

# KRIGING PREDICTION FROM A CIRCULAR GRID: APPLICATION TO WAFER DIFFUSION

GIOVANNI PISTONE AND GRAZIA VICARIO

**ABSTRACT.** In the production process of silicon wafers, which are crystalline slices used as substrate of electronic micro-circuits, the thickness of the  $\text{SiO}_2$  deposition on their top is a main characteristic to be controlled during the process. The experimental design which is commonly used to monitor the thickness to the target value consists of a regular array of points lying on concentric circles, the silicon wafer itself being a disk. In order to speed up the control process, the engineers aim to use just only a limited subset of such points. To reconstruct the values on untried locations of the silicon wafer, the Kriging interpolation has been proposed because of its recognized ability in providing fairly good predictions. In this paper, we consider two methodological issues, related to universal Kriging models. First, we discuss the modeling of the covariance structure among the measured points; in fact, spatial data usually show a strong correlation when they come from spatially near observed points. Second, we put forward an algebraic method to assess the identifiability of trend models, based both on the full experimental design and on special fractions of it. Our findings are illustrated with anonymous industrial data, kindly communicated to us by professor Diego Zappa (Università Cattolica del Sacro Cuore di Milano). The same data were analyzed in the paper [2].

## 1. INTRODUCTION

The processing of silicon wafers—the basic raw material in the production of integrated circuits— involves a good deal of chemistry and physics technology. One process step is the diffusion process, where a layer of oxide material is grown or deposited onto the wafer. The oxide deposition is measured for uniformity of deposition thickness; wafers are often sorted on an automated basis according to such measure in order to increase productivity in the next process step.

We will consider a case study based on industrial data, kindly communicated to us by professor Diego Zappa (Università Cattolica del Sacro Cuore di Milano), concerning the measurements of the thickness of the  $\text{SiO}_2$  deposition during the silicon wafer's diffusion process. The control quality operators, that make available the data, use a peculiar experimental design on the circular surface of each wafer. It may be regarded as the union of three sub-designs along with a central point, see Fig. 1 and 5. Each sub-design is a set of point lying on circles with radiuses proportional to 1:2:3. In order to speed the control (economical reasons demand for it), the industrial requirement is to reduce the number of the points to be inspected, identifying a subset of the 49 points of the usual experimental design. In absence of a specific physical model of the diffusion process, one phenomenological option would be to consider a regression model, e.g. see [7]. Here, we follow mainly [2], where the authors come to a 28 points design (a fraction of the full design) and use the ordinary Kriging model, a natural candidate for the interpolation algorithm and for predicting the thickness values at untried points.

Since the ordinary Kriging interpolating model does not require a specific design of testing points, it is usual to disregard the actual study of the aliasing properties of the set of points in use. In this paper, we observe that it is actually possible to make such a study with algebraic methods for the peculiar design used in the  $\text{SiO}_2$  case study. The aim is to assess how to use

this information in specifying the Kriging model trend (regression term) to be used and finding potential issues. Moreover, we face the problem of the correlation among the observed data, being spatial data. In fact, the original Krige's idea of an existing positive correlation that decreases with increasing distance between the experimental points demands for an accurate investigation of the covariance structure and accordingly its modelization. We use a parametric estimate of the correlation function.

The aim of this study is to discuss methods to be used in the process control activity. For this reason each individual wafer is treated on his own. The actual study of the physics of the deposition process is outside the scope of this paper and it would require a completely different sampling approach.

The paper is organized as follows. The Kriging methodology is briefly reviewed in Sec. 2. In Sec. 3 we present our case study. The algebraic methodology is described and applied to the case study in Sec. 4. In turn, the results of the design are used to discuss the Kriging model in Sec. 5. Final comments conclude the paper.

## 2. THE MODEL

The choice of the model has a prominent importance in predicting the response of an experiment, either physical or simulated, in the untried points. Moreover, it is a tricky issue to decide which is the best if many models are available and a complete physical description is not available.

In the paper we consider the Kriging model because of the recognized predictive capability of this class of models. These models have been put forward in 1951 by a South African engineer, Daniel G. Krige [11], who first referred to them in analyzing mining data. He used the statistical framework in facing the problem of making accurate prediction of a response based on a set of spatial data. He stated that response values that are spatially close are much more alike than values that are more distant. Therefore, the predictions at an untried location are highly affected by the observations closer to it.

His intuition was further developed in geo-statistics by other authors. Matheron [12] proposed the Gaussian Kriging model for modeling spatial data; the treatises by Cressie [5, 6] and by Goovaerts [10] gave a systemic approach to the subject. But at the end of the 1980s, a group of statisticians led by J. Sachs [18, 17] introduced the Kriging model for the deterministic output of a computer experiment. Once more, Kriging models have been adopted in industrial metrology to drive the online construction of sequential designs for inspecting industrial parts on Coordinate Measuring Machine, see Pedone, Vicario and Romano [13].

Let us consider  $n$  design points  $\mathbf{x}_i$ ,  $i = 1, \dots, n$ , over a  $d$ -dimensional experimental region  $\mathcal{X}_d$  and the observed output  $Y_i$  at  $\mathbf{x}_i$ . The Gaussian Kriging model is

$$Y(\mathbf{x}) = f'(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}), \quad (1)$$

where  $f'(\mathbf{x}) = [f_1(\mathbf{x}) \cdots f_m(\mathbf{x})]'$  is a set of specified trend functions,  $\boldsymbol{\beta}$  is a vector of parameters, and  $Z(\mathbf{x})$  is a Gaussian random field with zero mean and Stationary Correlation Function (SCF) over  $\mathcal{X}_d$ . This model differs from a regression model because instead of assuming that the error component is independent and identically distributed with null covariance, here it is assumed to be a Gaussian process. The model (1) is referred to as universal Kriging model. The most commonly used model is the ordinary Kriging model, that assumes  $f'(\mathbf{x})\boldsymbol{\beta} = \beta$ , so that  $Y(\mathbf{x}) = \beta + Z(\mathbf{x})$ , i.e. the trend is constant.

A common choice for the correlation function is to assume

$$\begin{aligned} E(Z(\mathbf{x})) &= 0, \\ \text{Cov}(Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})) &= \sigma_Y^2 R(\mathbf{h}; \boldsymbol{\psi}), \end{aligned}$$

where  $\sigma_Y^2$  is the field variance and  $R$  is its SCF depending only on the displacement vector  $\mathbf{h}$  between any pair of points in  $\mathcal{X}_d$  and on a vector parameter  $\boldsymbol{\psi}$ . The power exponential function is the most popular among the stationary family,

$$R(\mathbf{h}; \boldsymbol{\psi}) = \sigma_Y^2 \exp \left( \sum_{j=1}^d \left| \frac{h_j}{\theta_j} \right|^{p_j} \right), \quad (2)$$

where  $\boldsymbol{\theta} = (\theta_1 \cdots \theta_d)$  are positive inverse scale parameters of order  $p_j$ ,  $\mathbf{p} = (p_1 \cdots p_d)$  are smoothness parameters, and  $\boldsymbol{\psi} = (\boldsymbol{\theta}, \mathbf{p})$ . The conditions  $0 < p_j \leq 2$ ,  $j = 1, \dots, d$ , are necessary and sufficient for (1) to be positive definite and, therefore, for the existence of a stationary Gaussian field with an SCF of that form, see Berg, Christensen and Ressel [1]. The original Krige's idea of a positive correlation between the outputs that decreases with increasing distance between their sites is true for the given SCF. However, a further parameter called *nugget* is sometimes added to the model to accomodate uncorrelated errors in the response.

