Statistical methods applied in microelectronics

Università Cattolica del Sacro Cuore
June 13th, 2011

Fractions of a grid on a disk: their features in Kriging prediction

Giovanni Pistone Grazia Vicario

Collegio Carlo Alberto POLITECNICO DI TORINO

June 13, 2011

Preliminary abstract

In the paper we consider Gaussian responses on a regular grid of a disk. This peculiar experimental design appears in production process of wafers, surfaces of disks used in electronic equipments, in order to monitor the thickness of the SiO_2 deposition on their top. The circular symmetry of the used design prompts for an algebraic description of the Gaussian responses at each point of the grid. We discuss the relevant mathematical properties in order to:

1. describe Gaussian processes on a circular grid;
2. study the properties of fractions of the grid and their aliasing properties;
3. derive an algebraic form of the Kriging prediction in the unobserved points.

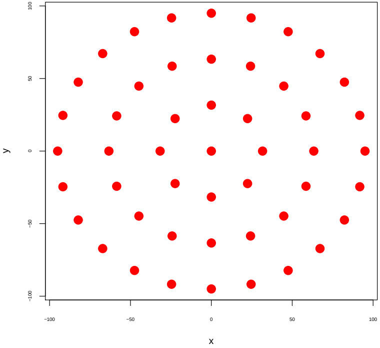
Plan

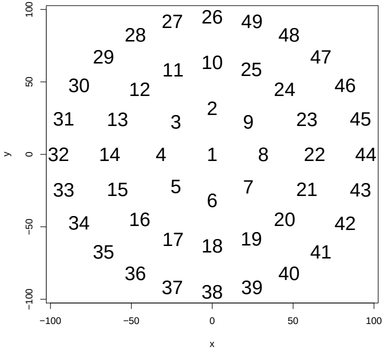
1. The data: training set and measurement.
2. The training set: algebraic theory of design.
3. Fractions of the training set and Kriging models.
4. Models for correlation.
5. Conclusions and future work.

- Seminal paper Riccardo Borgoni, Luigi Radaelli, Valeria Tritto, and Diego Zappa. Optimal reduction of amonitoring grid for SiO_2 deposition surface over a wafer for semiconductor devices. In *Atti della Riunione Scientifica della Società Italiana di Statistica*, Padova, 2010
- Algebraic Statistics Giovanni Pistone, Eva Riccomagno, and Henry P. Wynn. *Algebraic statistics*, volume 89 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, FL, 2001. Computational commutative algebra in statistics
- Algebraic Statistics in Kriging Giovanni Pistone and Grazia Vicario. Comparing and generating latin hypercube designs in kriging models. *ASTA ADVANCES IN STATISTICAL ANALYSIS*, 94:353–366, 2010
- Covariance modeling in technology Grazia Vicario, Suela Ruffa, Giusy Donatella Panciani, Francesco Ricci, and Giulio Barbato. Form tolerance verification using the kriging method. *Statistica Applicata*, 2011. Submitted

The training set

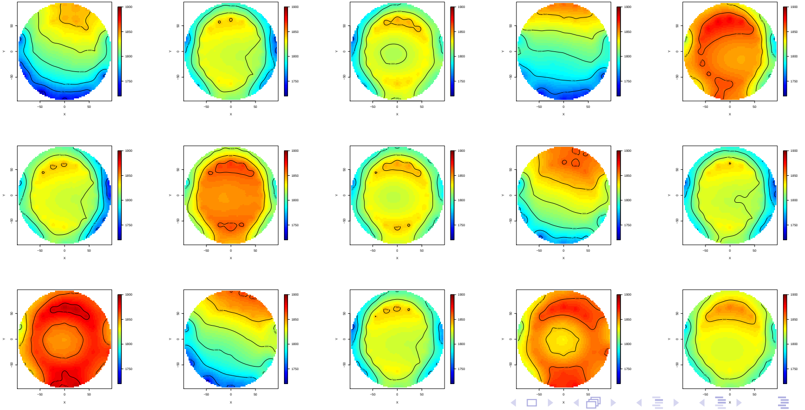
- The 49 training points for a design $\mathcal{D}$ of notable regularity.
- A measurement is taken at each point.
- 14 experiments are available, courtesy of professor Diego Zappa.





