

THE DAC STATISTIC: PROPERTIES AND USE IN DETECTION OF CLUSTERS

AI. PETRIȘOR, J. W. DRANE, L. DRAGOMIRESCU*

Deși utilizarea sa este limitată, testul DAC, definit ca diferența dintre distribuția empirică a cazurilor prezente într-un eșantion și cea a tuturor observațiilor, prezintă un potențial ridicat pentru a deveni un instrument de detectare a aglomerărilor. Studiile noastre anterioare au fost axate pe sensibilitatea testului la localizarea originii și orientarea axelor de coordonate și pe posibilul rol de instrument pentru detectarea aglomerărilor. Rezultatele arată că testul DAC nu depinde de localizarea originii, dar este influențat de orientarea axelor. Testul DAC nu poate fi un instrument de sine-stătător pentru detectarea aglomerărilor, dar poate fi folosit în acest scop în corelație cu alte metode. Metodele bazate pe tehnicile GIS au fost analizate ca potențial candidat. Rezultatele au indicat faptul că hărțile de densitate depind de parametrii aleși de utilizator, în timp ce diversele tehnici de kriging oferă rezultate mult mai ușor de repetat. Acest articol însumează rezultatele studiilor noastre anterioare, multe dintre acestea nefiind publicate, într-o încercare de a pune bazele unei metodologii care folosește testul DAC împreună cu alte tehnici spațiale pentru detectarea aglomerărilor spațiale.

Even though its use has been limited, the DAC statistic, defined as the difference between the empirical distribution for the cases and that of the total sample, presents a great potential in detecting spatial clusters. Our previous work focused on examining its sensitivity to the location of origin and orientation of axes, and on its possible role as an instrument to detect clusters. Results indicate that the DAC statistic does not depend on the location of the origin, but is influenced by the orientation of axes. The DAC statistic cannot be used alone to detect clusters, but in conjunction with our methods. GIS-based methods were examined as a possible candidate. Results indicate that results obtained using density maps depend on choices of the users, and kriging approaches are preferable in terms of repeatability. This article summarizes our previous results, most of which had not been published before, and attempts to build up a methodology using the DAC statistic in conjunction with other spatial techniques to detect spatial clusters.

Keywords: DAC statistic; spatial statistics; clusters; empirical cumulative distribution; kriging; geocoding; density maps; low birth weight; birth certificates data.

* Expert, General Directorate of Spatial Planning, Urbanism and Housing Policy, Ministry of Transports, Constructions and Tourism, Government of Romania, Bucharest, ROMANIA; Prof., Dept. of Epidemiology and Biostatistics, School of Public Health, University of South Carolina, Columbia, SC, USA; Prof., Dept. of Ecology, Faculty of Biology, University of Bucharest, Bucharest, ROMANIA

Introduction

Space-time analyses represent important issues, due to their wide area of application. In public health, they are used to detect disease space and time clusters [1-5], to increase the efficiency of health department's activity [3], or just to study the spatial pattern or distribution of a population dispersed over a continuous surface [6]. Different studies have indicated various approaches to space-time analyses over wide and expanding venues of applications. One approach was to work on disease risk from environmental hazard at three levels: analyses of distribution, analyses of sentinel events, and case cluster strategies [2]. The analysis of distribution refers to the DAC statistic; the analysis of sentinel events recognizes that some events are more important than others when used to draw attention, and case-cluster strategies permit the identification of disease clusters. The DAC statistic is defined as the difference between the empirical distribution for the cases and that of the total sample [2]. A simulation indicated that the location of the maximum DAC statistic is not unique, moreover there is a geometrical locus of it, and this varies as the orientation of the axes changes [7]. Other studies investigated the usefulness of the DAC statistic in suggesting spatial clusters. Sampling provided discrete data and the analysis could not point directly to potential clusters. SAS[®] yielded inconclusive results [8], whereas the location of clusters depended on classification if using ArcView GIS[®] [2, 9].

Spatial prediction, referred as kriging [10, 11] may provide a way to generate smooth continuous surfaces and predict the behavior of the DAC statistic at each location within the investigated area, suggesting also location of potential clusters. Various kriging procedures were developed in SAS[®] [11, 12] and ArcGIS[®] [10] and used in conjunction with the DAC statistic to suggest potential low birthweight clusters in Spartanburg county, SC [13].

The present study attempts to summarize our previous findings, most of which had not been published, and build up a methodology using the DAC statistic in conjunction with other spatial techniques to detect spatial clusters.

1. Paper contents

2. The DAC statistic
3. Birth data 1998-1990, Spartanburg County, SC
4. Sensitivity to random locations of origin
5. Sensitivity to random orientations of axes
6. The survivorship function and its connection with the DAC statistic
7. DAC and the theory of epidemiology
8. DAC and GIS-based density maps
9. DAC and kriging

10. Discussion

- 10.1. Sensitivity of the DAC statistic to the location of origin and orientation of axes
- 10.2. Epidemiological considerations suggested by GIS-based density maps
- 10.3. Epidemiological considerations suggested by kriging
- 10.4. Software limitations

2. The DAC statistic

The DAC statistic was introduced for the first time in the statistical literature through a study by Drane, Creangă, Aldrich, and Hudson [14]. The purpose of introducing the DAC statistic was to provide an instrument to detect spatial clusters, or, more generally, areas with health problems. The computation of the DAC statistic is based on the empirical cumulative distribution function.

