advantage of not spreading the effect of outliers in computing the maps. This estimator (CRE) is defined as follows (Cressie & Hawkins, 1980):

$$\gamma(\mathbf{h}) = \frac{1}{2} \times \frac{\left\{ \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} |Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \mathbf{h})|^{1/2} \right\}^4}{0.457 + \frac{0.494}{N(\mathbf{h})} + \frac{0.045}{[N(\mathbf{h})]^2}}. \quad (24)$$

Once the sample variogram has been calculated, a function (called the variogram model) has to be fit to it. First, because the matrices \mathbf{C} and \mathbf{D} may need semivariance values for lags that are not available from the sample data. And second, because the use of the variogram does

not guarantee the existence and uniqueness of solution to the ordinary kriging system. The most commonly used variogram models are the spherical model, the exponential model, the Gaussian model and the Matern model (Isaaks & Srivastava, 1989).

2.5 Cross-validation

Cross-validation is a procedure used to compare the performance of several competing models (Webster & Oliver, 2007). It starts by splitting the data set into two sets: a modelling set and a validation set. Then the modelling set is used for variogram modelling and kriging on the locations of the validation set. Finally the measurements of the validation set are compared to their predictions (Bivand et al., 2008). If the average of the cross-validation errors (or Mean Error, ME) is close to 0,

$$ME = \frac{1}{m} \sum_{i=1}^m [Z(\mathbf{x}_i) - \hat{Z}(\mathbf{x}_i)] . \quad (25)$$

we may say that apparently the estimates are unbiased ($Z(\mathbf{x}_i)$ and $\hat{Z}(\mathbf{x}_i)$ are, respectively, the measurement and estimate at point x_i and m is the number of measurements of the validation set). A significant negative (positive) mean error can represent systematic overestimation (underestimation). The magnitude of the Root Mean Squared Error (RMSE) is particularly interesting for comparing different models (Wackernagel, 2003; Webster & Oliver, 2007):

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m [Z(\mathbf{x}_i) - \hat{Z}(\mathbf{x}_i)]^2} . \quad (26)$$

The RMSE value should be as small as possible indicating that estimates are close to measurements. The kriging standard deviation represents the error predicted by the estimation method. Dividing the cross-validation error by the corresponding kriging standard deviation allows to compare the magnitudes of both actual and predicted error (Wackernagel, 2003; Webster & Oliver, 2007). Therefore, the average of the standardized squared cross-validation errors (or Mean Standardized Squared Error, MSSE)

$$MSSE = \frac{1}{m} \sum_{i=1}^m \frac{[Z(\mathbf{x}_i) - \hat{Z}(\mathbf{x}_i)]^2}{\sigma_{R(\mathbf{x}_i)}^2} , \quad (27)$$

should be about one, indicating that the model is accurate. A scatterplot of true versus predicted values provides additional evidence on how well an estimation method has performed. The coefficient of determination R^2 is a good index for summarizing how close the points on the scatterplot come to falling on the 45-degree line passing through the origin (Isaaks & Srivastava, 1989). R^2 should be close to one.

3. Results

3.1 Study site

A map of the study site is shown in Fig. 2(a). Foz do Arelho outfall is located off the Portuguese west coast near Óbidos lagoon. In operation since June 2005, is presently discharging about 0.11 m³/s of mainly domestic wastewater from the WWTPs of Óbidos, Carregal, Caldas da Rainha, Gaeiras, Charneca and Foz do Arelho, but it can discharge up to 0.35 m³/s. The total length of the outfall, including the diffuser, is 2150 m. The outfall pipe,

made of HDPE, has a diameter of 710 mm. The diffuser, which consists of 10 ports spaced 8 or 12 meters apart, is 93.5 m long. The ports, nominally 0.175 m in diameter, are discharging upwards at an angle of 90° to the pipe horizontal axis; the port height is about 1 m. The outfall direction is southeast-northwest (315.5° true bearing) and is discharging at a depth of about 31 m. In that area the coastline itself runs at about a 225° angle with respect to true north and the isobaths are oriented parallel to the coastline. A seawater quality monitoring program for the outfall has already started in May 2006. Its main purposes are to evaluate the background seawater quality both in offshore and nearshore locations around the vicinity of the sea outfall and to follow the impacts of wastewater discharge in the area. During the campaign the discharge remained fairly constant with an average flowrate of approximately 0.11 m³/s. The operation area specification was based on the outputs of a plume prediction model (Hunt et al., 2010) which include mixing zone length, spreading width, maximum rise height and thickness. The model inputs are, besides the diffuser physical characteristics, the water column stratification, the current velocity and direction, and the discharge flowrate. Information on density stratification was obtained from a vertical profile of temperature and salinity acquired in the vicinity of the diffuser two weeks before the campaign (see Fig. 3). The water column was weakly stratified due to both low-temperature and salinity variations. The total difference in density over the water column was about 0.13 σ -unit. The current direction of 110° was estimated based on predictions of wind speed and direction of the day of the campaign. A current velocity of 0.12 m/s was estimated based on historic data. The effluent flowrate consider for the plume behavior simulation was 0.11 m³/s. According to the predictions of the model, the plume was spreading 1 m from the surface, detached from the bottom and forming a two-layer flow. The end of the mixing zone length was predicted to be 141 m downstream from the diffuser. Fig. 2(b) shows the diffuser and a plan view of the AUV operation area (specified according to the model predictions), mainly in the northeast direction from the diffuser, covering about 20000 m². The vehicle collected CTD data at 1.5 m and 3 m depth, in accordance to the plume minimum dilution height prediction. During the mission transited at a fairly constant velocity of 1 m/s (2 knots) recording data at a rate of 16 Hz. Maximum vertical oscillations of the AUV in performing the horizontal trajectories were less than 0.5 m (up and down).