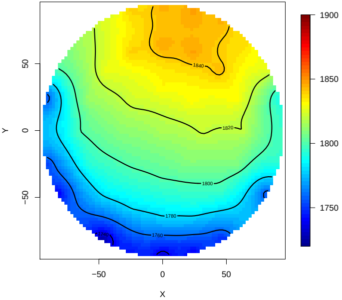
The data

- The 49×14 data matrix is represented below from a run of the function `Krig` of the R package `Fields`.
- The model is $Y(\zeta) = \beta + Z(\zeta)$, Z Gaussian process with exponential covariance.
- Each of the 14 displays represents one experiment. The last one is a picture of the mean data values.

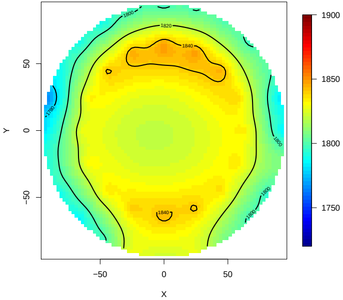


Clusters

- There is evidence of an interesting behavior both inter and infra.



Cluster A
Wafers 1, 4, 9, 12

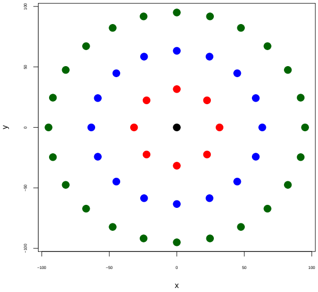


Cluster B
Wafers 2, 3, 5, 7, 8, 10, 11, 13, 14

- Cluster B is less homogeneous than cluster A
- Clusters could come from external factors.

Algebraic theory of design I

- A key feature of the algebraic approach to DoE is the representation of the set of points as the set of solutions of a system of algebraic equations.
- The design $\mathcal{D}$ is the union of sub-designs $\mathcal{D}_i$, $i = 0, 1, 2, 3$, i.e. the central point and 3 set of multiples of roots of unity.



Design	#
$\mathcal{D}_0$	1
$\mathcal{D}_1$	8
$\mathcal{D}_2$	16
$\mathcal{D}_3$	24

- The computations are conveniently performed with a symbolic software. E.g. CoCoA, from the University of Genoa

Algebraic theory of design II

- If the points of the plane are coded with complex numbers, the design $\mathcal{D}$ is the set of solutions of the equation $g(\zeta) = 0$ with

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

- The design ideal is the set of polynomials which are zero on the given points. In the example, these are the polynomials that are divisible by g.
- The possible advantage of this algebraic approach is a reduction of complexity with special arrays of points. In such a case it is easy to compute a monomial basis of the space of all responses on the design. In the example, the monomial basis is $1, \zeta, \zeta^2, \ldots, \zeta^{48}$.

Algebraic theory of design III

- In our example, each given function defined on $\mathbb{C}$ is interpolated on the design $\mathcal{D} = \{\zeta \in \mathbb{C} \colon g(\zeta) = 0\}$ by a polynomial of the form

$$b_0 + b_1\zeta + b_2\zeta^2 + \cdots + b_{48}\zeta^{48}.$$

- In particular, if the function 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 g, i.e. $\zeta^{49} = \zeta^{41} + 65536\zeta^{33} + \cdots$.
- For example, the indicator polynomial of $\mathcal{D}_1 = \left\{\zeta^8 = 1\right\}$ in $\mathcal{D}$ is

$$\frac{\zeta^{48}(\zeta^{48} - 2^{48})(\zeta^{48} - 3^{48})}{(1 - 2^{48})(1 - 3^{48})} = \frac{1}{18509019673216800}\zeta^{48} - \frac{2048}{578406864788025}\zeta^{32} - \frac{94143178827}{6169673224405600}\zeta^{24} + \frac{192805230237696}{192802288262675}\zeta^{8}$$

