# Using Field Subspaces for On-Line Survey Guidance

Maria-João Rendas  Matteo Perrone

Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis (I3S)
2000 Rte de Lucioles
06903 Sophia Antipolis cedex, FRANCE
{rendas,mperrone}@i3s.unice.fr

***Abstract-*** **We address the problem of surveying of oceanic parameters using autonomous instrumented mobile platforms. As an example, we consider the problem of current mapping in coastal areas. We study the impact on survey efficiency of using *a priori* knowledge concerning the surveyed field for on-line guidance of the sensors, as an alternative to the classical approach of executing a predefined trajectory, or to the more recently proposed perception-driven observation strategies. Availability of this *a priori* model enables extrapolation of the measurements, as well as the determination of the information yield by future observations, allowing the search for the *best next observation point*. In the paper, we present simulation results of the proposed on-line guidance based on information gain, and compare its efficiency to standard survey strategies.**

## I. INTRODUCTION

In this paper we address the problem of autonomous surveying of oceanic parameters using instrumented mobile platforms. As an example, we consider the problem of current mapping in coastal areas. Traditionally, surveys are conducted along a series of regular transects covering the region of interest, with inter-transect spacing being dictated by the characteristic variations of the observed field. In this paper we consider the case when *a priori* information concerning the surveyed field exists - as a result of previous surveys, or of mathematical models - and study the impact on survey efficiency of use of statistical models learned from this knowledge for on-line guidance of the sensors.

The rationale behind our approach is that most natural fields of interest are strongly inhomogeneous, defining spatial patterns dictated by the geographic and environmental surrounding conditions. In this sense, of all possible spatial variation patterns inside a given area, only a very limited number actually occur, that is, the set of actual possible observations has reduced dimensionality. This means that the values of the field in some regions are highly correlated to its values on some other regions, and thus observation of the latter should allow its prediction over the former. In this case, actually requiring the vehicle to cover both regions is a waste of survey time and vehicle energy. As an alternative to systematic covering of the region of interest, we propose to use this *a priori* knowledge about the spatial correlation of the observed field to guide the sensor, on-line, to the regions that yield the largest information gain about the overall field, or the feature of interest to the user.

*A priori* knowledge about natural fields usually exists in the form of maps corresponding to different exogenous conditions, representing a large amount of data. The first problem for implementing on-line guidance based on this information is its efficient representation. We propose to compress it using the notion of field subspaces. Each map (over a discrete grid) is represented as a column of a large dimensional matrix collecting all available data. Singular Value Decomposition of this matrix yields the low-dimensional subspace on which all maps can be approximated up to a given error: the *field subspace*. In our study, we use 51 maps of the horizontal current components at the mouth of the river Rhone, predicted by mathematical models over a grid of 15x22=330 points. The identified field has dimension 28, for an error of .05%, and can be further decreased if larger uncertainty can be tolerated (this threshold should be set according to accuracy of measurements).

Availability of this (linear) *a priori* model enables extrapolation of the measurements in one sub-region to its values on unobserved regions, as well as determination of the associated uncertainty by simple matrix computations. Moreover, the determination of the information yield by *future* observations can also be determined, allowing the search for the *best next observation point*. In the paper, we present simulation results of the proposed on-line guidance based on information gain, and compare its efficiency to other survey strategies.

The paper is organized in the following way. In Section II we assess the problem of knowledge representation, describing the subspace model used. In Section III we present how this model allows extrapolation of measures from one area to another, and in Section IV how the information gain of each observation can be determined.

Finally, we present in Section V an algorithm able to guide the sensor to the most informative areas of the sampled region, and in Section VI we present some conclusions and ideas for future work.

## II. KNOWLEDGE REPRESENTATION

In this section we present the mathematical model of the a priori knowledge about the observed field, extracted from learning examples.

