Adaptive Sampling Design for Bayesian Spatial Risk Mapping of Maritime Surveillance in Central Portugal

Belchior Miguel[1,2(✉)], Isabel Natário[1,3], Rui Gonçalves de Deus[4,5], and Paula Simões[1,6,7,8]

[1] Center for Mathematics and Applications (NOVA Math),NOVA University of Lisbon, Lisbon, Portugal
b.miguel@alumni.fct.unl.pt, {icn,pc.simoes}@fct.unl.pt
[2] Instituto Superior Politécnico de Tete, Tete, Mozambique
[3] Department of Mathematics of NOVA School of Science and Technology, Caparica, Portugal
[4] Centro de Investigação Naval (CINAV), Instituto Universitário Militar, Lisbon, Portugal
[5] Direção de Análise e Gestão da Informação, Marinha Portuguesa, Lisbon, Portugal
goncalves.deus@marinha.pt
[6] Centro de Investigação, Desenvolvimento e Inovação da Academia Militar (CINAMIL), Lisbon, Portugal
[7] Instituto Universitário Militar, Lisbon, Portugal
[8] Department of Mathematics, Lisbon Institute of Engineering (ISEL), Polytechnic University of Lisbon, Lisbon, Portugal

Abstract. The Portuguese Navy faces the important operational challenge of contributing to more effective and coordinated management of inspection and control across the Exclusive Economic Zone and national borders, as part of its mission to ensure maritime surveillance and enforce the law.

From an operational perspective, patrol activity tends to concentrate in areas characterised by intense maritime traffic, proximity to ports, and zones traditionally associated with fishing activities. The spatial distribution of inspected events or detected infringements reflect not only the pressure exerted by maritime activity but also the intensity and frequency of enforcement actions. Consequently, the spatial distribution of enforcement operations and recorded infringements contributes both the level of maritime activity and the surveillance strategy adopted. Within this framework, factors such as distance from the coastline, proximity to ports, and the spatial extent of the Exclusive Economic Zone (EEZ) act as structural drivers of the variability observed in enforcement data. Using geostatistical methods combined with INLA (Integrated Nested Laplace Approximation) approach jointly with SPDE (Stochastic Partial Differential Equations) formulation, this study explores the spatial distribution of suspected fishing infractions along the central coast of Portugal.

Some sampling design approaches are proposed for the surveillance of

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16760, pp. 51–68, 2027.
https://doi.org/10.1007/978-3-032-30521-3_4

fishing activities, providing adequate support for the Portuguese Navy's maritime surveillance work, thereby supporting efficient and properly coordinated protection of the maritime domain and borders.

Keywords: Bayesian geostatistics · Presence/absence spatial data · Adaptive sampling design · Fisheries-related infractions · Maritime surveillance · Risk mapping

1 Introduction

The surveillance and control of maritime areas under Portuguese sovereignty and jurisdiction are based on an integrated institutional framework combining naval capabilities, maritime authority, and specialised policing. This structure aims to ensure the State's presence at sea, guarantee navigational safety, protect marine resources, and enforce legislation applicable to maritime activities, including fisheries, transport, recreational boating, and resource exploitation.

The naval component is ensured by the Portuguese Navy, which operates assets specifically designed for patrol, inspection, and support to public-interest operations. Among these, the offshore patrol vessels stand out, as they are designed to operate for extended periods at sea and to cover large areas of the Exclusive Economic Zone (EEZ). These vessels perform State authority functions at sea, including surveillance, inspection, support to search and rescue (SAR) operations, and cooperation with other national and international entities [1,2].

In parallel, within the framework of the Maritime Authority System, the National Maritime Authority and the Maritime Police exercise inspection and specialised policing powers in maritime areas, port zones, and the maritime public domain. Their actions focus on ensuring compliance with legal and regulatory provisions, contributing to maritime safety, environmental protection, and the orderly management of activities conducted along the coastline and within the territorial sea [2].

At the coastal level, the Coastal Control Unit of the National Republican Guard complements this framework through surveillance, patrol, and interception activities, both on land and within the territorial sea. This institutional coordination enables broader operational coverage and strengthens the capacity to respond to infringements, irregularities, and risk situations [3].

From an operational perspective, patrol activity tends to concentrate in areas characterised by intense maritime traffic, proximity to ports, and zones traditionally associated with fishing activities. Consequently, the spatial distribution of inspected events or detected infringements may reflect not only the pressure exerted by maritime activity but also the intensity and frequency of enforcement actions. This aspect is particularly relevant in quantitative and spatio-temporal modelling, where event occurrence may be associated both with exposure to risk and with the probability of inspection.

The organisational and operational structure of the Portuguese maritime system thus reflects an integrated model of ocean governance, combining strategic naval presence, specialised policing, and interinstitutional coordination, with

objective of ensuring State authority, maritime safety, and the sustainability of economic activities conducted within national maritime space [4,5].

This study explores the spatial distribution of suspected fishing-related infractions along the central coast of Portugal using geostatistical methods. Based on georeferenced binary observations (presence/absence), the analysis seeks to estimate and visualise the probability of infraction occurrence across the study area. In addition, the study evaluates sampling design strategies derived from geostatistical modelling specifically the Maximum Risk Sampling Design (MaxRSD) and the Maximum Variance Risk Sampling Design (MaxVRSD) to support the analysis of binary data on fishing-related infractions along the Portuguese coast.

Regression models for binary response variables are widely used in many scientific applications, with logistic regression representing one of the most common modelling approaches [6]. These models fall within the class of generalized linear models (GLMs), which relate the expected value of the response variable to a set of covariates through an appropriate link function [7].