The empirical cumulative distribution function is:

$$F_n(x_1, x_2) = m(x_1, x_2) / n \quad (1)$$

where $m(x_1, x_2)$ is the number of points of the sample of size n such that $x_{1i} \leq x_1$ and $x_{2j} \leq x_2$. As (x_1, x_2) covers the entire sample from $(0, 0)$ to $(\max x_1, \max x_2)$, $m(x_1, x_2)$ spans the interval $[0, n]$.

The DAC statistic is, for all permissible values of (x_1, x_2) ,

$$DAC(x_1, x_2) = F_m(x_1, x_2) - F_n(x_1, x_2) \quad (2)$$

F_m is the empirical cumulative distribution function of all cases, and F_n is the empirical cumulative distribution function of the total population [14]. If within the sample of size n there are m cases and $n-m$ non-cases, F_{n-m} may be substituted for F_n .

The maximum absolute value of the DAC statistic represents the Kolmogorov-Smirnov statistic for two samples [15].

3. Birth data 1998-1990, Spartanburg County, SC

The data came from a demonstration project sponsored by the Robert Woods Johnson Foundation. The object of the effort was to demonstrate the usefulness of geographically coded health events. The one legal paper, which had a great promise of nearly a 100% response rate, was the birth certificate. It was chosen. For the period 1989-1992 nearly all of the live births in Spartanburg County SC were geocoded. The longitude and latitude of the mother's home was affixed to the birth certificate data of the baby. For this particular biostatistical methodological investigation the only data used were the longitude, latitude and the baby's birth weight.

In 1990 the population of Spartanburg County was 226,800. There were 3,762 live births in Spartanburg County in 1990, out of which 302 were low birth weights [16]. The results of a previous study [14] are presented below. Low birthweights were defined as those less than or equal to 2500 grams. Even if the two distributions presented in Figs. 1 and 2 appear similar to the naked eye, their differences, however small, are displayed in Fig. 3. The graphs displayed in the following were produced using a Turbo-Pascal[®] application.

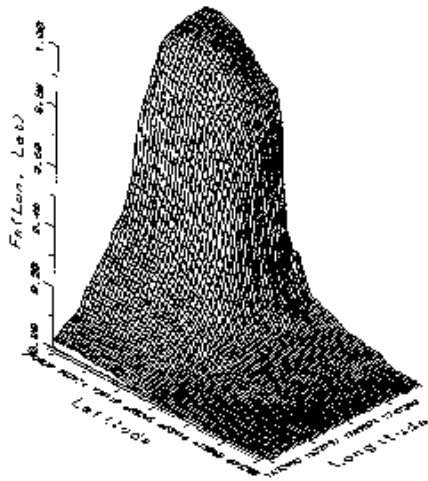


Fig. 1. Empirical distribution of live births, $N=6434$

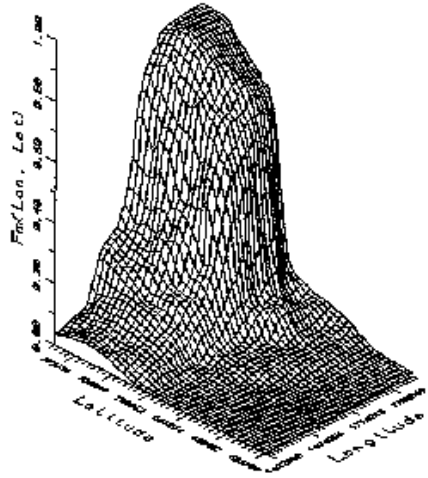


Fig. 2. Empirical distribution of low birth weights, $N=591$

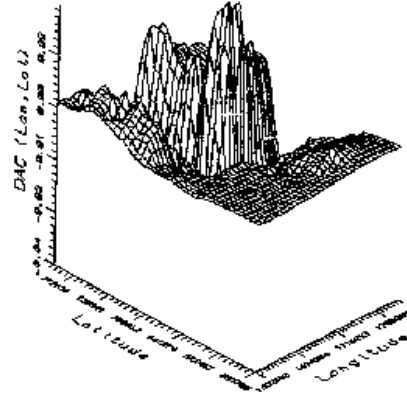


Fig. 3. Empirical distribution of the DAC statistic

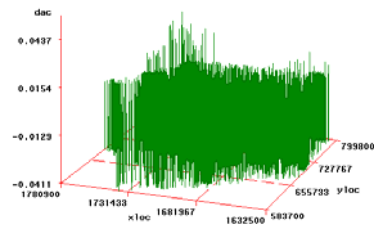


Fig. 4. Values of the DAC statistic for the Spartanburg data

Fig 4 displays the empirical distribution of the DAC statistic plotted using SAS[®]. The graph does not reveal any particular aspects that could be connected to a base map suggesting potential clusters.

