





# Sampling design for binary geostatistical data, application to inspection actions of fishing activity in Portugal

Belchior Miguel <sup>a,b,c</sup> ,\* , Paula Simões <sup>d,b</sup> ,\*\* , Rui Gonçalves de Deus <sup>e,f</sup> , Isabel Natário <sup>a,b</sup> 

<sup>a</sup> Nova School of Science and Technology, NOVA University of Lisbon, Caparica, Portugal

<sup>b</sup> Center for Mathematics and Applications (NOVA MATH), NOVA University of Lisbon, Caparica, Portugal

<sup>c</sup> Instituto Superior Politécnico de Tete, Tete, Mozambique

<sup>d</sup> Military Academy Research Center, Military University Institute (CINAMIL), Lisbon, Portugal

<sup>e</sup> Naval Research Centre (CINAV), Navy School, Military University Institute, Lisbon, Portugal

<sup>f</sup> Information management and analysis division, Portuguese Navy, Lisbon, Portugal

## ARTICLE INFO

### Keywords:

Sampling design  
Binary spatial data  
Presumed fishing infractions  
INLA  
SPDE

## ABSTRACT

The definition of surveillance routes is a very important but complex issue. The Portuguese Navy, in its common form of operation is in charge of the Naval Standard Device, which is distributed throughout the various coastal areas of the country. Enforcement actions can involve very high costs, so a good plan for the sampling designs used are in order, as to maximize the efficiency in obtaining information from the data of the actions developed over the area under consideration. The main objective of this study is to propose sampling design criteria based on geostatistical models, in the context of binary data on presumed maritime infractions in the Portuguese coast, that are advantageous in the optimization of maritime surveillance actions, in terms of efforts employed in their execution, in the maritime area of Portugal's responsibility. Two sampling design selection criteria are proposed: Maximum Risk Sampling design and Maximum Variance Risk Sampling Design. These are compared to the simple random design by the root mean square error (RMSE). A comparison of the designs at different sample sizes is made and the estimated risk maximization sampling design presents the best RMSE value. The proposed sampling designs may assist in the creation of alternative enforcement Portuguese Navy routes, optimizing the scheduling that maximizes the probability of finding a higher number of presumed fishing perpetrators with less resource efforts.

## 1. Introduction

Portugal, as a coastal country, has sovereignty and jurisdiction over 14 069 km<sup>2</sup> of internal waters, 50 957 km<sup>2</sup> of territorial sea and 1 660 456 km<sup>2</sup> of Exclusive Economic Zone (ZEE), totaling 1 725 482 km<sup>2</sup> of maritime space for which is responsible, [Fig. 1](#). The Portuguese Navy is the branch of the Portuguese Armed Forces in charge of the country's maritime defense and security, as well as protecting national interests at sea. The Navy's duties include control and surveillance, monitoring and patrolling territorial waters, ensuring maritime security.

\* Corresponding author.

\*\* Corresponding author at: Center for Mathematics and Applications (NOVA MATH), NOVA University of Lisbon, Caparica, Portugal.

E-mail addresses: [b.miguel@alumni.fct.unl.pt](mailto:b.miguel@alumni.fct.unl.pt) (B. Miguel), [pc.simoese@fct.unl.pt](mailto:pc.simoese@fct.unl.pt) (P. Simões).

<https://doi.org/10.1016/j.spasta.2025.100919>

Received 20 December 2024; Received in revised form 10 July 2025; Accepted 10 July 2025

Available online 21 July 2025

2211-6753/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

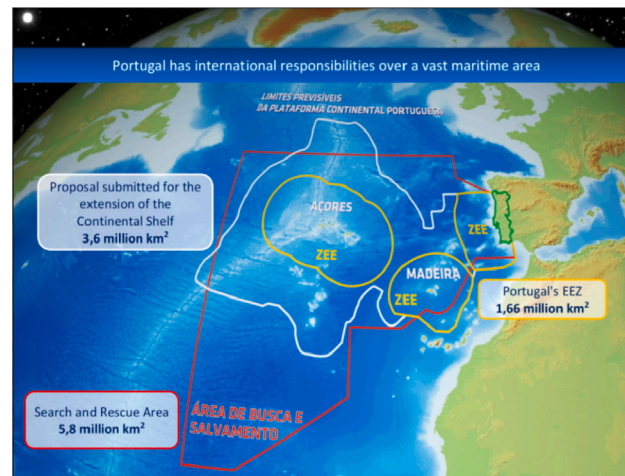


Fig. 1. Exclusive economic zone.  
Source: Portuguese Navy.

Monitoring fishing activity is very important for the preservation of the species that are fished as well as to ensure their continuity. The Portuguese Navy, in its common form of operation, is liable for the Naval Standard Device, which is distributed throughout the various coastal areas of the country. The Fishing Activities Inspection and Control System (SIFICAP), was created in the Portuguese Navy, aiming to ensure the surveillance, inspection, and control of activities.

Defining surveillance routes is a very important yet complex task. Since enforcement actions can be extremely costly, it is essential to develop well-designed sampling plans that maximize the efficiency of information gathering from the Navy's actions conducted in the target area (Liu et al., 2009). It is essential to find a survey design that optimizes maritime surveillance actions, in terms of effort employed in their execution, in the maritime area under Portugal's responsibility.

Sampling design has the guiding function of how carefully selecting the elements of the population that will be part of the sample to be used to make inferences about the population (Cochran, 1977; Rivoirard et al., 2008; Stein and Ettema, 2003). There are two main approaches to sampling design: design-based approach (Petitgas, 2001) and model-based approach (Brus and De Gruijter, 1997). In the design-based approach, requiring fewer assumptions, sampling sites are selected by probability sampling and statistical inference is independent of the sampling design, whereas in the model-based approach no requirements are needed for the selection of sampling sites, this is usually done using intentional (directed) sampling, such as in a centralized grid (Särndal et al., 1978; De Gruijter and Ter Braak, 1990).

Many scientific works have been developed related to sampling designs for spatial data. Simbahan and Dobermann (Simbahan and Dobermann, 2006) used two different sampling design approaches to optimize sampling based on secondary information and its application in soil carbon mapping. In their study, a design approach based on stratified random sampling schemes was used, and three model-based optimization alternatives were evaluated: minimization of the shortest distance of the future sample locations and the observations (MMSD), uniform distribution of point pairs for variogram estimation (WM), and a combination of MMSD (2/3 of samples) and WM (1/3). Bogaert and Russo (Bogaert and Russo, 1999) proposed an optimization algorithm based on generalized least squares to reduce the variability of estimators by minimizing the determinant of their covariance matrix, to find an optimal finite set of locations in space where a random field should be sampled to minimize the variability of the parametric variogram estimator. Rocha et al. (2020) studied the role of sampling design in predicting spatially dependent ecological data using remote sensing. In their study, they used four sampling designs representing design-based approaches: simple random design (SRD); systematic sampling; lattice plus close pairs; and lattice plus in-fill, which are often used in remote sensing and geostatistics. Currently, advances in geostatistical methods have led to the development of adaptive sampling strategies that are particularly well suited for spatial binary data, as commonly observed in epidemiological and surveillance contexts. For example, Chipeta et al. (2016) proposed an adaptive geostatistical design for estimating disease prevalence, while Kabaghe et al. (2017) applied such strategies in repeated malaria surveys in rural Malawi.

Under the Portuguese Navy's actions of control and surveillance, data containing relevant information on presumed maritime infractions and on the fish caught are gathered, being considered in this study. These are binary data on the presence or absence of presumed infractions, observed in the collection of the locations visited during the inspections carried out by the Portuguese Navy patrols, in four periods of the day of six hours each, corresponding to the work shifts, 0:01–6:00; 06:01–12:00; 12:01–18:00 and 18:01–24:00.