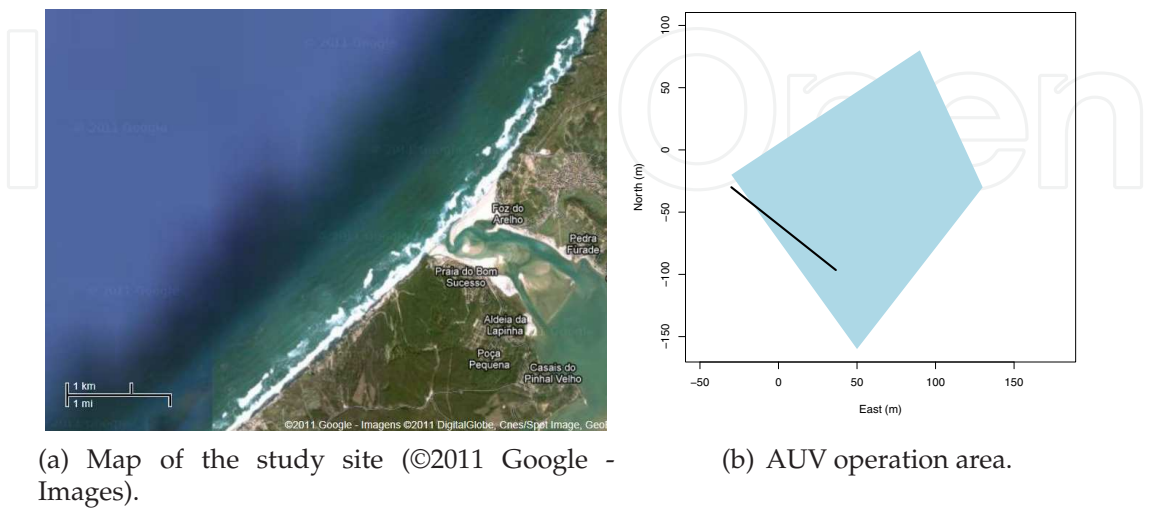


Fig. 2. Vicinity of Foz do Arelho sea outfall.

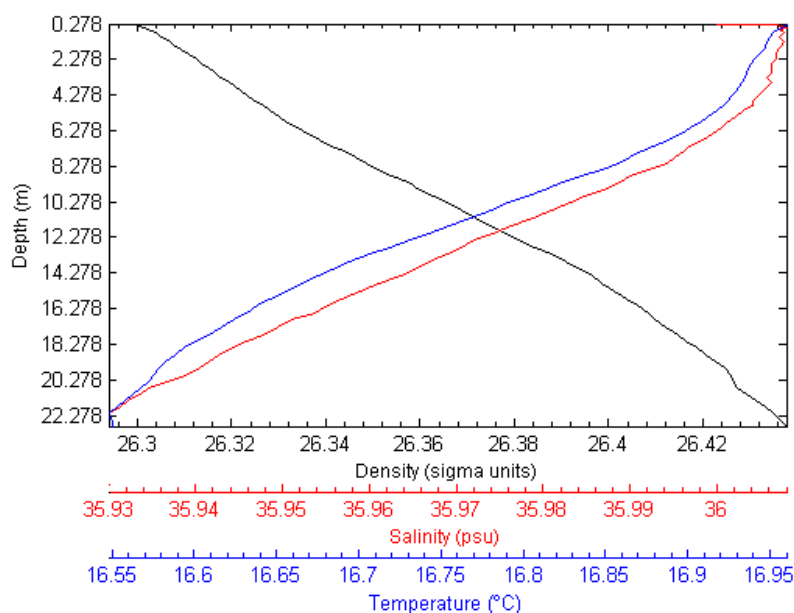


Fig. 3. Vertical STD profile used in the plume behavior simulation.

3.2 Exploratory analysis

In order to obtain elementary knowledge about the temperature and salinity data sets, conventional statistical analysis was conducted (see the results in Table 1 and Table 2). At the depth of 1.5 m the temperature ranged from 15.359°C to 15.562°C and at the depth of 3 m the temperature ranged from 15.393°C to 15.536°C. The mean value of the data sets was 15.463°C and 15.469°C, respectively at the depths of 1.5 m and 3 m, which was very close to the median value that was respectively 15.466°C and 15.472°C. The coefficient of skewness is relatively low (-0.309) for the 1.5 m data set and not very high (-0.696) for the 3 m data set, indicating that in the first case the histogram is approximately symmetric and in the second case that distribution is only slightly asymmetric. The very low values of the coefficient of variation (0.002 and 0.001) reflect the fact that the histograms do not have a tail of high values. At the depth of 1.5 m the salinity ranged from 35.957 psu to 36.003 psu and at the depth of 3 m the salinity ranged from 35.973 psu to 36.008 psu. The mean value of the data sets was 35.991 psu and 35.996 psu, respectively at the depths of 1.5 m and 3 m, which was very close to the median value that was respectively 35.990 and 35.998 psu. The coefficient of skewness is not too much high in both data sets (-0.63 and -1.1) indicating that distributions are only slightly asymmetric. The very low values of the coefficient of variation (0.0002 and 0.0001) reflect the fact that the histograms do not have a tail of high values. The ordinary kriging method works better if the distribution of the data values is close to a normal distribution. Therefore, it is interesting to see how close the distribution of the data values comes to being normal. Fig. 4 shows the plots of the normal distribution adjusted to the histograms of the temperature measured at depths of 1.5 m and 3 m, and Fig. 5 shows the plots of the normal distribution adjusted to the histograms of the salinity measured at depths of 1.5 m and 3 m. The density value in the histogram is the ratio between the number of samples in a bin and the total number of samples divided by the width of the bin (constant). Apart from some erratic high values it can be seen that the histograms are reasonably close to the normal distribution.

	Temperature@1.5 m	Temperature@3.0 m
Samples	20,026	10,506
Mean	15.463°C	15.469°C
Median	15.466°C	15.472°C
Minimum	15.359°C	15.393°C
Maximum	15.562°C	15.536°C
Coefficient of skewness	-0.31	-0.70
Coefficient of variation	0.002	0.001

Table 1. Summary statistics of temperature measurements.

	Salinity@1.5 m	Salinity@3.0 m
Samples	20,026	10,506
Mean	35.991 psu	35.996 psu
Median	35.990 psu	35.998 psu
Minimum	35.957 psu	35.973 psu
Maximum	36.003 psu	36.008 psu
Coefficient of skewness	-0.63	-1.1
Coefficient of variation	0.0002	0.0001

Table 2. Summary statistics of salinity measurements.