The data set consisted of 6434 lines of observations, corresponding to 6434 live births. Out of these, 591 were cases. Cases were low birthweight babies. Low birthweights were defined as those less than or equal to 2500 grams. Each line contains, in order, the following variables: a line (1-6434), the actual latitude and longitude, and the infant's birth weight [9].

4. Sensitivity to random locations of origin

The translation of origin is equivalent to adding constants to the coordinates of each data point. That is,

$$T(x_1, x_2) = T(x_1 + \alpha, x_2 + \beta) \quad (3)$$

for all (x_1, x_2) , where $-\infty < \alpha, \beta < \infty$.

Therefore, $m(x_1, x_2)$ is the number of points of the sample of size n such that $x_{1i} \leq x_1$ and $x_{2j} \leq x_2$, and becomes $m_1(x_1, x_2)$, the number of points of the sample of size n such that $x_{1i} + \alpha \leq x_1 + \alpha$ and $x_{2j} + \beta \leq x_2 + \beta$, which is equivalent to $x_{1i} \leq x_1$ and $x_{2j} \leq x_2$. Therefore,

$$m_1(x_1, x_2) = m(x_1, x_2) \quad (4)$$

and

$$F_n(x_1, x_2) = m(x_1, x_2) / n = F_n(x_1 + \alpha, x_2 + \beta) \quad (5)$$

and, similarly,

$$F_m(x_1, x_2) = F_m(x_1 + \alpha, x_2 + \beta) \quad (6)$$

Therefore:

$$DAC(x_1, x_2) = F_m(x_1, x_2) - F_n(x_1, x_2) = DAC(x_1 + \alpha, x_2 + \beta) \quad (7)$$

In summary, the change of the location of origin does not affect the order relationship between any possible set of data pairs. Only the measures of location, which change with a constant amount, are affected. As the cumulative distribution function is a step function and depends only on the order relationship between any possible set of data pairs, its shape is not influence by the change of the location of origin [7].

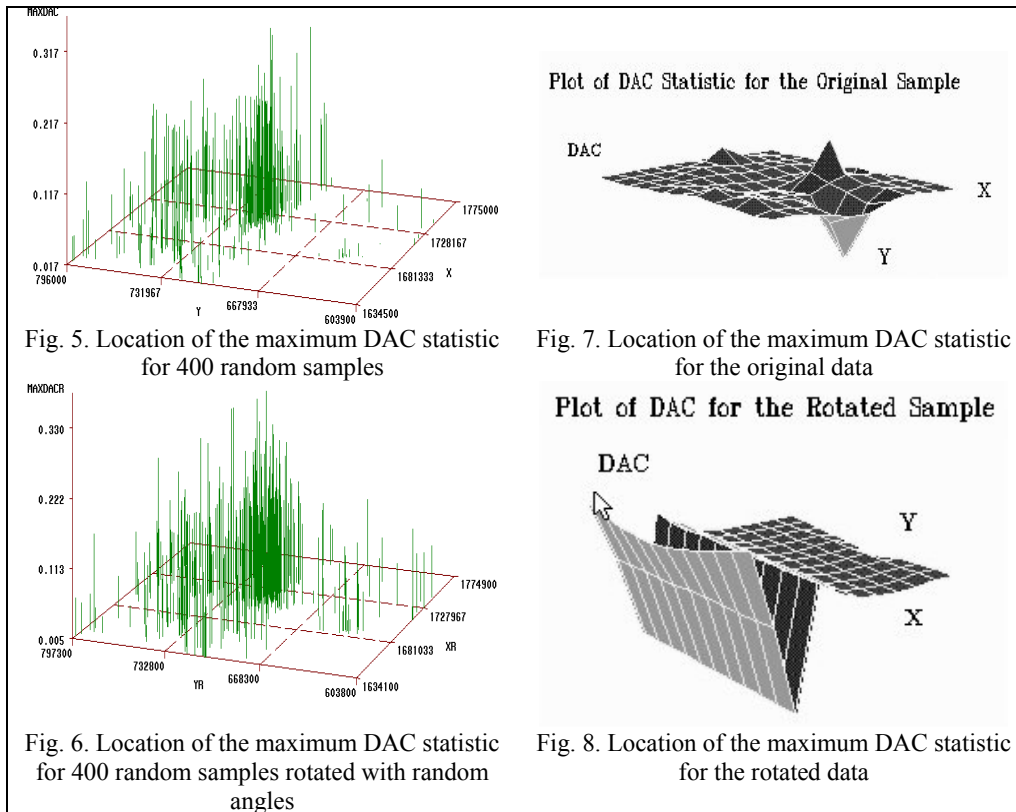
5. Sensitivity to random orientations of axes

For these simulations, a special program, called "DAC.EXE", was created in Microsoft Q-Basic[®]. In order to increase the efficiency of this program (in terms of memory usage and speed), it was converted to an executable program using Quick Basic[®]. The program reads the initial data in comma-delimited

format and produces an output file in the same format, containing as many lines as the number of samples indicates. Each line contains, in order:

- Maximum DAC statistic for respective sample (MaxDAC);
- The X value at which MaxDAC occurred;
- The Y value at which MaxDAC occurred;
- Maximum DAC statistic for rotated sample (Max DACr);
- The X value at which Max DAC occurred (in terms of original coordinates);
- The Y value at which Max DAC occurred (in terms of original coordinates) [7].

