

Spatial log-Gaussian Cox process models and sampling paths: towards optimal design

Kenneth Flagg^{a,*}, John Borkowski^a, Andrew Hoegh^a

^a*Department of Mathematical Sciences, Montana State University, Bozeman, MT 59717*

Abstract

Goal of this paper (placeholder abstract—add some results and the coverage statistic). Evaluate a wide variety of path designs in terms design-based heuristics and model-based criteria for spatial prediction using Bayesian LGCP models. Identify promising path designs. Illuminate any relationships among design characteristics and predictive criteria that will be helpful for constrained optimization.

Keywords: log-Gaussian Cox process, optimal sampling, model-based design, spatial sampling design

1. Introduction

Spatial point process models have generally been infeasible because of their computational demands, but recent advances in Bayesian computing have made the Log-Gaussian Cox process (LGCP) an attainable model in practice (Rue et al., 2009; Lindgren et al., 2011; Illian et al., 2012; Simpson et al., 2016). These advances make it possible to fit LGCP models easily, without time-consuming Monte Carlo methods. In some applications, the entire point pattern is not fully observed due to variable sampling effort. This is referred to as a degraded point pattern (Chakraborty et al., 2011). It is relatively simple to accommodate variable sampling effort in these models using modern Bayesian computing

*Corresponding author
Email address: kenneth.flagg@montana.edu (Kenneth Flagg)

tools (Yuan et al., 2017). However, the literature on optimal sampling for spatial point process models is in its infancy (Liu and Vanhatalo, 2020).

Point pattern data are routinely collected in species distribution studies and ordnance response projects. The data consist of the locations of events in some
15 spatial region. In many cases it is too costly to observe the entire study, region so a sampling procedure is employed. These applications may use quadrat sampling or line-transect sampling, with transect sampling being more common. Software is now available to fit spatial point process models to data acquired via distance sampling and simultaneously estimate the detection function (Johnson et al.,
20 2014; R Core Team, 2020). Various spatial mapping procedures have been used to map where events occur in space. However, these have traditionally involved aggregating the data to grid cell counts or computing moving averages. Aggregation has the downside of introducing arbitrary structure into the data by the choice of grid scheme or averaging window, and requires unnecessary
25 computation effort (Simpson et al., 2016).

In ecological settings, sampling plans are often designed around the goal of estimating total abundance. Ordnance response surveys are typically designed to provide enough data to detect (but not necessarily map) intensity hotspots (USACE, 2015; Flagg et al., 2020). However, to our knowledge, there
30 has been very little work done in deciding *where* to collect data when the goal is to map the intensity using a spatial point process model. While some ideas about the characteristics of a good point design apply to paths, creating an optimal path design is not as simple as connecting the points of a point design with line segments. There are many ways to connect points into a path, so optimal
35 design criteria must apply to the whole path and not only to the waypoints. In this paper, we develop a space-filling criterion for path designs, present a variety of sampling path designs, and assess their optimality for mapping intensity using LGCP models via a simulation study.

1.1. Log-Gaussian Cox process

40 The log-Gaussian Cox process is an inhomogeneous Poisson process where
the logarithm of the intensity function is a Gaussian process (Møller et al., 1998).
The LGCP provides a flexible model for mapping event intensity over space
using few parameters. Efficient Bayesian computation tools are available using
INLA to approximate the posterior marginal distributions (Rue et al., 2009),
45 a finite element approach to represent the Gaussian process (Lindgren et al.,
2011), and pseudodata to approximate the point process likelihood (Simpson
et al., 2016).

1.2. Spatial design

Most classical sampling and design work has been done for points, or for
50 small quadrats approximated as points, rather than for paths. In two-dimensional
(geostatistical) model-based design, regularity is optimal for spatial prediction
but randomness and a variety of interpoint distances are best for parameter
estimation (Diggle and Lophaven, 2006). Inhibitory plus close pairs designs are
55 a good compromise (Chipeta et al., 2017). Design-based approaches exist to
spread points through high-dimensional design spaces (Borkowski and Piepel,
2009), and Latin hypercube sampling has space-filling characteristics (McKay
et al., 1979; Husslage et al., 2011). Minimax and maximin distance are popular
60 criteria for ensuring deign points represent the entirety of a large (or high-
dimensional) design space (Johnson et al., 1990). These criteria are calculated
from the distances between design points and arbitrary points in the design
space, explicitely ensuring coverage by minimizing the maximum distance be-
tween the design space and the design (minimax) or maximizing the minimum
destance between pairs of design points (maximin).

1.3. Space-filling curves

65 Another relevant area of research is in deterministic space-filling curves.
These have used in the design of dense or stretchable circuits (Ogorzałek, 2009;

Ma and Zhang, 2016) and high-dimensional data visualization in bioinformatics (Anders, 2009). The Hilbert curve is simple to construct and the Peano curve is very flexible for filling irregular shapes (Fan et al., 2014). Space-filling
70 curves are one-dimensional paths constructed iteratively; as the number of iterations goes to infinity, the limiting path has nonzero area and actually fills the space (Sagan, 1994). For applications we stop after a finite number of iterations.

1.4. Paths as sampling designs

The small body of literature on spatial sampling design for point pattern data
75 has focused on line transects. Pollard et al. (2002) began with line transects and adaptively added zigzags in a species abundance survey.

The Visual Sample Plan software includes features to create systematic transect plans and augment plans with additional transects in regions lacking spatial coverage (Matzke et al., 2014). It helps the user choose the transect spacing
80 to maximize the probability of detecting the presence of a hotspot of specified size and intensity. However, it does not employ criteria to optimize spatial prediction.

Liu and Vanhatalo (2020) provided one of the first explicit discussions of design in the context of spatial LGCP models. They used narrow quadrats
85 (swaths along line-transects) as their sampling units. The transects were short relative to the size of the study region and not connected into a path.

2. Materials and methods

With an eye toward practical considerations of data collection, we present criteria to compare sampling strategies that impact LGCP posterior. We compare plans with (approximately) fixed path lengths, most of which avoid sharp turns.
90 Data collection equipment (e.g. metal detectors) may have limited mobility, requiring minimizing the number or angle of turns. This section describes the designs in detail, then introduces notation and design criteria, and finally summarizes the computational model-fitting approach.

95 2.1. Sampling design schemes

In this section, we present three variations of parallel line transect designs and three schemes that produce more complex designs. To clarify terminology, a *path* or *design* is a realized set of one or more connected components that has length but not area. The paths considered in this work are constructed as 100 sequences of line segments. A *design scheme*, or simply *scheme*, is procedure for generating designs with some shared characteristics. Figure 1 illustrates a selection of designs from these schemes.

2.1.1. Parallel line transects

Parallel straight-line transects are common in ordnance response studies and 105 in ecological studies using distance sampling. Systematic designs are common because they provide good spatial coverage in the sense that any point in the study site has an a priori known maximum distance from the path. For point designs, systematic designs are optimal for prediction, simple random samples are optimal for estimation, and inhibitory with close pairs designs are becoming 110 a popular compromise (Chipeta et al., 2017). We adapt all of these to the parallel line transect setting. We use line transects running north-south, with three methods of choosing the horizontal coordinate: simple random sample (SRS), systematic with a random starting point and even spacing, and inhibitory plus close pairs. Figure 1 (left column) shows an example of each scheme with 115 25 transects.

2.1.2. Parallel serpentine transects

One simple way to observe a greater variety of locations and different directions is to add lateral “zigzags” to transects. We include alternate right and left turns at right angles to create serpentine transects. This could decrease prediction variance because more of the path will be close to each point in the study 120 area than would be under a line transect design with similar total distance. They will also improve estimation of the covariance function in the presence of anisotropy. Figure 1, top right, shows two examples.

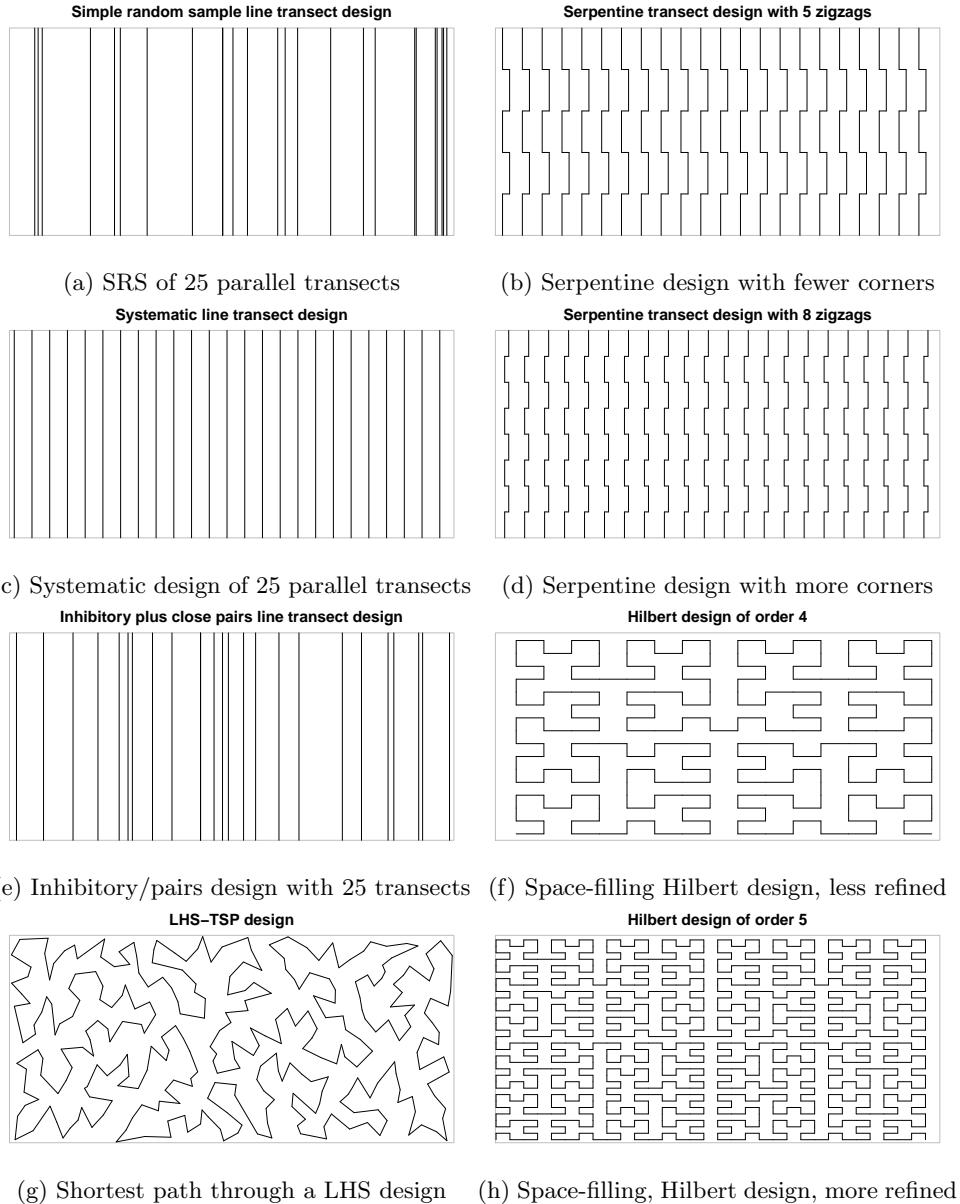


Figure 1: Examples of plans from six design schemes. Except for the Hilbert curve of order 5, all of these plans have approximately the same total length.

2.1.3. Latin hypercube sampling

125 Random Latin hypercube sampling (LHS) produces a design that spreads discrete points through a (potentially high-dimensional) design space, ensuring that the full range of each dimension is included while remaining balanced and keeping the number of points small (McKay et al., 1979). This is done by partitioning each dimension into a specified number k of intervals (thus stratifying a d -dimensional design space into k^d cells), selecting a Latin hypercube design to determine which k cells will contain a design point, and then drawing each design point from a uniform distribution over its cell. In two dimensions, this scheme produces point designs with good spatial coverage properties. We use the LHS design as waypoints for a path. Because longer distance typically 130 brings increased costs, we treat this as a traveling salesperson problem (TSP) and use the shortest path through the waypoints as our design. This LHS-TSP scheme produces paths that have many sharp corners but leaves few large voids (example in Figure 1, bottom left). Waypoints are generated by the `lhs` R package (Carnell, 2020) and connected into a the shortest path by the TSP 135 package (Hahsler and Hornik, 2020). A downside of this design scheme is that the length cannot be specified directly, and only certain distances are possible depending on the number of bins used.

2.1.4. Space-filling curves

As a representative of space-filling curves, we use the Hilbert curve scaled 145 to fit the study site. These designs have many short segments meeting at right angles. The only parameter of this design scheme is the order, or number of iterations used in refining the curve. Each iteration increases the length and complexity of the design. This is produces a deterministic design, so a random offset is added to vary which points are observed. The Hilbert curve is generated 150 by `HilbertVis` R package (Anders, 2009).

2.2. LGCP model and notation

The full point process is denoted \mathbf{X} and operates over the entire study site \mathcal{R} . The full realized point pattern is denoted \mathbf{x} and is a countable subset of \mathbf{R} . However, only a subregion of \mathcal{R} is surveyed. This subregion is \mathcal{S} , the set of all points within a specified radius of the survey path. The observed point pattern is defined as $\mathbf{x}_{\mathcal{S}} = \mathbf{x} \cap \mathcal{S}$.

In the LGCP model, \mathbf{X} is a Poisson process with intensity function $\lambda(u)$, $u \in \mathcal{R}$. This intensity function is itself a stochastic process, modeled on the natural logarithm scale as

$$\log[\lambda(u)] = \mu + \mathbf{e}(u), \quad (1)$$

with an intercept μ and spatial error term $\mathbf{e}(u)$. The error term is a Gaussian process with mean 0 and covariance function $C(u, v)$. In our simulation study, C is a stationary and isotropic Matérn covariance function with smoothness parameter ν , spatial scale ρ , and variance σ^2 .

2.2.1. Evaluation criteria

Another important characteristic is the concept of spatial coverage, or how well the design represents the study site in the sense that all points in the region are close to observed locations. We measure this by the average distance between points in the site and points on the path. We define the distance between a point and the path as the distance between that point and the nearest point on the path, then take the average over all points in the study region. This can be roughly understood as the radius of a typical unsurveyed void.

For model-based comparisons, we compare the posterior distribution of the intercept μ to the simulated true mean of the GP, and evaluate two spatial prediction criteria. The spatial prediction criteria that we evaluate are the mean squared prediction error (MSPE) of the log intensity surface,

$$E[(\mu + \mathbf{e}(u) - \log[\lambda(u)])^2 | \mathbf{x}_{\mathcal{S}}], \quad (2)$$

and the average prediction variance (APV) of the latent Gaussian process,

$$\text{Var}[\mathbf{e}(u) | \mathbf{x}_{\mathcal{S}}]. \quad (3)$$

In the above notation, λ is the true realized intensity function, which is known
170 in the simulations.

The first criterion captures accuracy in the overall scale of the intensity
surface. Large-magnitude errors in the mean imply that that model will predict
overall too many or two few events. The other criteria reflect two important
175 characteristics of a useful spatial prediction. First, deviation from the true
surface should be low, and is indicated by low MSPE. Second, the model's own
assessment of prediction uncertainty will be used in practice to build trust in
the inferences, so low APV is desireble.

2.2.2. Model fitting

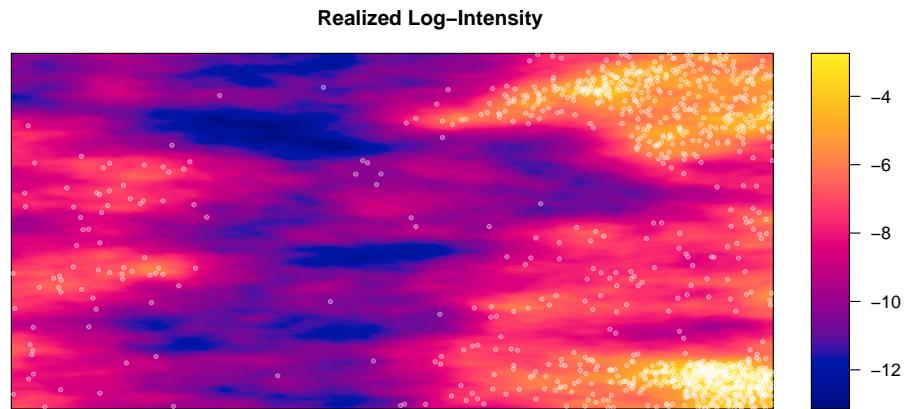
We fit the spatial LGCP model using nested integrated Laplace approximations
180 and the R-INLA package (Rue et al., 2009; Blangiardo and Cameletti, 2015). The Gaussian process is approximated using a finite element approach (Lindgren et al., 2011). The point pattern is modeled by pseudodata placed at the events and the finite element nodes (Simpson et al., 2016). This procedure allows fast and accurate approximation of the posterior distribution.

185 3. Simulation study

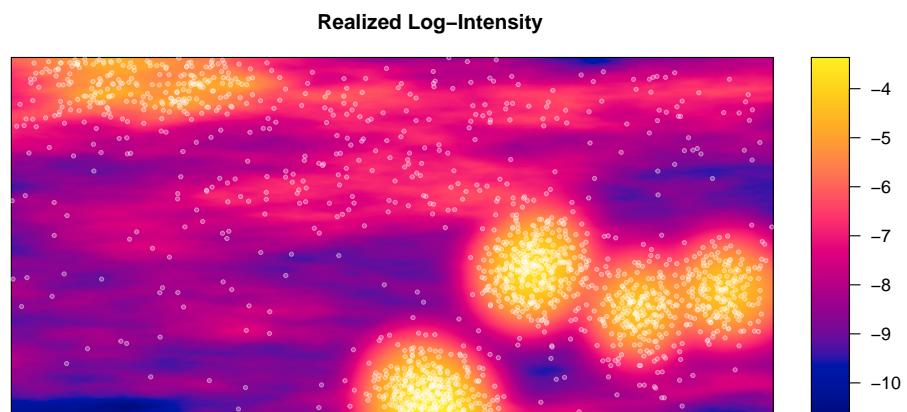
We simulate 100 designs from each of six schemes. All events within a 2 unit radius of the path are observed. The whole experiment is repeated for 5 realizations from each of two data generating models.

3.1. Study site

190 We consider a fictitious site \mathcal{R} with the simple shape of a 1500 unit by 700 unit rectangle. In this site, we simulate two data generating models that produce random intensity functions with hotspots. First, a LGCP with latent GP mean $\mu = \log(250/|\mathcal{R}|) = -8.34$ and Matérn covariance function with $\nu = 1$, $\sigma = 2$, and range = 200. This model produces relatively unstructured hotspots due to
195 large variability in the GP. Figure 2a shows an example realization from this process.



(a) Realized LGCP log-intensity



(b) Realized LGCP with Clusters log-intensity

Figure 2: The realized intensity function (natural log scale) and complete point pattern from a LGCP, and from a LGCP superposed with a cluster process.

Second, a two-stage cluster process and a LGCP are superposed. The cluster process (a Neyman-Scott or, more specifically, a Thomas process) is constructed as follows. The number of clusters is Poisson-distributed with mean 3. The 200 number of events per cluster is Poisson-distributed with mean 200. The cluster centers are distributed uniformly over \mathcal{R} . Events come from a bivariate normal distribution with mean equal to the cluster center and variance $\Sigma = \tau^2 \mathbf{I}$, $\tau = 50$. The LGCP component has $\mu = \log(250/|\mathcal{R}|) = -8.34$ and Matérn covariance with $\nu = 1$, $\sigma = 1$, and range = 200. This model is based upon the typical conceptual model of a firing range, with a background process (represented by the LGCP) and a small number of higher-intensity foreground clusters containing the events of interest (example in Figure 2b). As a shorthand, we refer to this generative process as LGCP with Clusters.

