

## Polar Gaussian Processes and Experimental Designs in Circular Domains\*

Esperan Padonou<sup>†</sup> and Olivier Roustant<sup>‡</sup>

**Abstract.** Predicting on circular domains is a central issue that can be addressed by Gaussian process (GP) regression. However, usual GP models do not take into account the geometry of the disk in their covariance structure (or kernel), which may be a drawback at least for industrial processes involving a rotation or a diffusion from the center of the disk. We introduce so-called polar GPs defined on the space of polar coordinates. Their kernels are obtained as a combination of a kernel for the radius and a kernel for the angle, based on either chordal or geodesic distances on the circle. Their efficiency is illustrated on two industrial applications. We further consider the problem of designing experiments on the disk. Two new Latin hypercube designs are obtained, by defining a valid maximin criterion for polar coordinates. Finally, an extension of the whole methodology to higher dimensions is investigated.

**Key words.** Kriging, polar coordinates, design of experiments, kernels, uncertainty quantification, disk

**AMS subject classifications.** 60G15, 65D05, 5D10, 62K20, 62K05

**DOI.** 10.1137/15M1032740

**1. Introduction.** This research aims at analyzing costly computer or physical experiments on a disk. The question was motivated by two industrial problems. The first one comes from the semiconductor industry where integrated circuits are produced on disks called wafers. Several technological processes such as lithography, heating, or polishing, exploit the geometry of the disk through rotations or diffusions from the center. A common issue is to reconstruct a quantity of interest over the whole disk, from a few measurements at specific locations. The second problem is related to air pollution modeling for environmental impact assessment. Greenhouse gas concentrations are simulated by a computer code. Among the input variables, the pair (speed, direction) of wind characteristics can be represented on a disk, the radius of which corresponds to the maximal speed. Here also, the goal is to predict the gas concentration from some simulated experiments.

Approximation problems on the disk have been considered since the works of Zernike [28] in optics. Zernike polynomials are orthogonal with respect to the usual scalar product on the unit disk, a useful property for linear models. For such models, it is shown that optimal designs of experiments (DoEs) are included in concentric circles [5]. More recently, a stochastic model consisting of a Gaussian process (GP), also called Kriging, has been proposed for microelectronics applications [24]. Among the existing interpolation and approximation methods, Kriging models are famous for their ability to provide both accurate prediction and uncertainty quantification, as pointed out in [14]. However their performance relies on the

---

\*Received by the editors August 7, 2015; accepted for publication (in revised form) May 4, 2016; published electronically August 30, 2016.

<http://www.siam.org/journals/juq/4/M103274.html>

<sup>†</sup>STMicroelectronics, 850 Rue Jean Monnet, 38920 Crolles, France ([padonou@emse.fr](mailto:padonou@emse.fr)).

<sup>‡</sup>Mines Saint-Etienne, UMR CNRS 6158, LIMOS, F-42023 Saint-Etienne, France ([roustant@emse.fr](mailto:roustant@emse.fr)).

choice of a covariance kernel, hereafter often simply called a kernel. Traditional kernels do not take into account the geometry of the disk. This may be a drawback, at least for technological or physical processes involving a diffusion from the center of the disk, or a rotation.

The main aim of the paper is to propose GP models that incorporate the geometry of the disk in their covariance kernel. For that purpose, we consider the parametrization of the unit disk in polar coordinates:  $\mathcal{D} = \{(\rho \cos \theta, \rho \sin \theta), \rho \in [0, 1], \theta \in \mathbb{S}\}$ , where  $\mathbb{S}$  represents the unit circle  $\mathbb{R}/2\pi\mathbb{Z}$ . The idea is to define a GP on the parametrization space  $\mathcal{C} = (0, 1] \times \mathbb{S}$  defined by  $(\rho, \theta)$ . This implies constructing a kernel on a product of the Euclidean space  $(0, 1]$  and of the circle  $\mathbb{S}$ , which can be done by algebraically combining kernels on these two spaces with sum, product, or ANOVA operations, for instance. The corresponding GPs will be called here *polar* GPs, and the usual ones based on Cartesian coordinates, *Cartesian* GPs.

The construction of kernels on  $\mathbb{S}$  can be achieved in several ways, and is connected to the literature of directional data (see, e.g., [18, 9]) and periodic functions (see, e.g., [25]). One possibility is to use so-called wrapped GP, obtained by transforming a multinormal density to a periodic one by applying an operator written as an infinite sum [17]. Here we focus on simpler approaches that provide explicit kernel expressions, either by considering restriction to  $\mathbb{S}$  of a 2-dimensional GP [25], involving the chordal distance on  $\mathbb{S}$ , or by using the recent results of Gneiting [12], involving the geodesic distance on  $\mathbb{S}$ . The geodesic distance on a general manifold was recently used in the context of free-form monitoring, with so-called geodesic GPs [4]. However, the goal and the approach are quite different here, where the form is fixed (the unit disk) and the geodesic distance known analytically. Furthermore, here the geodesic is relative to the manifold  $\mathbb{S}$  which is only an algebraic portion of the mapped space  $\mathcal{C}$ .

Second, we address the issue of defining an initial DoE for circular domains. Considering the space  $\mathcal{C}$  of polar coordinates is natural, but standard designs cannot be used directly due to its non-Euclidean structure. By considering a valid distance, we obtain maximin Latin hypercube designs (LHD, [20]) on  $\mathcal{C}$ . That class of designs is recommended when the process has a physical interpretation in polar coordinates. In order to deal with more general situations, we also propose a modified version, which still has the LHD structure with respect to  $\rho$  and  $\theta$ , and fills well the disk  $\mathcal{D}$ .

The paper is organized as follows. Section 2 presents the background and defines notations. Section 3 introduces so-called polar GPs. Section 4 shows the strength of the approach on two real applications, in microelectronics and environments. Section 5 is devoted to designs of experiments. Two new LHDs are introduced and compared to common designs, with respect to quality criteria. Their robustness in prediction is also investigated on a set of toy functions and models. Section 6 investigates an extension to higher dimensions, where the disk is replaced by a hyperball. Section 7 discusses the range of applicability of polar GPs and gives perspectives for future research.

**2. Background and notations.** Let  $\mathcal{D}$  denote the unit disk represented either in Cartesian or polar coordinates,

$$\mathcal{D} = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 \leq 1\} = \{(\rho \cos \theta, \rho \sin \theta), \rho \in [0, 1], \theta \in \mathbb{S}\},$$

where  $\mathbb{S} = \mathbb{R}/2\pi\mathbb{Z}$  is the unit circle. In various situations, one has to predict a variable of interest which is measured at a limited number of locations in  $\mathcal{D}$ . For that purpose, we will

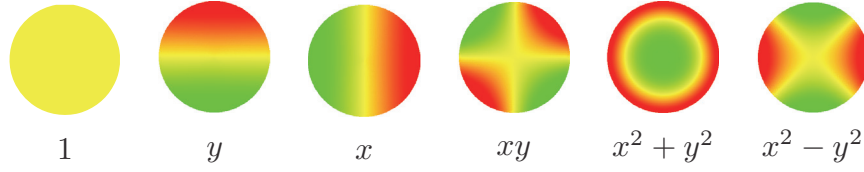


Figure 1. The six first Zernike polynomials.

consider the framework of GP regression [25], also called Kriging, in reference to its origins in geostatistics (see, e.g., [19]). The measurement locations, also called design points, will be denoted by  $X = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)})$ . In GP regression, the observed values at  $X$  are modeled by

$$(2.1) \quad Y_i = \mu(\mathbf{x}^{(i)}) + Z(\mathbf{x}^{(i)}) + \eta_i,$$

where  $\mu$  is a trend function,  $Z \sim GP(0, k)$  is a centered GP with covariance function—or *kernel*— $k$ , and  $\eta_1, \dots, \eta_n$  are Gaussian random variables representing noise. We now briefly detail the three parts of the model.

The trend function  $\mu$  is deterministic and often modeled as a linear combination of basis functions. Here, Zernike polynomials [28] are good candidates since they constitute an orthogonal basis for the usual scalar product on  $\mathcal{D}$ . Their shapes including regular patterns are suited to describe symmetries or rotations. They were recently used for predicting on a disk [24]. The first Zernike polynomials, up to order 2, are shown in Figure 1. The reader is referred to [28] for more details.

The stochastic part of model (2.1) comprises a GP and a noise. The GP  $Z$  takes into account the spatial dependence, which thus entirely depends on its kernel  $k$ . The choice of  $k$  is crucial for applications, and may be done in order to include knowledge such as smoothness, periodicity, symmetries, etc. There are many ways to construct a kernel, and a comprehensive presentation is found in [25, section 4]. A key idea is that multidimensional kernels can be obtained by algebraic operations, such as sum or products, of 1-dimensional kernels.

Finally the noise part represented by the  $\eta_i$ 's may have different purposes: modeling a measurement noise or potential discrepancies between the dataset and the kernel, and addressing numerical issues such as ill-conditioning [1, 13]. The  $\eta_i$ 's are modeled as independent  $N(0, \tau^2)$ , where  $\tau^2$  is an unknown homogeneous variance term often called “nugget” or “jitter.” When conditioning on the observed values, the model is an interpolator if  $\tau = 0$ . It is a smoother when  $\tau > 0$ , which gives more flexibility.