**2.1. Preliminary data analysis.** Along this paper a specific data set is used, courtesy of professor Diego Zappa (Università Cattolica del Sacro Cuore, Milano), see the related paper [2]. The data are masked for confidentiality reasons: we cannot identify the source nor give any further detail on the technology. The point data, 49 training points, are taken on a regular grid of a disk displayed in Fig. 1. This peculiar experimental design  $\mathcal{D} \subset \mathcal{X}_2$  is of notable regularity. A measurement was taken at each of the 49 point in 14 wafers. In order to pinpoint the peculiarities in the  $14 \times 49$  data matrix, we resorted a run of the function `km` of the R package `DiceKriging` [15, 16]. This implementation of the Kriging method allows for a variate choice of correlation functions and it is very well documented. The graphical output is displayed in Fig. 2. Because of the lack of knowledge of the trend, the ordinary Kriging model was implemented, with a Gaussian process with a tensor product of power exponential covariances and nugget effect. Each of the 14 displays represents one experiment. The last one is a picture of the mean data values.

There is visual evidence of an interesting variability both *within* each wafer and *across* the 14 wafers, see e.g. Fig. 2. At a first glance, the wafers may be grouped into two cluster: wafers 1, 4, 9 and 12 in one cluster and the remaining wafers (2, 3, 5, 7, 8, 10, 11, 13 and 14) in another one, see typical cases in Fig. 3. The largest cluster is the less homogeneous: it might be that homogeneity and number are mutually related. In our opinion, investigation about the existence of possible external factors that differently influence the behavior should be encouraged, since the uniformity of a product is a requirement of the Quality Process Control, and this is the case of the thickness of the  $\text{SiO}_2$  deposition on wafer's top. We didn't have the opportunity to deal with the heterogeneity matter because we didn't have access to the relevant setup. Therefore, our efforts are devoted in improving the predictions of the thickness, given the data set, in view of possible application to process control.

The estimated parameter in the ordinary Kriging model are shown in Table 1 and a graphical summary is displayed in Figure 4. The nugget is not shown as it shows values much lower than  $\sigma_Y^2$ . It can be observed that the cluster A has more homogeneous parameters and larger variance than cluster B. Values shown in the table are to be considered just an indication of an overall pattern as this analysis is preliminary and not intended as estimation of the model parameters.

Provisionally, we conclude that, while there is evidence of a clustering, the ordinary Kriging is probably not adapted in this situation to describe such a clustering. Therefore, we switch to the study of a regression model for the trend effect.

## 3. ALGEBRAIC THEORY OF DESIGN

A key feature of the algebraic approach to Design of Experiment (DoE) is the representation of the set of measured points as the set of solutions of a system of algebraic equations, see the monograph [14], and the collective volume [9]. This tool is based on one side on modern Computational Commutative Algebra, see the monograph [4], and on the other side on the availability of the specialized software suites, i.e. the system CoCoA [3]. This type of software is based on the Buchberger algorithm which is able to transform system of polynomial equations in a special form called Gröbner basis.

The design  $\mathcal{D}$  of Fig. 1 may be regarded as a set of points in the complex plane  $\mathbb{C}$ . More precisely, it is the union of sub-designs  $\mathcal{D}_i \subset \mathbb{C}$ ,  $i = 0, 1, 2, 3$ , i.e. the central point and 3 set of multiples of roots of unity, see Fig. 5.

Each of the sub-designs  $\mathcal{D}_i$ ,  $i = 1, \dots, 3$  may be represented as the set of zeros of polynomial equations,

$$\begin{aligned}\mathcal{D}_1 &= \{\zeta \in \mathbb{C} : \zeta^8 = 1\}, \\ \mathcal{D}_2 &= \{\zeta \in \mathbb{C} : \zeta^{16} = 2^{16}\}, \\ \mathcal{D}_3 &= \{\zeta \in \mathbb{C} : \zeta^{24} = 3^{24}\}.\end{aligned}$$

The 3 circle's radiuses are assumed in our discussion proportional to 1:2:3, as it is indeed in the experimental grid. Thus the union design  $\mathcal{D}$  is the set of solutions of the single equation  $g(\zeta) = 0$  with

$$\begin{aligned}g(\zeta) &= z(\zeta^8 - 1)(\zeta^{16} - 2^{16})(\zeta^{24} - 3^{24}) = \\ &\quad \zeta^{49} - \zeta^{41} - 65536\zeta^{33} - 282429470945\zeta^{25} + 282429536481\zeta^{17} \\ &\quad + 18509302102818816\zeta^9 - 18509302102818816\zeta.\end{aligned}$$

This algebraic approach is fully general. However, it shows possible advantages in the reduction of computational complexity in the case of special arrays of points, such as the present one, meaning that the coefficients in the last equation are integers.

In general, given the zero set of a system of polynomial equations, the *design ideal* is defined to be the set of all polynomials which are zero on all design points. From the special Gröbner form of the polynomial system, it is easy to derive a hierarchical monomial basis of the space of all responses on the design. In our case, the monomial basis is simply  $1, \zeta, \zeta^2, \dots, \zeta^{48}$ . In fact, each given function  $y$  defined on the design  $\mathcal{D} = \{\zeta \in \mathbb{C} : g(\zeta) = 0\}$  is interpolated on its points by a polynomial of the form

$$y(\zeta) = b_0 + b_1\zeta + b_2\zeta^2 + \dots + b_{48}\zeta^{48}.$$

Besides, if the function  $y$  itself is a polynomial of degree larger or equal to 49, the computation of the interpolator is made by recursively using the rewriting relation derived from the defining polynomial  $g$ , i.e. by writing  $g(\zeta) = 0$  as

$$\zeta^{49} = \zeta^{41} + 65536\zeta^{33} + \dots - 18509302102818816\zeta.$$

The design  $\mathcal{D}$  is in the complex plane; but it can be represented in cartesian coordinates, e.g. real part and complex part along with the imaginary unit by a design ideal whose generating equations have rational coefficients.

