Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Regular paper

A new PSO-optimized geometry of spatial and spatio-temporal scan statistics for disease outbreak detection

Hesam Izakian\textsuperscript{a,*}, Witold Pedrycz\textsuperscript{a,b}

\textsuperscript{a} Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, Canada
\textsuperscript{b} Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

ARTICLE INFO

Article history:
Received 11 September 2011
Received in revised form 23 December 2011
Accepted 4 February 2012
Available online 13 February 2012

Keywords:
Scan statistics
Particle swarm optimization
Scanning window

ABSTRACT

The spatial and spatio-temporal scan statistics proposed by Kulldorff have been applied to a number of geographical disease cluster detection problems. As the shape of the scanning window used in these methods is circular or elliptic, they cannot find irregularly shaped clusters, say clusters occurring along river valleys or in cases where disease transmission is linked to the road network. In this study, we propose a more flexible geometric structure to be used as a spatial or spatio-temporal scanning window. A particle swarm optimization (PSO) is used to optimize the scanning window to determine disease clusters. We evaluated the proposed method over a number of spatial and spatio-temporal datasets (Breast cancer mortality in Northeastern US 1988–1992 and different types of cancer in New Mexico 1982–2007). Experimental results demonstrate that the introduced approach surpasses the results produced by the circular and elliptic scan statistics in terms of efficiency, especially when dealing with irregularly shaped clusters.

1. Introduction

The 2001 anthrax attacks in the United States, the 2002 Severe Acute Respiratory Syndrome (SARS) in southern China and many other unusual health events have motivated public health communities to develop early outbreak detection systems [1]. These systems enable public health officials to manage the spread of disease by some prevention and control activities such as travel restrictions, movement bans on animals, vaccination, etc. In many cities in the United States and some other countries geographical early outbreak detection systems have been established. Also in livestock industries the importance and usage of such systems is on the rise [2].

In general, early outbreak detection techniques can be divided into two main categories, namely (a) statistical tests such as Scan statistics [3], cumulative sum methods [4], and (b) model-based methods such as generalized linear mixed models [5] and Bayesian models [6].

Statistical tests in disease surveillance systems attempt to determine whether the disease incidence in a spatial and/or temporal defined subset in comparison with the disease incidence in the whole study region is unusual. Thus, this class of methods is designed to detect clusters of disease in space and/or time [7]. Scan statistics originally developed by Naus [3] is used for temporal clustering to test whether the number of disease cases in a temporal subset exceed the expectation given a null hypothesis of no outbreak. Kulldorff [8] developed a spatial scan statistics to test geographical clusters and identify their approximate location. In this method, many circular windows with different centers and radii are checked to detect clusters, where the number of disease cases inside the cluster is maximized, while the number of disease cases outside the cluster is minimized. The spatial scan statistics is not suitable to detect spatio-temporal clusters. This method does not consider the time component and it cannot detect recently emerging clusters. To address this problem, Kulldorff [9] extended the spatial scan statistics to spatio-temporal scan statistics, which can be used to detect recent emerging (viz. alive) clusters. In this method, cylindrical search areas are used where, the spatial search area is quantified by a radius of the cylinder, and the temporal search area is defined by its height. Moreover Kulldorff et al. [10] proposed an elliptic version of the spatial scan statistics by using an elliptic scanning window of variable location, shape (eccentricity), angle, and size. This method is more efficient to find non-circular (narrow) shaped clusters. The idea has been implemented as a software package SaTScan [11], which is widely used as an efficient outbreak detection system.

As the shape of the scanning window used in scan statistics is circular or elliptic, it cannot find irregularly shaped clusters, say clusters occurring along river valleys or in cases where disease...
transmission is linked to the road network. In this study, we propose a more flexible geometric structure that can be used as an efficient scanner window to find irregular shaped clusters in spatial and spatio-temporal domains. More specifically, we build upon a generalization of a circular shape, which is divided into a certain number of sectors where each sector comes with its own radius. To take advantage of the significant flexibility introduced in this manner, the scanner windows need to be optimized. Given the nature of the optimization problem arising in this way, we consider the use of the Particle Swarm Optimization (PSO) being regarded as a flexible population-based optimization tool [12].

It is instructive to take a quick look at the methods used in forming clusters of irregular geometry, especially those exploiting mechanisms of Evolutionary Computing. Duczmal and Assunção, [13] used simulated annealing to find spatial clusters with irregular shapes. In this method, the regions of the studied map are considered as the nodes of a graph so that the adjacent regions are connected in the graph. Then the problem is converted to finding a connected sub-graph with maximum likelihood ratio. The idea of graph representation of the map, has gained a lot of attention among other researchers. Along with well-recognized advantages, the proposed method comes with some shortcoming. First, it usually finds peculiar (very irregular) shapes. Second, finding a medium to large size sub-graph with maximum likelihood ratio is a computational intensive task. To address this problem, Duczmal et al. [14] proposed a genetic algorithm that uses a fast offspring generation and an efficient evolution to generate different connected sub-graphs. The authors used a penalty function to avoid finding peculiar clusters. Tango and Takahashi [15] used graph representation as well and proposed a new method to construct connected sub-graphs. They used a \((k-1)\)-neighbor approach to select adjacent nodes in graphs, which controls the irregularity of clusters in shapes. In [16], Patil and Tailie used the notion of “upper level set” to reduce the size of scanning windows and proposed an “upper level set scan statistics”. The authors did not discuss how to select the level and no comparison with the circular scan statistics is reported.

Continuing along the ideas outlined in [15], Takahashi et al. [1] proposed a flexible spatio-temporal scan statistics. They compared their method with the circular scan statistics. Also they used the \((k-1)\)-nearest neighbor technique to control the irregularity of clusters in shape.

All of the above works (except [16]) used the graph-based representation which comes with computational intensive methods. Also there is another problem: in these methods the parameter “selectable population at risk”, which is the maximum rate of population that can be contained by an individual cluster, is not adjustable and instead the user can adjust only the number of selectable nodes (regions in maps).

The proposed method in this paper does not use the graph-based representation and can be used in both spatial and spatio-temporal form. Also in our method we can control the irregularity of the detected clusters so the algorithm does not have problems of finding peculiar clusters. Also the parameter “selectable population at risk” is adjustable by the user the same as circular and elliptic scan statistics.

This study is organized as follows. In Section 2, spatial and spatio-temporal scan statistics are briefly discussed. In Section 3, the PSO is briefly presented with emphasis on its use to the problem at hand. Section 4 describes the proposed method while Section 5 covers experimental results. Finally, Section 6 concludes this work.