The LGCP generating process matches the model fit to the realized data. 210 The model should both produce accurate predictions and have low bias in the posterior means of the parameters. On the other hand, the parameters of the LGCP with Clusters process will not correspond to the LGCP model parameters, but we anticipate that in general the latent GP in the LGCP model will still accurately predict the log-intensity function.

215 3.2. Path design schemes

The simulation uses each of the design schemes discussed in Section 2.1. The surveyed region \mathcal{S} consists of all points within 2 units of the design. That is, \mathcal{S} is a collection of strips along each segment of the design. All events within 220 2 units of the design are observed and all events farther than 2 units from the path are unobservable.

The design parameters for each scheme are varied to create four different levels of survey effort: low, medium, high, and very high. Within each effort level, the total length or distance traveled is comparable across all schemes. The inhibitory plus close pairs scheme and the serpentine scheme are further 225 varied to employ different numbers of pairs and different numbers of zigzags, respectively. Table 1 provides an overview of the different settings.

Table 1: Design schemes used in the simulation study. The table presents the design parameters, distance travelled along the path, percent of the study area surveyed, and average distance to the path under each of four survey effort levels. The average distance to the path varies due from design to design, and is presented as average (SD).

* Variable area due to overlapping segments, reported as average (SD).

** Variable distance surveyed, reported as average (SD).

Scheme	Low Effort	Medium Effort	High Effort	Very High Effort
SRS	10 transects Length: % Surveyed: Avg. Dist. to Path:	25 transects 7000 units 6.46% (0.202%)* 29.7 (6.01)	50 transects 35000 units 12.5% (0.331%)* 15.0 (1.99)	70 transects 49000 units 17.2% (0.459%)* 10.8 (1.44)
Systematic	10 transects Length: % Surveyed: Avg. Dist. to Path:	25 transects 7000 units 2.68% 6.70% 38.4 (0.866)	50 transects 35000 units 13.4% 7.56 (0.0307)	70 transects 49000 units 18.7% 5.40 (0.0164)
Inhibitory, 10% Close Pairs	9 transects 1 paired Length: % Surveyed: Avg. Dist. to Path:	23 transects 2 paired 7000 units 6.53% (0.172%)* 20.8 (1.82)	45 transects 5 paired 35000 units 12.8% (0.314%)* 10.6 (1.24)	63 transects 7 paired 49000 units 17.5% (0.427%)* 7.64 (0.576)
Inhibitory, 20% Close Pairs	8 transects 2 paired Length: % Surveyed: Avg. Dist. to Path:	20 transects 5 paired 7000 units 6.56% (0.157%)* 22.0 (2.97)	40 transects 10 paired 35000 units 12.8% (0.303%)* 11.4 (1.70)	56 transects 14 paired 49000 units 17.6% (0.368%)* 8.26 (0.826)
Serpentine, 5 Zigzags	7 transects 75 unit offset Length: % Surveyed: Avg. Dist. to Path:	22 transects 23.9 unit offset 7000 units 6.68% 43.2 (1.83)	47 transects 11.2 unit offset 35000 units 13.4% 7.79 (0.0508)	67 transects 7.84 unit offset 49000 units 18.7% 5.51 (0.0309)
Serpentine, 8 Zigzags	7 transects 42.9 unit offset Length: % Surveyed: Avg. Dist. to Path:	22 transects 13.6 unit offset 7000 units 6.67% 44.5 (2.03)	47 transects 6.38 unit offset 35000 units 13.3% 7.76 (0.0328)	67 transects 4.48 unit offset 49000 units 18.7% 5.50 (0.0271)
LHS-TSP	50 bins Length: % Surveyed: Avg. Dist. to Path:	300 bins 7177 (189) units** 2.73% (0.0722%)** 43.5 (1.44)	1200 bins 17191 (168) units** 6.51% (0.0632%)** 17.4 (0.186)	2400 bins 34196 (225) units** 12.9% (0.0835%)** 8.63 (0.0683) 48433 (342) units** 18.2% (0.127%)** 6.08 (0.0562)
Hilbert Curve	3rd Order Length: % Surveyed: Avg. Dist. to Path:	4th Order 8581 units 3.27% 37.3 (1.95)	5th Order 17442 units 6.634% 17.9 (0.480)	6th Order 35025 units 13.3% 8.77 (0.113) 70121 units 26.47% 4.37 (0.0382)

The parallel transect schemes have 10, 25, 50, or 70 line transects running north-south. We expect the simple random sample scheme to produce high prediction variance and large prediction error in big gaps between transects.

230 The systematic sample scheme uses a uniformly-distributed starting point and constant spacing between adjacent transects. We expect systematic transects to provide low bias and moderate prediction variance. However, this scheme can miss structures at certain sizes because no transects are close to each other in the east-west direction.

235 For the inhibitory plus close pairs line transect scheme, we vary the numbers of paired and unpaired transects. The total number of transects is 10, 25, 50, or 70, with 10% and 20% of the transects (rounded to the nearest integer) as redundant members of a pair. The remaining primary transects are placed according to a one-dimensional Strauss process (Strauss, 1975; Kelly and Ripley, 1976). The Strauss attraction parameter is set at $\gamma = 0.05$ and the radius for counting pairs is 1500 units divided by the total number of transects. Then each redundant transect is randomly paired to a primary transect, and placed within the pair radius of the primary transect according to a uniform distribution. We expect this scheme to have intermediate performance between the simple 240 random sample and the systematic line transect schemes.

The serpentine transect scheme has 7, 22, 47, or 67 transects running north-south with constant east-west spacing and a random starting point for the first transect. The number of zigzags (north-south segments) is 5 or 8 per transect, and the pedicular offset is set so the the total east-west distance equals the 250 length of three north-south line transects. Thus, the serpentine designs traverse the same length as the line transect designs. These designs should result in smaller prediction errors and lower variance farther from the path, compared to line-transect designs.

Our Latin hypercube sampling/traveling salesperson (LHS-TSP) scheme 255 uses 50, 300, 1200, or 2400 bins to generate the waypoints. Preliminary experimentation found that these bin numbers produced total lengths similar to the line-transect schemes. The LHS-TSP scheme is expected to result in small

prediction errors and low prediction variance per unit distance traveled. However, the designs will have many sharp corners and may leave some large voids.

The Hilbert curve scheme uses a random starting point and a Hilbert curve of order 3, 4, 5, or 6. The path length is a deterministic function of the order and differs greatly among curves of different orders. These orders yield lengths similar to the lengths of the transect designs. Hilbert designs should provide low prediction variance, but have lots of short segments.

3.3. Model specification

The same Bayesian LGCP model is fit to each observed dataset. The observed point pattern \mathbf{x} is a realization of \mathbf{X} , a Poisson process on \mathcal{R} with intensity $\lambda(u)$. The intensity is modeled as $\log[\lambda(u)] = \mu + \mathbf{e}(u)$. The spatial error term \mathbf{e} is a Gaussian process with mean $\mathbf{0}$ and a Matérn covariance function with fixed

$$\nu = 1.$$

The intercept μ has a $\text{Unif}(-\infty, \infty)$ prior. The covariance parameters σ and ρ have a PC prior with $\Pr(\sigma > 3) = 0.1$ and $\Pr(\rho < 100) = 0.1$ (Fuglstad et al., 2019; Simpson et al., 2017).

The Gaussian process prediction surface is approximated on the finite element mesh shown in Figure 3. The GP is predicted at the nodes (points) and is linearly interpolated elsewhere. The nodes are weighted according to the area of their dual cells (grey outlines) and used for numerical integration of the likelihood (Lindgren et al., 2011). This mesh is adequately fine to model the large-scale trends in the surface while keeping the computing time well under a minute for each model fit.

4. Results

4.1. Initial Observations

In describing the results, we focus on one LGCP dataset and one LGCP with Clusters dataset (Figure 2). The results are similar for all datasets (see the online supplement.)

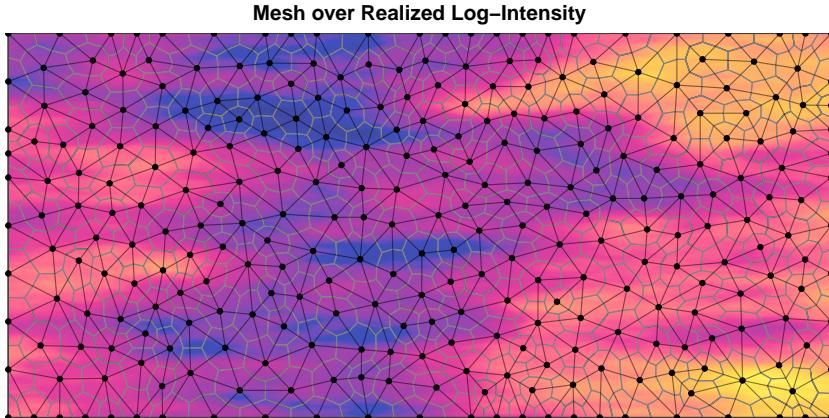
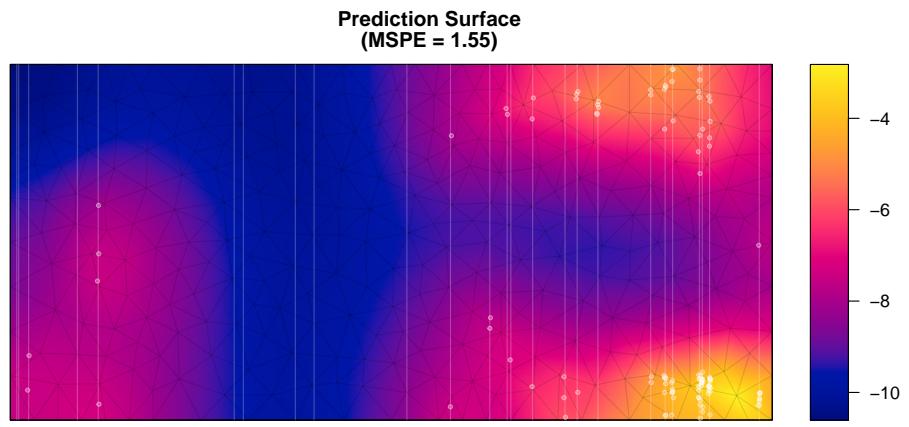


Figure 3: Illustration of the mesh used to approximate the latent GP. The mesh nodes are used in a numerical integration scheme where they are weighted by the area of their dual cells (outlined in grey).

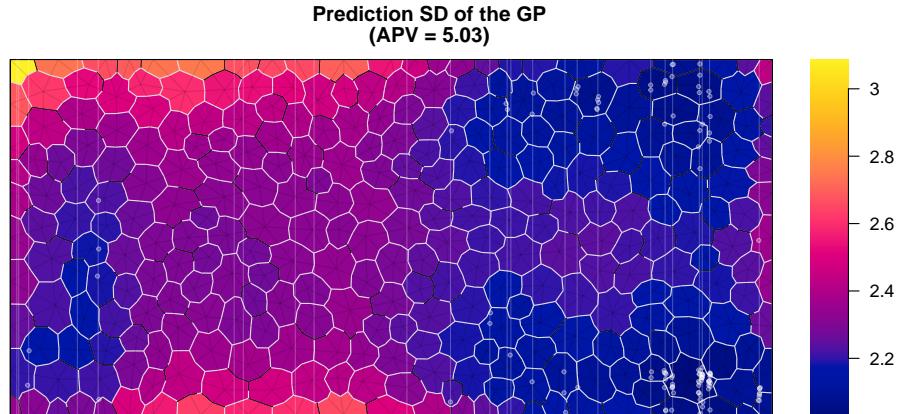
Figure 4 shows an example where the model does well at predicting the intensity of the realized LGCP from data observed along one of the SRS paths. In the figure, the path appears in white and the observed events are shown as white dots. The posterior predicted mean of the log-intensity (top panel) accurately captures the large-scale features, but smooths out much of the small-scale variation. The bottom panel shows the prediction standard deviation for each mesh node. The SD ranges from 2.0 to 3.1, and is lowest near observed events. SD increases farther from observed events, including in places where the surveyed area was observed to contain no events.

This same survey plan also did well for the LGCP with Clusters dataset, with the model accurately capturing the large-scale details of the intensitiy function, including two of the circular hotspots corresponding to clusters (Figure 5a). However, it also smoothed the surface quite a bit, notably merging the two overlapping clusters into a single oblong hotspot.

Most plans yielded similar prediction surfaces, capturing the large-scale trends, and having the least uncertainty near observed events. Results varied in accuracy at the most extreme peaks and valleys of the itensity function and in overall SD across the study region.

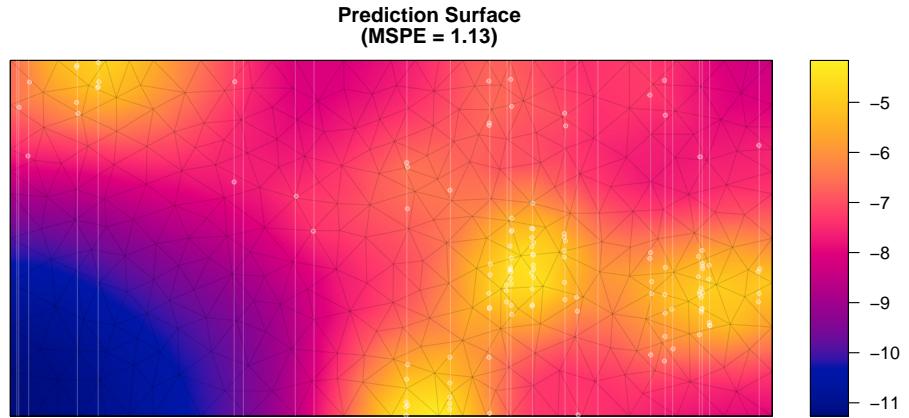


(a) Predicted log-intensity

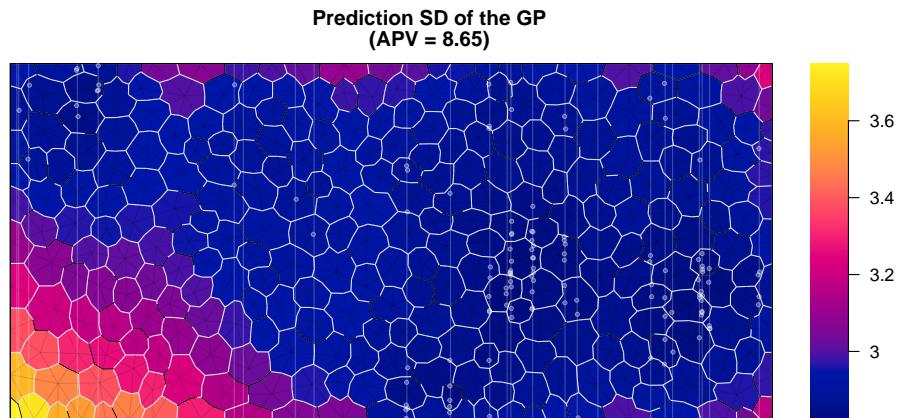


(b) Prediction SD

Figure 4: Predicted log-intensity function and prediction standard deviation using data observed via a SRS of line transects. The SD is shown for each finite element node. This example is a medium-effort plan applied to a LGCP dataset.



(a) Predicted log-intensity



(b) Prediction SD

Figure 5: Predicted log-intensity function and prediction standard deviation using data observed via a SRS of line transects. The SD is shown for each finite element node. This example is a medium-effort plan applied to a LGCP with Clusters dataset.

However, a small number of model fits suffered from apparent edge effects.
305 For example, Figure 6 shows the prediction surface resulting from a serpentine transect plan. The predicted log-intensity has a hotspot of extremely large values in the southeast corner (notice the color scale). The hotspot is driven by two nodes on the boundary with very large prediction values. Another, less extreme, edge effect is present in the northeast corner.

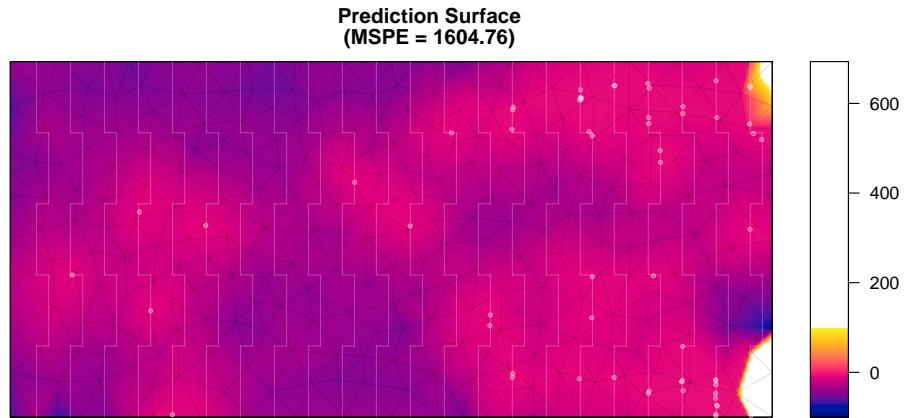
310 The next sections compare the schemes in terms of model-based criteria applied to one realized LGCP dataset and one realized LGCP with Clusters dataset.

4.2. Dealing with computational problems

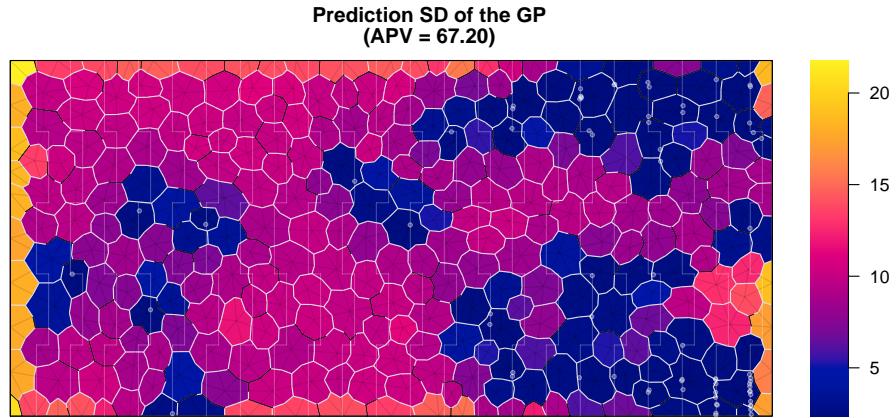
4.3. LGCP simulation results

315 Considered across all survey plans applied to one LGCP dataset, all design schemes tend to result in slightly underestimating the mean, but most 95% posterior intervals still capture the true mean value (Figure 10a). The capture rate increases with survey effort, with only the SRS and Hilbert schemes being observed to miss the true mean at the hight effort level. For each scheme at each
320 effort level, the median error in the posterior mean was between 0 and -1 . At any given effort level, there is little difference among the distribution of errors across schemes. However, at low and medium effort, all of these distributions have long right tails with severe overestimates (accompanied by wide posterior intervals).

325 Both MSPE and APV had right-skewed distributions. Thus we use logarithmic scales for plots and summarize them using the median and interquartile range (IQR). Median MSPE decreases with increasing path distance, leveling off around between the medium and high effort levels (Figure 11a). Variability (IQR) of MSPE also descreases as effort increases. Prediction surfaces with
330 edge effects form a cluster of large, outlying MSPE values. At all levels of effort, systematic designs had the least variability in MSPE. The lowest-MSPE prediction surfaces result from the longest Hilbert designs, but the distributions of MSPE are similar for all schemes at the high and very effort levels. Overall,

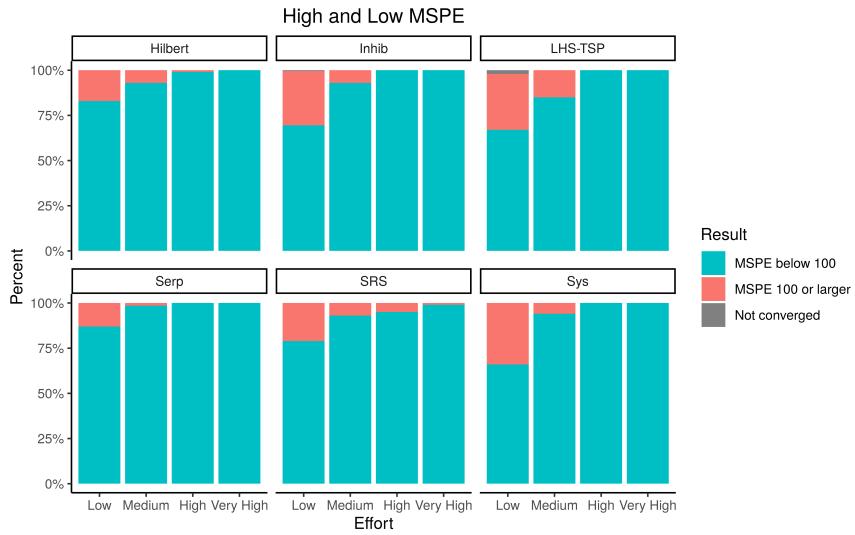


(a) Predicted log-intensity. The color scale is truncated at 100, but reaches a maximum of 692.

