Anomaly Detection in Space-Time (and higher dimensional) Point Processes

Michael D. Porter

porter@stat.ncsu.edu

VIGRE Post-Doc

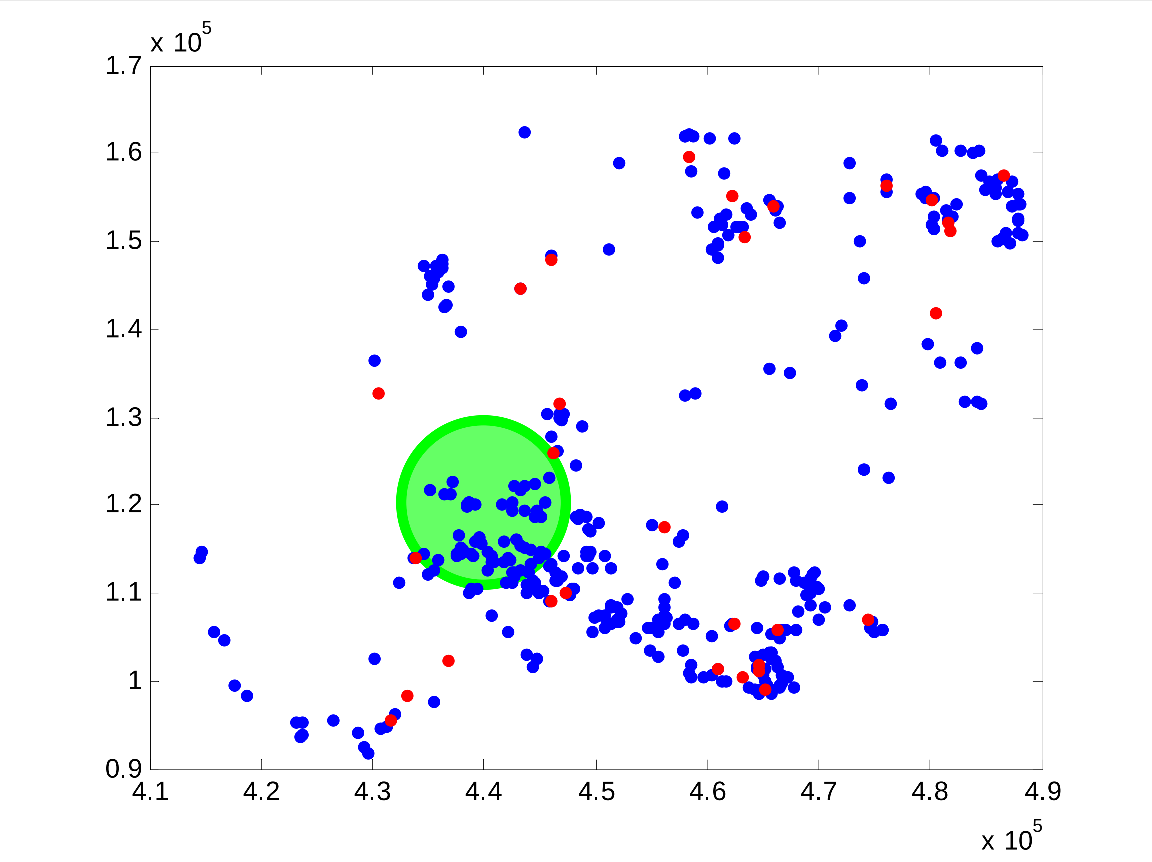
Department of Statistics

North Carolina State University

AEGISS Data

Diggle, P, & Rowlingson, B, & Su, T.L. (2005) Environmetrics

AEGISS - Ascertainment and Enhancement of Gastrointenstimal Infection Surveillance and Statistics



Blue: 1 Jan 2003 – 10 Feb 2003

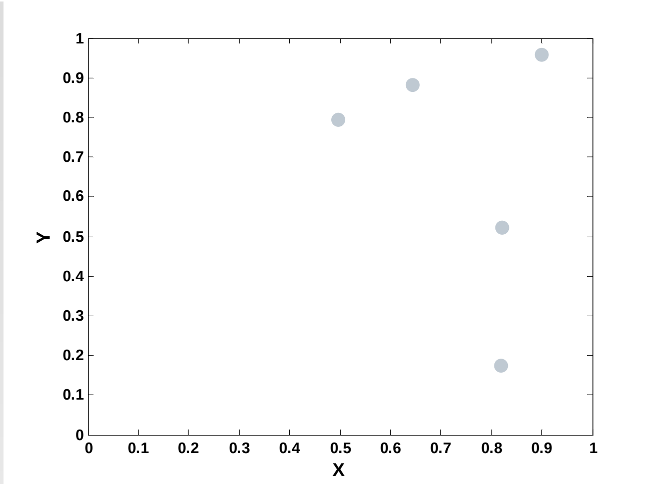
Red: 11 Feb 2003 – 16 Feb 2003

Problem Description

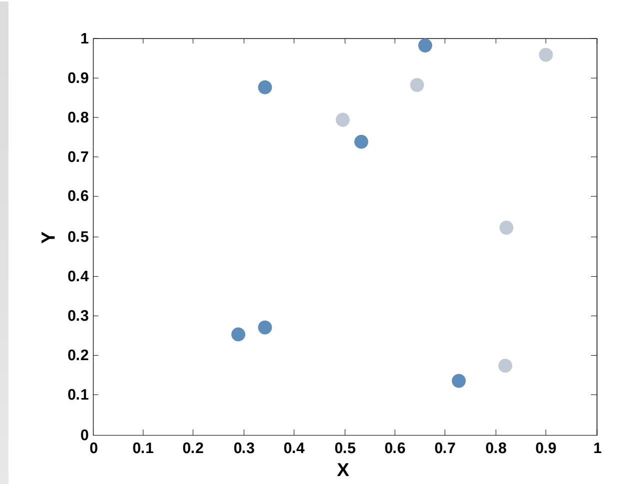
- **Quickly detect (space-time)**
  - Criminal/Terrorist outbreaks or change in tactics
  - Disease Outbreaks (Spatial and Syndromic)
  - Computer/Network Intrusions
  - Targets in Sensor Systems
- **Terminology: Change Detection**
  - Prospective Change Detection
  - Real-Time Change Detection
  - On-Line Change Detection
  - Surveillance
  - Stopping Time
  - Poisson Disorder Problem