2. Spatial and spatio-temporal scan statistics—an overview

The spatial and spatio-temporal scan statistics [8,9] are widely used in disease surveillance systems for geographical cluster detection and evaluation. Thyroid cancer among men in New Mexico [9], breast cancer mortality in Northeastern United States [17], Breast cancer mortality in Texas [18], late stage breast and colorectal cancer in Minnesota [19], are examples of such studies. In this section, spatial and spatio-temporal scan statistics are briefly discussed.

2.1. Spatial scan statistics

In spatial scan statistics, a window (circular, elliptic etc.) of variable shape and size moves across a geographical region (e.g. in Cartesian or Latitude/Longitude coordinates). Each window encompasses a set of regions in the map and defines a candidate cluster area. For each cluster area, the likelihood ratio is calculated based on the observed and expected number of disease cases inside and outside that area. The area with the maximum likelihood ratio defines the most likely cluster, which is the cluster, least likely to have occurred by chance [10].

Assume that a study region is partitioned into \(m\) fixed cells or sub regions (e.g. cities in a province or counties in a city) with a total of \(C\) disease cases and total \(N\) population. The zone \(z\) is a set of connected cells (sub regions), which are located inside the scanning window. The null hypothesis states that there is no cluster in the map. The alternative hypothesis states that there is at least one cluster in the map. Assume that the number of cases in each cell is distributed according to the Poisson distribution. The Likelihood ratio of \(z\), \(LR(z)\), is expressed as follows [8]:

\[
LR(z) = \frac{L(z)}{L_0} = \begin{cases} \frac{c_z}{\mu_z} & \frac{c_z - c_0}{\mu_z} > c_0 \mu_z \\ 1 & \text{otherwise} \end{cases}
\]

Here \(L_0\) and \(L(z)\) are the likelihood values under the null hypothesis and the likelihood under the alternative hypothesis, respectively. \(c_z\) is the number of cases occurred inside zone \(z\) and \(\mu_z\) is the expected number of cases inside that zone.

The scan statistics aims at detecting zones with the maximum likelihood ratio, which are most likely clusters. In practice, we use the logarithm of \(LR(z)\) abbreviated as \(LLR(z)\). After finding the most likely cluster (the cluster with the maximum likelihood ratio), we determine if this cluster has been obtained by chance or it is a real cluster. For this purpose the statistical significance of the detected cluster is determined by Monte Carlo simulation [20] expressed in terms of the \(p\)-value [8]. In this method, a large number (e.g. 99, 999 or 9999) of random datasets under the null hypothesis of no clustering is generated using the Poisson distribution whose parameters can be estimated on a basis of the available experimental dataset. Then the maximum likelihood for each random dataset is calculated in exactly the same manner as for the real data. The \(p\)-value is expressed as follows

\[
P-value = \frac{1 + \sum_{i=1}^{k} I(LLR_i \geq LLR(z))}{k + 1}. \tag{2}
\]

In this expression, \(k\) is the number of generated random datasets, \(LLR_i\) is the maximum logarithm likelihood ratio of the \(i\)th random dataset and \(I(.)\) is the indicator function. For the detected most likely cluster, if \(p\)-value < 0.05 the cluster is significant at the 5% significant level [8].

The circular scan statistics considers a circular window on the map and moves the center of the circle over the area so that at different positions the window contains different sets of neighboring census areas. Moreover, a census area is considered in the scanning window if its centroid falls in that window. In circular windows for each centroid of the circle, the radius of the window is varied continuously from 0 to a maximum radius so that the
window never includes more than 50% of the total population [9]. This method creates a large number of circular windows, with different location and radius.

The elliptic spatial scan statistics is similar to the circular spatial scan statistics, but instead of circular windows with different location and radius we have ellipses with different size, location, eccentricity, and orientation [10].

2.2. Spatio-temporal scan statistics

To detect clusters in spatio-temporal data, the spatial scan statistics can be used. But this method considers just the number of disease cases in each region and neglects the time component in spatio-temporal data. As a result, this method is not appropriate to detect recently emerging clusters. To address this problem, the spatio-temporal scan statistics has been proposed. In circular spatio-temporal scan statistics instead of a circular window located in two dimensions, a cylindrical window formed in three dimensions is used, where the spatial search area is defined by cylinder radius, and the temporal search area is defined by cylinder height. Since we are going to find “alive” clusters (prospective analysis), candidate cylinders are limited to those that start at any time during the study period and end at the current time period. In elliptic spatio-temporal scan statistics, like the cylindrical the spatial search area is defined by ellipses and the temporal search area is defined by its height.

An important limitation of the circular and elliptic spatial scan statistics, and circular and elliptic spatio-temporal scan statistics, is the use of circular or elliptic search area over the map, so the power of the scan statistics deteriorates as clusters become more irregular in shape.

3. Particle swarm optimization

Particle swarm optimization (PSO) is a population based stochastic optimization technique inspired by bird flocking and fish schooling originally designed and introduced by Kennedy and Eberhart [21]. The PSO paradigm has been applied successfully to many optimization problems [22–27].

The algorithmic flow of the PSO starts with a population of particles whose positions are the potential solutions of the problem, and the velocities are randomly initialized in the problem search space. In each iteration/generation, the search for optimal position (solution) is performed by updating the particles’ velocities and positions based on a predefined fitness function. The velocity of each particle is updated using two best positions, namely personal best position and neighborhood best position. The personal best position, pbest, is the best position the particle has visited and neighborhood best position, nbest, is the best position the particle and its neighbors have visited. Based on the size of the neighborhoods, two PSO algorithms can be developed. When all of the population size of the swarm is considered as the neighbor of a particle, nbest is called global best (gbest) and if the smaller neighborhoods are defined for each particle, nbest is called local best (lbest). Gbest uses the star neighborhood topology and lbest usually uses ring neighborhood topology. In this topology, each particle in the population is connected to its two immediate neighbors on the basis of particle indices as shown in Fig. 1 [28]. There are mainly two reasons why neighborhoods based on particle indices are preferred: It is computationally inexpensive, and helps the spread of information regarding good solutions to all particles [28].