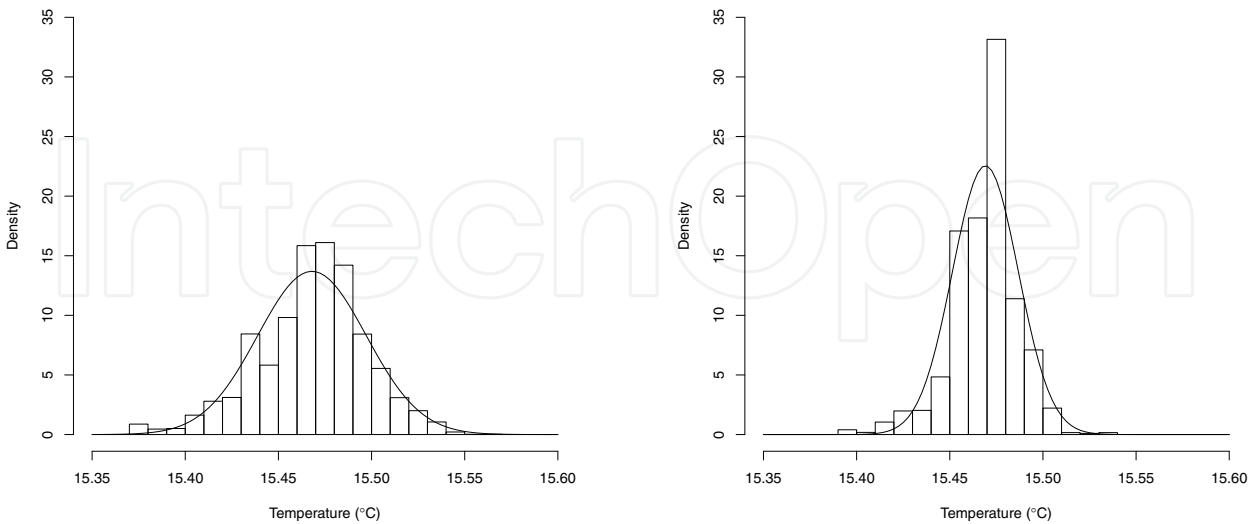


Fig. 4. Histograms of temperature measurements at depths of 1.5 m (left) and 3 m (right).

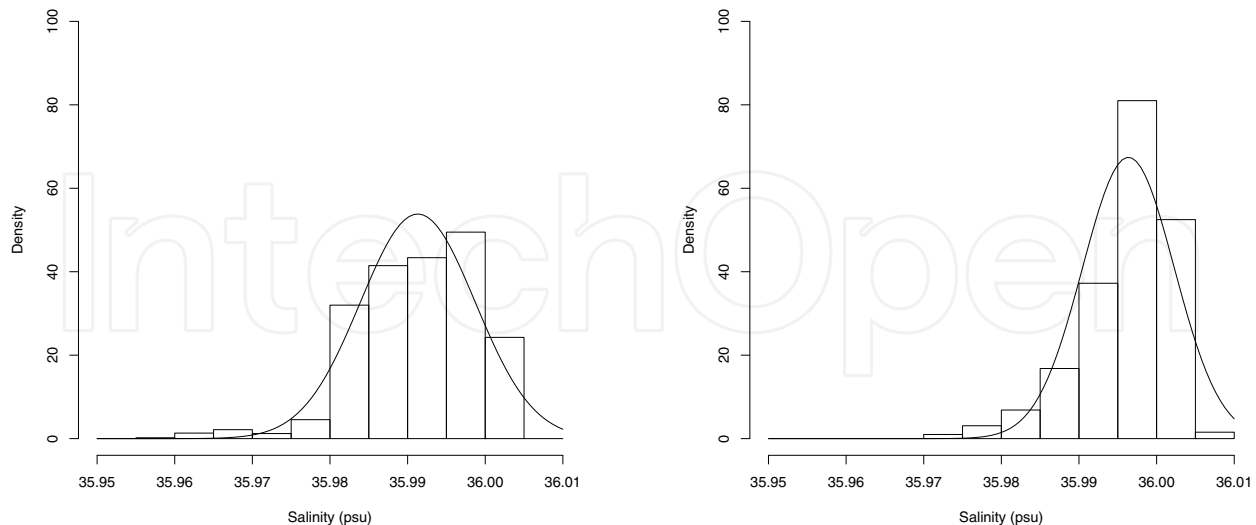


Fig. 5. Histograms of salinity measurements at depths of 1.5 m (left) and 3 m (right).

3.3 Variogram modeling

For the purpose of this analysis, the temperature and the salinity measurements were divided into a modeling set (comprising 90% of the samples) and a validation set (comprising 10% of the samples). Modeling and validation sets were then compared, using Student's-t test, to check that they provided unbiased sub-sets of the original data. Furthermore, sample variograms for the modeling sets were constructed using the MME estimator and the CRE estimator. This robust estimator was chosen to deal with outliers and enhance the variogram's spatial continuity. An estimation of semivariance was carried out using a lag distance of 2 m. Table 3 and Table 4 show the parameters of the fitted models to the omnidirectional sample variograms constructed using MME and CRE estimators. All the variograms were fitted to Matern models (for several shape parameters ν) with the exception to the salinity data measured at the depth of 3 m. The range value (in meters) is an indicator of extension where autocorrelation exists. The variograms of salinity show significant differences in range. The autocorrelation distances are always larger for the CRE estimator which may demonstrate the enhancement of the variogram's spatial continuity. All variograms have low nugget values which indicates that local variations could be captured due to the high sampling rate and to the fact that the variables under study have strong spatial dependence. Anisotropy was investigated by calculating directional variograms. However, no anisotropy effect could be shown.

3.4 Cross-validation

The block kriging method was preferred since it produced smaller prediction errors and smoother maps than the point kriging. Using the 90% modeling sets of the two depths, a two-dimensional ordinary block kriging, with blocks of $10 \times 10 \text{ m}^2$, was applied to estimate temperature at the locations of the 10% validation sets. The validation results for both parameters measured at depths of 1.5 m and 3 m depths are shown in Table 5 and Table 6. At both depths temperature was best estimated by the variogram constructed using CRE. Salinity at the depth of 1.5 m was best estimated by the variogram constructed using CRE and at the depth of 3 m was best estimated using the Gaussian model with the MME. The

Depth	Variogram Estimator	Model	Nugget	Sill	Range
1.5	MME	Matern ($\nu = 0.4$)	0.000	0.001	75.0
	CRE	Matern ($\nu = 0.5$)	0.000	0.002	80.1
3.0	MME	Matern ($\nu = 0.3$)	0.000	0.0002	101.3
	CRE	Matern ($\nu = 0.7$)	0.000	0.002	107.5

Table 3. Parameters of the fitted variogram models for temperature measured at depths of 1.5 and 3.0 m.

Depth	Variogram Estimator	Model	Nugget	Sill	Range
1.5	MME	Matern ($\nu = 0.6$)	0.436	11.945	134.6
	CRE	Matern ($\nu = 0.6$)	0.153	10786.109	51677.1
3.0	MME	Matern ($\nu = 0.8$)	0.338	11.724	181.6
	CRE	Gaussian	0.096	120.578	390.1

Table 4. Parameters of the fitted variogram models for salinity measured at depths of 1.5 and 3 m.

Depth	Method	R^2	ME	MSE	RMSE
1.5	MBK	0.9184	2.0174e-4	8.0530e-5	8.9739e-3
	CBK ^a	0.9211	1.6758e-4	7.7880e-5	8.8248e-3
3.0	MBK	0.8748	1.0338e-4	3.6295e-5	6.0244e-3
	CBK ^a	0.8827	0.6538e-4	3.4008e-5	5.8316e-3

^a The preferred model.