### A. Prior Knowledge

As we said previously, we apply our approach to the observation of current fields by an autonomous sensor. More precisely, we consider the observation of the field of ocean currents at the mouth of the Rhône river in the South-East of

France. The prior knowledge consists on current maps over a $15\times22$ horizontal point grid under a variety of external conditions (river discharge and wind strength and direction) predicted by mathematical models. We have a total of 51 examples, corresponding to different combination of the values of the control variables in the following table :

| Wind direct. | N | NE | E | SE | S | SW | W | NW |
|---|---|---|---|---|---|---|---|---|
| Wind Stren. | 0 | .05 | .2 | .4 | | | | |
| River Disch. | 750 | 3000 | | | | | | |

## B. *Field subspace*

A subset of the 51 maps provided by MUMM (10) has been reserved to test the performance of the model. A reduced-rank representation has been estimated by computing the Singular Value Decomposition [2] of the $660\times41$ learning matrix *M*:

$$M=\left[col(m_1)col(m_2)\cdots col(m_i)\cdots col(m_{41})\right]$$

The $L$ =28 singular vectors $\{v_i\}_{i=1}^{28}$ of this matrix associated to the largest singular values define the field subspace

$$c=Va+V^{\perp}e, \qquad V=\left[v_1\,v_2\;\cdots v_L\right],$$

effectively compressing the information in all the 41 660-dimensional learning examples. Note that this represents for this case an effective compression rate of about 36%. In the previous equation, $V^{\perp}$ is a matrix whose columns are orthogonal to the vectors $\{v_i\}_{i=1}^{L}$, representing the non-modeled part of the field.

Using this geometric model, we then infer a statistical *a priori* model for the current vectors. By construction, the elements of $a$ are orthogonal, and have mean-square value equal to the square of the corresponding singular value. Moreover, vectors $a$ and $e$ are uncorrelated. To learn the covariance matrix of vector $e$, would require a huge number of learning examples. For this reason, we decided to model $c$ as a normal random vector, with mean value $Vm_a$, where $m_a$ is estimated as the sample mean using the 41 learning maps, and covariance matrix

$$\Sigma_c=V(D^2-m_a m_a{}^T)V'+l_{L+1}I$$

where $l_{L+1}$ is the singular value associated to $n_{L+1}$, which is strictly larger than the true covariance. In this way, we are sure that we do not overestimate the confidence that is attached to the model learned from prior knowledge.

## C. *Examples*

Figure 1 shows one of the learning current fields (top plot) and its projection on the estimated field subspace (bottom plot). As expected the difference is not noticeable, and the macroscopic geometric properties of the map have been preserved.

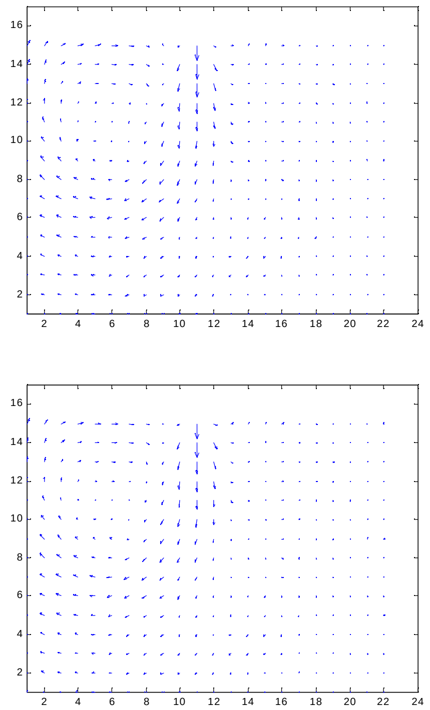

Fig. 1. Current field for one of the learning examples (top) and its projection in the field subspace (bottom).

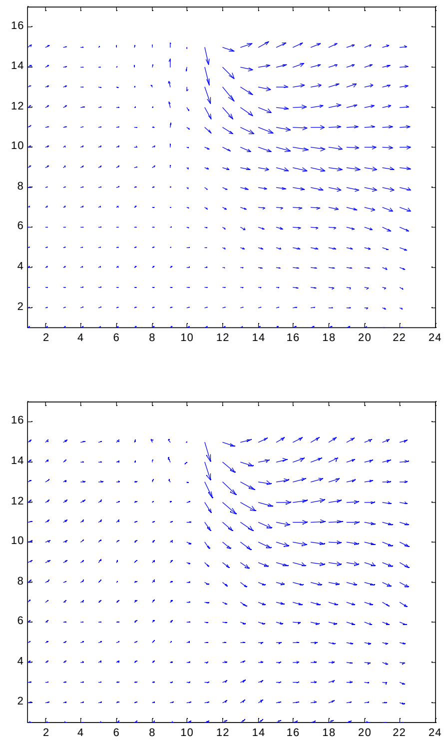

Fig. 2. Test field (top) and its projection on the field subspace (bottom).

