



HAL
open science

Two dimensional discrete scan statistics with arbitrary window shape

Alexandru Amârioarei, Michaël Genin, Cristian Preda

► **To cite this version:**

Alexandru Amârioarei, Michaël Genin, Cristian Preda. Two dimensional discrete scan statistics with arbitrary window shape. Handbook of scan statistics, 2019. <hal-02405886>

HAL Id: hal-02405886

<https://hal.science/hal-02405886v1>

Submitted on 11 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Two dimensional discrete scan statistics with arbitrary window shape

Alexandru Amărioarei^{1,2}, Michaël Genin³, and Cristian Preda^{4,5}

¹Faculty of Mathematics and Computer Science, University of Bucharest, Romania

²National Institute of Research and Development for Biological Sciences, Bucharest, Romania

³Univ. Lille, EA2694 - Santé publique : épidémiologie et qualité des soins, F-59000 Lille, France

⁴Laboratoire Paul Painlevé, Université de Lille, France

⁵Institute of Mathematical Statistics and Applied Mathematics of the Romanian Academy, Bucharest, Romania

May 27, 2019

Abstract

The definition of the the two-dimensional discrete scan statistic with rectangular window shape is extended to a more general framework. In particular, this approach allows to introduce different shapes for the scanning window (discretized rectangle, polygon, circle, ellipse or annulus). We provide approximation for the distribution of the scan statistic and illustrate their accuracy by conducting a numerical comparison study. The power of test based on the scan statistics is also evaluated by simulation.

1 Introduction

Let $T_1, T_2 \geq 2$ be positive integers, $\mathbf{R} = [0, T_1] \times [0, T_2]$ be a rectangular region and X_{s_1, s_2} , $1 \leq s_j \leq T_j$, $j \in \{1, 2\}$, be an array of independent and identically distributed random variables from a specified distribution (Bernoulli, binomial, Poisson, normal, etc.) associated to the elementary sub-region $r(s_1, s_2) = [s_1 - 1, s_1] \times [s_2 - 1, s_2]$. The two dimensional scan statistics introduced by [Chen and Glaz, 1996a] is defined as the largest number of events in any rectangular scanning window of size $m_1 \times m_2$, where $1 \leq m_1 \leq T_1$ and $1 \leq m_2 \leq T_2$ are positive integers, within the rectangular region \mathbf{R} , i.e.

$$S_{m_1, m_2}(T_1, T_2) = \max_{\substack{1 \leq i_1 \leq T_1 - m_1 + 1 \\ 1 \leq i_2 \leq T_2 - m_2 + 1}} \sum_{s_1 = i_1}^{i_1 + m_1 - 1} \sum_{s_2 = i_2}^{i_2 + m_2 - 1} X_{s_1, s_2} \quad (1)$$

The two-dimensional scan statistics have been widely used in several fields of application such as cosmology ([Darling and Waterman, 1986]), reliability theory ([Barbour et al., 1996]), epidemiology and public health ([Genin et al., 2013, Castra et al., 2019]). Since there are no exact formulas for the probability distribution of S , that motivates the researchers to look for accurate approximations. Nowadays, the study

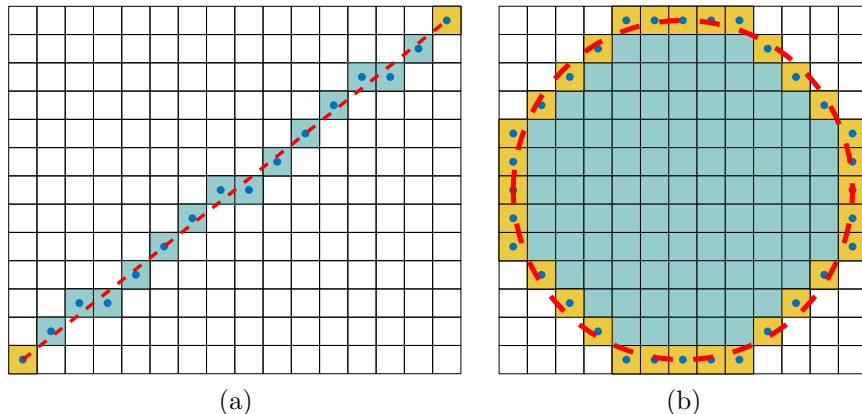


Figure 1: Example of discrete line (a) and discrete circle (b) using Bresenham's algorithms.

of the distribution of the scan statistics is an active research topic in statistics and several approximations and bounds have been provided in literature ([Chen and Glaz, 1996b, Boutsikas and Koutras, 2003, Haiman and Preda, 2006]).

From the definition of the discrete scan statistics, the shape of the scanning window is rectangular. That is the most common and easy to deal with among other possible choices. As far as we know, no other forms have been considered in literature and, one possible explanation is the discrete feature of data. Notice that for the two-dimensional *continuous* scan statistics, there exists works considering different shapes. Actually, [Naus, 1965, Loader, 1991] used rectangles, [Alm, 1997, Alm, 1998, Anderson and Titterington, 1997] considered rectangles and circles whereas [Alm, 1997, Alm, 1998] considered triangles, ellipses and other convex shapes [Assunção et al., 2006, Tango and Takahashi, 2005].

In this article we present a natural extension of the classical definition of the two dimensional discrete scan statistics by taking in (refeq1), instead of a moving sum, a moving score evaluated with the help of a function (matrix) applied on blocks of random variables. This score allows the definition of the two-dimensional discrete scan statistics with arbitrary scanning window shape. The article is organized as follows: Section 2 presents the definition of the two-dimensional discrete scan statistics with arbitrary scanning window shape. Section 2.2 presents the approximation methodology of the distribution of the scan statistic. In Section 3, we provide numerical applications for both the approximation of the distribution and power when considering different discrete shapes of the scanning windows. Lastly, the results are discussed in section 4.

2 Two dimensional discrete scan statistics with arbitrary window shape

2.1 Arbitrary discrete scanning window shape

Let G be a geometrical shape of the scanning window (rectangle, quadrilateral, ellipse, etc.) and \tilde{G} be its corresponding discrete form. The transformation of continuous shape to discrete shape are performed using rasterization algorithms. In the literature, several authors have proposed algorithm to rasterize line [Bresenham, 1965], circle [Bresenham, 1977] or Bezier curves [Foley et al., 1997]. Figure 1 illustrates the discretization of line and circle according the previous algorithm.

