

Chapter 3

Sand bank criteria

1. Goals

This Chapter discusses the specific criteria relevant for surveys for the sand bank volume. Its major goal is to produce models that relate the quality of the acquired raw data to the final confidence associated to the indexes that can be derived from it, describing the impact of the distinct kinds of errors that affect the collected data series.

In this section we give mathematical expressions for the ultimate quantities of interest to the end-user (in this case, MUMM) in terms of the true values of the observed field. The subsequent sections of the Chapter are organised as follows:

- Section 5.2 defines the performance criteria that evaluate the quality of the information produced to the end-user (according to the goals defined in section 5.1). Mostly, these will be first and second-order centred moments of the estimates produced using the data acquired during the survey.
- Sections 5.3--5.7 present the analytical equations that express the performance criteria presented in Section 5.2 in terms of the error processes that are present in the raw measurements, the sensor calibration errors and the errors induced by discrete spatial sampling of the continuous field.
- Section 5.8 presents the set of sampling strategies that can be implemented in an autonomous sensor, and derives expressions for the error moments appearing in the expressions presented in the previous section.

1.1 The total sand bank volume

As stated in the Technical Annex of the project, the ultimate quantity of interest to the end-user for this application is the estimation of the total volume of the sand bank below certain level lines.

We introduce some notation. Let

- $S^t \equiv \{S(z, t), z \in A\}$, $A \subset \mathfrak{R}^2$ be the depth of the sea-bottom in a given area of analysis A (a compact subset of \mathfrak{R}^2) at time t .
- $\{L_i\}_{i=1}^N \subset A$ a set of lines of interest in the region of study (these represent the DECA lines presently surveyed by MUMM with an oceanographic vessel).
- $\{h_i\}_{i=1}^K$ a set of reference depths (above which the sand bank volumes must be computed).

The ultimate quantity of interest for the sand bank application considered in SUMARE is the total volume of sand above a set of reference depths h_i inside area A at time t , which is given by

$$V^i(t) \equiv \int_A g_i(S(z,t)) dz = \int_{A_i} (h_i - S(z,t)) dz, = h_i \mu(A_i) - \int_{A_i} S(z,t) dz \quad (1)$$

where we defined

- $g_i(S) = \begin{cases} h_i - S, & S < h_i \\ 0, & \text{otherwise} \end{cases}$
- $A_i = \{z \in A : g_i(S(z)) > 0\}$
- $\mu(A_i) = \int_{A_i} dz$

and, for simplicity, we omitted the time dependency on the definition of the regions A_i , inside which the sea depth is above the i th reference depth h_i .

The other indexes of interest to MUMM are the variation of this volumes. Ideally, one would like to compute

$$\delta V^i(t) = \frac{dV^i(t)}{dt} = \frac{d}{dt} h_i \mu(A_i) - \frac{d}{dt} \int_{A_i} S(z,t) dz \cong - \int_{A_i} \frac{d}{dt} S(z,t) dz \quad (2)$$

where the last equality assumes that the regions A_i are stable (according to the available data, this is true at least for the lower cuts of the bank). The observation of temporal derivatives would imply a temporal sampling (survey rate) adapted to the rates of variation of the surface at any point of the area of study A . Since these rates are determined by the equations of transport of the sediment, subject to the external forces induced by currents and waves, they depend strongly on the temporal scales of the effects of storms and other exogenous factors. It would imply, for instance, that a survey be conducted after an external condition that may induce important transport mechanisms is observed. For this reason, instead of computing true time derivatives, it seems more realistic to compute the depth differences for each point of the area of study.

$$\Delta V^i(t_n) \equiv \int_{A_i} (S(z, t_n) - S(z, t_{n-1})) dz \quad (3)$$

where again we neglected the variation of the support set A_i , and $\{t_n\}$ are the (ordered) dates of the available surveys.

Equations (1) and (2) – or its difference analogue (3) – express the two types of indexes for interest to the management of the sand bank exploitation in terms of the true sea-bed bottom surface S^t , and are the main result of this subsection.

1.2 The sand bank volume along reference lines

The computation of the indexes $V^i(t)$ defined above requires observation of the complete sea-bed surface, or at least that it be spatially sampled at a rate that allows local reconstruction, implying impractical costs and survey durations. For this reason, a set of other indexes related to the previous ones have been defined, which require a reduced volume of observed data.

For each reference depth h_i and for a set of pre-specified observation lines L_k , compute the indexes

$$V_k^i(t) \equiv \int_{L_{ik}} (h_i - S(z, t)) d\ell, \cong h_i \mu(L_{ik}) - \int_{L_{ik}} S(z, t) dz \quad (3)$$

where $L_{ik} = \{z \in L_k : S(z) < h_i\}$, and their temporal variations:

$$\delta V_k^i(t) = \frac{dV_k^i(t)}{dt} \cong - \int_{L_{ik}} \frac{d}{dt} S(z, t) dz \quad (4)$$

where we again neglected the variation of the geometry of the lines L_{ik} with time. The remarks of the previous section concerning the estimation of time derivatives apply here, and in practice we will consider the problem of estimation temporal differences

$$\Delta V_k^i(t_n) = V_k^i(t_n) - V_k^i(t_{n-1}). \quad (5)$$

Equations (3) and (6) correspond to the indexes presently produced by MUMM for exploitation management purposes.

1.3 The variation of the bank shape

Another quantity that has been identified as of interest to the end-users is the variation of the shape of the sand banks. More precisely, instead of just computing the changes in volume, which can be zero even if the bank shape is varying, individually identify the areas where the sea depth is changing. This goal is attained if we can compute the temporal derivative of the sea bed surface (we assume that the sea-bed surface is a differentiable manifold parameterised by local geographical coordinates, which seems a reasonable hypothesis for the sand banks):

$$\delta S^t(z) = \frac{\partial S(z, t)}{\partial t}, z \in A$$

Again, the estimation of temporal derivatives will be replaced for the more realistic goal of assessing the differences in sea-bed height:

$$\Delta S^{tn}(z) \equiv S(z, t_n) - S(z, t_{n-1}), z \in A. \quad (6)$$

Besides the detailed surface of sea-bed variations, it may be of interest to compute a global (macroscopic) scalar index measuring how much the shape of the bank has changed between two consecutive surveys.

Definition of distance between geometric shapes is a subject of current research in statistics and computational geometry. In this study we will use the definition in ¹. According to this approach, a size-and-shape space $S\Sigma_n^k$ is defined as the collection of k -dimensional spaces in ambient Euclidean space \mathfrak{R}^n - with k varying in the integers. The elements of each *size-and-shape space* are defined as the quotient of \mathfrak{R}^{nk} modulo translation by vectors in \mathfrak{R}^n and rotations: each element of the shape space consists of all the translations and rotations of a representative set of k vectors in \mathfrak{R}^n , called the *pre-shape*, and defined such that their centroid is at the origin of \mathfrak{R}^n .

Given two pre-shapes S_1 and S_2 belonging to the same k -dimensional size-and-shape space, the distance between these shapes is defined as

$$d_s(s_1, s_2) = \min_{T \in SO^n} \|S_1 - TS_2\|^2 \quad (7)$$

where SO^n is the set of rotations and S_1 and S_2 are matrices that collect the points defining the corresponding shapes.

In our case, we will consider, around each point z , the 3-dimensional shapes (surfaces) defined by the coordinates in A and the sea-bed depth in an ordered discrete neighbourhood $G(z)$ of the point, and evaluate

$$\Delta S^{t_1, t_2}(z) = d_s(P_1(G(z)), P_2(G(z))) \quad (8)$$

where $P_i(G(z))$ is the pre-shape corresponding to

$$\begin{bmatrix} x_1 & & & x_{N_G} \\ y_1 & \cdots & & y_{N_G} \\ S(x_1, y_1) & & & S(x_{N_G}, y_{N_G}) \end{bmatrix},$$

and N_G is the size of the neighbourhood considered. Equation (8) defines the index of shape variation that will be used to assess the variation of the shape of the banks.

Note that when dealing with the total bank surface $n=3$ and the ambient space is the standard Euclidean space, and when observing only along lines, $n=2$. For this last case, and since we can identify \mathfrak{R}^2 with the complex plane C , the simpler expressions are obtained for distance between the elements of the size-and-shape space $S\Sigma_2^k$:

$$d_k(S_1, S_2)^2 = \|w_1\|^2 + \|w_2\|^2 - 2|\langle w_1, w_2 \rangle| \quad (9)$$

¹ *Shape and Shape Theory*, D.G. Kendall, D. Barden, T.K. Carne and H. Le, Wiley, 1999.

where the complex vectors w_1 and w_2 are the complex corresponding to each point in the shape:

$$w_i = x_i + jy_i.$$

This approach is detailed in section 5.7, where results are given for the observed shape variations for sets of 25 profiles for three distinct DECA lines.

2. Performance Criteria

In this section we present the performance criteria that measure for a given acquired data set the quality of the resulting estimates of the indexes defined in the previous section.

2.1 The total sand bank volume

We consider as performance criteria the estimates biases and their variance. A data set Y^t will be considered to be better than another if it leads to smaller values of

$$b_{V^i}(Y^t) = E[\varepsilon_{V^i}(Y^t)] = E[V^i(t) - \hat{V}^i(t, Y^t)] \quad (10)$$

and of

$$Cov_{V^i}(Y^t) = E\left[\left(\varepsilon_{V^i}(Y^t) - b_{V^i}(Y^t)\right)^2\right] \quad (11)$$

In the previous expressions, the use of “^” indicates the estimates of the corresponding quantities produced using the observed data Y^t , which will be detailed in the next section.

Equivalent criteria will be used for assessing the quality of the estimates of variation:

$$b_{\Delta V^i}(Y^t) = E[\varepsilon_{\Delta V^i}(Y^t)] = E[\Delta V^i(t) - \hat{\Delta V}^i] \quad (12)$$

$$Cov_{\Delta V^i}(Y^t) = E\left[\left(\varepsilon_{\Delta V^i}(Y^t) - b_{\Delta V^i}(Y^t)\right)^2\right]. \quad (13)$$

Note that this set of performance indexes may not define a *complete order relation* on the set of acquired data sets: it may happen that some observation strategies may lead to better estimation of some indexes while degrading another subset.

2.2 The sand bank volume along reference lines

Using the notation analogous to the one used in the previous paragraph, we consider that a data set Y^t is better than another if it leads to smaller values of

$$b_{V_k^i}(Y^t) = E \left[\varepsilon_{V_k^i}(Y^t) \right] = E \left[V_k^i(t) - \hat{V}_k^i(t, Y^t) \right] \quad (14)$$

and of

$$Cov_{V_k^i}(Y^t) = E \left[\left(\varepsilon_{V_k^i}(Y^t) - b_{V_k^i}(Y^t) \right)^2 \right] \quad (15)$$

with corresponding expressions for their variations.

2.3 The variation of the bank shape

As previously, we consider that a data set Y^t is better for observing shape variations than another if it leads to smaller values of

$$b_{\Delta S}(Y^t) = \left\| E \left[\varepsilon_{\Delta S}(z, Y^t) \right] \right\|_p = \left\| E \left[\Delta S(z) - \hat{\Delta S}(z, Y^t) \right] \right\|_p \quad (16)$$

where $\|\cdot\|_p$ denotes the L^p norm, and of some norm of its cross-correlation matrix

$$\left\| Cov_{\Delta S}(Y^t) \right\| = \left\| E \left[\left(\varepsilon_{\Delta S}(Y^t) - b_{\Delta S}(Y^t) \right) \left(\varepsilon_{\Delta S}(Y^t) - b_{\Delta S}(Y^t) \right)^T \right] \right\| \quad (17)$$

where we considered that the surface of variation is computed over a discrete grid.