In a spatial framework, binary geostatistical data represent observations indicating the presence or absence of a phenomenon at specific locations, allowing the study region to be partitioned into areas where the event occurs and areas where it does not [8]. Spatial dependence in such data is commonly addressed using hierarchical Bayesian spatial models, where spatial effects are introduced through latent Gaussian random fields incorporated into the linear predictor, leading to generalized linear mixed models.

In this study, geostatistical modelling for presence/absence data is implemented using the INLA (Integrated Nested Laplace Approximation) approach combined with SPDE (Stochastic Partial Differential Equations) formulation. This framework provides an efficient representation of the spatial random field and enables computationally efficient Bayesian inference. The models are fitted using the R-INLA framework, allowing the estimation of the spatial field (posterior mean and standard deviation) and the construction of spatial risk maps associated with the analysed infractions [9].

A related study was previously conducted by Miguel et al. [10] for the Algarve region. The present work extends this research by using new datasets, additional explanatory variables, and updated inspection routes, allowing the proposed models to be evaluated in a different geographical context. Furthermore, new sampling design criteria are introduced and compared with a stratified sampling design. The predictive performance of the competing approaches is assessed using RMSE, Log-Score, and LGOCV.

The remainder of the paper is structured as follows. Section 2 outlines the theoretical background of geostatistical models for binary data. Section 3 presents the modelling framework applied to the maritime surveillance data, including the spatial models and the proposed adaptive sampling design criteria, together with the corresponding results. Section 4 discusses the main findings and concludes the study.

2 Sampling Designs for Surveillance Fishing Activities

Sampling design plays a fundamental role in spatial studies because the configuration of sampling locations directly influences both parameter estimation and spatial prediction. Two main paradigms are commonly considered in spatial sampling: the *design-based* approach and the *model-based* approach [11,12].

In the design-based approach, inference relies on the randomization embedded in the sampling design. The population values are considered fixed but unknown, and the validity of statistical inference depends on the probabilistic properties of the sampling scheme. This approach is typically preferred when the objective is to estimate global or regional characteristics of a population, such as spatial cumulative distribution functions or averages over predefined areas, where statistical validity is prioritized over efficiency.

In contrast, the model-based approach treats the spatial phenomenon as a realization of an underlying stochastic process [12]. Within geostatistics, it assumes that the observed spatial data correspond to a single realization of an underlying spatial process defined over the study region [13]. The only source of stochasticity is the assumed spatial process itself, and inference depends on the adequacy of the specified model. This approach is particularly suitable when the objective is to produce spatial maps or predict values at unsampled locations, where predictive accuracy and efficiency are of primary importance.

Sampling design within this framework often involves selecting sampling locations in order to optimize a given objective function [14,15]. Common criteria include the minimization of prediction variance, the minimization of estimation variance for model parameters, or the improvement of spatial prediction accuracy [16]. Alternative strategies have also been proposed that do not require prior knowledge of the covariance structure, focusing instead on improving the estimation of the semivariogram, which is a key component of geostatistical analysis [14,17,18]. In practice, preliminary data exploration or empirical semivariogram estimates may guide the design without requiring a fully specified covariance model.

Although many optimal design approaches assume that the parameters governing spatial dependence are known, this assumption is often unrealistic in practice. When little or no prior information about the spatial structure is available, preliminary data may be used to estimate the covariance structure before refining the sampling design. This flexibility is particularly important in applications where the spatial dependence of the phenomenon is initially unknown.

Importantly, probability-based sampling designs can also be applied when the variable of interest is binary, such as in prevalence studies or in the detection of rare events [19]. In such cases, the spatial distribution of occurrences (e.g., presence or absence of an event) can be analyzed within a geostatistical modelling framework.

Under the model-based paradigm adopted in this work, the sampling locations define the spatial support on which the stochastic model is observed. Consequently, the design of the sampling network influences the amount of information available to estimate the parameters of the geostatistical model and the latent

spatial field introduced in the previous section. In particular, the spatial configuration of the sampled locations affects the estimation of the covariance structure and the accuracy of spatial predictions. Therefore, sampling design and geostatistical modelling should be considered jointly when developing spatial statistical analyses.

2.1 Geostatistical Model for Infractions Detection

The traditional geostatistical formulation can be extended to non-Gaussian outcomes within a generalized linear mixed model framework, as proposed by Diggle *et al.* [20]. In the case of binary responses (as for infraction detection), the observation model assumes that

$$Y_i \sim \text{Bernoulli}(\pi_i), \quad i = 1, \ldots, n, \tag{1}$$

where $\pi_i = P(Y_i = 1 \mid s_i)$ represents the probability of an event (e.g., an alleged infraction) occurring at location s_i.

The Bernoulli success probability is linked to a linear predictor through an appropriate link function. While the logit link is commonly adopted, the complementary log-log (cloglog) link is often preferred when modeling rare events. Under the cloglog specification, the model is written as

$$\text{cloglog} = \log\left(-\log(1-\pi_i)\right) = \beta_0 + \mathbf{x}_i'\boldsymbol{\beta} + u_i, \quad i = 1, \ldots, n, \tag{2}$$

where β_0 denotes the intercept, $\mathbf{x}_i$ is a vector of covariates observed at location s_i, $\boldsymbol{\beta}$ is the corresponding vector of regression coefficients, and $u_i = U(s_i)$ represents a spatial random effect. $\{U(s) : s \in D \subset \mathbb{R}^2\}$ is a latent Gaussian stochastic process describing the underlying spatial phenomenon, with zero-mean and Matérn covariance function defined, for two locations s_i and s_j separated by Euclidean distance $h = \|s_i - s_j\|$, as,

$$C(h) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\kappa h)^\nu K_\nu(\kappa h), \tag{3}$$

where σ^2 denotes the marginal variance of the spatial field, $\nu > 0$ is a smoothness parameter controlling the differentiability of the process, $\kappa > 0$ is a scale parameter related to the spatial range, and $K_\nu(\cdot)$ denotes the modified Bessel function of the second kind. The Matérn family provides a flexible class of covariance functions capable of representing a wide range of spatial dependence structures [21,22].

