SCIREA Journal of Agriculture

https://www.scirea.org/journal/Agriculture

September 18, 2019

Volume 4, Issue 4, August 2019

# Data Features of the Weighted Standard Deviational Curve

**Roger L Goodwin**

Summit Point, USA

Email: rogergoodwin AT frontiernet DOT net

## Abstract

This article presents the weighted data features of the standard deviational curve (SDC). Similar data features exist for the standard deviational ellipse (SDE). This paper presents weighted data features for the SDC which include the angle of rotation, the minimum and maximum standard deviation, the area, and the circulatory index. Performed correctly, weighted features give a better use of the areal data points. The additional computations of weighted data features are no more difficult than those of the unweighted data features.

**Keywords:** Spatial Analysis, data features, curve, angle of rotation, area, standard deviations, circulatory index.

## 1.0 Introduction

In 1926, Lefever presented a paper on the standard deviational ellipse. One year later, Furfey criticized that article because it did not have an elliptical pattern. Yuill pointed-out that Lefever's SDE provides orientation, as well as the average location and the dispersion to areal point data.

After a forty-five-year black-out period, Yuill treated parts of the values of the curve as a an "ellipse" that becomes a "line" when the minor axis is zero. When the point set is evenly dispersed about the average location, the ellipse is a circle. The minimum and maximum standard deviations determine the "shape" of the ellipse. It becomes a boundary value problem. Yuill also introduced weighted statistics into the formulas. He argues, for example, that the mean weighted center of hay production is different than the distribution of farms. Hay production, the random variable, can increase and decrease from farm-to-farm.

As stated, after forty-five years from Lefever's publication, Yuill corrected Lefever's ellipse using Furfey's suggestions and extending the results with weighted statistics.  Neither Gong nor any successors extended the standard deviational curve with weighed statistics. This paper presents a weighted standard deviational curve in [2].

References [2] and [4] present research developed on the Cartesian coordinate system. To have any practical use, the concepts must be applied to the geospatial coordinates latitude and longitude. This simple realization leads us to include a random variable of interest in the spatial estimators.  This leads us to the weighted estimators [7]. Yuill presents weighted estimators for the following data features:

1) Angle of rotation
2) Minimum and maximum of the variance
3) Area
4) Eccentricity

After seventy-six years from Lefever's publication, Gong proposed a set of statistics on Lefever's paper. The statistics are un-weighted. Gong follows the same thoughts. The "curve" needs boundary conditions. Gong uses the minimum and maximum standard deviations to measure the *circulatory index.*

The characteristics sought from arear point data did not change since Lefever's original paper: 1) the center of the system, 2) the direction or trend of the system, 3) the concentration or scatteration of the system, and 4) the relative concentration of the data within the ellipse.

This paper presents the weighted standard deviational curve (WSDC). For similar surveys on the same observation units, we can re-use the latitude and longitude $(x, y)$ observations. For the random variable (also called the weight variable) $w$, $w$ changes from survey to survey.

## 2.0 Estimators

This section lists the estimators for the WSDC. The estimators are not difficult to calculate for a given data set $(x, y, w)$ where $x$ represents the latitude, $y$ represents the longitude, and $w$ represents the weight (or the random variable).

Equation (1) gives the weighted mean center for $(x_i, y_i, w_i)$ for $i=1,\ldots,n$ observations where $x_i$ represents the $i^{th}$ latitude observation, $y_i$ represents the $i^{th}$ longitude observation, and $w_i$ represents the $i^{th}$ observation of the random variable [7]. We can find the same estimators in most advanced textbooks on linear regression [1], [3].

$$(\bar{x},\bar{y}) = \left(\frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i} \quad \frac{\sum_{i=1}^{n} y_i w_i}{\sum_{i=1}^{n} w_i}\right) \tag{1}$$

Wood's numeric example shows the subtle difference between un-weighted and weighted statistics. Wood only provided the latitude $\theta^o$ and the longitude $\psi^o$. The study encodes the randomness in the paired observations $(\theta_i^{\mathrm{o}},\psi_i^{\mathrm{o}})$ --- our $(x_i, y_i)$ pairs without weights. As Yuill pointed-out, had this been a farm survey in a chosen geographic region, the $(\theta_i^{\mathrm{o}},\psi_i^{\mathrm{o}})$ observations would have been constant regardless if he measured housing or hay.