If the global shape variation indexes are of interested, than we shall also consider their first and second order moments as the optimisation criteria.

3. Statistical analysis of volume estimation without prior model

The goal of this section is to present expressions for the criteria presented in the previous section (equations (7)—(14)) in terms of the errors affecting the data sets Y^t and the post-processing operations that produce the final quantities of interest in each case.

Before we start presenting the detailed equations, let us introduce some notation. Let

- $\{p_t^a\}_{t \in T}$ be the *true* vehicle trajectory in the horizontal plane during the entire observation (T is the length of the observation interval).
- $\{\theta_t^a\}_{t \in T}$ be the *true* vehicle attitude angles (the three Euler angles)
- $\{p_t\}_{t \in T}$ be the *observed* vehicle trajectory.
- $\{r_t\}_{t \in T}$ be the *estimated* vehicle depth and altitude above sea bottom
- $\{\theta_t\}_{t \in T}$ be the *observed* vehicle attitude angles.
- $\{t_k\}_{k=1}^N$ be the sampling instants
- $x_k = x(t_k)$ denote generically the value of the signal x at time t_k

3.1 Errors due to discrete sampling

Both for "total volume" and "section volumes" the quantities of interest are integrals and the temporal variations of these integrals. Since the manifolds (either a 2D line or a 3D surface, depending on the problem considered) are observed only at the discrete grid $\{p_k^a\}_{k=1}^N$ these integrals must be approximated, using any of the available integration methods, by finite sums. All these methods rely on the fact that an analytical signal can be approximated, in the neighbourhood of any point p_k by a finite Taylor expansion:

$$S(p_k + dp) = S(p_k) + \nabla_S(p_k)^T dp + \frac{1}{2} dp^T H_S(p_k) dp + h.o.t.$$

where we omit the time index t , for simplicity, $\nabla_S(p_k)^T$ is the gradient of the surface at point p_k and $H_S(p_k)$ is the Hessian matrix:

$$[H_S(p_k)]_{nm} = \frac{\partial^2 S(p_k)}{\partial p_n \partial p_m}, \quad n, m = 1, 2; \quad p_n = \begin{bmatrix} x_n \\ y_n \end{bmatrix}.$$

3.2 Total volume

Let \tilde{V} be the discrete grid approximation to V . In Appendix 2 we derive the expression for the errors, considering a simple 1st order expansion around the true samples points:

$$\tilde{V} = \sum_{k=1}^N \left(S(p_k^a) \mu(C_k^a) + \nabla_S(p_k^a)^T I_k^a \right),$$

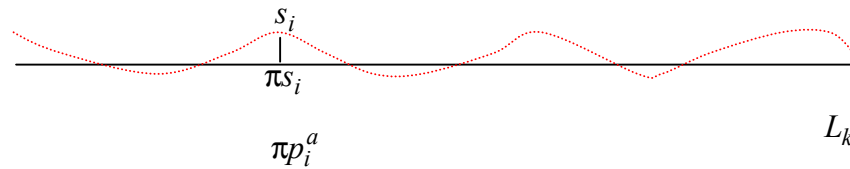
where $\mu(C_k^a)$ is the Lebesgue measure of the k -th Voronoi cell determined by the set of sampled points $\{p_k^a\}_{k=1}^N$, and $I_k^a / \mu(C_k^a)$ is the mass centroid of the k -th cell with the origin placed at point p_k^a . Using these expressions we can write the error due to the approximation of the integral by a finite sum:

$$\tilde{\epsilon}_V = \tilde{V} - V = - \sum_{k=1}^N \int_{C_k^a} (dp^T H_S(p_k^a) dp + h.o.t) dp \quad (15)$$

3.3 Section volume

In Appendix 2 we derive also an equivalent expression for the discrete grid error for sectional volumes. In this case, we must take into account that the actual sampled points do not fall in the line L_k

Let $\{p_i^a\}_{i=1,\dots,N} \subset A_i$ be the observed points, and $\{\pi p_i^a\}_{i=1,\dots,N}$ their orthogonal projections on the line



As far as the error or discrete sampling is concerned, this case is equivalent to the previous, if we consider the Voronoi cells determined by the orthogonal projections of the sampled points in the target line. We obtain, in Appendix 2, the following equation:

$$\tilde{\epsilon}_{V_k} = \tilde{V}_k - V_k = - \sum_{j=1}^N \int_{\pi C_j^a} \left(\frac{1}{2} \frac{\partial^2 S}{\partial \ell^2} \Big|_{\pi p_j^a} \ell^2 + h.o.t \right) d\ell \quad (16)$$

where

$$\tilde{V}_k = \sum_{j=1}^N \left(S(\pi p_j^a) \mu(\pi C_j^a) + \frac{\partial S(\pi p_j^a + \ell u)}{\partial \ell} \Big|_{\pi p_j^a} I_j^a \right),$$

u is the unit vector along the observed line, and μ and I have the same meaning as before but for the Voronoi diagram for the projected points.

3.4 Errors due to errors on depth measures

We assess now the way in which we can approximate the true depths appearing in the expressions of \tilde{V} of the previous paragraph, using the available measures Y^t .

3.5 Errors at each sampled point

The reconstructed sea depth for the i th measure is

$$\hat{S}_i = T(\theta_i)r_i + t_i,$$

where t_i is the tidal correction for point p_i and matrix $T(\theta_i)$ depends, besides the vehicle attitude angles, on the implantation of the depth sensor and of the altimeter in the vehicle.

The true value at the same point can be written as

$$S_i^a = T(\theta_i^a)r_i^a + t_i^a \Rightarrow r_i^a = T(\theta_i^a)^{-1} [S_i^a - t_i^a]$$

where super-index a indicates use of the true quantities. Expressing the observed values in terms of the true quantities and of the measuring errors:

$$r_i = r_i^a + \varepsilon_i^r, \theta_i = \theta_i^a + \varepsilon_i^\theta$$

we obtain

$$\hat{S}_i = T(\theta_i) \left(T(\theta_i^a)^{-1} S_i^a + \varepsilon_i^r \right) = S_i^a + \Delta T_i [S_i^a - t_i^a] + T(\theta_i) \varepsilon_i^r + \varepsilon_i^t \quad (17)$$

where we defined

$$\Delta T_i \equiv T(\theta_i) T(\theta_i^a)^{-1} - I$$

and

$$\varepsilon_i^t = t_i - t_i^a$$

is the error on tidal correction at point i .

The total error on the measures of depth can thus be decomposed as

$$\varepsilon_i^s = \hat{S}_i - S_i^a = \Delta T_i [S_i^a - t_i^a] + T(\theta_i) \varepsilon_i^r + \varepsilon_i^t.$$

3.6 Errors due to errors on position

Besides the errors on the estimation of the values of the depth at the visited points, we have to take into account also the errors due to the departure of the geometry of the actually set of sampled points with respect to the true visited points, which affect the quantities $\mu(C_i)$ and I_i .

We express these quantities as the true ones and an error term:

$$\mu(C_j^o) = \mu(C_j^a) + d\mu_j$$

$$I_j = I_j^a + dI_j$$

Note that these errors depend on the horizontal positioning errors at the point and all its neighbours, depending thus strongly on the correlation structure of the positioning error. Later, we will determine equations in terms of the individual positioning errors at each point.

3.7 Global volume estimation

Using the previous expressions, we can finally write the following total error model:

$$\varepsilon_V = \tilde{\varepsilon}_V + \varepsilon_{\Delta} + \varepsilon^{d\mu} + \varepsilon^{T\theta} + \varepsilon^{r,\theta} + \varepsilon^t,$$

where we defined the global error terms

$$\varepsilon_{\Delta} = -\sum_{i=1}^N \nabla S_i^T I_i$$

and

$$\varepsilon^{d\mu} = \sum_{j=1}^N S_j d\mu_j$$

is an error term due to the deformation of the estimated Voronoï cells wrt to the one defined by the really observed points,

$$\varepsilon^{T\theta} = \sum_{j=1}^N \Delta T_j S_j \mu(C_j)$$

is an error term due to the errors in the attitude angles,

$$\varepsilon^{r,\theta} = \sum_{j=1}^N T(\theta_j) \varepsilon_j^r \mu(C_j)$$

is the error term due to the error in the depth, altitude and attitude errors, and

$$\varepsilon^t = \sum_{j=1}^N \varepsilon_j^t \mu(C_j)$$

is the error due to tidal correction, and $\tilde{\varepsilon}_V$ has been defined before.

The equivalent expression

$$\boxed{\varepsilon_V = \tilde{\varepsilon}_V + \varepsilon_{\Delta} + \varepsilon^{d\mu} + \varepsilon^{\hat{s}}}$$
 (18)

where the definition of $\varepsilon^{\hat{s}}$ is obvious, will also be useful in the sequel, when noise reduction techniques will be studied.

This error decomposition can now be used in the performance criteria of section 5.2 to derive expressions for these criteria in terms of the 1st and 2nd order moments of noise processes:

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \tilde{\boldsymbol{\varepsilon}}_V \\ \boldsymbol{\varepsilon}_\Delta \\ \boldsymbol{\varepsilon}^{d\mu} \\ \boldsymbol{\varepsilon}^{\hat{s}} \end{bmatrix} \quad (19)$$

in terms of which

$$\boldsymbol{\varepsilon}_V = [1 \ 1 \ 1 \ 1] \boldsymbol{\varepsilon}. \quad (20)$$

3.8 Sectional volume estimation

When estimating the sectional volumes, an additional error term must be taken into account, due to the fact that the sampled points do not fall in the lines L_k .

Making a Taylor expansion around πs_j in the direction orthogonal to the observed line,

$$\begin{aligned} \sum_{j=1}^N S_j \boldsymbol{\mu}(\boldsymbol{\pi} C_j) &= \sum_{j=1}^N \left(S(\pi s_j) + (s_j - \pi s_j) \frac{\partial S(\pi s_j + v\ell)}{\partial \ell} \Big|_{\pi s_j} + h.o.t. \right) \boldsymbol{\mu}(\boldsymbol{\pi} C_j) \\ &= \sum_{j=1}^N S(\pi s_j) \boldsymbol{\mu}(\boldsymbol{\pi} C_j) + \sum_{j=1}^N \left((s_j - \pi s_j) \frac{\partial S(\pi s_j + v\ell)}{\partial \ell} \Big|_{\pi s_j} + h.o.t. \right) \boldsymbol{\mu}(\boldsymbol{\pi} C_j) \\ &= \tilde{V}_k - \sum_{j=1}^N \frac{\partial S(\pi s_j + v\ell)}{\partial \ell} \Big|_{\pi s_j} I_j + \sum_{j=1}^N \left((s_j - \pi s_j) \frac{\partial S(\pi s_j + v\ell)}{\partial \ell} \Big|_{\pi s_j} + h.o.t. \right) \boldsymbol{\mu}(\boldsymbol{\pi} C_j) \\ &= \tilde{V}_k - \sum_{j=1}^N \frac{\partial S(\pi s_j + v\ell)}{\partial \ell} \Big|_{\pi s_j} (I_j - (s_j - \pi s_j) \boldsymbol{\mu}(\boldsymbol{\pi} C_j)) + h.o.t. \\ &= \tilde{V}_k + \boldsymbol{\varepsilon}_\Delta + h.o.t. \end{aligned}$$

where the definition of $\boldsymbol{\varepsilon}_\Delta$ is now

$$\boldsymbol{\varepsilon}_\Delta = \sum_{j=1}^N \frac{\partial S(\pi s_j + v\ell)}{\partial \ell} \Big|_{\pi s_j} (I_j - (s_j - \pi s_j) \boldsymbol{\mu}(\boldsymbol{\pi} C_j)).$$

We obtain the same expression for the total error,

$$\boldsymbol{\varepsilon}_{V_k} = \tilde{\boldsymbol{\varepsilon}}_{V_k} + \boldsymbol{\varepsilon}_\Delta + \boldsymbol{\varepsilon}^{d\mu} + \boldsymbol{\varepsilon}^{T\theta} + \boldsymbol{\varepsilon}^{r.\theta} + \boldsymbol{\varepsilon}^t$$

with the new definitions

$$\boldsymbol{\varepsilon}^{d\mu} = \sum_{j=1}^N S_j d\boldsymbol{\pi} \boldsymbol{\mu}_j$$

is an error term due to the deformation of the estimated Voronoï cells wrt to the one defined by the really observed points,

$$\varepsilon^{T\theta} = \sum_{j=1}^N \Delta T_j S_j \mu(\pi C_j)$$

is an error term due to the errors in the attitude angles, and

$$\varepsilon^{r,\theta} = \sum_{j=1}^N T(\theta_j) \varepsilon_j^r \mu(\pi C_j)$$

is the error term due to the error in the depth, altitude and attitude errors, and ε^t has an equivalent definition for the errors in the tidal correction, where πC_j , the Voronoï cells of the projected estimated points replaces the Voronoï cells of the true sampled points in the corresponding expressions for the total volume.