To each discrete shape \tilde{G} it corresponds an unique rectangle of smallest size $m_1 \times m_2$ which circumscribes \tilde{G} .

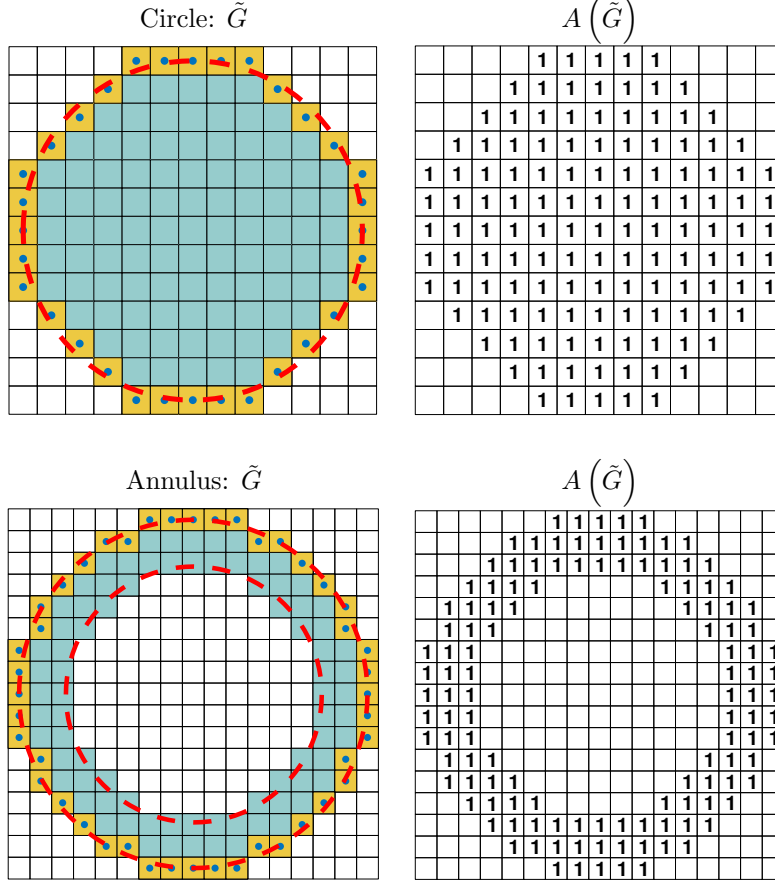


Figure 2: Discrete circle ($m_1 = m_2 = 13$) and discrete annulus ($m_1 = m_2 = 17$)

Therefore, let define the matrix $A(\tilde{G}) = (a_{ij})_{\substack{1 \leq i \leq m_1 \\ 1 \leq j \leq m_2}}$ with entries 0 and 1 ($a_{ij} = 1$ if there the elementary square $[i-1, i] \times [j-1, j]$ belongs to \tilde{G} and 0 otherwise). See Figure 2 for discrete circle ($m_1 = m_2 = 13$) and discrete annulus ($m_1 = m_2 = 17$).

We define the two-dimensional discrete scan statistic with scanning window \tilde{G} as the maximum of the moving sums

$$Y_{i_1, i_2} = \sum_{s_1=i_1}^{i_1+m_1-1} \sum_{s_2=i_2}^{i_2+m_2-1} A(s_1 - i_1 + 1, s_2 - i_2 + 1) X_{s_1, s_2}$$

over the region \mathbf{R} , i.e. $(i_1, i_2) \in [1, T_1 - m_1 + 1] \times [1, T_2 - m_2 + 1]$,

$$S_{\tilde{G}}(T_1, T_2) = \max_{\substack{1 \leq i_1 \leq T_1 - m_1 + 1 \\ 1 \leq i_2 \leq T_2 - m_2 + 1}} Y_{i_1, i_2}. \quad (2)$$

Remark that if, in particular, the shape G is a rectangle of size $m_1 \times m_2$ then its corresponding $\{0, 1\}$ matrix \tilde{G} has all the entries equal to 1. Thus, the definition of the two-dimensional discrete scan statistics (2) extends the classical one given in (1).

2.2 Approximation for the distribution of $S_{\tilde{G}}$

As far as we know, there are no available results on the distribution of the two-dimensional discrete scan statistics $S_{\tilde{G}}$ with general window shape.

For rectangular scanning window \tilde{G} of size $m_1 \times m_2$, various methods of approximation and bounds have been proposed for $\mathbb{P}(S_{\tilde{G}} \leq \tau)$, $\tau \in \mathbb{R}$. An overview of these methods as well as a complete bibliography on the subject are given in [Glaz et al., 2001] and [Glaz et al., 2009].

The approximation methodology proposed in [Haiman and Preda, 2006] has the advantage to provide sharp bounds for the associated error with reasonable computing time. The main idea of the method is to consider the scan statistics as the maximum of a 1-dependent stationary sequence of random variables. Therefore, one can use the results of [Haiman, 1999] (Theorem 1), improved by [Amărioarei, 2014], that provide approximations and bounds for the distribution of the maximum of 1-dependent stationary sequences.

In what follows, we describe this methodology for rectangular scanning window and use it to arbitrary scanning window shape.

Let $L_1 = \lfloor \frac{T_1}{m_1-1} \rfloor$ and $L_2 = \lfloor \frac{T_2}{m_2-1} \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of x . Denote by $Q_{i,j}$, $1 \leq i \leq L_1$, $1 \leq j \leq L_2$,

$$Q_{i,j} = S_{\tilde{G}}(i(m_1 - 1), j(m_2 - 1)),$$

the scan statistic considered on the rectangular region of size $[0, i(m_1 - 1)] \times [0, j(m_2 - 1)]$ with scanning window \tilde{G} . Notice that $S_{\tilde{G}}(T_1, T_2) = Q_{L_1, L_2}$.