Equation (2) gives the estimators for weighted variances for the observations $(x_i, y_i, w_i)$.

$$\left(\hat{\sigma}_{x,0}^2, \hat{\sigma}_{y,0}^2\right) = \left(\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 w_i}{\sum_{i=1}^{n} w_i} \quad \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2 w_i}{\sum_{i=1}^{n} w_i}\right) \tag{2}$$

Equation (3) gives the estimator for the weighted correlation coefficient for the observations ***(x<sub>i</sub>*** —

$$r_0 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})\, w_i}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 w_i}\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 w_i}} \tag{3}$$

From (2), we define the quantity ***a*** as

$$a = \frac{1}{2}\left(\hat{\sigma}_{x,0}^2 - \hat{\sigma}_{y,0}^2\right) \tag{4}$$

From (2) and (3), we define the quantity ***b*** as

$$b = r_0 \hat{\sigma}_{x,0} \hat{\sigma}_{y,0} \tag{5}$$

Previous literature gave interpretations for the quantities ***a*** and ***b*** which included ***a*** being the major axis and ***b*** being the minor axis. Likewise, we interpreted ***a*** to be the semi-major axis and ***b*** the semi-minor axis. In Gong's paper, he does not give any interpretation of ***a*** and ***b*** for the *deviational curve.* These are convenient quantities for calculation purposes, though.

For such simplistic estimators, the reader can implement these in Excel, VBA for Excel, or any iterative programming language with little difficulty. We use the estimators and quantities listed in this section in Section 3.

## 3.0 Data Features

This section lists the data features of the weighted standard deviational curve (WSDC). These data features are weighted extensions of the formulas in [2].

### 3.1 Weighted Angle of Rotation

Equation (6) defines the WSDC. We use he Arctan() function to find the value **θ.** Arctan() returns **θ** in units of radians.

$$\text{Tan}\,(\theta) = \frac{\sqrt{a^2 + b^2} - a}{b} \qquad (6)$$

Equation (7) gives the solution for **θ** which is conditional on ***a*** and ***b***.

$$\theta = \begin{cases} Arctan\left(\frac{\sqrt{a^2 + b^2} - a}{b}\right) & |a| + |b| \neq 0 \\ \text{No solution} & |a| + |b| = 0 \end{cases} \qquad (7)$$

### 3.2 Weighted Min and Max Variances

Equations (8) and (9) give the weighted minimum and maximum variances. We will use these to define the area of the curve and the circulatory index.

$$\widehat{\sigma}_{min} =- \sqrt{a^2 + b^2} + \frac{1}{2}\left(\widehat{\sigma}_{x,0}^2 + \widehat{\sigma}_{y,0}^2\right) \qquad (8)$$

$$\widehat{\sigma}_{max} = \sqrt{a^2 + b^2} + \frac{1}{2}\left(\widehat{\sigma}_{x,0}^2 + \widehat{\sigma}_{y,0}^2\right) \qquad (9)$$

### 3.3 Weighted Area

Equation (10) gives the weighted area of the WSDC.

$$\pi\left(\widehat{\sigma}_{min}^2 + \widehat{\sigma}_{max}^2\right) \qquad (10)$$

where $\left(\widehat{\sigma}_{min}^2, \widehat{\sigma}_{max}^2\right)$ are the weighted minimum and maximum variances from (8) and (9).

### 3.4 Weighted Circulatory Index

Reference [5] gives many formulas for the eccentricity of an ellipse. Both Gong and Furley argue that the index of the SDC does not describe the shape of the areal data points as an eccentricity measure would. For a comparable measure for a curve, Gong calls his measure the *circulatory index.* The equation presented here is not simply another addition to the eccentricity equations. Equation (11) gives the circulatory index of the WSDC.

$$0 \leqq \frac{\widehat{\sigma}_{min}^2}{\widehat{\sigma}_{max}^2} \leqq 1 \tag{11}$$

Values close to 1 indicate a circular distribution --- evenly spread-out data. Values close to 0 indicate a linear distribution (i.e. $\widehat{\sigma}_{min}^2$ << $\widehat{\sigma}_{max}^2$).