Due to the Quick Basic[®] processor, the maximum sizes allowed by the program ranged from either 20 samples of size 400 or 40 samples of size 200. This problem was overcome through a completely random device based on the computer clock. In successive steps, the program was able to draw 1,000 samples of size 400. The samples were rotated with random angles and the results are displayed below in Figs. 5 and 6.



In the next step, the DAC statistic was computed for all 6434 observations. Data were rotated arbitrarily and the DAC statistic was recomputed for the rotated

data. The results are displayed in Figs. 7 and 8 using a Turbo-Pascal[®] plotting application.

It may be noticed even with a naked eye that the maximum DAC statistic occurs at approximately the same location before and after rotating the samples arbitrarily. This may support the reliability of the maximum DAC statistic in terms of detecting spatial clusters.

6. The survivorship function and its connection with the DAC statistic

Recall that the cumulative distribution is:

$$F(x, y) = P(X \leq x \text{ and } Y \leq y) \quad (8)$$

whereas the survivorship function is:

$$S(x, y) = P(X > x \text{ and } Y > y) \quad (9)$$

Its compliment is:

$$P(X \leq x \text{ or } Y \leq y) = P(X \leq x) + P(Y \leq y) - P(X \leq x \text{ and } Y \leq y) = F(x) + F(y) - F(x, y) \quad (10)$$

$S(x, y)$ is on an intersection of $\{X > x\}$ and $\{Y > y\}$ with **strict** inequalities while $F(x, y)$ is on $\{X \leq x\}$ and $\{Y \leq y\}$ with non-strict inequalities.

Since:

$$F(x, \infty) = F(x) \text{ and } F(\infty, y) = F(y) \quad (11)$$

$$S(x, y) + [F(x) - F(x, y)] + [F(y) - F(x, y)] + F(x, y) = 1 \quad (12)$$

and

$$S(x, y) = 1 - F(x) - F(y) + F(x, y) \quad (13)$$

7. DAC and the theory of epidemiology

It is common practice in epidemiology to form the logarithm of the odds-ratio (OR) for an incremental (Δx) change in a variable thusly:

$$\log(OR(x + \Delta x | x)) = \log \left\{ \frac{P(y = 1 | x + \Delta x) / [1 - P(y = 1 | x + \Delta x)]}{P(y = 1 | x) / [1 - P(y = 1 | x)]} \right\} \quad (14)$$

because equation (13) gives rise to

$$\log(OR) = \beta_I \Delta x \quad (15)$$

if and only if $P(y = 1 | x)$ is the logistic dose response function. That is, if and only if:

$$P(y = 1 | x) = (1 + e^{-\beta_0 - \beta_I x})^{-1} \quad (16)$$

which is easily generalized to a multiplicity of x s.

This practice fails, and equation (15) does not follow for all other dose response functions. One needs only to rearrange equation (14) to obtain:

$$\log(OR) = \log[odds(x + \Delta x)] - \log[odds(x)] \quad (17)$$

where

$$odds(x) = P(x)/[1-P(x)] \quad (18)$$

Thus:

$$\lim_{\delta x \rightarrow 0} \frac{\log(OR(x + \Delta x : x))}{\delta x} = \frac{d}{dx} \log odds(x) = \frac{P'(x)}{P(x)} + \frac{P'(x)}{1-P(x)} = \frac{P'(x)}{P(x)[1-P(x)]} \quad (19)$$

This is named instantaneous log-odds ratio or ILOR [17]. Eq. (19) applies for all $P(y=I|x)$ on a continuum finite or the positive reals.

Is it $P(Y=I|X=x)$ or $P(Y=I|X \leq x)$? Every person examined and giving up data on themselves may ($Y=I$) or may not ($Y=0$) have a specific disease. In almost all cases the disease developed or occurred over a range of values of X no greater than the one measured at the physical exam. We therefore prefer $P(Y=I|X \leq x)$.

That being the case, let $F_0(x)$ be the cumulative distribution of the controls (or non-cases) on X , and accordingly $F_1(x)$ is the same for cases. Let θ , $0 < \theta < 1$, be the prevalence of the disease in the population. Then

$$F(x) = \theta F_1(x) + (1-\theta) F_0(x) \quad (20)$$

and

$$P(Y=1 | X \leq x) = \frac{\theta F_1(x)}{\theta F_1(x) + (1-\theta) F_0(x)} \quad (21)$$

The fundamental theorem of epidemiology is simply:

$$\lim_{x \rightarrow x_m} P(Y=1 | X \leq x) = \theta \quad (22)$$

The odds of a disease, given $X \leq x$, is:

$$odds(x) = \frac{\theta F_1(x)}{(1-\theta) F_0(x)} \quad (23)$$

and

$$ILOR(x) = \frac{F'_1(x)}{F_1(x)} - \frac{F'_0(x)}{F_0(x)} \quad (24)$$

If we take advantage of the fact that $S(x)$, the survivorship function, can be used because, in one dimension $S(x) = 1 - F(x)$.

$$1 - S(x) = \theta[1 - S_1(x)] + (1-\theta)[1 - S_0(x)] \quad (25)$$

$$S(x) = \theta S_1(x) + (1-\theta) S_0(x) \quad (26)$$

and

$$odds(x) = \frac{\theta S_1(x)}{(1-\theta) S_0(x)} \quad (27)$$

$$ILOR(x) = \frac{S'_1(x)}{S_1(x)} - \frac{S'_0(x)}{S_0(x)} = \lambda_0(x) - \lambda_1(x) \quad (28)$$

where $\lambda(x)$ is the hazard function found in reliability theory and life tables.