When all parameters are known, prediction with (2.1) is given in a closed form by a Gaussian conditional distribution knowing the observations  $Y_i$ ,  $i = 1, \dots, n$ . Its two moments are known as Kriging mean and Kriging variance. Analytical expressions are also available when the parameters are estimated, known as universal Kriging formulas that we use here (see, e.g., [25]). An important fact is that the Kriging mean at a new site  $\mathbf{x}$  is obtained as an affine combination of the observed values  $Y_i$  that are correlated to  $Z(\mathbf{x})$ . Though all the locations may be involved in the prediction, the neighboring locations, corresponding to high correlations, typically play a key role.

**3. Polar GPs.** One way to define a GP on the unit disk  $\mathcal{D}$  is to use the restriction of a GP on the square  $[0, 1]^2$ , defined in Cartesian coordinates. In this paper, we will call them

Cartesian GPs. In our work, we propose to further exploit the geometry of the disk by using the polar coordinates. The associated GPs will be called polar GPs.

When using the polar coordinates, the unit disk  $\mathcal{D}$  is connected to the cylinder  $\mathcal{C} = (0, 1] \times \mathbb{S}$ , where  $\mathbb{S}$  denotes the unit circle:

$$(3.1) \quad \Psi : (\rho, \theta) \in \mathcal{C} \mapsto (\rho \cos \theta, \rho \sin \theta) \in \mathcal{D} \setminus \{\mathbf{0}\}.$$

It is a one-to-one correspondence from  $\mathcal{C}$  to the unit disk without its center. The fact that the center is lost in the mapping may be a problem in theory. In practice a design point located at the center of the disk can be replaced by a set of design points placed on a closed concentric circle. A GP on  $\mathcal{D}$  can then be obtained by using  $\Psi^{-1}$ , resulting in kernels on  $\mathcal{D} \times \mathcal{D}$  of the form

$$(3.2) \quad k(\mathbf{x}, \mathbf{x}') = k_{\mathcal{C}}(\Psi^{-1}(\mathbf{x}), \Psi^{-1}(\mathbf{x}')),$$

where  $k_{\mathcal{C}}$  is a kernel on  $\mathcal{C} \times \mathcal{C}$ . Such transformations are referred to as “warping” in the context of GP modeling (see, e.g., [25, section 4.2.3]).

Kernels on the cylinder can be defined by exploiting its product structure. This can be done by combining a kernel  $k_r$  on  $(0, 1] \times (0, 1]$  and a kernel  $k_a$  on  $\mathbb{S} \times \mathbb{S}$ . A first way is by using the tensor product

$$(3.3) \quad k_{\text{prod}}(\mathbf{u}, \mathbf{u}') = k_r(\rho, \rho') k_a(\theta, \theta'),$$

where  $\mathbf{u} = (\rho, \theta)$  and  $\mathbf{u}' = (\rho', \theta')$  are in  $\mathcal{C}$ . This formulation implicitly assumes that the GP  $Z$  is the product of two independent components, a radial process  $R_\rho$  and an angular process  $A_\theta$  ( $Z_{\mathbf{u}} = R_\rho A_\theta$ ). It corresponds to a simple form of interaction. For processes that do not have interactions between these components ( $Z_{\mathbf{u}} = R_\rho + A_\theta$ ), an additive kernel should be more appropriate:

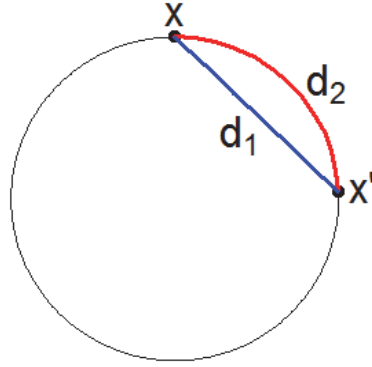
$$(3.4) \quad k_{\text{add}}(\mathbf{u}, \mathbf{u}') = k_r(\rho, \rho') + k_a(\theta, \theta').$$

A trade-off between these two extreme approaches is the ANOVA kernel defined as

$$(3.5) \quad k_{\text{ANOVA}}(\mathbf{u}, \mathbf{u}') = \left(1 + k_r(\rho, \rho')\right) \left(1 + k_a(\theta, \theta')\right).$$

The expanded form of (3.5) shows that a process  $Z_{\mathbf{u}}$  with ANOVA kernel can be viewed as a sum of four independent GPs: a constant process  $Z^0$ , a radial process  $R_\rho$  with kernel  $k_r$ , an angular process  $A_\theta$  with kernel  $k_a$ , and a process  $Z^{\text{inter}}$  on  $\mathcal{C}$  with kernel  $k_r k_a$ . From the ANOVA point of view, these processes are similar to constant term, main effects, and second-order interaction [8], but without respecting the unicity constraints such as centering. For more details on how to make new kernels from old, we refer the reader to [25].

Let us now define the kernels  $k_r$  on  $(0, 1] \times (0, 1]$  and  $k_a$  on  $\mathbb{S} \times \mathbb{S}$ . We recall that valid kernels must be positive definite. The domain  $(0, 1]$  is a segment of a 1-dimensional Euclidean space. As a consequence, traditional kernels are suitable for  $k_r$ . In particular, Matérn kernels



**Figure 2.** Chordal ( $d_1$ ) and geodesic ( $d_2$ ) distances on  $\mathbb{S}$ .

are attractive for their ability to control the smoothness of the process and to ensure numerical stability. In dimension 1, the Matérn  $\frac{5}{2}$  kernel is given by

$$(3.6) \quad k_m(x, x') = \left(1 + \frac{\sqrt{5} |x - x'|}{\ell} + \frac{5(x - x')^2}{3\ell^2}\right) \exp\left(-\frac{\sqrt{5} |x - x'|}{\ell}\right).$$

A simple way of defining kernels on  $\mathbb{S} \times \mathbb{S}$  is mentioned in [12]. They are based on the chordal distance  $d_1(\theta, \theta') = 2 \sin(\frac{\theta - \theta'}{2})$  and the geodesic distance  $d_2(\theta, \theta') = \text{acos}(\cos(\theta - \theta'))$  illustrated in Figure 2.

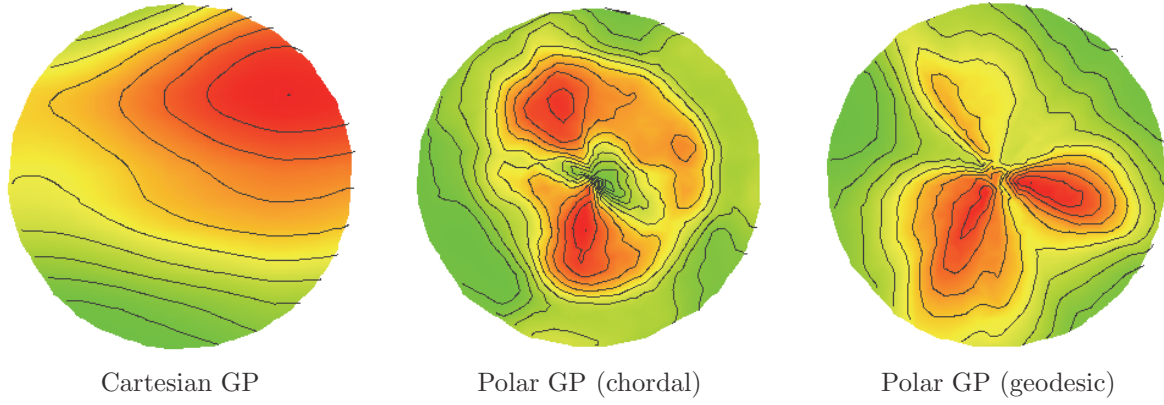
To define a kernel on  $\mathbb{S} \times \mathbb{S}$ , one could be tempted to compose usual kernels with  $d_1$  or  $d_2$ . Unfortunately, positive definiteness is not guaranteed for the resulting functions when  $d_2$  is used. As a counterexample, if the Gaussian kernel is chosen for  $k_a$ , then  $k_a \circ d_2$  is not positive definite [12, Thm. 8]. Alternatively, two sufficient conditions of positive definiteness over  $\mathbb{S} \times \mathbb{S}$  are provided by Gneiting [12]. Define  $F_d$  the class of continuous functions  $\varphi : [0, \infty) \rightarrow \mathbb{R}$ , with  $\varphi(0) = 1$  and such that the function  $(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^d \times \mathbb{R}^d \mapsto \varphi(\|\mathbf{x} - \mathbf{x}'\|)$  is positive definite. Then

- (i) if  $\varphi \in F_2$ , then  $\varphi \circ d_1$  is a kernel on  $\mathbb{S} \times \mathbb{S}$ ;
- (ii) if  $\varphi \in F_1$  is such that  $\varphi(t) = 0$  for  $t \geq \pi$ , then  $\varphi \circ d_2$  is a kernel on  $\mathbb{S} \times \mathbb{S}$ .

Kernels satisfying (i) were initially proposed by Yadrenko in 1983 and are often used to describe periodic functions (see, e.g., [25]). They correspond to restrictions of 2-dimensional isotropic GPs on  $\mathbb{R}^2$  to  $\mathbb{S}$ . The second result is due to Lévy in 1961. Kernels satisfying (ii) can be constructed from compactly supported functions on  $\mathbb{R}$  such as the  $C^2$ -Wendland function defined for  $0 \leq t \leq \pi$ :

$$(3.7) \quad W_c(t) = \left(1 + \tau \frac{t}{c}\right) \left(1 - \frac{t}{c}\right)_+^\tau, \quad c \in (0, \pi], \tau \geq 4.$$

For the geodesic distance, we use  $c = \pi$ , which is the largest possible value due to condition (ii) above. With this choice, the covariance between two angles  $\theta, \theta'$  is zero when  $d_2(\theta, \theta') = \pi$ , and strictly positive for  $d_2(\theta, \theta') < \pi$ . The same interpretation is possible for the chordal distance with  $c = 2$ , though it is not necessary to use a compactly supported function in that



**Figure 3.** Simulations of Cartesian and polar GPs with kernels (a)–(c).

case. From now on, we will use the Wendland function in both cases, resulting in the two following kernels on  $\mathbb{S} \times \mathbb{S}$ :

$$(3.8) \quad k_{\text{chord}}(\theta, \theta') = W_2(d_1(\theta, \theta')),$$

$$(3.9) \quad k_{\text{geo}}(\theta, \theta') = W_\pi(d_2(\theta, \theta')),$$

and the corresponding GPs will be denoted *polar GP (chordal)* and *polar GP (geodesic)*.