The application of geostatistical techniques aims to analyze and understand spatial patterns and variability in georeferenced data, such as these data related to maritime enforcement actions. This helps to identify areas with a higher concentration of infractions, understand risk factors and plan enforcement actions accordingly. A geostatistical model can be used to model the prevalence of

fishing infractions on the Portuguese coast. The probability of occurrence of the presumed infringements can thus be modeled through a Bayesian hierarchical model with a spatial component for the binary data.

Having said that, the main objective of this study is to propose sampling design criteria, based on geostatistical modeling of binary data, that prove to be advantageous in the goal of optimizing enforcement actions in terms of the effort employed in their execution. Additionally, we want to compare these proposals with conventional sampling designs. From the observed data obtained in enforcement actions, we intend to implement sampling schemes for future fishing enforcement based on two new criteria: maximization of infraction risk and exploration of areas where the variability associated with risk is high. These criteria are used to construct two sampling schemes that are compared to SRD using the root mean square error (RMSE).

This work is divided into 4 sections. Section 2 details various sampling designs, including both design-based and model-based approaches. It considers conventional sampling methods, designs for geostatistical data, and the newly proposed model-based sampling approach. Further, the criteria used for selecting and validating these sampling designs are detailed. Section 3 presents the results of applying the SRD, Maximum Risk Sampling Design (MaxRSD), and Maximum Variance Risk Sampling Design (MaxVRSD) to the binary geostatistical data considered in this study. Finally, Section 4 outlines the conclusions and provides recommendations for future research.

## 2. Sampling design

Sampling designs can be considered under two different approaches, the design-based approach (Petitgas, 2001) and the model-based approach (Brus and De Gruijter, 1997). Generally, the design-based approach has been the best choice if we are interested in the spatial cumulative distribution function for the area as a whole or for a restricted number of subareas, and furthermore, the validity of the result really matters (validity of the result is more important than efficiency), while the model-based approach is a better choice if we are interested in a map describing the values of many small areas, e.g. *pixels*, and we want to predict these values as accurately as possible (efficiency is more important than validity) (Brus and De Gruijter, 1997).

In this section, a description of the two aforementioned sampling design approaches is presented. It is also presented new adaptive sampling design criteria for binary data within the model-based approach.

### 2.1. Design-based approach

According to Brus and De Gruijter (1997) and Liu et al. (2009), the design-based approach is often referred to as classical sampling theory, where stochasticity is introduced by sampling, and sample locations are selected by a predetermined random selection criterion. In classical sampling, multiple samples can be selected in each area using the same sampling design. Unlike design-based approaches where inference is based on some random sampling scheme to choose a sample from the population units, the model-based approach considers distributional assumptions on the data-generating unknown process (the statistical model) (Dumelle et al., 2022).

There are several conventional sampling designs, but this paper only briefly describes the three types of sampling more widely used in research: SRD, systematic random sampling and stratified sampling (Cochran, 1977; Rocha et al., 2020).

Simple Random Design (SRD) is the most basic of probabilistic samplings and provides the theoretical foundation for more complex forms, where units are chosen from a list of observations in the population (Lohr, 2021). Systematic Sampling is widely used for its simplicity and convenience, involving the selection of elements at regular intervals after choosing a random starting point (Lohr, 2021). According to Cochran (1977), Stratified Random Sampling is commonly used in fisheries-related research and involves dividing the study area into internally homogeneous but heterogeneous strata, randomly selecting one point from each stratum to form the sample.

### 2.2. Model-based approach

In the model-based approach, the observation process is modeled as a stochastic process. This stochastic process is a mathematical abstraction used to describe reality, where the probabilities of occurrence of the elementary outcomes of this process, the value fields, are not known, but must be modeled (Brus and De Gruijter, 1997).

According to De Gruijter and Ter Braak (1990), in geostatistics, the model-based approach is mainly used for the prediction of spatial averages. This approach treats the value associated with any location not as a fixed value, but rather a random value. The set of values associated with all possible realizations in the study region is considered as just one realization of an underlying process, where some characteristics of this process are assumed to be known and these assumptions are formalized in a geostatistical model, which plays the role that is called the overpopulation model in sampling theory.

In the model-based approach, the only source of stochasticity is the underlying postulated process, and inference is based primarily on the formulated model (Brus and De Gruijter, 1997).

Also in the geostatistical context, a design involves the selection and optimization (maximizing or minimizing) of a suitable objective function (Liu et al., 2009; Diggle and Lophaven, 2006) that is developed based on a particular selected design criterion (such as minimizing the mean or minimizing variance in estimation) (Van Groenigen et al., 1999). However, there are other alternative approaches that do not require information on the covariance structure of the data (Liu et al., 2009; Russo, 1984). For instance, those focusing the quality of semivariogram estimation, the key element of geostatistical analysis (more details in Liu et al. (2009), Simbahan and Dobermann (2006), Van Groenigen et al. (1999) and Russo (1984)).

While the semivariogram is intrinsically linked to the covariance function under stationarity assumptions, as highlighted by Lohr (2021), some approaches, such as those discussed in Liu et al. (2009), use empirical or nonparametric semivariogram estimates as a starting point for sampling design. Although these methods may eventually involve selecting a parametric variogram model (e.g., spherical or exponential), this specification does not need to be made a priori. Instead, preliminary data modeling may guide such choices. Thus, the sampling design can be informed by an approximate spatial model without being rigidly constrained by a predefined spatial correlation structure. This is particularly advantageous in practical contexts where little or no information is available beforehand regarding the spatial dependence of the underlying phenomenon.

According to Van Groenigen et al. (1999), generally, most approaches assume that all parameters that define the covariance structure of the data are known and the goal is to identify the best design that can minimize the variance in the estimate for those parameters. There are cases where these approaches are not applicable, due to limited or even unknown knowledge, and it is necessary to resort to the data under analysis for the estimation of the covariance structure of the data.

According to Brus (2022), a wide variety of probability-based sampling designs are presented that are applicable to different types of data, including proportional and categorical variables. The author explicitly discusses the estimation of population proportions, which is directly applicable to the context of binary variables, and demonstrates that these sampling designs can be effectively applied even when the variable of interest is binary (0/1), such as in prevalence studies or the detection of rare events.

Next, some sampling design selection criteria for the model-based approach used in geostatistics, proposed in the literature, are described.

Firstly, the Minimum Mean Shortest Distance (MMSD) criterion (Simbahan and Dobermann, 2006; Lin and Clayton, 2005), which requires all sampling points to be regularly distributed throughout the study area and aims to minimize the average distances between any point within the region and its nearest sampling point. It utilizes all  $N$  points of a fine mesh as selection options. Secondly, the criterion for Uniform Distribution of Point Pairs for Variogram Estimation (WM), that uses variograms to achieve a uniform distribution of point pairs across different distance classes or groups, aiming for an accurate estimation of the variogram. It does not assume spatial correlation structures, relying solely on distances between sampling points (Liu et al., 2009; Simbahan and Dobermann, 2006; Van Groenigen et al., 1999; Russo, 1984). The combination of these two criteria, WM and MMSD, forms a third criterion, MMDS+WM, for additional optimization (more details see Liu et al. (2009), Simbahan and Dobermann (2006) and Van Groenigen et al. (1999)).