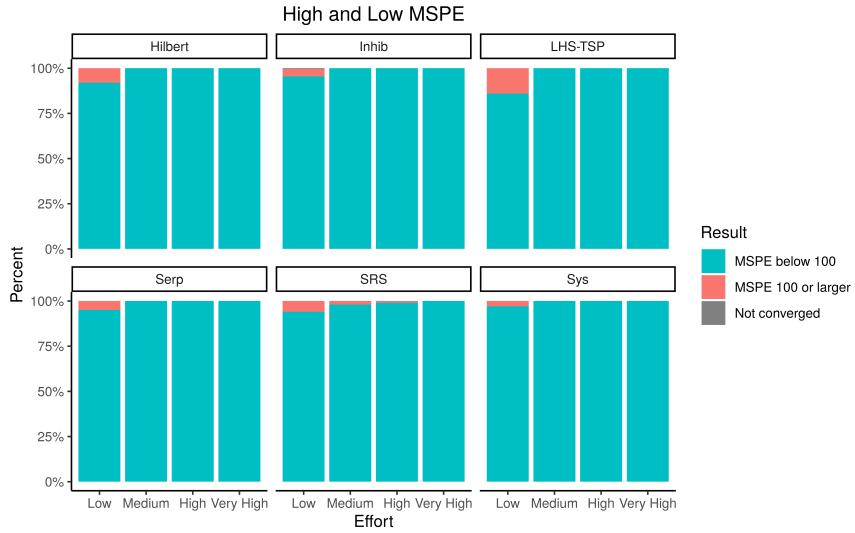


(b) Prediction SD

Figure 6: Predicted GP surface and prediction SD using data observed via a serpentine transect plan. The prediction has an apparent edge effect in the southeastern corner. The SD is high across much of the site. This example is a medium-effort plan applied to a LGCP dataset.



(a) MSPE distribution for one LGCP dataset



(b) MSPE distribution for one LGCP with Clusters dataset

Figure 7: Plots of the distribution of high and low mean squared prediction error (MSPE) vs survey effort for each plan applied to one realization of a LGCP and one realization of a LGCP with Clusters.

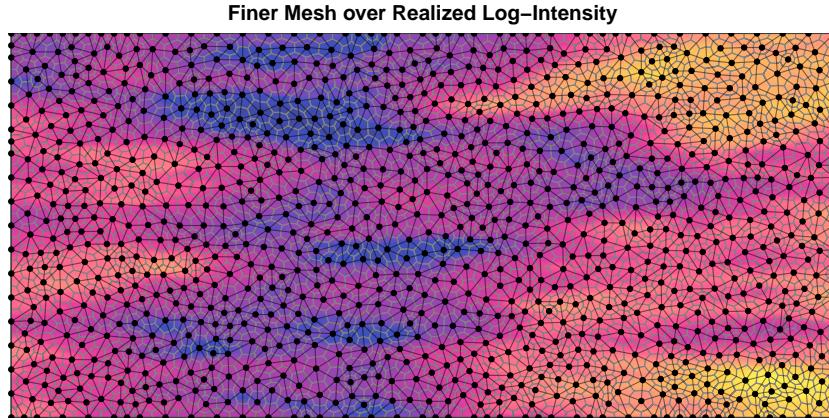


Figure 8: Illustration of a finer mesh than the one used in the full simulation. The mesh nodes are used in a numerical integration scheme where they are weighted by the area of their dual cells (outlined in grey).

the differences among the different schemes with respect to median MSPE are
³³⁵ much less than the variability in MSPE within each effort level. The results are largely the same for APV (Figure 11b).

4.4. LGCP with Clusters simulation results

For the LGCP with Clusters dataset, all schemes tended to do well at estimating the mean (Figure 10b). At low effort, all schemes occasionally fail to
³⁴⁰ capture the true mean in a 95% central posterior interval. Only the SRS scheme is observed to miss the mean at medium or high effort. The median errors for the scheme and effort combinations range between -0.13 and 0.16 . IQR decreases slightly with higher effort. At low effort, all schemes have long-tailed distributions containing some extreme errors; notably the serpentine and SRS
³⁴⁵ schemes have left-tailed distributions while the others are right-tailed.

MSPE and APV again had right-skewed distributions. Median MSPE decreases as effort increases (Figure 12a). Variability in MSPE is roughly constant across effort levels. The systematic scheme had the lowest median MSPE for low and medium effort and low IQR for all effort levels, while the serpentine and Hilbert schemes have the lowest median at high and very high effort. However,
³⁵⁰

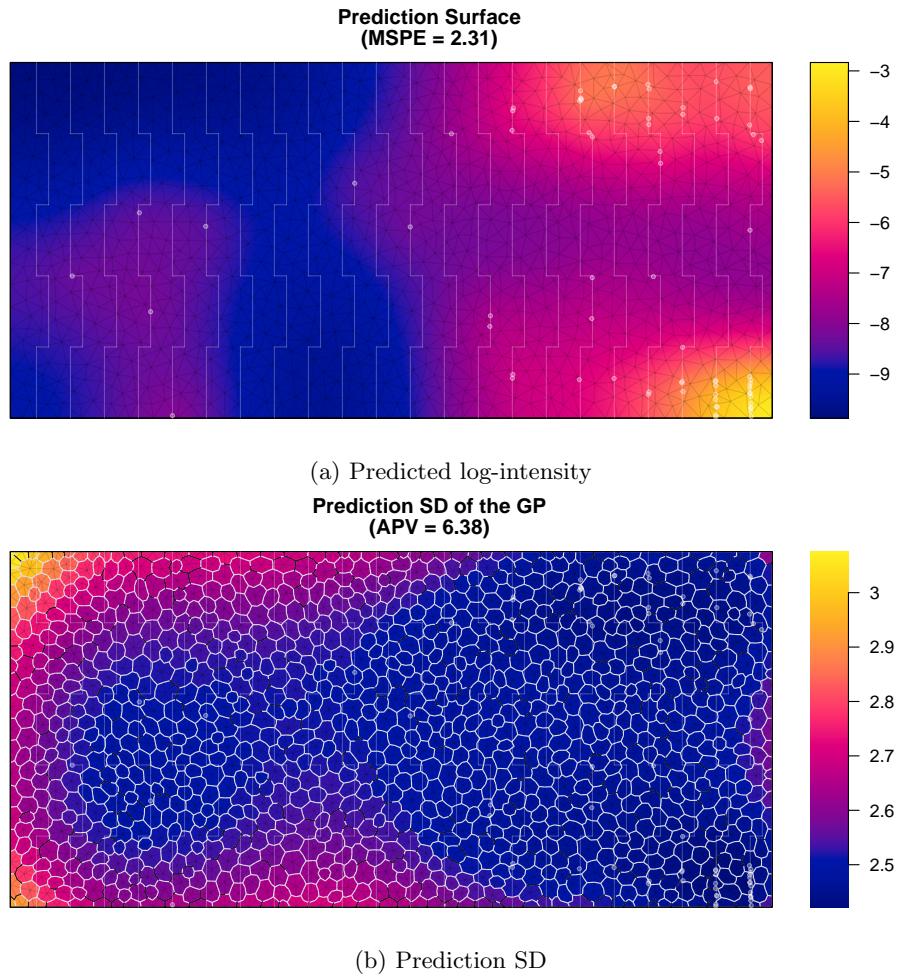
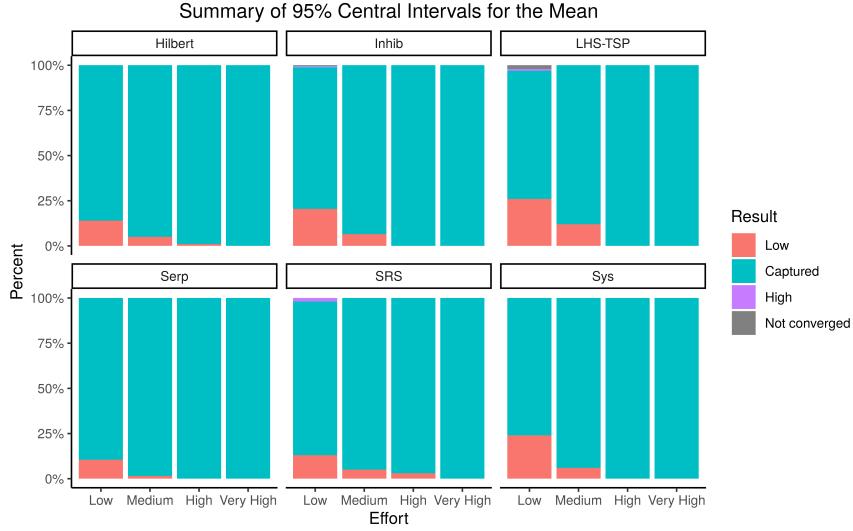
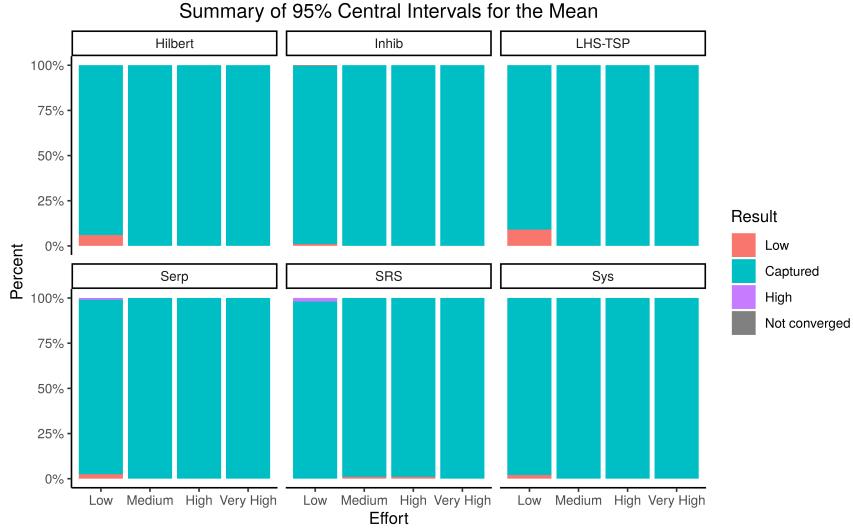


Figure 9: Predicted GP surface and prediction SD using data observed via a serpentine transect plan. The prediction has an apparent edge effect in the southeastern corner. The SD is high across much of the site. This example is a medium-effort plan applied to a LGCP dataset.

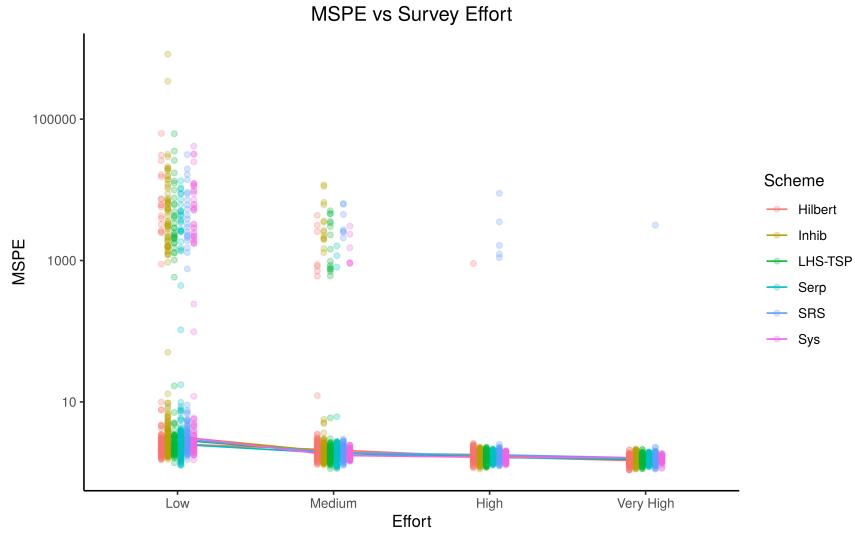


(a) Comparisons of posterior intervals for the intercept to the true mean for one LGCP dataset

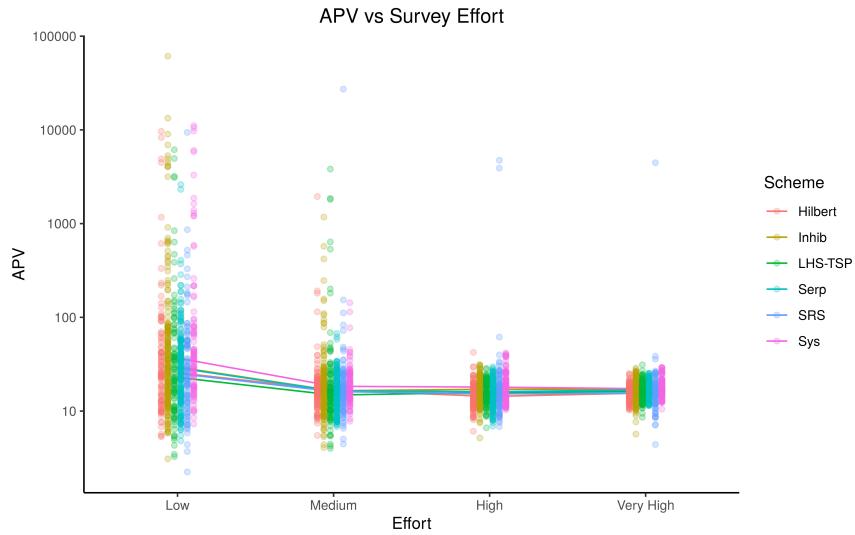


(b) Comparisons of posterior intervals for the intercept to the true mean for one LGCP with Clusters dataset

Figure 10: Plots summarizing comparisons between central 95% posterior intervals for the intercept and the true mean log-intensity of the realized GP. “Captured” indicates that the true mean is inside the interval. “High” and “Low” indicate that the entire interval is, respectively, above or below the true mean.



(a) MSPE vs survey effort for one LGCP dataset



(b) APV vs survey effort for one LGCP dataset

Figure 11: Plots of mean squared prediction error (MSPE) and average prediction variance (APV) vs survey effort for each plan applied to one realization of a LGCP. Line segments connect the median at each effort level.

differences between schemes are much less than differences across survey effort. At low effort, the distribution of APV has a long tail. Otherwise there is little difference in distribution of APV across schemes or effort (Figure 12b).

4.5. Spatial coverage

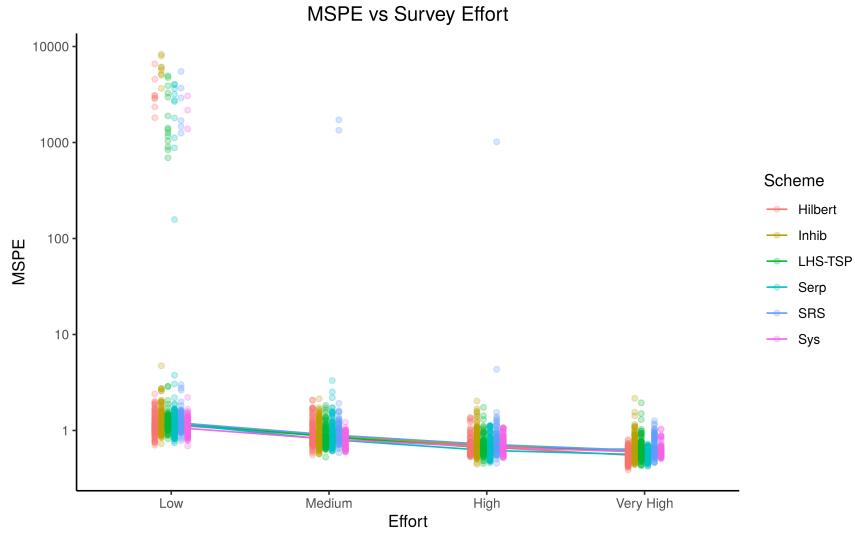
355 While the above results suggest the choice of design is relatively unimportant and the distance traveled is the main driver of the quality of the spatial predictions, it is important to consider that all of the design schemes ensure the path is distributed across the entire study region. Designs that leave large unexplored voids will not perform as well.

360 As an example, Figure 13 shows the results of using a systematic sample of 50 parallel transects in the western 20% of the study site. This design is the same length as our high-effort designs, but it leaves most of the site far from the surveyed path. As a result, the predicted log-intensity is flat near the GP posterior mean of -8.71 over most of the site, rendering the prediction mostly 365 useless.

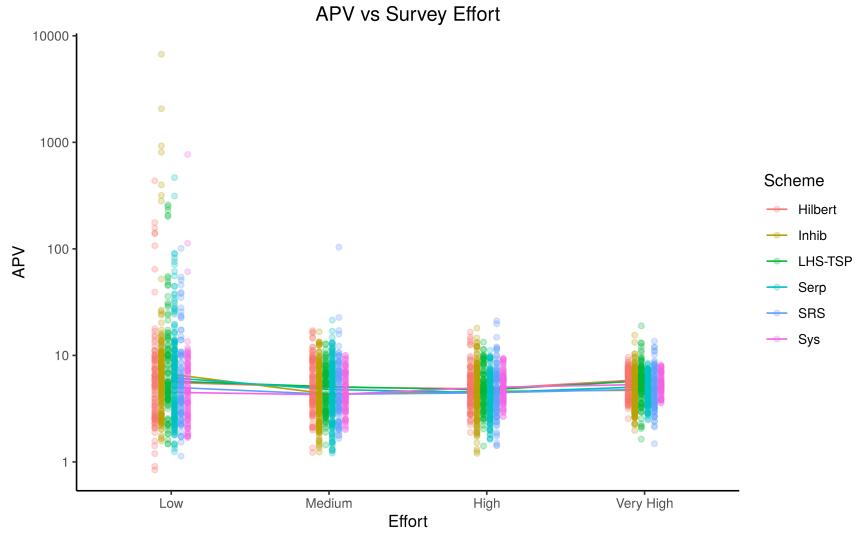
The problem with this example is easily explained in terms of the average distance between the path and arbitrary points in the study site. This example design has an average distance from the path of 486 units; the designs used in the simulation study all had average distances to the path under 130 (with most 370 under 50). In the simulation study, there was little association between MSPE and average distance to the path for paths of comparable length, but the average distance to path decreased as effort (length) increased (Figure 14). This means that all of these schemes do well at distributing the path around the site at all of these effort levels.