**GP simulations on the unit disk.** In order to better understand the specificities of polar GPs, it is useful to draw simulated surfaces. For the sake of simplicity, we propose to focus on the ANOVA combinations. We consider a Cartesian GP and the two polar GPs (chordal, geodesic) defined in (3.8), (3.9). Their expressions are written below, including variance factors  $s^2$ ,  $\alpha_1^2$ ,  $\alpha_2^2$ :

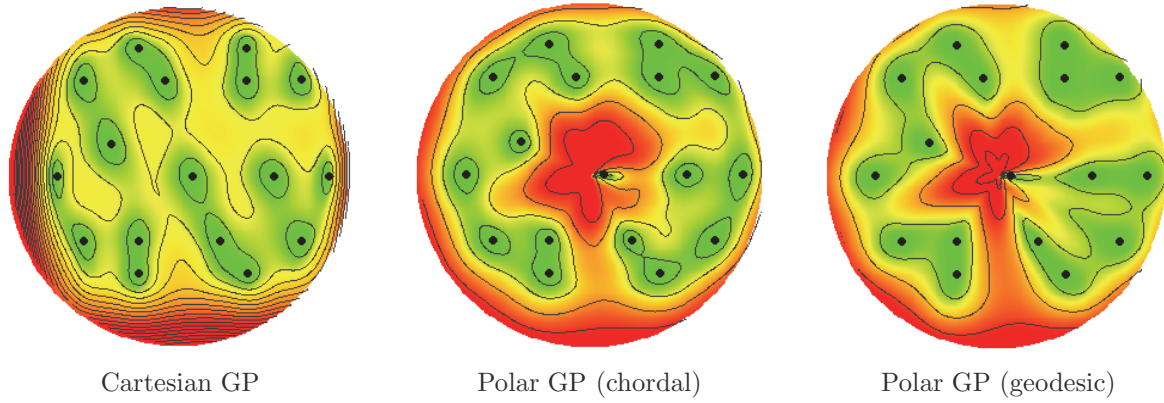
- (a)  $k(\mathbf{x}, \mathbf{x}') = s^2(1 + \alpha_1^2 k_m(x, x'))(1 + \alpha_2^2 k_m(y, y'))$ ,
- (b)  $k(\mathbf{x}, \mathbf{x}') = s^2(1 + \alpha_1^2 k_m(\rho, \rho'))(1 + \alpha_2^2 k_{\text{chord}}(\theta, \theta'))$ ,
- (c)  $k(\mathbf{x}, \mathbf{x}') = s^2(1 + \alpha_1^2 k_m(\rho, \rho'))(1 + \alpha_2^2 k_{\text{geo}}(\theta, \theta'))$ .

Simulation results are displayed in Figure 3. We can see that with polar GPs, the simulated surface exhibits radial and angular patterns around the center of the disk. Such kernels may thus be suitable to describe physical phenomena involving such effects. Figure 4 shows via Kriging standard deviation how model uncertainty varies over  $\mathcal{D}$ , given a design of 17 points. Two striking differences are visible, especially between the Cartesian GP and the polar GP (geodesic), about uncertainty at the center of the disk, and uncertainty regions at the vicinity of design points. On one hand, the neighborhoods produced by the Cartesian GP look like elliptical regions at any location of the circular domain. On the other hand, those produced by the polar GP (geodesic) look like pie chart sectors, oriented towards the center of the disk, which plays a particular role. This is also true for the polar GP (chordal), though less pronounced.

## 4. Applications.

**4.1. Quality control in microelectronics.** In microelectronics, integrated circuits are produced on circular slices of semiconductor materials called wafers. For quality monitoring,





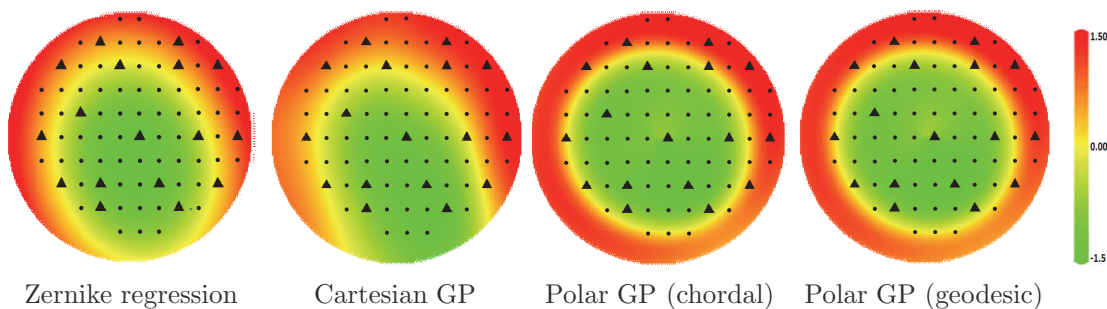
**Figure 4.** Kriging standard deviations for Cartesian and polar GPs (kernels (a)–(c)).

physical and electrical variables are collected on a set of locations of these wafers. In this example, the characteristic of interest is thickness, a key parameter affecting performance of integrated circuits. In our industrial background, only 17 predefined points are measured for economic reasons. The statistical challenge consists in predicting nonmeasured locations in order to assess the spatial risk of default from this dataset. For the purpose of this study, thickness is further measured at 64 new locations to serve as a test grid. For the sake of confidentiality, the technological process is not detailed and the thickness values are rescaled. It produces here data with a pronounced radial pattern. However, we will not assume that the model is purely radial, which is a too strong assumption in practice, due to the numerous successive operations on a wafer, and the possible slacks in processing. The aim of this section is to compare the Cartesian and polar GPs (chordal, geodesic), obtained with 3 types of algebraic combination (product, sum, ANOVA). The Cartesian GPs considered here are obtained by tensor product, tensor sum, or ANOVA product of the 1-dimensional Matérn kernel of (3.6). For the polar GPs, we use the same combinations for a kernel  $k_r$  on  $(0, 1]$  and a kernel  $k_a$  on  $\mathbb{S} \times \mathbb{S}$ , according to (3.3), (3.4), (3.5). For  $k_r$ , we use again the Matérn kernel, whereas for  $k_a$  we choose  $k_{\text{chord}}$  or  $k_{\text{geo}}$  (see (3.8), (3.9)). The range parameters  $\tau$  and  $\theta$ , as well as the variance factors  $s^2$ ,  $\alpha_1^2$ ,  $\alpha_2^2$  are estimated by maximum likelihood with R package `kergp` [6]. The optimizer used is the method L-BFGS-B proposed in the `optim` function in R: an adaptation of the quasi-Newton method BFGS for boundary constraints. To improve its performances, we added a multistart step: 10 initial points were sampled at random, and for each of them a separate optimization was performed. The best result among the ten was finally chosen. The model accuracy is computed on the 64 test points, with the root mean squared error (RMSE) criterion. The results are summarized in Table 1 when  $\mu$  is constant in (2.1). They show that the smallest prediction errors are obtained with the polar GPs, corresponding to gains around 20% compared to the Cartesian GP. Adding Zernike polynomials as a trend slightly improves the result for the Cartesian GP, but the untrended polar GPs still outperform with a gain of 15%. Actually the trend captures the main part of the phenomenon and the GP part has then a minor effect: results are the same as for a pure linear model based on Zernike polynomials of order 2.

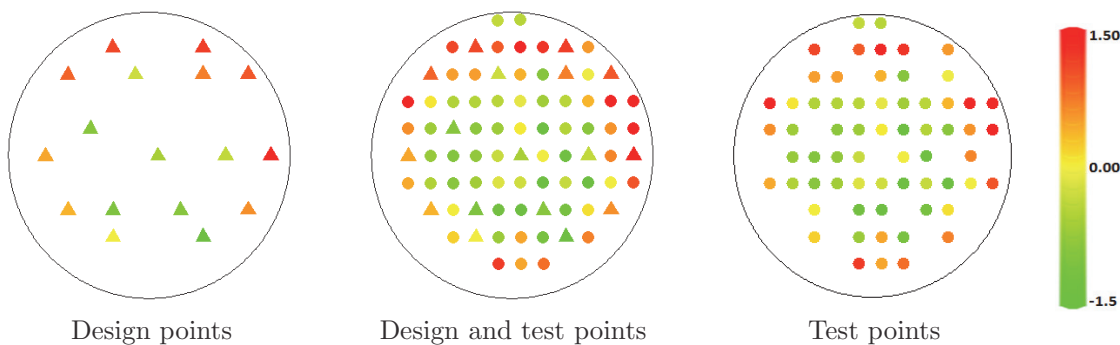
Table 1

RMSE computed on 64 test points for several GPs with a constant trend. For each GP type, the combination resulting in the smallest RMSE is marked by an asterisk. When a Zernike trend is added, the best RMSE is equal to 0.71 for all GP types, corresponding to the score of the trend only.