This model has a hierarchical representation consisting of the data model $[\boldsymbol{Y} \mid \boldsymbol{w}]$, in (1) and (2), and of the latent spatial structure $[\boldsymbol{w} \mid \boldsymbol{\theta}]$, given by the vector of latent Gaussian effects $\boldsymbol{w} = (\beta_0, \boldsymbol{\beta}, \mathbf{u})$. Vector $\boldsymbol{\theta}$ denotes the set of hyperparameters controlling the covariance structure. Under a Bayesian framework, prior distributions are assigned to all components of $\boldsymbol{w}$ and $\boldsymbol{\theta}$, completing the hierarchical specification through the prior model $[\boldsymbol{w} \mid \boldsymbol{\theta}]\,[\boldsymbol{\theta}]$.

2.2 Adaptive Sampling Design Criteria for Presence/absence Spatial Data

Design-based and model-based approaches are widely used in classical geostatistical problems, particularly when the response variable follows a Gaussian distribution. However, the present study considers presence/absence spatial data, and therefore some sampling design criteria proposed for continuous data may not be directly applicable in this context.

Adaptive geostatistical sampling provides a flexible framework in which new sampling locations are selected sequentially according to the information obtained in previous stages of the study. As discussed by Chipeta et al. [23], Pearse et al. [24], Simões et al. [25], and Silva et al. [26], this strategy allows the data collection process to focus on regions that are expected to provide more informative observations for the objectives of the analysis.

In general, constructing an adaptive geostatistical sampling design for binary data involves three main steps: (i) fitting a geostatistical model to the observed data; (ii) using the fitted model to generate predictions over a grid of potential sampling locations; and (iii) selecting new sampling points according to a predefined selection criterion.

Within this framework, two adaptive sampling criteria are used for binary geostatistical data: the Maximum Risk Sampling Design (MaxRSD) and the Maximum Variance Risk Sampling Design (MaxVRSD), introduced by Miguel, et al. [27].

Maximum Estimated Risk Sampling Design (MaxRSD).

The MaxRSD criterion guides the sampling process towards regions where the estimated probability of occurrence of the phenomenon is larger, meeting the objective of monitoring areas with higher risk of event occurrence. Being the spatial dependence structure initially unknown, a geostatistical model is firstly fitted to the available data and used to predict probability of the event under consideration (risk) on a dense grid of points covering the study region, defining a set of candidate sampling locations. A fixed number of sampling points n is subsequently selected from this grid, assigning to each point a selection probability proportional to the predicted probability of occurrence of the event. As a result, sampling tends to concentrate in areas with higher estimated risk. This criterion depends on the number of sampling points n and on the fitted model m, and is denoted by $\text{MaxRSD}(n, m)$.

Maximum Variance Risk Sampling Design (MaxVRSD).

As an alternative strategy, the MaxVRSD criterion directs the sampling process towards regions where the uncertainty associated with the estimated risk is higher. Similarly to the previous method, a geostatistical model is first fitted to the observed data and risk predictions are obtained over a grid of candidate locations covering the study region. These predictions provide both the estimated probabilities of event occurrence and their associated standard errors. New sampling locations are then selected with probability proportional to the magnitude of the standard errors of the predicted probabilities, thereby concentrating data

collection in areas where the model presents greater uncertainty. This criterion also depends on the number of sampling points n and the fitted model m, and is denoted by MaxVRSD(n, m).

For comparison purposes, a Stratified Random Sampling (SRS) design is considered as a reference method. This sampling strategy is widely used in fisheries-related studies and is well supported by classical sampling theory [28]. In a spatial context, this approach divides the study region into several sub-regions or strata that are relatively homogeneous with respect to the variable of interest. A sampling location is then randomly selected within each stratum, ensuring that all parts of the study area are represented in the sample and providing balanced spatial coverage.

In this study, stratified random sampling is used as a benchmark to evaluate the performance of the proposed adaptive criteria, namely MaxRSD and MaxVRSD. Unlike these adaptive approaches, which select new sampling locations based on the estimated risk or the uncertainty of model predictions, the stratified design follows a fixed strategy based solely on the spatial partition of the study area.

3 Sampling Design Performance Evaluation

The performance of the different sampling designs considered in this study is assessed based on the predictive quality of the geostatistical models fitted to the data. For this purpose, three commonly used predictive performance measures are considered: the Root Mean Squared Error (RMSE), the Logarithmic Score (Log-Score), and the Leave-Group-Out Cross-Validation (LGOCV). These metrics allow different aspects of predictive performance to be evaluated, namely the accuracy of point predictions, the quality of the predictive distribution, and the model's generalization ability.

3.1 Root Mean Squared Error (RMSE)

The Root Mean Squared Error (RMSE) is a widely used measure to quantify the discrepancy between observed values and model predictions. This metric provides an indication of prediction accuracy, assigning greater penalties to larger errors due to the squaring of the differences. Formally, RMSE is defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{4}$$

where y_i denotes the observed value at location s_i, $\hat{y}_i$ represents the predicted value obtained from the geostatistical model for the same location, and n is the total number of observations considered. Lower RMSE values indicate better predictive performance [29,30].

