
Sensor Placement for Spatial Gaussian Processes with Integral Observations

Krista Longi¹, Chang Rajani¹, Tom Sillanpää², Joni Mäkinen²,
Timo Rauhala³, Ari Salmi², Edward Hægström², Arto Klami¹

¹Department of Computer Science, University of Helsinki, Finland

²Electronics Research Laboratory, Department of Physics, University of Helsinki, Finland

³Altum Technologies, Finland

Abstract

Gaussian processes (GP) are a natural tool for estimating unknown functions, typically based on a collection of point-wise observations. Interestingly, the GP formalism can be used also with observations that are integrals of the unknown function along some known trajectories, which makes GPs a promising technique for inverse problems in a wide range of physical sensing problems. However, in many real world applications collecting data is laborious and time consuming. We provide tools for optimizing sensor locations for GPs using integral observations, extending both model-based and geometric strategies for GP sensor placement. We demonstrate the techniques in ultrasonic detection of fouling in closed pipes.

1 INTRODUCTION

GPs are widely used for modeling unknown functions, thanks to the closed-form posterior inference conditional on noisy observations. Importantly, this property holds also for observations corresponding to any linear operator of the function, such as derivatives and integrals [Rasmussen and Williams, 2006]. Derivative observations are frequently used in modeling dynamical systems [Solak et al., 2003], whereas use of integral observations has been limited to a few recent applications [O’Callaghan and Ramos, 2011, Jidling et al., 2018, Purisha et al., 2019, Jidling et al., 2019, Tanaka et al., 2019, Law et al., 2018, Hamelijneck et al., 2019].

We consider problems where the goal is to estimate an unknown spatial (in our case two-dimensional) function based on a collection of observations that are integrals of that function (Figure 1), computed along some known paths defined by the sensing configuration. Examples

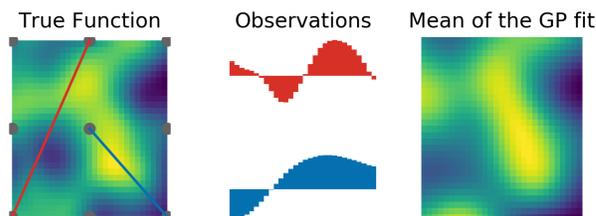


Figure 1: Solving a spatial inverse problem with line integral observations. **Left:** An unknown function is monitored by a collection of 9 sensors, each transmitting a signal that can be recorded by each of the other sensors. **Middle:** Illustration of two slices of the function along two example integration paths. **Right:** Mean estimate of the function, conditional on noisy observations of all 36 line integrals induced by the 9 sensors.

that map to this general problem formulation include laser scanners [O’Callaghan and Ramos, 2011], tomographic reconstruction [Jidling et al., 2018, Purisha et al., 2019], and ultrasonic structural health monitoring [Lu and Michaels, 2009] and fouling detection. For example, in the case of X-ray computed tomography, the observations measure the attenuation of the x-rays when transmitted through an object [Purisha et al., 2019].

In most cases the observations are collected by pairs of transmitting and receiving sensors: each observation corresponds to a signal originating from the transmitter that is recorded at the receiver, and the underlying function has point-wise effect on some property of the propagating signal, e.g., modulating its amplitude, frequency content, phase or velocity. For a practical inverse problem, a critical question concerns the placement of the sensors. Individual sensors are often expensive and setting up the sensors can be laborious and time consuming. In the case of CT scans, each additional measurement also increases the patient’s radiation exposure. In this work we focus on selecting optimal sensor locations before any mea-

measurements are made, as required in typical scenarios that require engineering effort for setting up the sensing configuration, but note that the techniques could be extended for active selection on additional sensors conditional on current measurements [Krause and Guestrin, 2007].

Sensor placement has been extensively studied for point observations [Krause et al., 2008, Sakiyama et al., 2016]. Model-based solutions evaluate the expected improvement of the function estimate for possible locations, measuring e.g. reduction of entropy or increase in mutual information as a proxy. Model-independent solutions ignore the GP and optimize for locations using geometric arguments [González-Banos, 2001] or black-box optimization of an objective characterizing the quality of the sensor set [Garnett et al., 2010]. The latter are often computationally more efficient, but naturally disregard information about the underlying model.

We extend both strategies of sensor placement for integral observations. Now each sensor induces several observations corresponding to multiple propagation paths, determined by the type of the sensors and the underlying geometry. The model-based solutions build directly on the earlier results, extended here to support multiple new observations being induced by individual sensors. The geometric approaches build on the *line arrangement* (see Agarwal and Sharir [2010] for an overview) that characterizes the sensor configurations in terms of intersections (vertex) of the lines and the segments (edges) and polygons (faces) induced by those, and optimize fitness functions derived for the arrangement.

We empirically characterise the behavior of the various sensor placement strategies for a few 2D geometries. The main experiment considers ultrasonic localization of fouling in closed metal pipes [Sillanpää et al., 2019]. Ultrasonic Lamb waves sent from a transmitter propagate along the pipe with group velocity depending on the material properties and thickness, and any fouling on the (inner) surface influences the velocity. Hence, time-of-flight difference between measurements of fouled and clean sections of a pipe can be analyzed as line integrals over the spatial fouling distribution. For practical use we want to minimize the number of sensors. As a sensor optimization problem this is interesting since each receiver records multiple integrals corresponding to different helical paths along the surface, as illustrated in Figure 2. Here we study the problem on simulated data, but the feasibility of ultrasonic fouling localization with line-integral GPs has been demonstrated for real measurement data on one sensor setting [Sillanpää et al., 2019].

2 BACKGROUND

A Gaussian process [Rasmussen and Williams, 2006] $f \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ is specified by a prior mean function $m(\mathbf{x})$ and a prior symmetric positive-definite kernel function $k(\mathbf{x}, \mathbf{x}')$. Most kernels have hyperparameters that determine the prior correlation structure, typically learned by maximizing the marginal likelihood.

We want to predict the function at test points x^* by conditioning on the observed data, and for conjugate normal likelihood we obtain the distribution in closed form. For observations $y_i = f(\mathbf{x}_i) + \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, the predictive distribution is a normal $\mathcal{N}(\mu, K)$ with

$$\begin{aligned} \mu &= K_{x^*x}(K_{xx} + \sigma^2 I)^{-1} \mathbf{y}, \\ K &= K_{x^*x^*} - K_{xx^*}(K_{xx} + \sigma^2 I)^{-1} K_{xx^*}^T \end{aligned} \quad (1)$$

where $[K_{xx}]_{ij} = k(x_i, x_j)$, $[K_{xx^*}]_{ij} = k(x_i, x_j^*)$ and $[K_{x^*x^*}]_{ij} = k(x_i^*, x_j^*)$.

2.1 GP WITH INTEGRAL OBSERVATIONS

GPs are closed under linear operators [Rasmussen and Williams, 2006]: If \mathcal{L} is a linear operator and $f \sim GP(m(x), k(x, x'))$, then $\mathcal{L}f$ is GP with

$$\mathbb{E}[\mathcal{L}f(x)] = \mathcal{L}m(x),$$

$$\text{cov}(\mathcal{L}f(x), \mathcal{L}f(x')) = \mathcal{L}x \mathcal{L}x'^T K(x, x').$$

Consequently, we can make predictions of the function $f(x)$ based on observations of linear operators, such as derivatives [Solak et al., 2003, Wahlström, 2015] or integrals [Jidling et al., 2018, Law et al., 2018].

We consider setups where the observations y_i are noisy values for line integrals

$$y_i = \|\mathbf{w}_i\| \int_0^1 f(\mathbf{w}_i t + \mathbf{p}_i) dt + \epsilon_i,$$

where \mathbf{p}_i is the start point of the line and \mathbf{w}_i defines its direction and length, so that the set of points along the line is provided by $\mathbf{p}_i + t\mathbf{w}_i$ for $t \in [0, 1]$. Throughout the work we assume $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ to retain conjugacy, and denote by l_i the line corresponding to the observation.

The covariance between a line l_i and a point x_i^* is

$$k(l_i, x_i^*) = \|\mathbf{w}_i\| \int_0^1 k(\mathbf{w}_i t + \mathbf{p}_i, x_i^*) dt, \quad (2)$$

and the covariance between two lines l_i and l_j is

$$k(l_i, l_j) = \|\mathbf{w}_i\| \|\mathbf{w}_j\| \int_0^1 \int_0^1 k(\mathbf{w}_i t + \mathbf{p}_i, \mathbf{w}_j s + \mathbf{p}_j) dt ds. \quad (3)$$

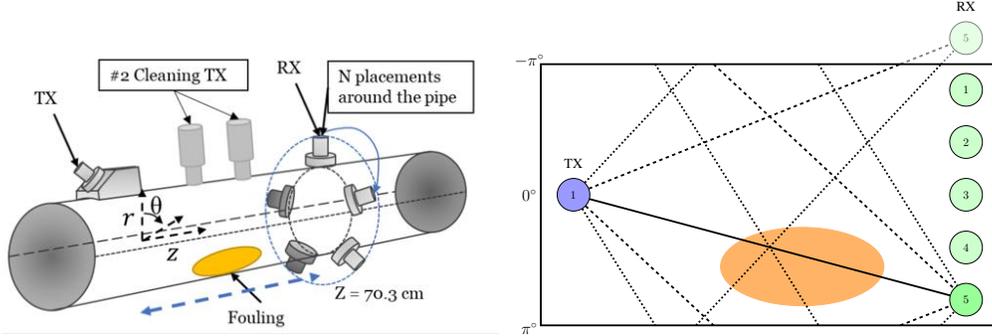


Figure 2: Illustration of the measurement setup of the ultrasonic fouling detection task (**left**, with additional high-power transducers attached for eventual cleaning), and the corresponding line integral GP formulation operating on flattened surface of the pipe (**right**). Ultrasonic waves are transmitted from, here, a single transducer on the left and the arrival times of the waveform is recorded at receiving sensors, so that a single line integral observation is obtained for each possible helical path along the surface, here shown up to 2nd order (solid: direct, dashed: 1st order, dotted: 2n order) for one receiver. A virtual copy of the 5th sensor is drawn on a virtual continuation of the surface to illustrate how the helical paths are formed. The task is to select a small number of sensor locations that allow accurate estimation of the location and thickness of potential fouling. The left figure is reproduced from [Sillanpää et al., 2019] with permission.