Figure 2 represents the same plots for one of the test fields, not included in the learning set. Again, we see that the map can be efficiently represented in the identified field subspace.

## III. EXTRAPOLATION

The model described in the previous section allows the extrapolation of measures taken over a limited given region, yielding estimates of the current field over the entire grid. Let $z$ be a $2N$ dimensional vector of noisy current measurements

$$z = \left[ z_x{}^T \; z_y{}^T \right]^T$$

$$z_x = \left[ c_x(p_1) \; c_x(p_2) \cdots c_x(p_N) \right]^T + [n_{x1} \, n_{x2} \cdots n_{xN}]^T$$

where $p_i$ are the points at which the measurements have been taken. For simplicity, we consider all over this paper that the observation points coincide with the grid points. This condition can be relaxed with a slight increase in the analytical complexity of the method. The observation noise $n$ has zero mean and known covariance matrix $S_n$, independent of the order by which the points are visited. Note that this is a realistic assumption in the case where the vehicle position is obtained for instance by triangulation with respect to an acoustic baseline array.

The observation vector $z$ can be written in terms of the entire current field as

$$z = Sc + n$$

where $S$ is an $2N \times 660$ selection matrix, whose $i$th and $(i+N)$th rows are the Euclidean vector $e_{p(i)}{}^T$ and $e_{p(i)+330}{}^T$, respectively.

The MAP (maximum a posteriori) estimate of the entire vector is defined by

$$\hat{c}_{MAP} = \arg\max_c \; p(c \mid z) = \arg\max_c \; p(z \mid c)\, p(c)$$

The first factor, the likelihood of the observations given a particular current field is a Gauss density of mean $Sz$ and known covariance matrix $\Sigma_n$. The second term is the statistical *pdf* learnt from *a priori* knowledge (gauss density with mean $m_a$ and covariance $\Sigma_c$) presented in the previous section.

This statistical model yields the following unbiased estimate

$$\hat{c}_{MAP} = \left( S^T \Sigma_n{}^{-1} S + \Sigma_c{}^{-1} \right)^{-1} \left( S^T \Sigma_n{}^{-1} z + \Sigma_c{}^{-1} V m_a \right).$$

Note that even if a single point is observed, i.e. $N=1$, we can simultaneously update the entire current map.

## IV. INFORMATION GAIN

The covariance matrix of this estimate (the variance of the estimation error) is easily found to be given by

$$Var(c \mid z) = \left( S^T \Sigma_n{}^{-1} S + \Sigma_c{}^{-1} \right)^{-1},$$

showing how matrix $S$ – the location of the observed points – impacts the performance of the overall estimated field.

The previous equations allows us to compute the information gain yield by the observation of any individual grid point $p$.

In this way, we can determine which point will yield a larger decrease in the error of the overall field estimate. The numerical complexity of simultaneous determination of the $N$ points leading to the best performance (smaller covariance matrix of the estimate) is prohibitive, even for off-line survey design. We use a sub-optimal approach that consists in determining these points iteratively, using a greedy strategy: first the most informative point is found, assuming observation at this point, the second best point is then determined, and so on. There are several optimality criteria that are typically used in the context of experience design, to quantify what is meant by *small covariance matrix* [3]. The most widely used criterion is by far the determinant of the error covariance matrix. Using this criterion the iterative search the next-best 40 points yield the map shown in Figure 3.