Moreover, there are two main differences between gbest and lbest with respect to their convergence characteristics [29]. Due to the larger particle interconnectivity of the gbest PSO, it converges faster than the lbest PSO, but lbest PSO is less susceptible to being trapped in local optima. The following expressions are used to update the velocity and position of the particles, respectively.

\[
V_{ki}^{t+1} = V_{ki}^t + c_1 r_1 (pbest_{ki}^t - X_{ki}^t) + c_2 r_2 (nbest_{ki}^t - X_{ki}^t); \quad k = 1, 2, \ldots, P, \quad i = 1, 2, \ldots, D, \quad V_{ki}^t \in R
\]

(3)

\[
X_{ki}^{t+1} = X_{ki}^t + V_{ki}^{t+1}, \quad X_{ki} \in R.
\]

(4)

Here \(V_{ki}^t\) is the \(i\)th component of the velocity of the \(k\)th particle in \(t\)th step, \(X_{ki}^t\) is the \(i\)th element of the position of the \(k\)th particle in \(t\)th step of the algorithm, \(P\) is the number of particles in the swarm, \(D\) is the dimensionality of the search space, while \(r_1\) and \(r_2\) are random values in range 0, 1 sampled from a uniform distribution. Furthermore \(c_1\) and \(c_2\) are positive constants, called acceleration coefficients and control the influence of pbest and nbest on the search process.

In initial studies, both \(c_1\) and \(c_2\) are taken to be equal to 2.0 yielding good results [21]. In some situations for example, when the particles are positioned far from their nbest, the velocities quickly reach very large values and as a result, after updating the position of particle it will leave the boundary of the search space. To control this problem, velocity clamping is used in (3) so that, if the right side of this expression exceeds a specified maximum value \(V_{max}\) then the velocity should be clamped to \(V_{max}\). In [30], Maurice proposed the use of a constriction factor \(\chi\) and a modified version of (3) as follows:

\[
V_{ki}^{t+1} = \chi [V_{ki}^t + c_1 r_1 (pbest_{ki}^t - X_{ki}^t) + c_2 r_2 (nbest_{ki}^t - X_{ki}^t)]; \quad k = 1, 2, \ldots, P, \quad i = 1, 2, \ldots, D, \quad V_{ki}^t \in R
\]

(5)

where

\[
\chi = 2 \left(2\varphi - \sqrt{\varphi^2 + 4\varphi^2}\right)^{-1}, \quad \varphi = c_1 + c_2 > 4.
\]

(6)

The intuition behind the use of the constriction factor is to prevent the velocity from growing out of bounds, thus the velocity clamping is not required [31]. However, the best performance can be achieved with the constriction factor while using velocity clamping [32].

4. PSO-based spatial and spatio-temporal scan statistics

In this study, we propose an efficient approach, which can find irregular shaped spatial and spatio-temporal clusters. For this purpose, we proposed a flexible geometric structure that can be used as a scanning window to find irregular shaped clusters.

4.1. A flexible geometric shape

We form a flexible geometric shape by dividing a circle into a number of sectors with equal size angles and different size radii.
The shape can be represented in the spatial domain by the following vector

\[ [x, y, r_1, r_2, \ldots, r_m]^T \]  

where \( x \) and \( y \) represent the center of the circle in the map, \( r_i \) is the radius size of the \( i \)th sector and \( m \) is the number of sectors. This geometric structure is very flexible to build irregular shapes. This flexibility comes with tangible advantages. We can create a very irregular shape by dividing the circle into more sectors and adjusting their radii. Moreover, since equal size angles are used for each sector, the proposed structure can be used simply in the map as a spatial scanner. In this method, by knowing the center of the circle, the number of sectors and radius size for each sector, we can determine which regions in the map are inside the scanner and which regions are not.

In the above structure, by adding time coordinate, \( \text{time}_\text{len} \) to (7), the proposed structure can be converted to a spatio-temporal scanner as follows:

\[ [x, y, \text{time}_\text{len}, r_1, r_2, \ldots, r_m]^T. \]  

Similar to circular and elliptic spatio-temporal scan statistics, \( \text{time}_\text{len} \) in (8) represents the height of the scanner. We will use (7) and (8) as spatial and spatio-temporal scanners in the PSO based scan statistics.

4.2. PSO in the determination of irregularly shaped clusters

As mentioned in Section 2, in scan statistics many scanners with different locations and sizes should be investigated in the map. The set of regions in the map that are inside each scanner, form a cluster. For each cluster the likelihood ratio can be calculated using (1), and the cluster with the highest likelihood ratio is considered as the most likely cluster. In circular scan statistics all scanner windows with different locations and radii (and height for spatio-temporal scan statistics) should be checked. Also for elliptic scan statistics, all scanner windows with different location, radii, angle and eccentricity (and height for spatio-temporal scan statistics) should be investigated. Unlike circular shapes, the number of elliptic shapes can be very high. To cope with this problem, Kulldorff et al. [10] proposed that, checking only a little number of different eccentricities to cover different elliptic scanners is enough. Moreover based on the size of eccentricity, only a little small number of angles should be investigated.

Therefore the number of circles or ellipses in circular and elliptic scan statistics is finite and investigating all of the scanners in these methods is possible. In our proposed shape, with increasing the number of sectors, the number of various scanners will increase exponentially and will be infinite. Therefore investigating all of such different scanners in the map is not possible and is an NP-complete problem. Thus we consider the problem, finding the optimal shaped scanner window with optimal location in the map, as an optimization problem, and use particle swarm optimization to solve it.

Particle encoding

One of the essential issues in designing a PSO algorithm is finding an appropriate mapping between problem solution and the particle’s representation. In this study, each particle represents an irregular scanner window (see Fig. 2) in a location of the map. For the spatial scan statistics, two elements are needed to represent the centroid (location) of the scanner and \( m \) elements are required to represent the radius of sectors (\( m+2 \) in total). In addition, for the spatio-temporal scan statistics the height of the scanner should be considered too. Fig. 3 illustrates the structure of each particle in PSO algorithm to find spatio-temporal clusters. In this structure \( x_{\text{axis}} \) and \( y_{\text{axis}} \) can be any arbitrary \( x-y \) coordinates in the map, \( \text{sector} \), represents the radius size of the \( i \)th sector which can be an integer between 0 and the distance between two farthest regions in the map. \( \text{time}_\text{len} \) is an integer between 1 and the range of time length of the study period and is used for spatio-temporal clusters. In this paper, since we are going to find alive clusters (clusters that occurred recently), the \( \text{time}_\text{len} \) element represents the range \( [T-\text{time}_\text{len}, T] \) where \( T \) is the current time. Notice that this element is not required for spatial clusters.