Algebraic theory of design IV

- The design $\mathcal{D}$ can be represented in cartesian coordinates, e.g, real part, complex part $\times$ the imaginary unit as an ideal of points whose generators have rational coefficients. Note the increase in combinatorial complexity, from integers to rationals.
- A monomial basis of 49 terms is:

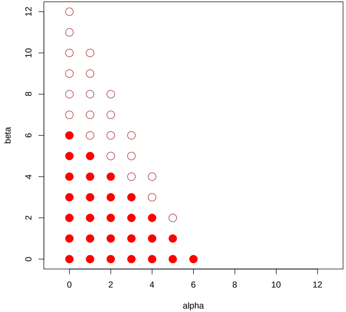
$$1; \quad y, x; \quad y^2, xy, x^2; \quad y^3, xy^2, x^2y, x^3;$$
$$y^4, xy^3, x^2y^2, x^3y, x^4; \quad y^5, xy^4, x^2y^3, x^3y^2, x^4y, x^5;$$
$$y^6, xy^5, x^2y^4, x^3y^3, x^4y^2, x^5y, x^6;$$

and

$$y^7, xy^6, x^2y^5, x^3y^4, x^4y^3, x^5y^2;$$
$$y^8, xy^7, x^2y^6, x^3y^5, x^4y^4; \quad y^9, xy^8, x^2y^7, x^3y^6;$$
$$y^{10}, xy^9, x^2y^8; \quad y^{11}, xy^{10}; \quad y^{12}.$$

Response surface model

- The available monomials are



- The Taylor expansion of a response surface $z = f(x, y)$ up to order n is

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

- Terms in the Taylor expansion are identifiable up to order $n = 6$.

Example of symbolic computation

```
Use R::= Q[z,x,y]; -- Ring
Eqs := [z^8-1, 2x-(z+z^7), 2y-(z-z^7)];
I := Ideal(Eqs); -- Ideal of points
QuotientBasis(I); -- Monomial basis
J := Elim(z,I); -- Elimination
Use S::=Q[x,y]; -- New ring
D:=Ideal(BringIn(Gens(J)));
ReducedGBasis(D); -- Groebned basis
QuotientBasis(D); -- Monomial basis
```

- Design in complex notation Giovanni Pistone and Maria Piera Rogantin. Algebraic statistics of level codings for fractional factorial designs. *J. Statist. Plann. Inference*, 138(1):234–244, 2008.

Response surface model + Gaussian field

- The list of terms in the monomial basis is complete up to the degree 6. This suggests the safe use of a response surface model up to degree 6.
- This result cannot be improved because the first missing term of degree 7 is $x^6 y$ which is aliased with

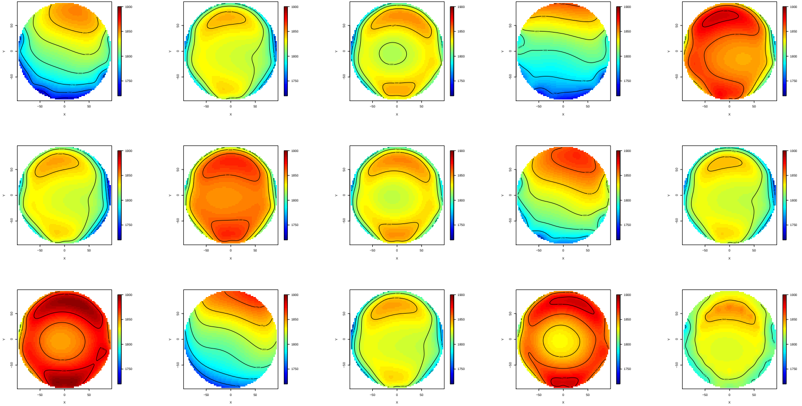
$$3x^4y^3 - 3x^2y^5 + y^7 + 6x^4y - 12x^2y^3 + 6y^5 - 11x^2y + 11y^3 + 6y.$$