3.2 Logarithmic Score (Log-Score)

While RMSE evaluates only the distance between observed values and point predictions, the Logarithmic Score (Log-Score) assesses the quality of the full predictive distribution associated with the model. This metric is based on the probability assigned by the model to the observed values. The Log-Score is defined as

$$\text{Log-Score} = -\frac{1}{n}\sum_{i=1}^{n} \log f(y_i \mid \hat{\theta}), \tag{5}$$

where $f(y_i \mid \hat{\theta})$ denotes the predictive probability density (or probability mass) of the observed value y_i under the model estimated with parameters $\hat{\theta}$. Smaller values of the Log-Score indicate better predictive performance, since they imply that the model assigns higher probability to the observed data [9,31].

3.3 Leave-Group-Out Cross-Validation (LGOCV)

Leave-Group-Out Cross-Validation (LGOCV) is used to evaluate the generalization capability of the model. In this procedure, the dataset is divided into K groups. At each iteration, one group is removed from the training data, the model is fitted to the remaining data, and predictions are then generated for the excluded observations. The overall LGOCV error can be expressed as

$$\text{Error}_{LGOCV} = \frac{1}{K}\sum_{k=1}^{K} L\left(y^{(k)}, \hat{y}^{(-k)}\right), \tag{6}$$

where $y^{(k)}$ represents the observations belonging to group k, $\hat{y}^{(-k)}$ denotes the predictions obtained from the model fitted without that group, and $L(\cdot)$ represents an appropriate loss function, such as RMSE or Log-Score [32].

This procedure provides a robust way to assess predictive performance, reducing the risk of overfitting and offering a more realistic estimate of the model's ability to predict new observations.

4 Maritime Fishery Surveillance at Central of Portugal

The Portuguese naval deployment is characterised by operational flexibility, with no permanent public allocation of assets to a specific geographical area. Resources are distributed according to strategic priorities, the intensity of maritime activity, and enforcement requirements. In the central region of Portugal, extending between Figueira da Foz, Lisbon, and Sines, maritime surveillance is ensured by different types of vessels operated by the Portuguese Navy and by units of the National Maritime Authority [2].

Within the Navy, Coastal Patrol Vessels (CPVs) include vessels such as NRP[1] Zaire, NRP Tejo, NRP Douro, and NRP Mondego (See the Fig. 1). These smaller

[1] Portuguese Republic Ships - NRP

and highly manoeuvrable vessels are particularly suited to missions within the territorial sea and coastal waters. Their operational role focuses on fisheries inspection, routine boarding operations, maritime traffic control, and reinforcing the State's presence in areas with a high concentration of vessels [33].

In addition, Offshore Patrol Vessels (OPVs) of the Viana do Castelo class, namely NRP Sines and NRP Figueira da Foz, operate in the region. These vessels have greater autonomy, extended endurance at sea, and the capability to operate both in the territorial sea and across the Exclusive Economic Zone (EEZ). Their missions include large-scale surveillance, offshore inspection and control, support to search and rescue (SAR) operations, environmental monitoring, and interinstitutional cooperation. Their enhanced operational capacity allows coverage of extensive areas of the central coastline and adjacent offshore waters [33].

From a spatial standpoint, patrol resources are distributed according to their operational roles. Coastal patrol units and Maritime Police vessels are typically deployed near the shoreline and around major ports, whereas offshore patrol vessels cover wider maritime areas, including distant offshore zones. Consequently, the spatial distribution of enforcement operations and recorded infringements may reflect both the level of maritime activity and the surveillance strategy adopted. Within this framework, factors such as distance from the coastline, proximity to ports, and the spatial extent of the Exclusive Economic Zone (EEZ) can act as structural drivers of the variability observed in enforcement data.



Fig. 1. Vessels of the naval deployment

In this section data are presented and modelled, allowing the sampling design methods to be applied.

4.1 Data

The dataset used in this study consists of georeferenced records of inspection activities carried out by the Portuguese Navy along the central region of the Portuguese coast over a five-year period (2013−2017). Inspection operations were organized into four daily periods of six hours each: 00:01−06:00, 06:01−12:00, 12:01−18:00, and 18:01−24:00.

The analysis focuses on three types of fishing-related infractions, recorded in binary form (presence/absence): the use of prohibited fishing gear (infraction 3),

fishing in restricted areas (infraction 4), and fishing prohibited due to excessive engine power or vessel tonnage (infraction 5).

Figure 2 illustrates the spatial distribution of inspection locations recorded during the study period. Red points indicate locations where suspected infractions were detected, while black points represent locations where no infractions were identified.

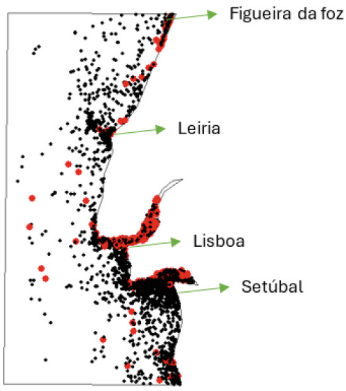


Fig. 2. Maritime data distribution map of the Figueira da Foz, Lisbon, and Sines coastal region

4.2 Geostatistical Model for Presence/absence Data

In this study, the available data correspond to binary observations indicating the presence or absence of alleged fishing infractions at specific locations [10]. As discussed in Sect. 2.1, these can be described by the Geostatistical model for binary data in (1), (2) and (3).

A total of 19 models were fitted, as presented in Table 1, with the aim of comparing different model specifications. The models were fitted efficiently using R-INLA and compared using the Deviance Information Criteria, DIC. Among these, Model O stood out by achieving the lowest DIC value (5687.46), indicating the best balance between model fit and complexity. This model incorporates distance, year, and period as covariates, along with a spatial random effect, allowing for an effective representation of the spatial variability of the phenomenon under study. Consequently, this model was selected as the most appropriate, and its results are presented next.