\( \text{pbest}, \text{nbest} \) and velocity

Since in the PSO algorithm, \( \text{pbest} \) and \( \text{nbest} \) are two positions that include the personal best position and neighborhood best position of each particle, therefore their encoding is similar to the particles’ position. In this paper, we used ring topology for \( \text{nbest} \) (Best PSO). Also for the velocity, the same as position, \( m+2 \) elements is used for spatial scan statistics and \( m+3 \) elements are used for spatio-temporal scan statistics whose elements are in range \([-V_{\text{max}}, V_{\text{max}}]\).

Fitness function

The fitness function is the logarithm of the likelihood ratio (LLR(\( z \)) for the selected region, \( z \), which is a set of regions in the scanning window represented by the particle. The higher likelihood ratio, the greater fitness value, and hence a quality of the solution is regarded higher.

Parameter settings

In order to optimize the performance of the proposed method, fine-tuning has been performed and the following ranges of the parameter values were tested:

\[ c_1 \text{ and } c_2 = 2, 3, P = [100, 5000], \text{maximum iterations/generations} = [100, 5000], \text{number of sectors} = [1, 2, 4, 8, 16, 32, 64]. \]

Based on experimental results on the two different benchmarks (refer to the next section), the PSO performs the best for the following settings: \( c_1 = c_2 = 2.05, \chi = 0.729 \) (see (6)), number of particles \( P = 1000 \), the stopping condition is the maximal number of generations \( k = 1500 \). For the velocity, \( V_{\text{max}}(x_{\text{axis}}) \) is the maximum \( x \)-axis in the map, \( V_{\text{max}}(y_{\text{axis}}) \) is the maximum \( y \)-axis in the map, \( V_{\text{max}}(\text{time}_\text{len}) \) is the length of time component in data. For example, if the data were collected on a yearly basis from 1990 to 2010, the time length is 20. \( V_{\text{max}}(\text{sector}) \) is the distance between two farthest regions in the map so that each sector based on its direction can be extended in the map. We set the parameter “selectable population at risk” to 50% and the number of sectors to 64. We also investigate the effect of these parameters on detected clusters.

Selecting regions within the scanning window

To consider a region inside the scanner, there are three alternatives. In the first method, a region is considered as a region inside the scanner, if the entire region is inside the scanner. In the second method, if any part of the region is inside the scanner then the whole region will be considered inside the scanner. In the third method which is used in the circular and elliptic scan statistics by default [11], a region will be considered inside the scanner,
if its centroid is inside the scanner. Fig. 4 shows the selected regions in the map based on the three mentioned methods when number of sectors is one (circular scanner). For implementing the first and second methods, one way is defining set of coordinates in different parts of each region to check whether a region (its defined coordinates) is inside a cluster or not. The advantage of these methods is that we can make sure that the selected regions by the scanner are connected. In other words the selected regions are set of neighbors in the map. The limitation of this method is that the flexibility of selecting various regions will be decreased. Also checking a lot of coordinates for each region to see whether a region is inside a cluster or not, is a time consuming process.

Using the third method, increases the flexibility of selecting different regions in different locations of the map, but it can lead to selecting a set of disconnected regions, especially when the number of sectors in the scanning window is high. In this method, before evaluating the selected regions by scanner, the connectivity of the regions should be tested. In this paper, we used the third method and to check the connectivity of the regions inside a scanner we use the algorithm outlined in Table 1.

In the proposed method, similar to the circular and elliptic scan statistics, the scanner never includes more than 50% the total population at risk. This will lead to an important effect on the detected clusters that is further investigated in our experiments.

**Overall scheme of the proposed method**

In summary, our proposed method works as follows: at the first step of the algorithm, numbers of particles with their velocities are generated randomly. Each particle represents an irregular scanner which encompasses set of regions in the studied map. In the second step, for each particle, the regions inside its corresponding scanner should be considered. If the selected regions for each particle are connected regions, and the population of the selected regions is less than or equal to 50% of the total population of the studied map, then the fitness value (LLR(z)) should be calculated. Next, the pbest and lbest for each particle should be updated and based on their values the velocity will be updated using (5), and then the particle position can be updated using (4). Table 2 shows the pseudo-code of the method.

### 5. Experimental studies

In this section, we present the experimental results produced by the proposed method on Breast cancer mortality in northeastern US 1988–1992 which is a spatial dataset. The second dataset concerning different types of cancer in New Mexico 1982–2007 is the spatio-temporal dataset. Both datasets are available from the Surveillance, Epidemiology, and End Results (SEER) Program at the National Cancer Institute (www.seer.cancer.gov/data/). The SEER program collects cancer incidence and mortality from the cancer registries in the United States.

We start with a description of the data and discuss the effect of “number of sectors” on the likelihood ratio and the shape of the detected clusters. In the sequel we investigate the effect of “selectable population at risk by the scanner” on the detected clusters. Finally, we compare the results obtained by running the method developed in the study with those produced by circular and elliptic scan statistics.

#### 5.1. Datasets

**Breast cancer mortality in Northeastern United States 1988–1992**

To investigate the performance of our proposed method as spatial scan statistics, we applied it to the breast cancer mortality data from Northeastern United States, 1988–1992. In this dataset the northeastern US map contains 245 counties in ten states and the District of Columbia, with a total population of 29,535,210 women with 58,943 breast cancer deaths. The mortality rate was 39.9 per 100,000 women per year [10].

All analyses completed for this dataset use the Poisson model, and the observed number of deaths in a county is modeled according to Poisson distribution. The analyses are adjusted for age applying indirect standardization [8] with 18 distinct five-year age groups: 0–4, 5–9, . . . , 80–84, and 85+. This data has been previously analyzed using the circular spatial scan statistics [8] as well as the elliptic spatial scan statistics [10].

**Cancer in New Mexico 1982–2007**

To investigate the performance of our proposed method as spatio-temporal scan statistics, we used the incidence of different types of cancers including Breast, Liver, Lung, Lymphoma, Prostate, Skin, Stomach and Thyroid in New Mexico during 1982–2007.

These data are geographically aggregated into 33 counties in New Mexico. Similar to the mortality dataset, all analyses in these datasets use the Poisson model, where, for each year the observed number of cancers in a county is modeled following the Poisson
distribution. The analyses are adjusted for age, sex, and race applying indirect standardization with 18 distinct five-year age groups: 0–4, 5–9, . . . , 80–84, and 85+, 2 distinct sex groups: male and female, and 3 distinct race groups: white, black and others [8].