Table 5. Cross-validation results for the temperature maps at depths of 1.5 and 3 m.

difference in performance between the two estimators: block kriging using the MME estimator (MBK) or block kriging using the CRE estimator (CBK) is not substantial. Fig. 6 shows the omnidirectional sample variograms for temperature at the depth of 1.5 m and 3 m fitted by the preferred models. Fig. 7 shows the omnidirectional sample variograms for salinity at the depth of 1.5 m and 3 m fitted by the preferred models. Fig. 8 and Fig. 9 show the scatterplots of true versus estimated values for the most satisfactory models. The dark line is the 45° line passing through the origin and the discontinuous line is the OLS (Ordinary Least Squares) regression line. These plots show that observed and predicted values are highly positively correlated. The R^2 value for the temperature at the depth of 1.5 m was 0.9211 and the RMSE was 0.0088248°C, and at the depth of 3 m was 0.8827 and the RMSE was 0.0058316°C (Table 5). The R^2 value for the salinity at the depth of 1.5 m was 0.9513 and the RMSE was 0.0016435 psu, and at the depth of 3 m was 0.8982 and the RMSE was 0.0019793 psu (Table 6).

Depth	Method	R^2	ME	MSE	RMSE
1.5	MBK	0.9471	3.1113e-5	2.8721e-6	1.6947e-3
	CBK ^a	0.9513	-3.1579e-5	2.7010e-6	1.6435e-3
3.0	MBK ^a	0.8982	-7.1735e-5	3.9175e-6	1.9793e-3
	CBK	0.7853	-8.1264e-5	8.2589e-6	2.8738e-3

^a The preferred model.

Table 6. Cross-validation results for the salinity maps at depths of 1.5 and 3 m.

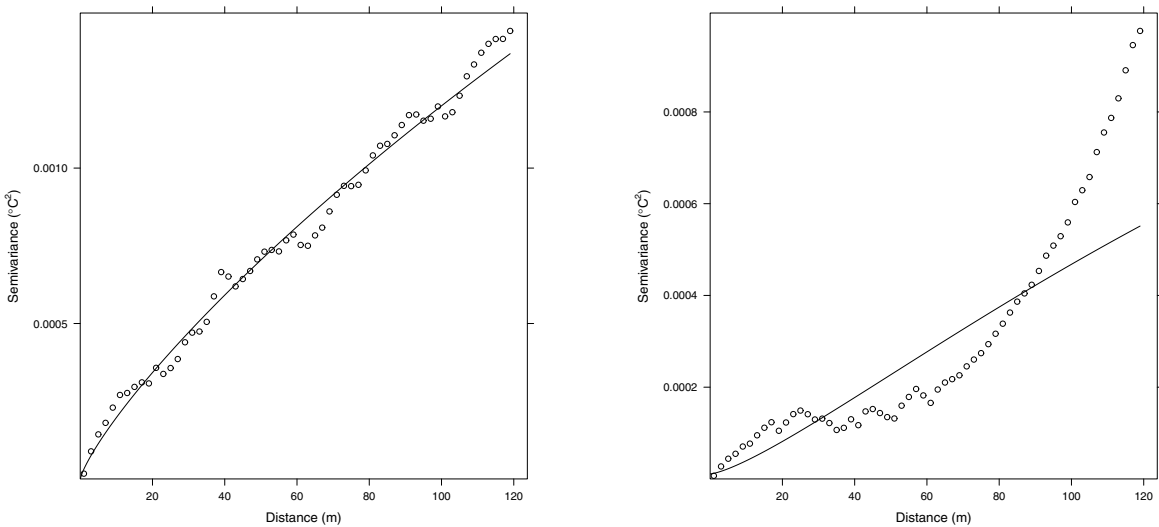


Fig. 6. Variograms for temperature at depths of 1.5 m (left) and 3 m (right).

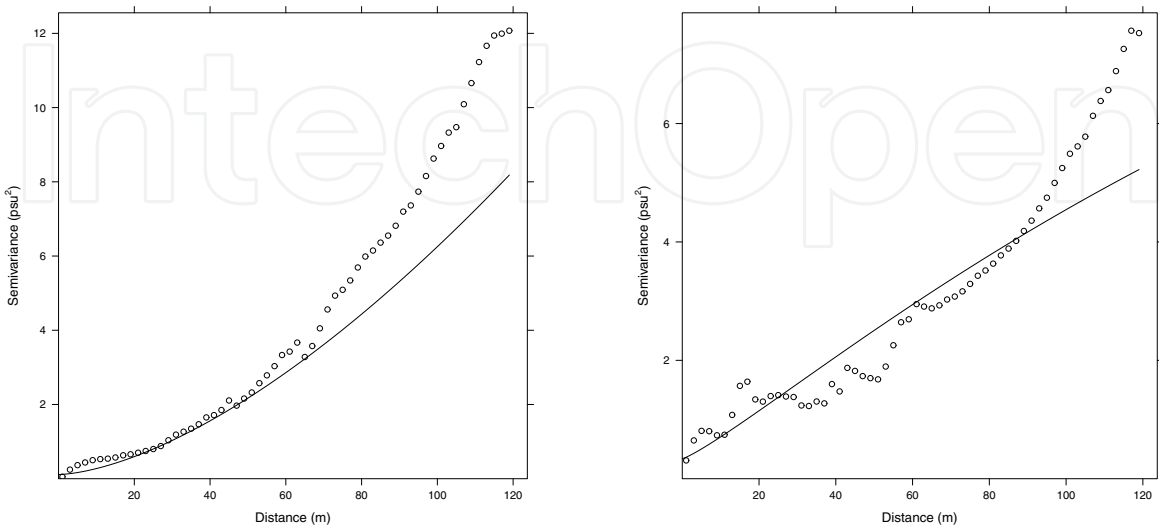


Fig. 7. Variograms for salinity at depths of 1.5 m (left) and 3 m (right).