- The model is

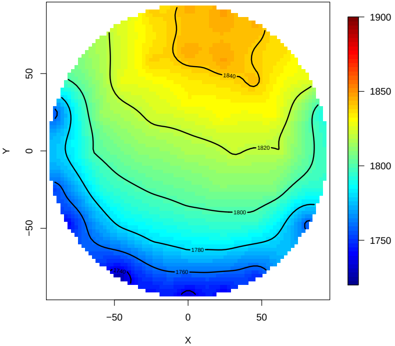
$$Y(x,y) = \sum_{k=0}^{6} \sum_{\alpha+\beta=k} x^\alpha y^\beta \eta(\alpha,\beta) + Z(x,y)$$

Degree 6, $\theta = 20$, $p = 1$

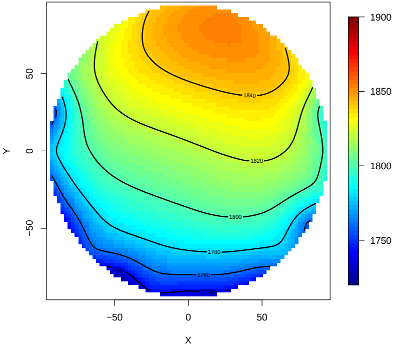
- The following panel is the image representation of the Kriging prediction with a full regression model of degree 6, exponential isotropic covariance with scale parameter $1/20 = .05$ and smoothness parameter $p = 1$.



Degree 0 vs degree 6: Wafer 1



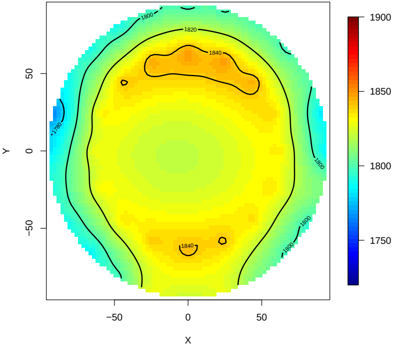
Degree 0



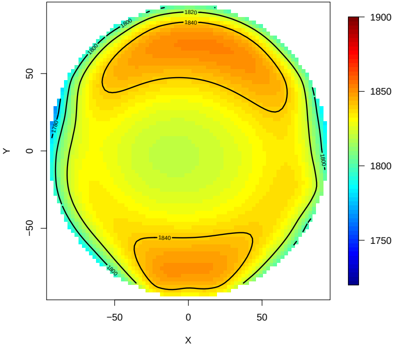
Degree 6

- Degree 6 is smoother

Degree 0 vs degree 6: Wafer 8



Degree 0

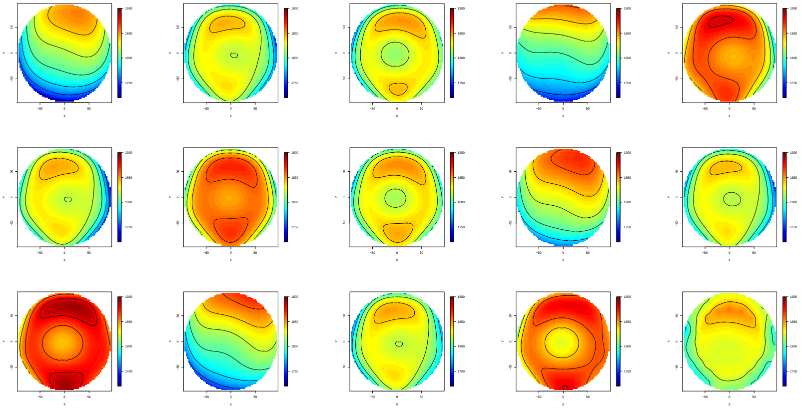


Dergee 6

- Degree 6 is smoother

Degree 6, $\theta = 0.2$

- The following panel is the image representation of the Kriging prediction with a full regression model of degree 6 and delta covariance i.e. scale parameter $1/.2 = 5$.