5.2. Effect of “number of sectors” on the likelihood ratio and cluster shape in PSO scan statistics

It is worth investigating the effect of “number of sectors” in our proposed shape as scanner on likelihood ratio using the mentioned datasets. For the Mortality dataset spatial analysis is used, and for the other datasets spatio-temporal analysis is performed. Since in meta-heuristic algorithms such as PSO, the final solutions are not deterministic, we ran the PSO algorithm 10 times for each case to obtain representative results as well as evaluate the associated diversity of the solutions. Table 3 shows the fitness values (LLR) for several selected number of sectors, namely 1, 2, 4, 8, 16, 32, and 64. For each entry in this table, the first number is the best result between 10 independent runs, the second is the average and the third is the standard deviation.

Evidently, as it could have been expected, in most cases the highest number of sectors (64) produces the best results. By increasing the number of sectors, the ability to find clusters with higher irregularity in shape increases and if the real cluster in the map is very irregular, a higher value of the likelihood ratio can be achieved. As it can be seen from this table, for the Breast dataset, the produced likelihood ratios are the same for different number of sectors. The reason is that the cluster in this dataset is circular in shape. Moreover, in some datasets (e.g., Lymphoma and Stomach) four sectors can achieve equally good results as those cases that use 64 sectors. The reason is that the cluster in these datasets is quite regular in shape so only four sectors are sufficient.

Furthermore, as it can be seen from Table 3, in most cases the PSO approach produces a low standard deviation which indicates that the proposed method produces stable solutions. However with increasing the number of sectors in the proposed geometric structure, the standard deviation increases. The reason is that the length of each particle increases, and as a result, the search space becomes larger. This leads to finding less stable solutions quantified by the associated higher standard deviation. Moreover in some datasets, fewer sectors have higher standard deviations. The reason is that for these datasets, the corresponding scanner gets to local optima and therefore the PSO finds more variable solutions.

Table 3

| Dataset       | Number of sectors |
|---------------|-------------------|
|               | 1     | 2     | 4     | 8     | 16    | 32    | 64    |
| Mortality     | 48.05 | 63.23 | 67.03 | 94.20 | 108.39| 121.89| 121.89|
|               | 0.00  | 1.69  | 0.00  | 4.09  | 1.82  | 7.90  | 5.65  |
| Breast        | 345.39| 345.39| 345.39| 345.39| 345.39| 345.39| 345.39|
|               | 82.41 | 82.41 | 89.21 | 92.71 | 95.43 | 95.43 | 95.43 |
| Liver         | 122.98| 122.98| 128.15| 149.11| 156.06| 161.67| 160.05|
| Lung          | 122.98| 122.98| 124.10| 142.06| 153.71| 157.45| 160.07|
| Lymphoma      | 24.82 | 24.82 | 27.01 | 27.01 | 27.01 | 27.01 | 27.01 |
| Prostate      | 222.25| 222.25| 226.82| 226.82| 226.82| 226.82| 226.82|
| Stomach       | 12.33 | 18.20 | 21.46 | 21.46 | 21.46 | 21.46 | 21.46 |
| Skin          | 25.11 | 25.72 | 27.58 | 28.26 | 28.26 | 28.26 | 28.26 |
| Thyroid       | 171.40| 178.43| 181.67| 181.67| 181.67| 181.67| 186.70|
|               | 0.00  | 2.85  | 0.85  | 1.33  | 0.00  | 0.64  | 2.19  |
7

(a) Number of sectors = 1, 2. (b) Number of sectors = 4. (c) Number of sectors = 8. (d) Number of sectors = 16.

Fig. 5. Detected clusters versus the number of sectors for Prostate dataset.

Table 4

| Dataset   | Population rate at risk |
|-----------|-------------------------|
|           | 10% | 20% | 30% | 40% | 50% |
| Mortality | 73.28 104.66 121.89 121.89 121.89 |
|           | 69.86 96.59 112.46 110.08 109.50 |
| Breast    | 58.75 90.83 90.83 281.26 345.39 |
|           | 58.75 90.83 90.83 281.26 345.39 |
| Liver     | 16.38 24.43 35.68 76.27 112.46 |
|           | 16.34 24.41 33.91 73.69 110.08 |
| Lung      | 104.195 137.901 149.41 163.48 166.05 |
|           | 104.195 134.813 149.41 160.11 160.07 |
| Lymphoma  | 7.58 8.96 14.10 21.63 27.74 |
|           | 7.58 8.74 13.84 21.01 27.71 |
| Prostate  | 63.45 86.48 100.05 180.53 244.21 |
|           | 63.45 82.92 96.06 178.91 240.19 |
| Stomach   | 21.46 21.46 21.46 21.46 21.46 |
|           | 21.46 21.46 21.46 21.46 21.46 |
| Skin      | 5.75 8.01 8.57 17.98 28.26 |
|           | 5.75 7.60 7.90 17.62 28.26 |
| Thyroid   | 38.70 57.79 104.84 152.19 186.70 |
|           | 38.02 56.71 99.33 151.68 182.57 |
|           | 0.70 1.18 13.48 0.37 2.19 |

Fig. 5 shows the best detected clusters (highest LLR value) by different number of sectors, for the Prostate dataset. It can be seen that the parameter “number of sectors” can control the irregularity of the detected clusters and lower sectors will lead to less irregular clusters in shape.

5.3. Effect of “selectable population at risk” on likelihood ratio and cluster shape

In this sub-section we investigate the impact of the parameter “selectable population at risk” on the detected clusters. In circular and elliptic scan statistics this parameter is set to 50% by default (SaTScan [11]). This means that the scanner window never includes more than 50% of the total population of the studied regions.

Table 4 shows the fitness value (LLR) obtained for different datasets and for selectable population at risk 10%, 20%, 30%, 40%, and 50%. For each entry, the first number is the best result between 10 independent runs, the second is the average and the third is the standard deviation. As it can be seen, for the dataset Stomach, the likelihood ratios for different population rates are the same. The reason is that the cluster occurred in regions with low population. The situation for other datasets is different and with increasing the population rate, the likelihood ratio increases because the cluster occurred in higher population rates. For the mortality dataset, 30% of the population rate has the highest average, because the cluster has occurred in this rate, and since the problem search space for this rate is less than the problem search space for the 40% and 50% rates, PSO can detect the clusters with high fitness values in different runs without trapping into local optima.

In Table 4, we cannot observe any relationship between the level of standard deviation and the selectable population rate. Actually, this amount depends upon the shape of the cluster. For some datasets the parameter, “selectable population rate”, has a large effect on the shape of the cluster (see Fig. 6). Increasing the irregularity of the clusters leads to higher values of the standard deviation.