4. Statistical analysis of volume estimation with structural model

In the previous section, we considered the problem of estimating the quantities of interest assuming no prior model for the observed field (the sand banks).

In this section we consider that a *structural model* is available, expressing the sand bank shape as a finite linear combination of known spatial functions, such that the sea depth at time t can be written over the area of study as:

$$S^t(z) = \sum_i \alpha_i^t \phi_i(z), \quad z \in A$$

where $\{\phi_i(\cdot)\}_{i=1}^L$ are *known* functions in \mathfrak{R}^2 (or in \mathfrak{R} in the case only the shape along lines is observed), independent of t , and $\{\alpha_i^t\}$ are unknown coefficients. In Appendix 3 we discuss using wavelets as the basis functions $\{\phi_i\}$, and derive a finite dimensional model for the shape of the Kwintebank along three DECA lines. Using this model enables extrapolation of local measures to the rest of the profile line, and thus estimation, based on a limited interval, of the total volume along that line. We discuss this possibility in section 5.8, where observation strategies are discussed.

Assuming the finite-dimensional model above, the total volume is directly obtained from knowledge of the coefficients α_i^t . Using the definitions given before,

$$V(t) = \int_{A_j} S(z, t) dz = \sum_i \alpha_i^t \int_{A_j} \phi_i(z) dz$$

The Maximum Likelihood (ML) estimate of the volume given the observations is given by

$$\hat{V}_t^{ML} = \sum_i \hat{\alpha}_i^t \int_{S_j} \phi_i(z) dz$$

where $\hat{\alpha}_i^t$ are the ML estimates of the coefficients given the data, determined by the equation

$$\hat{\alpha}_i^t = \arg \max_{\{\alpha_i\}} p\left(Y \mid \left\{\alpha_i\right\}_{i=1}^L\right),$$

and Y denotes all the observed data. The error of the reconstructed volume is

$$\varepsilon_V^{(2)} = \sum_i \varepsilon_{\alpha_i} c_i$$

where ε_{α_i} is the error on coefficient α_i , and we defined the constants

$$c_i \equiv \int_{S_i} \phi_i(z) dz.$$

Maximum likelihood estimators have interesting asymptotic (large sample size) properties. In particular, they are asymptotically unbiased, and - if an efficient estimator exists - their covariance is equal to the inverse of the Fisher Information matrix (Cramér-Rao theorem). It means that we can write for the errors in the coefficients the following equation²:

$$p(\varepsilon_a \mid Y) \rightarrow N(0, J(\alpha)^{-1})$$

where the element (i,j) of J is given by

$$[J(\alpha)]_{ij} = -E \left[\frac{\partial^2 \ln p(Y \mid \alpha)}{\partial \alpha_i \partial \alpha_j} \right].$$

To be able to build the likelihood function, we must condition on the true values of the platform state vector (X), which is not directly observed:

$$p(Y \mid \alpha) = \int_X p(Y, X \mid \alpha) dX = \int_X p(Y \mid X, \alpha) p(X \mid \alpha) dX.$$

The first factor in the integral above is the likelihood of the sensor data for each possible position, attitude and depth surface.

$$p(Y \mid X, \alpha) = \prod_{i=1}^N p(y_i \mid x_i, \alpha)$$

$$p(y_i \mid x_i, \alpha) = p(\varepsilon_i^{\hat{s}})$$

where $\varepsilon_i^{\hat{s}}$ has been defined before.

² $N(\mu, \Sigma)$ denotes a normal distribution with mean μ and covariance matrix Σ .

The second term is the probability of the sensor trajectory for a given sea bottom surface. To proceed further, we must have a model of the controlled platform, that dictates how its position and attitude evolves over time. The pertinent model depends on the type of guidance that is used, and we will continue this study on section 5.8, for each type of guidance strategy considered in the project.

4.1 Pertinence of low-dimensional models for the sand bank

We present here preliminary data analysis results that validate the possibility of identifying from available data this type of low-dimensional models. The plots below show the observed volume for 25 profiles of the DECA lines rG18, rG21 and rH01, and their volumes computed using the coefficients in the 10-dimensional basis for each of these lines presented in appendix 3.1.

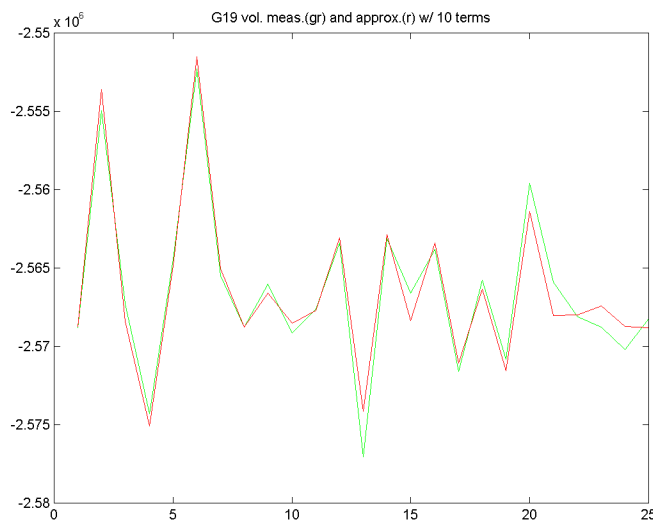


Figure 1: observed total volumes along line rG19 for 25 profiles (green) and their estimates using a model of dimension 10 (red).

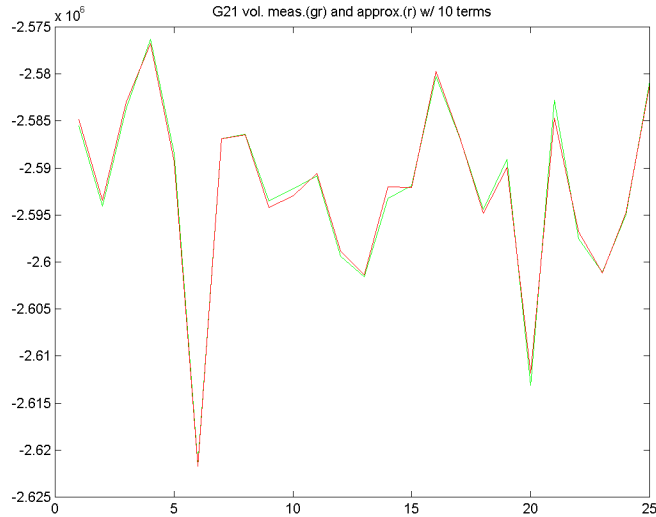


Figure 2: observed volumes for 25 profiles along rG21 (in green) and their estimates using a 10-dimensional model (in red).

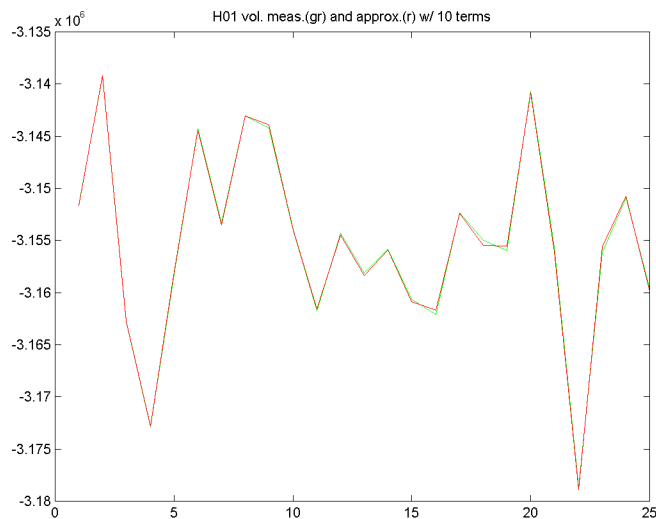


Figure 3: observed volumes along rH01(in green) and their estimates using a model of dimension 10 (in red).

These plots show that the 10-dimensional model effectively capture the data information necessary for determining the sand bank volume. If we consider for instance the plot corresponding to rG19, which from visual analysis (confirmed by MUMM) is a stable bank, we can see that the error due to the un-modelled details of the bank shape is much smaller than the noisy variation obtained from raw data.

5. Statistical analysis of volume estimation with statistical model

In Appendix 2.2 we identified an a priori statistical model for the shape of the profiles along each DECA line, by assigning a probability distribution to the vector of coefficients α of the structural model of the previous section. Using normal distribution model for the model coefficients

$$p(\alpha) = N(\mu_\alpha, \Sigma_\alpha)$$

and assuming white normal observation noise with variance σ_n^2 , the Bayesian estimator that yields the minimum mean-square error estimate of the coefficients is :

$$\hat{\alpha} = \mu_\alpha + \left(\sigma_n^2 I - \Phi_i \Sigma_\alpha^{-1} \Phi_i^T \right)^{-1} S_i^T y_i$$

where y_i are the observations, and Φ_i is the observation operator for the linear model

$$y_i = \Phi_i \alpha + \omega_i = [\phi_1(x_i, y_i) \cdots \phi_l(x_i, y_i)] \alpha + w_i$$

and w_i is observation noise.

The corresponding estimator of the total volume along that track is

$$\hat{V}_i(t) = I_\phi^T \hat{\alpha},$$

and its variance is

$$\sigma_v^2 = I_\phi^T \left(\Sigma_\alpha^{-1} + \frac{1}{\sigma_n^2} \Phi_i^T \Phi_i \right)^{-1} I_\phi$$

In the previous equations

$$I_\phi \equiv \int_{A_i} \phi(\ell) d\ell,$$

is the volume of the basis functions.

Note that this equation can be recursively applied, as more observations are gathered, by replacing the prior model by the current moments of the estimate of the coefficients. When the positions of the sampled points are noisy, matrix Φ_i in the previous equations is not precisely known. In that case, a possible approach is to consider an additional error term (that superimposes to the observation noise) that takes into account the increased uncertainty due to these errors.