### 2.3. Proposal of adaptive sampling designs

The design-based and the model-based approaches described before are suitable for traditional geostatistical data where the response of interest is Gaussian distributed. However in this work, we are dealing with binary data, so the previous criteria may not be able to apply in these cases.

#### 2.3.1. Geostatistical model for binary data

In geostatistics it is assumed that there is a phenomenon of interest modeled by a spatial continuous stochastic process  $\{U(s) : s \in D \subseteq \mathbb{R}^2\}$  that is imperfectly observed at some defined locations  $s = (s_1, \dots, s_n)$  in a certain domain  $D$  according to a sampling scheme, resulting in observations  $\mathbf{Y} = (Y_1, \dots, Y_n)$ ,  $Y_i = Y(s_i) \in \mathbb{R}$ ,  $i = 1, \dots, n$ . The geostatistical model for Gaussian data is given by

$$Y_i = \mu + U(s_i) + \epsilon_i = \mu + u_i + \epsilon_i, \quad s_i \in D, \quad i = 1, \dots, n. \tag{1}$$

where  $\mu \in \mathbb{R}$  is a constant mean parameter,  $U$  is a Gaussian process, with zero mean, variance  $\sigma_u^2$  and Matérn covariance function dependent of  $(k, \nu, \sigma_u^2)$ . Independent and identically distributed (i.i.d.) normal random errors  $\epsilon_i$  have  $E[\epsilon_i] = 0$  and  $Var(\epsilon_i) = \sigma_\epsilon^2$ , the nugget effect, and  $U(s_i) = u_i$ .

The data in hands are binary geostatistical data representing the presence or absence of fishing alleged infractions at specific locations, Miguel et al. (2024a), driven by an underlying spatial stochastic process  $\{U(s) : s \in D \subseteq \mathbb{R}^2\}$  as before, and described by a Bernoulli distribution:

$$y_i \sim \text{Bernoulli}(\pi_i) \tag{2}$$

where  $\pi_i$  is the probability of an alleged infraction in site  $s_i$ . This probability (the expectation of the Bernoulli distribution) is modeled through a link function to fit the success probability of the Bernoulli model by means of covariates and spatial dependence effects. The link function can be the logit function (widely used) or the complementary log–log function, which is more suitable for scarce events or imbalanced data, as is the case here. The form of the complementary log–log function is given below, as well as the corresponding predictor:

$$cloglog(\pi_i) = \log(-\log(1 - \pi_i)) = \beta_0 + \sum_{m=1}^M \beta_m x_{m,i} + u_i \tag{3}$$

where  $\beta_0$  is the intercept,  $\beta_m$  is the effect of covariate  $x_m$  and  $\mathbf{u} = \{u_i\}_{i=1}^n$  is a realization (at  $n$  locations) of the spatial random field with Matérn covariance function with range  $R$  and variance  $\sigma_u^2$ .

### 2.3.2. Proposed adaptive sampling design criteria

Next, we describe the sampling criteria we proposed and used in this work to construct the respective optimized sampling schemes. These are considered adaptive approaches. According to Chipeta et al. (2019), Pearse et al. (2020), Simões et al. (2021) and Silva et al. (2023), adaptive geostatistical designs are those in which sampled sites are chosen in groups in a temporal sequence. The sites chosen in any group optimize for the study objective, based on a criterion calculated considering the previous group. The adaptive sampling design criterion ensures that data are collected only from sites that may be useful in providing additional information. The traditional approaches mentioned before are not adaptive.

The approaches presented in this subsection are sampling designs for binary geostatistical data, built on adaptive geostatistical sampling, in order to obtain the most suitable criterion for the problem under analysis, thus contributing to the implementation of new selection methods of sampling design for binary geostatistical data.

#### Maximum estimated risk (MaxRSD).

We are interested in building a geostatistical sampling design for binary data (models with unknown parameters), based on a criterion that seeks to sample more intensively where a higher prevalence is predicted (adaptive).

Since the covariance structure is unknown, we first estimate a model that best fits an observed data set; then, we predict the model on a fine grid of points covering the area under study, which constitutes the set of potential points to be included in the sample, obtaining the estimated probabilities at each point; finally, we sample from these a fixed number of points  $n$ , with a probability of selection proportional to these estimated probabilities (risk). Like this, we choose to sample at the locations of higher estimated risk.

We call this design selection criterion Maximum Risk Sampling design (MaxRSD), which depends on  $n$  (number of points to sample) and  $m$  (estimated model),  $\text{MaxRSD}(n, m)$ , and can be obtained as follows:

1. Considering  $X_0 = (s_1, \dots, s_{n_0})$  the initial sample locations chosen using the SRD from previously observed points, corresponding to data  $Y_0$ , the geostatistical model given by Eq. (5) is estimated;

$$y_i \sim \text{Bernoulli}(\pi_i) \tag{4}$$

$$\text{cloglog}(\pi_i) = \log(-\log(1 - \pi_i)) = \beta_0 + \mathbf{x}'_i \boldsymbol{\beta} + u_i, \quad i = 1, \dots, n \tag{5}$$

where  $\mathbf{x}$  are covariates and  $\mathbf{u}$  are spatial random effects and  $n$  is number of points to sample.

2. One then estimates the risk at points different from  $X_0$ , the potential sampling locations on a fine grid (mesh) covering the study region  $X_1 = (s_1, \dots, s_{n_1})$  (unobserved), and determines the probabilities of inclusion of each of the  $X_1$  points as the predicted probabilities at those points, standardized to sum to 1 (estimated risk); and
3. Select  $n$  points from  $X_1$ ,  $n < n_1$ , according to the inclusion probabilities obtained in 2, obtaining a proposed sampling design with the locations of higher estimated risk.

#### Maximum variance risk sampling design (MaxVRSD).

An interesting alternative to a geostatistical sampling design for binary data (models with unknown parameters), is to use a criterion that seeks to sample with greater intensity where prevalence is estimated with greater uncertainty (adaptive). Since the covariance structure is unknown, we first estimate a model that best fits the observed data under analysis; then, we predict the model on a fine grid of points covering the area under study (which constitute the set of potential points to be included in the sample) obtaining the estimated probabilities at each point and determining the corresponding estimated standard errors; finally, we sample from the potential choices a fixed number of points  $n$ , with the probability of selection proportional to the standard errors of the risk. In this way, we choose to sample the locations of greatest uncertainty in the estimated risk.

We call this design selection criterion Maximum Variance Risk Sampling Design (MaxVRSD), which depends on  $n$  (number of points to sample) and  $m$  (model to estimate)  $\text{MaxVRSD}(n, m)$  and can be obtained as follows:

1. Considering  $X_0 = (s_1, \dots, s_{n_0})$  the initial points sample locations chosen using the SRD from previously observed points, corresponding to data  $Y_0$ , we estimate the geostatistical model given by Eq. (5);
2. The risk is then estimated at the points of the potential sampling locations on a fine grid (mesh) covering the study region  $X_1 = (s_1, \dots, s_{n_1})$  (unobserved), determining the probabilities of inclusion of each of these points as the standard errors of the projected risk at these points, standardized to form a probability distribution — given by the standard error at point  $i$  divided by the sum of all standard errors at the grid points; and
3. Select  $n$  points from  $X_1$ ,  $n < n_1$ , according to the inclusion probabilities obtained in 2, obtaining a proposed sampling design with the locations of maximum estimated variance of the estimated risk.

Note that we choose to analyze the standard error of all projected risks, and not only the highest values, as we pretend to globally be able to sample more where the variability of the predicted risk is larger.