Type of Change Considered

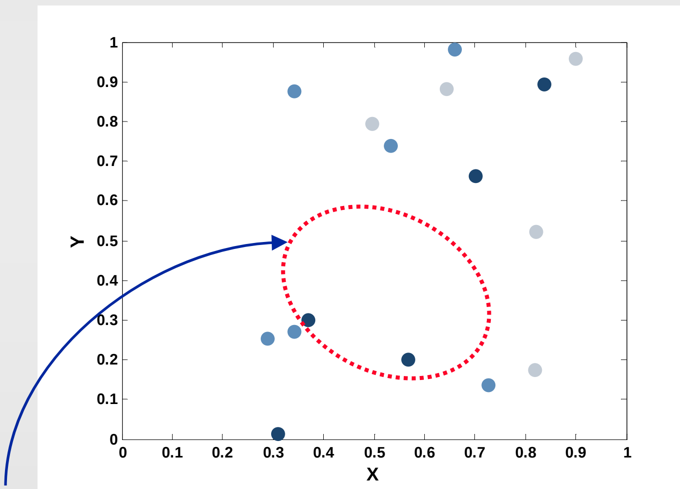
- **Anomaly (a specific type of change)**
  - **A localized (in space) region of unusual activity – in this case a change in intensity measure of the point process**
- **Anomaly Detection**
  - **Detection of *when* an anomaly occurs and the specific region *where* it occurs**
- **Spatial Anomaly vs. Time series**
- **Methodological Goal**
  - **`Quick Detection of Anomalies in Space-Time Point Processes'**



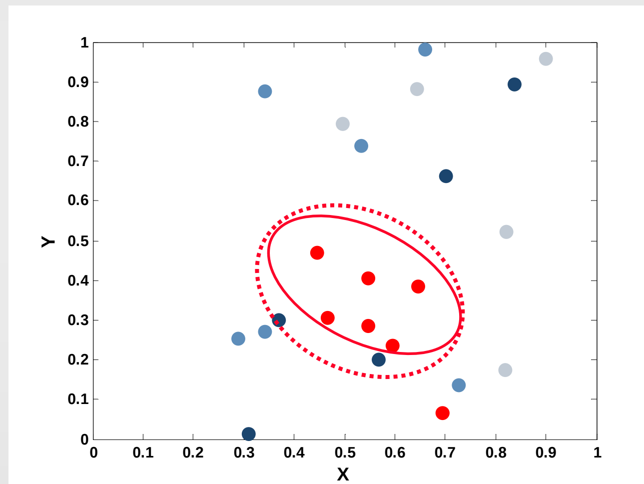
T=1, N(1)=5



T=2, N(2)=6

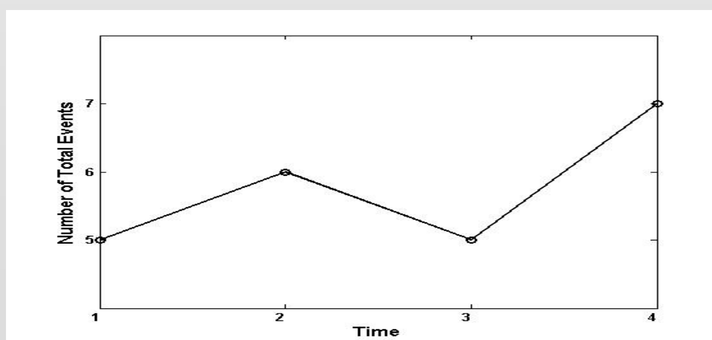


T=3, N(3)=5



T=4, N(4)=7

Or change at time T=3, not detected until T=4?



Problem Formulation

- **$N_t(B)$ is space-time Poisson process with intensity $\lambda_0(t,B)$**
  - **$t \in Z^+$ (Discretize time)**
  - **$B \subseteq W \subset R^d$ (or more generally R^k x C^l)**

$$H_o : N_t(B) \sim \lambda_0(t,B) \qquad \{t \in \mathcal{R}^+, B \subseteq W\}$$

$$H_a : \exists\, B^*, t^* :$$

$$\text{s.t.} \quad N_t(B^*) \sim \delta \cdot \lambda_0(t, B^*) \quad \{t \geq t^*, \delta \neq 1\}$$

<u>Change parameters</u>:

t* - Temporal change point

B* - Spatial change region

δ – change rate

GLR for Space-Time Anomalies

- **Lorden's Generalized Likelihood Ratio (GLR)**
  - **When unknown post-change parameter**

$$U_t^* = \max_{1<k<t} \left[\sup_{\theta_1 \in \Theta_1} \prod_{i=k}^{t} \Lambda(i, \theta_1, \theta_0) \right]$$

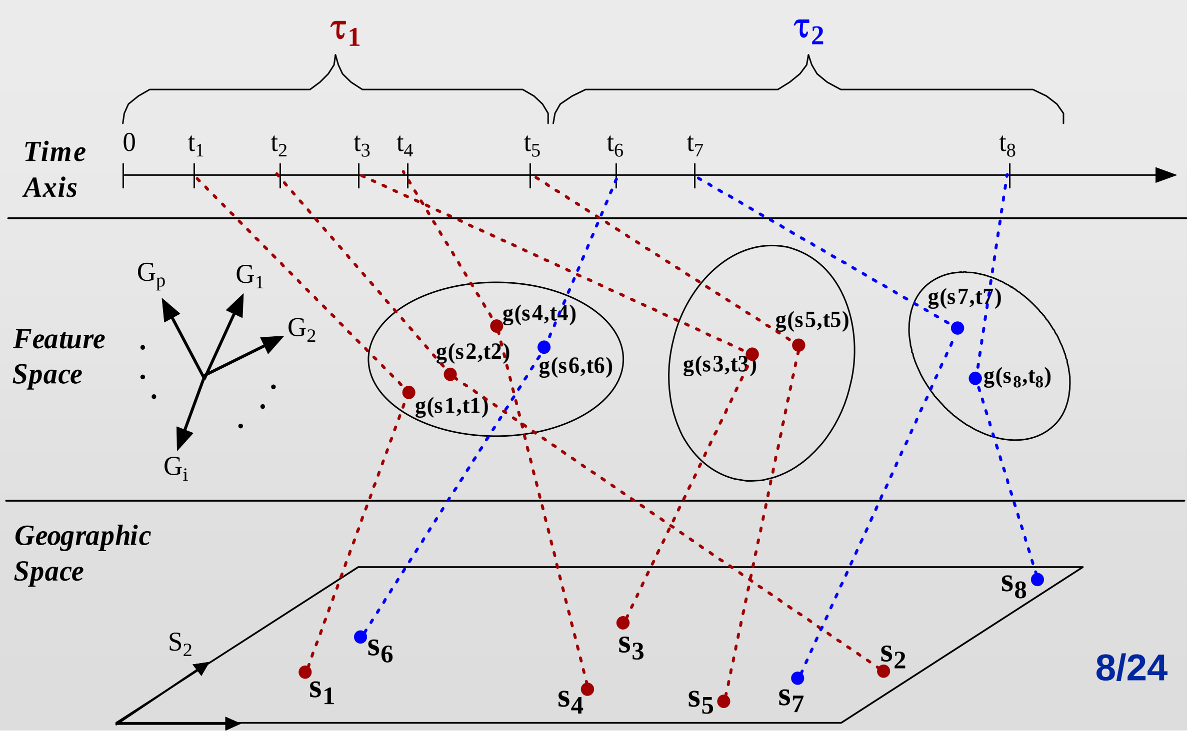
$$\text{Alternative: } \theta_1 = [B, \delta, \lambda_0(B)]$$