Let  $L_i(\zeta)$ ,  $i = 0, 1, 2, 3$ , be a polynomials with rational coefficients such that  $L_i(\zeta) = 1$  if  $\zeta \in \mathcal{D}_i$  and zero if  $\zeta$  belongs to the other sub-designs. For example,

$$L_1(\zeta) = \frac{\zeta^{48}(\zeta^{48} - 2^{48})(\zeta^{48} - 3^{48})}{(1 - 2^{48})(1 - 3^{48})}.$$

Each  $L_i$  can be reduced to its normal form  $S_i$ , i.e. the polynomial of degree less or equal to 48 that takes the same values on  $\mathcal{D}$ . For example  $S_1(\zeta)$  is

$$\frac{1}{18509019673216800}\zeta^{48} - \frac{2048}{578406864788025}\zeta^{32} - \frac{94143178827}{6169673224405600}\zeta^{24} + \frac{192805230237696}{192802288262675}\zeta^8.$$

The complex conjugate  $\bar{\zeta}$  is  $\zeta$  if  $\zeta = 0$  and  $1/\zeta$  if  $\zeta$  is a root of unity. Therefore, the complex conjugate is the polynomial

$$\bar{\zeta} = \zeta S_0(\zeta) + \frac{S_1(\zeta)}{\zeta} + \frac{2^2 S_2(\zeta)}{\zeta} + \frac{3^2 S_3(\zeta)}{\zeta}, \quad \zeta \in \mathcal{D}.$$

From the expressions of  $g$  and  $\bar{\zeta}$  we derive a system of the three polynomial equations, namely

$$g(\zeta) = 0, \quad x = \frac{\zeta + \bar{\zeta}}{2}, \quad y = \frac{\zeta - \bar{\zeta}}{2}.$$

The elimination of the indeterminate  $\zeta$  produces a system of polynomial equations whose solutions are the 49 points of the design expressed in term of  $(x, y) \in \mathbb{R}^2$ . Here, we are not interested in the solutions themselves, but to use the Buckberger algorithm to derive a monomial basis of 49 monomial terms. The result is presented in Fig. 6. The terms are ideally classified into two groups. The first group is comprehensive of all the terms whose degrees are from 0 to 6; in the second group, with degrees from 7 to 12, there are terms with those degrees that are missing. There is a nice graphical representation of the monomial basis. In Fig. 6 the dots represent the monomial terms according to the degree in the cartesian axis.

This preliminary study is functional to the detection of the trend component in the Kriging model. In fact, since the trend is used to be represented with a Taylor expansion of a response surface  $z = f(x, y)$ , the monomial basis pinpoints that the terms in the Taylor expansion are identifiable up to order 6:

$$f(x, y) = \sum_{k=0}^6 \sum_{\alpha+\beta=k} \frac{1}{\alpha!\beta!} \frac{\partial^k f(0,0)}{\partial x^\alpha \partial y^\beta} x^\alpha y^\beta + R_n(x, y)$$

This result cannot be improved because the higher order terms are partially aliased with lower order terms on the given design  $\mathcal{D}$ .

As the space of all polynomial in  $x$  and  $y$  up to total degree 6 is identified by the design  $\mathcal{D}$ , any basis of such a space produce an identifiable linear model. In particular, the monomials of degree up to 6 can be orthogonalized with respect of the uniform measure on a disk producing the orthogonal system of Zernike polynomials up to order 6. Zernike polynomials, whose use has been suggested in [7], are adapted here because of the ease of interpretation of regression coefficients in the case of a circular domain and because the model matrix on our design is expected to be well conditioned as the design itself is a good discrete approximation of the uniform distribution.

Zernike's polynomials are polynomials in the ring  $\mathbb{R}[x, y]$  whose form in polar coordinates is either  $R_n^m(\rho) \cos(m\theta)$  or  $R_n^m(\rho) \sin(m\theta)$ , with  $n \geq m$ . The radial part is characterized by  $R_m^m(\rho) = \rho^m$  and assuming  $R_n^m(\rho)$ ,  $n = m, m+1, \dots$  monic of degree  $n$  and orthogonal with weight  $\rho$  on  $[0, 1]$ . In the following we use the implementation in the R package **Rfringe**, where the function **Zernike** actually computes the orthonormalized version.

**3.1. A fraction.** A fraction  $\mathcal{F}$  of the design  $\mathcal{D}$  is any non-empty subset of  $\mathcal{D}$ . In the algebraic setting, we consider the fraction ideal  $\text{Ideal}(\mathcal{F})$  of all polynomials which are zero on all points of the fraction, so that  $\text{Ideal}(\mathcal{F}) \supset \text{Ideal}(\mathcal{D})$ . In general, the coefficients of the polynomial equations describing the fraction will belong to the number field generated by the coordinates of the points of the original design. In special cases, only the coefficients will be simple, e.g. integer.

As an example, consider the two fractions in Fig. 7. The fraction with 25 points is the solution

set of the equation

$$\zeta(\zeta^4 - 1)(\zeta^8 + 2^8)(\zeta^{12} - 3^{12}) = 0$$

and the same algorithm used before produces a list of 25 monomials that is full up to degree 5. The other fraction, of the type suggested in [2], is obtained from the function `cover.design` of the R library `Fields`. In the former case the identifiability of the linear model up to degree 5 is checked by an algebraic computation while in the latter has to be checked by a simple numerical computation, e.g. evaluating the model matrix  $X$  and computing the determinant of the  $X^t X$  matrix.

**3.2. Zernike regression, all 49 points.** We conclude the preliminary analysis of the case study by analyzing the data with a regression model up to order 5. In fact, we proved that such a model is identifiable both on the full design of 49 points and on the fraction of 25 points. This is preliminary to a joint analysis of the use of the fraction together with a universal Kriging model to be done in the next section. The model matrix constructed with the Zernike polynomials on  $\mathcal{D}$  is well conditioned.

Moreover, it would be interesting to check if the regression coefficients can be used to classify the wafers. Fig. 8 gives a display of the result. Comparison with the Kriging interpolation in Fig. 2 confirms that the general behavior is similar. As the regression is a smoothing, while the Kriging is an interpolation, the image of the Kriging should show a better evidence of extreme values—compare for example the two images of wafer 2—. The detection of low values at the border and/or high values at the center could be relevant for process control.

The coefficient of the regression are displayed in Tab. 2. The interpretation of the results confirms that Zernike’s regression is adapted to this application, see Fig. 9. In fact, by inspection of the table, we observe:

- Cluster A is characterized by mode **1s1**.
- Cluster B is characterized by a combination of modes **2n0** and **4n0**.
- Modes of order 5 are not relevant.

The mode’s estimates are only approximately orthogonal as the Zernike model matrix is not orthogonal on the design points, but it is approximately orthogonal because of the uniform distribution of the sampling points on the circle.

## 4. UNIVERSAL KRIGING

The preliminary analysis of the wafer data told us a number of facts that are able to lead to a more detailed analysis:

- A full polynomial regression model of order 6 is identifiable on the design with 49 points  $\mathcal{D}$ ;
- A full polynomial regression model of order 5 is identifiable on the fraction with 25 points  $\mathcal{F}$ ;
- Zernike’s polynomial regression is adapted to the case;
- Both ordinary Kriging and polynomial regression capture the essential features of the sample of 14 wafers;

Next step shall be a combined approach, using a refined trend model together with anisotropic correlation function, that is universal Kriging.

**4.1. Full design, linear model of order 4, anisotropic power-exponential covariance, nugget.** The universal Kriging analysis with Zernike regression up to order 4 and power exponential covariance produces the predictions in Fig. 10. The R package `DiceKriging` provides a

graphical tool for analyzing the goodness of the fit with a leave-one-out method. Results in this case are in Fig. 11.