### 3.5 **Built-In Checks**

Like the authors before him, [2] provides checks on his calculations. Those checks extend to weighted formulas, as well.

1) $-\,\mathrm{aSin}\,(2\theta)\ +\ \mathrm{bCos}\,(2\theta)\ = 0.$
2) $\mathrm{Sin}\,(2\theta) = \frac{b}{\sqrt{a^2+b^2}}$ (in radians).
3) $\mathrm{Cos}\,(2\theta) = \frac{a}{\sqrt{a^2+b^2}}$ (in radians).

Obviously, the data checks differ than his predecessors'. If the underlying assumption is that the shape is not an ellipse, and a curve, then the we cannot verify the area with ***F= πab*** where ***a*** is the semi-major axis and ***b*** is the semi-minor axis.

## 4.0 Comparison of the WSDE and WSDC

*Table 1: This table compares the weighted standard deviational ellipse (WSDE) to the weighted standard deviational curve (WSDC).*

| | **WDSE** | **WSDC** |
|---|---|---|
| Location | Same. See (1). | Same. See (1). |
| Rotation | $\frac{-b\pm\sqrt{b^2-4ac}}{2a}$. See below for explanation of ***a***, ***b***, and ***c***. | $\frac{\sqrt{a^2+b^2}-a}{b}$. See (4) and (5). |
| Area | ***πab*** where ***a*** is the semi-major axis and ***b*** is the semi-minor axis. | $\pi\left(\widehat{\sigma}_{min}^2 + \widehat{\sigma}_{max}^2\right)$. See (8) and (9). |
| Data Shape | Eccentricity. Found in most calculus textbooks. | Circulatory index. See (11). |

Table I shows a comparison between the two weighted proposals of Yuill and Gong. Yuill's formula for the rotation (***Tanθ***) are the roots to the quadratic equation ***a (Tanθ)²* + *b Tanθ* + *c* = *0*** where

$$Tan\,\theta = \frac{-\,b \pm \sqrt{b^2 - 4ac}}{2a}$$

and

$$a = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})\, w_i\,,$$

$$b = \sum_{i=1}^{n} (x_i - \bar{x})^2 w_i - \sum_{i=1}^{n} (y_i - \bar{y})^2 w_i ,$$

$$c = 1.$$

## 5.0 An Application: June 2003 Area Survey

Table 2 gives the planted corn acreage for Kentucky from the 2003 June Area Survey. We supplement the county names with the latitude (***$x_i$***) and longitude (***$y_i$***) coordinates for computational purposes. The final measurements of planted corn acreage (***$w_i$***) from the survey for each county ***i***, ***1=1, 2,...,90*** counties did not change. The computations do not require county names. Hence, we drop county names from the data set.

*Table 2: Supplemented Data from the 2003 June Area Survey*

| Lat. $x_i$ | Long. $y_i$ | Plant. $w_i$ | Lat. $x_i$ | Long. $y_i$ | Plant. $w_i$ | Lat. $x_i$ | Long. $y_i$ | Plant. $w_i$ |
|---|---|---|---|---|---|---|---|---|
| 38.95249 | -84.6811 | 77300 | 38.356 | -85.4788 | 4100 | 37.02473 | -88.0901 | 6000 |
| 36.84106 | -87.4664 | 75200 | 37.75161 | -85.1479 | 3900 | 38.61941 | -83.8897 | 5600 |
| 37.84211 | -87.5832 | 63900 | 38.21706 | -84.2279 | 3800 | 36.75719 | -84.8568 | 5500 |
| 36.68877 | -88.7109 | 60400 | 38.42684 | -85.1479 | 3500 | 36.74849 | -85.7256 | 5000 |
| 36.86983 | -86.8622 | 56900 | 38.4333 | -84.3542 | 3300 | 37.00324 | -85.6435 | 4900 |
| 37.7308 | -87.1024 | 56000 | 38.01232 | -85.3136 | 3200 | 37.25701 | -85.5612 | 4800 |
| 36.83386 | -87.1423 | 44700 | 37.61093 | -82.7284 | 3000 | 38.50546 | -82.6988 | 4800 |
| 36.56766 | -89.1866 | 38800 | 38.10682 | -83.7199 | 2600 | 37.31013 | -85.8486 | 4400 |