The estimation results for the fixed effects associated with the covariates distance to the coast, year, and period in Model O are presented in Table 2, together with the estimated hyperparameters. The variance explained by the Matérn spatial field (σ_u^2) was estimated at 3.725, while the posterior mean of the range parameter (R) was 8.477 km, with a 95% credibility interval of (5.616,

Table 1. Fitted models and corresponding DIC

Model	Linear Predictor	DIC
A	$\beta_0 + \boldsymbol{u}$	6593.49
B	$\beta_0 +$ period $+ \boldsymbol{u}$	6524.66
C	$\beta_0 +$ year $+ \boldsymbol{u}$	6551.73
D	$\beta_0 +$ year + period $+ \boldsymbol{u}$	6482.10
E	$\beta_0 +$ bathymetry + year + period $+ \boldsymbol{u}$	6543.86
F	$\beta_0 +$ bathymetry $+ \boldsymbol{u}$	6646.52
G	$\beta_0 +$ bathymetry + year $+ \boldsymbol{u}$	6604.05
H	$\beta_0 +$ bathymetry + period $+ \boldsymbol{u}$	6584.34
I	$\beta_0 +$ bathymetry + year + period	7970.26
J	$\beta_0 +$ bathymetry + year	8032.84
K	$\beta_0 +$ bathymetry + period	8017.94
L	$\beta_0 +$ bathymetry	8076.14
M	$\beta_0 +$ bathymetry + slope + distance + year + period $+ \boldsymbol{u}$	5689.46
N	$\beta_0 +$ slope + distance + year + period $+ \boldsymbol{u}$	5688.88
O	$\beta_0 +$ distance + year + period $+ \boldsymbol{u}$	**5687.46**
P	$\beta_0 +$ distance + period $+ \boldsymbol{u}$	5723.40
Q	$\beta_0 +$ distance + year $+ \boldsymbol{u}$	5738.05
R	$\beta_0 +$ distance $+ \boldsymbol{u}$	5772.27
S	$\beta_0 +$ year + period $+ \boldsymbol{u}$	5691.35

12.235). The effects of the working periods of the day, as well as all years from 2013 to 2017, were found to be statistically significant.

Figure 3 presents the estimated risk surfaces associated with fishing-related infractions for the year 2014. Overall, a clear spatial pattern is observed, with higher-risk areas concentrated along the coastal strip. Spatial dependence is pronounced, as reflected in the smooth transitions between low- and high-probability zones. During nighttime periods, shown in panels (a) and (b) for the year of 2014, higher-intensity areas are particularly evident in the northern (Leiria) and central (Lisbon) coastal regions, appearing slightly more extensive than in previous years, which suggests an expansion of the risk area. In contrast, daytime periods, represented in panels (c) and (d) for the year of 2014, exhibit a more diffuse spatial distribution, although localized increases in probability are observed in areas coinciding with nighttime hotspots. This indicates that certain high-risk locations remain consistently active across different periods of the day. The estimated risk surfaces associated with presumed fishing infractions for the years of 2013, 2015, 2016 and 2017 are consistent with that of year 2014 just described, only with a somewhat decreased density of occurrences for the year of 2013.

Considering the overall study period (2013−2017), the results reinforce the existence of a consistent and persistent spatial pattern, with recurring hotspots

Table 2. Summary of the parameter's posterior distributions of the model O

Variable	Mean	SD	Median	95% Credibility Interval
(Intercept)	−1.715	0.193	−1.705	(−2.124, −1.365)
dist_coast_std	−0.225	0.158	−0.226	(−0.535, 0.086)
Year 2014	0.389	0.106	0.389	(0.180, 0.597)
Year 2015	0.521	0.117	0.521	(0.292, 0.750)
Year 2016	0.755	0.120	0.755	(0.519, 0.992)
Year 2017	0.266	0.134	0.266	(0.004, 0.529)
06:01–12:00	−1.122	0.157	−1.122	(−1.431, −0.813)
12:01–18:00	−1.083	0.161	−1.083	(−1.400, −0.767)
18:01–24:00	−0.660	0.186	−0.660	(−1.024, −0.296)
σ_u^2	3.725	0.790	3.637	(2.422, 5.515)
R	8.477	1.691	8.319	(5.616, 12.235)

along the coastline. Risk levels are systematically higher and more concentrated during nighttime periods, while daytime patterns tend to be more diffuse and less intense.

4.3 Sampling Design

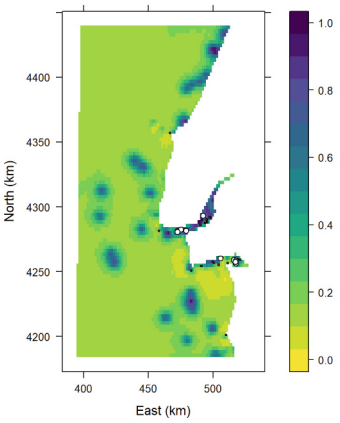
After obtaining the fitted value maps of fishing-related infractions, the same model is used to construct the sampling designs. In this context, the MaxRSD and MaxVRSD designs proposed by Miguel et al. [27] are considered and compared with the SRS design, assuming a success probability fixed at $p = 0.2$.

Figure 6 in the appendix, compares the SRS, MaxRSD and MaxVRSD sampling designs for 50 points, highlighting distinct spatial patterns. SRS provides a random coverage, MaxRSD focuses on high-risk areas, and MaxVRSD balances risk and uncertainty, supporting improved model estimation.