6. Statistical Analysis of volume variation

For the analysis of volume variation, we consider that the errors of the two surveys are statistically independent. This assumption holds only if there are no systematic causes of error (it would for instance be contradicted if the positioning errors for the two surveys are correlated). In this case, we can simply assume that the average error is the difference of the average errors of the individual surveys, and its covariance is the sum of the individual covariances. Each 1st and 2nd order moment is computed using the expressions of the previous section.

Note that when precise spatial positioning is not of primary importance, we can use the shape of the sand bank itself to reduce the relative positioning errors of two surveys. We have presented this approach in the document that presents the analysis of the survey data for three DECA lines provided by MUMM. This amounts to applying a pre-processing to the acquired data, to maximize their spatial alignment. Proper quantification of the residual co-registration error is required to enable assessment of its impact on the overall errors when observing volume variations.

A statistical model for the bank shape As presented in Appendix 2.2 provides a means to assess the average and standard deviation of the errors $\tilde{\epsilon}_V$ and $\tilde{\epsilon}_{V_k}$ due to integration on a finite grid, discussed at the beginning of section 5.3. The corresponding expressions are easy to derive, and involve the derivatives of the basis function, and the average values and standard deviations of the coefficients

7. Statistical Analysis of shape estimation

As for the previous section, to be able to infer shape, and its variation, it is important that the depth measures be precisely registered spatially. When observing shape variation, one additional request is to be able to establish the spatial correspondence between the points sampled in two distinct surveys. We consider that a discrete grid of analysis is defined, and that the shape variation is measured in that grid (see the discussion on distances between shapes in section 5.2). This requires the ability to interpolate the points of each survey to the depth values in this reference grid. This interpolation can either be done using any standard polynomial or spline numerical fitting method, or using Bayesian methods for estimating shape of the sand bank given the observed data.

As we said in the previous section, the shape of the bank can be used to co-register two distinct surveys. It can also be exploited in a different manner for reducing the relative positioning errors for a single survey by using contour following techniques, at the additional cost of requiring the robot to observe some parts of the bank more than once. In fact, as we showed before in the context of terrestrial robotics, the fact that we are observing directly a contour of the bank can be used to smooth the positioning estimates in a first step, and enable the estimation of the vector α using the simple additive noise model in a second step. The realization of these two steps is strongly dependent on the observation strategy and will be detailed in section 5.8.

7.1 Use of metrics in size-and-shape spaces

We presented in Section 5.1 the definition of a metric in size-and-shape spaces. We applied this definition to the survey data of three lines of the Kwintebank provided by MUMM. In the figure below we plot the distances between successive observed profiles (after automatic spatial alignment) :

$$d(S_{i-1}^k, S_i^k) / \sqrt{N_{ki}}, \quad i = 2, \dots, 25; \quad k = 'rG19', 'rG21', 'rH01'$$

where N_k is the number of points in k -the profile.

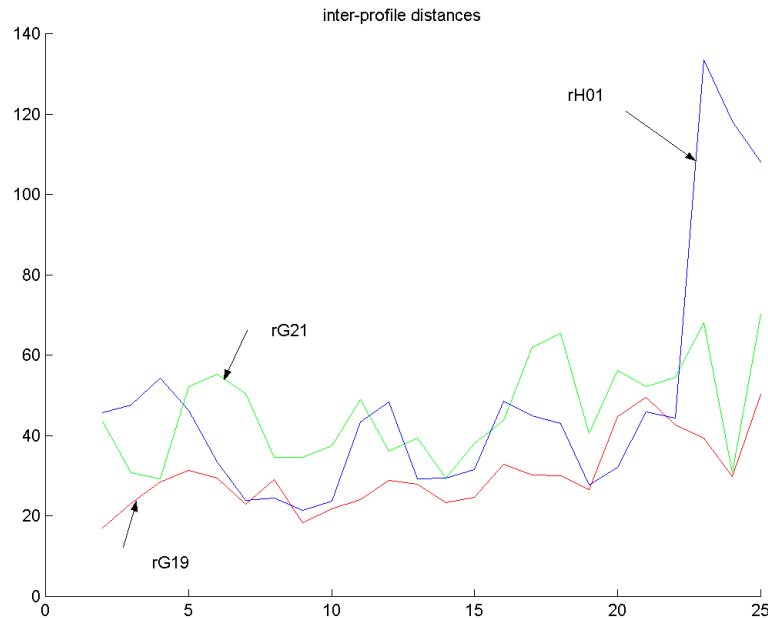


Figure 4 : distance between the shape of successive profiles

The plots above show that, as visual inspection of the profiles has suggested, the line rG19 is the most stable, that it appears that the last profiles of rH01 are outliers – with distances much larger than the subsequent ones – and that the line rG21 is the one that displays the larger variations in shape. Note that even if the profiles have been ordered temporally, the distance between successive profiles is far from constant, and thus that no change rate can be inferred from this plots.

The interest of this metric – or of a local version of it, enabling the assessment of the local changes of the shape of the different regions of the bank – must be further discussed with the end-users.

To complete this analysis, we should identify the distribution induced by measurement noises in the distance measured, and establish a test to decide whether the observed changes are statistically meaningful or not.

8. Sampling strategies

We presented in the previous sections expressions that allow the analysis of the impact of individual error terms on the total error figure for volume computations. The goal of this section is to relate these errors to the sampling strategies and data post-processing algorithms effectively used.

We consider four distinct types of observation strategies:

- **Fixed reference trajectory:** survey over n regular spatial grid: this is the most common approach: desired horizontal and vertical sampling rates are chosen according to prior knowledge of the typical rates of variation of the observed field, and a nominal trajectory is designed with the grid points as way points.
- **Adapted reference trajectory:** in this case, detailed a priori knowledge about the inhomogeneity of the observed field and about its spatial correlation structure is used to design an efficient reference trajectory that minimizes the error of the estimate of interest based on observations along that trajectory.
- **Contour guidance:** the vehicle directly maps selected lines of constant sea depth. This observation mode is specially interested if one wishes to observe shape variations of the bank.
- **Adaptive information observation:** in this case, the vehicle chooses its trajectory on-line on the basis of already observed values, identifying the most interesting unobserved regions.

For all types of observation strategies, and even if we consider as known the true sea-bed surface, the actual vehicle trajectory is random, due to positioning errors (for the first two) and also to the random noise of the sea depth measures (for the last two cases). Our goal in this section, is, for each of the four types of strategies mentioned above, derive expressions that relate the positioning errors of the platform to the errors identified in the previous sections of this Chapter. We will consider each kind of observation strategy separately.

8.1 Fixed reference trajectory

Let us introduce notation first. Let

- $\{p_t^0\}_{t \in T}$ be the nominal vehicle trajectory. We assume that the survey has been designed such that the points $\{p_i^0\}_{i=1}^N$ fall on a regular grid of spacing Δ_g . Note that Δ_g is a two dimensional vector if the total bank volume is being observed and is a simple scalar if only the sectional volumes are to be estimated.
- $\{\epsilon_t^{pos}\}_{t \in T}$ be the error of the navigation system on-board the vehicle:

$$p_i = p_i^a + \epsilon_i^{pos}.$$

The statistical characteristics of the error process are determined by the characteristics of the positioning system : existence and performance of acoustic equipment, quality of inertial sensors, and by the presence of unobserved sea currents.

Our goal is to analyse the impact of these two entities on the error processes affecting our criteria.

We assume a **perfect guidance assumption** and admit that the guidance system achieves

$$p_i = p_i^0$$

i.e., *the estimated position coincides with the reference trajectories.*

In this case, we can assume that the estimated position vehicle is deterministic, coinciding with the nominal one. The size of the Voronoi cells $\mu(C_i)$ and the coefficients I_i appearing in the previous sections can then be computed using knowledge of the nominal grid of points. For the total volume

$$\mu(C_i^0) = \Delta_{g_x} \Delta_{g_y}, \quad I_i^0 = 0.$$

And for sectional volumes

$$\mu(C_i^0) = \Delta_g, \quad I_i^0 = 0.$$

8.2 Intensive acoustic positioning

If the acoustic positioning system is intensively used, we can assume that the positioning errors at sampling points is an inhomogeneous white process, with a variance at each point which is dependent on the accuracy of the measurements of the transponder response times and on the estimates sound velocity. The true positions visited by the vehicle are:

$$p_i^a = p_i^0 + \varepsilon_i^{pos},$$

which we can model as a white noise process, with known mean and known covariance. In fact, expressions for these errors derive easily from analysis of the relative geometry of the vehicle and the baseline array of transponders, and is already an integral part of the present navigation system of our AUV. We do not address the issue of modelling the positioning errors here.

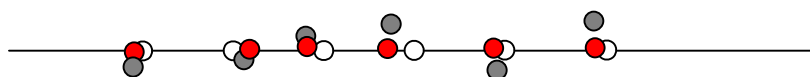
This process determines the following entities entering the error equation:

- $\mu(C_i^a)$ the size of the Voronoi cells of the set of true measured points
- I_i^a
- S_i^a the measured depth values.

To obtain statistical models for these errors, we consider each of them separately first. In the analysis, we assume that the positioning error is normal, and with known covariance matrix Σ_i^{pos} at each point.

Size of the Voronoi cell around point i .

We consider here only the **one-dimensional case**. The next diagram illustrates the nominal points p_i^0 (regular grid of white dots), the true sampled points p_i^a (grey dots on the plane) and their projections on the line, πp_i^a .



The size of the Voronoi cells in this case (the length of the intervals defined by the midpoint between the projected sample points) is given by

$$\mu(\pi C_i^a) = \frac{\pi p_i^a - \pi p_{i-1}^a}{2} + \frac{\pi p_{i+1}^a - \pi p_i^a}{2} = \frac{1}{2} (\pi p_{i+1}^a - \pi p_{i-1}^a)$$

Writing the nominal grid points as

$$p_i^0 = p_0 + i * \Delta_g \vec{u}$$

where p_0 is the first observed point on the line and \vec{u} is the unit vector along the line, yields

$$p_i^a = p_0 + i * \Delta_g \vec{u} + \varepsilon_i^{pos}$$

and its projection on the line is

$$\pi p_i^a = p_0 + i * \Delta_g \vec{u} + \vec{u} \vec{u}^T \varepsilon_i^{pos} = p_0 + i * \Delta_g \vec{u} + \alpha_{\varepsilon_i} \vec{u}$$

where we have defined

$$\alpha_{\varepsilon_i} = \vec{u}^T \varepsilon_i^{pos}$$

which , under our assumptions, is a normal random variable with zero mean, and variance $\vec{u}^T \Sigma_{pos} \vec{u}$ that gives the deviations with respect to the nominal points.

We can now write the size of the Voronoi cells in terms of the errors:

$$\mu(C_i^a) = \Delta_g + \frac{1}{2} (\alpha_{\varepsilon_{i+1}} - \alpha_{\varepsilon_{i-1}}),$$

and we can thus identify

$$d\mu_i = \frac{1}{2} (\alpha_{\varepsilon_{i+1}} - \alpha_{\varepsilon_{i-1}}),$$

which implies that $d\mu_i$ is a normal random variable with zero mean

$$E(d\mu_i) \equiv 0$$

and variance

$$E[d\mu_i^2] = \frac{1}{4} (\sigma_{i+1}^2 + \sigma_{i-1}^2) \cong \frac{1}{2} \sigma_i^2$$

if the noise field statistics varies slowly along the grid size.