**4.2. Fraction of 25 points, linear model of order 4, power exponential covariance, nugget.** As a final analysis, we consider the universal Kriging on the 25 points of the algebraic fraction of Fig. 7 (left panel). Prediction results are shown in Fig. 12 and confirm the overall behavior of the 14 wafers that has been obtained in the previous analysis using the full design. Displays of the leave-one-out analysis are good (not shown here). The prediction errors on the unused  $49-25=24$  points are satisfying, as confirmed in Tab. 3, where the means and standard deviation of the errors in each wafer are displayed.

## 5. CONCLUSION, FINAL REMARKS AND ACKNOWLEDGMENTS

The case study consists of deposition data measured on 49 points in 14 silicon wafers. We have studied the data set in order to find a statistical treatment suitable for Process Control. A first descriptive analysis has been performed with ordinary Kriging. This statistical methodology was used here as a black-box to provide data interpolation in the spirit of Computer Experiment.

We have discussed the identification capability of the experimental design and of a fraction of 25 points by algebraic arguments. The analysis proves that a full regression model of degree 5 is identifiable on the suggested fraction. However, a Zernike regression analysis gives no evidence of notable effects on the terms of degree 5, therefore suggesting the use of degree up to 4.

While the regression model gives a good picture of the wafers, we feel that the universal Kriging with regression up to order 4, as it is done in the final analysis, is to be preferred, because it is an interpolating procedure, therefore more sensible to local deviation for normal values. This fact could be relevant in Process Control. Moreover, the Kriging allows an evaluation of correlations, which would be important for a complete study of the technology. This study is not done here, where each wafer is individually treated.

Other further developments are clearly prompted by this study. We could mention, among others: the use a correlation model based on a circular and a radial component; improvement of the estimation of the mean by the use of quadrature formulæ based on Zernike's orthogonal system and the given interpolation nodes. Generally speaking, a better statistical modeling would be useful, but would require a much extensive data set, together with the knowledge of other covariates associate with the specific technological process, such as temperature in the diffusion oven, position of the individual wafer in the oven, gas flux.

We have used the R function `lm` and the R package `DiceKriging` to perform 4 different analysis of our data set: Ordinary kriging; Zernike regression up to order 5; Universal Kriging with a regression up to order 4; Universal Kriging with a regression up to order 4 on a fraction of 25 points.

The Authors thanks professor Diego Zappa for suggesting this topic and providing the data. The Authors had a very helpful feedback from the Journal referees and the gratefully acknowledge the anonymous referees for their comments. In particular, the section of Zernike polynomial was absent in the previous version and was developed following the suggestion on one referee.

## REFERENCES

1. Christian Berg, Jens Peter Reus Christensen, and Paul Ressel, *Harmonic analysis on semigroups. theory of positive definite and related functions*, Graduate Texts in Mathematics, vol. 100, Springer-Verlag, New York, 1984. MR MR747302 (86b:43001)
2. Riccardo Borgoni, Luigi Radaelli, Valeria Tritto, and Diego Zappa, *Optimal reduction of a monitoring grid for SiO<sub>2</sub> deposition surface over a wafer for semiconductor devices*, Atti della Riunione Scientifica della Società Italiana di Statistica (Padova), 2010.

3. CoCoATeam, *CoCoA: a system for doing Computations in Commutative Algebra*, Available at <http://cocoa.dima.unige.it>, online.
4. David Cox, John Little, and Donal O'Shea, *Ideals, varieties, and algorithms: An introduction to computational algebraic geometry and commutative algebra*, second ed., Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1997. MR MR1417938 (97h:13024)
5. Noel A. C. Cressie, *Statistics for spatial data*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Inc., New York, 1993, Revised reprint of the 1991 edition, A Wiley-Interscience Publication. MR 1239641 (94h:62155)
6. ———, *Spatial prediction and ordinary kriging*, *Mathematical Geology* **20** (1997), no. 4, 407–421.
7. Alessandro di Buccianico and B.J. Janssen, *Linear models, Zernike polynomials and applications to wafer data*, ENBIS-11 Coimbra 4–8 September 2011, 2011.
8. Reinhard Furrer, Douglas Nychka, and Stephen Sain, *Fields: tools for spatial data*, 2011, v. 6.6.1 <http://www.image.ucar.edu/Software/Fields>.
9. Paolo Gibilisco, Eva Riccomagno, Maria Piera Rogantin, and Henry P. Wynn (eds.), *Algebraic and geometric methods in statistics*, Cambridge University Press, Cambridge, 2010. MR 2640515 (2011a:62007)
10. Pierre Goovaerts, *Geostatistics for natural resources evaluation*, Oxford University Press, New York, 1997.
11. Daniel G. Krige, *A statistical approach to some basic mine valuation problems on the witwatersrand*, *J. of the Chem., Metal. and Mining Soc. of South Africa* **52** (1951), no. 6, 119139.
12. Georges Matheron, *The theory of regionalized variables and its applications*, Ecole nationale superieure des mines, Paris, 1971.
13. Paola Pedone, Grazia Vicario, and Daniele Romano, *Kriging-based sequential inspection plans for coordinate measuring machines*, *Applied Stochastic Models in Business and Industry* **25** (2009), 133–149.
14. Giovanni Pistone, Eva Riccomagno, and Henry P. Wynn, *Algebraic statistics: Computational commutative algebra in statistics*, Monographs on Statistics and Applied Probability, vol. 89, Chapman & Hall/CRC, Boca Raton, FL, 2001. MR 2332740 (2008f:62098)
15. R Development Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008, <http://www.R-project.org>.
16. Olivier Roustant, David Ginsbourger, and Yves Deville, *DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization*, *Journal of Statistical Software* **51** (2012), no. 1, 1–55.
17. J. Sacks, W. J. Welch, T. J. Mitchell, and Henry P. Wynn, *Design and analysis of computer experiments (with discussion)*, *Statistical Science* **4** (1989), 409–435.
18. Jerome Sacks, Susannah B. Schiller, and William J. Welch, *Designs for computer experiments*, *Technometrics* **31** (1989), no. 1, 41–47. MR MR997669 (90e:62112)

G. PISTONE: COLLEGIO CARLO ALBERTO, MONCALIERI, ITALY

G. VICARIO: DIPARTIMENTO DI SCIENZE MATEMATICHE, POLITECNICO DI TORINO, ITALY



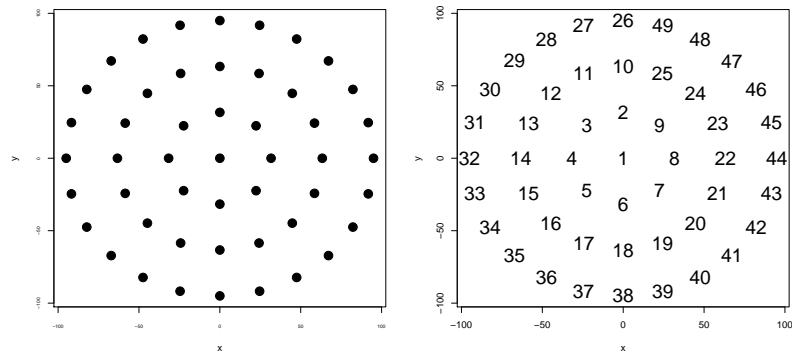


FIGURE 1. The experimental design  $\mathcal{D}$  and the ordered design points.

