Spatial Cluster Analysis and Cancer Mortality Hot Spots in Texas

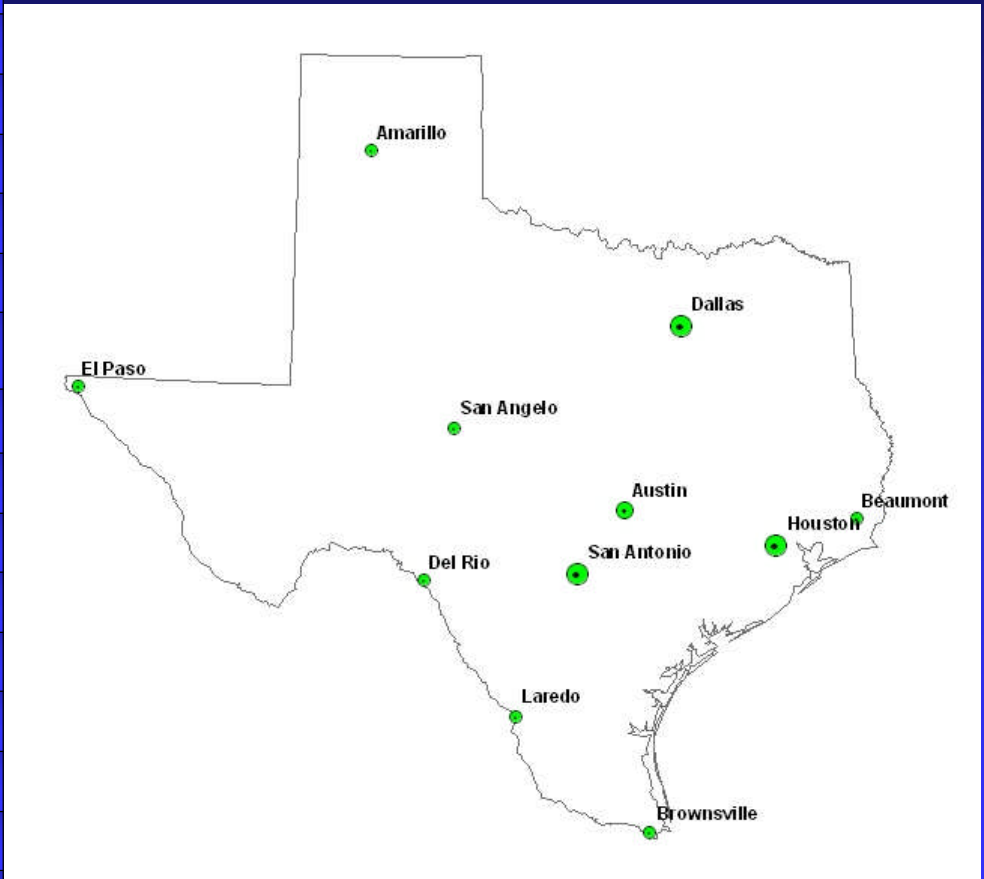
F. Benjamin Zhan

Professor and Director
Texas Center for Geographic Information Science
Texas State University—San Marcos
and
Chang Jiang Scholar Guest Chair Professor
School of Resource and Environmental Science
Wuhan University

zhan@txstate.edu

Background

Type of Cancer
Lung and bronchus
Colon
Female Breast
Prostate
Pancreas
Non-Hodgkin's lymphoma
Brain and other nervous system
Liver
Cervix Uteri
Leukemia
Stomach
Kidney
Bladder
Melonoma of skin
Larynx
Gallbladder



Epidemiology of Cancer and Its Complexity

- **Individual susceptibility**
  - Genetic factors
  - Behavioral issues
  - Diet and nutrition
- **Environmental exposures**
  - Environmental agents
  - Socio-economic conditions
- **Time**
- **Interaction of the above 3 dimensions**

Research Question

Where are the mortality clusters (hot-spots) for some 16 selected cancers?