$$\text{Null: } \theta_0 = \lambda_0(B)$$

- **For each possible change point k, estimate the parameters under the alternative model, and calculate the likelihood ratio.**
- **Choose the value of k which maximizes the LR**
- **Space-Time Scan Statistic (Kulldorff, JRSS-A 2001)**

Finding Most Likely Change Region

- At every time t, and possible change point k, scan for the (connected) region B that maximizes the U_t^* statistic
- Scan in geographic space or feature space?
- Examples
  - Disease - space or features (covariates, risk factors)
  - Computer Intrusion – space or features
  - Site selection - feature
  - Sensors $\mathbb{R}^2$ or $\mathbb{R}^3$



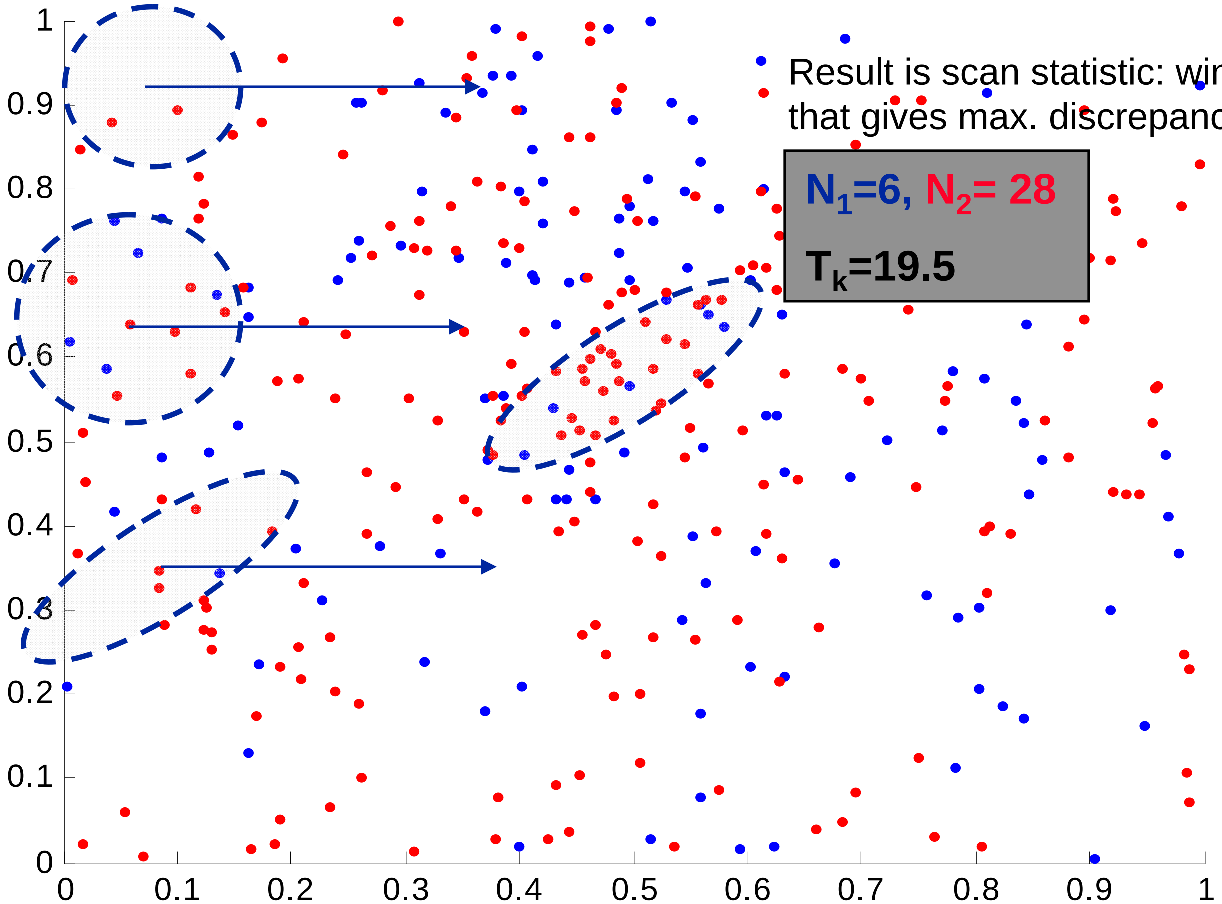
Generalized Scan Process

- A scan process can used to detect a (connected) region of unusual observations
- Create a window, W_s of some geometry and move it over the entire region of interest ($\forall$ s $\in$ W) and calculate some discrepancy score for each window, $D(W_s)$.
- Testing will be carried out on the ***scan statistic***, $T=\max_{s \in W} D(W_s)$.
  - Can include multiple window geometries
- The GLR statistics becomes $U_t^*=\max_{1<k<t} T_k$
- Problems:
  - Dependent upon the geometries and sizes of the windows
  - Gets difficult in high dimensions