To finish this section, consider two or more variables X . If one uses the multivariable survivorship function of, then

$$ILOR(x) = \nabla \log \frac{S_1(x)}{S_0(x)} \cdot \bar{r} \quad (29)$$

where ∇ is the del-operator $\left(\frac{\partial}{\partial x_1} i + \frac{\partial}{\partial x_2} j + \dots + \frac{\partial}{\partial x_m} m \right)$, and \bar{r} is the directional unit vector $(i \cos \varphi_1 + \dots + m \cos \varphi_m)$.

The initial objective of creating the empirical spatial distributions or survivorships was to use smoothing splines in order to use equation (29) and calculate isofects, lines of constant hazard or threat to the population overlaying the region of investigation on a map.

8. DAC and GIS-based density maps

To answer the question remained whether the DAC statistic is a reliable instrument to detect spatial clusters, a new application was created to use the DAC statistic with the Spartanburg data to detect clusters of low birthweight [7, 8]. The program read the initial data in comma-delimited format from an input file, prompted for the weight limit for normal births, and produced an output file in the same format, containing as many lines as the number of observations indicated. Each line contained, in order, the location (latitude and longitude) and the value of the DAC statistic, as well as the values of the cumulative distributions for the cases and for the entire sample. The results were used to create the map displayed in Fig. 9 using ArcView GIS[®]. This figure presents a choropleth map of the positive values of the DAC statistic in Spartanburg County, SC. The shading intensity is directly proportional to the density of positive values in the area. Cities are displayed as black dots. It may be easily noticed that the peaks of the DAC statistic concentrate around the cities. Positive values occur in the northeast part of Spartanburg and around Cowpens, Chesnee, Landrum, Campobello, and Inman. The highest values can be found around Spartanburg.

Fig 10 is a three-dimensional representation of the positive values of the DAC statistic in Spartanburg County, SC, in relationship to the position of the cities. The height of each peak and the shading intensity is directly proportional to the density of positive values in the area. Main cities are displayed as black full dots. The area of the county appears as a semitransparent gray shape.

It is expectable to find more DAC values around the large cities, and even more expectable for the peaks of the DAC statistic, to occur around these places. Our results show that maximum values tend to occur mostly in the Northwestern part of the county. This may be an indication of clustering. Furthermore, the peaks detected around cities, especially the larger ones, Spartanburg and Greer, may indicate problems in these areas. Epidemiological studies conducted in these areas might explain the causes of these clusters.

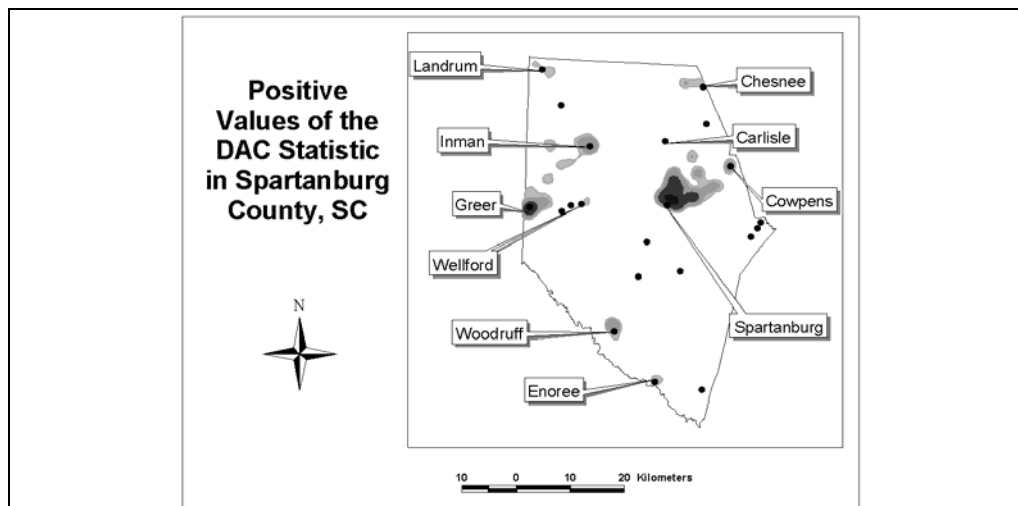


Fig. 9. Map of the positive DAC statistic values in Spartanburg County, SC in relationship with the position of the main cities.