Fig. 6 shows the selected regions by the PSO (best solution obtained for 10 independent runs) based on different population rates for Prostate dataset. Notice that the county “Bernalillo” contains about 29% of the population of New Mexico in our dataset. We can see that the parameter “selectable population at risk” has an important effect on detected clusters and to find true clusters, this parameter at least should be set to the population of the county with maximum population in the map (e.g., 29% for New Mexico datasets).

5.4. Comparative analysis

In this experiment, as in case of the circular and elliptic scan statistics we set the parameter “selectable population at risk” to 50%. Also we set the number of sectors to 64. Table 5 shows the results of comparison between circular, elliptic and PSO scan statistics on spatial (mortality) and spatio-temporal (different types of cancer in New Mexico) datasets. For the PSO approach, the result presented here is the best one among 10 independent runs. We can see that in comparison with circular and elliptic scan statistics, in most cases our method has achieved a very higher likelihood ratio.

For Breast cancer, the cluster is circular in shape so the LLR value, which achieved by PSO is the same as circular and elliptic methods. Also in some datasets e.g. Stomach, the LLR value achieved by the PSO is very close to the one produced by the elliptic scan statistics. This is not surprising as the real cluster shape is elliptic.

In Table 5 for the two datasets, Lung and Thyroid, the cluster period in PSO scan statistics is different with circular and elliptic methods. When the adjacent counties are selected in an irregular form using PSO, the optimal likelihood ratio of the selected regions...
Fig. 6. Selectable population at risk versus selected regions in Prostate dataset. Note that the county Bernalillo contains about 29% population of New Mexico.

### Table 5
Comparison of circular, elliptic and PSO-based scan statistic for spatial and spatiotemporal datasets. The mortality dataset is spatial and the others are spatiotemporal ones.

| Dataset | Method | Counties | Cluster period | Cases | Expected | LLR  | p-value |
|---------|--------|----------|----------------|-------|----------|------|---------|
| Mortality | Circular | PADelaware, PAMontgomery, PAPhiladelphia | NA | 3507 | 2972.10 | 48.05 | 0.001 |
|         | Elliptic | NJBergen, NJEssex, NJSomerset, NJUnion, NYWestchester, PAPhiladelphia | NA | 4502 | 3936.79 | 78.90 | 0.001 |
|         | PSO     | NJAtlantic, NJBergen, NJBurlington, NJEssex, NJMercer, NJMiddlesex, NJMonmouth, NJOcean, NJUnion, NYNassau, NYWestchester, PAPhiladelphia | NA | 11562 | 10107.30 | 121.89 | 0.001 |
| Breast  | Circular | Bernalillo, LosAlamos, Sandoval, SantaFe | 1987–2007 | 12766 | 10634.01 | 122.98 | 0.001 |
|         | Elliptic | Bernalillo, LosAlamos, Sandoval, SantaFe | 1987–2007 | 12766 | 10634.01 | 122.98 | 0.001 |
|         | PSO     | Bernalillo, LosAlamos, Sandoval, SantaFe | 1987–2007 | 12766 | 10634.01 | 122.98 | 0.001 |
| Liver   | Circular | Bernalillo, Cibola, McKinley, Sandoval, Socorro, Torrance, Valencia | 1992–2007 | 979 | 692.91 | 82.41 | 0.001 |
|         | Elliptic | Bernalillo, Cibola, Guadalupe, McKinley, Sandoval, Socorro, Torrance, Valencia | 1992–2007 | 988 | 697.06 | 84.98 | 0.001 |
|         | PSO     | Bernalillo, Guadalupe, McKinley, RioArriba, Sandoval, Taos, Torrance, Valencia | 1992–2007 | 1029 | 718.62 | 95.43 | 0.001 |
| Lung    | Circular | Chaves, Curry, DeBaca, Eddy, Lea, Quay, Roosevelt | 1987–2007 | 3336 | 2568.93 | 122.98 | 0.001 |
|         | Elliptic | Chaves, Curry, DeBaca, Eddy, Lea, Quay, Roosevelt | 1987–2007 | 3336 | 2568.93 | 122.98 | 0.001 |
|         | PSO     | Chaves, Cibola, Curry, DeBaca, Eddy, Guadalupe, Lea, Otero, Quay, Roosevelt, Sandoval, SanJuan, Sierra, Torrance, Valencia | 1987–2007 | 7872 | 6662.39 | 166.05 | 0.001 |
| Lymphoma | Circular | Bernalillo, LosAlamos, Sandoval, SantaFe | 1990–2007 | 2473 | 2201.75 | 24.82 | 0.001 |
|         | Elliptic | Bernalillo, LosAlamos, Sandoval, SanJuan, SantaFe | 1990–2007 | 2723 | 2434.20 | 27.01 | 0.001 |
|         | PSO     | Bernalillo, Curry, DeBaca, Harding, Lincoln, LosAlamos, Mora, Quay, Roosevelt, Sandoval, SantaFe, Union | 1990–2007 | 2821 | 2526.51 | 27.74 | 0.001 |
| Prostate | Circular | Bernalillo, LosAlamos, Sandoval, SantaFe, Valencia | 1990–2007 | 10779 | 9138.03 | 222.25 | 0.001 |
|         | Elliptic | Bernalillo, Chaves, Lincoln, LosAlamos, Sandoval, SantaFe, Torrance | 1990–2007 | 11496 | 9771.17 | 239.61 | 0.001 |
|         | PSO     | Bernalillo, Chaves, Cibola, Lincoln, LosAlamos, Sandoval, SantaFe, Socorro | 1990–2007 | 11786 | 10037.4 | 244.21 | 0.001 |
| Stomach | Circular | Colfax, Harding, LosAlamos, Mora, RioArriba, SanMiguel, Taos | 1982–2007 | 347 | 266.13 | 12.33 | 0.068 |
|         | Elliptic | Guadalupe, Mora, RioArriba, SanMiguel, Taos | 1982–2007 | 347 | 266.13 | 12.33 | 0.068 |
|         | PSO     | Guadalupe, Harding, Mora, Quay, RioArriba, SanMiguel, Taos | 1982–2007 | 335 | 232.90 | 21.46 | 0.049 |
| Skin    | Circular | Bernalillo, Catron, Cibola, Sandoval, SanJuan, Socorro, Valencia | 1986–2007 | 329 | 243.40 | 25.11 | 0.017 |
|         | Elliptic | Bernalillo, Lincoln, Sandoval, SanJuan, Socorro, Valencia | 1986–2007 | 331 | 242.85 | 26.63 | 0.021 |
|         | PSO     | Bernalillo, Lincoln, LosAlamos, Sandoval, SanJuan, Socorro, Valencia | 1986–2007 | 340 | 249.07 | 28.26 | 0.006 |
| Thyroid | Circular | Bernalillo, Cibola, LosAlamos, RioArriba, Sandoval, SantaFe | 2000–2007 | 1070 | 620.90 | 171.40 | 0.001 |
|         | Elliptic | Bernalillo, Catron, Grant, Hidalgo, LosAlamos, RioArriba, Sandoval, Socorro, Valencia | 2000–2007 | 1035 | 589.11 | 174.54 | 0.001 |
|         | PSO     | Bernalillo, Cibola, DeBaca, Guadalupe, Lincoln, LosAlamos, RioArriba, Roosevelt, Sandoval, SantaFe, Socorro | 1998–2007 | 1325 | 816.01 | 186.70 | 0.001 |
can be in a different time period rather than the detected clusters using circular and elliptic methods. Notice that the likelihood ratio of a cluster at any time period is dependent on the number of disease cases and the number of expected disease cases inside and outside cluster at that time period as discussed in (1).