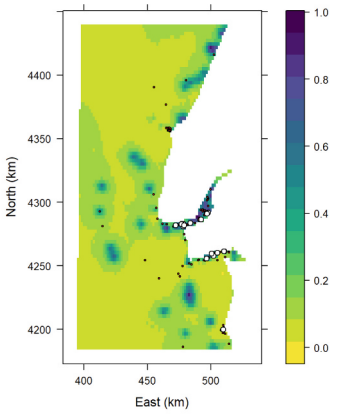
Figure 4 shows the three sampling designs (SRS, MaxRSD and MaxVRSD) for 200 points, where the differences between them become less evident. SRS maintains a more random distribution, MaxRSD continues to focus on high-risk areas, and MaxVRSD provides a balance between risk and uncertainty.

4.4 Surveillance Routes

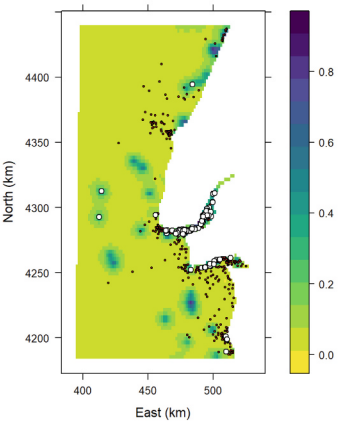
The resulting risk maps highlight areas where infractions are more likely to occur, thereby assisting inspection planning, which combined with the proposed sampling designs, allow to improve the allocation of patrol resources, potentially reducing operational costs, providing support for the Portuguese Navy's maritime surveillance activities related to fisheries inspection. Figure 5 (a) and (b) illustrate two distinct operational strategies based on the MaxRSD sampling design for 200 points, highlighting how route organisation can influence the efficiency of surveillance activities.



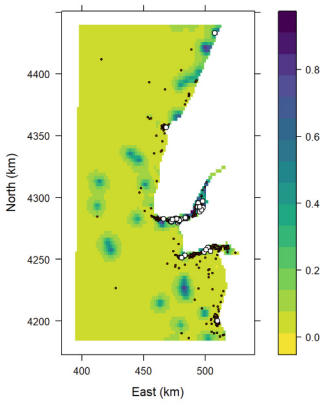
(a) Night: 0:01–6:00



(b) Night: 18:01–24:00



(c) Day: 06:01–12:00

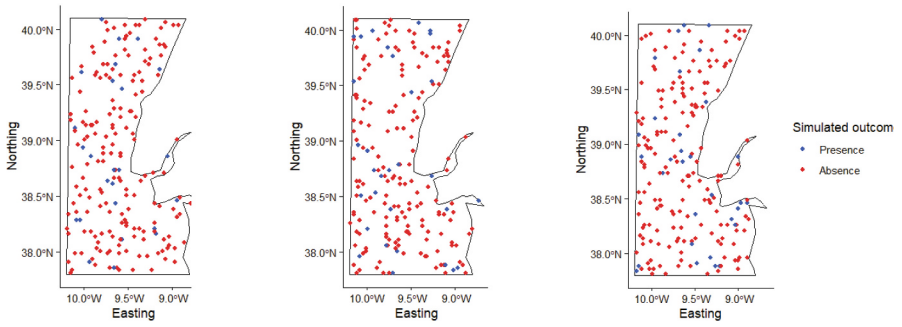


(d) Day: 12:01–18:00

Fig. 3. Map of fitted values for the 1st period (top-left) and 4th period (top-right), and below the 2nd period (bottom-left) and 3rd period (bottom-right) for the year 2014.

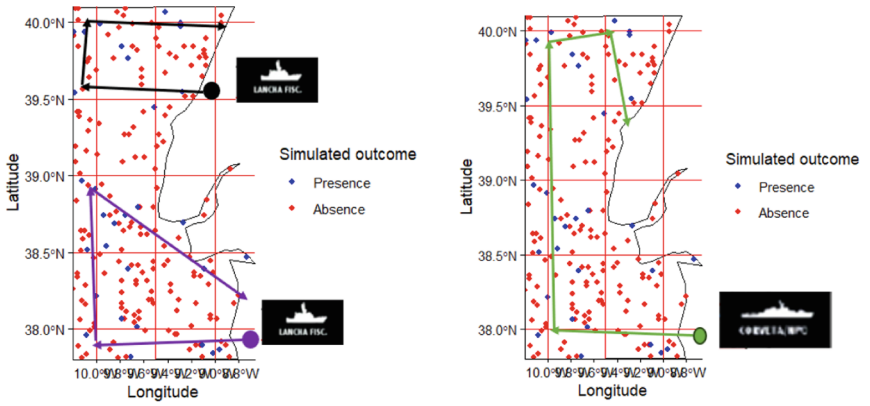
In Fig. 5 (a), two independent routes are proposed, starting from Setúbal and Leiria. This configuration enables simultaneous operations in different areas, particularly in zones with a higher concentration of occurrences, thereby improving response time and optimising travel efficiency.

In contrast, Fig. 5 (b) presents a single continuous route extending from Setúbal to Figueira da Foz. Although this approach may involve a longer trajectory, it provides a more uniform spatial coverage and allows for a more integrated monitoring of the study area along the coastline.



(a) SRS (b) MaxRSD (c) MaxVRSD

Fig. 4. Comparison of sampling designs (200 simulated points, Year 2014).



(a) MaxRSD sampling maps (200 points) with two proposed enforcement routes.

(b) MaxRSD sampling maps (200 points) with one proposed enforcement routes.