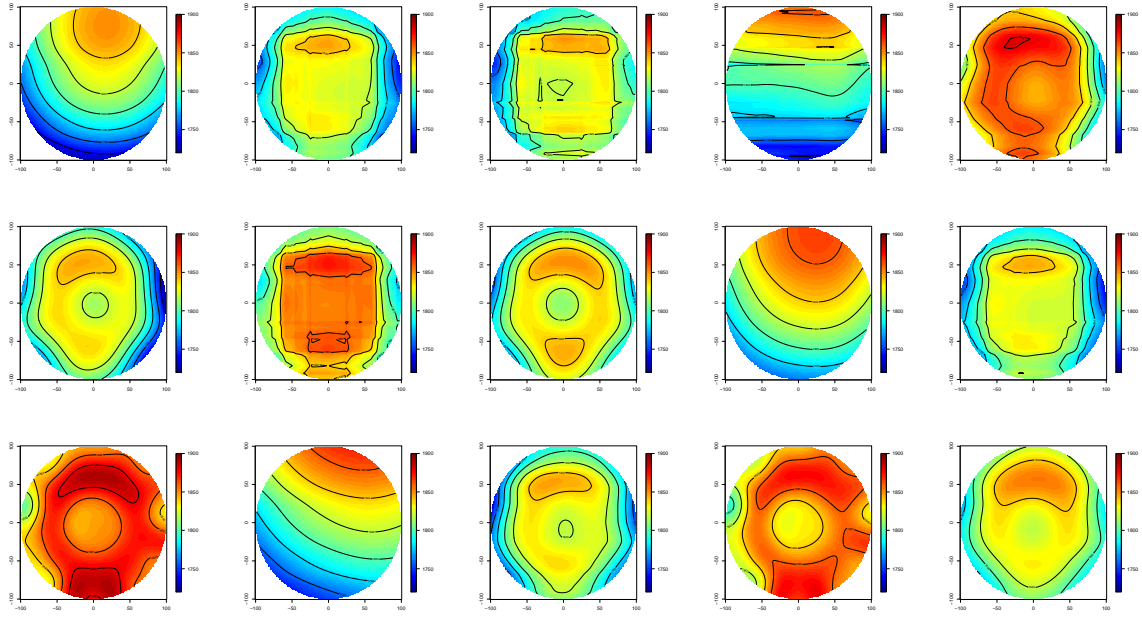
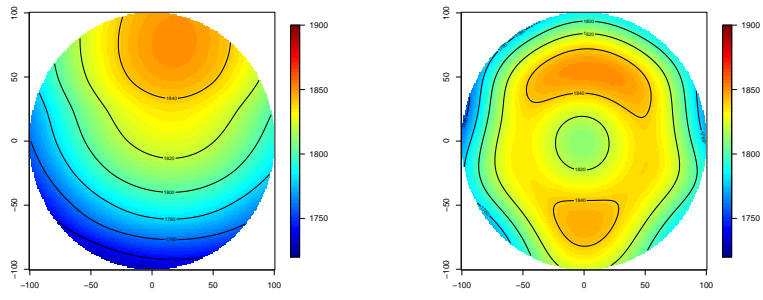


FIGURE 2. Graphical output of the run of the function `km` from the R library `DiceKriging`. Options: ordinary Kriging, tensor product of power exponential covariance, nugget. Results are displayed with the function `plot.surface` of the R library `Fields` [8].



Cluster A  
Wafers **1**, 4, 9, 12

Cluster B  
Wafers 2, 3, 5, 7, **8**, 10, 11, 13, 14

FIGURE 3. Possible clusters in the output of Fig. 2.

TABLE 1. Estimated values of parameters in the analysis show in Figures 2. The estimated nugget effect is not displayed.

|          | $\beta$ | $\theta_1$ | $\theta_2$ | $p_1$ | $p_2$ | $\sigma_Y^2$ |
|----------|---------|------------|------------|-------|-------|--------------|
| Wafer 1  | 1569.7  | 327.6      | 322.8      | 2     | 2     | 24071.0      |
| Wafer 2  | 1719.0  | 77.1       | 139.2      | 2     | 1.7   | 4672.0       |
| Wafer 3  | 1720.6  | 80.2       | 140.2      | 2     | 1.7   | 6083.7       |
| Wafer 4  | 1758.7  | 380        | 249.8      | 1.9   | 1.8   | 4659.8       |
| Wafer 5  | 1794.0  | 65.4       | 230.5      | 2     | 1.2   | 2570.6       |
| Wafer 6  | 1720.7  | 74.5       | 279.3      | 2     | 1.4   | 5278.6       |
| Wafer 7  | 1781.0  | 266.9      | 338.1      | 1.0   | 0.8   | 2821.1       |
| Wafer 8  | 1744.0  | 71.8       | 377.5      | 2     | 1.1   | 4686.8       |
| Wafer 9  | 1692.9  | 263.3      | 320.7      | 2     | 2     | 8947.8       |
| Wafer 10 | 1716.6  | 77.3       | 173.3      | 2     | 1.6   | 4971.1       |
| Wafer 11 | 1857.3  | 33.1       | 35.5       | 2     | 2     | 356.0        |
| Wafer 12 | 1762.5  | 380        | 196.2      | 1.8   | 1.9   | 4574.2       |
| Wafer 13 | 1750.4  | 217.8      | 282.1      | 1.1   | 1     | 2818.2       |
| Wafer 14 | 1833.0  | 55.6       | 376.3      | 2     | 0.8   | 1027.1       |
| Wafer M  | 1775.4  | 68.7       | 243.8      | 2     | 1.2   | 1715.1       |

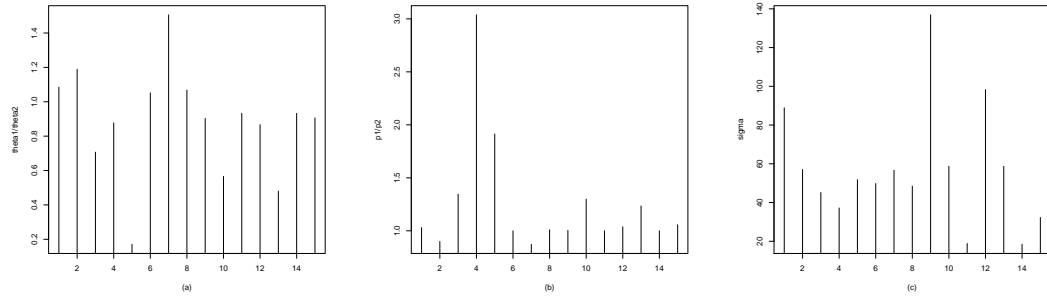
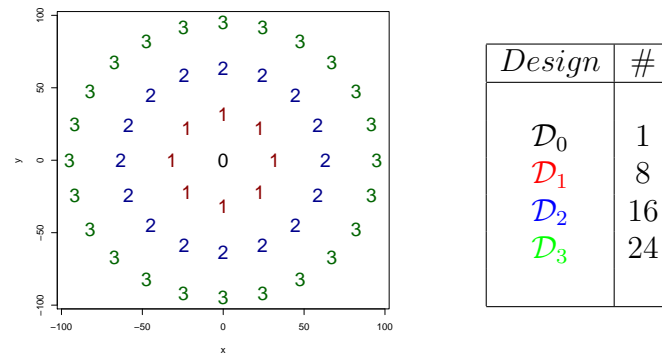


FIGURE 4. Graphical summary of Table 1. Display (a) shows the values of  $\theta_1/\theta_2$ . Display (b) shows the values of  $p_1/p_2$ . Display (c) shows the values of  $\sigma_Y^2$ .