Moreover, in Table 5 we can see that usually the number of selected counties by the PSO is more than the number of selected counties in both, circular and elliptic methods. Since PSO can consider irregular shaped clusters using the proposed flexible structure, it is able to remove the counties with low disease rates from the cluster. Therefore it can add a number of other counties that have high disease rates one by one to the final cluster. As a result the likelihood ratio of the final cluster will be increased without including more than 50% of the total population. On the other hand, because of the shape restrictions in circular and elliptic scanners, the areas with low disease rates cannot be removed from the cluster. Furthermore in these methods, by adding counties with higher disease rates to the cluster, some other counties will be added to the cluster (due to the circular or elliptic shape of cluster) that can lead to decrease in the likelihood ratio in the final cluster, or it will contain more than 50% of the total population in the studied regions (which is not allowed to be considered).

Figs. 7 and 8 show the clusters found by circular, elliptic and PSO scan statistics for Prostate and Liver datasets.

5.5. Convergence behavior of the PSO scan statistics

In this sub-section, we compare the required run time of circular, elliptic and PSO scan statistics. The run time of the circular and elliptic scan statistics depend directly on some parameters that the user has to select. In circular scan statistics, for some predefined locations (x-y coordinates in the map), a circle with centroid (x, y) starts with a radius size of zero. In each step the size of the radius is increased until the maximum radius size has been reached. The increasing amount for each step is selected by the user. Also, to select locations in the map, one method is dividing the map into a n × m grid, and to consider the centroid of blocks as the centroid of the circles. Moreover, for the elliptic scan statistics there are two more parameters; namely eccentricity and angle; the growth of their values that occurs at each step, has to be selected by the user as well.

To compare the run time of the method with circular and elliptic methods, we consider three settings. In the first setting we set the increase size of each parameter in circular and elliptic scan statistics to 5% of the maximum size of that parameter. For example, if the maximum radius size is 100, in this setting in each step we increase the size of radius by 5. The second and third settings are 10% and 20% of the maximum amount of each parameter. Notice that the above settings to select the parameters of circular and elliptic scan statistics is used for comparison studies and is not the best way. Kuldorff et al. in [10] discussed how to select the ellipses parameters in elliptic scan statistics.

For the PSO we set the stopping condition as the maximum generation $k = 1500$ or no change in fitness value in 500 consecutive generations. All approaches are implemented in VC++ + 6 running on 2.4 GHz processor. Table 6 shows the run time (expressed in seconds) and LLR value of circular, elliptic and PSO scan statistics for different datasets. For the PSO the run time is taken as an average of 10 independent runs and the LLR value is the best one among them.

From Table 6, we can see that by selecting a small amount of increment for different parameters in circular and elliptic scan statistics, it can find clusters with higher $LLR$, but the run time of the method will be increased and vice versa. The PSO scan statistics is free from these parameters. The circular scan statistics has the lowest run time in different datasets. Notice that this method, even when using the first setting (5% increment in parameters), it cannot find its optimal solution in the Skin and Thyroid datasets (see Table 5) and a smaller amount of increment is required for these datasets. Also, for the elliptic method there is the same problem for the mortality dataset.

In Table 6 we can see that in spatio-temporal datasets, our method is faster than the elliptic scan statistics with the increase amount of 5% for its parameters. Our method is slower than the elliptic method in the mortality dataset, but the achieved $LLR$ value in our method is much higher.

The run time of our method for the mortality dataset is higher than the cancer datasets. The reason is that the number of counties in the mortality dataset is much higher than the cancer datasets, and as a result, the problem search space for PSO in mortality dataset is larger. On the other hand, unlike to our method, the required run time for mortality dataset in circular and elliptic methods is less than the cancer datasets. The reason is that the three settings (5%, 10% and 20%) that have been used in circular and elliptic methods, their parameters are independent from the size of the map and number of counties. Moreover, in the mortality dataset we do not have the time component and as a result the required run time for the mortality dataset is lower in circular and elliptic methods.

Fig. 9 shows the fitness function values ($LLR$) obtained in successive generations in the PSO approach for the mortality, Prostate and Liver datasets. Most of the improvement appears in the first few hundreds of generations.

The most important advantage of the proposed method is that it can find irregular shaped clusters, with higher likelihood ratio, in comparison with those that are obtained for circular and elliptic scan statistics. We showed that the irregularity of the resulting clusters can be controlled by adjusting the number of sectors in the proposed flexible shape. Moreover, in contrast of the circular and elliptic scan statistics, where the user has to select and tune some parameters (e.g., the increasing amount of radius, eccentricity and angle), our approach is free from such parameters.

On the other hand, since the method is based on a meta-heuristic approach (PSO), the results are not deterministic and can differ from one run to another. We showed, however, that the algorithm comes with a low amount of the standard deviation.
of the produced results, which speaks of their high stability (repeatability). There is also an associated computing overhead; the method runs longer than the circular scan statistics.

6. Conclusions

In this paper, we proposed an irregular shaped scan statistics to detect irregular shaped disease clusters in the map. For this purpose, a flexible geometric structure is used as an irregular scanner window. As the checking all possible irregular shapes in the map resulting from this proposed flexible structure, is not possible and it is an NP-complete problem, a PSO method has been applied. We considered the impact of the two parameters, namely “number of sectors in the proposed structure” and “selectable population at risk” on the likelihood ratio and the shape of the resulting clusters.