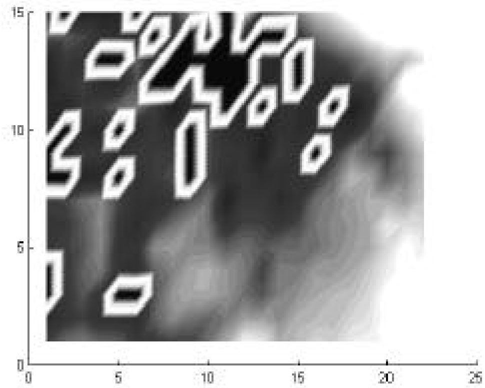


Fig. 3 : Information distribution.

In this plot, the (40) dark grid points surrounded by a white line correspond to the most informative points, while the background gray level represents the distribution of acquired (new) information (darkest corresponding to strongest).

Using the 20 best points to observe the test field (not in the learning set) represented on Figure 4 yields the global (extrapolated) map shown in Figure 5.

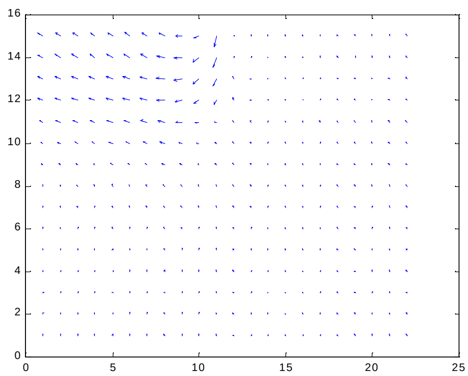


Fig. 4. True current field.

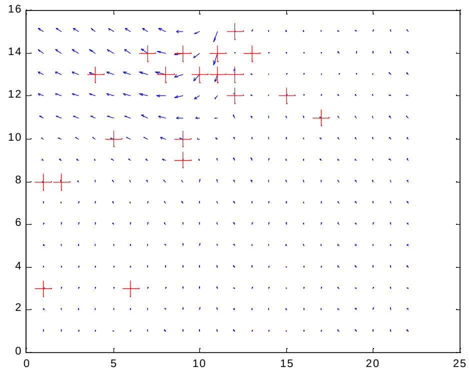

Fig. 5. Estimated field using 20 measures.

The error distribution (in the same scale) is given in Figure 6.

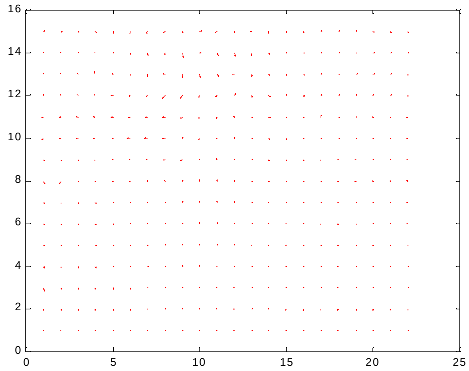

Fig. 6. Error field.

Note that the fact that the points were chosen to yield the largest possible information gain, together with the structure of the model built from the learning examples allow a very rapid global assessment of the overall situation. Figure 7 represents, for the same example, the map extrapolated from the 2 most informative points.

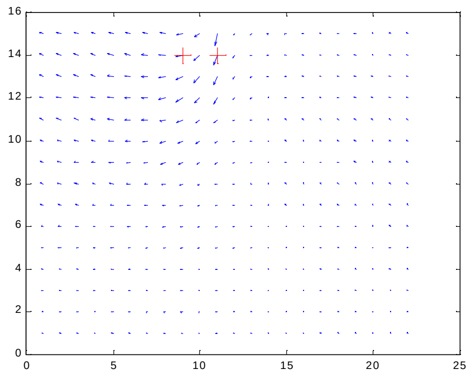

Fig. 7. Map estimated from 2 most informative points.

## V. INFORMATION GUIDANCE

In the previous section we saw that one can determine, prior to any measurement, the best strategy to sample a given field. However, in many situations, we may be interested in some particular characteristic of the current field, for instance, the lines of constant current intensity or of strongest variation. In this case, as the simulations presented in a latter section show, a rough global map can be obtained using a small number of points, enabling guidance of the vehicle towards the most relevant regions.

Below, we present our guidance approach in the context of an iso-line acquisition mission: contour of constant current intensity: $\|c\|^2 = C^{te}$.