Therefore, in order to obtain a Geostatistical sampling design for binary data, firstly estimate a geostatistical model that best fits the observed data set; then, predict the model on a fine grid of points covering the area under study, the potential sampling points; After, obtain the predicted probabilities at each point and sample from these a fixed number of points  $n$ , considering the proposed adaptive sampling design criteria — the Maximum Risk Sampling Design (MaxRSD) or the Maximum Variance Risk Sampling Design (MaxVRSD).

However, the operationalization of these designs may depend on their context. For example, if the context is surveillance on vessels at sea, and sampling is carried out using a vessel of a certain size, a policy of scanning the area may need to be considered whereby the vessel travels a certain route and the sampling points provided by the layout are visited using smaller boats or dinghies.

#### 2.4. Sampling design evaluation

This section describes the use of the root of the mean square error (RMSE) to compare and identify the optimal design.

Using the estimated values of the different parameters of the adopted model with  $X_0$  initial points in step 1 of the MaxRSD and MaxVRSD criteria above and re-estimating the model for the selected  $n$  points of  $X_1$  obtained in step 3 of each criteria, it is possible to evaluate and compare the results in terms of the estimated risk and thus identify the optimal design. For that, the root of the mean square error (RMSE) of prediction is used:

$$RMSE = \sqrt{\frac{1}{n} \sum_{\alpha=1}^n (y_{\alpha}^* - y_{\alpha})^2}, \quad (6)$$

where  $n$  is the number of sampled points,  $y_{\alpha}$  and  $y_{\alpha}^*$  are, respectively, the predicted values of response variable at site  $s_{\alpha}$ , based on the geostatistical model fitted for the observed data, and the estimated value of the response at site  $s_{\alpha}$  from the fit of the points from the proposed sampling design, using the previously mentioned predicted data.

Note that the sampled predicted values cannot be used directly to obtain the estimated values, they must be converted into successes (“1”) or failures (“0”). This is done by comparing the predicted value with a probability  $p$  which ideally would be 0.5 but which in low prevalence studies may have to be decreased to avoid the issue of too few successful data points (“1”) (Simbahan and Dobermann, 2006).

The most appropriate design proposal is the one with the lowest RMSE value.

### 3. Sampling designs for inspection actions of the fishing activity off Algarve coast

In this section, we apply the criteria described in 2.3 for selecting a sampling design for binary geostatistical data concerning fishing infractions, obtained in maritime surveillance actions in the Southern Portugal Zone Command (according to the Portuguese Navy’s operational areas division) in the years 2013 to 2017. The proposals obtained are compared with the design obtained by SRD.

#### 3.1. Study area

The available information refers to georeferenced data regarding enforcement actions carried out by the Portuguese Navy throughout the Portuguese coast. Data collection was carried out in four periods of the day of six hours each corresponding to working shifts, 0:01–6:00; 06:01–12:00; 12:01–18:00, and 18:01–24:00.

For this study, five years were considered (from 2013 to 2017), for the Southern area command (Algarve coastal region), restricted to the second period of the day between 06:01–12:00 h, for being a period with a higher number of infractions.

This study is focused on the southern part of the country, more specifically the Algarve coastal region, and on 3 types of presumed infractions (presence/absence) related to fishing (from now on, named presumed fishing infractions), namely prohibited gear (infringement 3), fishing in a prohibited zone (infringement 4) or fishing prohibited by motor power or excessive tonnage (infringement 5).

Fig. 2 shows the spatial distribution of the sites of inspection actions in the five years (from 2013 to 2017), for the second period of the day (06:01–12:00 h), in the coastal zone of Algarve, and identifies the larger towns in this area. The red dots represent the inspected locations where suspected infractions were found and the black dots represent the sites where no alleged infractions were detected.

The total number of inspected sites analyzed in the five years for second period of the day was 13 674, of which 13 394 (98.0%) correspond to places where no suspected violations were found (“infr. = 0”) and 280 (2.0%) correspond to places where at least one of the presumed fisheries infractions considered was found (“infr. = 1”). There are sites where more than one suspected infraction was found. The contribution of each of the three alleged infractions analyzed in this study for the five years for second period can be seen in Table 1. Of the 280 alleged infractions found, 112 correspond to infraction 3, which represents 0.8% of the data, 182 correspond to infraction 4, which represents 1.3% and 3 correspond to infraction 5, which represents 0.02%. Of the 3 alleged infractions under analysis in this study, it can be seen that infraction 4 had the greatest contribution, that is, it was the most found during the inspection for the considered infractions carried out by the Portuguese Navy, and infraction 5 was the least found.

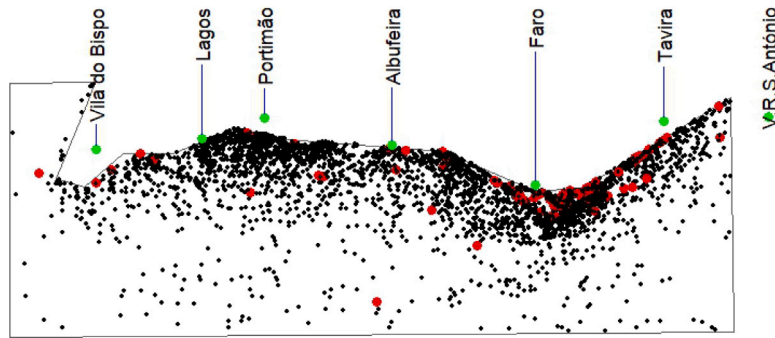


Fig. 2. Distribution map of the maritime inspections actions on the Algarve coastal region for the second period of the day (06:01–12:00 h), red dots are locations of suspected infractions.

Table 1

Contribution of each alleged infringement to the data for the second period of the day between 06:01–12:00 h.

Infringement	Total	infr. = 0	infr. = 0(%)	infr. = 1	infr. = 1(%)
Infringement 3	13 674	13 562	99.2	112	0.8
Infringement 4	13 674	13 492	99.7	182	1.3
Infringement 5	13 674	13 671	99.97	3	0.02
All infr.	13 674	13 394	98.0	280	2.0

### 3.2. Sampling distribution of fisheries infractions

The model considered in this application is given by

$$y_i \sim \text{Bernoulli}(\pi_i) \tag{7}$$

where  $\pi_i$  is the probability of the presence of that presumed infraction. Then, a linear model including intercept  $\beta_0$ , and a spatial random effect  $\mathbf{u}$  is specified in

$$c\log\log(\pi_i) = \log(-\log(1 - \pi_i)) = \beta_0 + u_i, \quad i = 1, \dots, n \tag{8}$$

where  $\beta_0$  is the intercept, and  $\mathbf{u}$  are spatial random effects. The parameter values ( $\beta_0$ ,  $\sigma_u^2$ , and  $R$ ) considered for the model (8) are estimated using the data described in Section 3.1. This random field is to be approximated by a Gaussian Markov random field, for which the corresponding domain is triangulated for the application of the approximation method of finite elements. Fig. 3 presents the considered mesh where the spatial domain of the study area is triangulated, to estimate this geostatistical model through the SPDE approach (Blangiardo and Cameletti, 2015).<sup>1</sup> The mesh has 1775 vertices and 6886 data points.