The approximation of the distribution of $S_{\tilde{G}}$ developed in [Haiman and Preda, 2006] is based on the distributions of $Q_{2,2}$, $Q_{2,3}$, $Q_{3,2}$ and $Q_{3,3}$. Unlike the case of $S_{\tilde{G}}(T_1, T_2) = Q_{L_1, L_2}$, the estimation of the distributions of $Q_{i,j}$, $i, j \in \{2, 3\}$, by Monte Carlo techniques is efficient and computationally feasible because of the small size of the scanned region and the use of importance sampling techniques in this framework (see [Amărioarei and Preda, 2014] for more details).

The approximation of the distribution of $S_{\tilde{G}}(T_1, T_2)$ is then obtained as follows ([Haiman and Preda, 2006], [Amărioarei and Preda, 2014]).

Theorem 1. *Let $i, j \in \{2, 3\}$, $\tau \in \mathbb{R}$ and $Q_{i,j}(\tau) = \mathbb{P}(Q_{i,j} \leq \tau)$. If $\hat{Q}_{i,j} \doteq \hat{Q}_{i,j}(\tau)$ is an estimate of $Q_{i,j}(\tau)$ with $|\hat{Q}_{i,j} - Q_{i,j}(\tau)| \leq \beta_{i,j}$ and τ is such that $1 - Q_{2,2}(\tau) \leq 0.1$ then,*

$$\left| \mathbb{P}(S_{\tilde{G}}(T_1, T_2) \leq \tau) - \left(2\hat{Q}_2 - \hat{Q}_3\right) \left[1 + \hat{Q}_2 - \hat{Q}_3 + 2(\hat{Q}_2 - \hat{Q}_3)^2\right]^{1-L_1} \right| \leq E_{sf} + E_{sapp}, \quad (3)$$

where, for $t \in \{2, 3\}$

$$\begin{aligned} \hat{Q}_t &= \left(2\hat{Q}_{t,2} - \hat{Q}_{t,3}\right) \left[1 + \hat{Q}_{t,2} - \hat{Q}_{t,3} + 2(\hat{Q}_{t,2} - \hat{Q}_{t,3})^2\right]^{1-L_2} \\ E_{sf} &= (L_1 - 1)(L_2 - 1) (\beta_{2,2} + \beta_{2,3} + \beta_{3,2} + \beta_{3,3}) \\ E_{sapp} &= (L_1 - 1) \left[F_1 \left(1 - \hat{Q}_2 + A_2 + C_2\right)^2 + (L_2 - 1)(F_2 C_2 + F_3 C_3) \right] \\ A_2 &= (L_2 - 1) (\beta_{2,2} + \beta_{2,3}) \\ C_t &= (L_2 - 1) F_t \left(1 - \hat{Q}_{t,2} + \beta_{t,2}\right)^2. \end{aligned}$$

and F_1 , F_2 and F_3 are constants defined in [Amărioarei, 2014].

The result in Theorem 1 shows that the distribution of the scan statistics $S_{\tilde{G}}(T_1, T_2)$ can be efficiently approximated as

$$\mathbb{P}(S_{\tilde{G}}(T_1, T_2) \leq \tau) \approx \left(2\hat{Q}_2 - \hat{Q}_3\right) \left[1 + \hat{Q}_2 - \hat{Q}_3 + 2(\hat{Q}_2 - \hat{Q}_3)^2\right]^{1-L_1}, \quad (4)$$

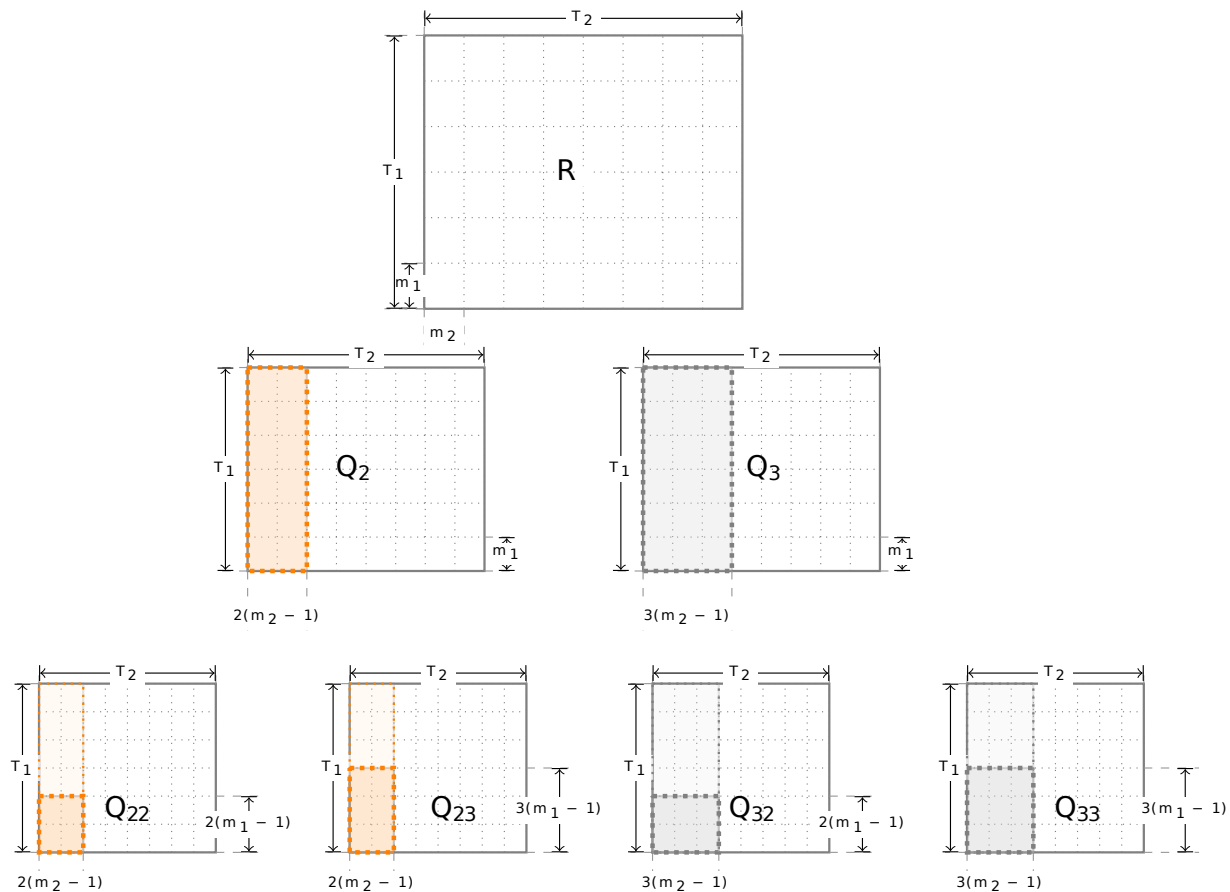


Figure 3: Approximation process for $\mathbb{P}(S_{\tilde{C}} \leq \tau)$

provided accurate approximations of $Q_{ij}(\tau), i, j \in \{2, 3\}$ are available. The approximation methodology is illustrated in Figure 3.