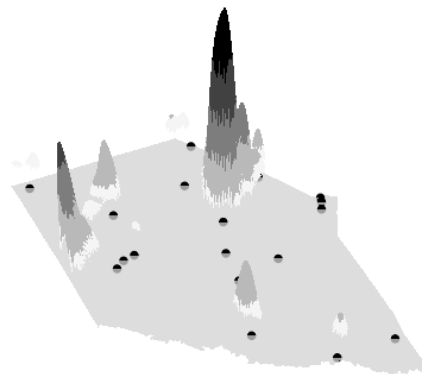


Fig. 10. Three-dimensional representation of the positive DAC statistic values in Spartanburg County, SC in relationship to the position of the main cities.

9. DAC and kriging

This study used three approaches to kriging available in ArcGIS®. Ordinary kriging uses semivariogram or covariance models relying on spatial relationships among the data, assuming intrinsic stationarity and that the true mean of the data (i.e. mean DAC value) is constant but unknown [10]. Fig. 11 displays the semivariogram corresponding to using ordinary kriging for the DAC data. It may be argued that the assumption of a constant mean does not hold in this case.

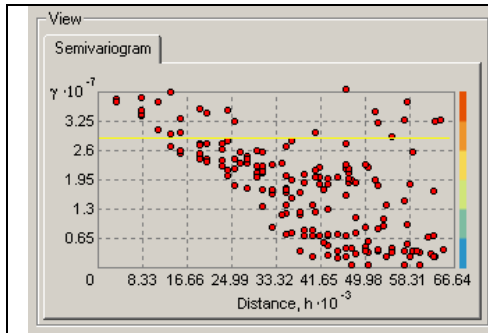


Fig. 11. Ordinary kriging: semivariogram

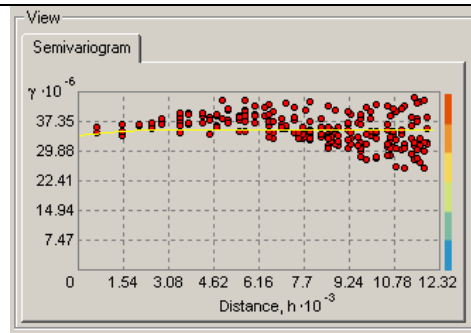


Fig. 13. Simple kriging: semivariogram

Ordinary Kriging of Spartanburg Data

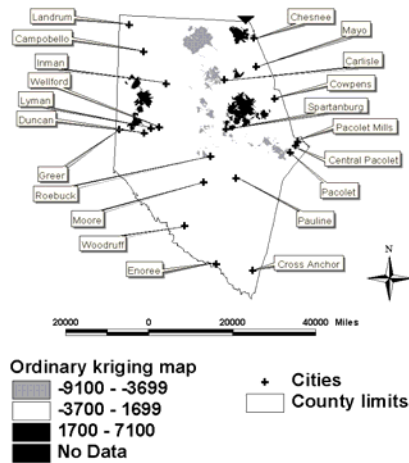


Fig. 12. Ordinary kriging: prediction map

Simple Kriging of DAC Data

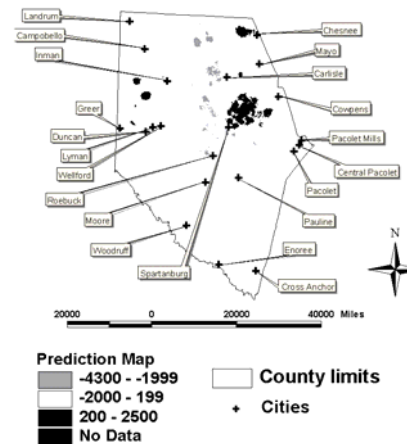


Fig. 14. Simple kriging: prediction map

Fitting this model provides the map displayed in Fig. 12. Grey shades indicate negative values of the DAC statistic. Of interest for our study are black shades suggesting clusters of positive values indicating low birthweights.

Simple kriging, as defined in ArcGIS[®], uses semivariogram or covariance models relying on spatial relationships among the data, assuming intrinsic stationarity and that the true mean of the data (i.e. mean DAC value) is constant and known [10]. Fig 13 displays the semivariogram corresponding to using simple kriging for the DAC data, which apparently fits the data satisfactorily.

Fitting this model provides the map displayed in Fig 14. Grey shades indicate negative values of the DAC statistic. Of interest for our study are black shades suggesting clusters of positive values indicating low birthweights.

Universal kriging, also available in ArcGIS[®], uses semivariogram or covariance models relying on spatial relationships among the data, assuming that the true mean of the data (i.e. mean DAC value) is some deterministic function [10]. Fig. 15 displays the semivariogram corresponding to using universal kriging for the DAC data. It may be argued that the semivariogram does not fit the model.