For the estimation, Bayesian inference is used with the INLA approach (Rue et al., 2009), implemented in R-INLA, [www.r-inla.org](http://www.r-inla.org). The prior distributions for the parameters are the ones R-INLA uses as default: the coefficients of the linear effects are assigned Gaussian prior distributions with zero mean and precision 0.001. For  $\theta_1 = \log(\tau)$ , where  $\tau^2 = \Gamma(\nu)/[\Gamma(\alpha)(4\pi)^{d/2}k^{2\nu}\sigma_u^2]$ , and  $\theta_2 = \log(k)$ , independent Gaussian standard prior distributions  $\mathcal{N}(0, 1)$  are specified. It might be more useful to define priors for the standard deviation  $\sigma_u$  and for the range  $R$ , where  $R = \sqrt{8\nu/k}$  instead of  $\theta_1$  and  $\theta_2$ , and in these cases one can use the following parameterization,

$$\log(\sigma_u) = \log(\sigma_0) + \theta_1 \quad \text{and} \quad \log(R) = \log(R_0) + \theta_2,$$

where  $\log(\sigma_0)$  and  $\log(R_0)$  represent some baseline values, Blangiardo and Cameletti (2015) and Krainski et al. (2018). We did not run a sensitivity analysis here as the primary focus of this work was to develop a sampling design rather than to achieve the best possible model fit. This has been considered before in the work of Miguel et al. (2024b).

According to Table 2, the estimated hyperparameters of the model for the second daily period (2013–2017) are presented. The variance explained ( $\sigma_u^2$ ) by the Matérn variogram is estimated at 4.029, while the posterior mean for the range parameter ( $R$ ) is 4.644 km, with a 95% credibility interval of (2.559; 7.546).

<sup>1</sup> The mesh was constructed based on a balance between spatial accuracy and computational efficiency, following the guidelines proposed by (Lindgren et al., 2011). The mesh resolution was adjusted to adequately capture the spatial variability of the observed data while avoiding excessive model complexity. The following criteria were taken into consideration during the mesh design: 1-Density of observed data points, 2-Spatial extent of the study area, 3-Size and scaling of internal and external triangles, and 4-Numerical stability of the SPDE approximation. The final mesh configuration was validated through visual inspection, analysis of spatial residuals, and assessment of the sensitivity of posterior estimates to alternative mesh specifications. These steps aimed to ensure the robustness of the spatial model without compromising performance.

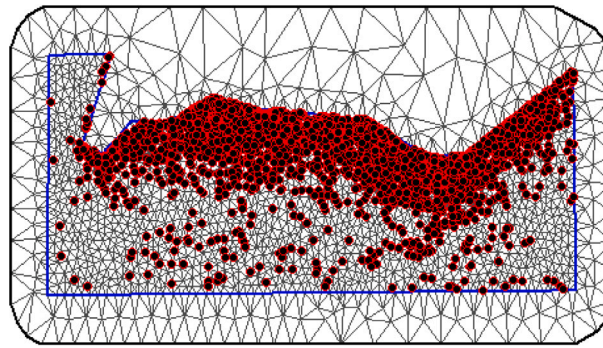


Fig. 3. Mesh chart for the five years data under review (2013–2017).

Table 2

Summary of the posterior probability distributions for the 2nd period of the day for the years 2013 to 2017.

Parameter	Mean	SD	Quantile 0.025	Quantile 0.5	Quantile 0.975
$\beta_0$	-2.767	0.246	-3.378	-2.733	-2.388
$\sigma_n^2$	5.432	2.483	2.371	4.828	11.889
$R$	3.924	0.831	2.465	3.871	5.708

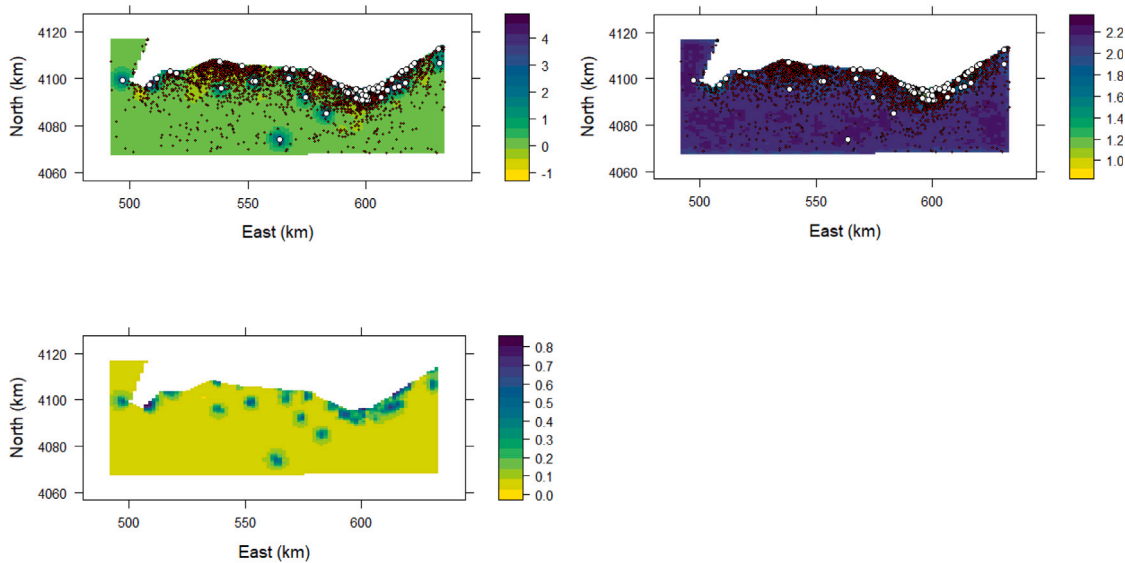


Fig. 4. Map of the posterior mean of the random field (top-left); of the posterior standard error of the random field (top-right); of the predicted values (bottom), for 2nd period of the day for the years 2013–2017.

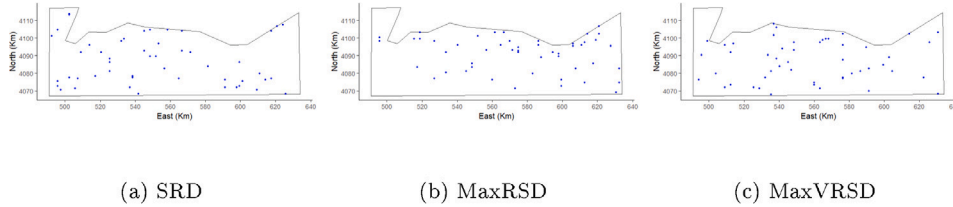
Fig. 4 shows the estimated posterior mean and posterior standard error of the random field and the predicted values for the grid points considered, based on the model adopted in (8) as well as the corresponding standard errors. We can also see that white dots represent the locations where presumed infractions were detected during inspection activities.

### 3.3. Risk maps vs. proposed adaptive sampling designs

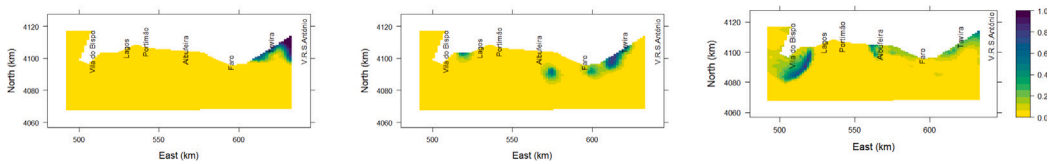
Based on these results, three sampling schemes were derived and are presented below, SRD( $n$ ); Maximum Risk Sampling Design, MaxRSD( $n$ , Model(8)); and Maximum Variance Risk Sampling Design, MaxVRSD( $n$ , Model(8)). For the evaluation of this study, we considered the success defining probability value  $p = 0.4$ , as described in Section 2.4, and three different sample sizes (50, 100, and 200).