Generalized Scan Process

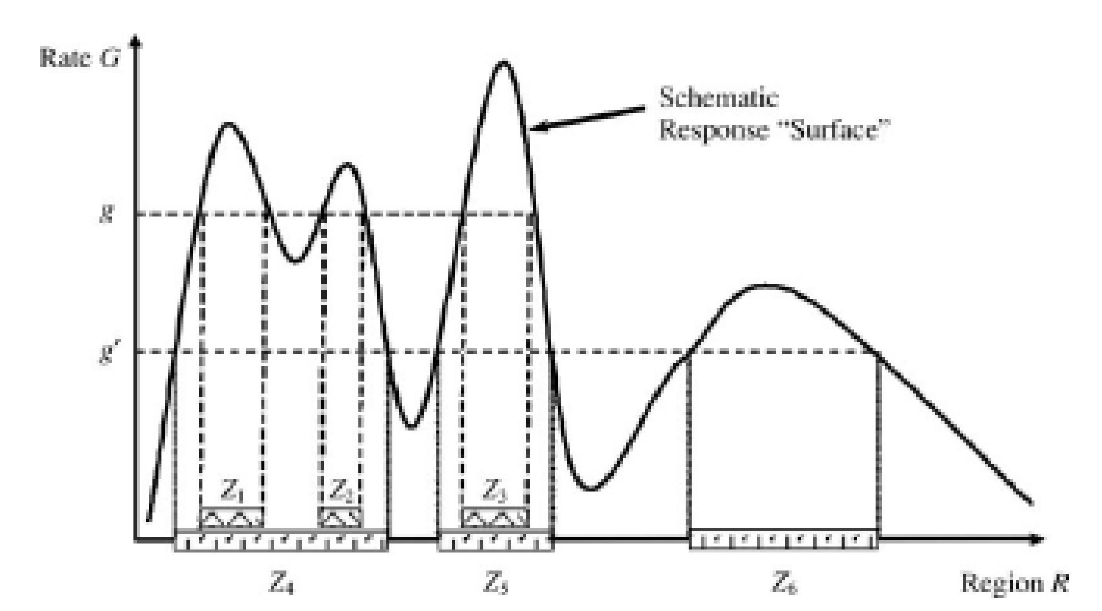
Choose windows and scan over entire region (for each window)

Result is scan statistic: window and location that gives max. discrepancy score



Alternatives to Scan Process

- ***Random Searches*** **(Porter and Brown, CSDA 2007)**
  - Modified PRIM for high dimensions (mixed variables)
  - Maximum of Random Partition Search
- ***Upper Level Set (ULS) Approaches*** **(Patil and Tallie, Env. & Ecol. Stats 2004)**
  - Kernel intensity estimation
  - Random Partition Search