375 4.6. Augmenting a poor-performing design

Even a poor-performing design could be used as a starting point for sequential design. As a simple illustration, we augment the design from Figure 15 with some additional sampling effort in the western part of the site, where the original design left a large void. The total distance surveyed increases from 8581

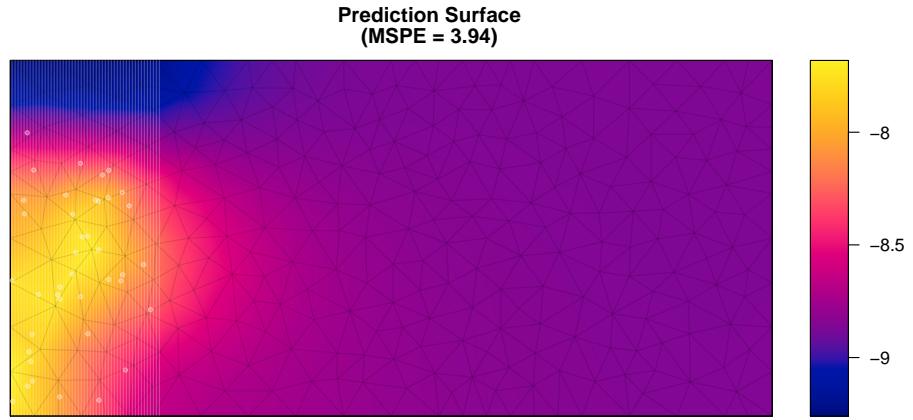


(a) MSPE vs survey effort for one LGCP with Clusters dataset

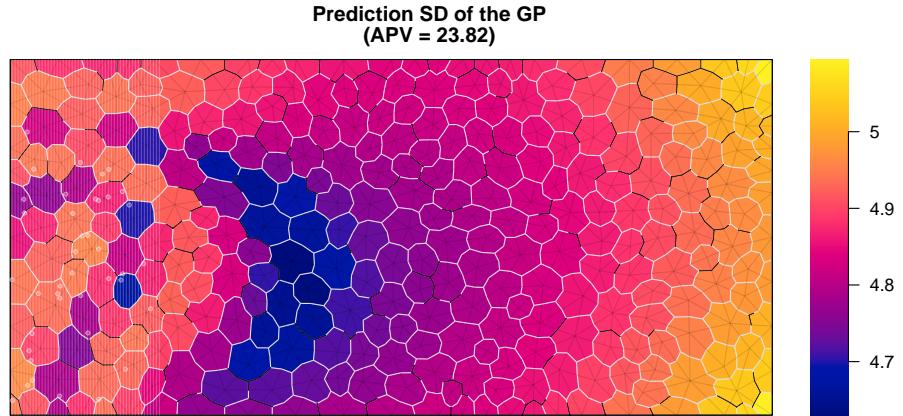


(b) APV vs survey effort for one LGCP with Clusters dataset

Figure 12: Plots of mean squared prediction error (MSPE) and average prediction variance (APV) vs survey effort for each plan applied to one LGCP with Clusters dataset. Line segments connect the median at each effort level.

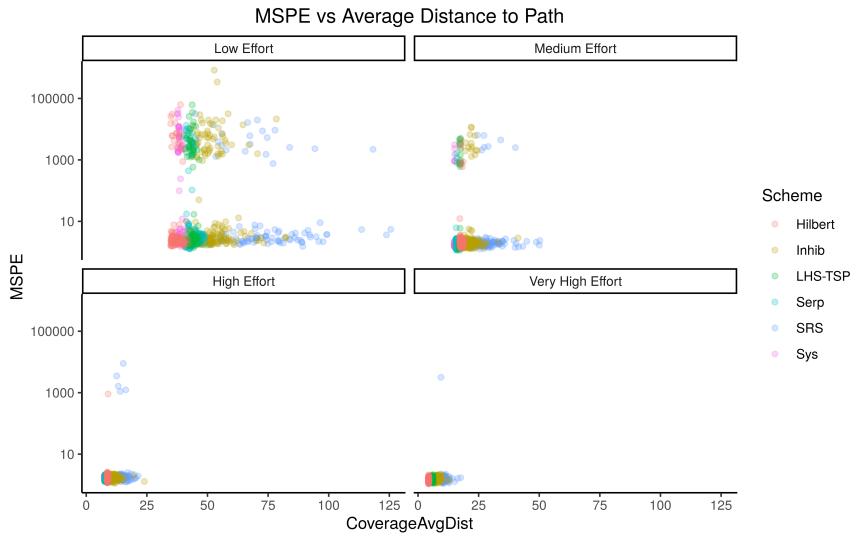


(a) Predicted log-intensity

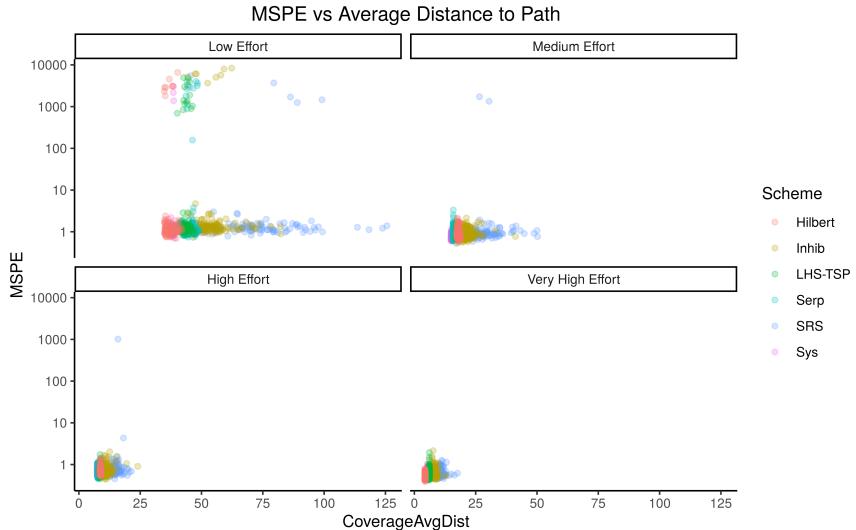


(b) Prediction SD

Figure 13: Predicted log-intensity function and prediction standard deviation using data observed via a systematic sample of a small section of the site. The SD is shown for each finite element node.



(a) MSPE vs coverage for one LGCP dataset



(b) MSPE vs coverage for one LGCP with Clusters dataset

Figure 14: MSPE plotted against the maximum distance from any point in the site to its the nearest point on the path. As survey effort increases, the cloud of points drifts down (lower MSPE) and to the left (lower average distance to the path).

³⁸⁰ to 12261 units, while the predicted log-intensity surface is much more accurate
³⁸⁵ (Figure 16). MSPE decreases from 9.98 to 1.44 and APV improves from 38.16 to 9.55. Across the site, the prediction standard deviation is lower than before, and is now highest around in the middle. If we were to continue adding segments to the path, giving some attention to the central portion of the site could further improve the prediction.

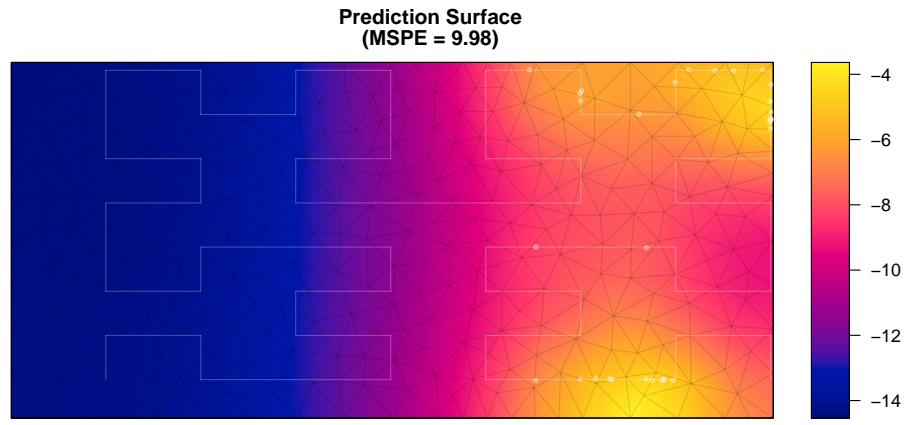
5. Discussion

5.1. Implications for sampling

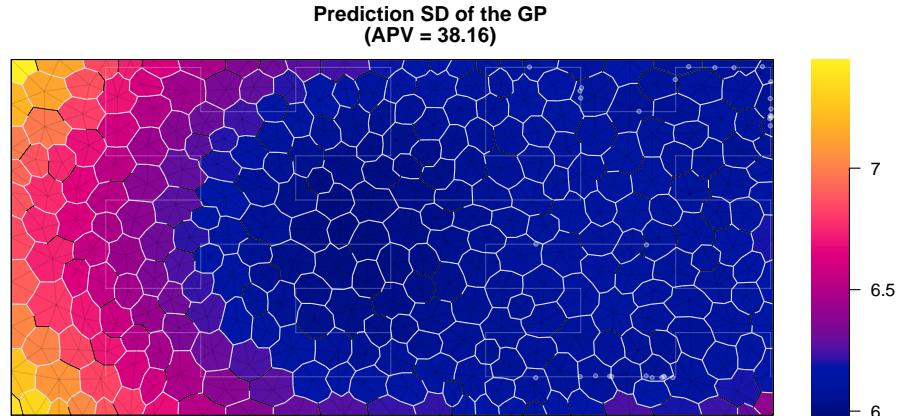
This study provides several insights about optimal sampling for LGCP models. First of all, prediction uncertainty increases with distance from observed ³⁹⁰ events, even along the path (consider the region of low mean and high standard deviation in the southwestern corner of Figure 5b). This behavior is different from the behavior of Gaussian process models in geostatistical studies, where observed design points all provide information that reduces uncertainty of nearby predictions, regardless of the values observed. In the LGCP model setting, an ³⁹⁵ "observed" strip that is observed to contain no events does not reduce the uncertainty in nearby predictions. This is reasonable behavior with sparse sampling because there may be events just out of detection range. Intuitively, observing no events tells us the intensity should be "low" but says little about how low, especially when intensity is modeled on a logarithmic scale.

⁴⁰⁰ This has consequences for adaptive or sequential sampling. A greedy algorithm based entirely on minimizing variance may excessively sample regions of low intensity, behavior that is unhelpful if the goal is to map hotspots of high intensity. We recommend including design-based considerations of spatial coverage in future studies on multi-stage design.

⁴⁰⁵ Additional considerations of applying path design to real-world studies include practical issues of data collection. Generally the path will be smoothed, without the perfectly-straight segments and instantaneous direction changes of our simulated data collection. Equipment may have limitations which ren-

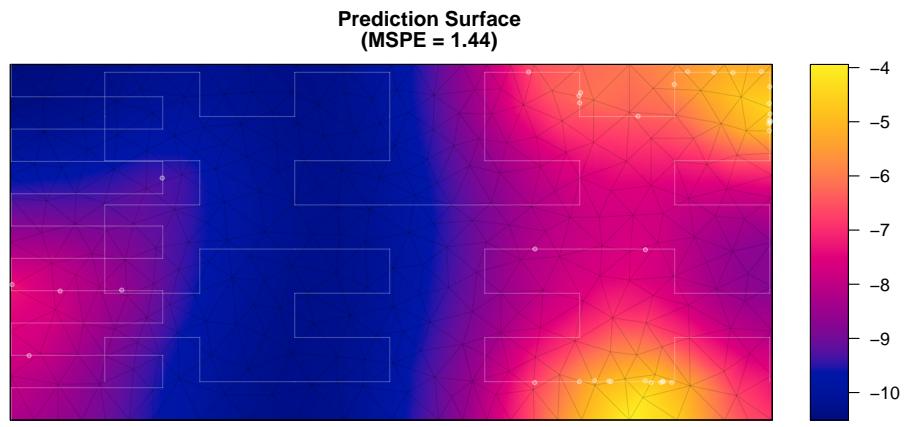


(a) Predicted log-intensity

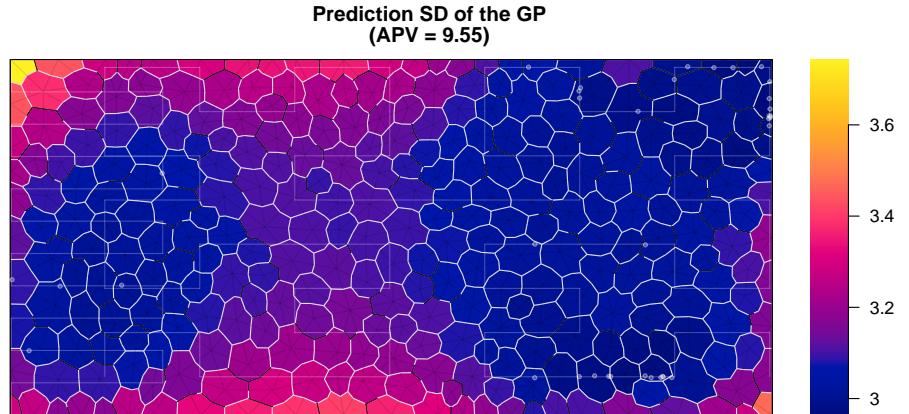


(b) Prediction SD

Figure 15: Predicted log-intensity function and prediction standard deviation using data observed via a Hilbert design. The SD is shown for each finite element node. This example is a medium-effort plan applied to a LGCP dataset.



(a) Predicted log-intensity



(b) Prediction SD

Figure 16: Predicted log-intensity function and prediction standard deviation using data observed via a Hilber design augmented post-hoc. The SD is shown for each finite element node.

der impossible the sharp corners of LHS-TSP designs or the repetitious short
410 segments and right angles of Hilbert curves. Future optimal design work could incorporate turn angles into a loss function or use multi-objective optimization (Lark, 2016).

5.2. Computing

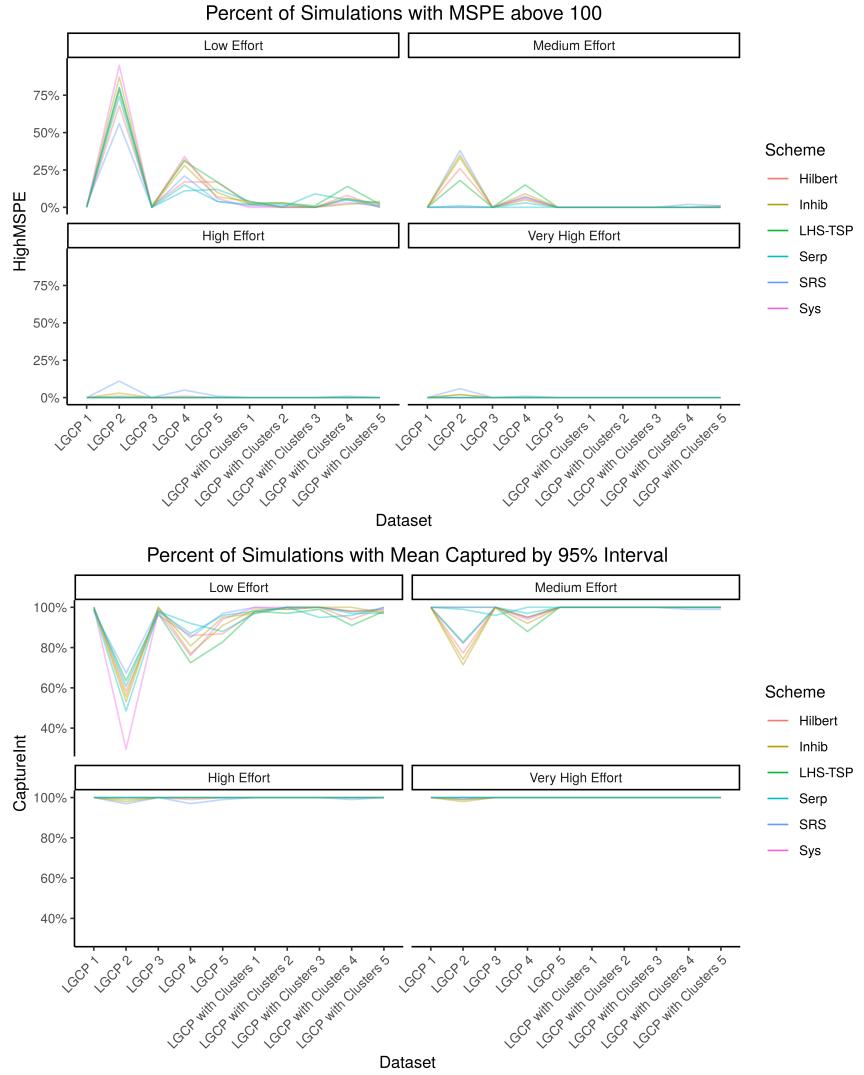
Bayesian computing (especially INLA) has brought LGCP models into the
415 realm of practicality. However, this is not yet a fully “push button, get results” procedure; in a handful of our simulations, the gradient descent step of INLA failed to converge with the default settings. In many other simulations, the posterior predictions appear numerically unstable at nodes on the boundary.

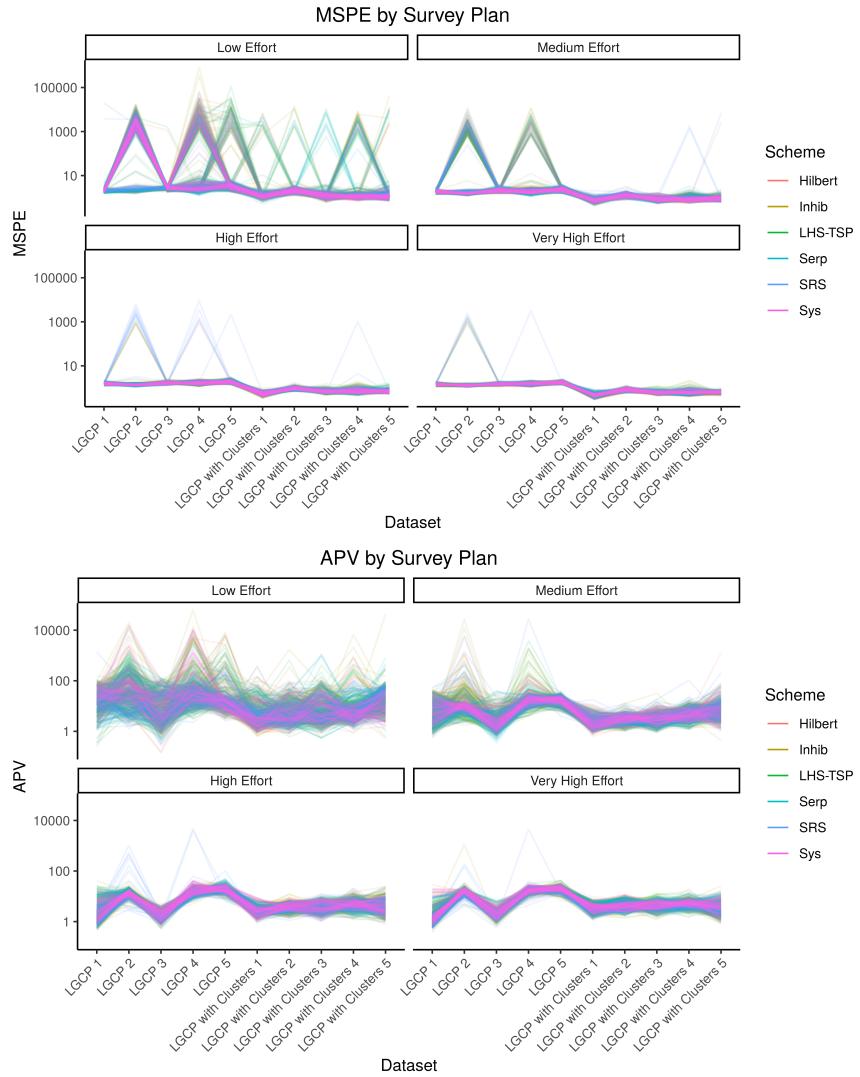
One aspect that would affect both of these issues is the choice of finite
420 element mesh. A finer mesh than the one used in these simulations would have more flexibility to model steep changes in the GP surface. We hypothesize that this would make convergence problems and edge effects less likely. The downside of a finer mesh is increased computing time because each mesh node becomes a pseudodata point that must be processed by the model-fitting software.