Fig. 5. Proposed surveillance routes

4.5 Performance Design Evaluation

Table 3 presents the comparison of the SRS, MaxRSD, and MaxVRSD sampling designs using different predictive performance criteria. The results show that the effectiveness of each sampling strategy strongly depends on sample size and on the specific objective of the analysis.

Under sparse sampling conditions (50 points), substantial discrepancies were observed among the models, particularly in the LGOCV values, which indicate reduced predictive stability and increased uncertainty when limited spatial information is available. In this scenario, SRS provides greater overall stability but shows limited ability to identify high-risk areas and capture complex spatial patterns. In contrast, the adaptive designs, MaxRSD and MaxVRSD, more effec-

Table 3. Comparison of sampling design criteria for different sample sizes

Criteria	50 points				200 points			
	RMSE	LogScore	LGOCV	DIC	RMSE	LogScore	LGOCV	DIC
MaxVRSD	0.6051	−6.4451	616.4678	326.7230	0.0429	−0.1572	0.4458	174.4539
MaxRSD	0.2136	−0.3118	705.8629	258.8679	0.3163	−0.5850	0.3749	145.2507
SRS	0.3341	−1.2825	708.0625	91.8115	0.1332	−0.2879	0.5162	156.3470

tively direct sampling towards relevant locations, improving the representation of the spatial structure of the phenomenon.

Among the adaptive approaches, MaxRSD achieved the best balance between predictive accuracy and probabilistic calibration, presenting lower RMSE and better LogScore values. This suggests that the design is more effective for identifying and predicting high-risk areas under constrained sampling conditions. Conversely, MaxVRSD demonstrated a stronger capacity for uncertainty exploration, although at the expense of reduced local prediction precision, as reflected by its higher RMSE and lower LogScore values.

As the sample size increased to 200 points, all models exhibited substantial improvements in predictive stability, with LGOCV values decreasing considerably. In this case, the differences between the sampling strategies became less pronounced, although the model-based designs continued to show superior overall performance. MaxVRSD achieved the highest predictive accuracy, while MaxRSD maintained a strong ability to identify critical areas. Although SRS also improved with denser spatial coverage, it remained less efficient in highly heterogeneous spatial contexts.

Overall, the results suggest that adaptive sampling designs based on risk and predictive uncertainty provide a more efficient framework for optimising surveillance resources and improving enforcement strategies, particularly in operational scenarios characterised by limited sampling capacity.

4.6 Conclusions and Recommendations

Optimising enforcement efforts remains a significant challenge for maritime authorities, and sampling design plays a key role in improving resource management and reducing operational costs. The results of this study show that the choice of sampling design has a substantial impact on both the quality of spatial inference and the effectiveness of field operations.

The comparison between SRS, MaxRSD and MaxVRSD demonstrates the advantages of model-based approaches. For smaller sample sizes, SRS provides balanced spatial coverage but shows limitations in identifying critical areas. In contrast, MaxRSD directs sampling towards regions with higher probabilities of infractions, enhancing detection capability, while MaxVRSD incorporates both risk and uncertainty, leading to improved estimation quality.

As the sample size increases, the differences between the sampling designs become less pronounced. Nevertheless, approaches that incorporate predicted probability and variability continue to offer advantages, particularly in resource-constrained settings. Furthermore, the integration of sampling design with operational route planning proves to be crucial, as the spatial organisation of enforcement activities can significantly affect overall efficiency.

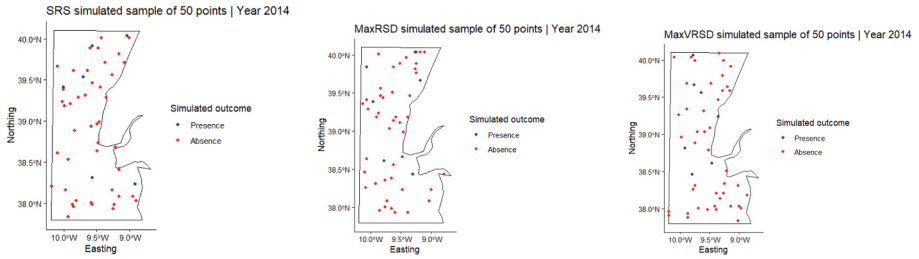
Overall, the findings confirm that model-based sampling designs are effective tools for supporting fisheries enforcement. MaxRSD stands out for its ability to prioritise high-risk areas, while MaxVRSD is particularly useful for improving estimation by accounting for uncertainty. Although SRS is simple and widely used, it shows limitations in contexts with high spatial heterogeneity. In this sense, combining risk maps with adaptive sampling strategies enables more efficient use of resources and supports better-informed operational decisions.

Based on these findings, the following recommendations are proposed: adopt MaxRSD when the primary objective is to maximise detection of infractions; use MaxVRSD to improve estimation and reduce uncertainty; integrate sampling design with operational route planning; develop patrol routes tailored to specific types of infractions; and explore simulation-based approaches and decision-support tools to optimise resource allocation and enforcement strategies. With regard to integrate sampling design with operational route planning, it is expected that in the near future, along with the development of appropriate software will be able to simulate patrol routes allowing the selection of the target zone for simulate a feasible patrol route depending on the infraction, or set of infractions, takingo into accout available the operational resources, to achive the highest probability of success for missions carried out by the Portuguese maritime surveillance.

Acknowledgments. This work is funded by national funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of the projects UID/00297/2025 (https://doi.org/10.54499/UID/00297/2025) and UID/PRR/00297/2025 (https://doi.org/10.54499/UID/PRR/00297/2025) (Center for Mathematics and Applications). We thank the NOVA Math and the Portuguese Navy for making this work possible, within the scope of the protocol established between them.