| | | |
|---|---|---|
| 36.64028 | -88.285 | 38300 |
| 36.7772 | -86.6208 | 36100 |
| 37.56444 | -87.2618 | 34800 |
| 37.48922 | -87.737 | 30800 |
| 36.9886 | -86.4997 | 27900 |
| 37.31531 | -87.5791 | 27100 |
| 36.51281 | -88.8818 | 26300 |
| 37.03599 | -89.0179 | 26100 |
| 36.76887 | -88.3029 | 24500 |
| 37.150 | - | 24400 |
| 37.19417 | -86.2159 | 2600 |
| 37.82586 | -84.8985 | 2600 |
| 38.06065 | -84.4803 | 2500 |
| 37.7143 | -84.3121 | 2500 |
| 37.6526 | -84.8151 | 2300 |
| 38.99406 | -84.7316 | 2200 |
| 37.98421 | -85.6846 | 2200 |
| 38.31724 | -84.5641 | 2200 |
| 38.07217 | -84.7316 | 1700 |
| 36.980 | - | 1600 |
| 38.39876 | -83.6774 | 4300 |
| 37.31575 | -84.8985 | 4100 |

|  |  |  |  |  |  |
|---|---|---|---|---|---|
| 1 | 87.8942 |  | 88 | 82.9853 |  |
| 37.51085 | -86.822 | 24300 | 37.64132 | -84.5641 | 1600 |
| 38.31218 | -84.0276 | 23100 | 38.57256 | -82.8314 | 1600 |
| 36.84736 | -87.7763 | 21300 | 38.50797 | -83.3789 | 1600 |
| 36.96768 | -85.8486 | 15700 | 38.03146 | -83.8897 | 1600 |
| 37.51896 | -85.7256 | 15200 | 38.7117 | -84.059 | 1500 |
| 38.17781 | -85.2308 | 14300 | 37.95995 | -84.1435 | 1500 |
| 38.78842 | -84.3688 | 12900 | 37.06935 | -84.1857 | 1400 |
| 37.03306 | -88.7109 | 12500 | 38.46247 | -85.3042 | 1400 |
| 37.32912 | -87.051 | 12500 | 37.37431 | -84.312 | 1300 |

| | | | | | |
|---|---|---|---|---|---|
| | 4 | | | 1 | |
| 37.17725 | -87.1423 | 12100 | 36.66528 | -88.9936 | 1200 |
| 37.7982 | -86.4592 | 11700 | 36.73108 | -86.5783 | 1200 |
| 37.33299 | -88.0813 | 11100 | 38.37676 | -84.059 | 1200 |
| 37.96015 | -86.2159 | 10700 | 38.19381 | -85.6435 | 1100 |
| 36.85738 | -88.4016 | 9300 | 38.17707 | -83.4644 | 1100 |
| 38.7801 | -84.6062 | 9000 | 38.60133 | -85.3136 | 1100 |
| 37.08535 | -84.5222 | 8800 | 37.89557 | -84.5641 | 900 |
| 37.47513 | -84.6479 | 8600 | 38.67484 | -85.0649 | 800 |
| 37.33288 | -85.3136 | 8200 | 36.92606 | -83.8897 | 800 |

| | | | | | |
|---|---|---|---|---|---|
| 37.0764 | -85.3136 | 7800 | 36.72281 | -84.4711 | 800 |
| 38.33204 | -82.9455 | 7200 | 38.72951 | -84.8776 | 700 |
| 37.2986 | -84.2146 | 7000 | 38.89519 | -84.3963 | 600 |
| 37.82809 | -86.7617 | 6500 | 37.75509 | -83.4644 | 600 |

### 5.1 Applying the WSDE

Table 3 gives the data features for Yuill's weighted standard deviational ellipse using the June Area Survey data from 2003. Fig. 1 shows the weighted standard deviational ellipse.