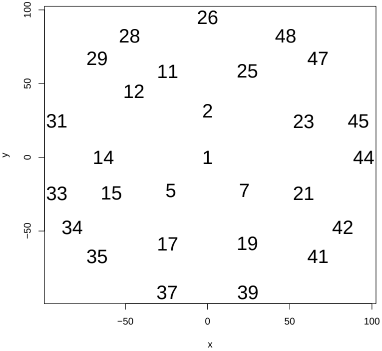


Criteria for a sub-sample

- The experiment consists in observing the response in any point of the wafer surface and the aim is to investigate how the response changes varying the locations: the two cartesian axis directions are considered as two factors that affect the response.
- If the designs proposed by DoE considering principles as replication, blocking, randomization, orthogonality and optimality of different alphabet letters are suitable for estimating the trend term of the Kriging model, the Gaussian field term is better estimated by the space-filling designs, i.e. designs based on measures of distance between locations (to quantify how evenly the points are spread out), designs based on the evenness of the locations throughout the region, etc.

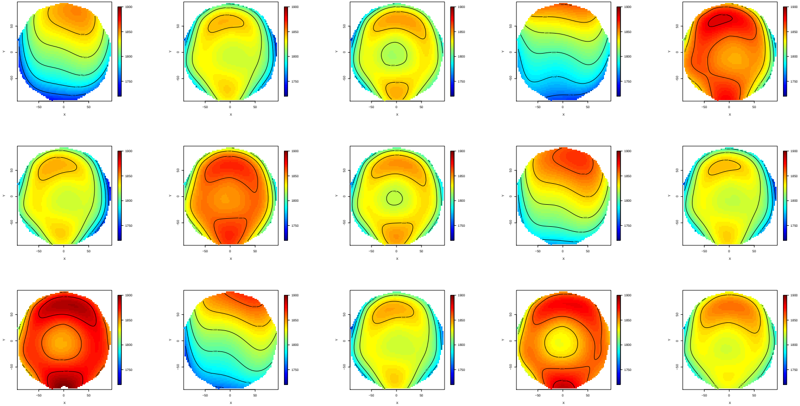
Minimax sample: random search

- The response surface argument suggests to try a random sample of 28 points.
- A random sample is generated with the function `cover.design` of the library `fields`.



The data: 28 points, degree 5, $\theta = 20$

- The following panel is the image of the Kriging prediction with 28 points, a full regression model of degree 5 and exponential isotropic covariance with scale parameter $1/20 = .05$ and smoothness parameter $p = 1$.
- Degree 6 is not identifiable.



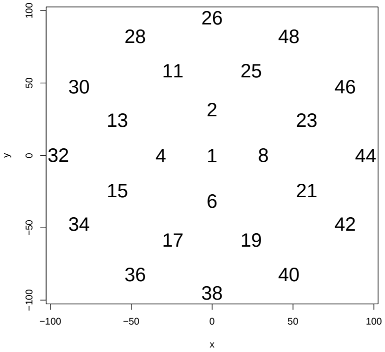
Algebraic sample of 25 points

- If $\zeta^8 = 1$, then $0 = \zeta^8 - 1 = (\zeta^4 - 1)(\zeta^4 + 1)$, so that

$$\mathcal{D}_1 = \left\{\zeta \in \mathbb{C} \colon \zeta^4 = -1\right\} \cup \left\{\zeta \in \mathbb{C} \colon \zeta^4 = 1\right\}.$$