**Table 3**  
RMSE results for the proposed sampling schemes for  $p = 0.4$ .

Sample size	RMSE		
	SRD	MaxRSD	MaxVRSD
50	0.3362	<b>0.0868</b>	0.2721
100	0.3415	<b>0.0695</b>	0.0946
200	0.3450	<b>0.0622</b>	0.0929



**Fig. 5.** Sampling schemes for 50 points, from left to right, SRD sampling scheme; MaxRSD sampling scheme; MaxVRSD sampling scheme ( $p = 0.4$ ).



**Fig. 6.** Risk maps of simulated fisheries infractions data from three sampling schemes for 50 points, SRD sampling scheme (left); MaxRSD sampling scheme (center); and MaxVRSD sampling scheme (right) ( $p = 0.4$ ).

For each of the considered sampling designs, we proceed to evaluate the most suitable design for each of the chosen dimensions, using the RMSE measure according to Section 2.4.

Table 3 presents a comparison of the results for the sampling designs SRD, MaxRSD, and MaxVRSD, based on root mean square error (RMSE), considering three distinct sample sizes. The analysis reveals significant differences between the sampling strategies, in terms of point estimation accuracy. Design MaxRSD yielded the lowest values across all sample sizes, indicating greater accuracy in point predictions. For instance, with a sample size of 100, the RMSE was 0.0695 for design MaxRSD, lower than the values of 0.3415 and 0.0946 observed for designs SRD and MaxVRSD, respectively. These results suggest that the structure of design MaxRSD more efficiently captures the underlying process variability, leading to more accurate point estimates.

These results were found despite the low number of successful observations (“1”) for all sample sizes and all designs. It could have been a choice to reduce the success probability considered.<sup>2</sup>

Figs. 5, 7 and 9, illustrate the three sampling schemes analyzed and Figs. 6, 8 and 10 show the fitted values, for  $n = 50$ ,  $n = 100$  and  $n = 200$ , respectively.

Fig. 5 shows the three sampling schemes SRD, MaxRSD, and MaxVRSD for 50 points with  $p = 0.4$ . The results indicate no significant differences in spatial point distribution among the three designs.

Fig. 6 illustrates the spatial distribution of adjusted risk values derived from simulated data on fishing violations, based on three distinct sampling designs: Simple Random Design (SRD; criterion 1), Design MaxRSD (criterion 2), and Design MaxVRSD (criterion 3), using 50 sampling points. The results indicate that Designs MaxRSD and MaxVRSD outperformed the SRD in detecting areas of elevated risk (identified by the nearer larger towns). Specifically, Design MaxRSD delineated high-risk zones extending from the lake region to Tavira, while Design MaxVRSD identified critical risk areas in the vicinity of Vila do Bispo and Albufeira. In contrast, the SRD approach, employed as the reference design, detected only a single high-risk area, located in the Albufeira region.

Fig. 7 shows the three sampling schemes SRD, MaxRSD, and MaxVRSD for 100 points. The results indicate no significant differences in spatial point distribution among the three designs.

Fig. 8 illustrates the spatial distribution of adjusted risk estimates derived from simulated data on fishing violations, using three distinct sampling designs: Simple Random Sampling (SRD; criterion 1), Design MaxRSD (criterion 2), and Design MaxVRSD (criterion 3), based on 100 sampling points. The results indicate that Designs MaxRSD and MaxVRSD were more effective in detecting high-risk areas compared to SRD. Design MaxRSD identified elevated risk zones primarily between the regions of Vila do Bispo and Tavira, whereas Design MaxVRSD revealed critical areas around Lagos, Albufeira, and Tavira. In contrast, the reference design (SRD) detected only a single high-risk zone located in the Albufeira region.

<sup>2</sup> The choice of these specific threshold values was based on historical information on this type of infraction, which is rare and lacks thorough scientific analysis. As such, the thresholds were primarily informed by Navy’s experts working in this field. Their insights allowed us to replicate the most realistic operational scenarios.

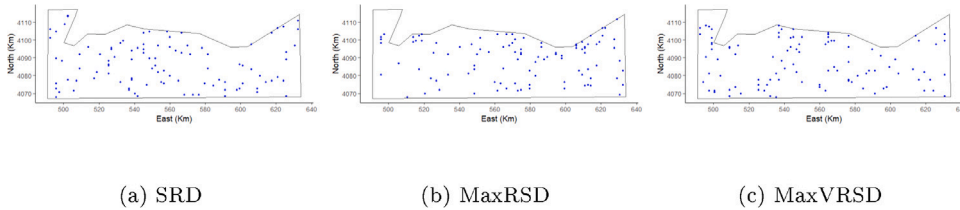


Fig. 7. Sampling schemes for 100 points, from left to right, SRD sampling scheme; MaxRSD sampling scheme; MaxVRSD sampling scheme ( $p = 0.4$ ).

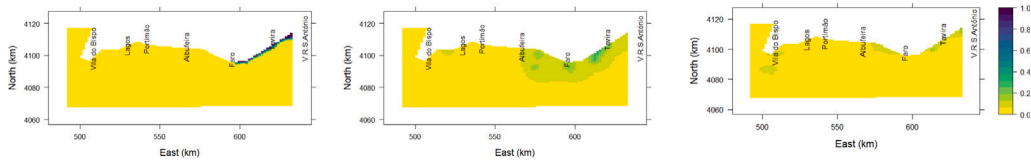


Fig. 8. Risk maps of simulated fisheries infractions data from three sampling schemes for 100 points, SRD sampling scheme (left); MaxRSD sampling scheme (center); and MaxVRSD sampling scheme (right) ( $p = 0.4$ ).

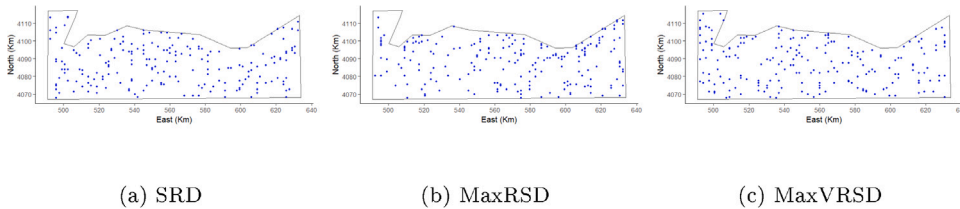


Fig. 9. Sampling schemes for 200 points, from left to right, SRD sampling scheme; MaxRSD sampling scheme; MaxVRSD sampling scheme ( $p = 0.4$ ).

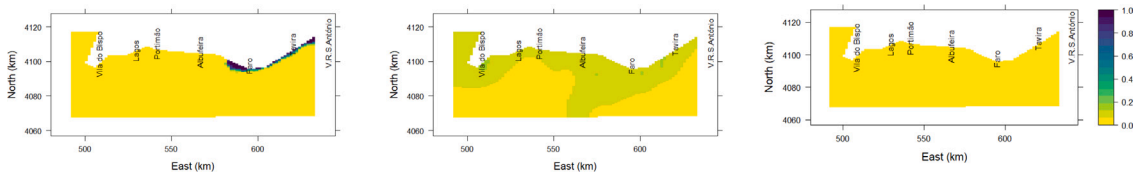


Fig. 10. Risk maps of simulated fisheries infractions data from three sampling schemes for 200 points, SRD sampling scheme (left); MaxRSD sampling scheme (center); and MaxVRSD sampling scheme (right) ( $p = 0.4$ ).

Fig. 9 shows the three sampling schemes SRD, MaxRSD, and MaxVRSD for 200 points. The results indicate no significant differences in spatial point distribution among the three designs.