FIGURE 5. Partitioning of the design  $\mathcal{D}$  into 4 sub-designs.

| terms  | totdeg | no terms |
|--|--------|----------|
| 1  | 0      | 1        |
| $y, x$   | 1      | 2        |
| $y^2, xy, x^2$                                 | 2      | 3        |
| $y^3, xy^2, x^2y, x^3$                         | 3      | 4        |
| $y^4, xy^3, x^2y^2, x^3y, x^4$                 | 4      | 5        |
| $y^5, xy^4, x^2y^3, x^3y^2, x^4y, x^5$         | 5      | 6        |
| $y^6, xy^5, x^2y^4, x^3y^3, x^4y^2, x^5y, x^6$ | 6      | 7        |
| $y^7, xy^6, x^2y^5, x^3y^4, x^4y^3, x^5y^2$    | 7      | 6        |
| $y^8, xy^7, x^2y^6, x^3y^5, x^4y^4$            | 8      | 5        |
| $y^9, xy^8, x^2y^7, x^3y^6$                    | 9      | 4        |
| $y^{10}, xy^9, x^2y^8$                         | 10     | 3        |
| $y^{11}, xy^{10}$                              | 11     | 2        |
| $y^{12}$                                       | 12     | 1        |

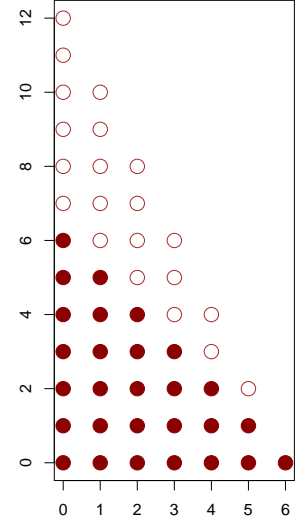


FIGURE 6. On the left, the identifiable monomials in cartesian coordinates on the design  $\mathcal{D}$ . On the right, a graphical display of the exponents of identified monomials. Filled dots belong to the full model of degree 6.

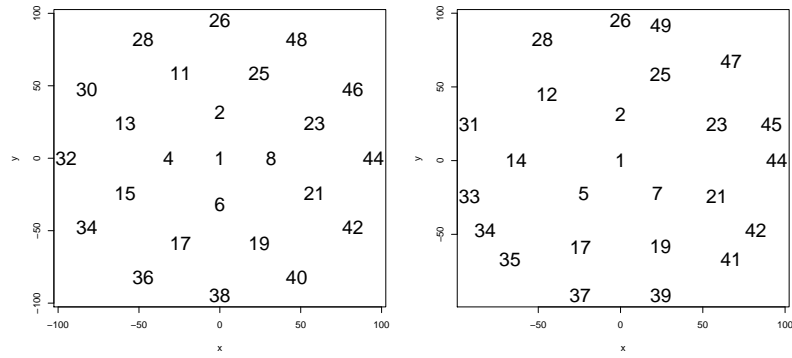


FIGURE 7. The left display shows a fraction with 25 points which has a simple algebraic description. The right display shows a fraction with has 28 points, a complicated algebraic description, but possibly better covering properties.



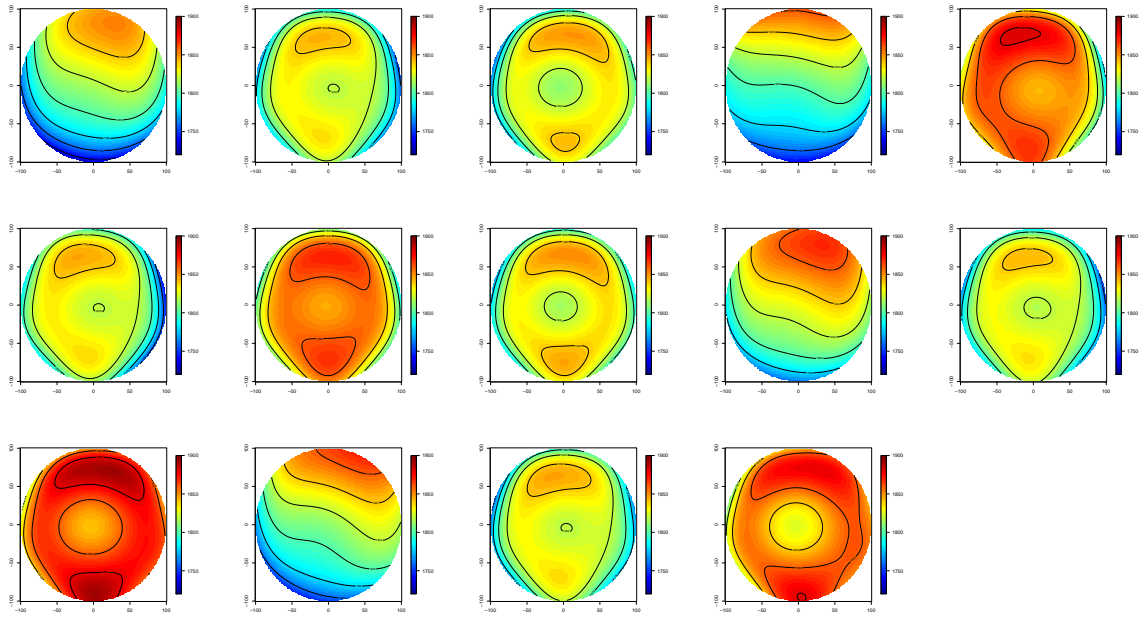


FIGURE 8. Graphical output of the run of the R function `lm`. The regression model includes all Zernike polynomials up to order 5. We use the orthonormal Zernike polynomials provided by the R library `Rfringe`. Results are displayed with the function `plot.surface` of the R library `Fields`. The mean wafer is not considered here.