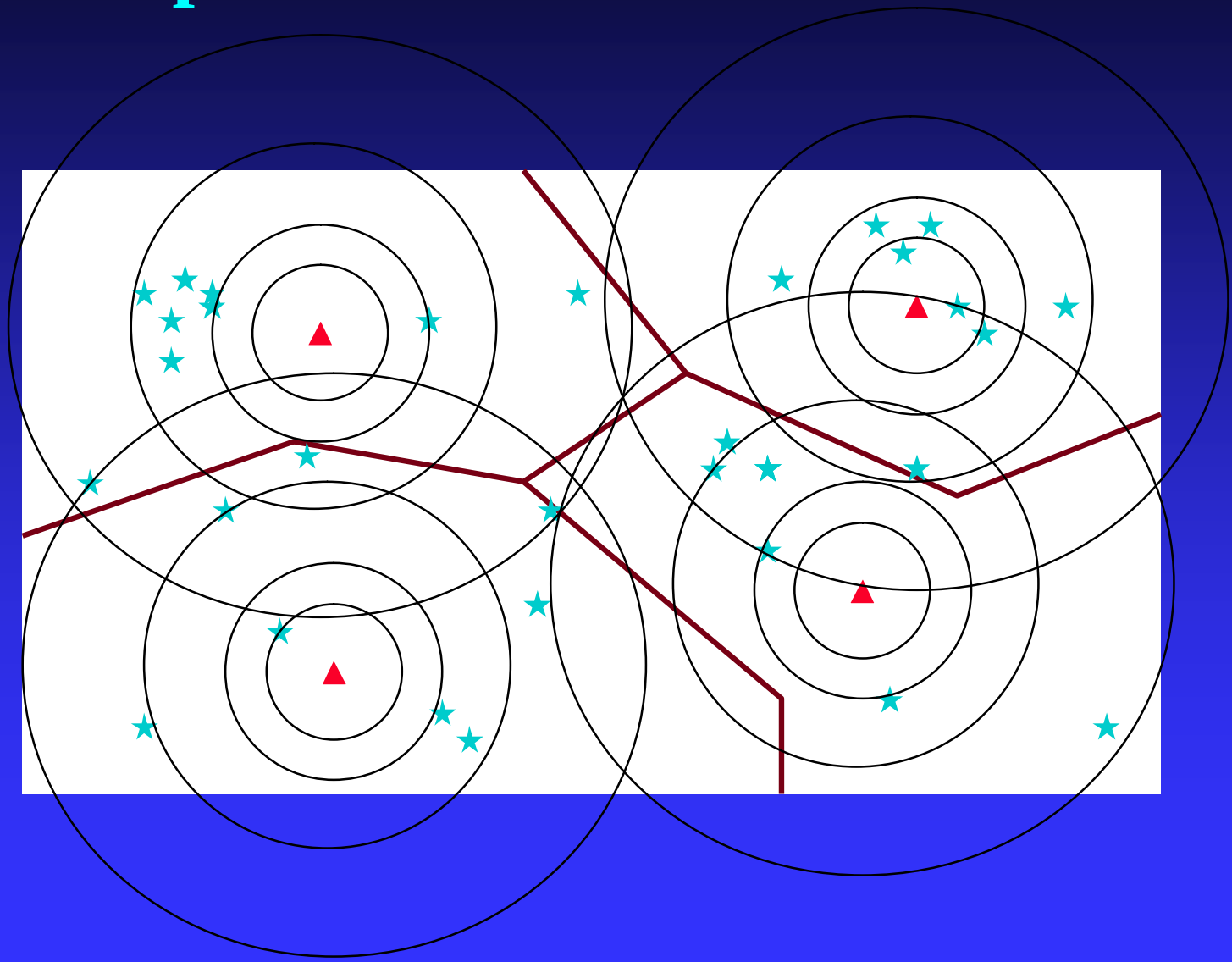
The Role of Spatial Cluster Analysis

- Help understand the processes leading to different types of cancer
- Improve our understanding about the role of environmental and socio-economic factors involved in the processes
- Availability of vast amount of spatial data
- Maturity of GIS technology
- Starting point – Where are the hot-spots?

Analysis Methods Available

- **Traditional methods:** Variance-mean ratio; nearest neighbor
- **Exploratory analysis methods:** Openshaw et al. (1987); Fotheringham and Zhan (1996)
- **Spatial clustering methods suitable for hypothesis testing:** Turnbull et al. (1990); Kulldorff (1997)
- **Method to use:** Kulldorff's Spatial Scan Statistic

The Spatial Scan Statistic



The Spatial Scan Statistic (continued)

- Computation of likelihood ratio
- Determination of the most likely cluster
- Determination of secondary clusters
- Evaluation of significance of a cluster
- The p value and likelihood ratio