A global on-line guidance criterion which would search, over the entire set of grid points, for the one that yields a better overall performance, can lead to an erratic path of the vehicle, switching observation between regions of the workspace widely spaced apart. Instead, we use a *local criterion*, searching amongst the neighbors of the current point which are also neighbors of the contour, for the one that optimizes the quality of estimation of the contour point that is worst estimated (*min-max* approach). The neighborhood considered is expanded incrementally until at least one good candidate point is found. Let $p$ be the current point grid, and $I_p$ the set of neighbors of $p$ which are also neighbors of the contour points $j$. To each contour point $j$ we associate the two grid points $(j_1,j_2)$ which are its closest neighbors (one on each side of the contour). Let $\ell$ be the distance between the contour point and point $j_1$. This distance is a non-linear function of the *North* and *East* components of the current estimates at $j_1$ and $j_2$, and the determination of statistics of its error require intensive numerical computations. By using a 1st order Taylor series and neglecting the correlation between the estimates of the two components of the currents, we can approximate

$$\Sigma(j \mid i) \equiv E\left[\left(\ell_j - \bar{\ell}_j\right)^2 \mid z_i\right] \cong$$

$$\frac{1}{\left(c(j_1) - c(j_2)\right)^4}\left[c(j_1) \quad -c(j_2)\right]^T \Sigma_{j_1 j_2}(z_i) \begin{bmatrix} c(j_1) \\ -c(j_2) \end{bmatrix}$$

where $\Sigma_{j_1 j_2}(z_i)$ is the covariance matrix of the error of the intensity estimates at points $j_1$ and $j_2$ if point $i$ is observed, determined from the covariance of the horizontal and vertical components (see Section IV) by neglecting their cross-correlation.

The next point to be observed is then given by

$$p^* = \arg \min_{i \in I_p} \max_j \Sigma(j \mid i).$$

In Figures 8-11 we show the evolution of the estimated contour level using this strategy during an observation where the goal is to map the iso-intensity line $\|c\|^2 = 0.14$.

The true current field used in the simulations shown below is one of the 10 maps reserved for testing. The star in the plots indicates the current vehicle position. The true contour is the boundary of the gray area, and its estimate, obtained using the prior model and all the observations up to the current one, is drawn as a solid line. The hexagram  indicates the target point, identified as the most informative with respect to the neighboring region of the contour as explained above.  Figure 8 shows the estimated contour after a single random;y chosen point has been observed. Since the vehicle is started in a very uninformative region, the estimate is very poor.

We can see that our strategy effectively guides the vehicle towards the most informative region, which it reaches after 5 observations, see Figure 9. The vehicle then follows an almost linear path at the interior of the contour, until its boundary is reached, see Figure 10, which shows the contour estimated after 13 observations.  Note that at this point the vehicle has already a good knowledge of the overall contour shape. Our algorithm then guides the vehicle in a zig-zagging path alternating on both sides of the contour.

Finally note that with only 27 a good global representation has been obtained, see Figure 11. We stress that in this example, the observations are contaminated by a strong level of noise. This can be appreciated by the fact that the estimated contour in the region actually observed is at a considerable distance from the true one.

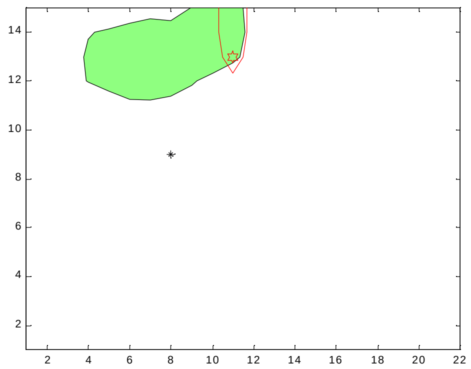


Fig. 8. Initial contour estimation.

The efficiency of our approach should be contrasted to purely perception-driven (not model-based) observation approaches. Often, as in the example shown, the initial observation point is not in the neighborhood of the feature of interest, and with purely perception driven approaches the vehicle has no information regarding its location. A random wandering approach, or a systematic geometric pattern of search (e.g., spirals) must then be used to detect the region of interest, see e.g. [1], penalizing the efficiency of the observation. Moreover, we note that the use of *a priori* information enables observation of the contour without requiring oscillation of the vehicle on each side of the contour, as for perception-guided control strategies [1], which are prone to lost of tracking problems which decrease their performance.