TABLE 2. Coefficients (modes) of the Zernike regression. The coding of columns is the order  $n$ , followed by a letter for the type of trigonometric function (**s** for sin, **c** for cos, **n** for none i.e.  $\cos 0$ ), followed by the integer  $m$ .

|    | 1c1  | 1s1  | 2c2  | 2n0   | 2s2  | 3c3  | 3c1  | 3s1  | 3s3  | 4c4  | 4c2  | 4n0   | 4s2  | 4s4  | 5c5  | 5c3  | 5c1  | 5s1  | 5s3 | 5s5  |
|----|------|------|------|-------|------|------|------|------|------|------|------|-------|------|------|------|------|------|------|-----|------|
| 1  | 7.4  | 25.7 | -2.4 | -9.2  | 0.7  | -0.1 | 0.1  | 1.4  | -1.3 | -1   | -0.6 | -3.3  | 0    | -0.1 | 0    | 0.1  | 0.2  | 0    | 0.1 | -0.7 |
| 2  | -4.3 | 2    | -7.5 | -10.9 | 0.8  | -0.2 | 0.2  | -1.9 | 1    | 0.6  | -0.1 | -8.5  | -0.2 | 0    | -0.3 | 0    | 0.2  | -0.8 | 0.1 | -2.3 |
| 3  | 4.8  | 1.1  | -5.6 | -5.5  | 0.5  | -0.2 | 2    | -2.5 | 1.1  | 0.9  | 0.5  | -10.4 | 0    | 0.1  | -0.2 | -0.2 | 0.9  | -1.1 | 0.1 | -2.1 |
| 4  | 2.1  | 26   | -2.1 | -1    | 0.6  | -0.2 | -1.7 | 2.2  | -1.4 | -0.9 | -0.5 | -2.1  | 0    | -0.1 | 0    | 0    | -0.2 | 0.5  | 0.1 | -0.7 |
| 5  | -8.2 | 3.7  | -8.1 | -5.6  | 0.4  | 0    | -1.7 | -3.2 | 0.9  | 0.6  | -0.1 | -8.6  | -0.1 | 0    | -0.1 | 0    | -0.5 | -1.1 | 0.1 | -2.8 |
| 6  | -7.5 | 3.5  | -7.7 | -10.7 | 0.3  | -0.1 | -1.6 | -0.9 | 1    | 0.8  | -0.2 | -8.7  | -0.1 | 0    | -0.3 | -0.1 | -0.2 | -0.4 | 0.1 | -2.5 |
| 7  | -0.4 | -0.4 | -6.9 | -10.3 | 0.5  | -0.2 | -0.9 | -2.3 | 1.1  | 0.8  | 0.1  | -9.2  | 0    | 0.2  | -0.2 | -0.1 | 0    | -0.9 | 0.1 | -2.4 |
| 8  | 2.3  | 0.3  | -6.1 | -6    | 0.3  | -0.1 | 0.3  | -2.4 | 1    | 0.9  | 0.4  | -10.3 | -0.1 | 0.2  | -0.2 | -0.1 | 0.4  | -1   | 0.2 | -2.1 |
| 9  | 7.2  | 23.5 | -2.1 | -6.1  | 0.5  | -0.2 | -0.1 | 0.5  | -1.1 | -1.3 | -0.6 | -3.3  | 0    | -0.1 | 0.1  | 0.1  | 0.3  | -0.1 | 0.1 | -0.7 |
| 10 | -3.5 | 1.2  | -7.7 | -10.5 | 0.6  | -0.2 | 0.1  | -2.4 | 1.1  | 0.6  | -0.1 | -8.6  | -0.1 | 0    | -0.3 | -0.1 | 0.3  | -1   | 0.1 | -2.4 |
| 11 | 4.3  | 0.8  | -6.6 | 2.9   | 0.5  | -0.1 | 1.5  | -2.7 | 1.2  | 0.8  | 0.4  | -9.5  | 0    | 0.1  | -0.2 | -0.1 | 0.8  | -1.1 | 0.2 | -2.2 |
| 12 | 10.3 | 25.3 | -2   | -2    | -0.5 | 0    | 0.9  | 1.6  | -1.1 | -1   | -0.7 | -2.9  | -0.1 | -0.2 | -0.1 | 0.1  | 0.4  | 0.1  | 0.1 | -0.6 |
| 13 | -1.5 | 3    | -7.9 | -10.2 | 0.7  | -0.2 | 1.7  | -1.8 | 1.1  | 0.7  | 0    | -8.5  | 0.2  | 0    | -0.3 | -0.1 | 0.8  | -0.8 | 0.1 | -2.4 |
| 14 | 4.6  | 1.4  | -6.3 | 5.1   | 0.7  | -0.3 | 2.3  | -2.2 | 1.3  | 0.8  | 0.4  | -9.1  | 0    | 0.1  | -0.3 | -0.2 | 1.2  | -1   | 0.2 | -2.2 |

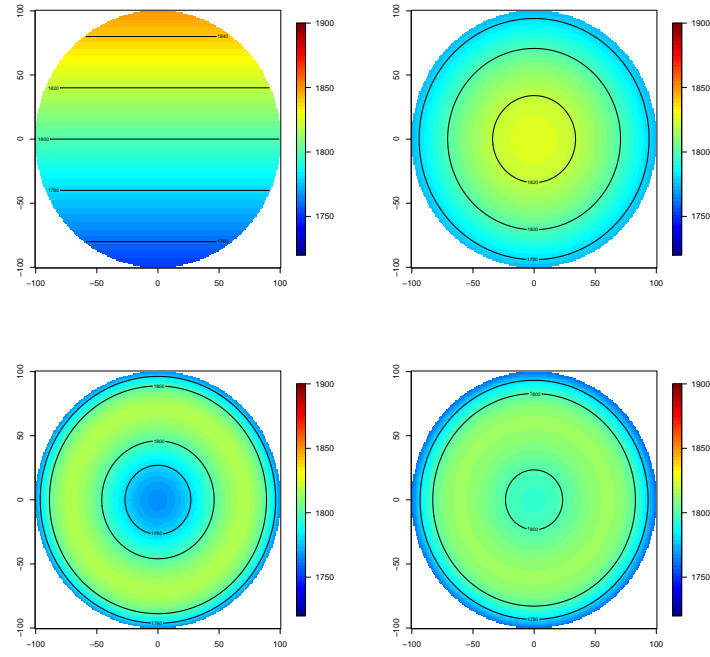


FIGURE 9. Visual illustration of the type contribution of modes  $1s1$ ,  $2n0$ ,  $4n0$ ,  $2n0+4n0$ .