Given these quantities, the posterior GP can be computed by (1) simply using l_i in place of x_i . However, evaluating these integrals is challenging. For arbitrary kernels pure numerical integration can be used, but for (3) it has quadratic complexity in terms of the number of evaluation points L along the line [O’Callaghan and Ramos, 2011]. For special cases more efficient strategies are possible: Hendriks et al. [2018] provides algorithm with linear complexity in L for the RBF kernel by analytical integration of the inner integral, and in the Supplement we derive fully analytic expressions for rational quadratic kernel with $\alpha = 2$, useful for cases that require extremely fast computation. For sufficiently smooth stationary kernels a practical alternative to direct evaluation is using a finite spectral approximation on a Hilbert space [Solin and Särkkä, 2014, Jidling et al., 2017].

2.2 SENSOR PLACEMENT FOR GP

The question of sensor placement for spatial functions is widely studied for direct point observations [Krause et al., 2008, Krause and Guestrin, 2007, Zhu and Stein, 2006, Seo et al., 2000, Garnett et al., 2010]. The sensor locations can be selected before making any measurements [Krause et al., 2008], or by sequentially selecting the next location based on observations for previously selected sensors [Krause and Guestrin, 2007]. We focus on the former case because for most physical applications setting up the sensors is time-consuming and needs to be carried out in advance, but note that the strategies presented here could be extended for sequential procedures.

Next, we outline the research on model-based methods

and geometric approaches for point observations, as necessary for understanding the rest of the paper.

2.2.1 Model-based approaches

An optimal sensor configuration maximizes some measure of information content. For most measures solving for optimal combination is NP-complete [Ko et al., 1995, Krause et al., 2008], and practical algorithms are typically greedy, selecting one sensor at a time from a finite subset of possible locations \mathcal{V} .

Various selection criteria have been presented. The entropy criterion [Cressie, 1991, Shewry and Wynn, 1987] measures the conditional entropy of a set of sensors \mathcal{A} for the unobserved locations $\mathcal{V} \setminus \mathcal{A}$, providing sensors \mathcal{A} that are most uncertain about each other:

$$\begin{aligned} \mathcal{A}^* &= \operatorname{argmin}_{\mathcal{A} \subset \mathcal{V}: |\mathcal{A}|=N} H(\mathcal{V} \setminus \mathcal{A} | \mathcal{A}) \\ &= \operatorname{argmax}_{\mathcal{A} \subset \mathcal{V}: |\mathcal{A}|=N} H(\mathcal{A}) \end{aligned}$$

In practice, the sensor locations y are added sequentially to maximize the increase in conditional entropy:

$$y^* = \operatorname{argmax}_y H(y | \mathcal{A}) = \operatorname{argmax}_y \frac{1}{2} \log(2\pi e \sigma_y^2 | \mathcal{A}). \quad (4)$$

Algorithms maximizing mutual information $I(\cdot, \cdot)$ (MI) [Caselton and Zidek, 1984] search for sensors \mathcal{A} that are most informative about the unsensed locations $\mathcal{V} \setminus \mathcal{A}$:

$$\mathcal{A}^* = \operatorname{argmax} I(\mathcal{A}; \mathcal{V} \setminus \mathcal{A}).$$

In practice, we solve for [Krause et al., 2008]:

$$y^* = \operatorname{argmax}_y H(y|\mathcal{A}) - H(y|\mathcal{V} \setminus (\mathcal{A} \cup y)) \propto \frac{\sigma_{y|\mathcal{A}}^2}{\sigma_{y|\mathcal{V} \setminus (\mathcal{A} \cup y)}^2}. \quad (5)$$

Entropy only considers the selected locations, and thus tends to place sensors far from each other (often along the borders of the area of interest) [Ramakrishnan et al., 2005]. MI often leads to higher prediction accuracies, but also requires more computation and becomes unfeasible with large sets of possible locations, although computational tricks like lazy evaluation with priority queues can help [Krause et al., 2008].

2.2.2 Geometric approaches

Geometric approaches use the geometric properties of the measurement space and heuristics characterizing the receptive field to select the sensors, and provide arrangements independent of the modeling approach. For example, González-Banos [2001] assumed every sensor can sense perfectly within a fixed radius while providing no information outside of it, leading to art-gallery optimization problem of covering the area with as few sensors as possible. For such formulations there often exist classical geometric analysis and algorithms, here the disk model of Kershner [1939].

3 OPTIMIZING LINE SENSORS

In this section we discuss aspects that need to be accounted for with integral observations. Eventually we will be able to utilise much of the mathematical machinery already developed for point-observation sensors, but the details depend on the sensing configuration and geometry, characterized by the following aspects illustrated in Figure 3: (a) The set of possible transmitter and receiver locations, provided either as discrete sets (\mathcal{V}_t and \mathcal{V}_r) or continuous areas (\mathcal{T}_t and \mathcal{T}_r). (b) The type of sensors; see Section 3.1 for details. (c) The area of interest \mathcal{I} , typically a subset of some simple Euclidean geometry, for which we want to estimate the function.

The solution also depends on choices such as the optimization principle (model-based or model-independent) and its details (e.g. information measure and choice of kernel for model-based). We start with the general aspects, and then proceed to elaborate the possibilities of how the problem can be solved in Sections 4 and 5.

3.1 SENSING CONFIGURATION

Each point-observation sensor provides one observation. For line integral observations, instead, each measure-

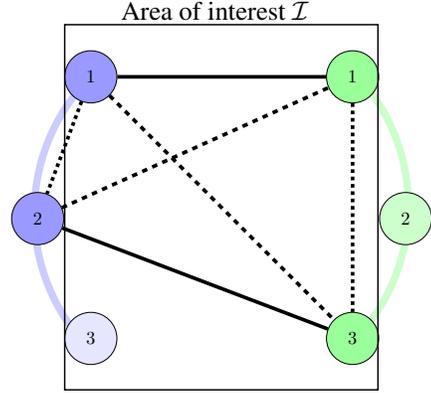


Figure 3: The task is to select small number of transmitters (blue) and receivers (green), amongst discrete sets (nodes) or along a continuous curve (colored line), with possible locations either outside or inside the area of interest \mathcal{I} . The solid lines indicate observations induced by the selected sensors (darker color) for the *Paired Sensors* type. The dashed lines indicate additional observations obtained if using *Separate Transmitters and Receivers*, and the dotted lines further observations recorded by *Universal Sensors*.

ment is induced by a pair of sensors, one transmitting the signal and the other receiving it. The relationship between sensors and measurements depends on the specific sensing setup, with three distinct alternatives as illustrated in Figure 3:

1. **Paired Sensors:** Each receiver listens to only one transmitter. Each observation matches a single pair and hence $2N$ sensors induce $M = N$ observations.
2. **Separate Transmitters and Receivers:** The sensors are split into two groups and each receiver listens to all transmitters. For N_t and N_r sensors in the two groups, we get $M = N_t N_r$ observations.
3. **Universal Sensors:** We have a collection of N sensors, each acting in turn both as transmitter and receiver, listening to all other sensors. This results in a total of $M = N(N - 1)$ observations.

The first case applies e.g. to cases where physically moving an object from one location is required for making a single observation. For example, to study the friction of a surface we might slide a disk across the surface, and separate measurement is required for each observation. This scenario is analogous to the point-observation case, since each choice (of two sensors) independently induces a measurement. Consequently, classical algorithms can be used as such, by simply making choices amongst the tuples of transmitters and receivers.

Most sensing systems based on wave propagation map to the second case. An example most readers can relate to would be a set of microphones (receivers) recording audio signals originating from a set of loudspeakers (transmitters). Every microphone receives all signals (but at different times and amplitudes), and the loudspeakers can be scheduled so that only one of them sends a signal at any given time. The third case generalizes the second one, by using sensors that can both transmit and receive a signal, e.g. because they consist of multiple physical devices (microphone and loudspeaker) or they can inherently perform both tasks (e.g. human observer). Again the signals can typically be encoded in time so that all measurements corresponding to one transmitter can be recorded simultaneously, without inference from others.

For the latter two cases, the optimization problem does not reduce to that of point-observation sensors. The optimization is carried out in the space of sensors, but the evaluation depends on the measurements that depends on multiple sensors. That is, we are optimizing for sensors via a set of measurements induced by them.

3.2 LINE ARRANGEMENT

For analysing the set of lines induced by the sensors, the geometric concept of *line arrangements* is useful. Instead of a collection of start- and endpoints, any set of line segments can equivalently be represented by the set of vertices, edges and faces formed by the geometric pattern, a partitioning of the space, they induce [Agarwal and Sharir, 2010]. Here vertex refers to an intersection of two lines, an edge is a line segment between two vertices, and a face is a polygon induced by edges and vertices. The arrangement can be constructed, starting with the sensor locations, in $\mathcal{O}(M^2)$ time for M lines using various algorithms, such as topological sweeping [Edelsbrunner and Guibas, 1989] that constructs the arrangement one line at a time.

The arrangement provides e.g. the locations of line intersections and the number and length distribution of edges, which can be useful: Intuitively, covering the area with many lines, vertices and short edges could be good. The computational geometry literature also has pre-computed arrangements (or families of arrangements) for example with maximal number of triangle faces [Grunbaum, 2009] that can be used. In other words, line arrangements provide computationally feasible basis for determining sensor locations based on geometric arguments.

3.3 AREA OF INTEREST

The optimal placement is naturally relative to the area of interest \mathcal{I} , in our case some subset of the 2D space,

and its geometry. Even though we limit the analysis here to 2D surfaces, they may be wrapped in 3D spaces (e.g. surface of an object), which influences the line arrangement induced by the sensors. The set of possible sensor locations \mathcal{V} (or \mathcal{T}) can be outside or inside the area; the case of sensors outside the area is more common in physical sensing configurations, but all of the algorithms presented here are applicable also for cases where (some of) the sensor locations can be within the area.

An important aspect regarding \mathcal{I} concerns the assumed support \mathcal{S} of the function $f(x)$, the area for which $f(x)$ can be non-zero. The case of $\mathcal{S} \in \mathcal{I}$ is easy, requiring sensor arrangement that covers $\mathcal{S} \cap \mathcal{I}$. For cases where $f(x)$ can be non-zero outside \mathcal{I} it is important to remember that the sensors correspond to integral over the whole line l_i , and hence sensors outside \mathcal{I} are influenced also by the points for which $\mathbf{p}_i + t\mathbf{w}_i \in \mathcal{S}$ but $\mathbf{p}_i + t\mathbf{w}_i \notin \mathcal{I}$. Consequently, we need to estimate $f(x)$ also outside \mathcal{I} to make sure we can estimate it within it.