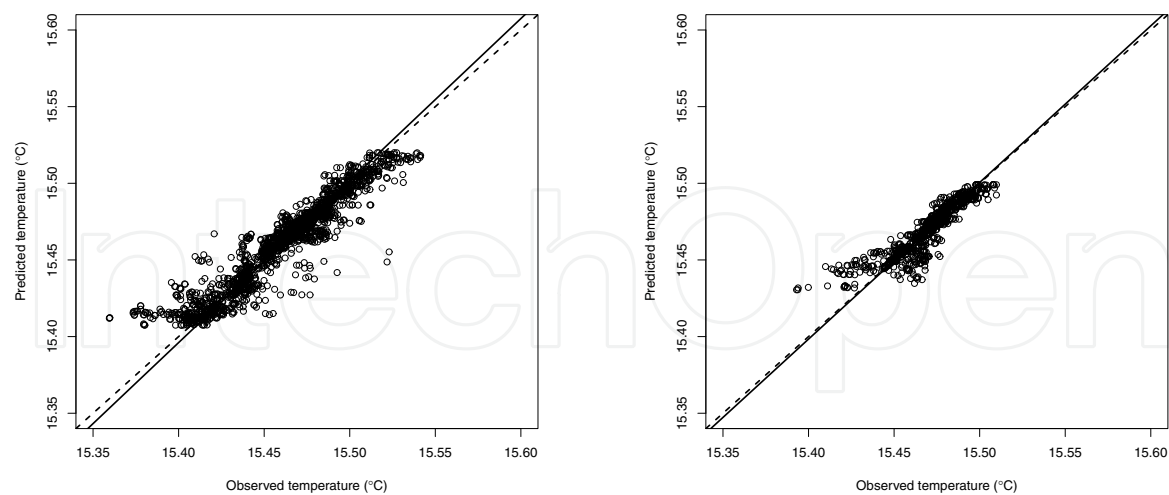


Fig. 8. Predicted versus observed temperature at the depths of 1.5 m (left) and 3 m (right) using the preferred models.

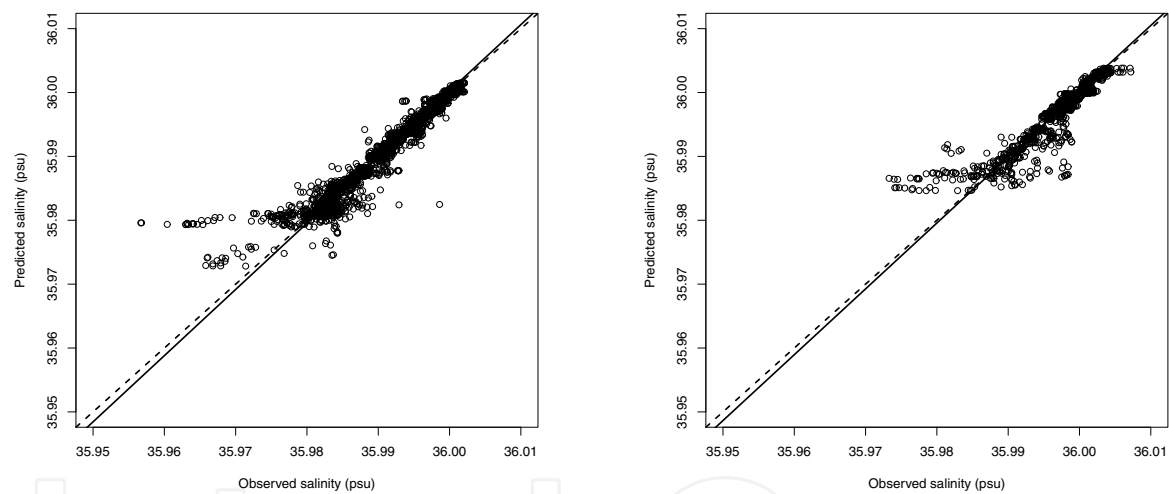


Fig. 9. Predicted versus observed salinity at the depths of 1.5 m (left) and 3 m (right) using the preferred models.

3.5 Mapping

Fig. 10 shows the block kriged maps of temperature on a $2 \times 2 \text{ m}^2$ grid using the preferred models. Fig. 13 shows the block kriged maps of salinity on a $2 \times 2 \text{ m}^2$ grid using the preferred models. In the 1.5 m kriged map the temperature ranges between 15.407°C and 15.523°C and the average value is 15.469°C (the measured range is 15.359°C – 15.562°C and the average value is 15.463°C). In the 3 m kriged map the temperature ranges between 15.429°C and 15.502°C and the average value is 15.467°C (the measured range is 15.393°C – 15.536°C and the average value is 15.469°C). We may say that estimated values are in accordance with the measurements since their distributions are similar (identical average values, medians, and quartiles). The difference in the ranges width is due to only 5.0% of the samples in the 1.5 m depth map (2.5% on each side of the distribution) and only 5.3% of the samples in the 3.0 m depth map

(3.1% on the left side and 2.2% on the right side of the distribution). These samples should then be identified as outliers not representing the behaviour of the plume in the established area. In the 1.5 m kriged map the salinity ranges between 35.960 psu and 36.004 psu and the average value is 35.992 psu, which is in accordance with the measurements (the measured range is 35.957 psu – 36.003 psu and the average value is 35.991 psu). In the 3 m kriged map the salinity ranges between 35.977 psu and 36.004 psu and the average value is 35.995 psu, which is in accordance with the measurements (the measured range is 35.973 psu – 36.008 psu and the average value is 35.996 psu). As predicted by the plume prediction model, the effluent was found dispersing close to the surface. From the temperature and salinity kriged maps it is possible to distinguish the effluent plume from the background waters. It appears as a region of lower temperature and lower salinity when compared to the surrounding ocean waters at the same depth. At the depth of 1.5 m the major difference in temperature compared to the surrounding waters is about -0.116°C while at the depth of 3 m this difference is about -0.073°C . At the depth of 1.5 m the major difference in salinity compared to the surrounding waters is about -0.044 psu while at the depth of 3 m this difference is about -0.027 psu. It is important to note that these very small differences in temperature and salinity were detected due to the high resolution of the CTD sensor. (Washburn et al., 1992) observed temperature and salinity anomalies in the plume in the order, respectively of -0.3°C and -0.1 psu, when compared with the surrounding waters within the same depth range. The small plume-related anomalies observed in the maps are evidence of the rapid mixing process. Due to the large differences in density between the rising effluent plume and ambient ocean waters, entrainment and mixing processes are vigorous and the properties within the plume change rapidly (Petrenko et al., 1998; Washburn et al., 1992). The effluent plume was found northeast from the diffuser beginning, spreading downstream in the direction of current. Using the navigation data, we could later estimate current velocity and direction and the values found were, respectively, 0.4 m/s and 70°C , which is in accordance with the location of the plume.