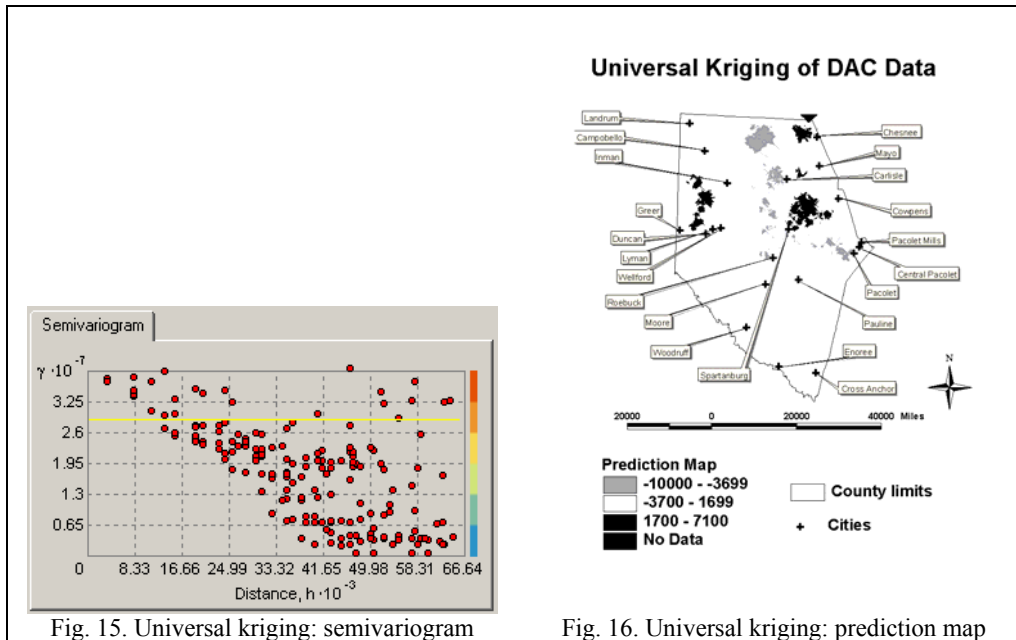


Fig. 15. Universal kriging: semivariogram

Universal Kriging of DAC Data

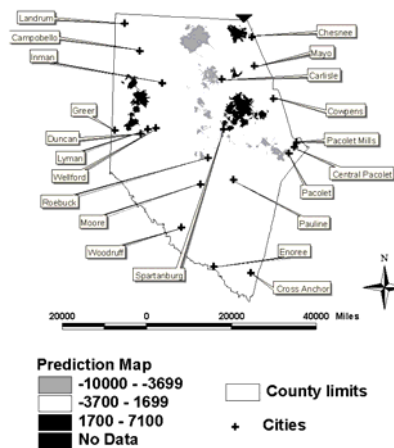


Fig. 16. Universal kriging: prediction map

Fitting this model provides the map displayed in Fig 16. Grey shades indicate negative values of the DAC statistic. Of interest for our study are black shades suggesting clusters of positive values indicating low birthweights.

There were two attempts to kriging with SAS[®]. The first one was designed to perform an ordinary kriging. The first problem was to generate a prediction grid. Given the size of the original data set (6434 observations), generating a fine resolution prediction grid resulted into exceeding the allocated memory for a proper running of the program. Therefore, no results were obtained before limiting the prediction grid to a low-resolution one.

At this stage, SAS[®] provided a unique estimate with the same standard error for all the locations within the predicted grid. It could be argued that this value represents an estimate of the true mean of the data [10, 11].

The next step involved an attempt to universal kriging. Again, the allocated memory for a proper running of the program was exceeded and no results were obtained.

In summary, all the attempts to kriging with SAS[®] provided inconclusive results for the Spartanburg data.

10. Discussion

10.1. Sensitivity of the DAC statistic to the location of origin and orientation of axes

The results indicated that the DAC statistic does not depend on the location of the origin. However, the dependence on the orientation of axes has an analytical expression that may not be easily detected. In this example, the maximum DAC statistic appears to be a reliable instrument in detecting spatial clusters independently of the orientation of axes.

In real life example, the maximum DAC statistic does not have necessarily an analytical expression, therefore it is almost impossible to find its geometrical locus. The question remains whether it will still remain a reliable instrument in detecting spatial or temporal clusters.

10.2. Epidemiological considerations suggested by GIS-based density maps

It is expectable to find more DAC values around the large cities, and even more expectable for the peaks of the DAC statistic, to occur around these places. Our results show that maximum values tend to occur mostly in the Northwestern part of the county. This may be an indication of clustering. Furthermore, the peaks detected around cities, especially the larger ones, Spartanburg and Greer, may indicate problems in these areas. Epidemiological studies conducted in these areas might explain the causes of these clusters.

10.3. Epidemiological considerations suggested by kriging

Despite of the kriging method used in ArcGIS[®], all the results indicated that the predicted clusters of low birthweights occurred close to Spartanburg, Chesnee, and possibly Inman and Greer. In this example, the DAC statistic appears to be a useful instrument in suggesting spatial clusters if used in conjunction with spatial prediction methods. Nevertheless, the DAC statistic should be used with caution, but its usefulness as a set of spatial descriptive statistic is not diminished in the least.

10.4. Software limitations

After unsuccessful attempts to compute the DAC statistic using SAS[®], a Quick Basic[®] application had been developed. The application was used mainly for simulation purposes. Its usage was limited by the size of the Quick Basic[®] processor. This limited sampling to a formula where the product between the number of samples to be selected and the sample size could not exceed 8000, therefore the program is not suitable for large data sets. An Excel spreadsheet was developed lately to compute the DAC statistic. The number of observations limits its usage, and the results for the entire data set differ from those obtained using the Quick Basic[®] application for some locations, even though results obtained using smaller data sets for testing purposes are the same.