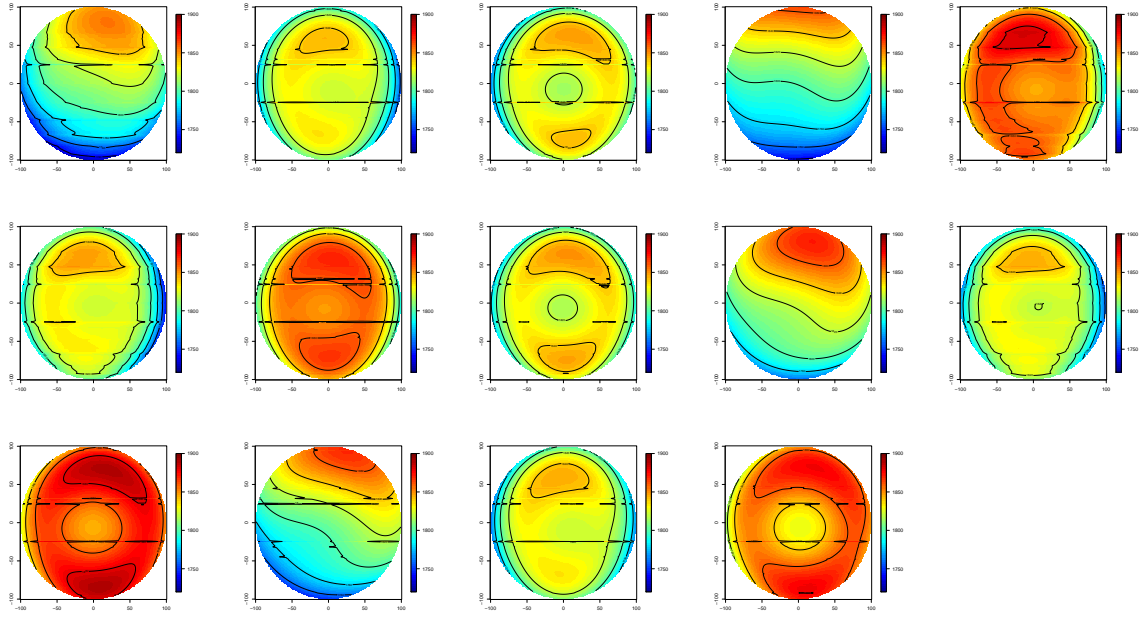


FIGURE 10. Graphical output of the run of the function `km` of the R library `DiceKriging` with Zernike regression of order 4, anisotropic power exponential correlation, nugget. Note the instability of level curves.

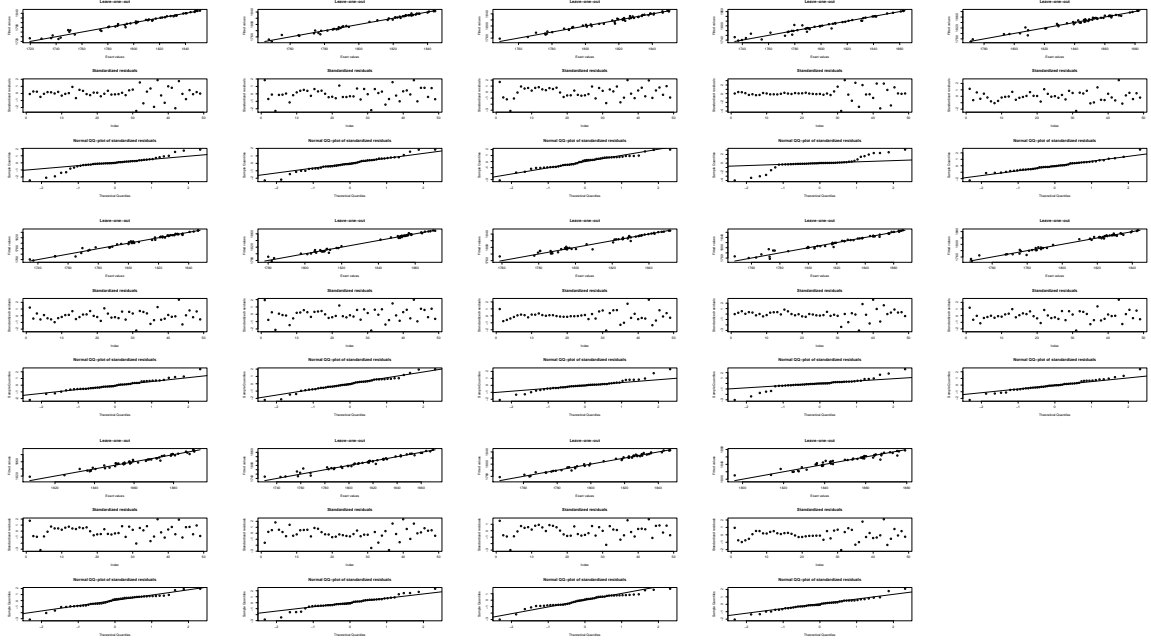


FIGURE 11. Leave-one-out-results of the case in 10. For each 14 wafers, 3 plots are presented: the residuals at each of the 49 location (top), the standardized variance of residuals (middle), the QQ plot of residuals (bottom).

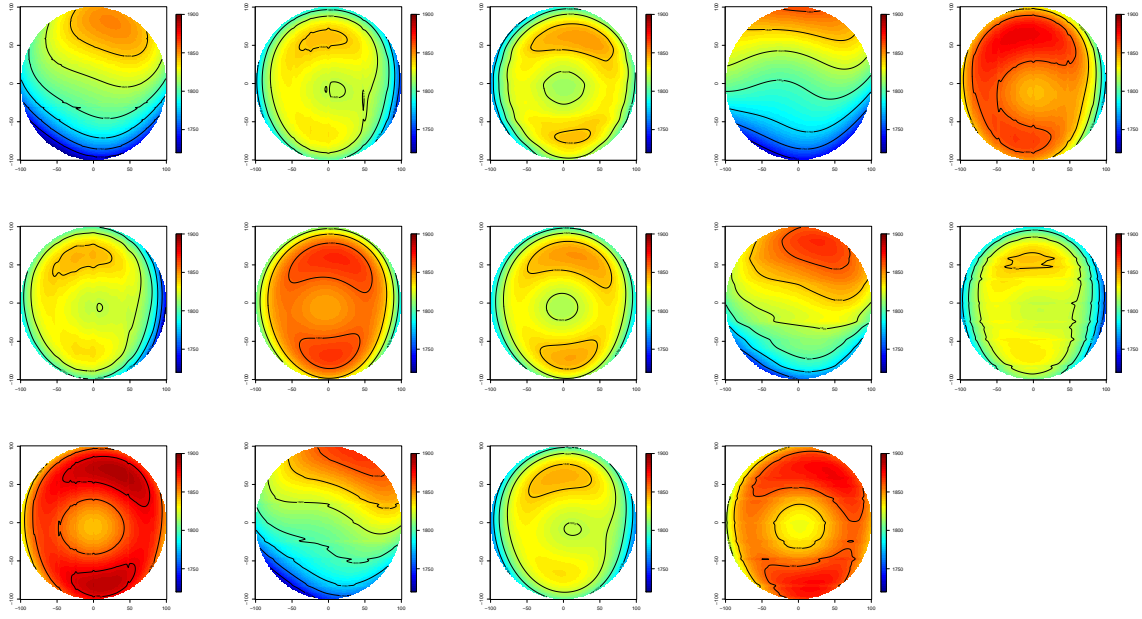


FIGURE 12. Graphical output of the run of the function `km` of the R library `DiceKriging` with Zernike regression of order 4, anisotropic power exponential correlation, nugget, on the 25 points of the algebraic fraction in Fig. 7 (left panel).

TABLE 3. Mean and standard deviation of the prediction error on the 24 unused points.

|      | W 1  | W 2  | W 3 | W 4  | W 5  | W 6  | W 7  | W 8 | W 9  | W 10 | W 11 | W 12 | W 13 | W 14 |
|------|------|------|-----|------|------|------|------|-----|------|------|------|------|------|------|
| mean | 1.4  | 0.4  | 0.4 | 1.4  | 0.4  | 0.5  | 0.7  | 0.7 | 1.7  | 0.2  | 0.7  | 1.6  | 0.2  | 0.6  |
| sd   | 14.8 | 10.1 | 9.3 | 15.8 | 11.8 | 10.6 | 10.2 | 9.8 | 15.7 | 10.2 | 9.9  | 15.2 | 10.2 | 9.7  |