425 5.3. Conclusions

For optimal MSPE and APV, the choice of design scheme does not matter much as long as it provides spatial coverage. For short paths, including corners and regular spacing help avoid model-fitting problems, as in the Hilbert and serpentine transect designs. On the other hand, systematic line-transect designs
430 provide the best spatial coverage but require many transects for good model performance. There is a tradeoff between obtaining a useable posterior and employing a simple design. We recommend the Hilbert or serpentine designs if path length is heavily constrained, but systematic line transects are fine if the path can be long. A simulation using available prior information for a specific
435 study will be helpful in assessing if a given length is “long enough” to use a simple design.

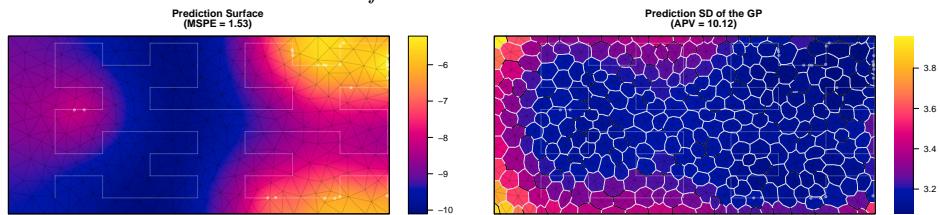
Appendix A. Profiles of designs across all dataset