GP type	Cartesian			Polar (chordal)			Polar (geodesic)		
Kernel type	$k_{\text{prod}}$	$k_{\text{add}}$	$k_{\text{ANOVA}}$	$k_{\text{prod}}$	$k_{\text{add}}$	$k_{\text{ANOVA}}$	$k_{\text{prod}}$	$k_{\text{add}}$	$k_{\text{ANOVA}}$
RMSE	0.75 *	0.77	0.76	0.69	0.60 *	0.62	0.68	0.61 *	0.65



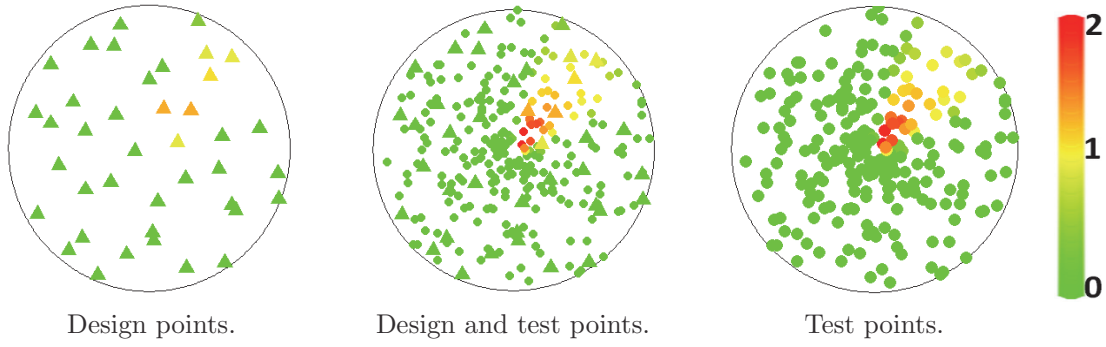
**Figure 5.** Prediction surface for the best untrended GP models of Table 1. When adding a Zernike trend, the prediction surface is approximately the same as for a pure Zernike regression represented on the left. Black bullets correspond to test points, triangles to design points.



**Figure 6.** Rescaled thickness values. The 81 measurement locations are shown in the middle, including 17 design points (triangles, left) and 64 test points (bullets, right).

In order to further analyze the results, we select for each GP type the kernels corresponding to the best combination, indicated by an asterisk in Table 1. The prediction surfaces obtained with these 3 kernels are shown on Figure 5. All the GPs succeed in recovering the radial pattern of the dataset, visible on Figure 6, middle. However, it is less faithfully identified by the Cartesian GP. The differences on the predicted values can be explained by looking at the space in which the kernel is defined. For polar GPs, prediction at one location will particularly involve the locations corresponding to a high correlation according to  $\rho$  or  $\theta$ . Typically, the resulting neighborhoods in  $\mathcal{D}$  may look like pie chart sectors (high radial correlation) or ring portions (high angular correlation). Here, a closer look at estimated parameters reveals that there is a high angular correlation. Therefore, prediction at the bottom of the disk involves

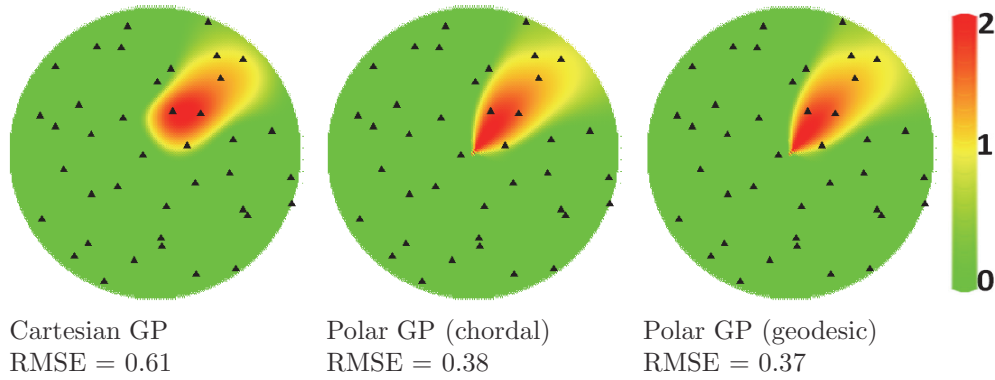




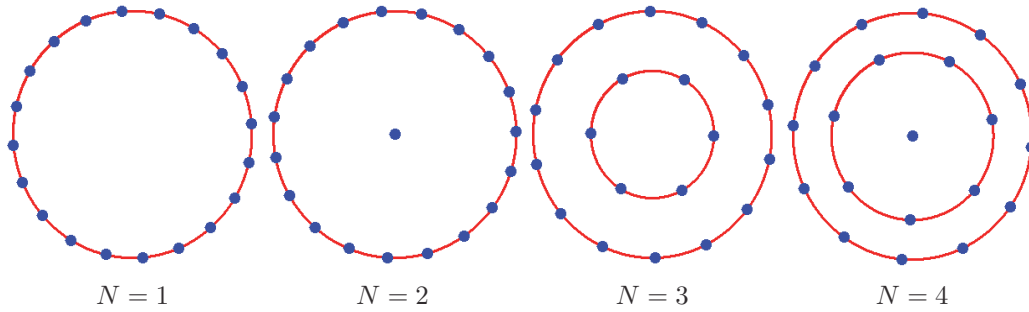
**Figure 7.** Rescaled gas concentrations. The 242 simulation locations are shown in the middle, including 30 design points (triangles, left) and 212 test points (bullets, right).

the other points that are close to the boundary. On the other hand, for the Cartesian GP, the predicted thickness has a low value, since the measurement points around, in the  $(x, y)$  space, have a low value. Finally notice that the predicted value at the extreme boundary of the disk should be considered with care, since no test points are defined on this region due to technical constraints.

**4.2. Air pollution modeling with a directional input.** The problem tackled here is an environmental question. A greenhouse gas emitted by a known source, usually an industrial plant, is measured at a given location for air quality monitoring. In the absence of sensors, gas concentration must be predicted. For simple landscapes, analytical expressions are available based on transport and diffusion equations. However, for complex landscapes, gas concentration is simulated by numerical codes [2]. The input variables include the emitted flow, landscape characteristics, and meteorological variables. Here we focus on wind speed and wind direction. In this short study, 242 simulations were carried out, 30 of which serve as design points and the other ones are used for tests, as illustrated in Figure 7. The wind speed, initially given on the range  $[0; 12]$  ( $m.s^{-1}$ ), is rescaled to  $[0, 1]$ . With this transformation, the domain of the variables (speed, direction) is the unit disk. The aim of this study is simply to compare the prediction accuracy of Cartesian and polar GPs without using a priori information. In particular, we do not specify the constraints of positivity or nullity of the gas concentration on a known subregion. We use the same kernels as in the first application, corresponding to 3 algebraic combinations (product, sum, ANOVA). Here, the best model is obtained for the tensor-product combination for all kinds of GPs. This claims in favor of an interaction speed direction for the wind on gas concentration. Notice that adding a Zernike polynomial trend does not improve the results here, since the angular shape is restricted to a region of the disk, which is hard to capture with Zernike polynomials. The results are displayed in Figure 8. In terms of prediction accuracy (measured by the RMSE criterion) the polar GPs are clearly outperforming, corresponding to gains around 40% compared to the standard tensor-product Matérn kernel. Furthermore, for the polar GPs the influence of wind direction on gas concentration has an angular shape, which is intuitive, and corresponds to the true shape visible in Figure 7 (middle). On the other hand, this shape is rectangular for the Cartesian GP.



**Figure 8.** Estimated gas concentrations according to wind speed ( $\rho$ ) and direction ( $\theta$ ), for untrended Cartesian and polar GPs. Adding a Zernike polynomial trend does not improve the results. Triangles correspond to design points.



**Figure 9.** 20-point D-optimal DoEs for Zernike polynomials of degree  $N$ .

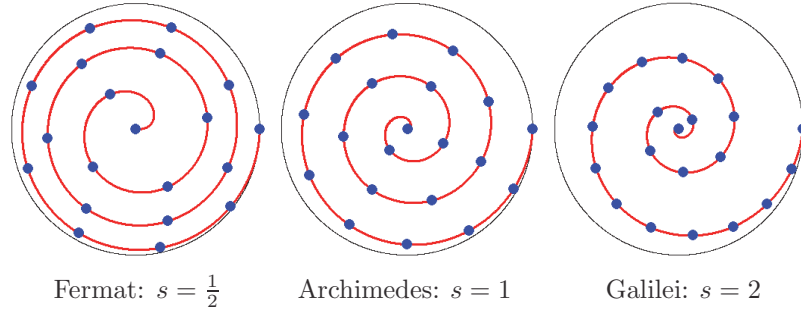
## 5. DoEs on the disk.

**5.1. Optimal designs for Zernike polynomials and spirals.** Among the DoEs that are specific to the disk, there are optimal designs for Zernike polynomials. The D-optimal designs were investigated in [5] and were found to be contained in a few concentric circles, as illustrated in Figure 9.