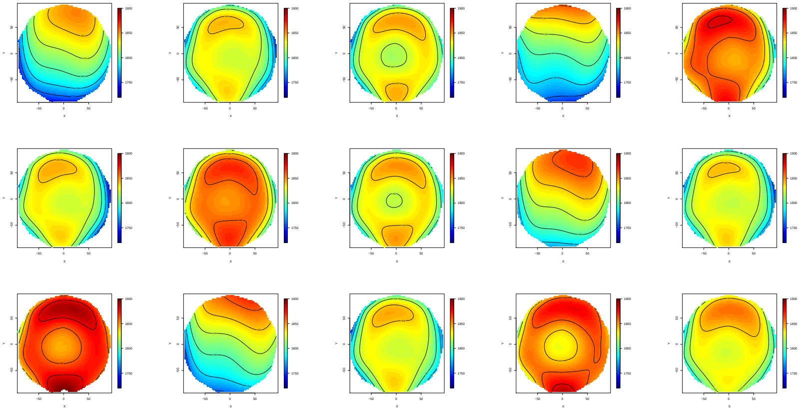
- Similarly on $\mathcal{D}_2, \mathcal{D}_3$.
- We define a fraction of $\mathcal{D}$ with the equation

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



Kriging with the algebraic sample

- The following panel is the image representation of the Kriging prediction with 25 points, a full regression model of degree 5 and exponential covariance.



Why Kriging with spatial data

- Spatial data are not spatially independent in most of the physical settings. They show a strong correlation when the data come from spatially near observed points; and the correlation vanishes when the points are far away from each other.
- The Kriging model is a parametric gaussian statistical model with state-space representation Y of the form

$$Y(\zeta) = \mathbf{f}^T(\zeta)\boldsymbol{\eta} + Z(\zeta), \quad \zeta \in \mathcal{R} \subset \mathbb{R}^2 \simeq \mathbb{C},$$

  where

  - $\mathbf{f}$ is a given linear model with regression parameters $\boldsymbol{\eta}$;
  - Z is a centered gaussian random field and Z is stationary with respect to translations

$$\mathrm{Cov}\,(Z(\zeta_1), Z(\zeta_2)) = \sigma_Z^2 R(\zeta_1 - \zeta_2);$$

  - σ_Z^2 is the field variance, R is the real auto-correlation function depending on the displacement vector $h = \zeta_1 - \zeta_2$ only, $R(\zeta_1 - \zeta_2) = R(h)$, $R(h) = R(-h)$, $R(0) = \sigma_Z^2$.

Modeling the variance

- The spatial correlation has to be measured and modeled; George Matheron (1962) has introduced the use of the variogram

$$2\gamma_Y(\zeta_1 - \zeta_2) = \mathrm{Var}\left(Y(\zeta_1) - Y(\zeta_2)\right),$$

  γ_Y being the semivariogram.

- In Matheron's theory the variogram is assumed to be continuous at 0; if it is not the case, we have a nugget effect which is possibly the case in our working example.

- Let us consider n spatial locations

$$\mathcal{D} = \left\{\zeta_i = (x_i, y_i) \colon (x_i, y_i) \in \mathcal{R} \subset \mathbb{R}^2\right\}$$

  in the region $\mathcal{R}$, together with the observations $Y(\zeta_i)$.

- $Y_{|\mathcal{D}} = (Y(\zeta_1), Y(\zeta_2), \ldots, Y(\zeta_n))$ is a sample from a Gaussian field with zero expected value and covariance $\mathrm{Cov}\left(Y(\zeta_i), Y(\zeta_j)\right)$ depending on the undirected displacements $\overline{\zeta_i \zeta_j} \in \binom{\mathcal{D}}{2}$ only.