Note that the values of $d\mu_i$ are correlated for points spaced of $2\Delta_g$:

$$E[d\mu_{i-1}d\mu_{i+1}] = \frac{1}{4} E[(\alpha_i - \alpha_{i-2})(\alpha_{i+2} - \alpha_i)] = -\frac{1}{4} E[\alpha_i^2] = -\frac{1}{4} \sigma_i^2.$$

In summary:

$$E[d\mu_i d\mu_j] = \begin{cases} \frac{1}{4} (\sigma_{i-1}^2 + \sigma_{i+1}^2), & i = j \\ \frac{1}{4} \sigma_{\frac{i+j}{2}}^2, & |i-j| = 2 \\ 0, & \text{otherwise} \end{cases}$$

Centroid of the cells

To compute the statistics of the centroid of the cells, we remember its definition:

$$I_i^a = \int_{\pi C_i^a} (\ell - \pi p_i^a) d\ell = \frac{\pi p_i^a - \pi p_{i-1}^a}{2} \int_{\frac{\pi p_i^a + \pi p_{i-1}^a}{2}}^{\pi p_i^a} p(\ell - p_i^0 - i\Delta_g - \alpha_{\varepsilon_i}) d\ell$$

Using the expressions for the sampled points in terms of the nominal points, and carrying out the integration, we obtain

$$I_i^a = \frac{1}{8} \left[\alpha_{\varepsilon_{i+1}}^2 - \alpha_{\varepsilon_{i-1}}^2 + 2\Delta_g (\alpha_{\varepsilon_{i+1}} - \alpha_{\varepsilon_{i-1}}) + 2\alpha_{\varepsilon_i} (\alpha_{\varepsilon_{i-1}} - \alpha_{\varepsilon_{i+1}}) \right]$$

Assuming again that the errors are uncorrelated yields and zero mean at each point

$$E[I_i^a] = \frac{1}{8} (\sigma_{\varepsilon_{i+1}}^2 - \sigma_{\varepsilon_{i-1}}^2)$$

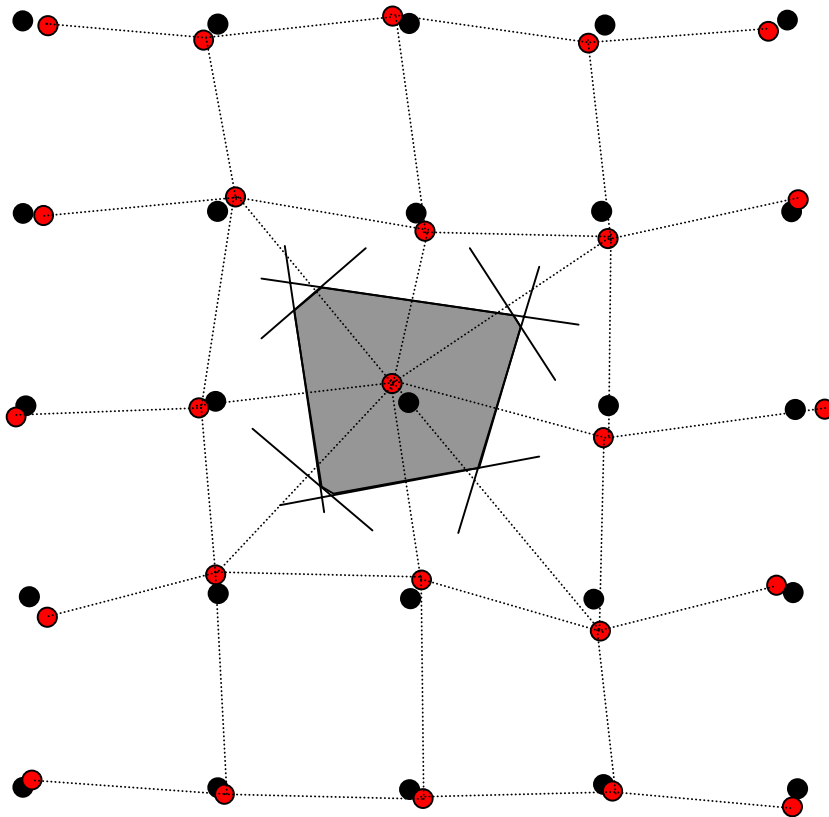
and the second order moments can also be easily computed in terms of the variance of the reduced positioning errors.

We formulate below the problem for the case of **2-dimensional observation surface**. We assume that in this case the nominal observation grid is a rectangular grid aligned with the local Cartesian coordinates, such that the generic nominal observation point is given by

$$p_{i,j}^0 = \begin{bmatrix} p_{i_x}^0 \\ p_{j_y}^0 \end{bmatrix} = \begin{bmatrix} p_x^0 \\ p_y^0 \end{bmatrix} + \begin{bmatrix} (i-1) & 0 \\ 0 & (j-1) \end{bmatrix} \begin{bmatrix} \Delta_x \\ \Delta_y \end{bmatrix}$$

The points that are actually observed are noisy versions of the nominal points:

$$p_{i,j}^a = p_{i,j}^0 + \varepsilon_{i,j}^{pos}.$$



The next figure illustrates the deformation of the grid nominal grid (in black in the figure) due to the positioning errors, resulting in the grid shown in red. The Voronoi cell associated to the central point, which was a square for the nominal grid, has now 7 sides. We are interested in computing the mean value and the variance of the area of the cells.

Consider cell (i,j) . Its geometry is determined by the 8 points in its neighbourhood (we assume that the error is small, such that the probability that the initial ordering be disturbed is zero). Write the position of those points in spherical coordinates with respect to position of the observed point (i,j) as a complex number;

$$s_{i+\delta, i+g\lambda}^a \equiv p_{i+\delta, i+g\lambda}^a - p_{i,j}^a = r_{i+\delta, j+\lambda} e^{j(\theta_{i+\delta, j+\lambda})}$$

If the positioning noise is Gauss, these variables are complex normal, with mean

$$p_{i+\delta, j+\lambda}^0 - p_{i,j}^0 = \begin{cases} \Delta_y, & \delta = 0 \\ \Delta_x, & \lambda = 0 \\ \sqrt{\Delta_x^2 + \Delta_y^2}, & |\delta| = |\lambda| = 1 \end{cases}$$

and covariance matrix

$$\begin{aligned}
\Sigma_{\delta,\lambda;\delta_1,\lambda_1} &= E\left[(\alpha_{i+\delta,j+\lambda} - \alpha_{i,j})(\alpha_{i+\delta_1,j+\lambda_1} - \alpha_{i,j})\right] \\
&= E\left[\alpha_{i+\delta,j+\lambda}\alpha_{i+\delta_1,j+\lambda_1}\right] - E\left[\alpha_{i+\delta,j+\lambda}\alpha_{i,j}\right] - E\left[\alpha_{i+\delta_1,j+\lambda_1}\alpha_{i,j}\right] + E\left[\alpha_{i,j}^2\right] \\
&= \delta_{\delta-\delta_1}\delta_{\lambda-\lambda_1}\sigma_{i+\delta,j+\lambda}^2 + \sigma_{i,j}^2 = \sigma_{i,j}^2 I + \text{diag}\left(\sigma_{i+\delta,j+\lambda}^2\right)
\end{aligned}$$

where δ_x is the Kroenecker symbol.

There are several possibilities for the number of neighbours that will define the Voronoi cell around each point, depending on the relative size of the errors. The determination for the values of the average size of the Voronoi cell and for the inertia parameter are quite involved except for the uninteresting case where the error process is isotropic and spatially homogeneous, and is currently under study.

Depth measures

Since

$$S_i^a = S(p_i^0 + \varepsilon_i^{pos})$$

we can use the formula for the probability distribution of a function of two random variables to find the exact distribution of S_i^a in terms of the distribution of the positioning error.

A simpler approach considers only a first order Taylor expansion of the sea bottom surface locally around each point, in which case, we obtain that the depth measurements are Gaussian random variables with mean

$$E[S_i^a] = S(p_i^0)$$

and

$$E\left[\left(S_i^a - S(p_i^0)\right)^2\right] = \nabla_S^T \Sigma_{pos} \nabla_S$$

where $\nabla_S^T = [\partial S / \partial x \quad \partial S / \partial y]$.

8.3 Sparse acoustic positioning

When acoustic positioning is not possible at every observed point, our assumption concerning the independence of the positioning error does not hold, requiring the consideration of cross-covariances in the previous analysis. Expressions for these covariances, considering that in between acoustic fixes the vehicle uses inertial navigation has been done in the past, and is not addressed here.

8.4 Using prior models of the bank shape

As we mentioned previously, the quality of the volume reconstruction, or more generally, of the observation of the sand bank depends strongly on the accuracy of the estimates of the location of the sampled points. When a statistical a priori model of the bank is available

(see discussion in Appendix 2) we may use the depth measures them selves to correct the estimate of the sampled point obtained from other sensors (acoustics, GPA, inertial,...).

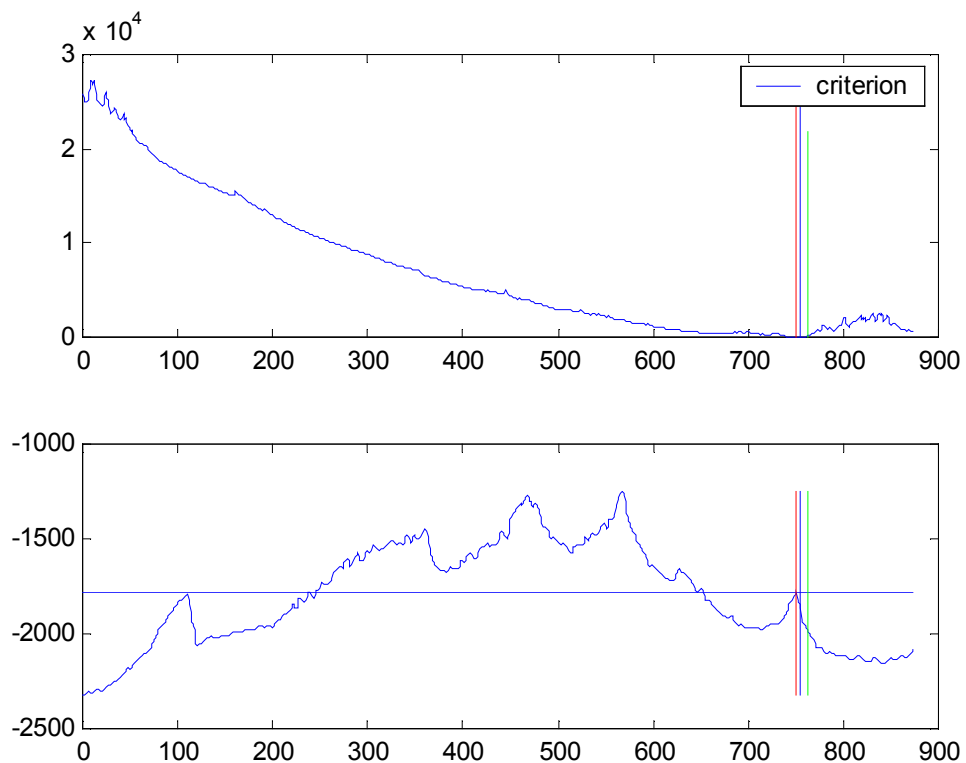
We formalised the problem as a mixed Maximum Likelihood (for the position) and Maximum a posteriori (for the bank surface). We consider observations along the DECA lines, for which we have built an a priori model. For the simple case of scalar measures (position is in this case a scalar coordinate along the line), and assuming Gaussian observation noises of zero mean both for depth and position, we obtain that the best estimate of the position, given the depth measure is

$$\hat{x} = \arg \min_x \left\{ \frac{\sigma_d^2}{\|\Phi(x)\|^2} \left(\frac{R_s(x)}{R_s(x) + \sigma_d^2} \right)^2 (d_{meas} - \bar{S}(x))^2 + \frac{1}{\sigma_x^2} (x_{meas} - x)^2 \right\}$$

where

- x_{meas}, d_{meas} are the measured position and depth
- σ_s^2, σ_x^2 are the covariance of the depth and position measures
- $\Phi(x)$ is the vector of basis functions of the priori model taken at point x
- $S(x), R_S(x)$ are the mean and variance of the model for point x

The simple example below shows that the prior model for depth can effectively correct the measured positions.