In this study, even if easier to control from a statistical viewpoint, kriging with SAS[®] was limited by the size of the data set resulting into exceeding the allocated memory for a proper running of the program. Therefore, given these limitations, the results provided by SAS[®] analyses are inconclusive. At the same time, limitations to kriging with ArcGIS[®] refer to the ability to control the modeling process statistically.

Conclusions

Results indicated that the DAC statistic is inflexible to the location of origin, but depends on the orientation of axes. In this regard, it should be used with caution when the coordinate system is subject to changes. Despite of this property, the DAC statistic could play a substantial role in detecting spatial clusters, especially when it is used in conjunction with other spatial prediction techniques, such as kriging. Generalizing, GIS techniques are also a good instrument in conjunction with the DAC statistic, but choice of parameters (when generating density maps or selecting the kriging approach) influence the results. Finally, it is important to emphasize that the DAC statistic is able to detect

clusters, but explanations on the occurrence of clusters are expected from other fields, such as epidemiology.

REFERENCES

1. *T. E. Aldrich and J.W. Drane*, "Cluster 3.1: Software System for Epidemiologic Analysis", U.S.D.H.H.S., A.T.S.D.R., Atlanta, GA, 1993
2. *T. E. Aldrich*, K. Krauthelm, E. Kinee, J. W. Drane and D. Tibara D., "Statistical Methods for Space-Time Cluster Analysis", Proceedings of the International Symposium on Computer Mapping in Epidemiology and Environmental Health, 1997, pp. 226-236
3. *T. Britton*, "Tests to Detect Clustering of Infected Individuals within Families", Research Report. Institute of Actuarial Mathematics and Mathematical Statistics, Stockholm University, 1995, pp. 1-18
4. *C. R. Stark and N. Mantel N.*, "Temporal-Spatial Distribution of Birth Dates for Michigan Children with Leukemia", in *Cancer Res.*, **vol. 27**, 1967, pp. 1749-1775
5. *E. H. Williams, P. G. Smith, N. E. Day, A. Geser, A. Ellice and P. Tukei*, "Space-Time Clustering of Burkitt's Lymphoma in the West Nile District of Uganda: 1961-1975", in *Br. J. Cancer*, **vol. 37**, 1978, pp. 109-122
6. *J. E. Paloheimo and A. M. Vokov*, "On measures of Aggregation and Indices of Contagion", in *Math. Biosci.*, **vol. 30**, 1976, pp. 69-97
7. *A. I. Petrișor, J. W. Drane, K. Jackson and L. Dragomirescu*, "Spatial Statistics: The DAC Statistics", in Proceedings of the Section on Statistics and the Environment, American Statistical Association, 2002, pp. 2683-2686
8. *A. I. Petrișor, J. W. Drane, L. Dragomirescu and K. Jackson*, "A Computer-Based Application to Detect Spatial Clusters Using the DAC Statistic", in Healthcare Telematics Support in Transition Countries. Proceedings of MIE 2001 Special Topic Conference, Bucharest, June 7-9, 2001, Ed.: Mihălaș GI, Engelbrecht R., Băzăvan M., Hofdijk J, Naszlady A., Publishing House EUROBIT Timișoara, 2002, pp. 138-141
9. *A. I. Petrișor, L. Dragomirescu, J. W. Drane, K. L. Jackson and D. Cowen*, "Using a Quick-Basic Application Based on the DAC Statistic to Detect Spatial Clusters", in *Control Engineering and Applied Informatics*, **vol. 4**, nr. 1, 2002, pp. 55-64
10. *K. Johnston, J. M. Ver Hoef, K. Krivoruchko and N. Lucas*, "Using ArcGIS Geostatistical Analyst", ESRI, Redlands, CA, 2001
11. *W. W. Piegorsch and D. Edwards*, "Notes on Temporal and Spatial Analysis for a Course in Environmetrics", Department of Statistics. University of South Carolina. Columbia, SC, 2001
12. *R. P. Cody and J. K. Smith*, "Applied Statistics and the SAS® Programming Language", Prentice Hall, Upper Saddle River, New Jersey, 1997
13. *A. I. Petrișor, J. W. Drane and L. Dragomirescu*, "Using Kriging and the DAC Statistic to Predict Low Birthweight Clusters in Spartanburg County, SC", in Proceedings of the Section on Statistics and the Environment, American Statistical Association, 2002, pp. 2687-2691
14. *J. W. Drane, D. L. Creangă, T. E. Aldrich and M. B. Hudson*, "Detecting Adverse Health Events via Empirical Spatial Distributions (Abstract)", Symposium on Statistical Methods 1995, U.S.D.H.H.S., P.H.S., C.D.C., Atlanta, GA, January 24-26, 1995
15. *M. Hollander, D. A. Wolfe*, "Nonparametric Statistics Methods", John Wiley & Sons, New York, 1973

16. *** The Division of Biostatistics, Office of Vital Records and Public Health Statistic, South Carolina Department and Environmental Control. South Carolina Vital and Morbidity Statistics 1990. Volume I: Annual Vital Statistics Series, 1993
17. *J. W. Drane*, "Instantaneous log odds ratio", International Biometric Society, ENAR, Richmond, 1996