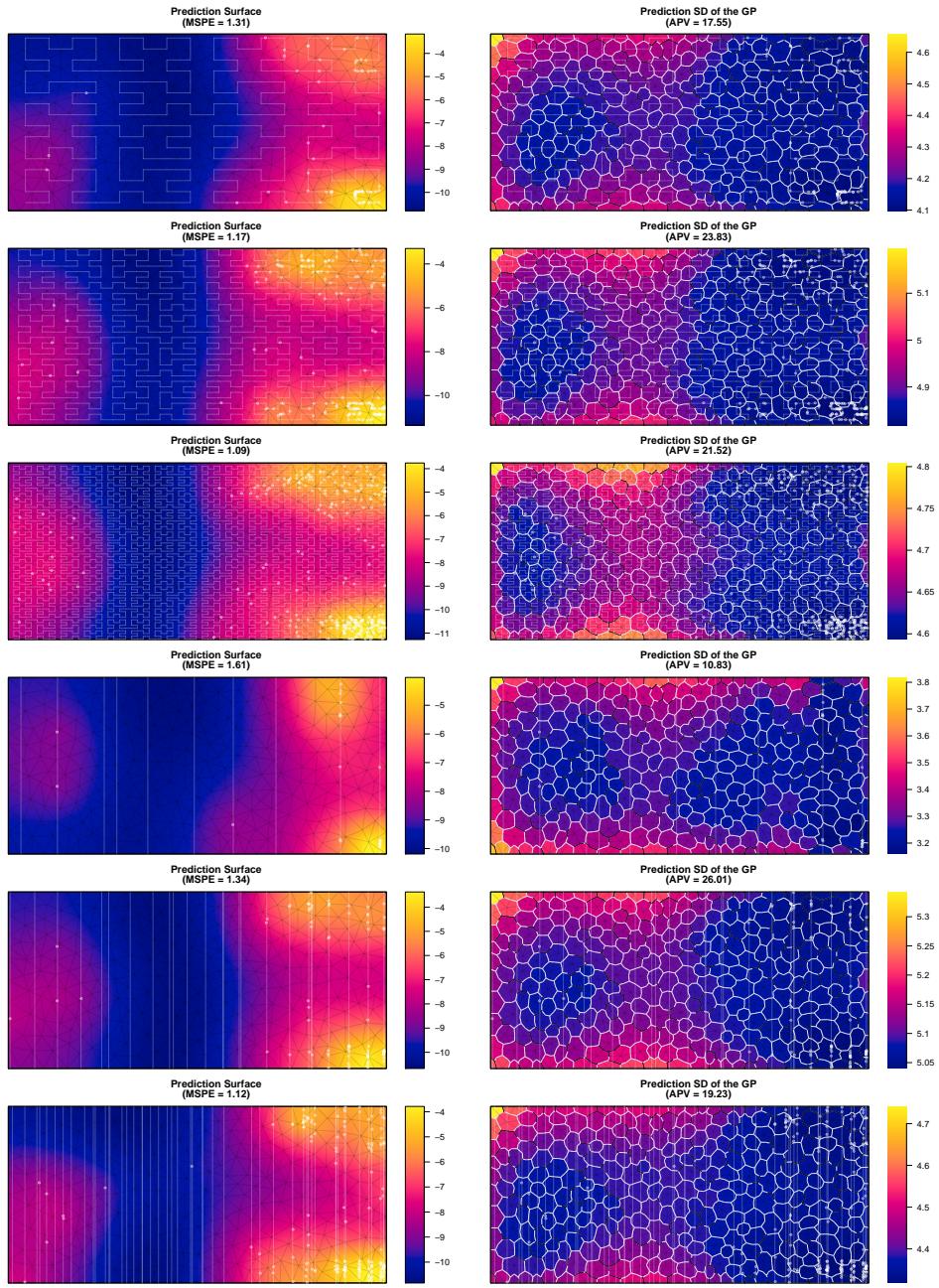


Appendix B. Additional plots of posterior predictions

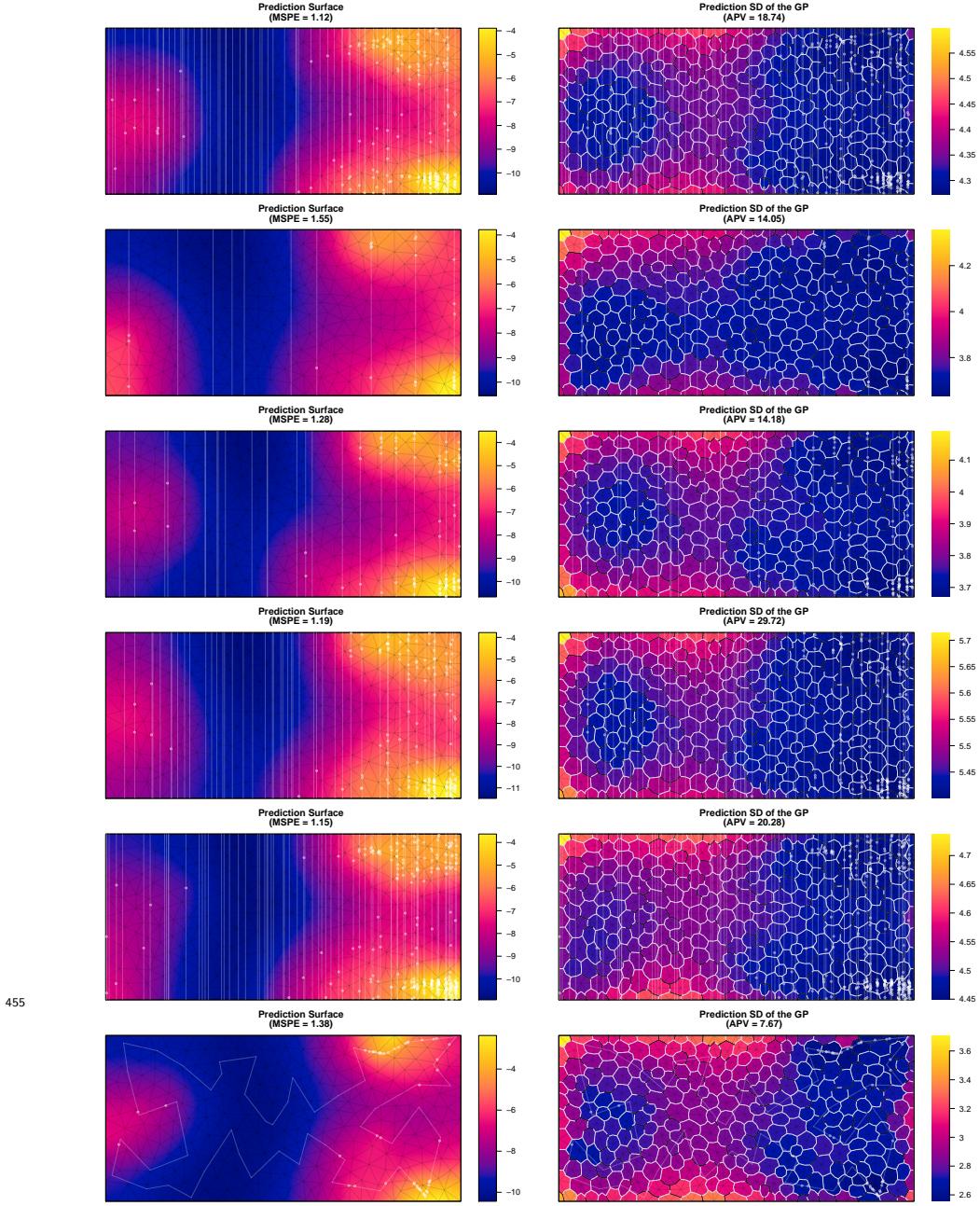
Appendix B.1. Minimum-MSPE surfaces

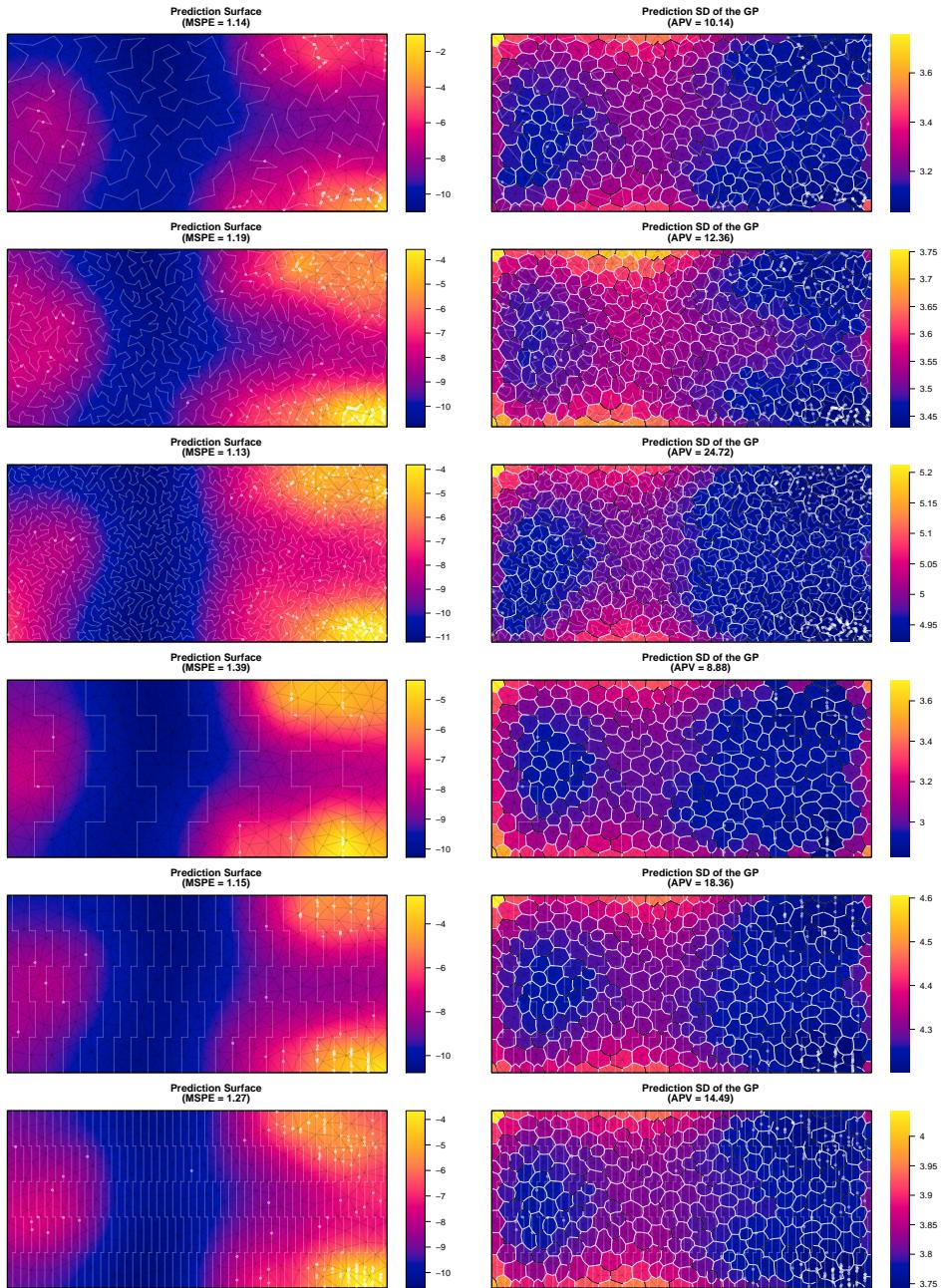


445

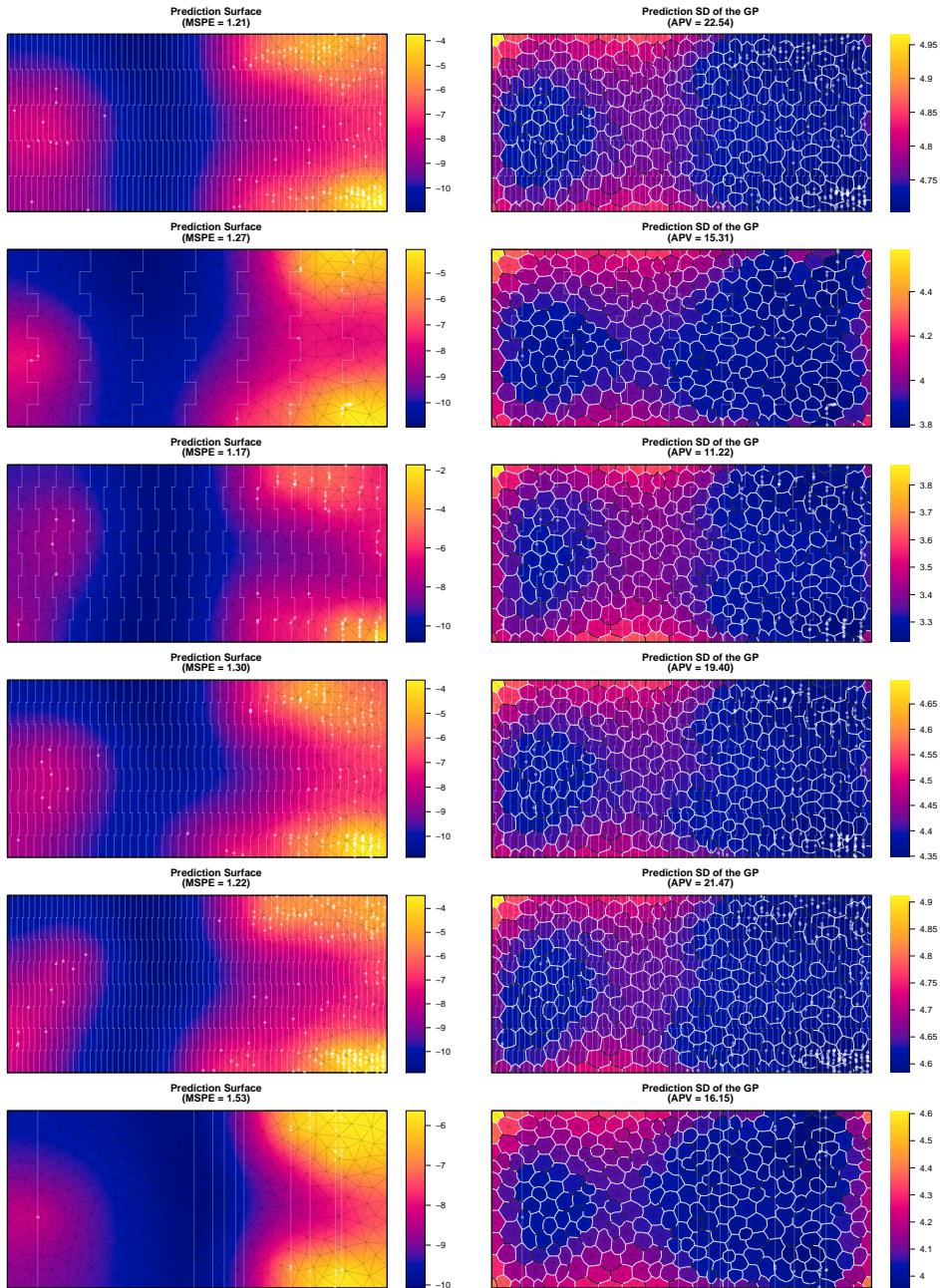


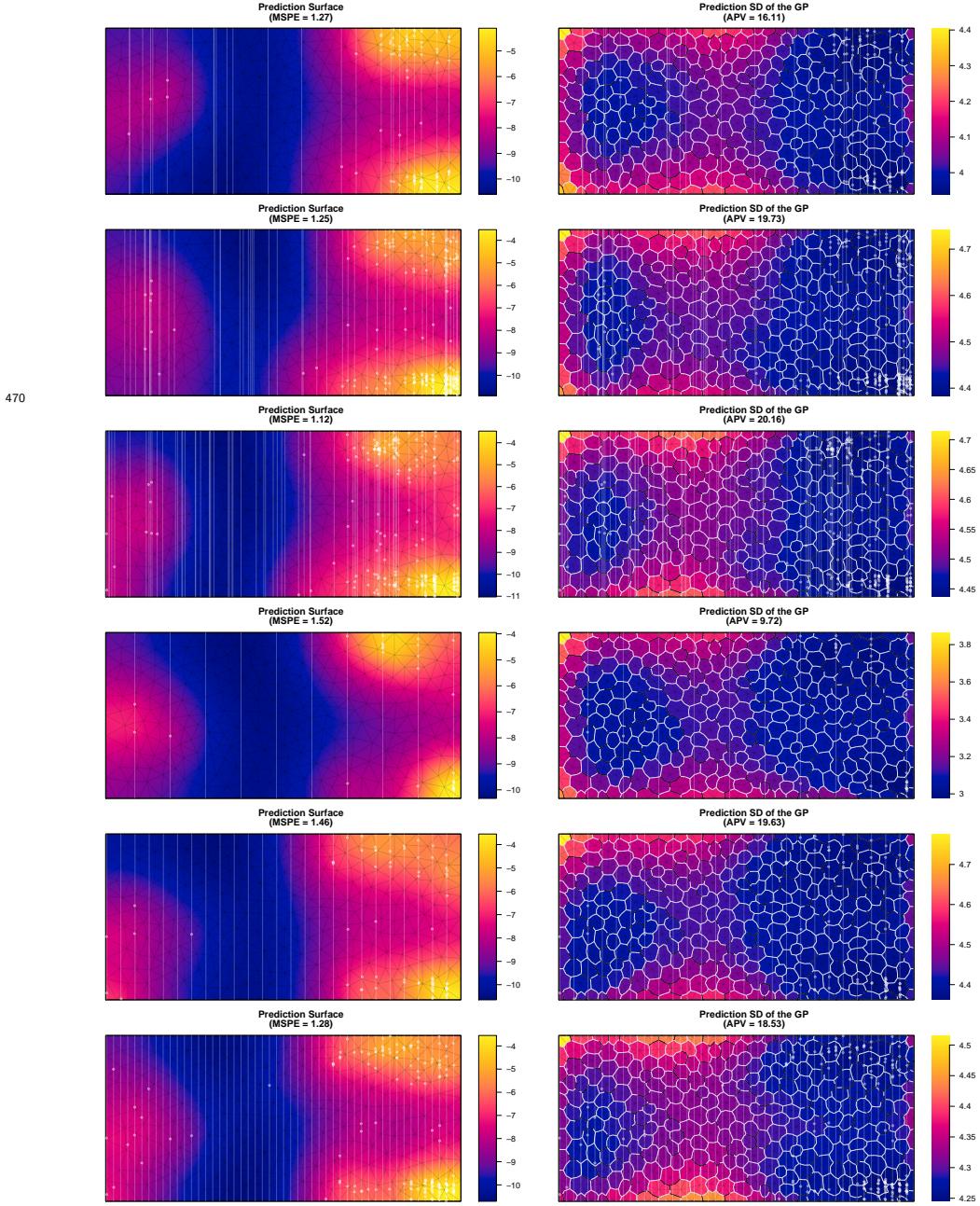
450

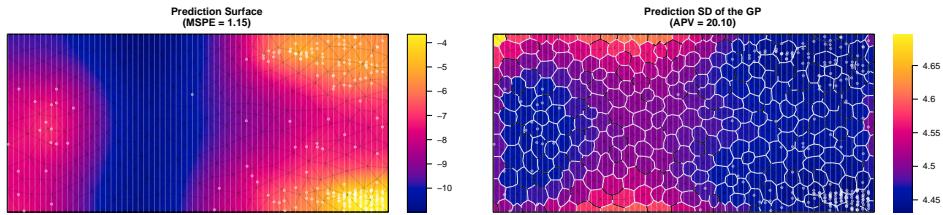




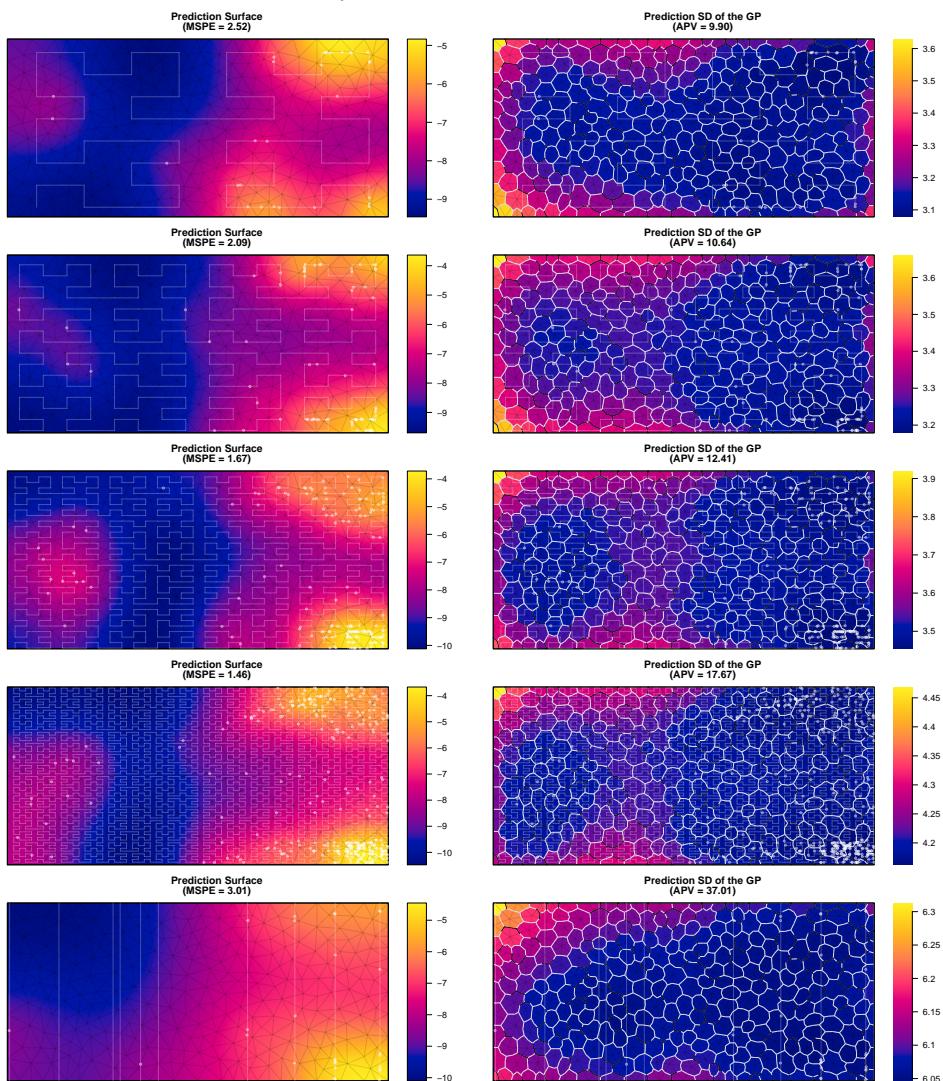
465



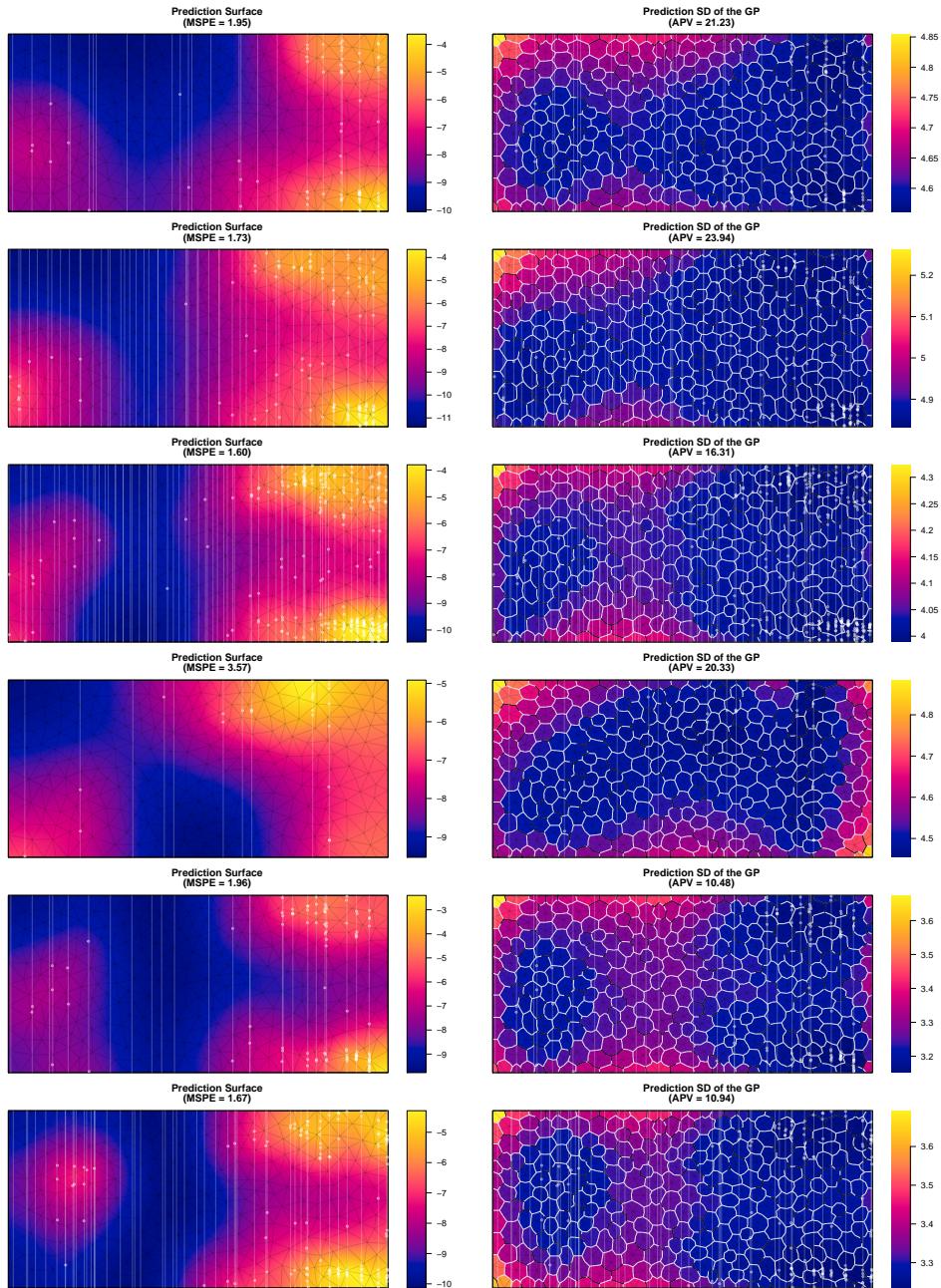




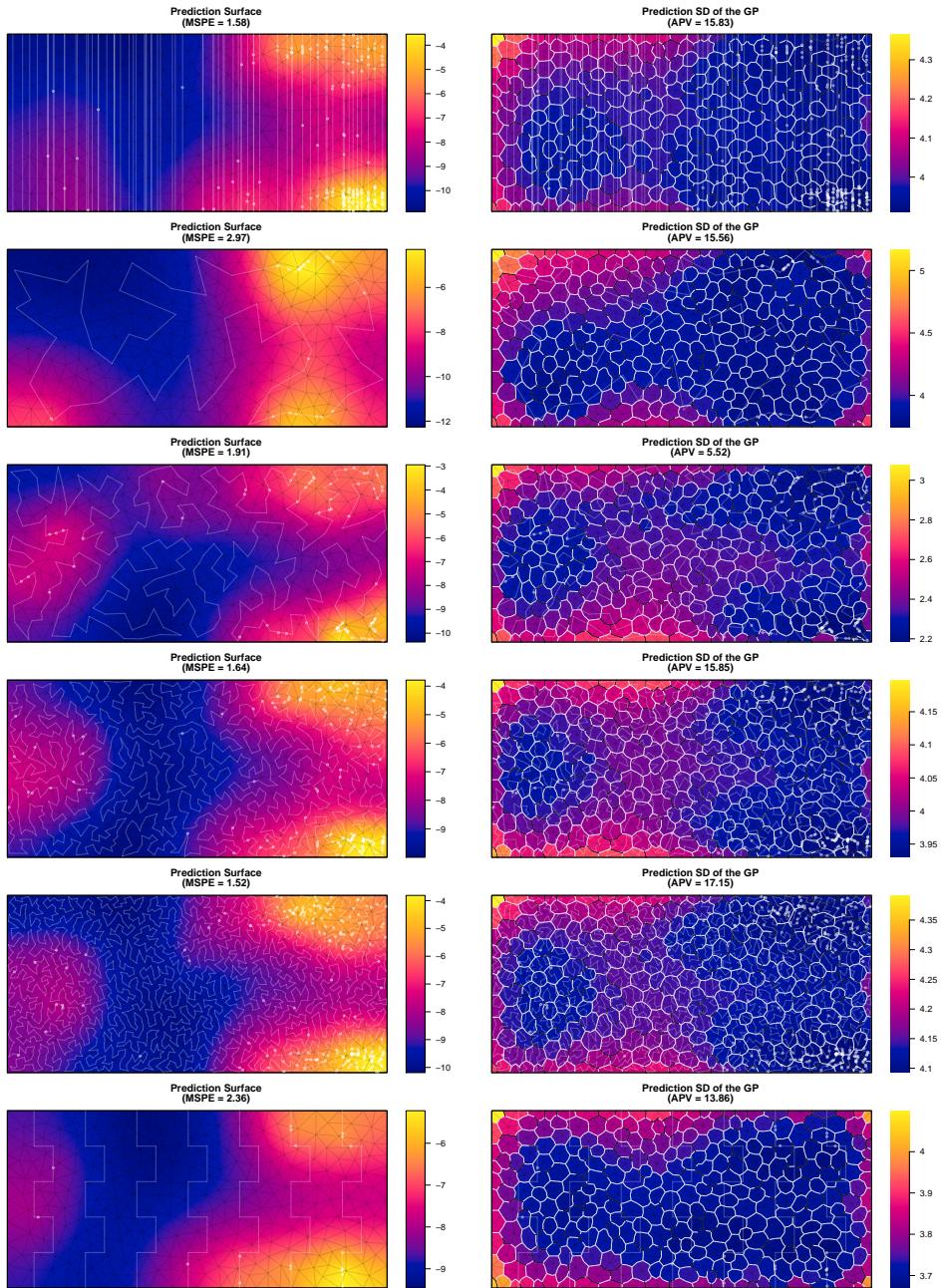
Appendix B.2. Median-MSPE surfaces

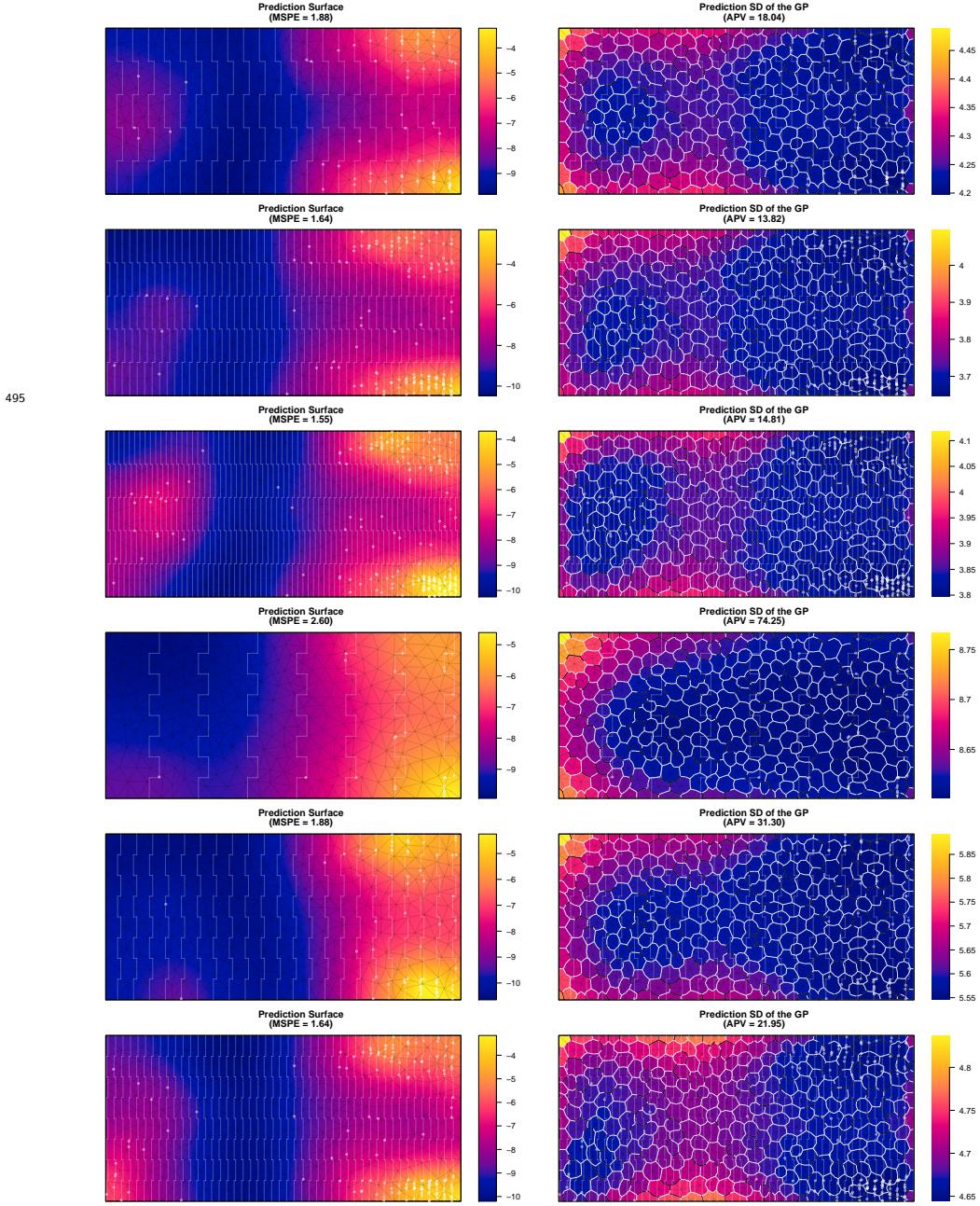


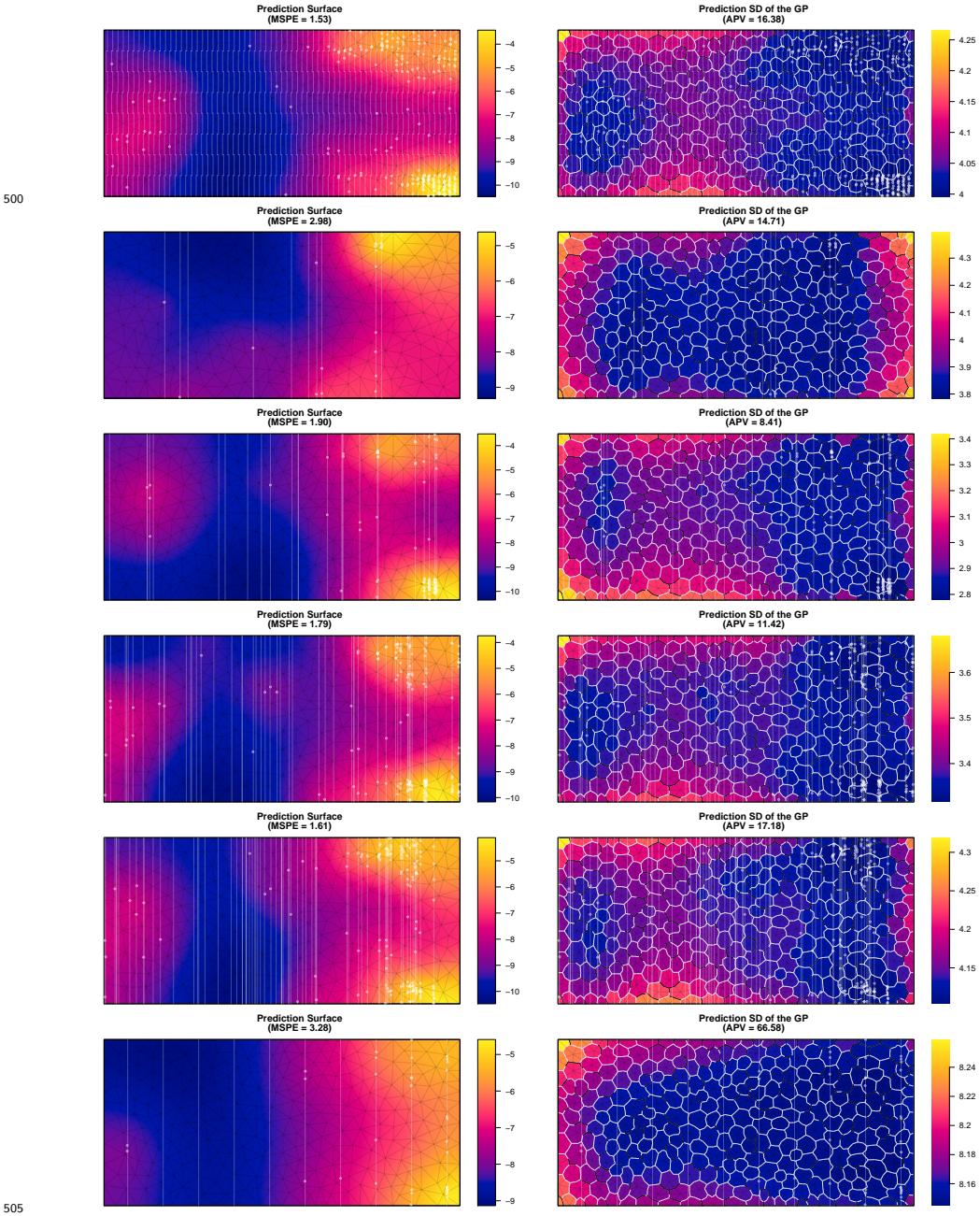
485

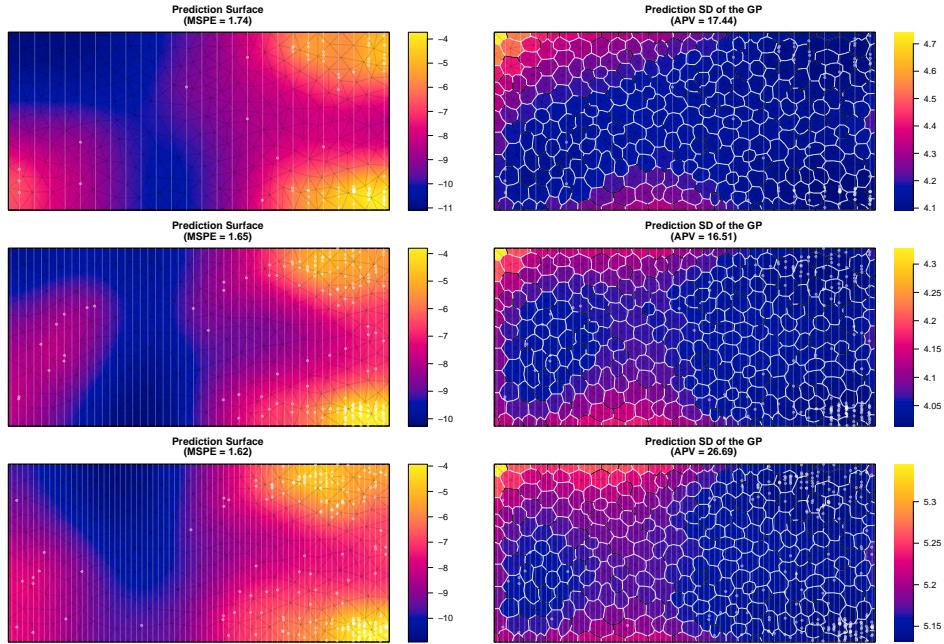


490

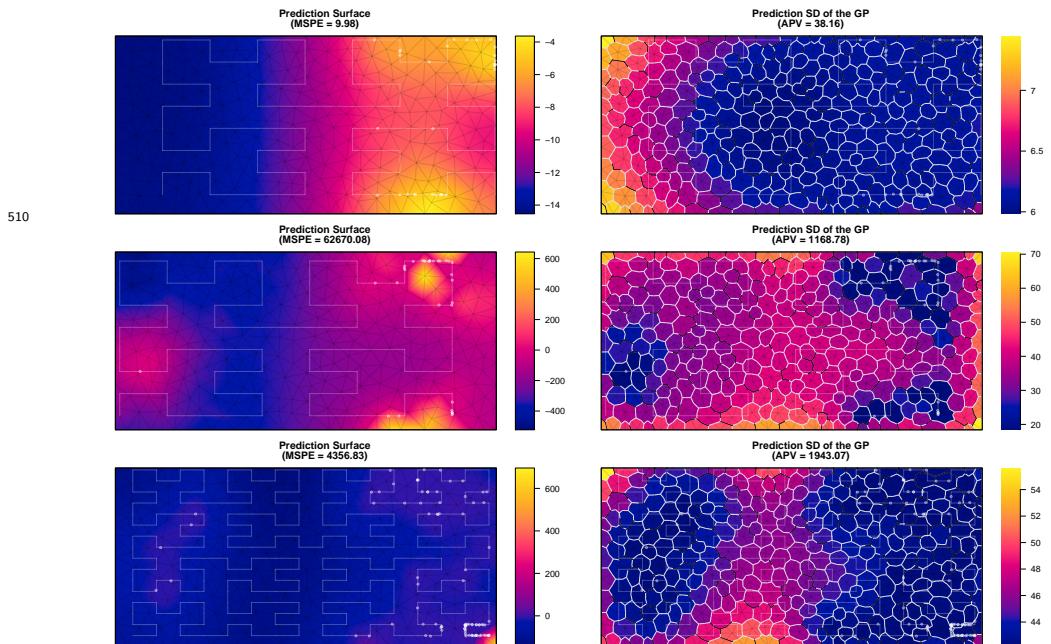


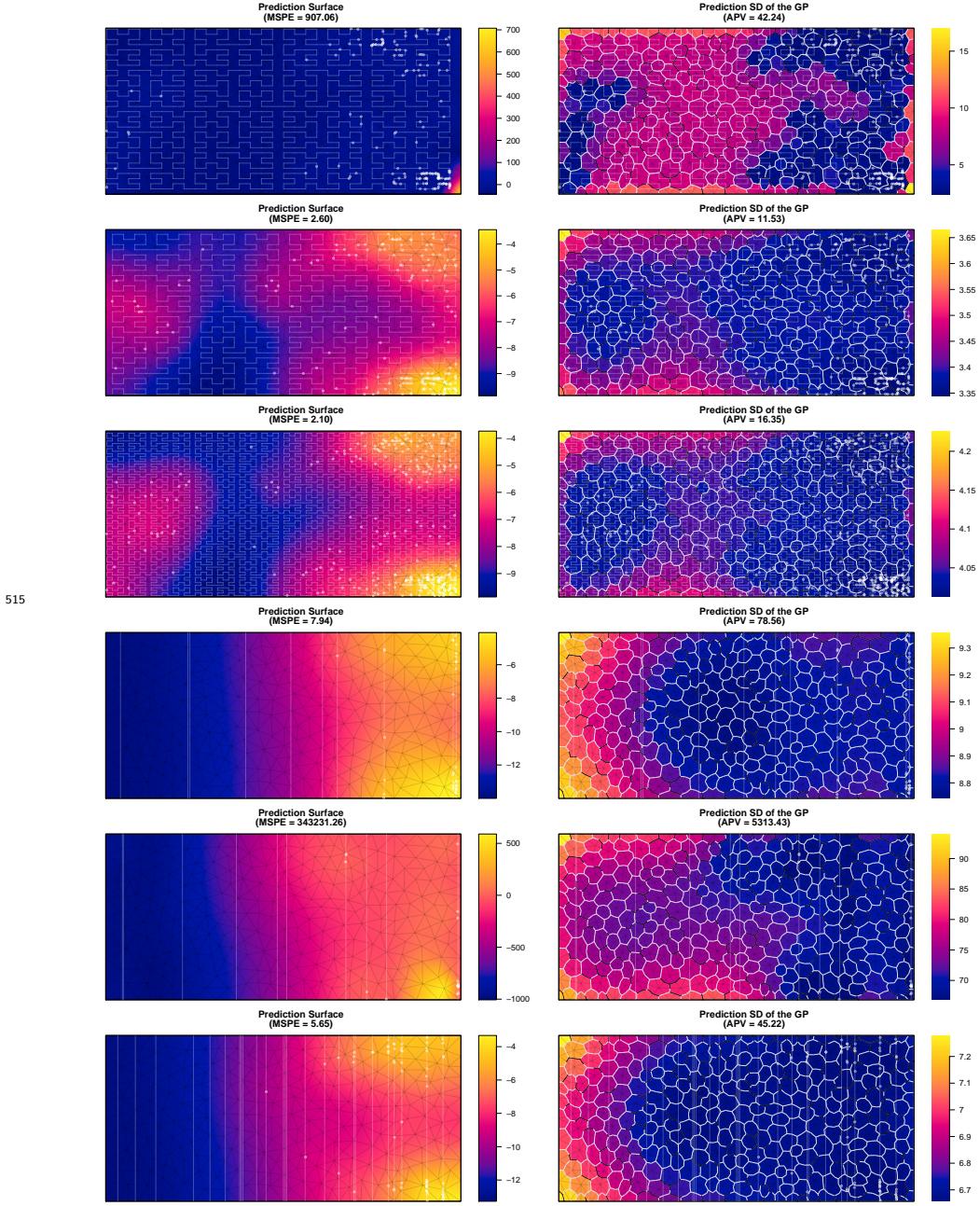




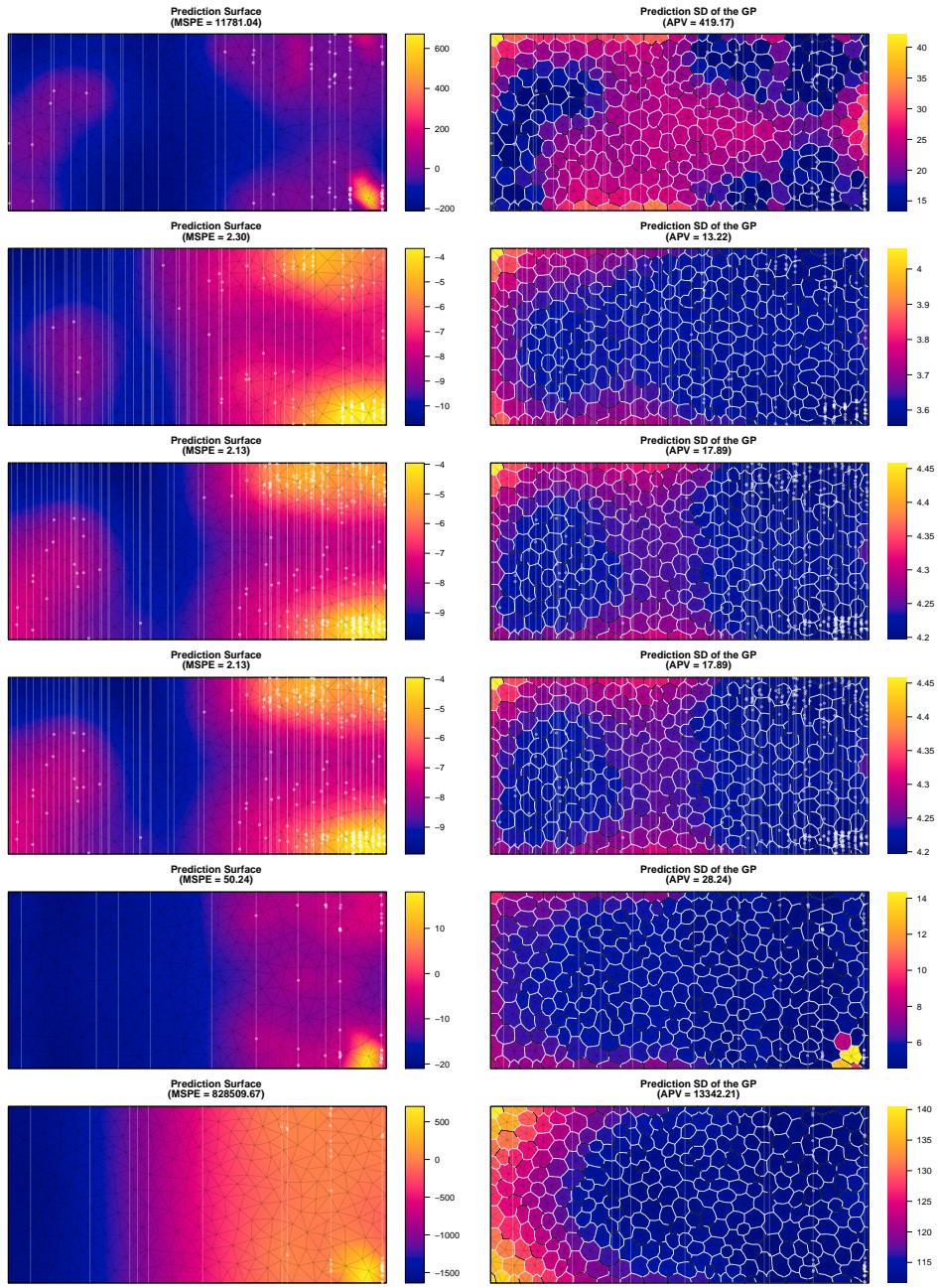


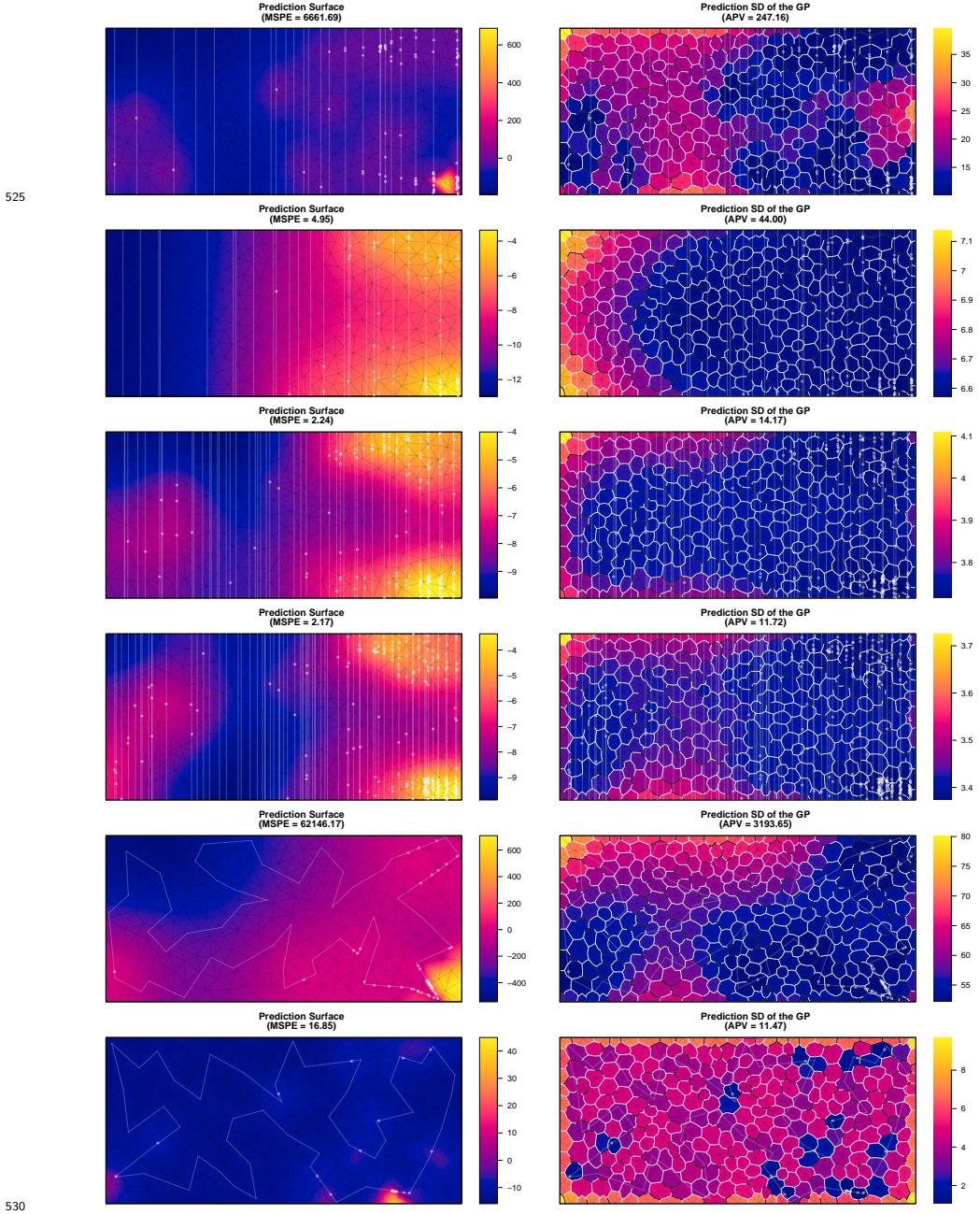
Appendix B.3. Max-MSPE surfaces

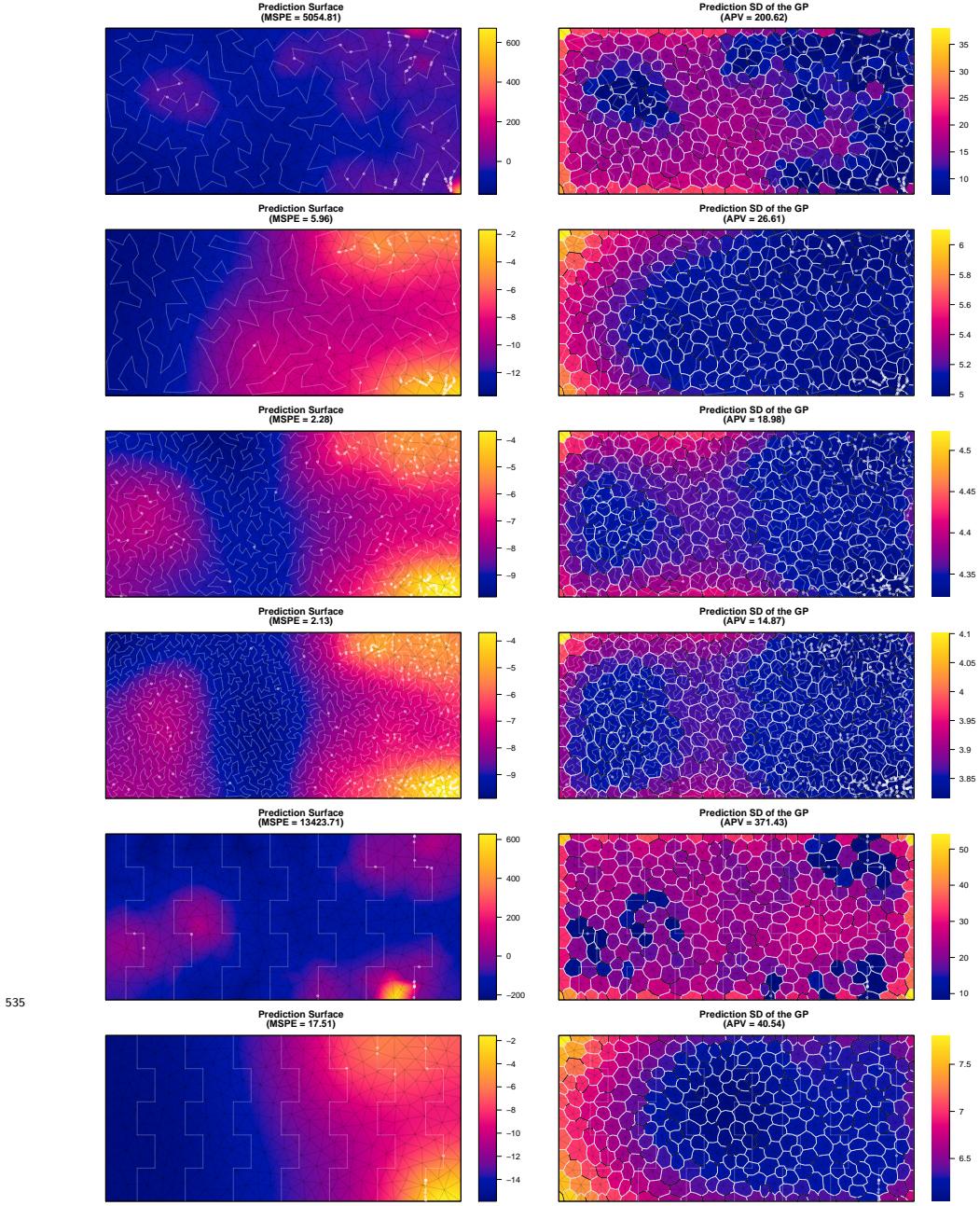


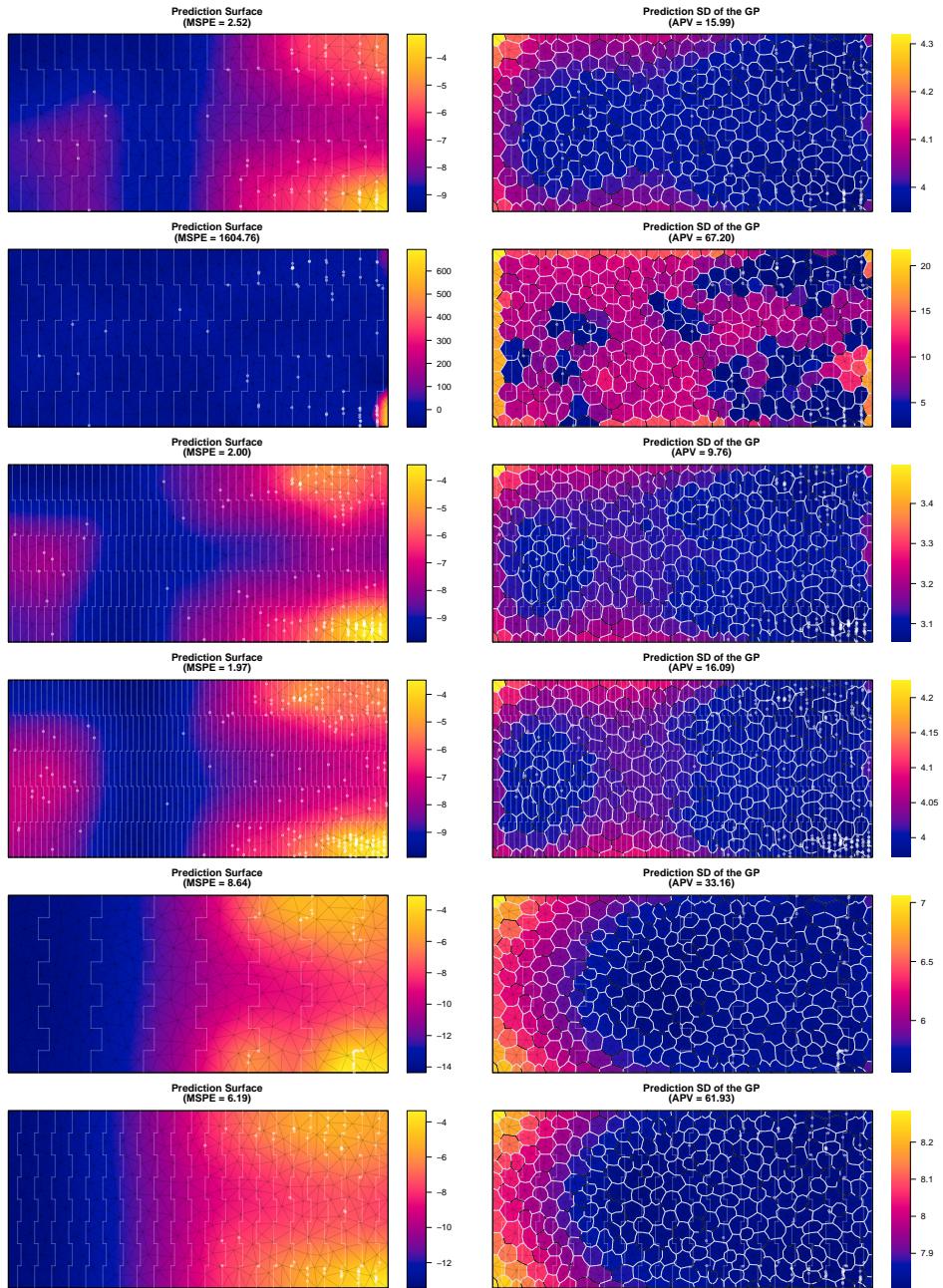


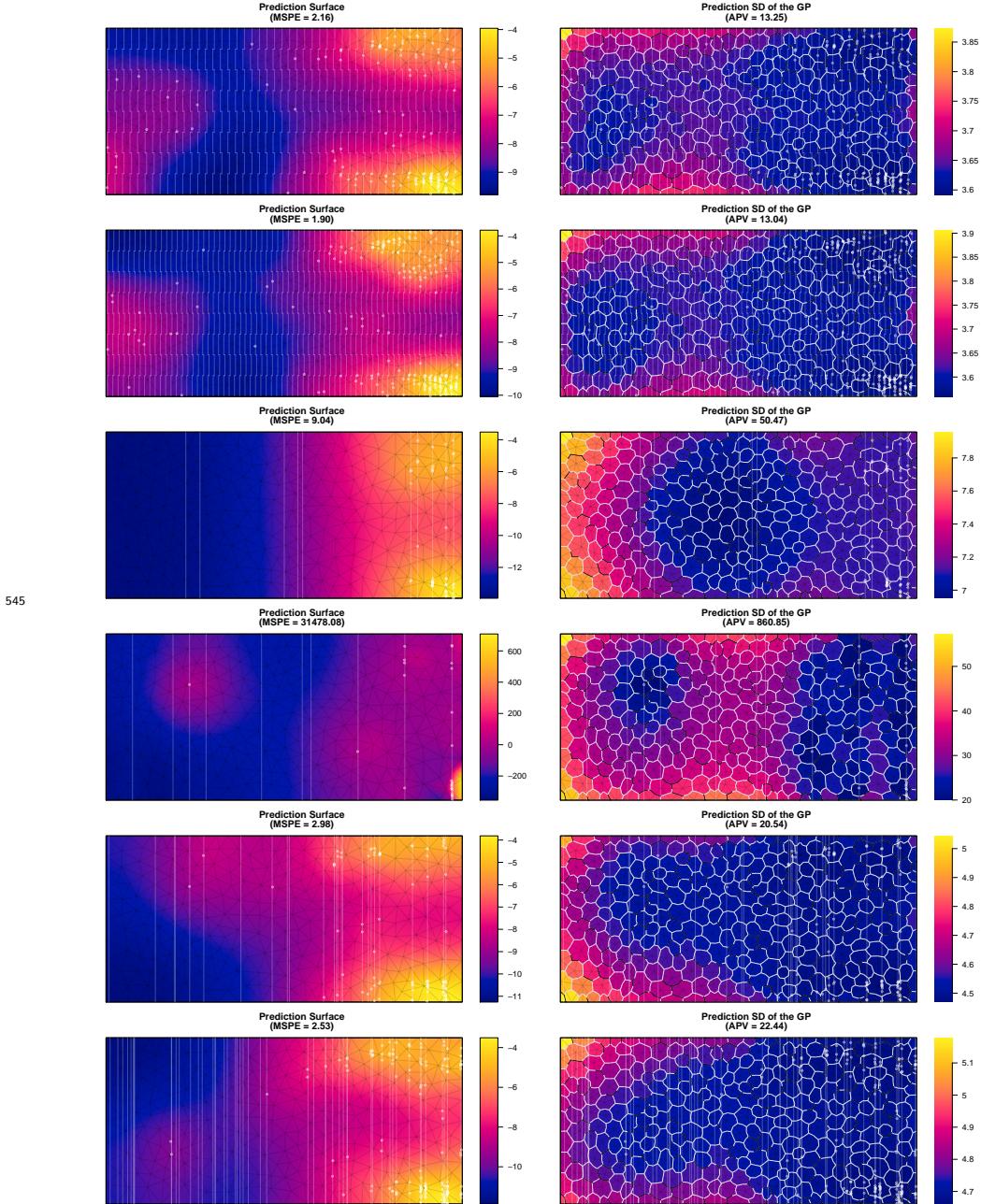
520

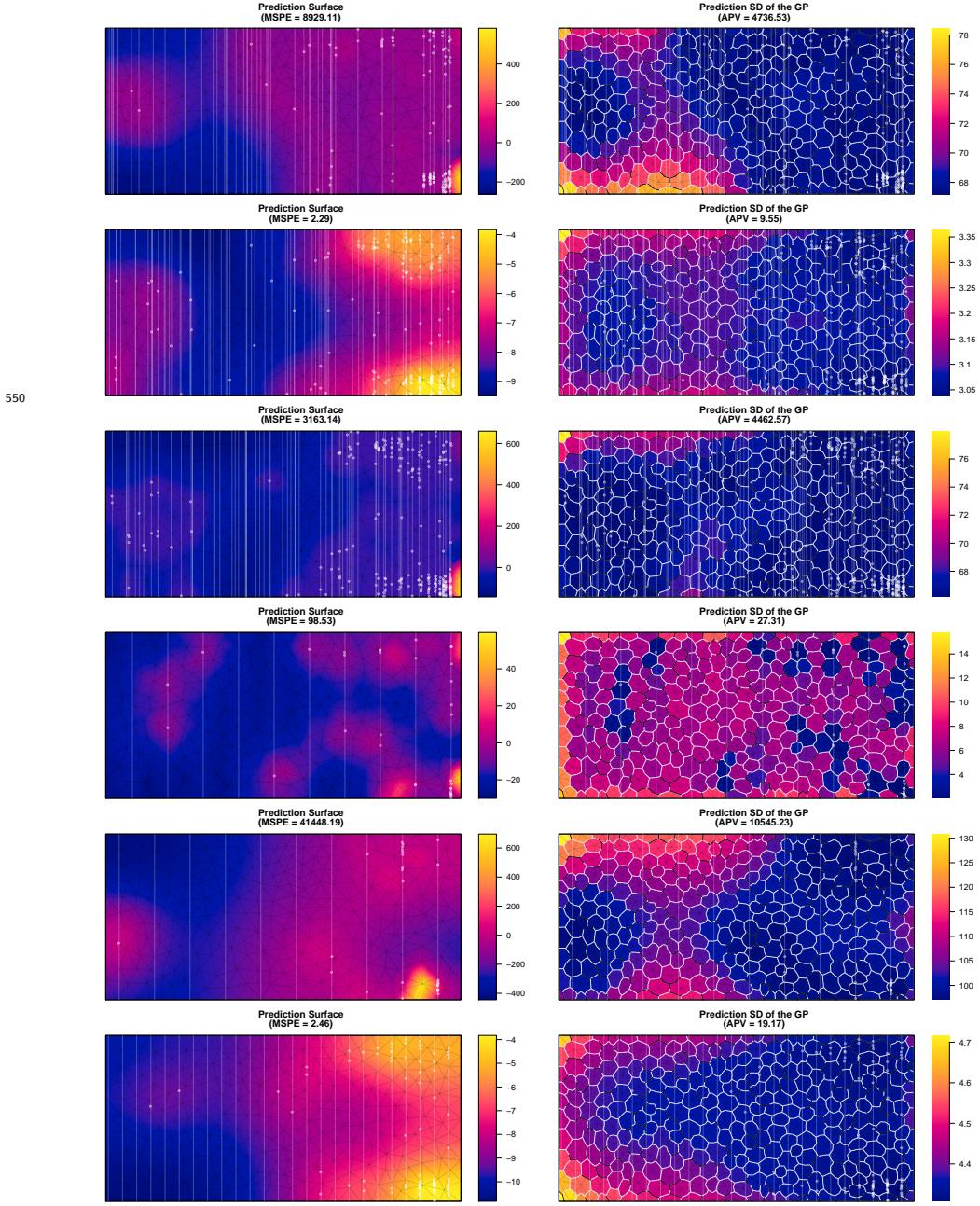


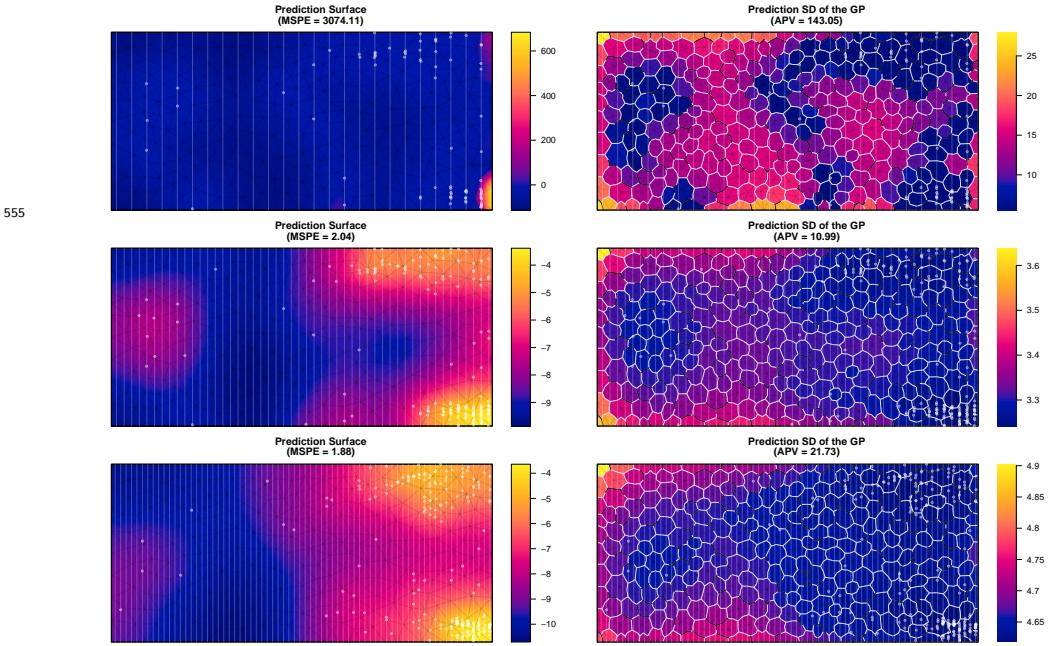












Appendix C. Tables summarizing results

Appendix D. Notation and terminology

- 560
- process defined on $\mathcal{D} \subset \mathbb{R}^d$, domain of the intensity function, in this manuscript $d = 2$
 - observation window $\mathcal{S} \subset \mathcal{D}$
 - define three regions:
 - the domain \mathcal{D} over which the process mathematically operates
 - the study region \mathcal{R} over which inferences are desired
 - the observed/sampled observation window \mathcal{S}
 - general relationship is $\mathcal{S} \subset \mathcal{R} \subset \mathcal{D} \subset \mathbb{R}^d$ where all of the subset symbols taken to mean “subset or equal”
- 565

- \mathcal{D} can be bounded or unbounded (often equal to \mathbb{R}^d), \mathcal{S} practically always bounded, \mathcal{R} bounded or unbounded depending on application and inferential goals
- the “fully surveyed” (censused) situation is $\mathcal{S} = \mathcal{R}$
- survey path \mathcal{P} is a one-dimensional subset of \mathcal{R}
 - set of one or more sequences of waypoints connected by line segments
 - \mathcal{S} is the set of all points within a fixed (and assumed known) radius of \mathcal{P}
- \mathbf{X} point process on \mathcal{R} , $\mathbf{x} = \{x_1, \dots, x_n\}$ realized point pattern
 - $\mathbf{X}_{\mathcal{S}} = \mathbf{X} \cap \mathcal{S}$ the restriction of \mathbf{X} to \mathcal{S} , $\mathbf{x} = \mathbf{X} \cap \mathcal{S}$ the realized observable point pattern
- point $x \in \mathbf{x}$ called an event
- intensity function $\lambda(u)$
- types of “points” in space:
 - x event in the point pattern
 - s numerical integration node
 - u arbitrary location in \mathcal{D} used to index intensity function and predictors