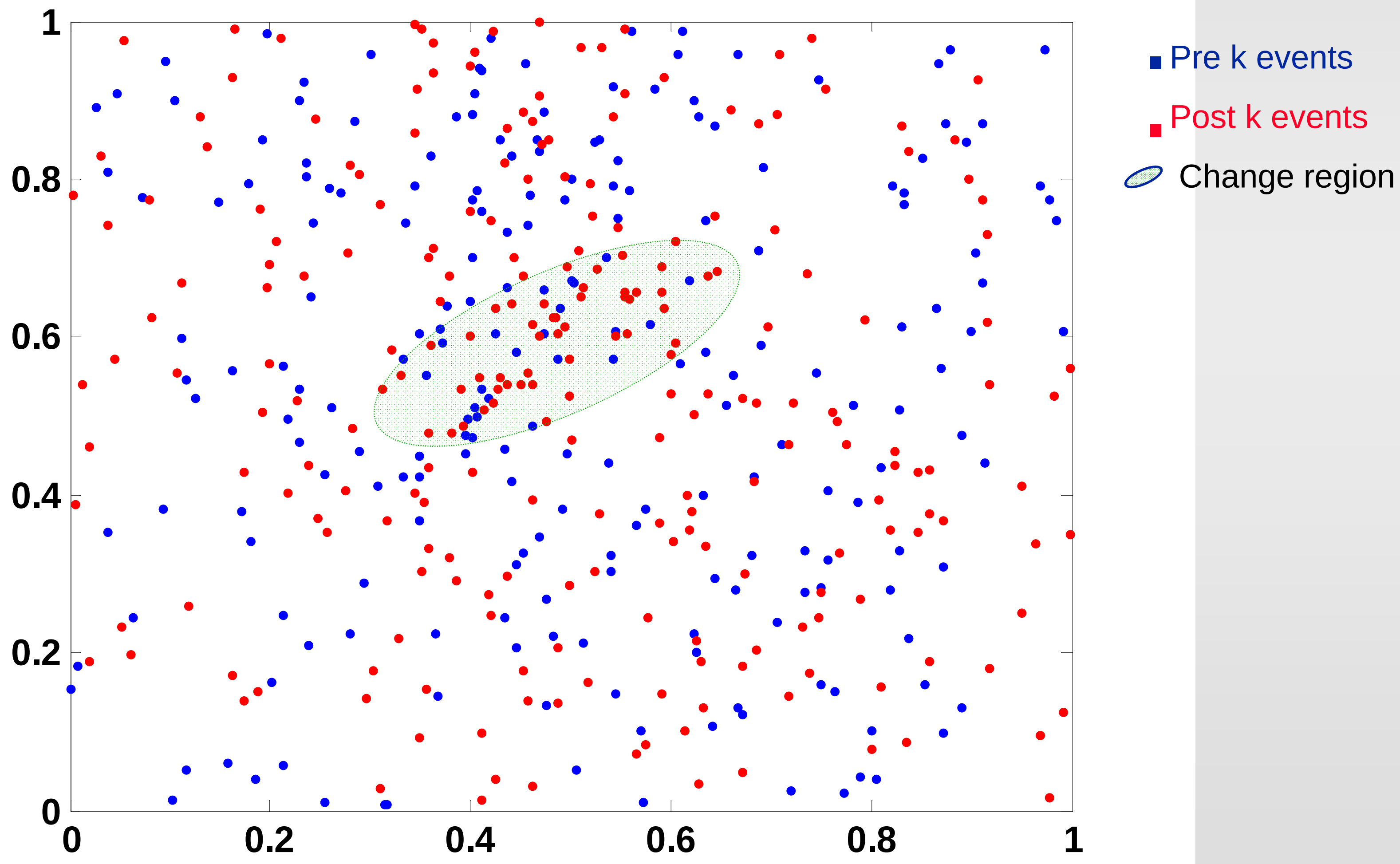


Patil and Tallie, Env. & Ecol. Stats 2004

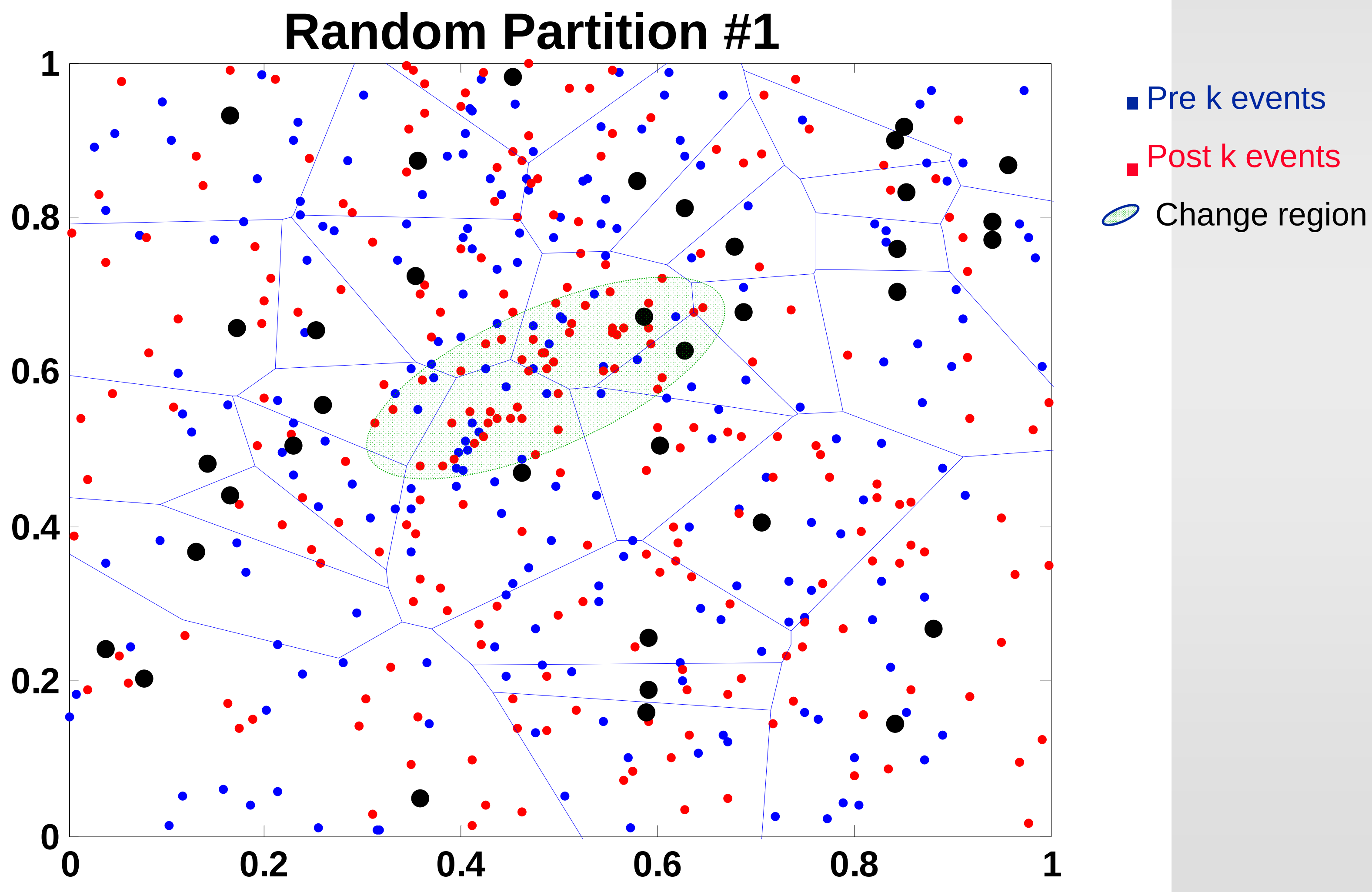
Random Voronoi Partition Search

- Mix of random searching and ULS to find region of maximum discrepancy
- Generate M Poisson Voronoi partitions of W
  - Generators are Poisson distributed with rate μ
  - For each partition, select the cell with the max discrepancy
- Identify B^* as either:
  - The max cell over all M
  - Score the region in each max cell with a value of 1, search through ULS to maximize U_t^*
    - » Here, only considering connected regions.