Notice that E_{sapp} , given in (3), rapidly tends to zero as $\mathbb{P}(S_{\tilde{C}}(T_1, T_2) \leq \tau)$ is close to 1. Moreover, for large L_1 and L_2 , the contribution of E_{sapp} is almost negligible with respect to E_{sf} . Thus, the accuracy of the approximation essentially depends on the number N of replications used for estimating the distribution of $Q_{2,2}$, $Q_{2,3}$, $Q_{3,2}$ and $Q_{3,3}$. For rectangular scanning window we used the importance sampling technique presented in [Amărioarei and Preda, 2014] which provides sharper error bounds than the usual Monte Carlo method we used for discrete circle. The approximation error is useful when one has to compare approximations for different shapes. A difference between two approximations can only be highlighted if the error bounds of those approximations are sufficiently small.

3 Numerical applications. Approximation and power of the scan statistic test

In this section, we first present a simulation study aiming to evaluate the change in the distribution of the two-dimensional discrete scan statistics when scanning with different shapes of the window. Secondly, we evaluate the power of the test based on the scan statistics accordingly to the shape of the scanning window and that of the simulated cluster (under the alternative hypothesis). We consider here the case of triangular, rectangular, quadrilateral, circular, ellipsoidal and annular scanning windows of same size (area).

3.1 Approximation and scanning window shape

We present approximations of the distribution of scan statistics, $\mathbb{P}(S_{\tilde{C}} \leq \tau)$, when the X_{ij} 's are distributed as binomial ($\mathcal{B}(\nu, p)$), Poisson ($\mathcal{P}(\lambda)$) and Normal ($\mathcal{N}(\mu, \sigma)$). The methodology used for obtaining such approximations is that presented in Section 2.2. For each model, we compare the distributions of the scan statistics when scanning with triangular, rectangular, quadrilateral, circular, ellipsoidal and annular windows. Tables 1 and 2 provide the distribution of the two-dimensional discrete scan statistic when scanning a rectangular region of size $T_1 \times T_2 = 250 \times 250$ with a scanning window of fixed size A ($A \approx 130$) and various shapes. Notice that, since the scanning region is a square, the distribution of the scan statistics with rectangular scanning window of size $m_1 \times m_2$ is the same as scanning with $m_2 \times m_1$.

We consider several models for X_{ij} : $\mathcal{B}(1, 0.01)$, $\mathcal{B}(5, 0.05)$, $\mathcal{P}(0.25)$ and $\mathcal{N}(0, 1)$. The results presented in the Tables 1 and 2 are obtained using the approximation (3). We used the importance sampling technique presented in [Naiman and Priebe, 2001] (see also [Amărioarei and Preda, 2013]) which provides sharper error bounds than the usual Monte Carlo method.

In Tables 1 and 2 we have :

- **Sim** column denotes the estimation obtained by Monte Carlo (IS is the number of replications with importance sample algorithm and IA otherwise),
- **AppH** denotes the "Haiman approximation" given by (4),
- **ETotal** denotes the total error of approximation given in Theorem 1, $ETotal = E_{sf} + E_{sapp}$.

Comparing the approximation values and taking into account the error bounds, one observes that the distributions of the scan statistics associated to different scanning window shapes are close to each other. However, as expected, there are significant small differences, as for example in the Poisson model when scanning with a square scanning window ($\mathbb{P}(S_{\tilde{C}} \leq 61) < 0.96$) and with a quadrilateral scanning window ($\mathbb{P}(S_{\tilde{C}} \leq 61) > 0.96$).

3.2 Power of the scan statistic test: window and cluster shapes

We evaluate the power of the test based on the scan statistics accordingly to the shape of the scanning window and that of the existing cluster under the alternative hypothesis. We consider here for X_{ij} 's the binomial model. The test based on scan statistics checks the null hypothesis $\mathcal{H}_0 : X_{ij} \sim \mathcal{B}(\nu, p_0)$ against the alternative \mathcal{H}_1 supporting the existence of a subregion (cluster) $\mathcal{C} \in \{1, \dots, T_1\} \times \{1, \dots, T_2\}$ such that for all $(i, j) \in \mathcal{C}$, the X_{ij} 's are distributed according to binomial distribution $\mathcal{B}(\nu, p_1)$, $p_1 > p_0$. Otherwise, i.e. outside \mathcal{C} , the X_{ij} 's are binomial distributed $\mathcal{B}(\nu, p_0)$.

In what follows we consider the shape of the cluster \mathcal{C} to be triangular, rectangular, quadrilateral, circular, ellipsoidal and annular. For a fixed shape of the cluster one evaluates the power of the scan statistic to detect the cluster when the scanning window is of various shapes, i.e. triangular, rectangular, quadrilateral, circular, ellipsoidal and annular. We adopt the following procedure:

- Fix the type 1 error to $\alpha = 0.05$.
- Generate a random field $\{X_{i,j}\}$, $\{i, j\} \in \{1, \dots, T_1\} \times \{1, \dots, T_2\}$ such that there is a change in the distribution parameters of $X_{i,j}$ in the cluster \mathcal{C} . The location of \mathcal{C} inside $\{1, \dots, T_1\} \times \{1, \dots, T_2\}$ is randomly generated.
- Under \mathcal{H}_1 , we consider p_1 in the interval $[p_0, p_{max}]$ taking all values with a discrete step of 0.01. The value p_{max} is such that $p_{max} \geq p_0$ and the power of the scan statistics does not change (i.e. equals to 1) for all considered scanning window shape.
- For each particular shape of the scanning window,
 - Let $\tau_{1-\alpha}$ be the smallest integer such that $\mathbb{P}(S_{\tilde{C}} \leq \tau_{1-\alpha}) \geq 1 - \alpha$. It is obtained as the $1 - \alpha$ quantile of the approximated distribution of $S_{\tilde{C}}$ given by (3).
 - Under \mathcal{H}_1 , for each specific shape of cluster and parameter (p_1), simulate $N = 10^6$ realizations of random field $\{X_{i,j}\}$. For each realization i , $1 \leq i \leq N$, the observed value of the scan statistic $S_{\tilde{C}_i}$ is compared to $\tau_{1-\alpha}$. The power $1 - \beta$ is estimated by :

$$1 - \beta = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{S_{\tilde{C}_i} \geq \tau_{1-\alpha}\}} \quad (5)$$