There are some considerations that can be sought to improve the method. In some situations (e.g. when the real clusters are circular and small in shape), it is likely that the method discussed here could be trapped in local optima. One could consider selecting the number of sectors dynamically using the PSO algorithm. In this case, we can start with a small number of sectors and gradually increase their number. By having a small number of sectors, the size of the problem search space could be reduced and this will facilitate the PSO search process. Then by increasing the number of sectors the quality of results can be improved. Another
modification that could improve the method is to use some other geometric structures to find irregular clusters, or by using the proposed geometric structure with irregular angles.

Acknowledgment

Support from the Natural Sciences and Engineering Research Council of Canada (NSERC) under a CRD Grant is gratefully acknowledged.

References

[1] K. Takahashi, M. Kulldorff, T. Tango, K. Yih, A flexibly shaped space–time scan statistic for disease outbreak detection and monitoring, Int. J. Health Geogr. (2008) 7–14.
[2] J. Berezowski, Alberta veterinary surveillance network, advances in pork production, in: Proceedings of the 2010 Banff Pork Seminar, vol. 21, 2010, pp. 87–93.
[3] J.J. Naus, The distribution of the size of the maximum cluster of points on a line, J. Amer. Statist. Assoc. 60 (1965) 532–538.
[4] P. Rogerson, Surveillance systems for monitoring the development of spatial patterns, Stat. Med. 16 (18) (1997) 2081–2093.
[5] K. Kleinman, R. Lazarus, R. Platt, A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism, Amer. J. Epidemiol. 159 (3) (2004) 217–224.
[6] N. Best, S. Richardson, A. Thomson, A comparison of Bayesian spatial models for disease mapping, Stat. Methods Med. Res. 14 (1) (2005) 35–59.
[7] C. Robertson, T.A. Nelson, Y.C. MacNab, A.B. Lawson, Review of methods for space–time disease surveillance, Spatial and Spatio-temporal Epidemiology 1 (2010) 105–116.
[8] M. Kulldorff, A spatial scan statistic, Comm. Statist. Theory Methods 26 (6) (1997) 1481–1496.
[9] M. Kulldorff, Prospective time periodic geographical disease surveillance using a scan statistic, J. R. Stat. Soc. Ser. A 164 (1) (2001) 61–72.
[10] M. Kulldorff, L. Huang, L. Pickle, L. Duczmal, An elliptic spatial scan statistic, Stat. Med. 25 (2006) 3929–3943.
[11] SaTScan. http://www.satscan.org.
[12] J. Kennedy, R.C. Eberhart, Swarm Intelligence, Morgan Kaufmann, 2001.
[13] L. Duczmal, R. Assunção, A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters, Comput. Statist. Data Anal. 45 (2004) 269–286.
[14] L. Duczmal, A.L.F. Cançadob, R.H.C. Takahashic, L.F. Bessegatob, A genetic algorithm for irregularly shaped spatial scan statistics, Comput. Statist. Data Anal. 52 (2007) 43–52.
[15] T. Tango, K. Takahashi, A flexibly shaped spatial scan statistic for detecting clusters, Int. J. Health Geogr. (2005) 4–11.
[16] G.P. Patil, C. Taillie, Upper level set scan statistics for detecting arbitrarily shaped hotspots, Environ. Ecol. Stat. 11 (2004) 189–197.
[17] M. Kulldorff, E. Feuer, B. Miller, L. Freedman, Breast cancer in northeast United States: a geographic analysis, Amer. J. Epidemiol. 146 (1997) 161–170.
[18] E.S. Hsu, H.E. Jacobson, F. Soto Mas, Evaluating the disparity of female breast cancer mortality among racial groups—a spatio-temporal analysis, Int. J. Health Geogr. (2004) 3–4.
[19] A.J. Thomas, B.P. Carlin, Late detection of breast and colorectal cancer in Minnesota counties: an application of spatial smoothing and clustering, Stat. Med. 22 (2003) 113–127.
[20] M. Dwass, Modified randomization tests for nonparametric hypotheses, Ann. Math. Statist. 28 (1957) 181–187.
[21] J. Kennedy, R.C. Eberhart, Particle swarm optimization, in: Proceedings of the IEEE International Conference on Neural Networks, 1995, pp. 1942–1948.
[22] H. Izakian, A. Abraham, Fuzzy C-means and fuzzy swarm for fuzzy clustering problem, Expert Syst. Appl. 38 (2011) 1835–1838.
[23] T. Navalértorne, N. Afrulpurka, Optimization of tile manufacturing process using particle swarm optimization, Swarm and Evolutionary Computation 1 (2011) 97–109.
[24] H. Izakian, A. Abraham, V. Snášel, Metaheuristic based scheduling meta-tasks in distributed heterogeneous computing systems, Sensors 9 (2009) 5339–5350.
[25] H. Izakian, B.T. Ladani, A. Abraham, V. Snášel, A discrete particle swarm optimization approach for grid job scheduling, Int. J. Innovative Comput. Inf. Control 6 (9) (2010) 4219–4233.
[26] R. Malviya, D.K. Pratihar, Tuning of neural networks using particle swarm optimization to model MIG welding process, Swarm and Evolutionary Computation 1 (2011) 223–235.
[27] J. Senthilnath, S.N. Omkar, V. Mani, Clustering using firefly algorithm: performance study, Swarm and Evolutionary Computation 1 (2011) 164–171.
[28] A.P. Engelbrecht, Computational Intelligence—An Introduction, John Wiley & Sons Ltd., 2007.
[29] R.C. Eberhart, P.K. Simpson, R.W. Dobbins, Computational Intelligence PC Tools, Academic Press Professional, 1996.
[30] C. Maurice, The swarm and queen: towards a deterministic and adaptive particle swarm optimization, in: Proceedings of the IEEE Congress on Evolutionary Computation, 1999, pp. 1551–1557.
[31] A.W. Mohamed, N.C. Sahoo, Particle swarm optimization combined with local search and velocity re-initialization for shortest path computation in networks, in: 2007 Swarm Intelligence Symposium, 2007, pp. 266–272.
[32] R.C. Eberhart, Y. Shi, Comparing inertia weight and constriction factors in particle swarm optimization, in: Proceedings of the IEEE Congress on Evolutionary Computation, 2000, pp. 84–88.