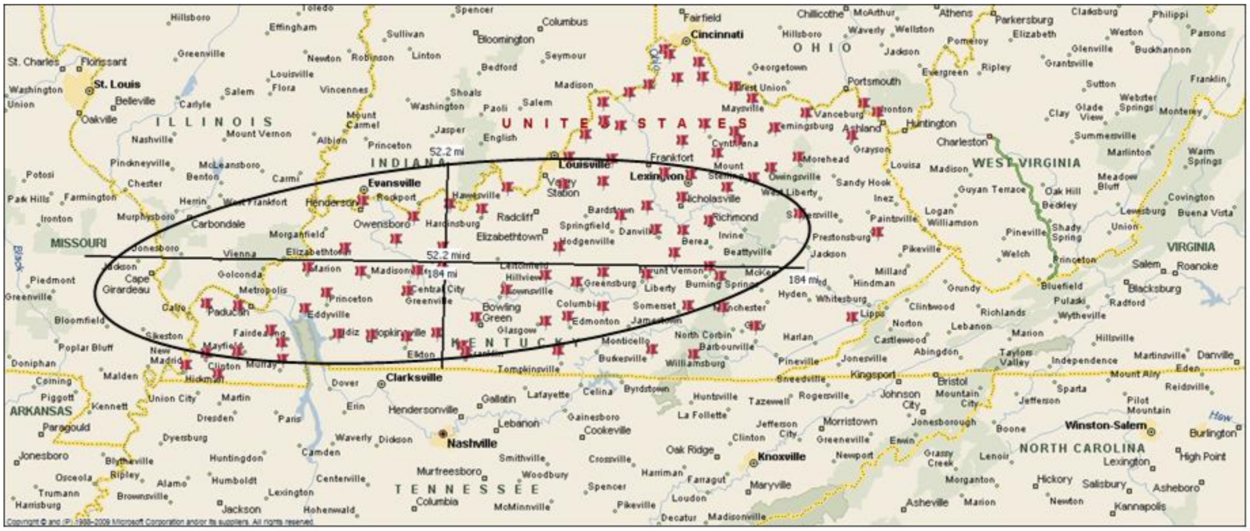

*Figure 1: This figure shows the Kentucky data from 2003 and the standard deviational ellipse. It shows a sketch of the standard deviational ellipse with Ohio County at the center. The semi-major axis is approximately 184 miles long and the semi-minor axis is ap approximately 52 miles long. We rotated the major axis roughly 70$^{o}$ from the Y-axis in an imaginary Cartesian coordinate system.*

*Table 3: WSDE 2003 Results on Corn acreage in Kentucky*

| Description | Measurement |
|---|---|
| Center | (37.38297399, -86.8129514) |
| Axes | $\boldsymbol{a}$ = 184.8734659, $\boldsymbol{b}$ = 52.23219209 |
| Area | 30,336.30686 sq. mi. |
| Standard Deviations | $\boldsymbol{\delta_x}$ = 1.29151776, $\boldsymbol{\delta_y}$ = 1.625551351 |
| Rotation | Y-Axis |
| Orientation | $\boldsymbol{\theta_x}$ =70.36257022$^{o}$, $\boldsymbol{\theta_y}$ = -19.63742978$^{o}$ |
| Eccentricity | 0.959258636 (linear) |

We can verify some of the results in this section with the alternative formula for the area $\boldsymbol{F = \pi ab}$ where $\boldsymbol{a}$ equals the semi-major axis and $\boldsymbol{b}$ equals the semi-minor axis. The alternative formula gives ***F = 30,336.30686*** which matches the area in Table 3. We used a more complex formula to arrive at the area in Table 3.

### 5.2 Applying the WSDC

Table 4 gives the data features for Gong's weighted standard deviational curve using the June Area Survey data from 2003. To draw a shape with $2\widehat{\sigma}_{max}$ as the major axis and $2\widehat{\sigma}_{min}$ as the minor axis [p. 161, (2)] is impracticable. The length of the shape is too long and the width of the shape is too narrow. Even using unweighted statistics for the curve, it does not alleviate the extreme values of $2\widehat{\sigma}_{max}$ and $2\widehat{\sigma}_{min}$. This is due to the use of the definitions of $2\widehat{\sigma}_{max}$ and $2\widehat{\sigma}_{min}$. Reference [p. 165, (2)] redefines these values later in his paper for finding the area and the circulatory index. Creating new notation is not a strong point in Gong's paper. Section 5.3 gives direction on drawing the WSDC without the need to redefine $2\widehat{\sigma}_{max}$ and $2\widehat{\sigma}_{min}$.