For $T_1 \times T_2 = 250 \times 250$, the estimation of the power function of the test based on scan statistics is presented in Figures 4 and 5. For both models and for each of the cluster shapes, we plot the power function corresponding to scan statistics with different scanning window shapes. With regard to Bernoulli's model (Figure 4), we observe that the power curves are very similar whatever the shape of the simulated cluster and that of the scanning window. However, cluster detection is slightly more effective when the shape of the scanning window is identical to that of the simulated cluster. This is more marked in the case of an annular scan window. The same comments are still valid for the binomial model (Figure 5).

4 Discussion

We proposed an extension of the classical definition of the two dimensional discrete scan statistics when considering an scanning window of arbitrary shape. A simulation study carried out with different discrete shapes of scanning window showed that the distributions of the associated scan statistics are very close to each other but significantly different. The power of the scan statistics is linked to the shape of the scanning window and that of the existing cluster under the alternative hypothesis throughout the simulation study.

References

- [Alm, 1997] Alm, S. E. (1997). On the distributions of scan statistics of a two-dimensional poisson process. *Advances in Applied Probability*, pages 1–18.
- [Alm, 1998] Alm, S. E. (1998). Approximation and simulation of the distributions of scan statistics for poisson processes in higher dimensions. *Extremes*, 1(1):111–126.
- [Amărioarei, 2014] Amărioarei, A. (2014). *Approximations for the multidimensional discrete scan statistics*. PhD thesis, University of Lille 1.

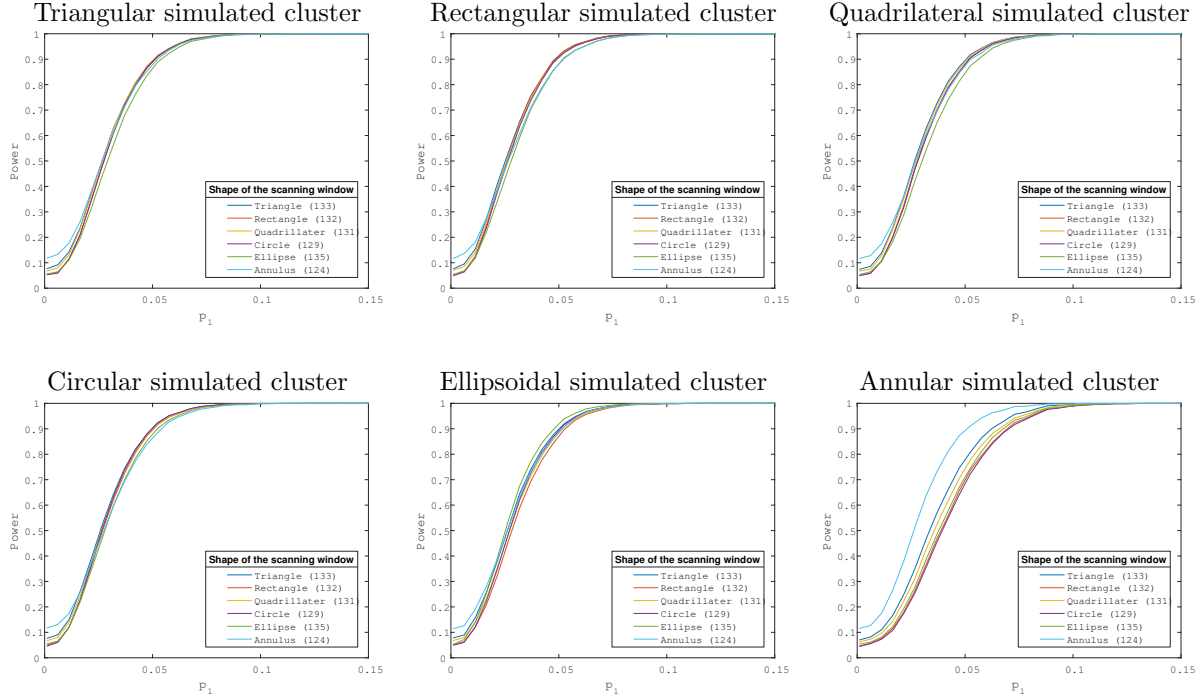


Figure 4: Power evaluation for $\mathcal{B}(1, 0.001)$ model

- [Amárioarei and Preda, 2013] Amárioarei, A. and Preda, C. (2013). Approximation for the distribution of three-dimensional discrete scan statistic. *Methodol Comput Appl Probab.*
- [Amárioarei and Preda, 2014] Amárioarei, A. and Preda, C. (2014). Approximations for two-dimensional discrete scan statistics in some block-factor type dependent models. *Journal of Statistical Planning and Inference*, 151-152:107–120.
- [Anderson and Titterington, 1997] Anderson, N. H. and Titterington, D. M. (1997). Some methods for investigating spatial clustering, with epidemiological applications. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(1):87–105.
- [Assunção et al., 2006] Assunção, R., Costa, M., Tavares, A., and Ferreira, S. (2006). Fast detection of arbitrarily shaped disease clusters. *Statistics in Medicine*, 25(5):723–742.
- [Barbour et al., 1996] Barbour, A., Chryssaphinou, O., and Roos, M. (1996). Compound poisson approximation in systems reliability. *Naval Research Logistics (NRL)*, 43(2):251–264.
- [Boutsikas and Koutras, 2003] Boutsikas, M. V. and Koutras, M. V. (2003). Bounds for the distribution of two-dimensional binary scan statistics. *Probability in the Engineering and Informational Sciences*, 17(4):509–525.
- [Bresenham, 1977] Bresenham, J. (1977). A linear algorithm for incremental digital display of circular arcs. *Communications of the ACM*, 20(2):100–106.
- [Bresenham, 1965] Bresenham, J. E. (1965). Algorithm for computer control of a digital plotter. *IBM Systems journal*, 4(1):25–30.