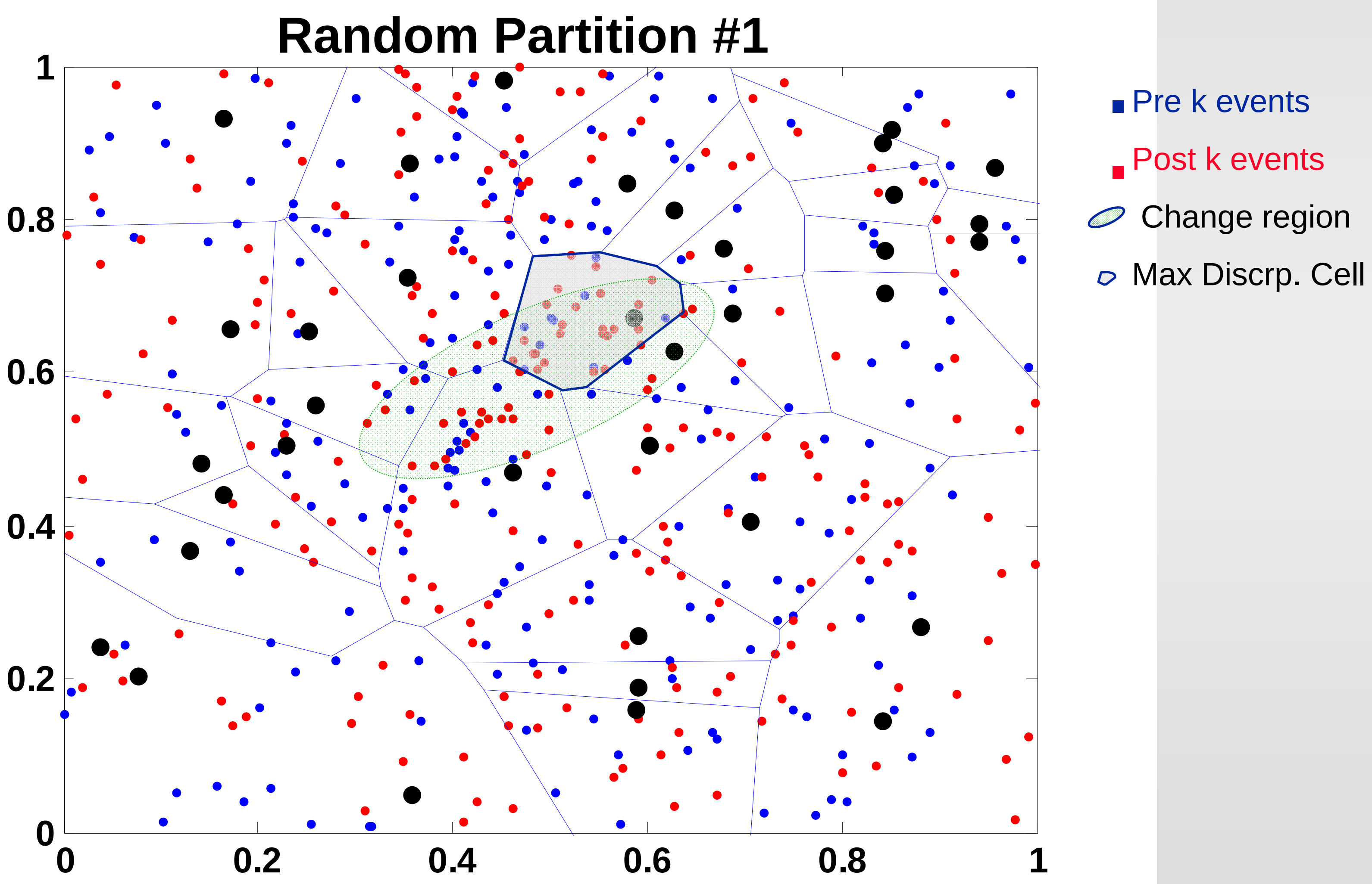
Realization of Space-Time PP



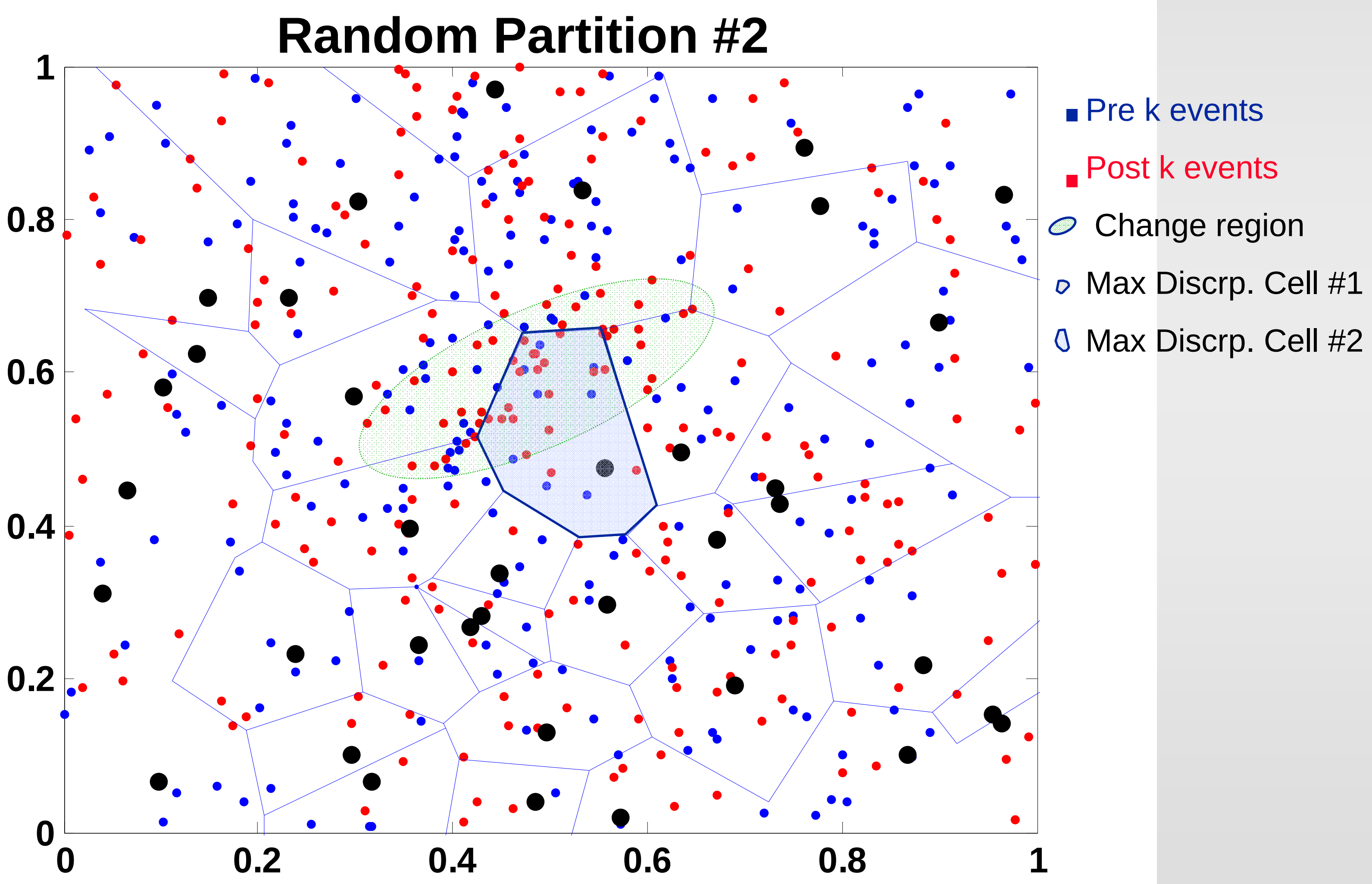
Random Voronoi Partition Search



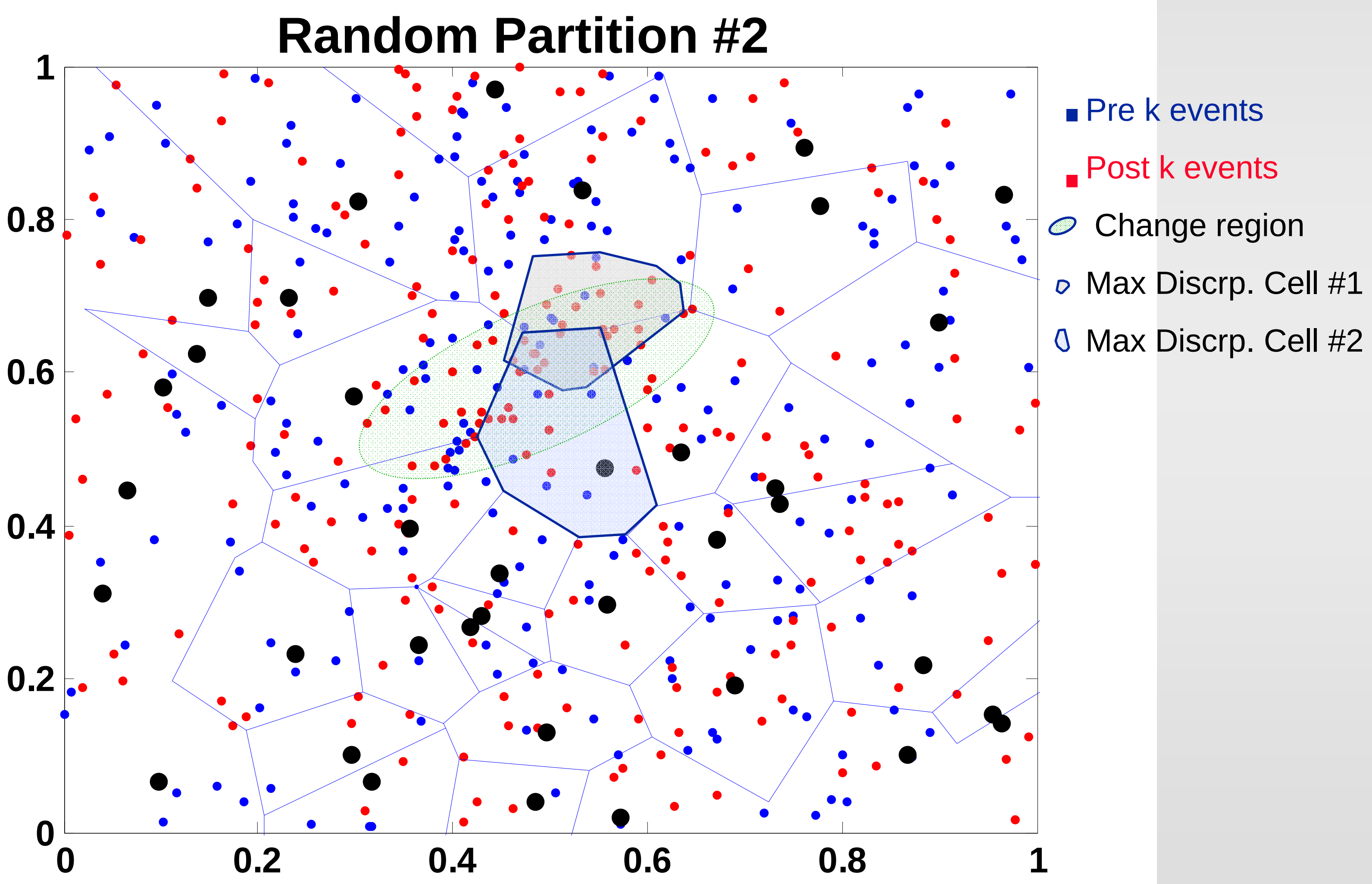
Random Voronoi Partition Search



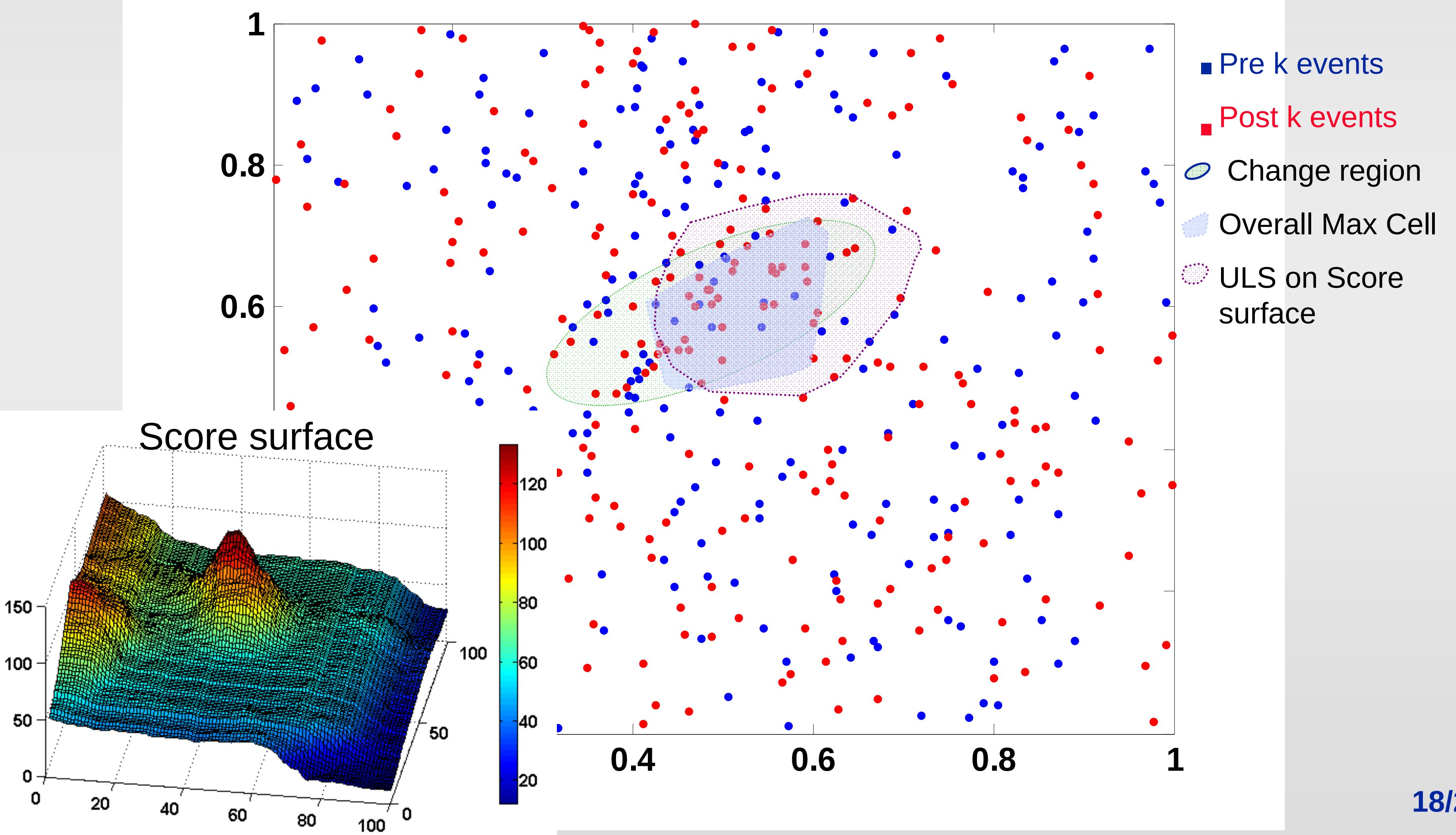
Random Voronoi Partition Search



Random Voronoi Partition Search



Random Voronoi Partition Search



Problems with GLR

- GLR can become very large, even under the null model
  - Especially if allowed to search over large class of shapes
- Anomalous region difficult to estimate precisely
- This can result in an increased false alarm rate as U_t^* will take large values even if no change
- Difficult to obtain performance bounds
  - Distribution of U^*_t unknown
- Mitigate these problems through *adaptive* procedures