Fig. 10 presents the spatial distribution of adjusted risk values based on 200 sampling points, again comparing SRD (criterion 1), Design MaxRSD (criterion 2), and Design MaxVRSD (criterion 3). Similar to the previous scenario, Designs MaxRSD and MaxVRSD demonstrated greater sensitivity in identifying high-risk areas. Notably, Design MaxRSD highlighted a continuous stretch of elevated risk along the coastal zone, extending from Vila do Bispo to Vila Real de Santo António. Meanwhile, Design MaxVRSD detected localized high-risk clusters in the vicinities of Vila do Bispo, Lagos, Olhão, and Tavira. The SRD approach, used as the baseline, identified three high-risk zones corresponding to the regions of Vila do Bispo, Lagos, and Vila Real de Santo António.

### 3.4. Surveillance routes

The definition of surveillance routes is a very complex issue. The Portuguese Navy, in its common form of operation, is in charge of the Naval Standard Device that is distributed throughout the various coastal zones of the country so that all means are committed throughout the year. In the case of the Southern Area Command, the requirement is 3 vessels on permanent duty. But this does not prevent a corvette or ocean patrol vessel (NPO) from being committed to the South for inspection, so that all means are committed throughout the year, Fig. 11.

There are several factors that must be taken into consideration when planning enforcement action. When a filter is made for a certain area of the country, the type of surveillance that is carried out in this area will be underlying, that is, the intrinsic characteristics of the vessels that circulate in this area will be taken into consideration. In this case, the graphs with the routes



Fig. 11. Portuguese Navy Naval standard device.  
Source: Portuguese Navy.

presented in Figs. 12 and 13, for the Algarve coastal region takes into account the area’s characteristics, essentially with regard to the available Naval Device resources. As for the factor of available sailing time to survey, there is a maximum period that is determined, not only by the fuel but also by the supplies. And the type of ship also influences the period of time at sea, where we can have a speedboat or a corvette.

Consider the following example: when we have a sample of 50 points, an idea of operationalization in practice is shown in Fig. 12. It shows three proposals of possible surveillance routes made by speedboats allocated or committed to each captaincy of the Algarve coastal region. The first proposed patrol route, defined by the black line, begins in the eastern part of the Algarve coastal region, specifically in Vila Real de Santo António, near the Spanish border, and ends in Olhão; the second proposed surveillance route, defined by the yellow line, starts in the coastal region of Faro and ends in the region of Portimão; and finally, the third proposal starts in the coastal region of Alvor and ends in the western part of the Algarve coastal region, specifically in the region of Vila do Bispo.

Fig. 13, presents another proposal for the definition of inspection routes made by three speedboats and a corvette, where the speedboats do the inspection by captaincies in areas near the coastline and the corvette does the inspection in areas further away from the coastline. The black continuous line represents a proposal of a possible inspection route done with a Corvette, with the

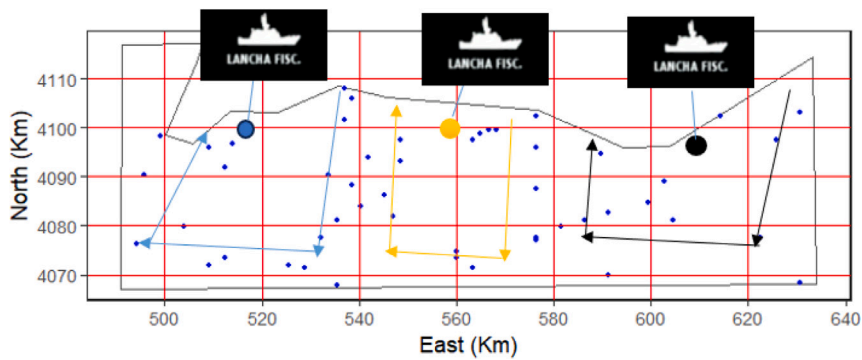


Fig. 12. Maps of the MaxVRSD sampling scheme, for 50 points with three proposed enforcement routes.

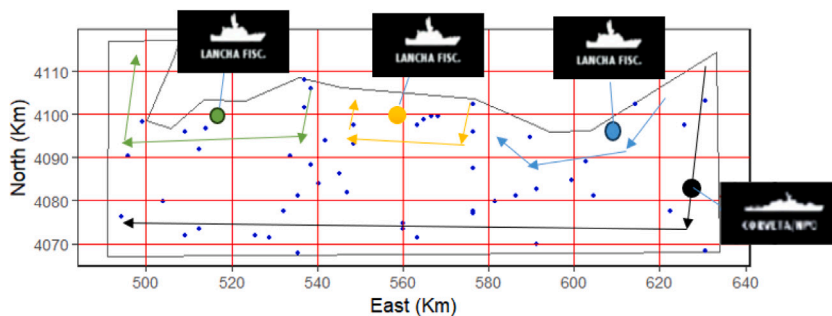


Fig. 13. Maps of the MaxVRSD sampling scheme, for 50 points with four proposed enforcement routes.

beginning of the activity in the eastern area of the Algarve coastal region, specifically in Vila Real de Santo António, near the border with Spain and with the end of the activity in the region of Sagres. The second proposed route of surveillance is represented by the continuous line of blue color, made by a speedboat, with the beginning of activity in the region of Tavira and with the end in the region Faro. The third proposed route is represented by the yellow line, with the beginning of activity in the coastal region of Quarteira and with the end in the region Albufeira. Fourth proposed surveillance route starts in Lagos and ends in the west of the Algarve coastal region, specifically in the region of Vila do Bispo.

#### 4. Conclusions and future work

In this paper, we present model-based sampling designs for geostatistical binary data.

In geostatistics one of the designs that is often used for sample selection is the systematic sampling design (Lin and Clayton, 2005). However, for this study, although this sampling scheme has been tried, since we have few sampling sites with high estimated probability, the scheme had difficulty in finding points that were informative in terms of the probability of the infraction, making it clear that this was not a good strategy for this particular problem.

Instead, results from the SRD (Lohr, 2021) were presented and used to compare to the proposed sampling design criteria, Maximum Risk Sampling design (MaxRSD) and Maximum Variance Risk Sampling Design (MaxVRSD) through RMSE (Simbahan and Dobermann, 2006) in order to verify the best design for the problem under analysis in this study.

Some other sampling design selection criteria have been used in geostatistical studies, such as MMSD and WM (Liu et al., 2009; Simbahan and Dobermann, 2006; Lin and Clayton, 2005; Warrick and Myers, 1987). These selection criteria are based on minimizing the mean of the shortest distance between two points and the uniform distribution of pairs of points for variogram estimation. The sampling design selection criteria proposed in this work are rather based on predicted probability and estimated variability. These sampling design selection criteria have a clear advantage of being able to sample more in places where the probability of committing presumed offenses is higher and where prevalence is estimated with higher uncertainty.

The sampling design is extremely important to answer the objectives of a given study, and its answer is limited mainly by the type of design employed. For this study, the model-based approach was used because it is the most appropriate in research of this nature. According to the RMSE values described in Table 3, the two proposed sampling design criteria are the best, and thus may be a good alternative for the problem under study.

Resorting to the risk maps associated with presumed fishing violations in the Southern Area Command with the simulated data, we can highlight which areas present a high risk of committing presumed fishing violations, which is reflected in the most optimal

routes, in order to maximize the probability of finding a higher number of presumed fishing perpetrators with less resource efforts, one of the main objectives of this work.

The application of these designs presupposes that we have these designs operationalized in practice, and this operationalization depends on a number of factors, such as the sailing time available for inspection, which depends on fuel and supplies; the availability of boats and ships for this type of activity and the characteristics of the area to be inspected.