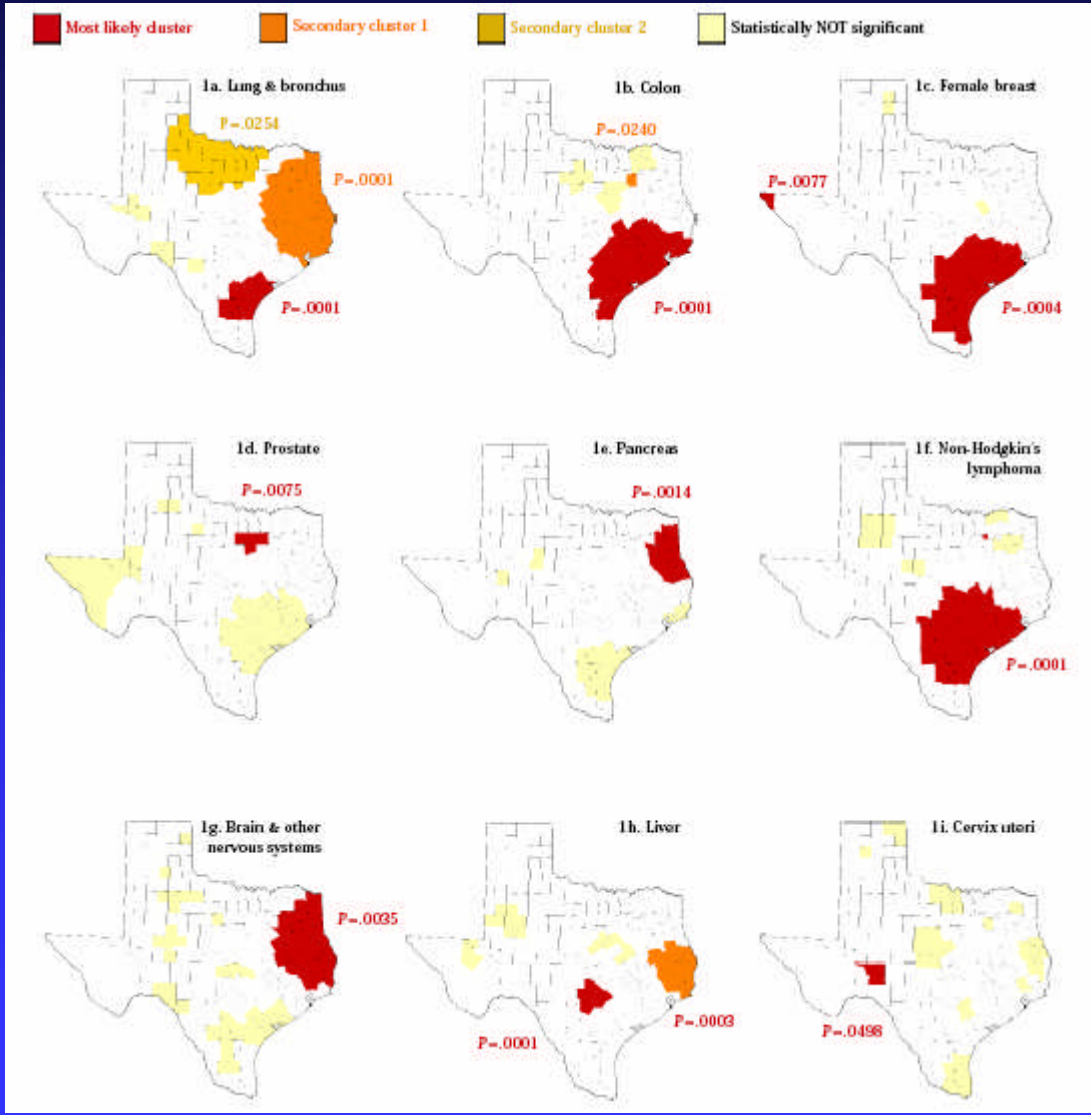
Data Preparation

- Cancer mortality data – geographic area unit resolution at the county level
- County level at-risk population data
- Geographic location data – county polygon centroids
- Mortality rates were adjusted for race (White, Black, Hispanic, and Other), age group, and sex

Overview of Data and Results, 1990-97

Type of Cancer	No. of Cases	Clusters?
All cancers combined	245,334	Yes
Lung and bronchus	71,678	Yes
Colon	21,352	Yes
Female Breast	18,993	Yes
Prostate	15,146	Yes
Pancreas	12,273	Yes
Non-Hodgkin's lymphoma	9,837	Yes
Brain & other nervous system	6,929	Yes
Liver	6,607	Yes
Cervix Uteri	2,658	Yes
Gallbladder	928	Yes/No
Leukemias	9,639	No
Stomach	6,536	No
Kidney	5,554	No
Bladder	4,384	No
Melonoma of skin	3,322	No
Larynx	1,915	No
Childhood cancer (ages 0-14)	1,112	No
Childhood cancer (ages 0-19)	1,590	No

Summary


Most likely cluster
Secondary cluster 1
Secondary cluster 2
Statistically NOT significant
1a. Lung & bronchus
P=.0254
P=.0001
P=.0001
1b. Colon
P=.0240
P=.0001
1c. Female breast
P=.0077
P=.0004
1d. Prostate
P=.0075
1e. Pancreas
P=.0014
1f. Non-Hodgkin's lymphoma
P=.0001
1g. Brain & other nervous systems
P=.0035
1h. Liver
P=.0001
P=.0003
1i. Cervix uteri
P=.0498

Answers to the Questions

- **Where are the cancer mortality clusters (hot-spots) in Texas?**
- **Answer:**
  - **Nine counties in the Houston area**
  - **Ten counties in the Corpus Christi area**
  - **Five counties in the San Antonio area**
  - **Twenty counties in northeast Texas**
  - **Val Verde county**

Concluding Remarks

- It is a screening process.
- The results will help public health officials target areas that warrant further investigation.
- Analysis using data at a finer resolution is desired.