Disclosure of Interest. The authors have no competing interests to declare that are relevant to the content of this article.

A Appendix



(a) SRS (b) MaxRSD (c) MaxVRSD

Fig. 6. Comparison of sampling designs (50 simulated points, Year 2014).

References

1. Portuguesa, M.: Navios Patrulha Oceânicos – Classe Viana do Castelo (2023). https://www.marinha.pt
2. Autoridade Marítima Nacional. Missão e Competências da Polícia Marítima (2023). https://www.amn.pt
3. Guarda Nacional Republicana. Unidade de Controlo Costeiro – Missão (2023). https://www.gnr.pt
4. Bueger, C., Edmunds, T.: Beyond seablindness: a new agenda for maritime security studies. Int. Aff. **93**(6), 1293–1311 (2017)
5. European Commission. EU Maritime Security Strategy Action Plan. European Commission (2021)
6. Jr, H., David, W., Lemeshow, S., Sturdivant, R.X.: Applied Logistic Regression, 398. John Wiley & Sons, New York (2013)
7. McCulloch, C.E., Searle, S.R.: Generalized, Linear and Mixed Models. John Wiley & Sons, New York (2004)
8. De Oliveira, V.: Models for Geostatistical Binary Data: Properties and Connections. The American Statistician, vol. 74, no. 1, pp. 72–79. Taylor & Francis (2020)
9. Rue, H., Martino, S., Chopin, N.: Approximate Bayesian inference for latent Gaussian models by using integrated nested laplace approximations. J. Royal Stat. Soc. Seri. b (Stat. Methodol.) **71**(2), 319–392 (2009). Wiley Online Library
10. Miguel, B., Simões, P., de Deus, R. G., and Natário, I.: Bayesian geostatistics modeling of maritime surveillance data. In: Proceedings of the International Conference on Computational Science and Its Applications, pp. 187–205. Springer (2024)
11. Petitgas, P.: Geostatistics in fisheries survey design and stock assessment: models, variances and applications. Fish Fish. **2**(3), 231–249 (2001)
12. Brus, D.J., De Gruijter, J.J.: Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). Geoderma **80**(1–2), 1–44 (1997)

13. De Gruijter, J.J., Ter Braak, C.J.F.: Model-free estimation from spatial samples: a reappraisal of classical sampling theory. Math. Geol. **22**(4), 407–415 (1990)
14. Liu, Y., Chen, Y., Cheng, J.: A comparative study of optimization methods and conventional methods for sampling design in fishery-independent surveys. ICES J. Mar. Sci. **66**(9), 1873–1882 (2009)
15. Diggle, P., Lophaven, S.: Bayesian geostatistical design. Scand. J. Stat. **33**(1), 53–64 (2006)
16. Van Groenigen, J.W., Siderius, W., Stein, A.: Constrained optimisation of soil sampling for minimisation of the kriging variance. Geoderma **87**(3–4), 239–259 (1999)
17. Russo, D.: Design of an optimal sampling network for estimating the variogram. Soil Sci. Soc. Am. J. **48**(4), 708–716 (1984)
18. Simbahan, G.C., Dobermann, A.: Sampling optimization based on secondary information and its utilization in soil carbon mapping. Geoderma **133**(3–4), 345–362 (2006)
19. Brus, D.J.: Spatial Sampling with R. CRC Press, Boca Raton (2022)
20. Peter, J.D., Jonathan, A.T., Rana, M.A.: Model-based geostatistics. J. Royal Stat. Soc. Seri. C (Appl. Stat.) **47**(3), 299–350 (1998). Wiley Online Library
21. Finn, L., Håvard, R., Johan, L.: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. Royal Stat. Soc. Seri. B (Stat. Methodol.) **73**(4), 423–498 (2011). Wiley Online Library
22. Elias, K., et al.: Advanced Spatial Modeling With Stochastic Partial Differential Equations Using R and INLA. Chapman and Hall/CRC (2018)
23. Chipeta, M.G., Rowlingson, B., Diggle, P.J.: Geosample: an R Package for Geostatistical Sampling Designs (2019). Under review
24. Pearse, A.R., et al.: SSNdesign: an r package for pseudo-Bayesian optimal and adaptive sampling designs on stream networks. PLoS ONE **15**(9), e0238422 (2020)
25. Simões, P., Carvalho, M.L., Figueiredo, I., Monteiro, A., Natário, I.: Geostatistical sampling designs under preferential sampling for black scabbardfish. In: Proceedings of the International Conference of the Portuguese Statistical Society, pp. 137–151. Springer (2021)
26. Silva, D., Menezes, R., Serra-Pereira, B., Azevedo, M., Figueiredo, I.: 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 (2023)
27. Miguel, B., Simões, P., de Deus, R.G., Natário, I.: Sampling design for binary geostatistical data, application to inspection actions of fishing activity in Portugal. Spatial Stat. 100919 (2025)
28. Cochran, W.G.: Sampling Techniques. John Wiley & Sons (1977)
29. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer, New York (2009)
30. Diggle, P.J., Ribeiro, P.J.: Model-Based Geostatistics. Springer, New York (2007)
31. Gneiting, T., Raftery, A.E.: Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. **102**(477), 359–378 (2007)
32. Vehtari, A., Gelman, A., Gabry, J.: Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat. Comput. **27**, 1413–1432 (2017)
33. Portuguese Navy. Patrol Vessels – NRP ZAIRE, NRP TEJO, NRP DOURO, NRP MONDEGO. Disponível em: https://www.marinha.pt/en/os_meios/patrulhas/Pages/default.aspx Acesso em 2026