- $z(u)$ a column vector of covariates/predictors at u (not used in this manuscript)
- “point” refers to a u unless clearly stated otherwise
- bold for sets and spatial processes, normal italics for spatial vectors
- y and variations will be used for objects derived from the point pattern, e.g. marks, pseudodata
- distance sampling fits into the framework with expansion of notation to include a (nontrivial) detection function and differentiate between the observed and observable point patterns

595 **References**

- Anders, S., 2009. Visualisation of genomic data with the Hilbert curve. *Bioinformatics* doi:10.1093/bioinformatics/btp152.
- Blangiardo, M., Cameletti, M., 2015. Spatial and Spatio-temporal Bayesian Models with R-INLA. Wiley.
- 600 Borkowski, J.J., Piepel, G.F., 2009. Uniform designs for highly constrained mixture experiments. *Journal of Quality Technology* 41, 35–47.
- Carnell, R., 2020. lhs: Latin Hypercube Samples. URL: <https://CRAN.R-project.org/package=lhs>. R package version 1.0.2.
- Chakraborty, A., Gelfand, A.E., Wilson, A.M., Latimer, A.M., Silander, J.A.,
605 2011. Point pattern modelling for degraded presence-only data over large regions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 60, 757–776.
- Chipeta, M., Terlouw, D., Phiri, K., Diggle, P., 2017. Inhibitory geostatistical designs for spatial prediction taking account of uncertain covariance structure.
610 *Environmetrics* 28.
- Diggle, P., Lophaven, S., 2006. Bayesian geostatistical design. *Scandinavian Journal of Statistics* 33, 53–64.
- Fan, J.A., Yeo, W.H., Su, Y., Hattori, Y., Lee, W., Jung, S.Y., Zhang, Y., Liu, Z., Cheng, H., Falgout, L., Bajema, M., Coleman, T., Gregoire, D., Larsen,
615 R.J., Huang, Y., Rogers, J.A., 2014. Fractal design concepts for stretchable electronics. *Nature communications* 5, 3266.
- Flagg, K.A., Hoegh, A., Borkowski, J.J., 2020. Modeling partially surveyed point process data: Inferring spatial point intensity of geomagnetic anomalies. *Journal of Agricultural, Biological and Environmental Statistics* 25, 186–205.

- 620 Fuglstad, G.A., Simpson, D., Lindgren, F., Rue, H., 2019. Constructing priors that penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association* 114, 445–452.
- Hahsler, M., Hornik, K., 2020. TSP: Traveling Salesperson Problem (TSP). URL: <https://CRAN.R-project.org/package=TSP>. R package version 1.1-10.
- 625 Husslage, B.G., Rennen, G., Van Dam, E.R., Den Hertog, D., 2011. Space-filling latin hypercube designs for computer experiments. *Optimization and Engineering* 12, 611–630.
- Illian, J.B., Sørbye, S.H., Rue, H., 2012. A toolbox for fitting complex spatial 630 point process models using integrated nested laplace approximation (inla). *The Annals of Applied Statistics* , 1499–1530.