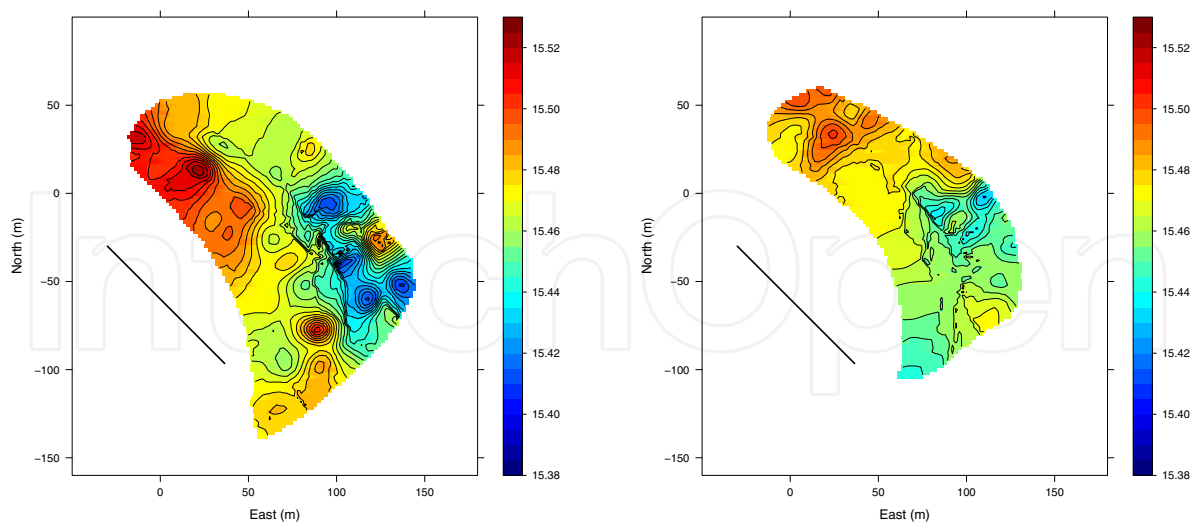


Fig. 10. Prediction map of temperature distribution ($^{\circ}\text{C}$) at depths of 1.5 m (left) and 3 m (right).

Fig. 12 shows the variance of the estimation error (kriging variance) for the maps of temperature distribution at depths of 1.5 m and 3 m. The standard deviation of the estimation error is less than 0.0195°C at the depth of 1.5 m and less than 0.0111°C at the depth of 3

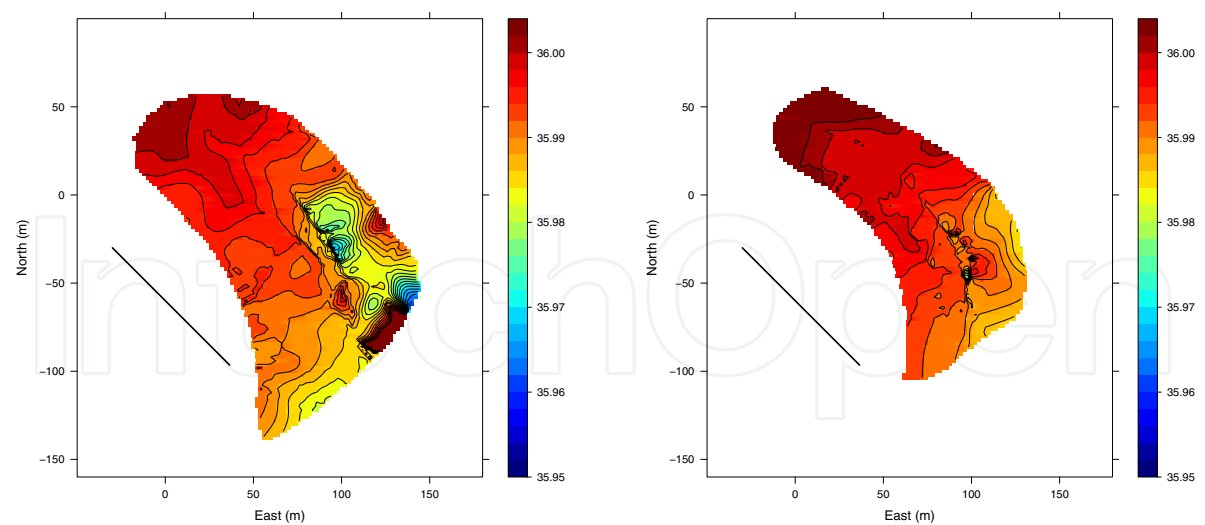


Fig. 11. Prediction map of salinity distribution (psu) at depths of 1.5 m (left) and 3 m (right).

m. Results of the same order were obtained for salinity. It’s interesting to observe that, as expected, the variance of the estimation error is less the closer is the prediction from the trajectory of the vehicle. The dark blue regions correspond to the trajectory of MARES AUV.

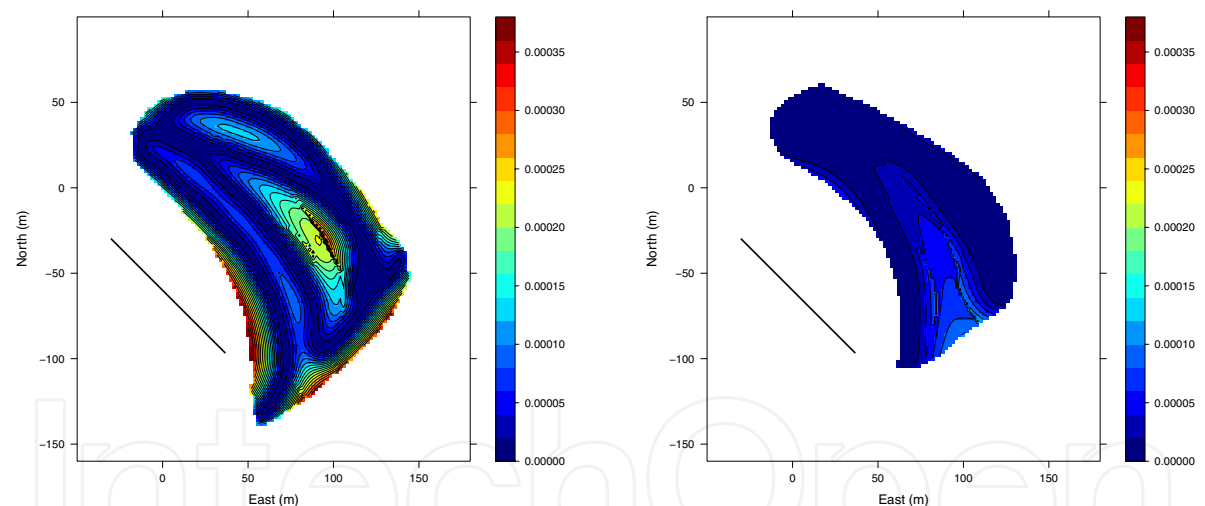


Fig. 12. Variance of the estimation error for the maps of temperature distribution at depths of 1.5 m (left) and 3 m (right)