- Seminal book: Georges Matheron. *Traité de géostatistique appliqué*. Number 14 in Mem. Bur. Rech. Geog. Minieres. Editions Technip, 1962.
- Discussion: `http://www.mail-archive.com/ai-geostats@jrc.it/msg01298.html`
- R: `http://www.mail-archive.com/ai-geostats@jrc.it/msg01298.html`

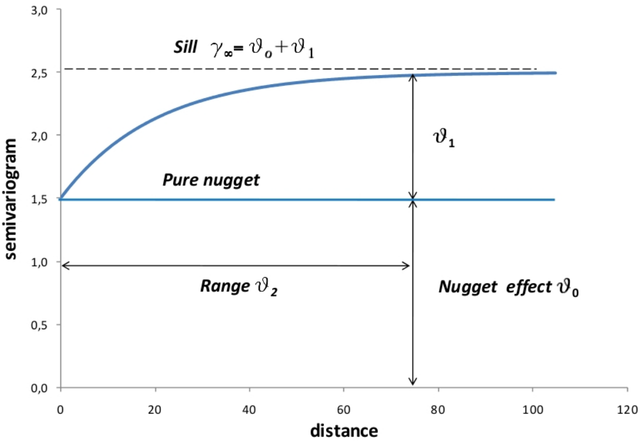
Variogram I

- Matheron suggested an empirical estimator of the variogram

$$2C(h) = \frac{1}{\#\binom{\mathcal{D}}{2}} \sum_{\overline{\zeta_1\zeta_2} \in \binom{\mathcal{D}}{2}:\ h=\zeta_1-\zeta_2} |Y(\zeta_1) - Y(\zeta_2)|^2$$

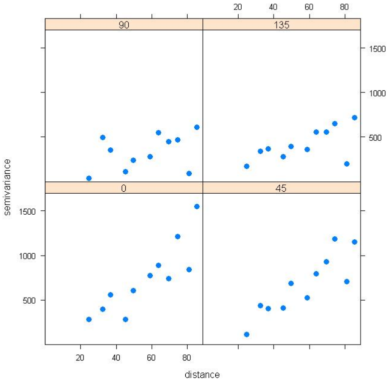
- This estimator is unbiased if the mean is constant; it lacks robustness as it is badly affected by outliers due to the presence of the ^{2} term.
- Materon discusses the following properties of the variogram.
  1. It is positive except at the origin where it is zero.
  2. It is assumed to be continuos at the origin.
  3. If not, it does not approach 0 as the distance between ζ_i and ζ_j approaches zero, the discontinuity produces the nugget effect.
  4. The presence of the nugget effect is usually due to measurement error: repeating the measurements a number of time, the measured values tend to fluctuate around the true value.

Variogram II

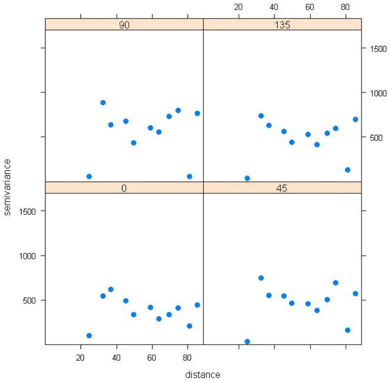


Two classes of variograms

- The 14 variograms obtained with the R library `Gstat` confirm the classification already obtained.
- The variograms appears to be inconsistent with an isotropy assumption, i.e. $R(h) = R(\|h\|)$.
- A nugget effect does not depend on measurement errors.



Variogram of wafer 1



Variogram of wafer 8

Conclusions and future work

- Given the Kriging model is favored, the two components, trend and stochastic process, demand a careful specification and choice.
- The algebraic theory is a very useful tool in detecting the response surfaces for modeling the trend.
- On the other hand, the use of the variograms, as geostatisticians do, is advisable for modeling a possible covariance structure.
- Also the choice of the experimental setting is crucial: a mixture between principles proper of DoE and the space filling designs seem to benefit the predictions.
- Further investigations are demanded for less regular designs and for the identification of the correlation function with possible anisotropy.