## VI. CONCLUSIONS

In this paper we presented an approach to the observation of natural phenomena which combines on-line perceptual guidance and a model-based approach. Note that our prior knowledge is a statistical model, and should not be confused with approximate (or partially known) deterministic prior maps which are frequently used in the context of the navigation of autonomous robots.

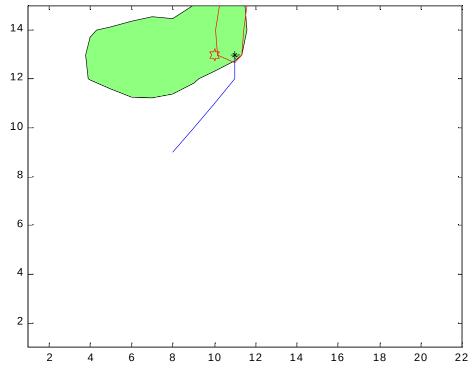


Figure 9: Estimated contour (5 observation points).

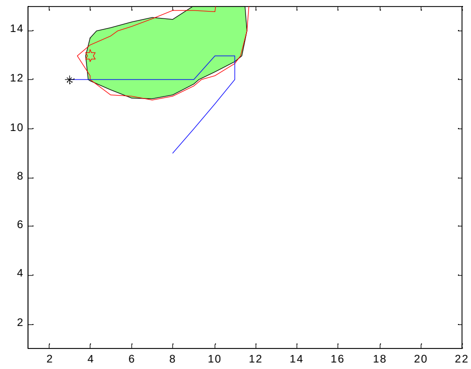


Fig. 10. Estimated contour (13 observation points).

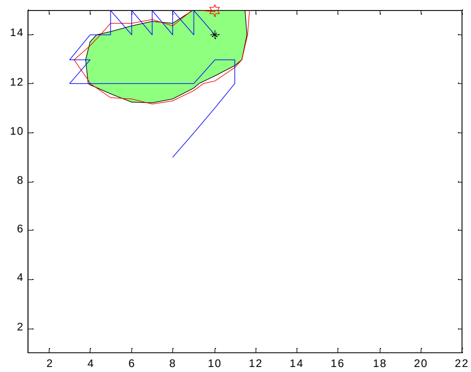


Fig. 11. Estimated contour (27 observation points).

For phenomena for which a considerable amount of prior knowledge exists, for instance in the form of maps, we have shown that it is possible to compress this information in such a way that it can be used to extrapolate localized measures to have a global estimate of the overall observed field. We show, by simulation examples, that use of this extrapolated global maps can considerably increase the efficiency of environmental surveys.

The work presented here considers several simplifying assumptions and constraints that will be dropped in the future. In particular, we constrained the observation points to be identical to the grid points at which the learning maps are defined. This constrain can be dropped with a corresponding increase of the numerical complexity of the method, by considering that the matrix $S$ that defines the observation vector is actually an interpolation matrix.

Our guidance strategy considers as candidate observation points, the points in the immediate neighborhood of the current point. Instead, a criterion combining distance to the target point and the increase in performance that it yields can be more efficient. We will analyze this possibility in the future.

## Acknowledgements

This work has been partially funded by the European Union through research contract IST-1999-10836 (SUMARE*: Survey of Marine Resources*). We acknowledge the collaboration of MUMM (*Management Unit of the North Sea Mathematical Models*), Brussels, Belgium, which provided the current maps at the Rhône mouth used in this study.

## References

[1] Andrew A. Bennet and John Leonard, “A Behavior-Based Approach to Adaptive Feature Detection and Following with Autonomous Underwater Vehicles,” *IEEE J. of Oceanic Eng.*, vol. 25, pp. 213-226, April 2000.

[2] Gene H. Golub and Charles F. van Loan, *Matrix Computations*, The John Hopkins University Press, 1989 (2nd edition).

[3] S. D. Silvey, *Optimal Design*, Chapman & Hall, Series Monographs on applied probability and statistics, 1980.