3.6 Dilution estimation

Environmental effects are all related to concentration C of a particular contaminant X . Defining C_a as the background concentration of substance X in ambient water and C_0 as the concentration of X in the effluent discharge, the local dilution comes as follows (Fischer et al., 1979):

$$S = \frac{C_0 - C_a}{C - C_a},$$

(28)

which can be rearranged to give $C = C_a \left(\frac{S-1}{S} \right) + \left(\frac{1}{S} \right) C_0$. In the case of variability of the background concentration of substance X in ambient water the local dilution is given by

$$S = \frac{C_0 - C_{a0}}{C - C_a}, \quad (29)$$

where C_{a0} is the background concentration of substance X in ambient water at the discharge depth. This expression in 29 can be arranged to give $C = C_a + \left(\frac{1}{S} \right) (C_0 - C_{a0})$, which in simple terms means that the increment of concentration above background is reduced by the dilution factor S from the point of discharge to the point of measurement of C . Using salinity distribution at depths of 1.5 m and 3 m we estimated dilution using Equation 29 (see the contour maps in Fig. 13). We assumed $C_0 = 2.3$ psu, $C_{a0} = 35.93$ psu, $C_a = 36.008$ psu at 1.5 m depth and $C_a = 36.006$ psu at 3 m depth. The minimum dilution estimated at the depth of 1.5 m was 705 and at the depth of 3.0 m was 1164 which is in accordance with Portuguese legislation that suggests that outfalls should be designed to assure a minimum dilution of 50 when the plume reaches surface (INAG, 1998). (Since dilution increases with the plume rising we should expect that the minimum values would be greater if the plume reached surface (Hunt et al., 2010)).

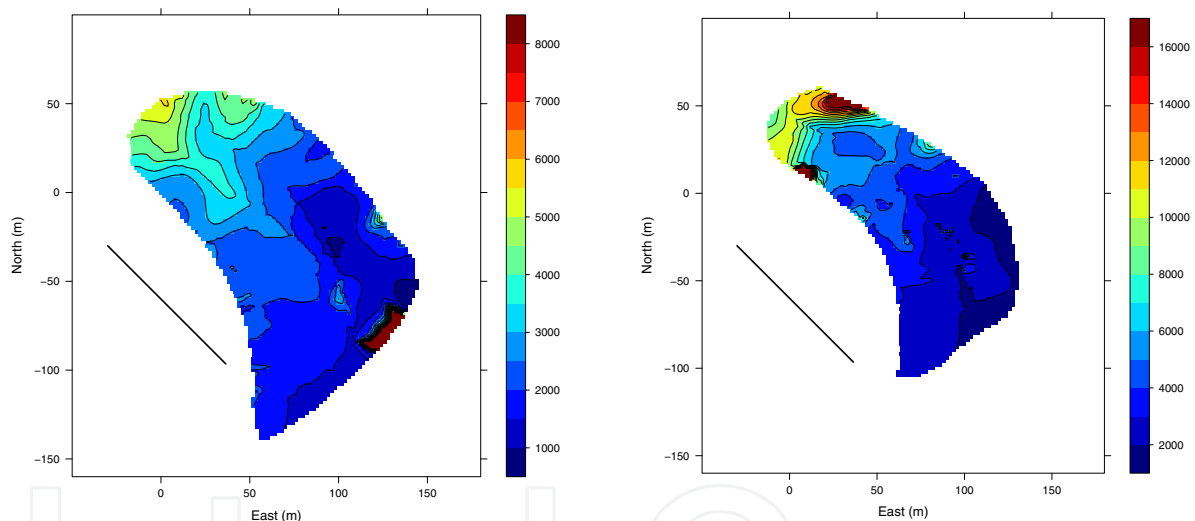


Fig. 13. Dilution maps at depths of 1.5 m (left) and 3 m (right).

4. Conclusion

Through geostatistical analysis of temperature and salinity obtained by an AUV at depths of 1.5 m and 3 m in an ocean outfall monitoring campaign it was possible to produce kriged maps of the sewage dispersion in the field. The spatial variability of the sampled data has been analyzed and the results indicated an approximated normal distribution of the temperature and salinity measurements, which is desirable. The Matheron's classical estimator and Cressie and Hawkins' robust estimator were then used to compute the omnidirectional variograms that were fitted to Matern models (for several shape parameters) and to a Gaussian model. The performance of each competing model was compared using a split-sample approach. In the case of temperature, the validation results, using a two-dimensional ordinary block

kriging, suggested the Matern model ($\nu = 0.5 - 1.5$ m and $\nu = 0.7 - 3.0$ m) with semivariance estimated by CRE. In the case of salinity, the validation results, using a two-dimensional ordinary block kriging, suggested the Matern model ($\nu = 0.6 - 1.5$ m and $\nu = 0.8 - 3.0$ m) with semivariance estimated by CRE, for the depth of 1.5 m, and with semivariance estimated by MME, for the depth of 3 m. The difference in performance between the two estimators was not substantial. Block kriged maps of temperature and salinity at depths of 1.5 m and 3 m show the spatial variation of these parameters in the area studied and from them it is possible to identify the effluent plume that appears as a region of lower temperature and lower salinity when compared to the surrounding waters, northeast from the diffuser beginning, spreading downstream in the direction of current. Using salinity distribution at depths of 1.5 m and 3 m we estimated dilution at those depths. The values found are in accordance with Portuguese legislation. The results presented demonstrate that geostatistical methodology can provide good estimates of the dispersion of effluent that are very valuable in assessing the environmental impact and managing sea outfalls.

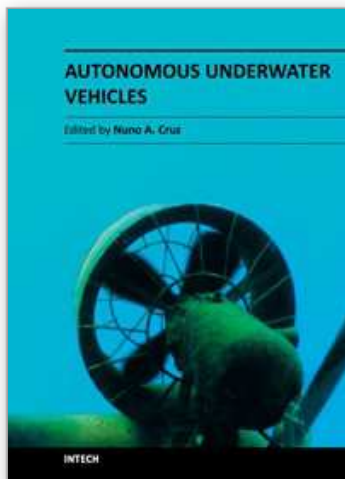
5. Acknowledgment

This work was partially funded by the Foundation for Science and Technology (FCT) under the Program for Research Projects in all scientific areas (Programa de Projectos de Investigação em todos os domínios científicos) in the context of WWECO project - Environmental Assessment and Modeling of Wastewater Discharges using Autonomous Underwater Vehicles Bio-optical Observations (Ref. PTDC/MAR/74059/2006).