*Table 4: WSDC 2003 Results on Corn acreage in Kentucky*

| Description | Measurement |
|---|---|
| Center | (37.38297399, -86.8129514) |
| Variances | $\widehat{\sigma}^2_{x,0}$= 1,397.755, $\widehat{\sigma}^2_{y,0}$ = 7,537.707 |
| Correlation Coeff. | $\boldsymbol{r_0}$ = -0.9994 |
| Quantities | ***a*** = -3,069.976, ***b*** = -3,244.057 |
| Min/Max Variances | min= 1.339, max = 8,934.123 |
| Area | 28,071.58 sq. mi. |
| Circulatory Index | 0.012242073 (linear) |

To verify the calculations in Table 4 are correct, we use the formulas listed in Section 3.5. The value of $\boldsymbol{\theta}$ is $\boldsymbol{\theta}$ = -1.164315553 (must use radians). Using ***-a Sin 2θ + b Cos 2θ***, we obtain 0.00000E+00 --- essentially zero in Excel. The other two identities, $\mathbf{Sin\ (2\theta)} = \frac{\boldsymbol{b}}{\sqrt{\boldsymbol{a}^2+\boldsymbol{b}^2}}$ and $\mathbf{Cos\ (2\theta)} = \frac{\boldsymbol{a}}{\sqrt{\boldsymbol{a}^2+\boldsymbol{b}^2}}$, give exact results for the left-hand side and the right-hand side.

Applying the built-in checks from Section 3.5, we demonstrated that:

1) The estimators are correct.
2) The data features are correct.
3) The calculations are correct.

Some additional observations from a data comparison perspective include:

1) The WSDC area is roughly the same as the WSDE area.
2) The WSDE eccentricity and the WSDC circulatory index *agree* regarding the shape of the data.

### 5.3 Drawing the WSDC

As alluded to in Section 5.2, drawing a credible SDC using Gong's instructions is an issue. Gong allocates the first part of his paper, pages 159 --- 162, to drawing the SDC. Gong allocates the second part of his paper, pages 163 --- 166, to defining the SDC data feature values. Reference [2] gives two figures to show how to draw the SDC:

1) [Fig. 3, p. 161, (2)] applies to a circle as a starting point.
2) [Fig. 4, p. 162, (2)] applies to an ellipse as a starting point.

Since the problem at hand is to provide a reasonable drawing of the WDSC, we will discuss the ellipse as a starting point. We begin by drawing two circles using the two ends of the major axis of the ellipse. We draw each circle such that the circumference touches each end of the major axis of the ellipse. The opposite ends of the circles' major axis reach as close as possible to the mean center. The intersections of the ellipse's circumference and the circles' circumference serve as the connecting points between the two circles. Fig. 2 shows a rough sketch of the weighted standard deviational curve. We laid it on top of the weighted standard deviational ellipse for the augmented Kentucky 2003 survey data.