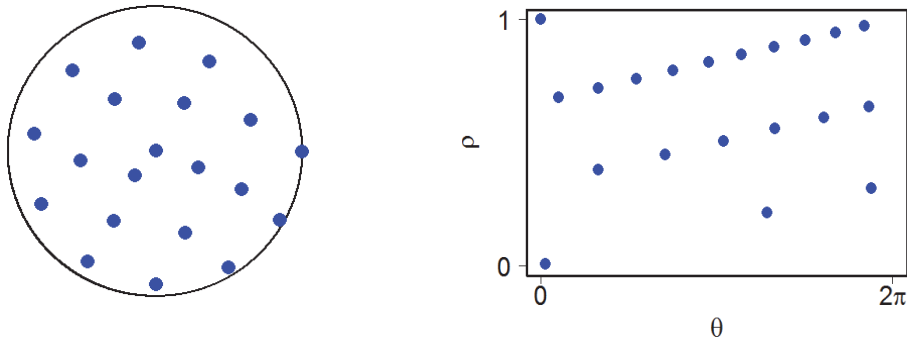
Spirals, hereafter denoted spiral DoEs, are used in various industrial settings: micro-electronics, optics, microbiology, etc. They allow us to control the density of the design (see, e.g., [23]). Some of them are represented in Figure 10, corresponding to the equation  $\rho = a\theta^p + b$ .

D-optimal DoEs for regression models are not robust to departures from the assumed shapes [16], and do not fill the space, a property usually required in the framework of GP modeling for capturing potential nonlinearities. Poor space-filling properties are also visible for spirals in the space  $(\rho, \theta)$  of polar coordinates (as shown in Figure 11), though they may correctly fill the disk.

**5.2. Maximin Latin hypercubes for polar coordinates.** For metamodeling a potentially complex phenomenon, two main properties are expected from a good DoE: space-filling, in order to capture nonlinearities, and uniformity of the marginal distributions, to avoid redun-



**Figure 10.** 20-point DoEs defined from spirals of the form  $\rho = a\theta^s + b$  with  $\theta \in [0, 6\pi]$ . The parameter  $s$  controls the speed with which the curve moves away from the center, and  $a, b$  are chosen such that the spirals start at the center and end at the boundary.



**Figure 11.** Cartesian (left) and polar (right) representations of the Archimedean spiral DoE. This DoE fills well the disk but not the cylinder of polar coordinates.

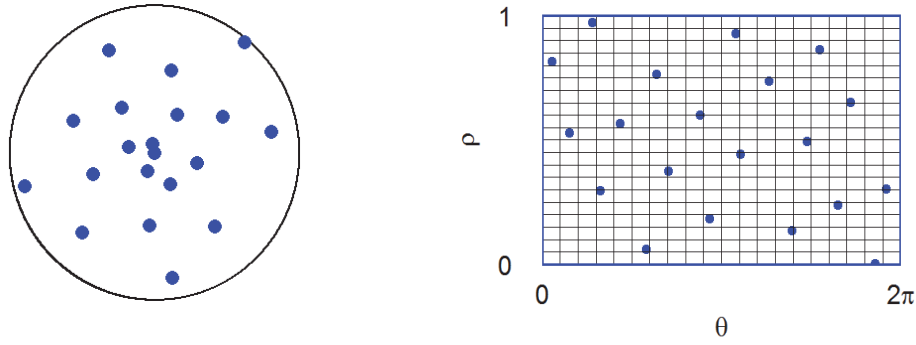
dancies in projection. Among the indicators used to assess space fillingness, the maximin criterion [21] is a common choice. In addition, LHDs, [20] provide good projection properties onto marginal dimensions. Thus, maximin LHDs are often proposed as initial DoEs. However such designs cannot be directly used in polar coordinates, due to the nonEuclidean structure of  $\mathcal{C}$ . The aim of this section is to adapt their construction.

Let us first recall the construction of a maximin LHD over the hypercubic domain  $[0, 1]^2$ . Given a design  $\mathbf{X} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)})$  of elements of  $[0, 1]^2$ , we denote  $\Phi_{\text{Mn}}(\mathbf{X})$  as the so-called maximin criterion, giving the minimal distance among design points,

$$(5.1) \quad \Phi_{\text{Mn}}(\mathbf{X}) = \min_{i \neq j} \left( \|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\| \right).$$

A maximin DoE is a design that maximizes  $\Phi_{\text{Mn}}$ . However,  $\Phi_{\text{Mn}}$  is hard to optimize and a regularized version  $\Phi_p$ , more suitable for optimization, was proposed in [22]:

$$(5.2) \quad \Phi_p(\mathbf{X}) = \left( \sum_{1 \leq i < j \leq n} \|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|^{-p} \right)^{\frac{1}{p}}.$$



**Figure 12.** Cartesian (left) and polar (right) representations of a 20-point maximin Latin cylinder design (LCD). The design fills well the cylinder  $\mathcal{C}$  of polar coordinates, displayed as a 2-dimensional map: In particular, the design points near the left and right boundaries are also spread out from each other.

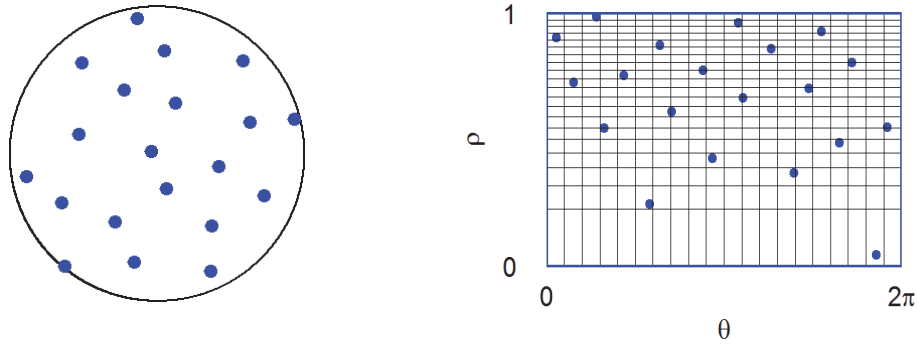
For  $p \rightarrow \infty$ , maximizing  $\Phi_{Mn}$  is equivalent to minimizing  $\Phi_p$ . Following [22, 3], we will use  $p = 50$ . In software, the algorithms used for optimization are often based on simulated annealing or evolutionary strategies (see, e.g., [10]). When the input variables are not provided in the same unit of measure, a maximin LHD is first designed over  $[0, 1]^2$ , corresponding to dimensionless variables.

Now let us consider the cylinder  $\mathcal{C}$  of polar coordinates. The construction of a Latin hypercube on  $\mathcal{C}$  is identical for a hypercubic domain, by considering discretizations of  $[0, 1]$  and  $\mathbb{S}$ . For the sake of clarity, we propose to call *polar Latin cylinder design (polar LCD)* or simply LCD, an LHD defined in polar coordinates, referring to the geometry of the polar space. As for the maximin criterion, two modifications are needed for polar coordinates. First, a valid distance on  $\mathcal{C}$  must fulfill the condition  $\|\mathbf{u} - \mathbf{u}'\| = 0$  for  $\mathbf{u} = (\rho, \theta)$  and  $\mathbf{u}' = (\rho, \theta')$ , with  $\theta = \theta' \pmod{2\pi}$ . In particular the Euclidean distance is no longer valid since it does not see that the points  $(\rho, 0)$  and  $(\rho, 2\pi)$  are the same in  $\mathcal{C}$ . Second, the range of the polar angle  $\theta$  is  $\pi$ , which is the maximum value of the geodesic distance over  $\mathbb{S}$ . Therefore, any distance over the dimensionless cylinder  $[0, 1] \times (\frac{1}{\pi}\mathbb{S})$  applies to the polar space  $(\rho, \theta)$ . A natural choice is the geodesic distance given by

$$(5.3) \quad \|\mathbf{x} - \mathbf{x}'\|_{\text{Polar}} = \sqrt{(\rho - \rho')^2 + \left(\frac{d_2(\theta, \theta')}{\pi}\right)^2}.$$

Notice that the factor  $\frac{1}{\pi}$  rescales  $d_2$  to  $[0, 1]$  and weighs equivalently the radius and the angle.

From now on we will denote  $\Phi_{\text{Polar}}$  (resp.,  $\Phi_{\text{Cartesian}}$ ) the  $\Phi_p$  criteria computed with  $\|\cdot\|_{\text{Polar}}$  (resp.,  $\|\cdot\|_2$ ). Minimizing  $\Phi_{\text{Polar}}$  leads to a maximin LCD. A 20-point maximin LCD is displayed in Figure 12, where the cylinder is represented as a 2-dimensional map. As expected it fills well the space of polar coordinates. Though it looks similar to a maximin LHD obtained in a hypercubic domain with the usual Euclidean distance, the difference is visible on the left and right boundaries which correspond to the same points in  $\mathcal{C}$ : The design points near the left and right boundaries are also spread out from each other. LCDs are recommended when the studied phenomenon has a physical interpretation with respect to polar coordinates. First, if the phenomenon is purely radial (resp., angular), the Latin structure



**Figure 13.** Cartesian (left) and polar (right) representations of the LCD obtained by transforming the maximin LCD of Figure 12 with  $\rho \mapsto \sqrt{\rho}$ .

ensures that all the design radius (resp., angles) values are different, so that no information is lost by projection. Furthermore, the maximin property helps in capturing nonlinearities with respect to  $\rho$  and  $\theta$ . However, when no a priori information about the phenomenon is known, the maximin LCD may be inappropriate, due to nonuniform filling that they produce on  $\mathcal{D}$ , as visible in Figure 12. Though it is not possible to optimize simultaneously maximin criteria based on distances in Cartesian and polar coordinates, a multicriteria approach could be investigated. In this paper, as a first study, we focus on a simple transformation of a maximin LCD which helps improving space fillingness on  $\mathcal{D}$  while preserving the Latin structure on  $\mathcal{C}$ . This is done by applying the transform  $\rho \mapsto \sqrt{\rho}$ , based on the well-known fact that if  $R, T$  are independent random variables with uniform distribution on  $[0, 1]$  and  $[0, 2\pi]$  respectively, then  $(\sqrt{R}, T)$  is uniform on  $\mathcal{D}$ . This transformation was applied to the design of Figure 12, resulting in the design displayed in Figure 13.