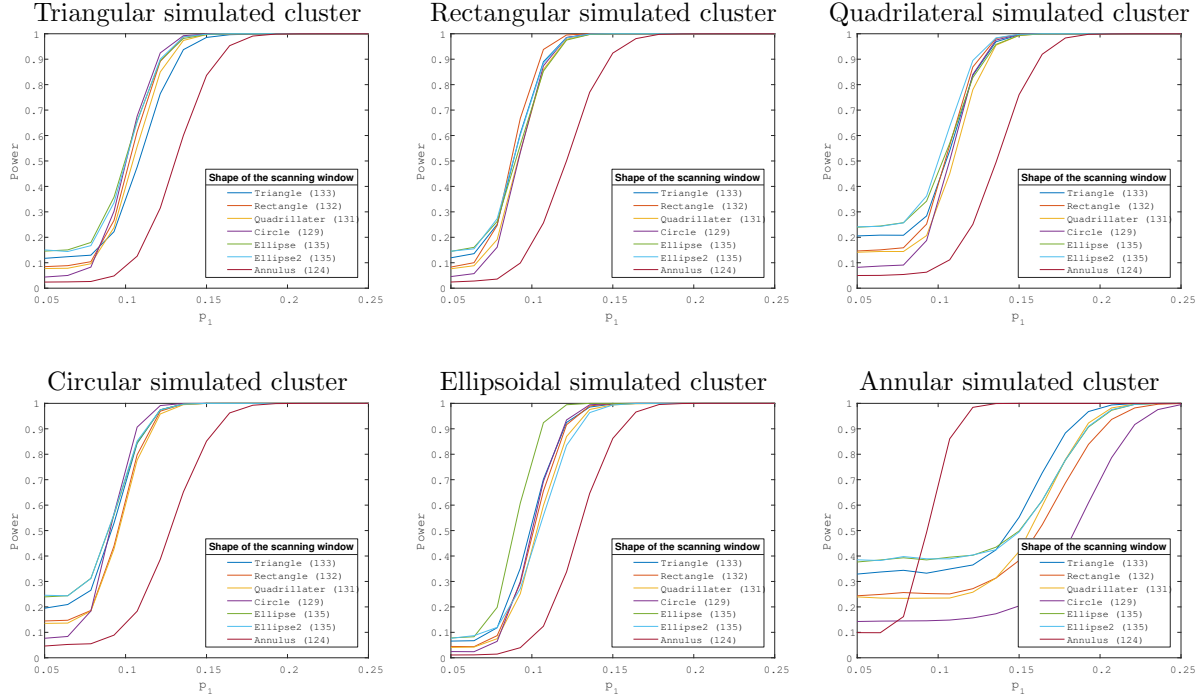


Figure 5: Power evaluation for $\mathcal{B}(5, 0.05)$ model

- [Castra et al., 2019] Castra, L., Genin, M., Escutnaire, J., Baert, V., Agostinucci, J.-M., Revaux, F., Ursat, C., Tazarourte, K., Adnet, F., and Hubert, H. (2019). Socioeconomic status and incidence of cardiac arrest: a spatial approach to social and territorial disparities. *Eur J Emerg Med*, 26.
- [Chen and Glaz, 1996a] Chen, J. and Glaz, J. (1996a). Two-dimensional discrete scan statistics. *Statist. Probab. Lett.*, 31:59–68.
- [Chen and Glaz, 1996b] Chen, J. and Glaz, J. (1996b). Two-dimensional discrete scan statistics. *Statistics & probability letters*, 31(1):59–68.
- [Darling and Waterman, 1986] Darling, R. W. R. and Waterman, M. S. (1986). Extreme value distribution for the largest cube in a random lattice. *SIAM J. Appl. Math.*, 46:118–132.
- [Foley et al., 1997] Foley, J. D., Dam, A. v., Feiner, S. K., Hughes, J. F., and Carter, M. P. (1997). Computer graphics: Principles and practice, in c. *Color Research and Application*, 22(1):65–65.
- [Genin et al., 2013] Genin, M., Duhamel, A., Preda, C., Fumery, M., Savoye, G., Peyrin-Biroulet, L., Salleron, J., Lerebours, E., Vasseur, F., Cortot, A., et al. (2013). Space-time clusters of crohn’s disease in northern france. *Journal of Public Health*, 21(6):497–504.
- [Glaz et al., 2001] Glaz, J., Naus, J., and Wallenstein, S. (2001). *Scan Statistics*. Springer.
- [Glaz et al., 2009] Glaz, J., Pozdnyakov, V., and Wallenstein, S. (2009). *Scan Statistics: Methods and Applications*. Birkhäuser Boston.

- [Haiman, 1999] Haiman, G. (1999). First passage time for some stationary processes. *Stochastic Process. Appl.*, 80:231–248.
- [Haiman and Preda, 2006] Haiman, G. and Preda, C. (2006). Estimation for the distribution of two-dimensional discrete scan statistics. *Methodology and Computing in Applied Probability*, 8(3):373–382.
- [Loader, 1991] Loader, C. R. (1991). Large-deviation approximations to the distribution of scan statistics. *Advances in Applied Probability*, pages 751–771.
- [Naiman and Priebe, 2001] Naiman, D. Q. and Priebe, C. E. (2001). Computing scan statistic p values using importance sampling, with applications to genetics and medical image analysis. *J. Comput. Graph. Statist.*, 10:296–328.
- [Naus, 1965] Naus, J. (1965). The distribution of the size of the maximum cluster of points on a line. *J. Amer. Statist. Assoc.*, 60:493–517.
- [Tango and Takahashi, 2005] Tango, T. and Takahashi, K. (2005). A flexibly shaped spatial scan statistic for detecting clusters. *Int J Health Geogr*, 4:11. Tango, Toshiro Takahashi, Kunihiko England Int J Health Geogr. 2005 May 18;4:11.