4 MODEL-BASED APPROACHES

Section 2.2.1 reviewed the greedy optimization for point sensors based on entropy and MI criteria. For *Paired Sensors* the equations (4) and (5) can be used as is, by simple re-formulation of the notation. Now the set of all possible sensors \mathcal{V} is formed by a cartesian product of \mathcal{V}_t and \mathcal{V}_r – the sets of possible locations for transmitters and receivers – and we denote the line induced by any particular choice of $s_1 \in \mathcal{V}_t$ and $s_2 \in \mathcal{V}_r$ by l . We denote by \mathcal{B} all the lines induced by already selected sensors, and the next optimal sensor pair can be solved by

$$(s_1, s_2)^* = l^* = \operatorname{argmax}_l H(l|\mathcal{B})$$

in the case of entropy criterion and analogously for MI.

For the *Universal Sensors* case every new sensor s induces M' new line observations $l_m^s \in \mathcal{J}^s$, one for all other sensors (*separate transmitters and receivers* is special case of this). We can compute the uncertainty reduction for each of these independently, but in order to select the best new sensor we need to summarize the information provided by all M' new observations. We consider three intuitive alternatives:

$$\begin{aligned} s_a^* &= \operatorname{argmax}_s \operatorname{argmax}_l H(l_m^s|\mathcal{B}), \\ s_b^* &= \operatorname{argmax}_s \operatorname{argmin}_l H(l_m^s|\mathcal{B}), \\ s_c^* &= \operatorname{argmax}_s \frac{1}{M'} \sum_m^{M'} H(l_m^s|\mathcal{B}). \end{aligned}$$

The first one favors sensors that induce one new observation we are maximally uncertain of, the middle one favors sensors for which all new observations are at least

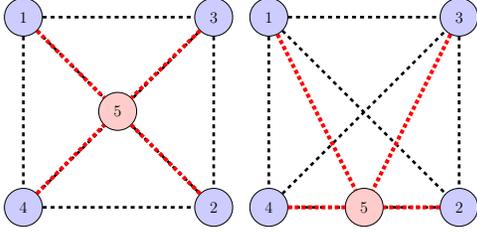


Figure 4: Example of different summaries resulting in different choices. In the case of universal sensors and square area, the 5th sensor is placed differently when we use minimum as the summary (left) compared to mean and maximum (right).

somewhat uncertain, and the last one makes a compromise by simply taking the mean. As these optimize for different things, different summaries might work better in different situations. Figures 4 and 9 show examples where the summary strategy makes a difference.

The above formulas are for the entropy criterion, but the MI criterion is completely analogous, replacing $H(l^s|\mathcal{B})$ with $H(l_m^s|\mathcal{B}) - H(l_m^s|\mathcal{P}\setminus(\mathcal{B}\cup\mathcal{J}_m^s))$, where \mathcal{P} denotes the set of all possible lines – the need to compress M' new observations into a single quantity remains. However, the computational complexity of MI is much higher than the entropy as it requires comparisons against all possible line induced by the allowed sensor locations.

5 GEOMETRIC APPROACHES

Geometric approaches for line integral sensor placement build on line arrangements explained in Section 3.2. The idea is to find such a sensor set that the arrangement spans \mathcal{I} well, at least by edges but ideally also by vertices that help in localizing the function along the lines.

Next, we describe a particular geometric optimization algorithm designed for cases where the sensors are to be placed on one-dimensional surfaces \mathcal{T}_t and \mathcal{T}_r , without discretizing the set of options¹. This is done for two reasons: (1) Model-based solutions necessarily require discrete locations and we want to demonstrate sensor optimization for a scenario not supported by them, and (2) optimizing over discrete locations reduces to binary programming and is easy to implement (but slow to optimize due to exponential complexity) with generic solvers.

For one-dimensional optimization surface \mathcal{T} we can formulate a continuous optimization problem: All locations are parameterized by a real-valued scalar θ_i that indicates the distance along the curve from some starting point.

¹We present the details for *Universal Sensors* with $\mathcal{T}_t = \mathcal{T}_r$; the other cases are analogous but require two parameter vectors.

For convenience of notation, we assume the length of the curve to be 1, such that the parameters all fall into $[0, 1]$. A collection of N sensors can then be parameterized in identifiable manner using unconstrained parameter vector $\phi \in \mathbb{R}^{N+1}$ representing the gaps between ordered sensors, such that

$$\theta_i = \frac{\sum_{j \leq i} e^{\phi_j}}{\sum_{j=1}^{N+1} e^{\phi_j}} \quad \text{for } i \in 1, \dots, N. \quad (6)$$

The last element in ϕ is required since there are $N + 1$ gaps between N sensors and the two end points.

We can now optimize for arbitrary functions $d(\cdot)$ of the line arrangement induced by the placements. For differentiable functions (e.g. distances of points and lines) we can use gradient-based optimization with automatic differentiation, but for more general cases (functions involving vertices) we resort to black-box global optimization methods, specifically, Bayesian optimization (BO) [Pelikan et al., 1999] that has previously been used for sensor set optimization by Garnett et al. [2010]. Since box constraints are easy in BO, we use $\phi_i \in [0, 1]$ and drop the exponent in (6).

Besides directly optimizing for the objective, one can also build regular grid arrangements based on intuitive reasoning or predetermined line arrangements with desirable properties [Grunbaum, 2009]. One interesting option for the case \mathcal{T} being a circle is to use regular n -gons that place the N sensors evenly. This reduces the number of parameters to be determined to one, the orientation of the n -gon, resulting in extremely fast optimization that still finds good sensor configurations (Section 6.1.1).

6 EXPERIMENTS

We conduct a range of experiments to characterize the differences of the algorithms and illustrate the value of sensor selection. We start with artificial data examples for a simple geometry, and then apply the methods for selecting sensor locations for two practical applications of CT scans and ultrasonic fouling detection.

6.1 ARTIFICIAL DATA

We compare both model-based and geometric algorithms for a simple geometry where \mathcal{I} is a square with $0 \leq x, y \leq 5$ and *Universal Sensors* are used. We consider two separate sensor placement settings: (a) the space is discretized into a 5x5 grid of evenly spaced points, allowing sensors to be placed at 25 different locations inside the area. (b) the sensors are placed on a smallest circle (\mathcal{T}) surrounding the area of interest, which is discretized into 48 locations for model-based algorithms. We use

the rational quadratic kernel amenable to closed-form integrals (Supplement), with lengthscale = 1, $\sigma_f^2 = 1$, and the noise variance $\sigma_{noise}^2 = 1$. We present here the results for estimating one particular $f(x)$, measuring the quality by root mean squared error (RMSE); see Supplement for results on multiple functions.

6.1.1 Model-based Sensor Optimization

Figure 5 (left) compares the entropy and MI criteria using different summary strategies (*mean, min, max*) in setting (a). Here entropy criterion, which prefers long lines, is more accurate, and the different summary strategies provide very similar results. We also compared the two strategies (using *mean* summary) against simple baselines of random choices (averaged over three runs) and placing the sensors according to the same criteria but evaluated for point sensors (Fig. 5 (middle)). It clearly pays off to explicitly search for sensors good for line integral observations, instead of simpler heuristics.

6.1.2 Geometric Sensor Optimization

We demonstrate two geometric algorithms for setting (b): (1) directly optimizing all the locations using the parametrization proposed in Section 5 (denoted by *Geometric*), and (2) placing sensors evenly (n-gon), while only optimizing for the rotation (denoted by *Optimized*). We experiment on two different heuristic loss functions $d(\cdot)$: the largest euclidean distance to the closest line, and the largest distance to the closest vertex (intersection of two lines), to illustrate possible arrangement-based objectives. Figure 6 gives the surrogate losses as a function of sensors (N), and illustrates how the two algorithms differ in their placement. Both losses provide similar placements, but the vertex one is slower since for M lines there are $\mathcal{O}(M^2)$ vertices.

Finally, Figure 5 (right) compares the results against model-based solution, applying the entropy criterion for discretized locations along the circle (*Entropy (circle)*) as well as to the sensor configuration (a) that allows placement also within the area (*Entropy*). Allowing sensors everywhere in the area results in highest accuracy, as expected, but the geometric algorithms outperform the model-based ones in setting (b) due to the discretization of possible locations for the latter.

6.2 SIMULATED CT SCAN

In X-ray computed tomography (CT) X-rays are sent through an object, and the measured observations represent the attenuation of the x-rays. Purisha et al. [2019] demonstrated that line-integral GPs improve on traditional methods such as filtered back projection (FBP)

[Purisha et al., 2019], for fixed sensor configuration of nine imaging angles (see Supplement for details).

Sensor placement here corresponds to optimizing the angles from which the beams are sent. We let the model-based algorithm with Entropy (mean) select nine sensors amongst a set of 18 evenly based possible angles, and show that the inference result matches in accuracy (measured by peak-signal-to-noise ratio, PSNR, and relative error, RE) the one computed for commonly used configurations for CT scan (Figure 7). That is, we demonstrated we can find essentially an optimal sensor configuration without requiring any knowledge of the problem domain.

6.3 ULTRASOUND PROPAGATION

We demonstrate the techniques in a cleantech application. The task is to non-invasively estimate a spatial map characterizing the thickness of fouling inside a closed structure, here a pipe, used for determining when the structure has to be cleaned (Fig. 2). We have previously shown the feasibility of line-integral GPs for this task on real physical sensing configuration for one manually chosen sensor setting [Sillanpää et al., 2019].

A transducer attached to the outer surface of the pipe transmits an ultrasonic Lamb wave, which travels along the pipe and can be measured at any location with another transducer. The group velocity of the wave packet for a given material and geometry is known, and hence time of arrival y_i^0 at any location for a clean pipe can be computed numerically (see, e.g., Lowe [1995]). Any fouling at the surface of the pipe changes the velocity at that location. If we denote by y_i^1 the arrival time corresponding to a wave traversing along a path l_i , then $y_i = y_i^1 - y_i^0$ corresponds to a line integral of the assumed fouling $f(x)$ with noise induced e.g. by the arrival time detection and physical coupling of the sensors². For details on ultrasound propagation on layered metal structures, see e.g. Rose [2014] and Brekhovskikh [2012].

We want to minimize the number of transducers, to reduce the cost and manual labor required for coupling the transducers. The configuration corresponds to *Separate transmitters and receivers*, but with additional observations induced due to the cylinder geometry – each receiver can record the arrival times for all helical paths up to some reasonable order (for high-order paths detecting arrival times becomes difficult).

We use here simplified simulated data that directly models the velocity changes induced by fouling, since collecting physical measurements or full simulations of

²The sensor records waveform, which can be converted into arrival time by detecting peaks of the envelope of the signal.

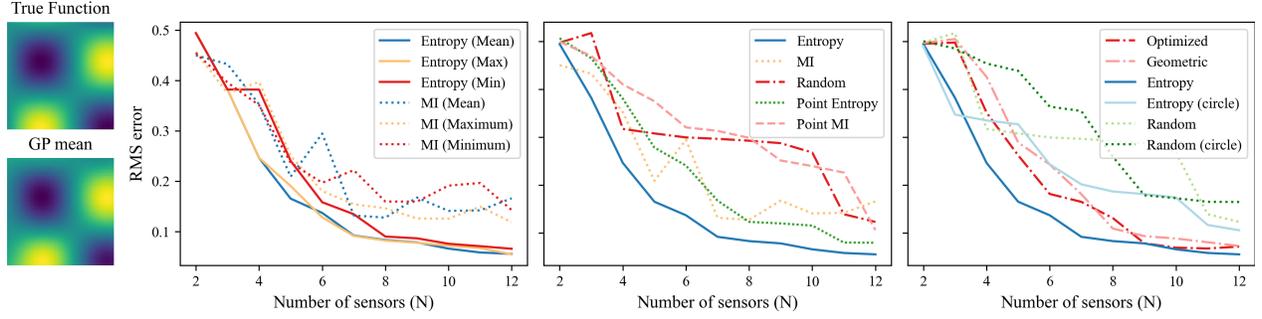


Figure 5: Estimation errors for different sensor placement strategies for one example function (depicting also the GP mean for 12-sensor solution). **Left:** Comparison of the summary strategies for entropy and MI. **Middle:** Comparison of entropy and MI for baseline methods. **Right:** Comparison of model-based and geometric approaches.

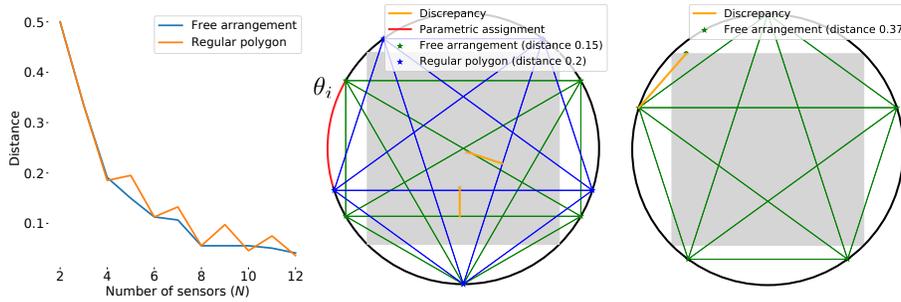


Figure 6: Illustration of geometric arrangements for setup where the gray square is the area of interest \mathcal{I} and sensors are placed on the black circle \mathcal{T} . **Left:** Optimizing the rotation of a n -gon (orange) matches the solution of freely optimized sensor placements (blue) for even number of sensors, but for odd numbers we can find better placements by explicitly optimizing for each sensor. **Middle and Right:** The optimal sensor arrangements for $N = 5$ for two optimization criteria; the first covers the area with lines, whereas the latter does it with vertices. For the latter case the two methods give identical results – only one is shown.

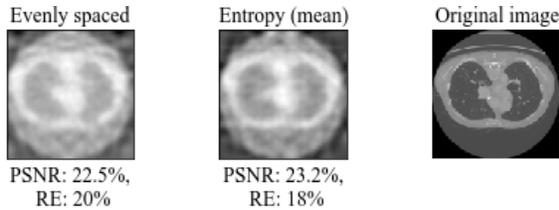


Figure 7: Means of the GP fit for nine sensor locations chosen using entropy (middle) and the even spacing typically used for CT scans (left). The true image (right) is reconstructed with comparable accuracy, with the proposed method here having slight edge.

the acoustic field for extensive set of possible locations would be infeasible. We generated six functions simulating possible fouling in a pipe with radius 35mm and length 300mm. We use the Matern32 kernel ($l = 15, \sigma_f^2 = 0.5, \sigma_{noise}^2 = 0.001$) with numerical integration to allow evaluation of the higher order paths; we use paths up to 2nd order.

For model-based sensor placement we allow for 20 possible transducer locations at five angular placements on four circles around the pipe, split so that two circles are dedicated for transmitters and the other for receivers, and consider two sensing configurations: In one we only use one transmitter, corresponding to the setup of Sillanpää et al. [2019], whereas in the other the algorithm can always decide whether the next sensor is transmitter or receiver. Figure 8 shows the RMS error for both settings, with intuitive results. Increasing the number of sensors always helps, and allowing for multiple transmitters provides significant advantage. It turns out to be optimal to maintain roughly balanced number of transmitters. Figure 9 shows the optimal placements for six transducers for different algorithms, including the result of geometric placement when allowing free placement of the sensors along two belts around the pipe.

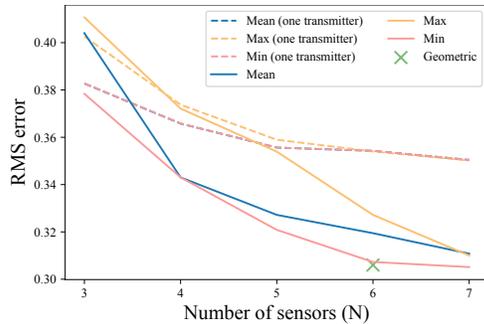


Figure 8: Mean function estimation accuracy over six functions for alternative sensor configurations optimized with the entropy criterion. The dashed lines assume only one transmitter, whereas the solid lines can freely select how many of the sensors are transmitters and receivers. For $N = 6$ the geometric algorithm can match the best model-based result.

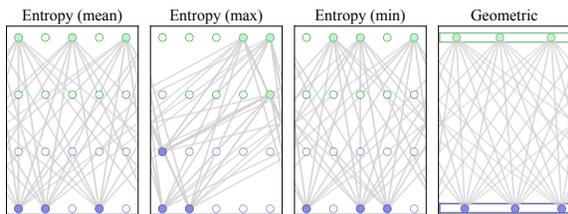


Figure 9: Optimal placements for six sensors for different algorithms. Entropy (max) selects also sensors within the area, while others favor sensors at the border.

7 DISCUSSION

Even though posterior inference of Gaussian processes for integral observations is relatively straightforward, they have been used only in few applications [O’Callaghan and Ramos, 2011, Purisha et al., 2019, Tanaka et al., 2019]. For increasing use of the promising machinery in empirical sciences, an important practical question concerns the choice of sensors used for gathering the information. This work is the first one considering this problem, characterizing the various aspects that need to be accounted for while selecting the sensors, including identification of possible sensor types and their relationship to the integral observations for a given geometry. We covered the concept of line arrangements [Agarwal and Sharir, 2010] as basis for model-independent selection strategies, and extended classical model-based strategies [Krause et al., 2008] for line integral sensors. For model-based strategies we demonstrated how the entropy criterion favors longer lines than mutual information, and in our experiments typically resulted in better configurations while also having clear speed advantage. We also showed qualitative differences

in final configuration corresponding to how the information for different observations induced by the sensor configuration is combined, but cannot draw clear conclusions on which strategy is to be preferred since it depends on the geometry and assumptions on the underlying function.

We illustrated how both approaches find good sensor configurations for example 2D geometries, and demonstrated their use in two practical applications. For CT scanning we found a configuration providing equally good reconstructions as the commonly used one (expecting to outperform it would be unreasonable), and for ultrasonic fouling detection we demonstrated how using balanced number of transmitters and receivers helps.

In this work we focused on pre-selection of sensors independent of the measurement data, matching typical application needs. However, the model-based techniques could be extended for active selection. In addition, existing knowledge of the physical properties are easy to account for similar to the point-observation sensors, such as different costs depending on the location or the type of the sensor [Krause et al., 2008] or weighting of lines based on measurement quality.

Acknowledgements

This work was supported by the Academy of Finland Flagship programme: Finnish Center for Artificial Intelligence, FCAI, and grant 324852.

References

- P. K. Agarwal and M. Sharir. Arrangements and their applications. In *Handbook of Computational Geometry*. Elsevier, 2010.
- L. Brekhovskikh. *Waves in layered media*, volume 16. Elsevier, 2012.
- W. F. Caselton and J. V. Zidek. Optimal monitoring network designs. *Statistics & Probability Letters*, 2(4), 1984.
- N. Cressie. *Statistics for spatial data*. Wiley, 1991.
- H. Edelsbrunner and L. Guibas. Topologically sweeping an arrangement. *Journal of Computer and System Sciences*, 38(1), 1989.
- R. Garnett, M. A. Osborne, and S. J. Roberts. Bayesian optimization for sensor set selection. In *Proc. of the 9th ACM/IEEE international conference on information processing in sensor networks*, 2010.
- H. González-Banos. A randomized art-gallery algorithm for sensor placement. In *Proc. of the 17th annual symposium on Computational geometry*, 2001.

- B. Grunbaum. A catalogue of simplicial arrangements in the real projective plane. *Ars Mathematica Contempronea*, 2, 2009.
- O. Hamelijnck, T. Damoulas, K. Wang, and M. Girolami. Multi-resolution multi-task Gaussian processes. In *NeurIPS*, 2019.
- J. N. Hendriks, C. Jidling, A. Wills, and T. B. Schön. Evaluating the squared-exponential covariance function in Gaussian processes with integral observations. *arXiv 1812.07319*, 2018.
- C. Jidling, N. Wahlström, A. Wills, and T. B. Schön. Linearly constrained Gaussian processes. In *NeurIPS*, 2017.
- C. Jidling, J. Hendriks, N. Wahlström, A. Gregg, T. B. Schön, C. Wensrich, and A. Wills. Probabilistic modelling and reconstruction of strain. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 2018.
- C. Jidling, J. Hendriks, T. B. Schön, and A. Wills. Deep kernel learning for integral measurements. *arXiv 1909.01844*, 2019.
- R. Kershner. The number of circles covering a set. *American Journal of mathematics*, 61(3):665–671, 1939.
- C.-W. Ko, J. Lee, and M. Queyranne. An exact algorithm for maximum entropy sampling. *Operations Research*, 43(4):684–691, 1995.
- A. Krause and C. Guestrin. Nonmyopic active learning of Gaussian processes: an exploration-exploitation approach. In *ICML*, 2007.
- A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9, 2008.
- H. C. Law, D. Sejdinovic, E. Cameron, T. Lucas, S. Flaxman, K. Battle, and K. Fukumizu. Variational learning on aggregate outputs with Gaussian processes. In *NeurIPS*, 2018.
- M. Lowe. Matrix techniques for modeling ultrasonic waves in multilayered media. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 42(4):525–542, 1995.
- Y. Lu and J. E. Michaels. Feature extraction and sensor fusion for ultrasonic structural health monitoring under changing environmental conditions. *IEEE Sensors Journal*, 9(11):1462–1471, 2009.
- S. T. O’Callaghan and F. T. Ramos. Continuous occupancy mapping with integral kernels. In *AAAI*, 2011.
- M. Pelikan, D. E. Goldberg, and E. Cantu-Paz. Boa: The Bayesian optimization algorithm. In *Proc. of the 1st Annual Conference on Genetic and Evolutionary Computation-Volume 1*, 1999.
- Z. Purisha, C. Jidling, N. Wahlström, T. Schön, and S. Särkkä. Probabilistic approach to limited-data computed tomography reconstruction. *Inverse Problems*, 2019.
- N. Ramakrishnan, C. Bailey-Kellogg, S. Tadepalli, and V. N. Pandey. Gaussian processes for active data mining of spatial aggregates. In *Proc. of the 2005 SIAM International Conference on Data Mining*, 2005.
- C. E. Rasmussen and C. K. Williams. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.
- J. L. Rose. *Ultrasonic Guided Waves in Solid Media*. Cambridge University Press, 2014.
- A. Sakiyama, Y. Tanaka, T. Tanaka, and A. Ortega. Efficient sensor position selection using graph signal sampling theory. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016.
- S. Seo, M. Wallat, T. Graepel, and K. Obermayer. Gaussian process regression: Active data selection and test point rejection. In *Mustererkennung 2000*, pages 27–34. Springer, 2000.
- M. C. Shewry and H. P. Wynn. Maximum entropy sampling. *Journal of applied statistics*, 14(2), 1987.
- T. Sillanpää, T. Rauhala, J. Mäkinen, C. Rajani, K. Longi, A. Klami, A. Salmi, and E. Hægström. Ultrasonic fouling detector powered by machine learning. In *IEEE International Ultrasonics Symposium*, 2019.
- E. Solak, R. Murray-Smith, W. E. Leithead, D. J. Leith, and C. E. Rasmussen. Derivative observations in Gaussian process models of dynamic systems. In *NeurIPS*, 2003.
- A. Solin and S. Särkkä. Hilbert space methods for reduced-rank Gaussian process regression. *arXiv 1401.5508*, Jan. 2014.
- Y. Tanaka, T. Tanaka, T. Iwata, T. Kurashima, M. Okawa, Y. Akagi, and H. Toda. Spatially aggregated Gaussian processes with multivariate areal outputs. In *NeurIPS*, 2019.
- N. Wahlström. *Modeling of magnetic fields and extended objects for localization applications*. PhD thesis, Linköping University Electronic Press, 2015.
- Z. Zhu and M. L. Stein. Spatial sampling design for prediction with estimated parameters. *Journal of agricultural, biological, and environmental statistics*, 11(1): 24, 2006.

Supplementary material for *Sensor Placement for Gaussian Processes with Integral Observations*

1 EXPERIMENTS

In this section we provide more detailed descriptions of some of the experiments presented in the paper as well as additional results.

1.1 ARTIFICIAL DATA

We perform experiments on nine additional functions to show that the results are consistent to the ones presented in the paper. The results replicating comparisons between different model-based methods are presented in Figure 1, and results replicating experiments for geometric approaches are presented in Figure 2.

The functions evaluated are

$$\begin{aligned}
 f_1(x, y) &= -\frac{1}{2} |\sin(x) \cos(y) e^{|1 - \sqrt{x^2 + y^2} / \pi|}| \\
 f_2(x, y) &= \frac{1}{600} (x + y^2 - 7)^2 + \frac{1}{600} (x^2 + y - 11)^2 \\
 f_3(x, y) &= \frac{1}{50} (\sin^2 3\pi x + (x - 1)^2 (1 + \sin^2 2\pi y) + \\
 &\quad (y - 1)^2 (1 + \sin^2 2\pi y)) \\
 f_4(x, y) &= 0.26(x^2 + y^2) - 0.48xy \\
 f_5(x, y) &= \frac{1}{120} (x + 2y - 7)^2 + \frac{1}{120} (2x + y - 5)^2 \\
 f_6(x, y) &= \sin(x/2)^2 \sin(y/2)^2 \\
 f_7(x, y) &= \cos(x/2)^2 \sin(y/2) \\
 f_8(x, y) &= \frac{1}{600} (x - y^2 - 7)^2 + \frac{1}{600} (x^2 - y - 11)^2 \\
 f_9(x, y) &= -20 \exp \left[-0.2 \sqrt{0.5 (x^2 + y^2)} \right] \\
 &\quad - \exp [0.5 (\cos 2\pi x + \cos 2\pi y)] + e + 20 \\
 f_{10}(x, y) &= \sin(x) \cos(y)
 \end{aligned}$$

where f_1 is the Hölder table function, f_4 Matyas function and f_8 is the Ackley function [Ackley, 1987], f_5 is based on the Booth function, f_3 on the Lévi function and f_2 on the Himmelblau’s function [Himmelblau, 1972]. The

observations are line integral values computed for these functions with added Gaussian noise (with 0 mean and 0.01 variance).

1.2 CT SCAN

In this experiment we follow the setup of Purisha et al. [2019]. As ground truth, we use one slice (128×128) of 3D Chest dataset [Matlab], which is shown in Figure 3 (right). We use 9 projections, and for each projection we get 128 lines, yielding a total of 1152 lines. The measurements are computed using the `radon` command in `scikit-image` and by adding Gaussian noise with zero mean and 0.001 variance. We use the Hilbert space approximation of the Matern kernel with $\nu = 1$ and amount of basis functions $m = 10^4$. For more information on GPs for CT scan see Purisha et al. [2019].

For the sensor placement, we use a kernel with lengthscale $l = 8.5$, and $\sigma_f^2 = 0.03$, $\sigma_{noise}^2 = 0.15$. We let the entropy algorithm select nine sensors amongst a set of 18 evenly based possible angles with different summary strategies. Filtered back projection (FBP) baseline is computed using the `iradon` command in `scikit-image`.

We show that the result corresponds roughly with the commonly used configurations for CT scan (even placement) (Figure 3). In fact, in this geometry even random placement performs fairly well when considering the measurements, however, the picture is clearly blurrier. The angles we retrieve with for example using entropy (mean) are [0, 90, 40, 130, 70, 160, 20, 110, 60].

1.3 ULTRASOUND PROPAGATION

In this experiment, the area of interest (\mathcal{I}) is a surface of a pipe with length 300mm and radius 35mm. We use the Matern32 kernel ($l = 15$, $\sigma_f^2 = 1$, $\sigma_{noise}^2 = 0.1$) with numerical integration to allow incorporating the second and/or third order flight paths. In practice, in the numeri-

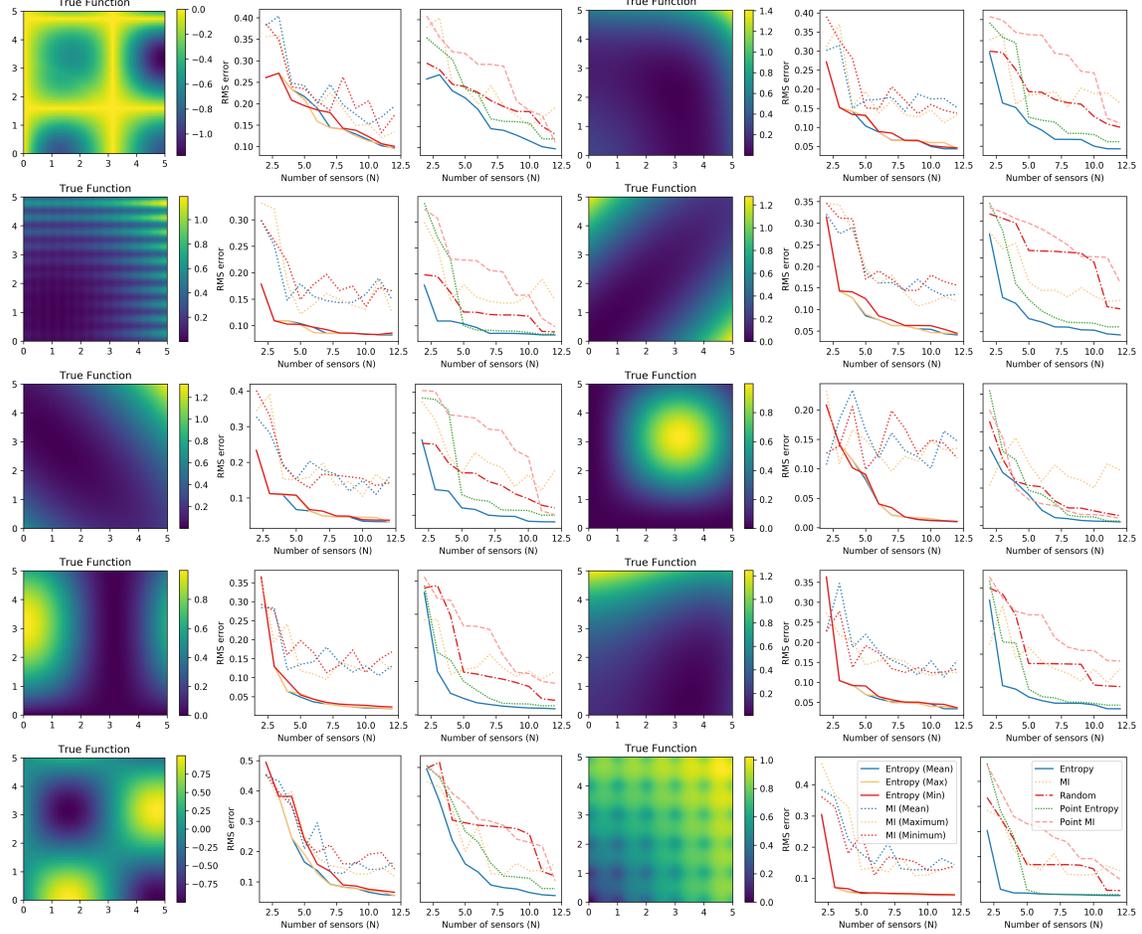


Figure 1: RMS error for different sensor placements for the true function presented in the left. **Middle:** Comparing strategies for computing MI and entropy. **Right:** Comparing MI and entropy to other baseline strategies.

cal integration, we move the points that are outside of the circumference of the pipe $([-35\pi, 35\pi])$ inside this area by adding/subtracting the period length (70π) . Another option would be to have a periodic kernel that takes these second and third order paths into account.

As artificial data we generated functions simulating possible fouling in the pipe. For this purpose, the fouling is represented by 5 relatively smooth continuous functions (scaled to $[0,1]$):

$$\begin{aligned}
 f_1(x, y) &= \cos(x/50)^2 \sin(y/100) \\
 f_2(x, y) &= \sin(x/90)^2 \sin(y/90)^2 \\
 f_3(x, y) &= \sin(x/35)^2 \sin(y/35)^2 \\
 f_4(x, y) &= \cos(x/100)^2 \sin(y/150) \\
 f_5(x, y) &= \sin(x \cdot 0.02857) \cos(y \cdot 0.01) \\
 f_6(x, y) &= \sin(x \cdot 0.02857) \sin(y \cdot 0.02857)
 \end{aligned}$$

To study the behaviour of the entropy criterion, we consider two different cases: 1) we allow transducers to be placed at 20 different locations at five angular placements on four circles around the pipe, so that transmitters are located on the two first circles and receivers on the two last. 2) By placing one transmitter at $[0, 0]$, and allowing the algorithm to place receivers on 20 different locations on two circles around the pipe.

Figure 4 shows the root mean squared error between the mean of a GP fit and the true function for the six function described above for these two different cases. Clearly allowing the algorithm to decide which type of sensors to place is beneficial. We also include the result of geometric placement, where the locations were optimized using the parametrization proposed in Section 5. We allowed 3 transmitters and 3 receivers to be placed on two continuous lines (the \mathcal{T} in this case) at the ends of the pipe.

In addition, we simulate cases where fouling is represented as different sized rectangles placed at different po-

sitions in the pipe. The true functions as well as the GP fit for these with six sensors using the described strategies are presented in Figure 6.

2 ANALYTIC COMPUTATION OF THE KERNEL

We consider setups where the observations \mathbf{y} are noisy values for line integrals

$$\mathbf{y}_i = \|\mathbf{w}_i\| \int_0^1 f(\mathbf{w}_i s + \mathbf{p}_i) ds + \epsilon_i,$$

where \mathbf{p}_i is the start point of the line and \mathbf{w}_i defines its direction and length, so that the set of points along the line is provided by $\mathbf{p}_i + t\mathbf{w}_i$ for $t \in [0, 1]$.

Then, the covariance between an line observation l_i and a test point x^* with a kernel function k is defined as

$$k(l_i, \mathbf{x}^*) = \|\mathbf{w}_i\| \int_0^1 k(\mathbf{w}_i t + \mathbf{p}_i, \mathbf{x}^*) dt, \quad (1)$$

and the covariance between two line observations l_i and l_j is defined as

$$k(l_i, l_j) = \|\mathbf{w}_i\| \|\mathbf{w}_j\| \int_0^1 \int_0^1 k(\mathbf{w}_i t + \mathbf{p}_i, \mathbf{w}_j s + \mathbf{p}_j) dt ds. \quad (2)$$

For a configuration of K observations and L evaluation points for the function, we need to compute integral (1) KL times and the integral (2) $K(K-1)$ times in order to obtain the posterior GP. The core challenge in practical applications of line integral GPs is efficient computation of these. As the line-vs-line case involves a double integral, it is in general considerably more challenging, but for large L also efficient evaluation of the simpler case is important.

Avoiding approximations and numerical integration would be ideal, but no analytic solutions for the line-vs-line case have been presented. The question of whether this can be done depends integrally on the kernel, and there are no clear indicators on when it is possible. In this work we show that for the special case of rational quadratic kernel with $\alpha = 2$ we can do it, and later make remarks on possible other alternatives. Table 1 illustrates the practical effect of this, by comparing the relative computational costs of the proposed analytic computation against semi-analytic [Hendriks et al., 2018] and numeric evaluation [O’Callaghan and Ramos, 2011]. Already for $s = 10$ the analytic solution has 40-fold speedup compared to full numeric integration, and for the more accurate $s = 100$ (which would be needed for accurate integration of e.g. near-parallel lines – see Table 1 in Hendriks et al. [2018]) the analytic integral is also 17 times faster than the semi-analytic solution.

Table 1: Comparison of running times (in milliseconds for each entry of the kernel matrix) for analytic, semi-analytic and fully numeric integration.

	Number	Analytic	Semi-Analytic		Numeric	
			s=10	s=100	s=10	s=100
K_{xx}	$\mathcal{O}(K^2)$	0.07	0.17	1.2	2.78	273.4
K_{xx^*}	$\mathcal{O}(KL)$	0.02	0.02	0.02	0.25	2.45

2.1 RATIONAL QUADRATIC

In this section, we present the full analytical integrations for the rational quadratic kernel (RQ) with α set to be 2:

$$\begin{aligned} k(\mathbf{x}_i, \mathbf{x}_j) &= \sigma^2 \left(1 + \frac{1}{2\alpha} r_{ij}^T \mathbf{V} r_{ij} \right)^{-\alpha} \\ &= \sigma^2 \left(1 + \frac{1}{4} r_{ij}^T \mathbf{V} r_{ij} \right)^{-2}. \end{aligned}$$

2.1.1 Line vs. Point

Following the notation of Hendriks et al. [2018], we denote the vector between an arbitrary point \mathbf{x}^* and a point along the line $\mathbf{p}_i + t\mathbf{w}_i$ by \mathbf{r}_i , which results in

$$\mathbf{r}_i^T \mathbf{V} \mathbf{r}_i = (\mathbf{w}_i + \mathbf{v}_i)^T \mathbf{V} (\mathbf{w}_i + \mathbf{v}_i),$$

where $\mathbf{v}_i = \mathbf{p}_i - \mathbf{x}^*$. This further simplifies into $c_1 t^2 + c_2 t + c_3$, where

$$c_1 = \mathbf{w}_i^T \mathbf{V} \mathbf{w}_i, \quad c_2 = 2\mathbf{w}_i^T \mathbf{V} \mathbf{v}_i, \quad c_3 = \mathbf{v}_i^T \mathbf{V} \mathbf{v}_i.$$

Thus, we are solving the equation

$$K(\mathbf{x}_i, \mathbf{x}_k) = \|\mathbf{w}_i\| \int_0^1 \left(1 + \frac{1}{4} (c_1 t^2 + c_2 t + c_3) \right)^{-2} dt,$$

for which the solution is given by

$$\begin{aligned} K(\mathbf{x}_i, \mathbf{x}_k) &= 16 \|\mathbf{w}_i\| \left(\frac{2c_1 + c_2}{A(c_1 + c_2 + c_3 + 4)} - \frac{c_2}{A(c_3 + 3)} \right. \\ &\quad \left. + \frac{4c_1 \tan^{-1}\left(\frac{2c_1 + c_2}{\sqrt{A}}\right)}{A^{\frac{3}{2}}} - \frac{4c_1 \tan^{-1}\left(\frac{c_2}{\sqrt{A}}\right)}{A^{\frac{3}{2}}} \right), \end{aligned}$$

where $A = 4c_1(c_3 + 4) - c_2^2$

2.1.2 Line vs. Line

Again following the notation of Hendriks et al. [2018], and writing $\mathbf{u}_{ij} = \mathbf{p}_i - \mathbf{p}_j$, we get

$$\begin{aligned} r_{ij}^T \mathbf{V} r_{ij} &= (\mathbf{w}_i t - \mathbf{w}_j s + \mathbf{u}_{ij})^T \mathbf{V} (\mathbf{w}_i t - \mathbf{w}_j s + \mathbf{u}_{ij}) \\ &= a - bs + ct - dst + et^2 + fs^2, \end{aligned}$$

where

$$\begin{aligned} a &= \mathbf{u}_{ij}^T \mathbf{V} \mathbf{u}_{ij}, & b &= 2\mathbf{u}_{ij}^T \mathbf{V} \mathbf{w}_j, & c &= 2\mathbf{u}_{ij}^T \mathbf{V} \mathbf{w}_i \\ d &= 2\mathbf{w}_j^T \mathbf{V} \mathbf{w}_i, & e &= \mathbf{w}_i^T \mathbf{V} \mathbf{w}_i, & f &= \mathbf{w}_j^T \mathbf{V} \mathbf{w}_j. \end{aligned}$$

Then the equation we are solving for is

$$K(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{w}_i\| \|\mathbf{w}_j\| \int_0^1 \int_0^1 \left(1 + \frac{1}{4} r_{ij}^T \mathbf{V} r_{ij}\right)^{-2} dt ds,$$

and the full result of the equation is given by

$$\begin{aligned} K(x_i, x_j) &= 16 \|\mathbf{w}_i\| \|\mathbf{w}_j\| \left(\frac{-B_{12} \tan^{-1}\left(\frac{c-d+2e}{A_{12}}\right) + B_{22} \tan^{-1}\left(\frac{2f-b-d}{A_{22}}\right)}{A_{12} D} + \frac{B_{21} \tan^{-1}\left(\frac{2f-b}{A_{21}}\right)}{A_{22} D} \right. \\ &\quad - \frac{-B_{11} \tan^{-1}\left(\frac{c-d}{A_{11}}\right) + B_{21} \tan^{-1}\left(\frac{2f-b}{A_{21}}\right)}{A_{11} D} \\ &\quad - \frac{-B_{12} \tan^{-1}\left(\frac{c-d}{A_{12}}\right) + B_{22} \tan^{-1}\left(\frac{2f-b}{A_{22}}\right)}{A_{12} D} \\ &\quad \left. + \frac{-B_{11} \tan^{-1}\left(\frac{c}{A_{11}}\right) + B_{22} \tan^{-1}\left(\frac{-b}{A_{21}}\right)}{A_{11} D} + \frac{B_{22} \tan^{-1}\left(\frac{-b}{A_{21}}\right)}{A_{21} D} \right) \end{aligned}$$

where

$$\begin{aligned} A_{11} &= \sqrt{-c^2 + 4e(4+a)} \\ A_{12} &= \sqrt{-c^2 + 2cd - d^2 + 4e(4+a-b+f)} \\ B_{11} &= cd - 2be \\ B_{12} &= B_{11} - d^2 + 4ef \\ A_{21} &= \sqrt{-b^2 + 4f(4+a)} \\ A_{22} &= \sqrt{-b^2 - 2bd - d^2 + 4f(4+a+c+e)} \\ B_{21} &= bd - 2cf \\ B_{22} &= B_{21} + d^2 - 4ef \\ D &= -bcd + (4+a)d^2 + b^2e + (c^2 - 4(4+a)e)f \end{aligned}$$

2.1.3 Lines with the same start point

Due to D becoming zero when the two lines start from the same point, we need to give special consideration to this case. We can then rewrite the equation as follows:

$$\begin{aligned} K(\mathbf{x}_i, \mathbf{x}_j) &= \|\mathbf{w}_i\| \|\mathbf{w}_j\| \int_0^1 \int_0^1 \left(1 + \frac{1}{4} (et^2 - dts + fs^2)\right)^{-2} dt ds \\ &= \frac{\tan^{-1}\left(\frac{2e-d}{A}\right)}{A} + \frac{\tan^{-1}\left(\frac{2f-d}{B}\right)}{B} \\ &\quad - \frac{\tan^{-1}\left(\frac{-d}{A}\right)}{A} - \frac{\tan^{-1}\left(\frac{-d}{B}\right)}{B} \end{aligned}$$

where

$$A = \sqrt{4e(f+4) - d^2} \quad B = \sqrt{4f(e+4) - d^2}$$

2.1.4 Parallel lines

We also need to give special consideration when the two lines are parallel. In this case, we can write $w_i = \gamma w_j$, and rewrite

$$c = 2\gamma b \quad d = 2\gamma f, \quad e = \gamma^2 f$$

The equation we are solving then becomes

$$\begin{aligned} K(\mathbf{x}_i, \mathbf{x}_j) &= \|\mathbf{w}_i\| \|\mathbf{w}_j\| \int_0^1 \int_0^1 \left(1 + \frac{1}{4} (a - bs + 2\gamma bt - 2\gamma f st + \gamma^2 ft^2 + fs^2)\right)^{-2} dt ds \\ &= 32 \|\mathbf{w}_i\| \|\mathbf{w}_j\| \frac{1}{A^{(\frac{3}{2})p}} \left[\right. \\ &\quad 2f \tan^{-1}\left(\frac{B_{11}}{\sqrt{A}}\right) + (b + 2fp) \tan^{-1}\left(-\frac{B_{11}}{\sqrt{A}}\right) \\ &\quad + b \tan^{-1}\left(\frac{-B_{00}}{\sqrt{A}}\right) - 2f \tan^{-1}\left(\frac{B_{10}}{\sqrt{A}}\right) \\ &\quad \left. - b \tan^{-1}\left(\frac{-B_{10}}{\sqrt{A}}\right) - (b + 2fp) \tan^{-1}\left(\frac{-B_{01}}{\sqrt{A}}\right) \right], \end{aligned}$$

where

$$\begin{aligned} A &= -b^2 + 4(4+a)f \\ B_{11} &= b - 2f + 2fp & B_{01} &= b + 2fp \\ B_{10} &= b - 2f & B_{00} &= b \end{aligned}$$

2.2 ON OTHER KERNELS

The above derivations show that RBF kernel is amenable to single integration in analytic fashion, whereas the RQ kernel with $\alpha = 2$ has analytic expression also for the line-vs-line case. Here we briefly discuss the feasibility of analytic integrals for other kernels of interest. We start by noting that similar derivation could be carried out also for e.g. $\alpha = 1$ for the RQ kernel, but doing it for $\alpha > 2$ seems difficult. The case of $\alpha \rightarrow \infty$ corresponds to the RBF kernel and hence has no fully analytic solution for the line-vs-line case.

Regarding other kernels, the first important observation is that both RBF and RQ are parameterized by $\mathbf{r}^T \mathbf{V} \mathbf{r}$, the squared distance. Many of the other frequently used kernels, such as the Matern kernels, the exponential kernel, and the Wendland's piecewise polynomial kernels [Wendland, 2005] are parameterized by $|\mathbf{r}|$ instead. The integrands are then functions of $\sqrt{c_1 s^2 + c_2 s + c_3}$ (for line-vs-point) or $\sqrt{a - bs + ct - dst + et^2 + fs^2}$ (for line-vs-line), where the square root considerably complicates analytic integration.

For the Matern kernel we have not been able to compute even the line-vs-point scenario analytically, and hence

recommend using the spectral approximation by Purisha et al. [2019] and Jidling et al. [2018]. The exponential kernel $e^{-|r|/l}$, in turn, is similar to the RBF kernel: it has closed-form solution for line-vs-point but only allows for semi-analytic solution for the line-vs-line case because of the error function.

Finally, for the low-order instances of the Wendland's piecewise polynomial kernels with compact support [Wendland, 2005], such as $(1 - |r|)_+^{[D/2]+q+1}$, we can compute both the line-vs-point and the line-vs-line cases analytically. However, the expressions are lengthy polynomials, reaching several dozens of terms for the latter case. Combined with the fact that geometric analysis would be needed to determine the area of integration due to the compact support, we do not recommend using analytic integrals; it would likely be less efficient than numerical integration.

References

- D. H. Ackley. A connectionist machine for genetic hill-climbing. In *Kluwer Academic Publishers, Boston MA*, 1987.
- J. N. Hendriks, C. Jidling, A. Wills, and T. B. Schön. Evaluating the squared-exponential covariance function in Gaussian processes with integral observations. *arXiv 1812.07319*, 2018.
- D. Himmelblau. Applied nonlinear programming. In *McGraw-Hill*, 1972.
- C. Jidling, J. Hendriks, N. Wahlström, A. Gregg, T. B. Schön, C. Wensrich, and A. Wills. Probabilistic modelling and reconstruction of strain. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 2018.
- Matlab. Segment lungs from 3-d chest scan. URL <https://se.mathworks.com/help/images/segment-lungs-from-3-d-chest-mri-data.html>.
- S. T. O'Callaghan and F. T. Ramos. Continuous occupancy mapping with integral kernels. In *AAAI*, 2011.
- Z. Purisha, C. Jidling, N. Wahlström, T. Schön, and S. Särkkä. Probabilistic approach to limited-data computed tomography reconstruction. *Inverse Problems*, 2019.
- H. Wendland. *Scattered Data Approximation*. Cambridge University Press, 2005.

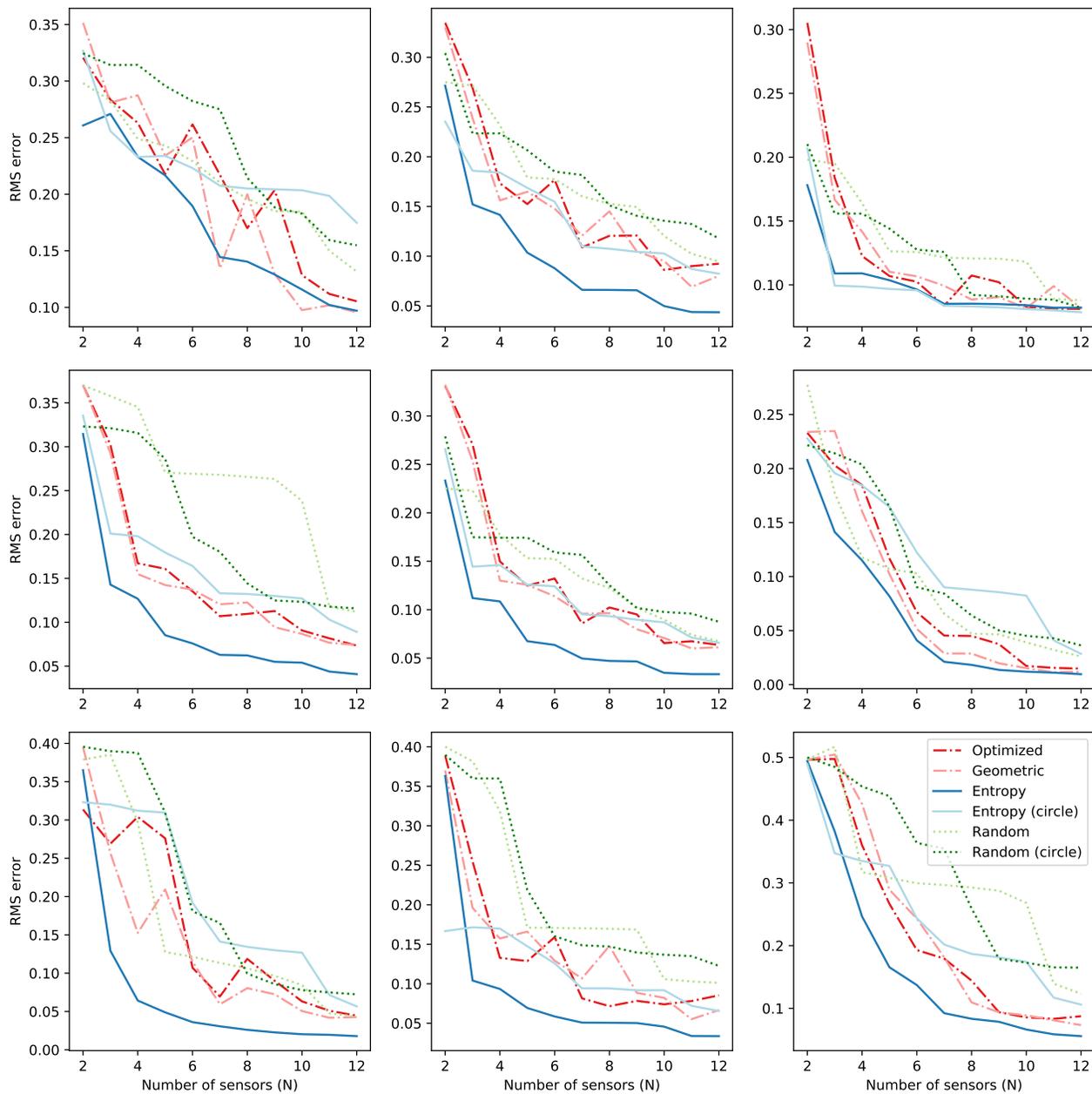


Figure 2: RMS error computed for different amount of sensors on 9 different functions using geometric and model-based sensor placement strategies.

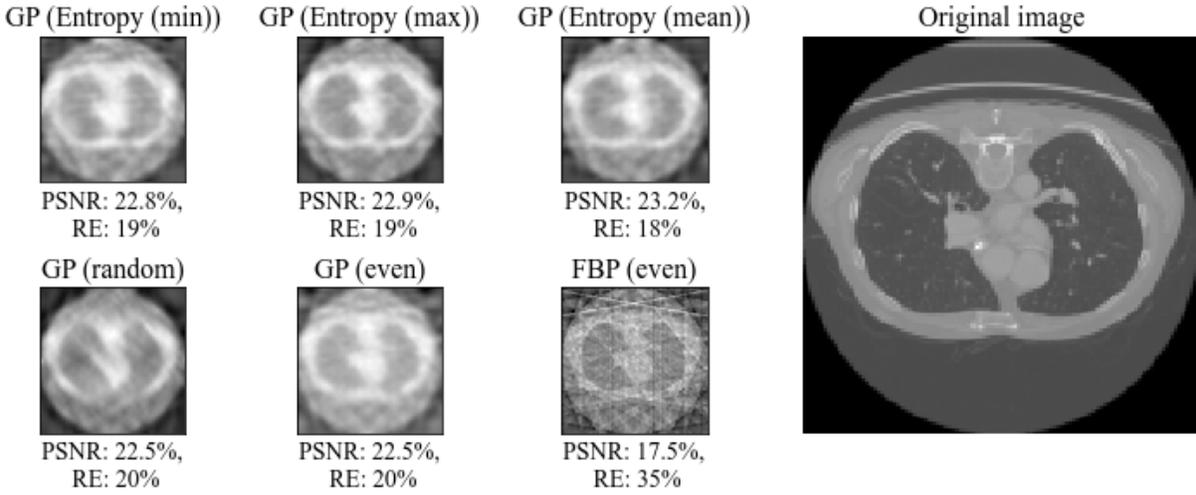


Figure 3: Means of the GP fit for 9 sensor locations chosen using entropy (mean, max, min), evenly spaced sensors as well as random selection. We also compare against filtered back projection (FBP).

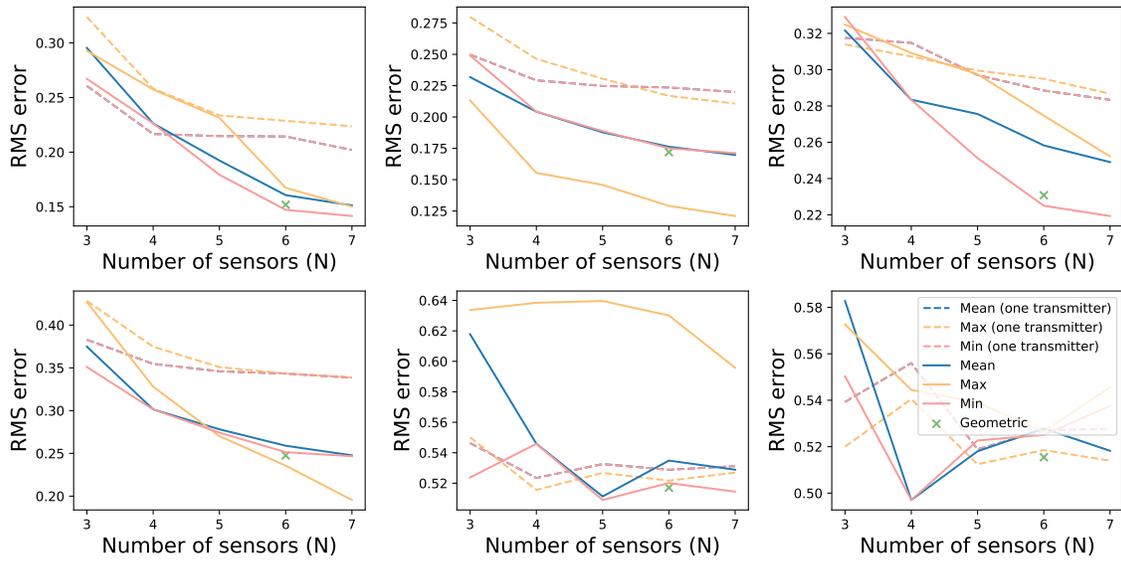


Figure 4: RMSE computed between a true functions and a GP mean with different sensor placement settings using the entropy criterion. *One transmitter* refers to a setting where we allow only one transmitter, whereas in the second case we let the algorithm to decide whether to add new transmitters or receivers at each step.

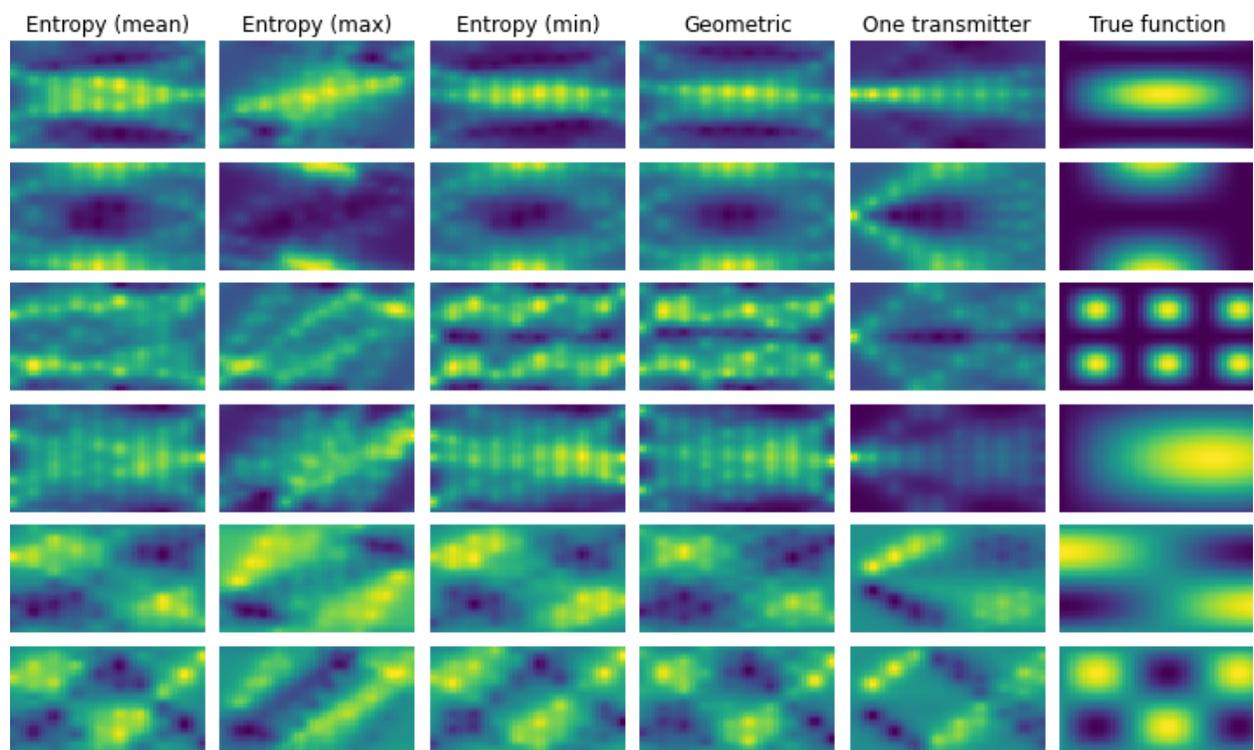


Figure 5: Mean of a GP fit with six sensors. The locations of the sensors have been obtained using entropy criterion with different summary strategies, as well as with the geometric approach. We also include the case where we allowed only one transmitter while optimizing for the receiver locations (*One transmitter*).

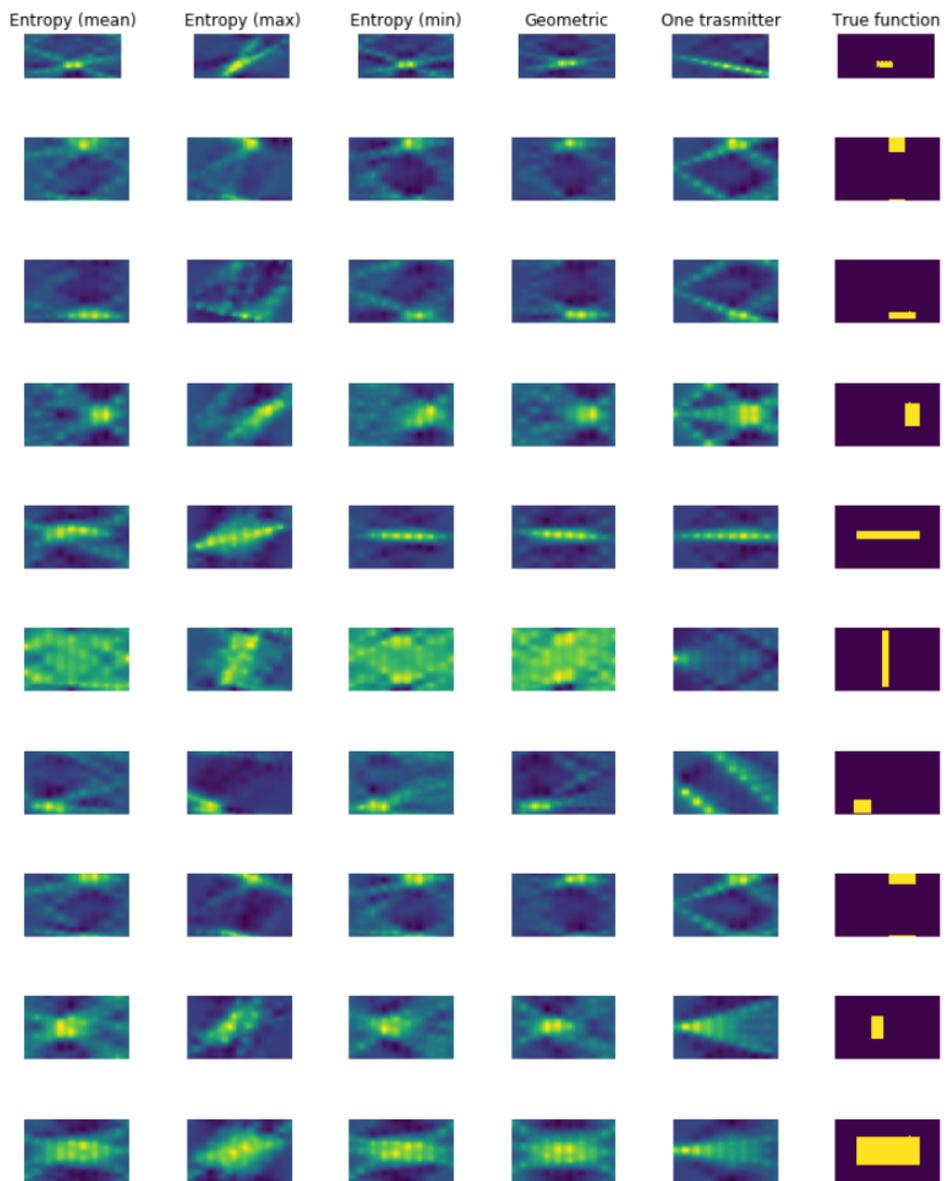


Figure 6: Mean of a GP fit with six sensors, where the locations have been obtained using the entropy criterion with different summary strategies.