**5.3. Comparison.** The aim of this section is to compare the DoEs presented above with respect to quality criteria, and to evaluate their performance on a set of toy functions. We will denote  $\text{Dopt1}, \dots, \text{Dopt4}$  the D-optimal DoEs for Zernike regression of order  $N$  ( $1 \leq N \leq 4$ ) shown in Figure 9, and Spiral-F, Spiral-A, Spiral-G the spiral DoEs (Fermat, Archimedes, Galilei) of Figure 10. We also denote  $\text{maxLCD}$  the maximin LCD of Figure 12 and  $\text{maxLCD}^*$  its transformed version with  $\rho \mapsto \sqrt{\rho}$  (Figure 13). All these 20-point DoEs are compared according to the following scheme:

- (i) An assessment is made according to space filling and D-optimality criteria. For space filling, two indicators are used, the minimum Euclidean distance, and the minimum geodesic distance (5.3) between design points. The D-optimality criterion for the  $N$ -order Zernike regression (see [5]) is given in log-scale.
- (ii) A comparison in terms of prediction accuracy. The RMSE over a test grid of 1,000 points is computed for the 6 analytical functions shown in Figure 14, illustrating various nonlinear patterns. For each DoE, the best model is chosen among Zernike polynomials up to order 4, Cartesian GPs, and polar GPs with kernels obtained by combination (sum, product, ANOVA) of 1-dimensional kernels as in section 4.

The results of Table 2 are consistent with the theory of D-optimality and exhibit the lack of robustness of D-optimal designs in the case of departure from their assumptions, especially

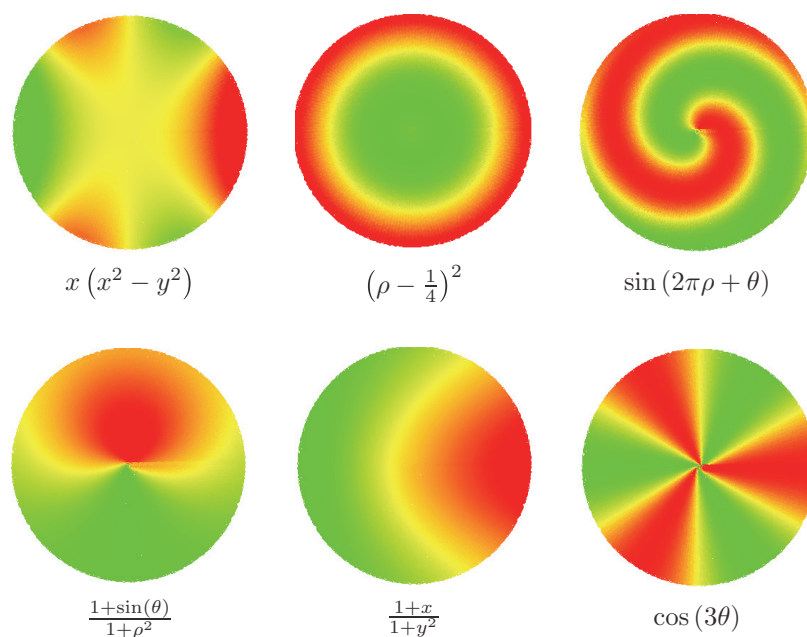


Figure 14. Color representation of test functions.

Table 2

Comparison of DoEs according to D-optimality and space-filling criteria.

	D-optimality			$\min_{i \neq j} (\  \mathbf{x}^{(i)} - \mathbf{x}^{(j)} \ )$	
	$D_{N=2}$	$D_{N=3}$	$D_{N=4}$	$\  \cdot \ _{\text{Polar}}$	$\  \cdot \ _2$
D-opt1	-159.9	-308.3	-448.3	0.01	0.31
D-opt2	<b>36.6</b>	-135.6	-353.1	0.01	0.33
D-opt3	35.4	<b>49.1</b>	-18.9	0.02	<b>0.45</b>
D-opt4	34.4	47.5	<b>63.5</b>	0.03	0.32
Spiral-F	29.7	37.2	44.1	0.04	0.20
Spiral-A	27.2	31.6	32.1	0.03	0.22
Spiral-G	23.3	19.3	-1.4	0.01	0.13
maxLCD	22.2	20.3	2.4	<b>0.06</b>	0.06
maxLCD*	27.5	32.8	33.0	0.04	0.28

when  $N$  is underestimated. The comparison also shows that spiral DoEs have rather good scores for all criteria. The best spirals for Zernike polynomials are the one that have the smaller  $p$  (Spiral-F), but the intermediate one (Spiral-A) has the best space-filling scores; It seems to be the best trade-off among spirals. As expected, the maximin LCD is interesting for the polar GPs because it optimally fills the polar space, but has the worst space-filling score in Cartesian coordinates. This weakness is overcome by its modified version maxLCD\*, which seems to accomplish the best trade-off for the different criteria among all the DoEs considered.

In Table 3, we see that D-optimal designs of low order (1, 2) have in general poor scores in term of RMSE for the functions considered here that present nonlinearities. Spirals and



Table 3

Comparison of DoEs in terms of predictive performance on toy functions.

	Prediction RMSE (as percentage of the standard deviation)					
	$x(x^2 - y^2)$	$(\rho - \frac{1}{4})^2$	$\sin(2\pi\rho + \theta)$	$\frac{1+\sin(\theta)}{1+\rho^2}$	$\frac{1+x}{1+y^2}$	$\cos(3\theta)$
D-opt1	14.0	153.1	84.7	2.6	2.9	2.8
D-opt2	14.4	46.6	84.0	3.0	1.7	0.2
D-opt3	0.0	9.1	62.1	1.1	<b>0.4</b>	<b>0.2</b>
D-opt4	0.0	9.1	50.0	1.9	0.8	1.4
Spiral-F	0.0	0.4	35.2	4.0	0.7	4.5
Spiral-A	0.0	0.1	46.0	2.0	0.7	2.8
Spiral-G	0.0	<b>0.0</b>	49.1	2.9	1.3	0.8
maxLCD	0.0	0.0	31	1.0	1.0	0.8
maxLCD*	0.0	0.3	<b>23.3</b>	<b>1.0</b>	0.4	2.6

maxLCD perform rather well. maxLCD met our expectations when radial and angular patterns are dominant (functions 1, 2, 3, and 6), and the modified maxLCD\* seems to adapt well to the range of functions and models considered here, confirming its robustness among other DoEs. Finally, notice that the function  $z = \sin(2\pi\rho + \theta)$  was poorly reconstructed by all models, whatever the DoE. Indeed, either neighborhoods in  $(x, y)$  or  $(\rho, \theta)$  coordinates are not suited to capture the spiral shape of the function, and an acceptable fit would require more than 20 points.

**6. Generalization to hyperballs.** In computer experiments, the problem dimension is often higher than in spatial statistics, and the aim of this section is to investigate an extension of polar GPs in higher dimensions. More precisely, we investigate situations where the angular part of the inputs is in a higher dimension. As an example, a force vector may be represented as a pair (magnitude, direction), where direction is a point on a sphere.

**6.1. Polar GPs on hyperballs.** Let us consider that the input domain is the unit  $d$ -dimensional ball  $\mathbb{B}^d$  ( $d > 1$ ), represented either in Cartesian coordinates by  $\mathbb{B}^d = \{\mathbf{x} \in \mathbb{R}^d, \|\mathbf{x}\| \leq 1\}$ , where  $\|\cdot\|$  is the Euclidian norm, or in spherical coordinates  $(\rho, \theta_1, \dots, \theta_{d-1})$ . As in section 3, we call a Cartesian GP any restriction to  $\mathbb{B}^d$  of usual GPs on  $\mathbb{R}^d$ . Polar GPs are generalized to  $\mathbb{B}^d$  by using the product structure  $\mathbb{B}^d = [0, 1] \times \mathbb{S}^{d-1}$ , where  $\mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d, \|\mathbf{x}\| = 1\}$  denotes the  $(d-1)$ -sphere. Their kernels are obtained by combining the kernel  $k_r$  on  $[0, 1] \times [0, 1]$  and a kernel  $k_a$  on  $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}$ .

A simple way to construct kernels on  $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}$  is to restrict a kernel on  $\mathbb{R}^d \times \mathbb{R}^d$ , remarking that positive definiteness is preserved by restriction. This gives, for instance, the kernels defined with the chordal distance,  $d_1(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$ , i.e.,  $k_a(\mathbf{u}, \mathbf{v}) = \phi(d_1(\mathbf{u}, \mathbf{v}))$ , where  $(\mathbf{x}, \mathbf{x}') \mapsto \phi(d_1(\mathbf{x}, \mathbf{x}'))$  is a kernel on  $\mathbb{R}^d \times \mathbb{R}^d$ . This also includes restriction of anisotropic kernels. For example,  $(\mathbf{u}, \mathbf{v}) \mapsto \sigma^2 \exp(-\sum_{j=1}^d (\frac{u_j - v_j}{\ell_j})^2)$  defines a kernel on  $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}$ . The drawback of this construction is that it does not involve the geometry of the sphere: When distances define correlations, they lie on the Euclidian space  $\mathbb{R}^d$  and not on the sphere.