6. References

- Abreu, N., Matos, A., Ramos, P. & Cruz, N. (2010). Automatic interface for AUV mission planning and supervision, *MTS/IEEE International Conference Oceans 2010*, Seattle, USA.
- Abreu, N. & Ramos, P. (2010). An integrated application for geostatistical analysis of sea outfall discharges based on R software, *MTS/IEEE International Conference Oceans 2010*, Seattle, USA.
- Bellingham, J. (1997). New Oceanographic Uses of Autonomous Underwater Vehicles, *Marine Technology Society Journal* 31(3): 34–47.
- Bellingham, J., Goudey, C., Consi, T. R. & Chrysostomidis, C. (1992). A Small Long Range Vehicle for Deep Ocean Exploration, *Proceedings of the International Offshore and Polar Engineering Conference*, San Francisco, pp. 151–159.
- Bivand, R. S., Pebesma, E. J. & Gómez-Rubio, V. (2008). *Applied spatial data analysis with R*. ISBN: 97-0-387-78170-9.
- Cressie, N. (1993). *Statistics for spatial data*, New York.
- Cressie, N. & Hawkins, D. M. (1980). Robust estimation of the variogram, I, *Jour. Int. Assoc. Math. Geol.* 12(2): 115–125.
- Fernandes, P. G., Brierley, A. S., Simmonds, E. J., Millard, N. W., McPhail, S. D., Armstrong, F., Stevenson, P. & Squires, M. (2000). Fish do not Avoid Survey Vessels, *Nature* 404: 35–36.
- Fischer, H. B., List, J. E., Koh, R. C. Y., Imberger, J. & Brooks, N. H. (1979). *Mixing in Inland and Coastal Waters*, Academic Press.
- Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*, *Applied Geostatistics Series*. ISBN13: 9780195115383, ISBN10: 0195115384.

- Hunt, C. D., Mansfield, A. D., Mickelson, M. J., Albro, C. S., Geyer, W. R. & Roberts, P. J. W. (2010). Plume tracking and dilution of effluent from the Boston sewage outfall, *Marine Environmental Research* 70: 150–161.
- INAG (1998). *Linhas de Orientação Metodológica para a Elaboração de Estudos Técnicos Necessários para Cumprir o Artigo 7º do D.L. 152/97. Descargas em Zonas Menos Sensíveis*, Instituto da Água. Ministério do Ambiente, Lisboa.
- Isaaks, E. H. & Srivastava, R. M. (1989). *Applied geostatistics*, New York Oxford. ISBN 0-19-505012-6-ISBN 0-19-505013-4 (pbk.).
- Kitanidis, P. (1997). *Introduction to geostatistics: applications in hydrogeology*, New York (USA).
- Matheron, G. (1965). *Les variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature*, Paris.
- Nadis, S. (1997). Real Time Oceanography Adapts to Sea Changes, *Science* 275: 1881–1882.
- 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(6): 565–571.
- Ramos, P. & Abreu, N. (2010). Spatial analysis of sea outfall discharges using block kriging, *6th International Conference on Marine Waste Water Discharges and Coastal Environment, MWWD2010*, Langkawi, Malaysia.
- Ramos, P. & Abreu, N. (2011a). Environmental impact assessment of Foz do Arelho sewage plume using MARES AUV, *IEEE/OES International Conference Oceans 2011*, Santander, Spain.
- Ramos, P. & Abreu, N. (2011b). Spatial analysis and dilution estimation of Foz do Arelho Outfall using observations gathered by an AUV, *International Symposium on Outfall Systems*, Mar del Plata, Argentina.
- Ramos, P. & Abreu, N. (2011c). Using an AUV for Assessing Wastewater Discharges Impact: An Approach Based on Geostatistics, *Marine Technology Society Journal* 45(2).
- Robinson, A., Bellingham, J., Chrystostomidis, C., Dickey, T., Levine, E., Petrikalakis, N., Porter, D., Rothschild, B., Schmidt, H., Sherman, K., Holliday, D. & Atwood, D. (1999). Real-time forecasting of the multidisciplinary coastal ocean with the littoral ocean observing and predicting system (LOOPS), *Proceedings of the Third Conference on Coastal Atmospheric and Oceanic Prediction Processes*, American Meteorological Society, New Orleans, LA.
- Stein, M. L. (1999). *Interpolation of spatial data: some theory for kriging*, New York.
- Wackernagel, H. (2003). *Multivariate geostatistics: an introduction with applications*, Berlin.
- Washburn, L., Jones, B. H., Bratkovich, A., Dickey, T. D. & Chen, M.-S. (1992). Mixing, Dispersion, and Resuspension in Vicinity of Ocean Wastewater Plume, *Journal of Hydraulic Engineering, ASCE* 118(1): 38–58.
- Webster, R. & Oliver, M. (2007). *Geostatistics for environmental scientists, 2nd Edition*. ISBN-13: 978-0-470-02858-2(HB).
- Yu, X., Dickey, T., Bellingham, J., Manov, D. & Streitlien, K. (2002). The Application of Autonomous Underwater Vehicles for Interdisciplinary Measurements in Massachusetts and Cape Cod Bayes, *Continental Shelf Research* 22(15): 2225–2245.



Autonomous Underwater Vehicles

Edited by Mr. Nuno Cruz

ISBN 978-953-307-432-0

Hard cover, 258 pages

Publisher InTech

Published online 17, October, 2011

Published in print edition October, 2011

Autonomous Underwater Vehicles (AUVs) are remarkable machines that revolutionized the process of gathering ocean data. Their major breakthroughs resulted from successful developments of complementary technologies to overcome the challenges associated with autonomous operation in harsh environments. Most of these advances aimed at reaching new application scenarios and decreasing the cost of ocean data collection, by reducing ship time and automating the process of data gathering with accurate geo location. With the present capabilities, some novel paradigms are already being employed to further exploit the on board intelligence, by making decisions on line based on real time interpretation of sensor data. This book collects a set of self contained chapters covering different aspects of AUV technology and applications in more detail than is commonly found in journal and conference papers. They are divided into three main sections, addressing innovative vehicle design, navigation and control techniques, and mission preparation and analysis. The progress conveyed in these chapters is inspiring, providing glimpses into what might be the future for vehicle technology and applications.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Patrícia Ramos and Nuno Abreu (2011). Mapping and Dilution Estimation of Wastewater Discharges Based on Geostatistics Using an Autonomous Underwater Vehicle, *Autonomous Underwater Vehicles*, Mr. Nuno Cruz (Ed.), ISBN: 978-953-307-432-0, InTech, Available from: <http://www.intechopen.com/books/autonomous-underwater-vehicles/mapping-and-dilution-estimation-of-wastewater-discharges-based-on-geostatistics-using-an-autonomous->

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen