XXIII IMEKO World Congress “Measurement: sparking tomorrow’s smart revolution”
August 30 - September 3, 2021, Yokohama, Japan

GENERATION OF NUMERICAL ARTEFACTS INCORPORATING SPATIALLY CORRELATED FORM ERROR

Alistair Forbes [1]

[1] National Physical Laboratory, Teddington, UK, alistair.forbes@npl.co.uk

Abstract - This paper is concerned with the generation of numerical artefacts, that is, reference data sets, for assessing the accuracy and fitness of purpose of software for computing minimum zone (MZ, Chebyshev) associated features. We describe an algorithm for generating datasets corresponding to a pre-specified best-fit surface and form error. We use a Gaussian process model to generate form errors that are spatially correlated. The form errors can be drawn from multivariate Gaussian or rectangular distributions. For the latter case a Gaussian copula is used construct a multivariate multivariate distribution with the pre-assigned correlation. We illustrate the data generation on MZ circle fitting problems. We also describe an approximate MZ circle fitting problem that can be solved using linear programming.

Keywords: coordinate metrology, form error, minimum zone, numerical artefacts, spatial correlation

1. INTRODUCTION

Coordinate metrology can be thought of as a two-stage process, the first stage using a coordinate measuring system to gather coordinate data $\boldsymbol{x}_i$, $i = 1, \ldots, m$, related to a workpiece surface, the second extracting a set of parameters or characteristics $\boldsymbol{a} = (a_1, \ldots, a_n)^\top$ from the data $\boldsymbol{x}_{1:m}$ using software implementing mathematical algorithms, e.g., determine the parameters associated with the best-fit cylinder to data. The computational software is involved in the traceability chain and for the outputs of the software to be considered traceable to the metre, it is necessary to assess the accuracy and fitness for purpose of the software [1, 2, 3]. A major tool for testing software is the generation of reference datasets or *numerical artefacts* for which the correct output is known to high accuracy. One approach for generating such datasets is to start with the output $\boldsymbol{a}$ and then generate the inputs $\boldsymbol{x}_{1:m}$, the so-called null space approach. In this paper we describe a null space approach for generating numerical artefacts associated with minimum zone (MZ) associated features [4], i.e., surface fitting according to the Chebyshev criterion . In particular, we show how datasets can be generated with a specified form error and spatially correlated form errors.

2. Reference data generation for MZ problems

2.1. Form error model incorporating spatial correlation

The basic model for form error can be described as

$$\boldsymbol{x}_i = \boldsymbol{s}_i + f_i \boldsymbol{n}_i,$$

where $\boldsymbol{s}_i = \boldsymbol{s}(\boldsymbol{u}_i, \boldsymbol{a})$ is a point of the surface of an ideal geometric element, $\boldsymbol{n}_i$ is the unit surface normal at $\boldsymbol{s}_i$ and f_i is the signed departure from the ideal geometry at $\boldsymbol{s}_i$, or the form error at $\boldsymbol{s}_i$. The parameters $\boldsymbol{a}$ specify the particular geometric element, for example, the centre coordinates and radius of a circle. The idea of spatial correlation is that if $\boldsymbol{s}_i$ is near $\boldsymbol{s}_j$ then f_i should not be too far from f_j. The notion of nearness is quantified by a length scale parameter λ: if $\boldsymbol{s}_i$ is close to $\boldsymbol{s}_j$, relative to λ, then f_i should be close to f_j. Models of spatial correlation can be implemented using statistical correlation. Using a Gaussian process model [5], the form errors f_i and f_j are regarded as samples from a joint statistical distribution and the covariance of f_i with f_j is controlled by a correlation kernel function $k(\boldsymbol{s}_i, \boldsymbol{s}_j)$ related to the distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, for example,

$$\mathrm{cov}(f_i, f_j) = \sigma_F^2 \exp\left\{-(\boldsymbol{s}_i - \boldsymbol{s}_j)^\top (\boldsymbol{s}_i - \boldsymbol{s}_j)/\lambda^2\right\}.$$

2.2. Least squares best fit elements and multivariate Gaussian distribution with spatial correlation

Given surface points $\boldsymbol{s}_i = \boldsymbol{s}(\boldsymbol{u}_i, \boldsymbol{a})$ with normal vectors $\boldsymbol{n}_i$, $i \in I = \{1, \ldots, m\}$, let J be the Jacobian matrix $J = QR$ where Q is $m \times n$ associated with the least-squares fit to $\boldsymbol{s}_i$ [6]. Construct the projection matrix $P = I - QQ^\top$. Suppose the spatial correlation is encoded by the $m \times m$ variance matrix V generated as

$$V(i, j) = k(\boldsymbol{s}_i, \boldsymbol{s}_j),$$

for some correlation kernel k. Then if

$$\boldsymbol{x}_i = \boldsymbol{s}_i + f_i \boldsymbol{n}_i, \quad \boldsymbol{f} \in \mathrm{N}(\boldsymbol{0}, PVP^\top),$$

the least-squares best-fit surface to $\{\boldsymbol{x}_i\}_1^m$ is specified by $\boldsymbol{a}$.

2.3. Minimum zone elements and multivariate rectangular distribution with spatial correlation

Suppose we are given surface points $\boldsymbol{s}_i = \boldsymbol{s}(\boldsymbol{u}_i, \boldsymbol{a})$ with normal vectors $\boldsymbol{n}_i$, $i \in I = \{1, \ldots, m\}$. Let the index set $I^* \subset I$ specifying contacting points that define a minimum zone solution [7, 4], with $f_i = \pm F$, $i \in I^*$. We assume, as before that the spatial correlation is encoded in the variance matrix V generated as

$$V(i, j) = k(\boldsymbol{s}_i, \boldsymbol{s}_j),$$

for some correlation kernel k, and that C is the correlation matrix derived from V. The following scheme generates a sample $\boldsymbol{f} \in [-F, F]^m$ from a multivariate rectangular distribution constructed using a Gaussian copula [8] with preassigned correlation. The inputs to the scheme are: parameter $F > 0$ specifying the rectangular marginal distributions $U(-F, F)$, integer $m > 0$ and an $m \times m$ correlation matrix C specifying the spatial correlation. The outputs are spatially correlated form errors f_i with $-F \leq f_i \leq F$. The steps in the scheme are:

1. Form the Cholesky factorisation $C = LL^\top$.

2. Set $\boldsymbol{z} = L\boldsymbol{e}$ where $\boldsymbol{e} \in \mathrm{N}(\boldsymbol{0}, I)$, and $\boldsymbol{u}$ such that $u_i = C_\mathrm{N}(z_i)$, $0 < u_i < 1$, $i = 1, \ldots, m$, where C_N is the cumulative distribution function for a standard normal distribution.

3. Set $\boldsymbol{f} = 2F\boldsymbol{u} - F$.

For generating data with known minimum zone solution, we can first generate contacting points determining a local minimum and then add as many additional points that lie within the minimum zone. The generation of contacting points for local solutions defined by first order optimality conditions, termed 'vertex solutions' is straightforward [4, 7]. In order to determine data sets that reflect spatially correlated form errors we would like to add additional points that are spatially correlated relative to each other but also relative to the pre-specified contacting points.

As a first step, assume the index set is partitioned so that $I = I_1 \cup I_2$ where I_1 specifies the first k points and I_2 the remaining $m - k$ points and that the form errors associated with I_1 are pre-assigned to be $\boldsymbol{f}_1$, $-F \leq f_i \leq F$, $i \in I_1$. Let the Cholesky factor of L of C be partitioned as

$$L = \left[\begin{array}{cc} L_{11} & \\ L_{21} & L_{22} \end{array} \right],$$

where L_{11} is $k \times k$, etc. Steps 2 and 3 above can be also be partitioned accordingly:

$$\boldsymbol{z}_1 = L_{11}\boldsymbol{e}_1, \quad \boldsymbol{z}_2 = L_{21}\boldsymbol{e}_1 + L_{22}\boldsymbol{e}_2,$$

$$\boldsymbol{u}_1 = C_\mathrm{N}(\boldsymbol{z}_1), \quad \boldsymbol{u}_2 = C_\mathrm{N}(\boldsymbol{z}_2),$$

and

$$\boldsymbol{f}_1 = 2F\boldsymbol{u}_1 - F, \quad \boldsymbol{f}_2 = 2F\boldsymbol{u}_2 - F.$$

If $\boldsymbol{f}_1$ is pre-assigned, then $\boldsymbol{u}_1$, $\boldsymbol{z}_1$ and $\boldsymbol{e}_1$ are all also pre-assigned according to

$$\boldsymbol{u}_1 = \frac{\boldsymbol{f}_1 + F}{2F}, \quad \boldsymbol{z}_1 = C_\mathrm{N}^{-1}(\boldsymbol{u}_1), \quad \boldsymbol{e}_1 = L_{11}^{-1}\boldsymbol{z}_1.$$

With $\boldsymbol{e}_1$ so defined, for $\boldsymbol{e}_2 \in \mathrm{N}(0, I)$, $\boldsymbol{f}_2$ can be generated as above to reflect the spatial correlation encoded in the correlation matrix C and the known from errors $\boldsymbol{f}_1$. This generation scheme is analogous to generating a plausible set of form errors $\boldsymbol{f}_2$ at surface points $\boldsymbol{s}_i$, $i \in I_2$, given that form errors $\boldsymbol{f}_1$ at locations $\boldsymbol{s}_i$, $i \in I_2$, have been measured accurately.

In order to generate data for which the MZ solution is known, I_1 is used to specify the contacting points with $f_i = \pm F$, $i \in I_1$. However, for this choice, u_i is assigned to 0 or 1 and the inverse normal cdf C_N^{-1} is not finite at these values. From a probability point of view, given a finite sample from a rectangular distribution $U(-F, F)$, there is zero probability that any member of the sample is equal to F in magnitude. We can use order statistics for the uniform distribution to assign more plausible values for $\boldsymbol{f}_1$ [9]. Given m samples from the rectangular distribution $U(0, 1)$, the expected value of the smallest sample is $1/(m + 1)$ and the expected value of the largest is $m/(m + 1)$. Given m samples from $U(0, 1)$ it is quite likely that one or more samples will lie outside the interval $[1/(m+1), m/(m+1)]$ but much less likely that they will lie outside an interval such as $[1/(2m + 1), 2m/(2m + 1)]$. On this basis we assign $f_i = \pm F_m$, where

$$F_m = \frac{2m - 1}{2m + 1}F.$$

With this value, there is now a finite but small probability that $\max_{i \in I_2} |f_i| > F_m$, given a random draw $\boldsymbol{e}_2 \in \mathrm{N}(0, I)$. If this happens, we perform a small adjustment in $\boldsymbol{f}_2$ so that the bound is satisfied according to:

$$f_i := \frac{Gf_i}{\sqrt{G^2 + (\gamma^2 - 1)f_i^2}}, \quad \gamma = \frac{G}{F_m}, \quad G = \max_{i \in I_2} f_i. \tag{1}$$

The effect of the adjustment is to reduce only those f_i that are above or close to the bound F_m.

2.4. MZ data generation algorithm

The following scheme generates data points $\boldsymbol{x}_i$ for which the minimum zone fit is given by $\boldsymbol{a}$ with contacting points $\boldsymbol{x}_i$, $i \in I^*$ and spatially correlated form error f_i with $|f_i| \leq F$. We assume that the data points $\boldsymbol{s}_i$ are ordered so that $I = I_1 \cup I_2$ with $I^* = I_1 = \{1, \ldots, k\}$ and $I_2 = \{k + 1, \ldots, m\}$. The inputs to the scheme are:

1. Surface points $\boldsymbol{s}_i = \boldsymbol{s}(\boldsymbol{u}_i, \boldsymbol{a})$ with normal vectors $\boldsymbol{n}_i$, $i \in I$ where $I = I_1 \cup I_2 = \{1, \ldots, m\}$, where I_1 specifies k contacting points that define a minimum zone solution.

2. Form errors $f_i = \pm F$, $i \in I_1$ associated with contacting points

3. Form error bound $F_\mathrm{max} \leq F$ for non-contacting points.

4. An $m \times m$ correlation matrix C specifying the spatial correlation.

The outputs are Spatially correlated form errors $-F \leq f_i \leq F$, $i \in I$, $-F_\mathrm{max} \leq f_i \leq F_\mathrm{max}$, $i \in I_2$, and data set $\boldsymbol{x}_i = \boldsymbol{s}_i + f_i\boldsymbol{n}_i$, such that the minimum zone fit to $\boldsymbol{x}_i$ is given by parameters $\boldsymbol{a}$. The steps in the scheme are:

1. Form the Cholesky factorisation $C = LL^\top$ and partition L as
$$L = \left[\begin{array}{cc} L_{11} & \\ L_{21} & L_{22} \end{array} \right],$$
where L_{11} is $k \times k$, etc.
2. Set $F_m = (2m-1)F/(2m+1)$ and $\boldsymbol{f}_{1,m} = F_m \boldsymbol{f}_1/F$.
3. Set
$$\boldsymbol{u}_1 = \frac{\boldsymbol{f}_{1,m} + F}{2F}, \quad \boldsymbol{z}_1 = C_{\rm N}^{-1}(\boldsymbol{u}_1), \quad \boldsymbol{e}_1 = L_{11}^{-1}\boldsymbol{z}_1.$$
4. Set $\boldsymbol{z} = L_{21}\boldsymbol{e}_1 + L_{22}\boldsymbol{e}_2$, $\boldsymbol{e}_2 \in {\rm N}(\boldsymbol{0}, I)$, and $\boldsymbol{u}_2 = C_{\rm N}(\boldsymbol{z}_2)$.
5. Set $\boldsymbol{f}_{2,m} = 2F\boldsymbol{u}_2 - F$.
6. If necessary, adjust $\boldsymbol{f}_{2,m}$ according to (1).
7. Set
$$\boldsymbol{f} = \left[\begin{array}{c} \boldsymbol{f}_1 \\ \boldsymbol{f}_{2,m} \end{array} \right].$$
8. Set $\boldsymbol{x}_i = \boldsymbol{s}_i + f_i \boldsymbol{n}_i$, $i \in I$.

3. EXAMPLE: MINIMUM ZONE CIRCLE

To illustrate the approach, we generate data for minimum zone (MZ) circle calculations. Determining an MZ circle fit can be posed as a constrained optimisation problem
$$\min_{x_0,y_0,r_0,R_0} R_0 - r_0 \tag{2}$$
subject to the constraints
$$R_0^2 \geq r_i^2 \geq r_0^2, \quad r_i^2 = (x_i - x_0)^2 + (y_i - y_0)^2.$$
As posed, while the objective function is linear in the optimisation parameters, the constraints depend nonlinearly on them and algorithmic solutions for this type of problem are not straightforward in general [10]. However, defining s and S by
$$s = r_0^2 - x_0^2 - y_0^2, \quad S = R_0^2 - x_0^2 - y_0^2,$$
the optimisation problem can be reformulated as
$$\min_{x_0,y_0,s,S} \frac{S - s}{R_0 + r_0} \tag{3}$$
subject to
$$\begin{aligned} x_i^2 + y_i^2 &\geq 2x_i x_0 + 2y_i y_0 + s, \\ x_i^2 + y_i^2 &\leq 2x_i x_0 + 2y_i y_0 + S, \end{aligned} \tag{4}$$
where
$$\begin{aligned} r_0 &= r_0(x_0, y_0, s) = (s + x_0^2 + y_0^2)^{1/2}, \\ R_0 &= R_0(x_0, y_0, S) = (S + x_0^2 + y_0^2)^{1/2}. \end{aligned}$$
The reformulated optimisation problem involves minimising a nonlinear objective function subject to linear constraints and can be addressed by more straightforward algorithms compared to those required for (2). The objective function in (3) is nonlinear due to the factor of $1/(R_0 + r_0)$. For data representative of a circle, this factor will be approximately constant over the region representing plausible solutions to the problem, suggesting that we can determine an approximate solution by solving
$$\min_{x_0,y_0,s,S} S - s \tag{5}$$
subject to the constraints (4). This last formulation is a linear programming problem for which there are straightforward and highly effective solution algorithms [10, 11], including Dantzig's celebrated simplex algorithm [12].

The MZ circle problem as posed in (4) can have a number of local solutions and most algorithms to solve it require a starting estimate for the circle centre and the algorithms will perform a sequence of iterations, reducing the radial separation at each step, until a local minimum is found. The approximate MZ (AMZ) circle problem has a unique global minimum which will coincide with a local minimum for the MZ problem, but not necessarily the global minimum for the MZ problem. The MZ data generation algorithm can be used to test the performance of an algorithm solving the AMZ problem on data involving a realistic form errors.

Figures 1–3 show three data sets generated using the scheme described in section 2.4 for $r_0 = 10$, $F = 1$, $m = 180$, and spatial correlation lengths $\lambda = 0.1$, 2.0 and 10.0, corresponding to weak correlation and near random form errors, moderate correlation and 'wavy' form errors and strong correlation and smoothly varying form errors, respectively. Also shown are the four contacting points that define the MZ circle solution, each showing the characteristic interlacing property in which the contacting points alternate between being situated on the inner and outer circle, going around the circle. We have implemented MATLAB software for solving the MZ problem using MATLAB optimisation component FMINCON.M implementing a sequential quadratic programming algorithm and for solving the AMZ problem using MATLAB optimisation component LINPROG.M implementing a dual simplex algorithm. Both software implementations found the optimal solution in a small number (about 10 or less) iterations. For datasets similar to those in the figures 1–3, the AMZ solution will be almost certainly be the same as the MZ solution.

4. CONCLUSIONS

This paper has been concerned with the generation of numerical artefacts (reference data sets) for assessing the accuracy and fitness of purpose of software for computing minimum zone (MZ, Chebyshev) associated features. We have described an algorithm for generating datasets with a pre-specified form error and spatial correlation, using a Gaussian copula to sample from a multivariate distribution with rectangular marginal distributions. We have illustrated

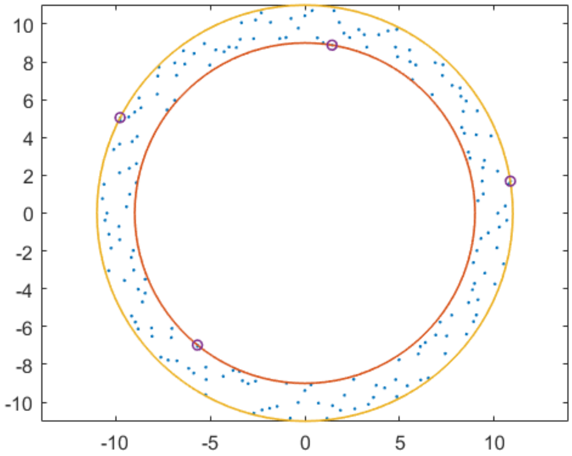

Fig. 1. Synthesized circle data with for $r_0 = 10$, $F = 1$, $m = 180$, and spatial correlation length $\lambda = 0.1$, corresponding to weak correlation and near random form errors. The circle mark the four 'contacting' points associated with the solution MZ circle