- Johnson, D., Laake, J., VerHoef, J., 2014. DSpat: Spatial Modelling for Distance Sampling Data. URL: <https://CRAN.R-project.org/package=DSpat>. R package version 0.1.6.
- 635 Johnson, M.E., Moore, L.M., Ylvisaker, D., 1990. Minimax and maximin distance designs. *Journal of statistical planning and inference* 26, 131–148.
- Kelly, F.P., Ripley, B.D., 1976. A note on Strauss's model for clustering. *Biometrika* 63, 357–360.
- Lark, R., 2016. Multi-objective optimization of spatial sampling. *Spatial Statistics* 640 18, 412–430.
- Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73, 423–498.
- 645 Liu, J., Vanhatalo, J., 2020. Bayesian model based spatiotemporal survey designs and partially observed log Gaussian Cox process. *Spatial statistics* 35, 100392.

- Ma, Q., Zhang, Y., 2016. Mechanics of fractal-inspired horseshoe microstructures for applications in stretchable electronics. *Journal of Applied Mechanics* 83.
- Matzke, B., Wilson, J., Newburn, L., Dowson, S., Hathaway, J., Sego, L., Bramer, L., Pulsipher, B., 2014. Visual Sample Plan Version 7.0 User's Guide. Pacific Northwest National Laboratory. Richland, Washington. URL: <http://vsp.pnnl.gov/docs/PNNL-23211.pdf>.
- McKay, M.D., Beckman, R.J., Conover, W.J., 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21, 239–245.
- Møller, J., Syversveen, A.R., Waagepetersen, R.P., 1998. Log Gaussian Cox processes. *Scandinavian journal of statistics* 25, 451–482.
- Ogorzałek, M.J., 2009. Fundamentals of fractal sets, space-filling curves and their applications in electronics and communications, in: Intelligent Computing Based on Chaos. Springer, pp. 53–72.
- Pollard, J., Palka, D., Buckland, S., 2002. Adaptive line transect sampling. *Biometrics* 58, 862–870.
- R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)* 71, 319–392.
- Sagan, H., 1994. Space-filling curves. Springer.
- Simpson, D., Illian, J.B., Lindgren, F., Sørbye, S.H., Rue, H., 2016. Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *Biometrika* 103, 49–70.

- Simpson, D., Rue, H., Riebler, A., Martins, T.G., Sørbye, S.H., et al., 2017. Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical science* 32, 1–28.
- Strauss, D.J., 1975. A model for clustering. *Biometrika* 62, 467–475.
- 680 USACE, 2015. Technical Guidance for Military Munitions Response Actions. Technical Report EM 200-1-15. United States Army Corps of Engineers. URL: http://www.publications.usace.army.mil/Portals/76/Publications/EngineerManuals/EM_200-1-15.pdf.
- Yuan, Y., Bachl, F.E., Lindgren, F., Borchers, D.L., Illian, J.B., Buckland, S.T.,
685 Rue, H., Gerrodette, T., et al., 2017. Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *The Annals of Applied Statistics* 11, 2270–2297.