A second way is to define a kernel on  $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}$  from a distance on the sphere. The theory is well developed for *isotropic* kernels, meaning that the covariance function depends only on the geodesic distance  $d_2(\mathbf{u}, \mathbf{v}) = \arccos(\cos\langle \mathbf{u}, \mathbf{v} \rangle)$ . In this context, positive definiteness

is harder to meet. Thus, the approach used in section 3 for  $d = 2$  consisting in plugging  $d_2$  into a compactly supported correlation function, is only valid for  $d \leq 3$  [12]. For  $d \geq 4$ , conditions for positive definiteness are provided in [12]. A first option is to plug the geodesic distance  $d_2$  into a completely monotonic function, i.e., a function  $f$  admitting derivatives at any order and with alternate derivative signs:  $(-1)^m f^{(m)} \geq 0$  for all integer  $m$ . As an example,  $(\mathbf{u}, \mathbf{v}) \rightarrow \exp(-\frac{d_2(\mathbf{u}, \mathbf{v})}{\tau})$ ,  $\tau > 0$ , is a kernel on  $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}$ . Another option is to use a correlation function which admits a representation as an infinite sum of cosine powers, called Gegenbauer expansion, with strictly positive coefficients (see [12] for more details). As an example,  $\varphi_{\sin} \circ d_2$  is a kernel over  $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}$  for  $d \geq 2$  with  $\varphi_{\sin}$  defined as

$$(6.1) \quad \varphi_{\sin}(t) = 1 - \left( \sin\left(\frac{t}{2}\right) \right)^\alpha, \quad \alpha \in (0, 2).$$

**6.2. Space-filling designs on hyperballs.** We now aim at extending the space-filling designs considered in section 5.2 to hyperballs. Let us first remark that there are two difficulties in extending the space-filling LCDs. Indeed, when  $d \geq 3$  the geometry of the hypersphere  $\mathbb{S}^{d-1}$  is more complex and the mapping to a hypercube with boundary constraints (of the kind  $2\pi = 0$ ) is not clear. Furthermore, although the  $\Phi_p$  maximin criterion can be generalized, its optimization in dimension  $nd$ , where  $n$  is the design size, seems much harder when  $d$  increases. For instance, when  $d = 10$  and with the rule of thumb  $n = 10d$ , the optimization problem is in dimension 1000.

On the other hand, it is easy to simulate uniform designs on hyperspheres. A simple procedure described in [27] consists in noting that if  $\mathbf{X} \sim \mathcal{N}(0, I_d)$  then  $\mathbf{T} = \frac{\mathbf{X}}{\|\mathbf{X}\|}$  is uniform on  $\mathbb{S}^{d-1}$ . Furthermore if  $R$  is a random variable drawn independently and uniformly on  $[0, 1]$  then  $R^{1/d} \mathbf{T}$  is uniform on  $\mathbb{B}^{d-1}$  (see, e.g., [7, Theorem 2.2.1]). This extends the case of the disk (section 5.2). Notice however that uniform designs on hyperballs may not be the best designs when radial or angular patterns are present: In the 2-dimensional case, we obtained better results when the radius  $R$  was sampled uniformly (section 5.3). This suggests two strategies:

- Common part: Simulate independently  $R \sim \mathcal{U}[0, 1]$ ,  $\mathbf{T} \sim \mathcal{U}(\mathbb{S}^{d-1})$ .
- Strategy 1 “ $\mathcal{U}_{\mathbb{B}}$ ” (uniform sampling on hyperballs): Compute  $R^{1/d} \mathbf{T}$ .
- Strategy 2 “ $\mathcal{U}_r \times \mathcal{U}_{\mathbb{S}}$ ” (uniform sampling of radial and angular parts): Compute  $R \mathbf{T}$ .

**6.3. Case study on toy functions.** In order to investigate the behavior of polar GPs in a dimension higher than 2, we consider the following test functions:

- $f_1 : (x_1, \dots, x_d) \mapsto \|\mathbf{x}\|^2$ .
- $f_2 : (\rho, \theta_1, \dots, \theta_{d-1}) \mapsto \sum_{i=1}^{d-1} \cos(3\theta_i)$ .
- $f_3 : (x_1, \dots, x_d) \mapsto \left( \sum_{i=1}^d x_i \right)^2$ .

The function  $f_1$  is purely radial, and  $f_2$  purely angular. On the other hand,  $f_3$  does not exhibit any radial or angular pattern.

We perform numerical tests with  $d = 10$ . For each test function, three GP models with a constant trend are tested. Recall that  $k_m$  denotes the Matérn  $\frac{5}{2}$  kernel (see (3.6)). Then we consider the following:

- A Cartesian GP with a tensor-product kernel with a common characteristic length  $l$ :  $\prod_{j=1}^d k_m(x_j, x'_j; \ell)$ .

Table 4

Model accuracy of three GP models and two design strategies on toy functions. Each experiment is repeated 100 times, and the median of the normalized RMSE (i.e., divided by the output standard deviation) is reported as well as the interquartile interval (into brackets).

Function	$f_1(\mathbf{x}) = \ \mathbf{x}\ ^2$			$f_2(\mathbf{x}) = \sum_{i=1}^{d-1} \cos(3\theta_i)$			$f_3(\mathbf{x}) = \left(\sum_{i=1}^d x_i\right)^2$		
Kernel	kCart	kchord	kgeo	kCart	kchord	kgeo	kCart	kchord	kgeo
$\mathcal{U}_{\mathbb{B}}$	28.9 (6.2)	<b>0.0</b> (0.0)	<b>0.0</b> (0.0)	15.1 (10.0)	<b>8.1</b> (0.6)	<b>8.1</b> (0.6)	23.6 (5.3)	91.6 (8.2)	97.9 (8.6)
$\mathcal{U}_r \times \mathcal{U}_{\mathbb{S}}$	14.2 (3.1)	<b>0.1</b> (0.2)	<b>0.2</b> (0.5)	11.1 (1.1)	<b>8.2</b> (0.8)	<b>8.1</b> (0.7)	<b>17.4</b> (6.6)	34.4 (9.9)	65.4 (14)

- A polar GP based on chordal distance, with kernel  $k = k_r + k_a$ , where  $k_r = k_m$  for the radius, and  $k_a = k_m \circ d_1$  (restricted to  $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}$ ) for the angular part.
- A polar GP based on the geodesic distance, with kernel  $k = k_r + k_a$ , where  $k_r = k_m$  for the radius, and isotropic sine power kernel (6.1) for  $k_a$ .

Their kernels are denoted by  $k_{\text{Cart}}$ ,  $k_{\text{chord}}$ , and  $k_{\text{geo}}$ . Notice that the other algebraic combinations among sum, product, and ANOVA have been tried for all kernels, without modifying the conclusions. Moreover, the proposed kernels take into account the symmetry of the problem in their definition (isotropy, common parameter value per dimension) and thus depend on a very small number of parameters. The numerical likelihood maximization is then highly reliable, and was carefully done using ten different initial values.

Finally, the two design strategies presented in the previous section are applied. The design size is fixed to  $n = 10d$ . In order to assess model accuracy, the RMSE criterion is computed over a test set of size 1,000, sampled uniformly in  $\mathbb{B}^d$ . For the sake of interpretability, the RMSE is shown as a percentage of the standard deviation of the output values on the test set.

Finally, the whole study is repeated  $N = 100$  times, and the box plot characteristics of the RMSE values over the  $N$  repetitions are shown in Table 4: median and interquartile values.

We observe that polar GPs give better results for the two functions that exhibit a radial and angular pattern, and a worse result for the other one. In particular, predicting a radial function is done much more accurately with a polar GP. This may be explained by the reconstruction process and geometry considerations in high dimensions. Indeed, as in the 2-dimensional case, polar GPs reconstruct a radial function by using the points located on a closed concentric hypersphere (high angular correlation) whereas Cartesian GPs use the neighbors (in the sense of Euclidian norm) which are very few in high dimensions (see, e.g., [15, section 2.5]). Notice that even if we double the number of experiments ( $n = 20d$ ) to learn this radial function, the performance of the Cartesian GP with the best design strategy has a median RMSE equal to 5% (not shown in Table 4), which is still worse than polar GPs.

The results about design strategy on these toy functions are in favor of uniformly sampling the radial part, rather than uniformly sampling on the hyperball. Finally polar GP construction with chordal distance  $d_1$  performs better than for geodesic distance  $d_2$ . In addition to the sine power kernel, we also tested the exponential kernel  $\exp(-\frac{d_2(\mathbf{u}, \mathbf{v})}{\tau})$ , but it gave worse results. However, other kernels could have been tried with a possible different conclusion, and a deeper investigation should be done in the future.

**7. Conclusions and discussion.** We addressed the issue of analyzing costly computer or physical experiments on a disk. Such problems are encountered in various industrial applications, where the geometry of the disk is exploited for several technological processes involving rotations or diffusions from the center. For prediction purposes, we introduced so-called polar GP models that take into account the geometry of the disk both in their mean and covariance kernel. The new kernels are defined in polar coordinates. They are obtained as a combination of a kernel for the radius using a Euclidean distance, and a kernel for the angle, based on either chordal or geodesic distances on the unit circle. It was shown in two industrial examples where radial and angular patterns are visible that the approach significantly improves prediction. The best algebraic combination was found to be either a tensor product or a tensor sum, which claims in favor of using a kernel mimicking the more general ANOVA decomposition [11]. Furthermore, in these applications there were only a few differences in the results obtained with the polar GPs based on chordal or geodesic distances. This can be explained by the strong monotonic relationship between the chordal and geodesic distance. However, in theory the geodesic distance does not distort distances on the circle, and should be preferred. Finally, though not reported here, similar results were obtained with other kernel choices such as Matérn  $\frac{3}{2}$  or exponential kernels for the Cartesian GP.