Table 1: Numerical results for $\mathbb{P}(S_{\bar{G}} \leq \tau)$ for triangular, rectangular and quadrilateral shapes of the scanning window.

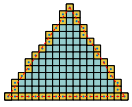
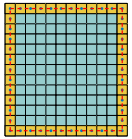

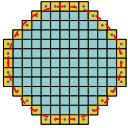
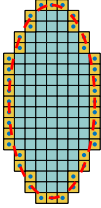
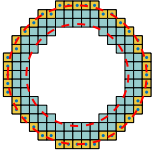
Window shape	Triangle ($m_1 = 14, m_2 = 18, A = 133, IS = 1e5, IA = 1e6$)							
	$X_{i,j} \sim \mathcal{B}(1, 0.01)$				$X_{i,j} \sim \mathcal{B}(5, 0.05)$			
	τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
	3	0.916397	0.918667	0.004333	59	0.863336	0.897101	0.004902
	4	0.997488	0.997483	0.000082	60	0.936684	0.948160	0.002010
	5	0.999947	0.999947	0.000002	61	0.971529	0.974938	0.000894
	6	0.999999	0.999999	0	62	0.987094	0.988279	0.000412
	7	0.999999	0.999999	0	63	0.994372	0.994664	0.000192
	8	1.000000	1.000000	0	64	0.997563	0.997643	0.000089
	9	1.000000	1.000000	0	65	0.998946	0.998971	0.000041
	10	1.000000	1.000000	0	66	0.999564	0.999567	0.000018
	11	1.000000	1.000000	0	67	0.999817	0.999820	0.000008
		$X_{i,j} \sim \mathcal{P}(0.25)$				$X_{i,j} \sim \mathcal{N}(0, 1)$		
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal	
59	0.739866	0.820277	0.011813	50	0.911170	0.934999	0.002737	
60	0.871829	0.902957	0.004483	51	0.945219	0.957777	0.001655	
61	0.939577	0.950977	0.001911	52	0.966494	0.972997	0.001026	
62	0.971673	0.975649	0.000890	53	0.979944	0.982929	0.000644	
63	0.987075	0.988206	0.000425	54	0.987840	0.989469	0.000406	
64	0.994104	0.994492	0.000204	55	0.992768	0.993477	0.000257	
65	0.997384	0.997452	0.000098	56	0.995667	0.996022	0.000162	
66	0.998821	0.998855	0.000046	57	0.997412	0.997574	0.000102	
67	0.999489	0.999490	0.000022	58	0.998509	0.998563	0.000063	
Window shape	Rectangle ($m_1 = 11, m_2 = 12, A = 132, IS = 1e5, IA = 1e6$)							
	$X_{i,j} \sim \mathcal{B}(1, 0.01)$				$X_{i,j} \sim \mathcal{B}(5, 0.05)$			
	τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
	3	0.947896	0.948226	0.002110	59	0.856993	0.857094	0.005485
	4	0.997983	0.997986	0.000058	60	0.919416	0.919589	0.002300
	5	0.999943	0.999943	0.000001	61	0.956420	0.956569	0.001024
	6	0.999999	0.999999	0	62	0.977142	0.977065	0.000471
	7	0.999999	0.999999	0	63	0.988228	0.988208	0.000220
	8	1.000000	1.000000	0	64	0.994130	0.994095	0.000103
	9	1.000000	1.000000	0	65	0.997121	0.997107	0.000048
	10	1.000000	1.000000	0	66	0.998610	0.998607	0.000022
	11	1.000000	1.000000	0	67	0.999342	0.999342	0.000010
		$X_{i,j} \sim \mathcal{P}(0.25)$				$X_{i,j} \sim \mathcal{N}(0, 1)$		
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal	
59	0.764372	0.764495	0.013285	50	0.865858	0.865484	0.004572	
60	0.857872	0.859289	0.005320	51	0.904555	0.904805	0.002691	
61	0.918972	0.918732	0.002307	52	0.933323	0.933206	0.001620	
62	0.954682	0.954579	0.001059	53	0.953950	0.953807	0.000993	
63	0.975271	0.975391	0.000502	54	0.968340	0.968483	0.000617	
64	0.986996	0.986966	0.000242	55	0.978632	0.978687	0.000386	
65	0.993240	0.993261	0.000117	56	0.985791	0.985752	0.000242	
66	0.996557	0.996551	0.000056	57	0.990621	0.990585	0.000152	
67	0.998290	0.998283	0.000027	58	0.993837	0.993815	0.000096	
Window shape	Quadrilateral ($m_1 = 14, m_2 = 18, A = 131, IS = 1e5, IA = 1e6$)							
	$X_{i,j} \sim \mathcal{B}(1, 0.01)$				$X_{i,j} \sim \mathcal{B}(5, 0.05)$			
	τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
	3	0.926068	0.927398	0.003806	59	0.914546	0.927613	0.002942
	4	0.997622	0.997627	0.000075	60	0.959599	0.963873	0.001255
	5	0.999946	0.999946	0.000002	61	0.981235	0.982506	0.000571
	6	0.999999	0.999999	0	62	0.991423	0.991796	0.000266
	7	0.999999	0.999999	0	63	0.996113	0.996233	0.000124
	8	1.000000	1.000000	0	64	0.998283	0.998337	0.000057
	9	1.000000	1.000000	0	65	0.999266	0.999266	0.000026
	10	1.000000	1.000000	0	66	0.999684	0.999684	0.000012
	11	1.000000	1.000000	0	67	0.999868	0.999869	0.000005
		$X_{i,j} \sim \mathcal{P}(0.25)$				$X_{i,j} \sim \mathcal{N}(0, 1)$		
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal	
59	0.835054	0.870351	0.006852	50	0.920004	0.935266	0.002571	
60	0.917972	0.931040	0.002768	51	0.950232	0.957711	0.001556	
61	0.960397	0.964711	0.001237	52	0.968755	0.972594	0.000964	
62	0.981228	0.982451	0.000585	53	0.980695	0.982566	0.000606	
63	0.991142	0.991510	0.000281	54	0.988110	0.989060	0.000383	
64	0.995855	0.995971	0.000136	55	0.992626	0.993110	0.000242	
65	0.998108	0.998124	0.000065	56	0.995569	0.995771	0.000153	
66	0.999135	0.999153	0.000031	57	0.997361	0.997394	0.000096	
67	0.999620	0.999622	0.000014	58	0.998379	0.998435	0.000060	

Table 2: Numerical results for $\mathbb{P}(S_{\tilde{G}} \leq \tau)$ for circular, ellipsoidal and annular shapes of the scanning window.