In the previous figure, the vertical lines indicate: the true position (red), the measured position (green) and the estimated position using the criterion given above. As we can see use of the measured depth allowed effective correction of the large positioning error. We note that the utility of the depth measure (the amount of effective correction that it allows) depends on the variability of the prior model around that point (expressed by the term $R_S(x)$), and on the shape of the bank in the measured region (the rapidity of variation of the surface). We derived analytical expressions for the confidence of the position estimate using this approach, which will be detailed in a future deliverable, along with the multi-variable version of these results. The important point is that it enables us to identify nominal observation trajectories for along which the positioning error will be minimal (increasing our confidence on the estimate of the observed bank shape).

After the position estimate is obtained, we can use the depth information to increase our knowledge of the bank surface (by estimating the coefficients of the linear model). This possibility will be further exploited in the future.

8.5 Adapted reference trajectory

By this generic name we designate sampling trajectories that are designed on the basis of knowledge of the possible appearances of the sand bank. This knowledge comes both from past existing survey data and from physical analysis of the dynamics of the sand bank, and knowledge of the forces acting upon it.

When the goal is to observe each DECA line, more precisely to estimate its volume and to assess their variability (identify areas where sand is accumulating, and those where it diminishes), then the discussion of a sampling strategy is not really relevant: the vehicle must always track each DECA line. One can still discuss, however, and in the light of the modelling study presented in Appendix 2, which areas of the line it should sample with more precision, or whether if sampling only a limited region would enable the prediction of the entire line, total volume estimation and, eventually, of what other regions still need, or not, to be sampled. Note that the issue of determining an appropriate sampling rate seems rather irrelevant in this context, as the temporal sampling rate of the basic sensor (acoustic device) is usually a fixed parameter during the entire survey. Only velocity of the baseline platform, either an oceanographic vessel or an autonomous robot, along the track can then be adjusted. For autonomous vehicles, the speed is principally determined by considerations involving the total mission length, since energy consumption varies with the cube of speed. For surface ships, adaptive variation of the speed during survey seems unrealistic.

We will address below the following question in the context of designing a reference trajectory for surveying the DECA lines:

Problem: *Which interval of length l provides, according to the identified model, the largest information with respect to the total volume? What is the uncertainty of the volume estimate given those observations? Can we repeat this step iteratively, identifying the next better region on the basis of the already observed regions?*

We presented in section 5.5 expressions for the error of the volume estimates given a set of data points. Using these equations, we now search for the interval of a given length that minimizes the error of the estimated volume, by computing

$$\sigma_v^2(S) = I_\phi^T \left(\Sigma_c^{-1} + \frac{1}{\sigma_n^2} S^T \Phi^T \Phi S \right)^{-1} I_\phi$$

considering that the observed points are a subset of the line A_i of the form

$$S = \{s + i\Delta \mid s \in A_i, i \in \{1, \dots, L\}\}$$

and studying the variation of $\sigma_V^2(S)$ as a function of the initial point s .

The next plots show the standard deviation of the error of the volume estimation for different choices for the location of the observed interval, for the three DECA lines used in this study, considering intervals of fixed length $L=100$.

We see that for DECA line **rG19**, which is very smooth, the best interval is on the steepest slope region, while for lines rG21 and rH01 it falls on the dune region in the centre of the bank. Note that the variation of the error with the observed interval is larger for the last two cases (about 130 and 250), for which the shape of the bank has larger variability from profile to profile, than for rG19 (about 60).

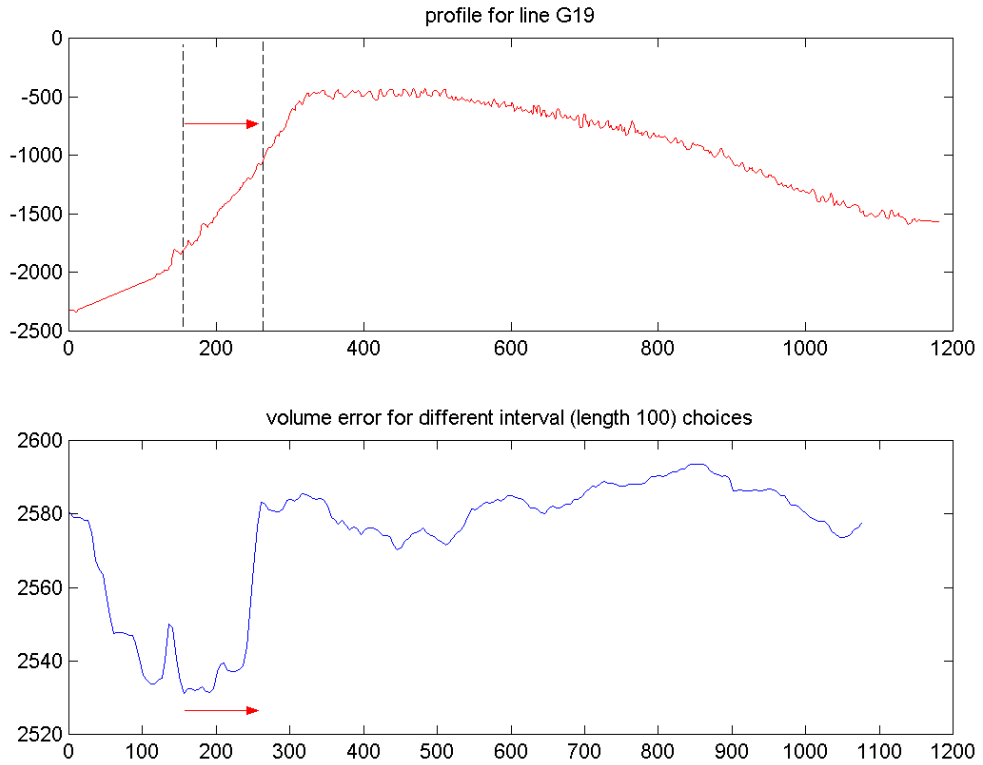


Figure 5: evolution of the predicted volume estimation error (bottom) and best observation region (top) for line rG19.

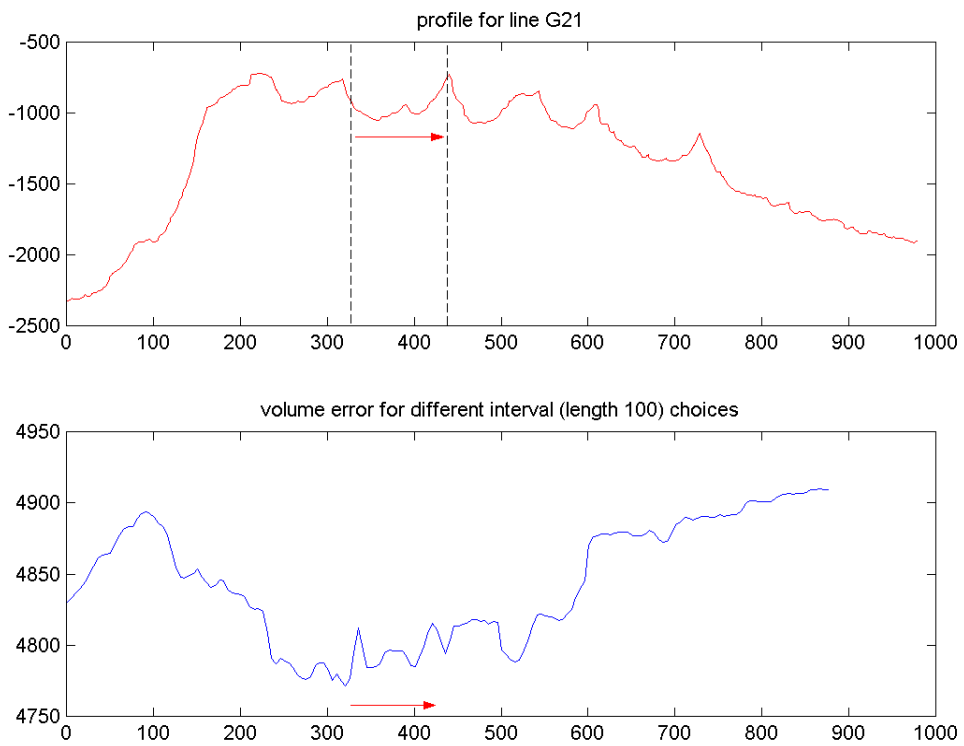


Figure 6: evolution of the predicted volume estimation error (bottom) and best observation region (top) for line rG20.

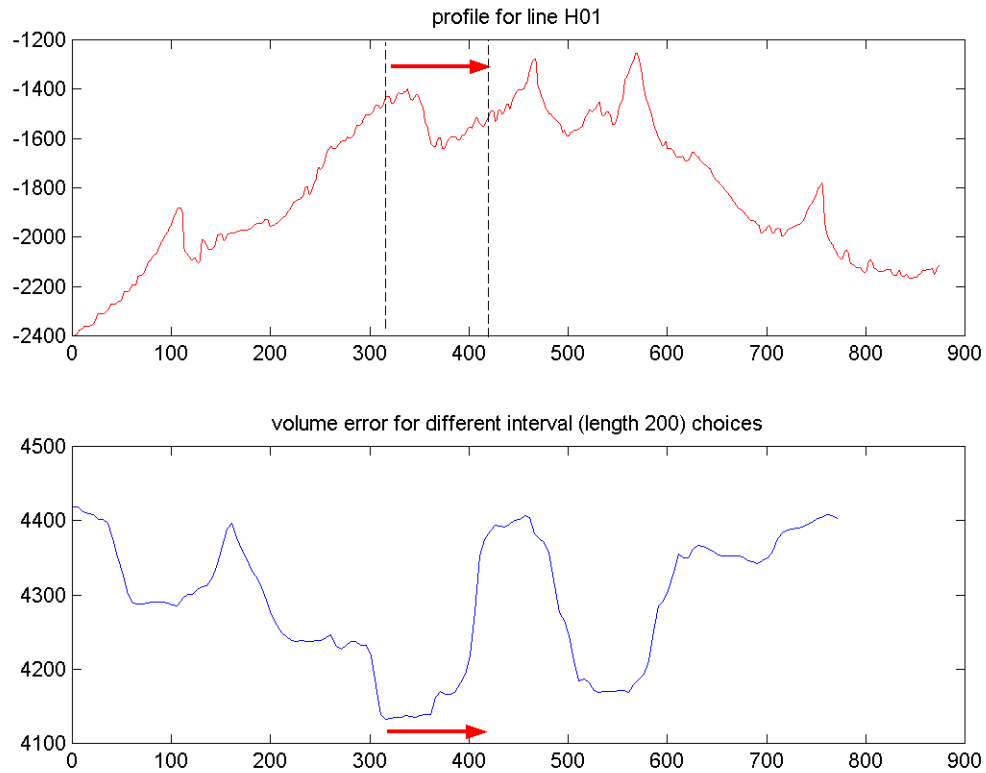


Figure 7: evolution of the predicted volume estimation error (bottom) and best observation region (top) for line rGH01.

Length of observation interval

To study the influence of the length of the observation interval, we plotted the minimum value of this error as a function of the length of the interval. We see that while when we increase the observation interval to the entire line the improvement for rG19 is only about 140, it is on the order of 450 for both rG21 and rH01.