The proposed sampling designs may assist in the design of possible enforcement routes, using the criteria presented in 2.3. We thus obtain a subset of points, taking into consideration the starting point of enforcement and the region of operation (subset of the southern zone command in this case) planned for that day/days on that route. Routes can be assigned to a small unit (Speedboats) or a large unit (Corvettes). After defining the enforcement route, the most appropriate route is assigned to a naval unit available to enforce this zone optimizing the scheduling that increases the probability of finding a high number of violators with less resource effort.

For future work, it is planned to extend the present approach to other regions of the Portuguese coast and to similar situations (other applications where it is necessary to identify potential offenders) or study of other types of infractions; and surveillance routes can be defined taking into account the factors of available sailing time, which depends on fuel and supplies; availability of speedboats and ships for surveillance; and the characteristics of the area to be surveyed. Since we are dealing with unbalanced binary data, the application of Inhomogeneous Spatial Poisson Point Process models might be recommended (Zuur et al., 2017). For future investigation, we can also consider extending and using two additional approaches: a marked Log-Gaussian Cox Process (LGCP) and a spatio-temporal approach. Both methodologies can provide more robust and flexible modeling for analyzing spatial and temporal patterns of infractions, particularly in contexts of surveillance and monitoring of illegal activities.

## Acknowledgments

This work was supported by national fund through FCT - Fundação para a Ciência e a Tecnologia, I.P., under projects PREFERENTIAL, PTDC/MAT-STA/28243/2017 and Center for Mathematics and Applications - NOVA MATH, UIDP/00297/2020. We thank the NOVA MATH and the Portuguese Navy for making this work possible, within the scope of the protocol established between them.

## References

- Blangiardo, M., Cameletti, M., 2015. *Spatial and Spatio-Temporal Bayesian Models with R-INLA*. John Wiley & Sons.
- Bogaert, P., Russo, D., 1999. Optimal spatial sampling design for the estimation of the variogram based on a least squares approach. *Water Resour. Res.* 35 (4), 1275–1289.
- Brus, D.J., 2022. *Spatial Sampling with R*. In: *The R Series*, CRC Press, Boca Raton.
- Brus, D., De Gruijter, J., 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma* 80 (1–2), 1–44.
- Chipeta, M.G., Rowlingson, B., Diggle, P.J., 2019. Geosample: an R package for geostatistical sampling designs. Under review.
- Chipeta, M., Terlouw, D., Phiri, K., Diggle, P., 2016. Adaptive geostatistical sampling enables efficient identification of malaria hotspots in rural Malawi. *Am. J. Trop. Med. Hyg.* 95 (1), 134–140. <http://dx.doi.org/10.4269/ajtmh.15-0333>.
- Cochran, W.G., 1977. *Sampling Techniques*. John Wiley & Sons.
- De Gruijter, J., Ter Braak, C., 1990. Model-free estimation from spatial samples: a reappraisal of classical sampling theory. *Math. Geol.* 22 (4), 407–415.
- Diggle, P., Lophaven, S., 2006. Bayesian geostatistical design. *Scand. J. Stat.* 33 (1), 53–64.
- Dumelle, M., Higham, M., Ver Hoef, J.M., Olsen, A.R., Madsen, L., 2022. A comparison of design-based and model-based approaches for finite population spatial sampling and inference. *Methods Ecol. Evol.* 13 (9), 2018–2029.
- Kabaghe, A., Chipeta, M., McCann, R., Phiri, K., van Vugt, M., Takken, W., Diggle, P., Terlouw, D., 2017. Adaptive geostatistical sampling for malaria hotspots identification in the context of a national malaria control program. *Malar. J.* 16 (1), 1–12. <http://dx.doi.org/10.1186/s12936-017-2044-0>.
- Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., Rue, H., 2018. *Advanced Spatial Modeling with Stochastic Partial Differential Equations using R and INLA*. Chapman and Hall/CRC.
- Lin, P.-S., Clayton, M.K., 2005. Analysis of binary spatial data by quasi-likelihood estimating equations. *Ann. Statist.* 33 (2), 542–555.
- Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *J. Royal Statist. Soc. Series B: Statist. Method.* 73 (4), 423–498.
- Liu, Y., Chen, Y., Cheng, J., 2009. A comparative study of optimization methods and conventional methods for sampling design in fishery-independent surveys. *ICES J. Mar. Sci.* 66 (9), 1873–1882.
- Lohr, S.L., 2021. *Sampling: Design and Analysis*. Chapman and Hall/CRC.
- Miguel, B., Simões, P., de Deus, R.G., Natário, I., 2024a. Bayesian geostatistics modeling of maritime surveillance data. In: *International Conference on Computational Science and Its Applications*. Springer, pp. 187–205.
- Miguel, B., Simões, P., de Deus, R.G., Natário, I., 2024b. Bayesian geostatistics modeling of maritime surveillance data. In: *International Conference on Computational Science and Its Applications*. Springer, pp. 187–205.
- Pearse, A.R., McGree, J.M., Som, N.A., Leigh, C., Maxwell, P., Ver Hoef, J.M., Peterson, E.E., 2020. Snsdesign: An R package for pseudo-Bayesian optimal and adaptive sampling designs on stream networks. *PLoS One* 15 (9), e0238422.
- Petitgas, P., 2001. Geostatistics in fisheries survey design and stock assessment: models, variances and applications. *Fish. Fish.* 2 (3), 231–249.
- Rivoirard, J., Simmonds, J., Foote, K., Fernandes, P., Bez, N., 2008. *Geostatistics for Estimating Fish Abundance*. John Wiley & Sons.
- Rocha, A.D., Groen, T.A., Skidmore, A.K., Willemen, L., 2020. Role of sampling design when predicting spatially dependent ecological data with remote sensing. *IEEE Trans. Geosci. Remote Sens.* 59 (1), 663–674.
- Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 71 (2), 319–392.
- Russo, D., 1984. Design of an optimal sampling network for estimating the variogram. *Soil Sci. Am. J.* 48 (4), 708–716.
- Särndal, C.-E., Thomsen, I., Hoem, J.M., Lindley, D., Barndorff-Nielsen, O., Dalenius, T., 1978. Design-based and model-based inference in survey sampling [with discussion and reply]. *Scand. J. Stat.* 27–52.

- Silva, D., Menezes, R., Serra-Pereira, B., Azevedo, M., Figueiredo, I., 2023. Adapting the sampling design of research surveys to improve the biomass estimation of secondary species—the case study of *Raja clavata*. *ICES J. Mar. Sci.* 80 (10), 2567–2578.
- Simbahan, G.C., Dobermann, A., 2006. Sampling optimization based on secondary information and its utilization in soil carbon mapping. *Geoderma* 133 (3–4), 345–362.
- Simões, P., Carvalho, M.L., Figueiredo, I., Monteiro, A., Natário, I., 2021. Geostatistical sampling designs under preferential sampling for black scabbardfish. In: *International Conference on Congress of the Portuguese Statistical Society*. Springer, pp. 137–151.
- Stein, A., Ettema, C., 2003. An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons. *Agric. Ecosyst. Environ.* 94 (1), 31–47.
- Van Groenigen, J.W., Siderius, W., Stein, A., 1999. Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma* 87 (3–4), 239–259.
- Warrick, A., Myers, D., 1987. Optimization of sampling locations for variogram calculations. *Water Resour. Res.* 23 (3), 496–500.
- Zuur, A.F., Ieno, E.N., Saveliev, A.A., 2017. *Spatial, Temporal and Spatial-Temporal Ecological Data Analysis with R-INLA*, vol. 1, Highland Statistics Ltd.