Window shape	Circle ($m_1 = 13, m_2 = 13, A = 129, IS = 1e54, IA = 1e6$)							
	$X_{i,j} \sim \mathcal{B}(1, 0.01)$				$X_{i,j} \sim \mathcal{B}(5, 0.05)$			
	τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
	3	0.950311	0.950461	0.002195	59	0.920229	0.920388	0.002318
	4	0.998118	0.998114	0.000059	60	0.956814	0.957143	0.001016
	5	0.999947	0.999947	0.000001	61	0.977460	0.977614	0.000462
	6	0.999999	0.999999	0	62	0.988568	0.988567	0.000214
	7	0.999999	0.999999	0	63	0.994312	0.994309	0.000099
	8	1.000000	1.000000	0	64	0.997229	0.997228	0.000046
	9	1.000000	1.000000	0	65	0.998678	0.998679	0.000021
	10	1.000000	1.000000	0	66	0.999380	0.999381	0.000009
	11	1.000000	1.000000	0	67	0.999715	0.999715	0.000004
$X_{i,j} \sim \mathcal{P}(0.25)$				$X_{i,j} \sim \mathcal{N}(0, 1)$				
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal	
59	0.858454	0.859178	0.005391	50	0.888137	0.887891	0.003485	
60	0.919182	0.919586	0.002310	51	0.921173	0.921549	0.002058	
61	0.955229	0.955388	0.001047	52	0.945761	0.945644	0.001243	
62	0.976023	0.975987	0.000491	53	0.962790	0.962825	0.000760	
63	0.987414	0.987344	0.000234	54	0.974848	0.974878	0.000470	
64	0.993477	0.993502	0.000112	55	0.983235	0.983263	0.000293	
65	0.996706	0.996703	0.000054	56	0.988885	0.988907	0.000182	
66	0.998372	0.998365	0.000025	57	0.992730	0.992734	0.000114	
67	0.999207	0.9992032	0.000012	58	0.995269	0.995287	0.000071	
Window shape	Ellipse ($m_1 = 19, m_2 = 9, A = 135, IS = 1e5, IA = 1e6$)							
	$X_{i,j} \sim \mathcal{B}(1, 0.01)$				$X_{i,j} \sim \mathcal{B}(5, 0.05)$			
	τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
	3	0.944001	0.944211	0.002297	59	0.764871	0.763482	0.009128
	4	0.997757	0.997758	0.000069	60	0.860903	0.860548	0.004127
	5	0.999935	0.999935	0.000002	61	0.921089	0.920882	0.001941
	6	0.999998	0.999998	0	62	0.956735	0.956866	0.000934
	7	1.000000	1.000000	0	63	0.977118	0.977094	0.000452
	8	1.000000	1.000000	0	64	0.988182	0.988152	0.000218
	9	1.000000	1.000000	0	65	0.994044	0.994012	0.000104
	10	1.000000	1.000000	0	66	0.997037	0.997037	0.000049
	11	1.000000	1.000000	0	67	0.998554	0.998558	0.000023
$X_{i,j} \sim \mathcal{P}(0.25)$				$X_{i,j} \sim \mathcal{N}(0, 1)$				
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal	
59	0.638962	0.639102	0.019442	50	0.843156	0.844113	0.004369	
60	0.768283	0.769666	0.008610	51	0.887477	0.887692	0.002755	
61	0.861614	0.860885	0.004012	52	0.920601	0.920385	0.001757	
62	0.919144	0.919301	0.001948	53	0.944398	0.944328	0.001127	
63	0.954941	0.954864	0.000965	54	0.961682	0.961667	0.000725	
64	0.975255	0.975369	0.000481	55	0.973797	0.973864	0.000468	
65	0.986869	0.986900	0.000240	56	0.982232	0.982330	0.000301	
66	0.993115	0.993127	0.000119	57	0.988183	0.988132	0.000193	
67	0.996485	0.996472	0.000058	58	0.992138	0.992134	0.000123	
Window's shape	Annulus ($m_1 = 17, m_2 = 17, A = 124, IS = 1e5, IA = 1e6$)							
	$X_{s_1, s_2} \sim \mathcal{B}(1, 0.01)$				$X_{s_1, s_2} \sim \mathcal{B}(5, 0.05)$			
	τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
	3	0.881798	0.882489	0.004812	59	0.951170	0.951245	0.000699
	4	0.995434	0.995465	0.000069	60	0.975772	0.975727	0.000255
	5	0.999883	0.999883	0.000001	61	0.988275	0.988270	0.000099
	6	0.999998	0.999998	0	62	0.994460	0.994466	0.000041
	7	0.999999	0.999999	0	63	0.997439	0.997440	0.000017
	8	1.000000	1.000000	0	64	0.998839	0.998840	0.000007
	9	1.000000	1.000000	0	65	0.999484	0.999484	0.000003
	10	1.000000	1.000000	0	66	0.999775	0.999775	0.000001
	11	1.000000	1.000000	0	67	0.999903	0.999903	0.000000
$X_{s_1, s_2} \sim \mathcal{P}(0.25)$				$X_{s_1, s_2} \sim \mathcal{N}(0, 1)$				
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal	
59	0.903956	0.903852	0.002128	50	0.860708	0.860416	0.004097	
60	0.949083	0.949059	0.000735	51	0.904651	0.904644	0.001977	
61	0.973814	0.973844	0.000277	52	0.936077	0.935938	0.000987	
62	0.986909	0.986876	0.000111	53	0.957564	0.957650	0.000508	
63	0.993576	0.993570	0.000047	54	0.972378	0.972311	0.000270	
64	0.996910	0.996907	0.000020	55	0.982139	0.982136	0.000148	
65	0.998539	0.998539	0.000009	56	0.988564	0.988587	0.000082	
66	0.999320	0.999320	0.000004	57	0.992769	0.992769	0.000047	
67	0.999689	0.999689	0.000002	58	0.995471	0.995466	0.000027	