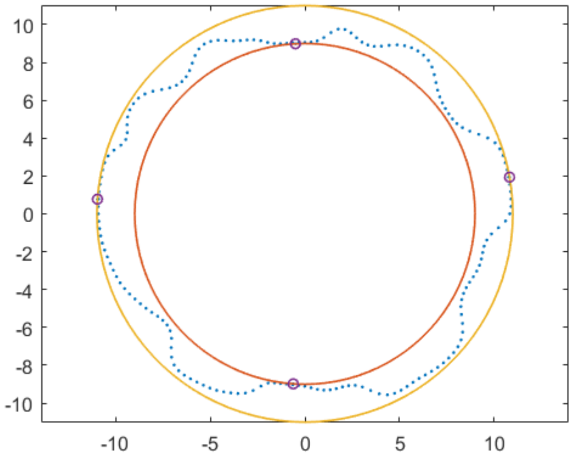

Fig. 2. As figure 1 but with $\lambda = 2$, corresponding to moderate correlation and 'wavy' form errors.

the data generation on MZ circle fitting. We have also described an approximate MZ circle fitting problem that can be solved using linear programming.

ACKNOWLEDGEMENTS

This work is supported by the UK Government Department for Business, Energy and Industrial Strategy (BEIS) through the National Measurement System programme for *Data Science* and through the EMPIR project 17NRM03, *Standards for the evaluation of the uncertainty of coordinate measurements in industry* (EUCoM). The EMPIR programme is co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

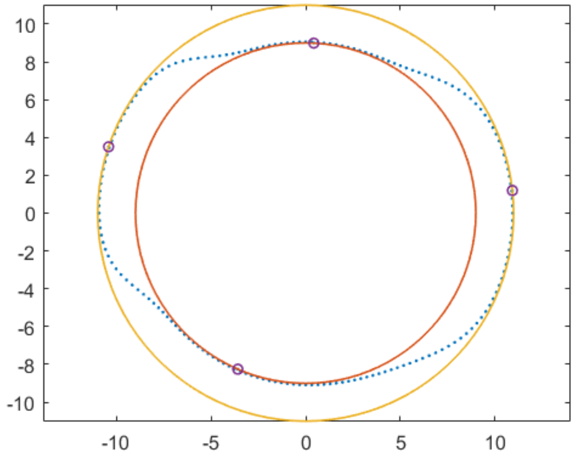

Fig. 3. As figure 1 but with $\lambda = 10$, corresponding to strong correlation and smoothly varying form errors.

REFERENCES

[1] F. Härtig, B. Müller, K. Wendt, M. Franke, A. B. Forbes, and I. M. Smith. Online validation of metrological software using the TraCIM system. In *XXI IMEKO World Congress, September 2015, Prague*, 2015.

[2] G. J. P. Kok, P. M. Harris, I. M. Smith, and A. B. Forbes. Reference data sets for testing metrology software. *Metrologia*, 53(4), 2016.

[3] J. M. Linares, G. Goch, A. Forbes, J. M. Sprauel, A. Clément, F. Härtig, and W. Gao. Modelling and traceability for computationally-intensive precision engineering and metrology. *CIRP Annals Manufacturing Technology*, August 2018. online version, DOIL 10.1016/j.cirp.2018.05.003.

[4] A. B. Forbes and H. D. Minh. Generation of numerical artefacts for geometric form and tolerance assessment. *Int. J. Metrol. Qual. Eng.*, pages 145–150, 2012.

[5] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, Mass., 2006.

[6] A. B. Forbes. Least-squares best-fit geometric elements. Technical Report DITC 140/89, National Physical Laboratory, Teddington, 1989.

[7] A. B. Forbes and H. D. Minh. Form assessment in coordinate metrology. In E. H. Georgoulis, A. Iske, and J. Levesley, editors, *Approximation Algorithms for Complex Systems*, Springer Proceedings in Mathematics, Vol 3, pages 69–90, Heidelberg, 2011. Springer-Verlag.

[8] Roger B. Nelsen. *An Introduction to Copulas*. Springer Publishing Company, Inc., New York, 2010.

[9] H. A. David and H. N. Nagaraja. *Order Statistics*. John Wiley & Sons, Inc., Hoboken, 3rd edition, 2003.

[10] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization*. Academic Press, London, 1981.

[11] M. J. Todd. The many facets of linear programming. *Mathematical Programming*, 91:417–436, 2002.

[12] G. B. Dantzig. *Linear programming and extensions*. Princeton University Press, Princeton, N.J., 1963.