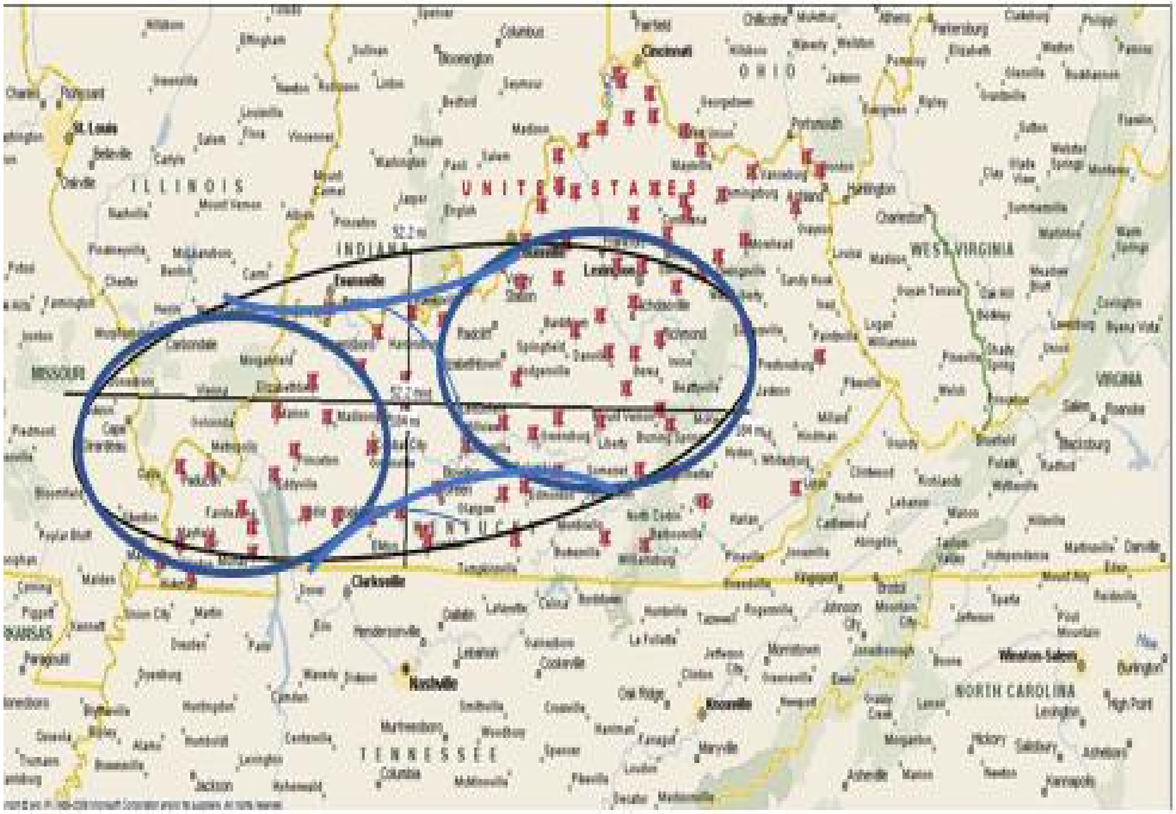

*Figure 2: This figure shows the Kentucky data from 2003. The standard deviational curve is laid on top of the standard deviational ellipse. Ohio County remains at the center.*

## 6.0 Concluding Remarks

Given the weighted estimators, it is a straightforward effort to find the weighted data features of the SDC. In the estimation section, the forms changed significantly. In the data features section, the forms did not change too dramatically. Since the data has the same shape under both the WSDE and WSDC, the eccentricity and the circulatory index seem to be a good comparison. Section 4 shows that Yuill and Gong decided to rotate the ellipse and the curve differently. Thus, any comparison between the two (or extension of the concepts) of $\boldsymbol{\theta}$ will not agree.

## References

[1] B. L. Bowerman and R. T. O'Connell, *Linear Statistical Models, An Applied Approach*, Duxbury Press, Wadsworth Publishing Company, Belmont, CA, 1990.

[2] J. Gong, "Clarifying the Standard Deviational Ellipse," *Geographical Analysis*, Vol. 34, No. 2, April 2002, pp. 155-167.

[3] E. T. Lee, *Statistical Methods for Survival Data Analysis, Second Edition*, John Wiley & Sons, Inc., New York, 1992.

[4] W. Lefever, "Measuring Geographic Concentration by Means of the Standard Deviational Ellipse," *The American Journal of Sociology*, Vol. 32, No. 1, Jul 1926, pp. 88-94.

[5] G. B. Thomas and R. L. Finney, *Calculus and Analytic Geometry, Fifth Edition*, Addison Wesley Publishing Company, Reading Massachusetts, 1981.

[6] A. Wood, "A Bimodal Distribution on the Sphere, *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, Vol. 31, No. 1, 1982, pp. 52-58.

[7] R. S. Yuill, "The Standard Deviational Ellipse: An Updated Tool for Spatial Description," *Geografiska Annaler, Series B. Human Geography*, Vol. 53, No. 1, 1971, pp. 28-39.