We also considered the problem of designing experiments for circular domains. We introduce so-called (polar) LCDs, which correspond to LHD in polar coordinates, and are suited to phenomena that have a physical interpretation in polar coordinates. By defining a valid distance in polar coordinates, we obtained a maximin LCD. A modified LCD was proposed for more general situations, with the additional property of filling well the disk. We assessed the performances of these two designs, compared to other common DoEs of the literature, on a set of toy functions aiming at representing various non-linear patterns. As expected, the maximin LCD is outperforming the others in the case of radial or angular patterns, but loses efficiency otherwise. The modified LCD is attractive for its robustness, showing good performances over the whole range of functions and models considered here.

It is important to be precise as to when polar GPs, based on distances on the unit circle, are relevant. One main difference between polar GPs and the usual ones, called here Cartesian GPs, is about the neighborhoods used for prediction. Since kernels of polar GPs are mapped to the polar space  $(\rho, \theta)$ , the prediction at one location will particularly involve the locations corresponding to a high radial or angular correlation with respect to  $\rho$  or  $\theta$ . Typically, the resulting neighborhoods in the disk may look like pie chart sectors (high radial correlation) or ring portions (high angular correlation). This explains why polar GPs give more accurate predictions when there are radial or angular patterns, as may happen for technological processes that involve a rotation or a diffusion from the center. In other situations, involving for instance translations, Cartesian GPs may give better results. These two cases might correspond to the “two clusters of profiles over a circular grid” mentioned in [24] without any additional information about their origin. A knowledge of the process or historical data may help to choose which kernel is appropriate. In any case, there remain a lot of degrees of freedom about a GP model definition, concerning at least the trend shape or the different kernels corresponding to a given distance. To address this problem, aggregation techniques may be a solution.

Finally, we investigated an extension of the whole methodology to higher dimensions, replacing the disk by a hyperball. We performed empirical tests on several toy functions in 10

dimensions. Similar general conclusions hold, i.e., that polar GPs give better results for the functions that exhibit radial or angular patterns. In particular radial functions are much better reconstructed with polar GPs. This may be explained by the fact that in high dimensions points are located on the boundaries. Now, reconstruction with a polar GP involves points located on closed concentric hyperspheres (high angular correlation) while reconstruction with a Cartesian GP involves the neighbors, which are very few. Among other conclusions, kernels based on geodesic distance are here more difficult to handle, and on our first trials they performed worse than kernels based on chordal distance.

The discussion above concerning the model choice raises several questions about DoEs. The study presented in this paper shows the possibility of adapting existing criteria to new distances. In the situations where there is limited information about the physical process being modeled, choosing a distance may be difficult, and there is a need for building DoEs that can be suitable for any distance. In such a case, a solution would be to consider a multicriteria approach, for instance, by aggregating the maximin criteria in Cartesian and polar coordinates. On the other hand, if a specific kernel is justified, then integrated-mean-squared-error-optimal designs could be computed with respect to this kernel.

**Acknowledgments.** We gratefully acknowledge an associate editor and two anonymous referees that helped us to improve the quality of the paper; in particular section 6 which was not in the first version. The authors would like to thank the company STMicroelectronics for giving us the opportunity to work on this research project, and in particular H. Duverneuil. We also thank J. Blue, from Mines Saint-Etienne for useful discussions. Part of this work has been conducted within the frame of the ReDice Consortium, gathering industrial (CEA, EDF, IFPEN, IRSN, Renault) and academic (Ecole des Mines de Saint-Etienne, INRIA, and the University of Bern) partners around advanced methods for computer experiments. In particular we would like to thank D. Ginsbourger and N. Durrande for their relevant remarks. Some software developments were also facilitated by two R packages produced by ReDice and Dice: DiceKriging [26], kergp [6], and DiceDesign [10]. Finally we thank A. Iyassu for his help on editing.

## REFERENCES

- [1] I. ANDRIANAKIS AND P.G. CHALLENGOR, *The effect of the nugget on Gaussian process emulators of computer models*, Comput. Statist. Data Anal., 56 (2012), pp. 4215–4228.
- [2] M. BATTON-HUBERT, M. BINOIS, AND E. PADONOU, *Inverse modeling to estimate methane surface emission with optimization and reduced models: Application of waste landfill plants*, in 13th Annual Conference of the European Network for Business and Industrial, Ankara, Turkey, 2013.
- [3] G. DAMBLIN, M. COUPLET, AND B. IOOSS, *Numerical studies of space filling designs: Optimization of Latin hypercube samples and subprojection properties*, J. Simul., 7 (2013), pp. 276–289.
- [4] E. DEL CASTILLO, B.M. COLOSIMO, AND S.D. TAJBAKSH, *Geodesic Gaussian processes for the parametric reconstruction of a free-form surface*, Technometrics, 57 (2015), pp. 87–99.
- [5] H. DETTE, V.B. MELAS, AND A. PEPELYSHEV, *Optimal designs for statistical analysis with Zernike polynomials*, Statistics, 41 (2007), pp. 453–470.
- [6] Y. DEVILLE, D. GINSBOURGER, O. ROUSTANT, AND N. DURRANDE, *kergp: Gaussian Process Laboratory*, 2015, <http://CRAN.R-project.org/package=kergp>.
- [7] V. DUBOURG, *Adaptive Surrogate Models for Reliability Analysis and Reliability-Based Design Optimization*, Ph.D. thesis, Université Blaise Pascal, Clermont-Ferrand, Paris, 2011.

- [8] N. DURRANDE, D. GINSBOURGER, O. ROUSTANT, AND L. CARRARO, *ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis*, J. Multivariate Anal., 115 (2013), pp. 57–67.
- [9] N.I. FISHER, *Statistical Analysis of Circular Data*, Cambridge University Press, Cambridge, 1995.
- [10] J. FRANCO, D. DUPUY, O. ROUSTANT, G. DAMBLIN, AND B. IOOSS, *DiceDesign: Designs of Computer Experiments*, 2014, <http://www.jstatsoft.org/v65/i11/>.
- [11] D. GINSBOURGER, O. ROUSTANT, D. SCHUHMACHER, N. DURRANDE, AND N. LENZ, *On ANOVA decompositions of kernels and Gaussian random field paths*, Springer Proc. Math. Statist. 163, Springer, New York, 2016, pp. 315–330.
- [12] T. GNEITING, *Strictly and non-strictly positive definite functions on spheres*, Bernoulli, 19 (2013), pp. 1327–1349.
- [13] R.B. GRAMACY AND H.K.H. LEE, *Cases for the nugget in modeling computer experiments*, Statist. Comput., 22 (2012), pp. 713–722.
- [14] R.B. GRAMACY, J. NIEMI, AND R.M. WEISS, *Massively parallel approximate Gaussian process regression*, SIAM/ASA J. Uncertain. Quantif., 2 (2014), pp. 564–584.
- [15] T. HASTIE, R. TIBSHIRANI, AND J.H. FRIEDMAN, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Ser. Statist., Springer, New York, 2001.
- [16] P.J. HUBER, *Robust Statistics*, Wiley Series in Probability and Statistics, Wiley-Interscience, New York, 1981.
- [17] G. JONA-LASINIO, A. GELFAND, AND M. JONA-LASINIO, *Spatial analysis of wave direction data using wrapped Gaussian processes*, Ann. Appl. Stat., 6 (2012), pp. 1478–1498.
- [18] K.V. MARDIA AND P.E. JUPP, *Directional Statistics*, Wiley, Chichester, England, 2000.
- [19] G. MATHERON, *Principles of geostatistics*, Econ. Geol., 58 (1963), pp. 1246–1266.
- [20] M.D. MCKAY, R.J. BECKMAN, AND W.J. CONOVER, *A comparison of three methods for selecting values of input variables in the analysis of output from a computer code*, Technometrics, 21 (1979), pp. 239–245.
- [21] M.D. MORRIS, *Gaussian surrogates for computer models with time-varying inputs and outputs*, Technometrics, 54 (2012), pp. 42–50.
- [22] M.D. MORRIS AND T.J. MITCHELL, *Exploratory designs for computational experiments*, J. Statist. Plann. Inference, 43 (1995), pp. 381–402.
- [23] R. NAVARRO AND J. ARINES, *Complete Modal Representation with Discrete Zernike Polynomials - Critical Sampling in Non Redundant Grids*, INTECH, 2011, [doi:10.5772/24631](https://doi.org/10.5772/24631).
- [24] G. PISTONE AND G. VICARIO, *Kriging prediction from a circular grid: Application to wafer diffusion*, Appl. Stoch. Models Bus. Ind., 29 (2013), pp. 350–361.
- [25] C.E. RASMUSSEN AND C.K.I. WILLIAMS, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA, 2006.
- [26] O. ROUSTANT, D. GINSBOURGER, AND Y. DEVILLE, *DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by Kriging-based metamodeling and optimization*, J. Stat. Softw., 51 (2012), pp. 1–55.
- [27] R.Y. RUBINSTEIN, *Generating random vectors uniformly distributed inside and on the surface of different regions*, European J. Oper. Res., 10 (1982), pp. 205–209.
- [28] F. ZERNIKE, *Diffraction theory of the cut procedure and its improved form, the phase contrast method*, Physica, 1 (1934), pp. 689–704.