We consider the variation of $\sigma_V^2(S)$ for observation sets of the form

$$S(s, L) = \{s + i\Delta \mid s \in A_i, i \in \{1, \dots, L\}\}$$

and considering

$$\sigma_V^2(L) = \min_s \sigma_V^2(S(s, L)).$$

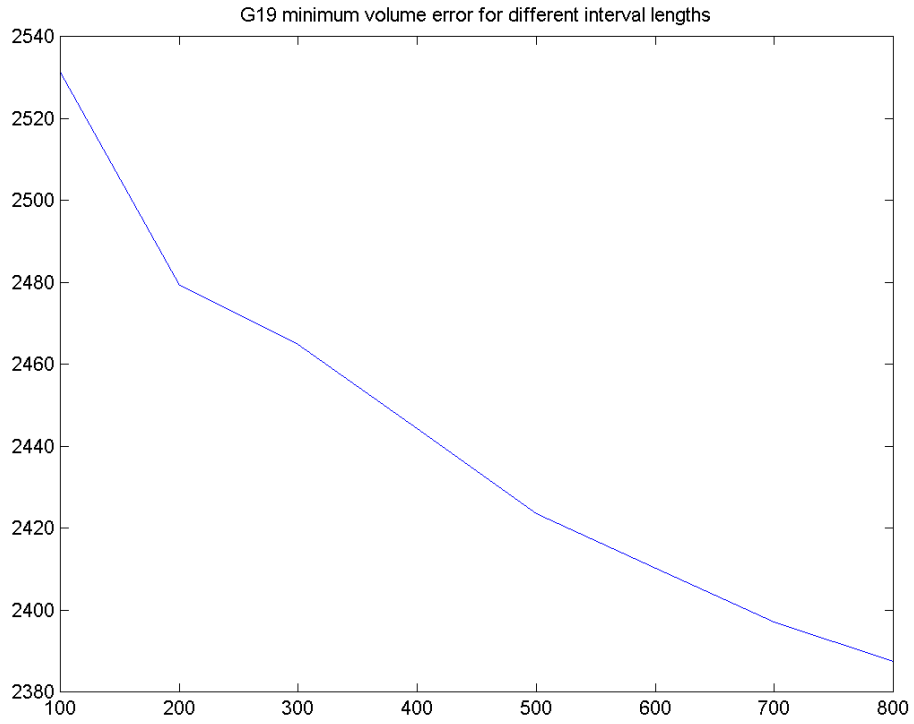


Figure 8: $\sigma_V^2(L)$ for line rG19.

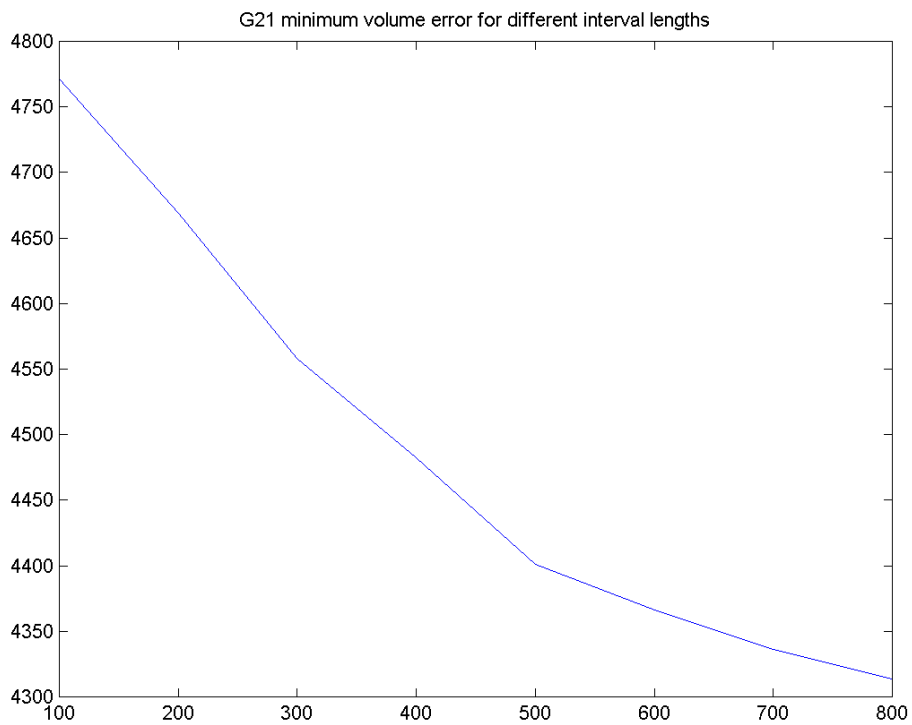


Figure 9: $\sigma_V^2(L)$ for line rG21.

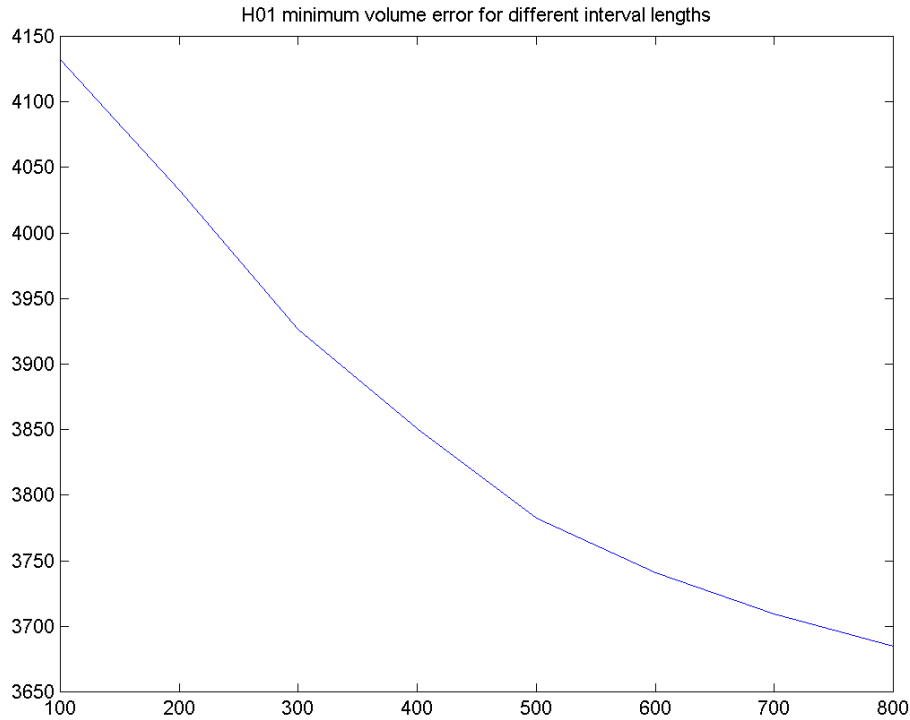


Figure 10: $\sigma_V^2(L)$ for line rH01.

We can see that the line whose volume is estimated with better accuracy is rG19, and the one whose estimate is the most uncertain is rG21. This is somewhat surprising, since rH01 seems at first glance to be the line that undergoes the most profound modifications.

Note that the errors predicted here assume the error model learned from the available survey data, presented in Appendix 2.2, and do not stem from the detailed error budget analysis that we presented in section 5.3. Note also that use of the low dimensional model enables noise reduction when compared to simple integration of the raw observed values, as the following table shows.

DECA line	rG19	rG21	rH0z
Sample noise (std)	415.6	476.1	388.1
Std (raw)	4.9042e+005	4.6559e+005	3.3922e+005
Asy. Std w/ model	2.3515e+003	4.2894e+003	3.6802e+003

The small differences between the values observed for the different DECA lines are mainly due to the distinct sample noise levels that have been estimated for each one (true noise level plus un-modelled components). Note that the units are given in cm (for the 1st

line) and in cm.m for the variances of the integral variables. The gain obtained from using the model even over short data sequences is much larger than the one that is obtained by increasing the observation interval to the entire line.

Note that the standard deviations of the observed values (variance of the sum of the profiles) is $2.6571e+003$ for rG19, $5.0084e+003$ for rG21, and $4.5397e+003$ for rH01, still larger than the values obtained using the model for small observation intervals.

Spatial sampling rate

We studied also the influence of the spatial sampling rate (the basic rate assumed is 2 meters) on the accuracy of the volume estimates, by considering observation sets of the form

$$S(k) = \{ik\Delta, i = 1, \dots, \lfloor N/(k\Delta) \rfloor\}$$

and computing

$$\sigma_V^2(k) \equiv \sigma_V^2(S(k)).$$

The plots below show the standard deviation $\sigma_V(k)$ of the volume estimates using increasingly sparse sampling grids, multiple of the original 2 meters grid.

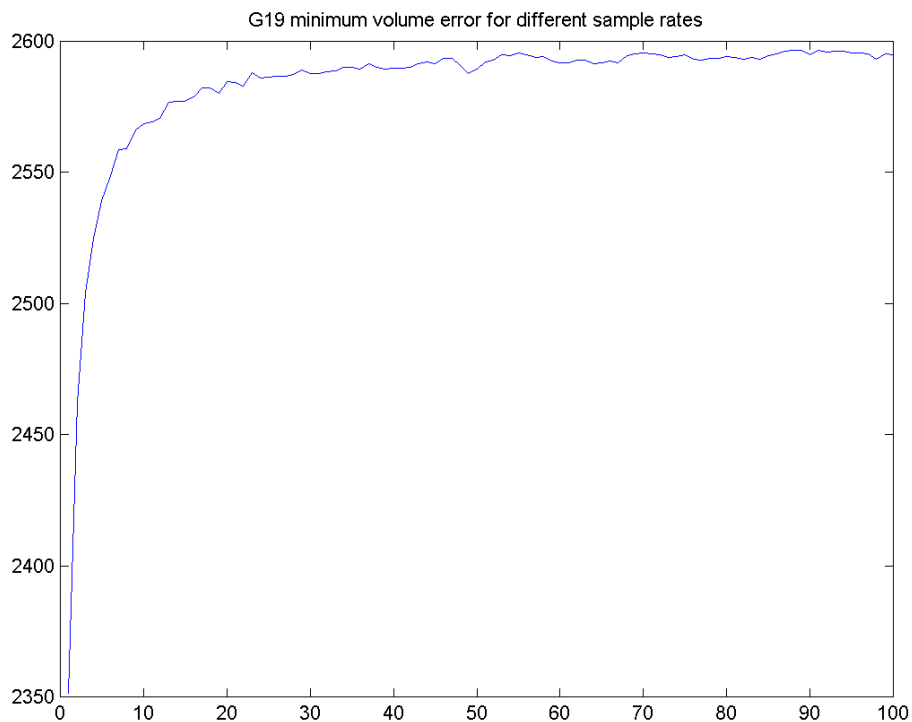


Figure 11: $\sigma_V(k)$ for line rG19.

Note that the asymptotic (coarse sampling) limit on this plot is on the order of the covariance of the volume estimates for this line (which is stable, enabling the

determination of the covariance of the raw estimates of volume), showing that fine sampling does not increase precision on the determination of the volumes. The explanation for this fact is that the high frequency components of the bank shape, which have very small amplitudes compared to the lower frequency ones, have an effective signal-to-noise ratio which is very small, making them, in practical terms, unobservable.

This effect is still more pronounced for the other two profiles, for which the asymptotic values of the statistical method are considerably lower than the observed variance.

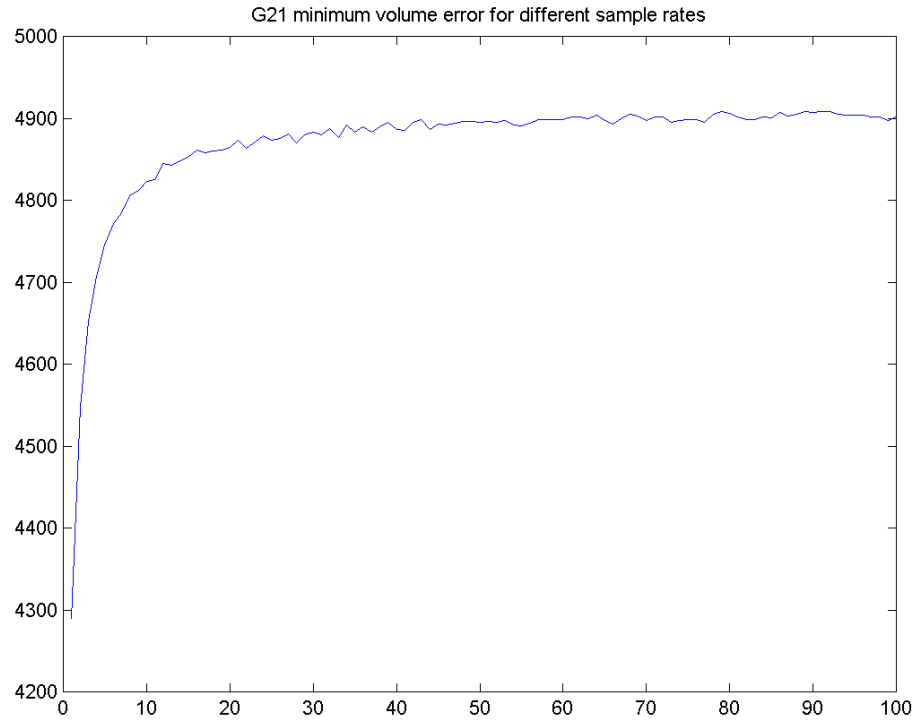


Figure 12 $\sigma_V(k)$ for line rG21.

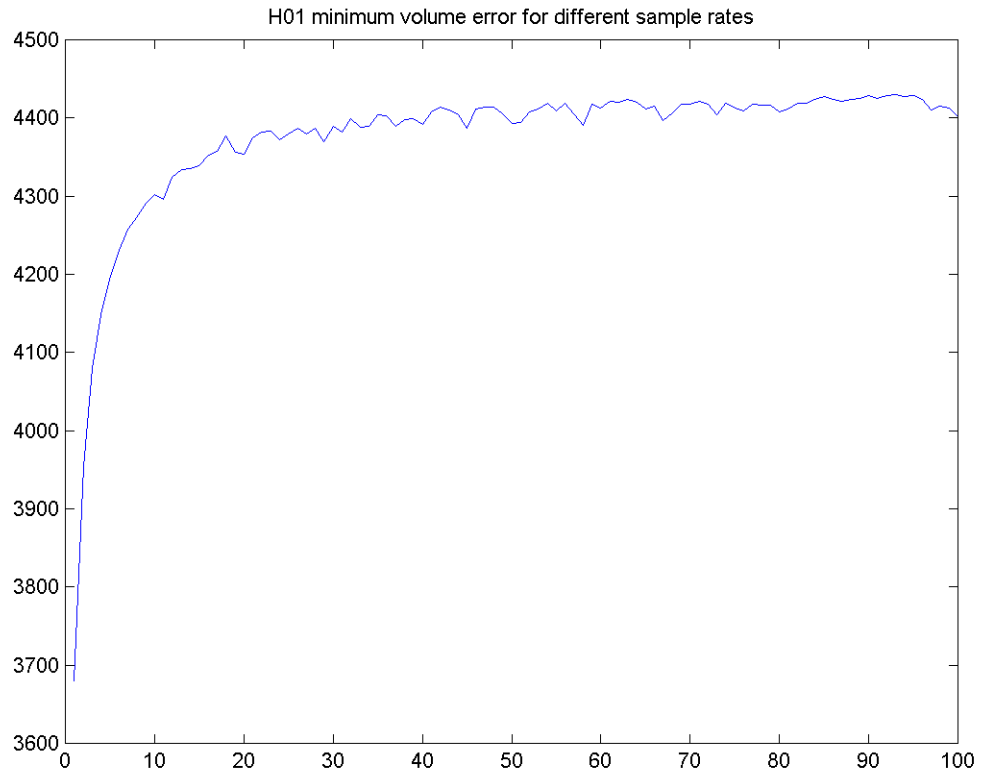


Figure 13 $\sigma_V(k)$ for line rH01.

We see that the error evolves exponentially to an asymptotic value which is still much smaller than the error obtained by straightforward integration of the raw data, showing that if a residual error can be tolerated, the sampling rate can be significantly decreased.

Note that the study that we performed here along a single dimension, based on the statistical model built for 3 DECA lines on the basis of 75 learning examples, could equally be performed for the entire surface, if a consistent set of “snapshots” of the bank was available. Instead of 1D wavelets, the study would be based on 2D wavelets, and a reduced dimension statistical model would be built for the coefficients of the wavelets. Existence of this model would allow us to perform an interesting study on the design of adaptive sensing strategies, where the vehicle would decide on where to move next in order to obtain the maximum possible increase in the confidence on its current estimate of the total volume of the sand bank, along the same lines as we used on our study of the observation of current fields in [oceans2000]. In the case of one dimensional manifolds, and given the continuity and differentiability of the admissible trajectories of the vehicle, there is no space left for optimising the survey trajectory, which must necessarily coincide with the surveyed line, the only possibilities left being either to observe just a subset of the entire line or to sub-sample it, as we did above.

8.6 Contour guidance

The ability of driving an autonomous platform on line, depending on the value of the signals acquired by its sensors, enables direct tracking of the shape characteristics of the sand banks. This feature seems desirable especially when the main interest of the user is to observe the shape of the bank, or its variations.

Consider the recursive estimator of the vector of shape coefficients, and add as unknown to the estimation problem the initial localization of the vehicle when it starts observing the contour and a locally stable unobserved displacement (accounting for unobserved currents).

Consider the problem of choosing the bank region such that the vehicle is able to rapidly estimate its initial position and the external disturbance. It can be shown that the best regions of the sand bank for this purpose are those that show the least variability (the most stable regions).

It is important to note that the vehicle is using its sensor (altitude and depth) data to guide himself along the contour level, but we assume that its positioning information comes from independent sensors (on-board inertial sensors, external acoustic positioning sensors). It does not use the map to locate itself.

If intensive external positioning is possible, the location and scale observation model presented in section 5.7 can be used to obtain optimal estimates of the coefficients of a reduced-dimension model and thus to extrapolate the complete sea bed surface.

The major advantage of this guidance mode with respect to the other discussed in this section is that it allows the exploitation of the shape of the bank itself to eliminate biases due to unobserved velocity components during inertial navigation periods.

Assume that the control algorithms perform well, such that we can assume that the actual vehicle position when tracking a contour line is centred around the true line, with deviations of known variance (coming from analysis of the perception based closed loop control system):

$$(x_i^a, y_i^a) = (x_i^{Lk}, y_i^{Lk}) + \varepsilon_i^{cont}$$

For inertial positioning, we can assume that

$$(x_i y_i) = (x_i^a, y_i^a) + \varepsilon_i^{pos} = (x_i^{Lk}, y_i^{Lk}) + \varepsilon_i^{pos} + \varepsilon_i^{con}$$

where the positioning error is a process with independent increments (but not necessarily stationary) of known statistics. Assume now that we can make the vehicle travel twice along the same region of the contour (this is always possible: either it goes around a closed line, or it makes a U-turn and tracks a second time the same line, in the opposite direction). For simplicity, we assume in this study that in the second time it visits exactly the same points of the line:

$$(x_i^{Lk}, y_i^{Lk}) \Leftrightarrow (x_i^{(1)}, y_i^{(1)}) (x_i^{(2)}, y_i^{(2)})$$

where the correspondence is made by

$$\begin{aligned} (x_i^{(1)}, y_i^{(1)}) &= (x_i, y_i) \\ (x_{N-i}^{(2)}, y_{N-i}^{(2)}) &= (x_{N+\delta+i}, y_{N+\delta+i}) \end{aligned}, \quad i = 1, \dots, N$$

when it makes a U-turn (N is the number of sample sin the contour and δ is the samples taken during the U-turn, which have no correspondent on the two turns - , and by

$$\begin{aligned} (x_i^{(1)}, y_i^{(1)}) &= (x_i, y_i) \\ (x_i^{(2)}, y_i^{(2)}) &= (x_{N+i}, y_{N+i}) \end{aligned}$$

when it closes a turn around a closed contour line.

We can now solve for the maximum likelihood estimator of the alignment index N (for the two series and, if required, the shift δ) and for the true visited points:

$$\hat{N}, \left\{ \hat{x}_i^{Lk}, \hat{y}_i^{Lk} \right\} = \max p \left(\left\{ x_i^{(1)}, y_i^{(1)} \right\} \left\{ x_i^{(2)}, y_i^{(2)} \right\} N, \left\{ x_i^{Lk}, y_i^{Lk} \right\} \right).$$

Or, for U-turn trajectories

$$\hat{N}, \hat{\delta}, \left\{ \hat{x}_i^{Lk}, \hat{y}_i^{Lk} \right\} = \max p \left(\left\{ x_i^{(1)}, y_i^{(1)} \right\} \left\{ x_i^{(2)}, y_i^{(2)} \right\} N, \hat{\delta}, \left\{ x_i^{Lk}, y_i^{Lk} \right\} \right)$$

The performance of these estimators can be assessed by computing the Fisher information for the problem, and can be easily seen to depend on the correlation function of the contour line itself. We see thus that the performance of these estimates depends strongly on the characteristics of the observed field. Using prior models for the shape of the bank, we can then compute expected values of the performance of the spatial registration method.

The smoothed estimates obtained in this way must then be used in the expressions given in section 5.7 to produce estimates of the shape of the bank, and, if required, of its total volume. The performance analysis of the previous sections apply there, if we replace the raw measurement errors by the errors of the smoothed estimates.

Note that this approach decomposes the estimation problem into two steps: in the first one the unknown but fixed shape of bank is used to filter out positioning drifts. In a second step, the corrected trajectory is used to infer the shape of the bank and its total volume, using the prior model of the bank shape.

8.7 Adaptive information observation

We consider in this subsection a last kind of vehicle guidance, where the statistical model is fully exploited to determine, on-line, the vehicle trajectory, according to the previous observed data, and its confidence on its current estimate of its goal. This observation mode is specially interesting when the user goal is dependent on only a subset of the bank, for instance, the location of a given iso-depth line. Otherwise, when the user goal is dependent on the entire surface (as it is the case for the volume evaluation), the best sampling trajectory can be determined off-line as we discussed in the section "Adapted reference trajectory" above.

The approach has been presented before in [Oceans2000] applied to the estimation of current fields, for a mission of delineation of lines of constant current strength. It involves predicting the information gain yield by visiting unobserved regions of the bank with respect to the knowledge of the feature of interest, and recursively guide the vehicle to those that provide the largest gain.

Application of this kind of goal is interesting only if a complete 2D model of the sand bank shape is identified, allowing the vehicle to move in each horizontal plane. Until now, only 1D models for three DECA lines have been identified. We hope, however, to be able in the near future to build equivalent 2D models – at least linking the values of distinct DECA

lines at the same time - and evaluate the interest of this kind of guidance in the context of this application.