Adaptive Procedures

Dragalin, EQC 1997; Lorden & Pollak, Annals 2005; Tartakovski, et al, Stat. Meth. 2006

- **Uses parameter estimates from previous data only**
  - **At time t-1, an estimate of θ_1, θ_0 is made**
  - **These values are used in the likelihood ratio at time t**
- **Adaptive CUSUM (AdC)**

$$\tilde{U}_t^* = \max_{1<k\leq t} \prod_{i=k}^{t} \Lambda(i, \theta_1^{i-1}, \theta_0^{i-1})$$

- **Adaptive SR (ASR)**

$$\tilde{R}_t^* = \sum_{k=1}^{t} \prod_{i=k}^{t} \Lambda(i, \theta_1^{i-1}, \theta_0^{i-1})$$

where θ^{t-1} is an estimate of θ based only on observations $\{N_1, \ldots, N_{t-1}\}$

Adaptive Properties

- $E[\Lambda(t, \theta_1^{t-1}, \theta_0)|\mathcal{H}_0] = 1$
  for any value of θ_1^{t-1}.

- **If t<t* then multiplying the $\Lambda(t, \theta_1^{t-1}, \theta_0^{t-1})$ should remain relatively small**

- **However, if t $\geq$ t* then $\theta_1^{t-1} = (B^{t-1}, \delta^{t-1}, k^{t-1})$ are the ML estimates of the post change parameters**
  - $\tilde{U}_t^*$ **and** $\tilde{R}_t^*$ **will accumulate quickly**

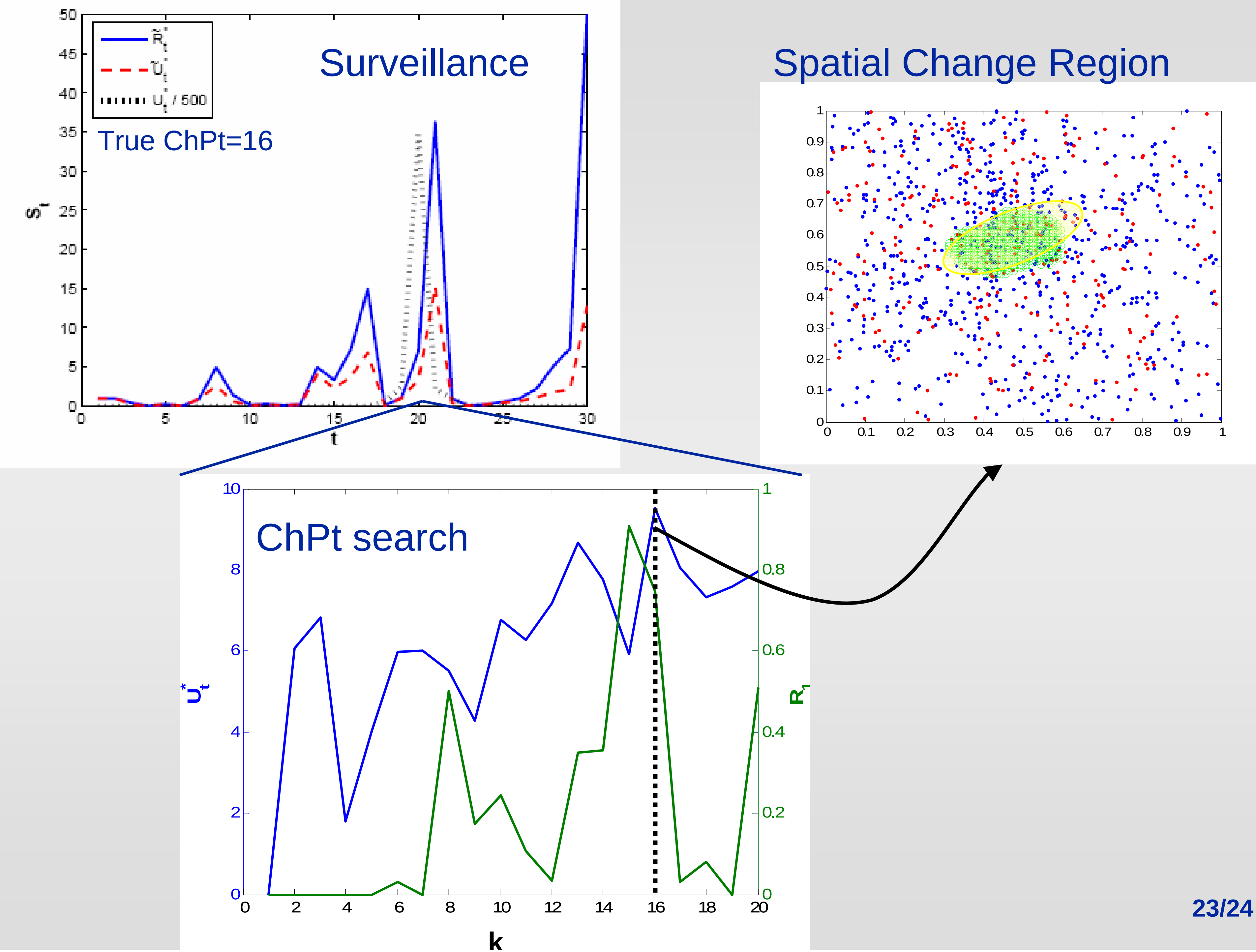
Estimating Adaptive Parameters

- **At time t-1, estimate $(\widehat{t^*}, \hat{B}, \hat{\delta}, \hat{\lambda}_0(\hat{B}))$ and use these as the parameter estimates at time t**

$$(\widehat{t^*}, \widehat{B^*}, \hat{\delta}) = \arg\left(\max_{0<k\leq t-1}\left[\sup_{B\in\mathcal{B}_W}\sup_{\delta\geq 0}\Lambda_0^{t-1}(k, B, \delta, \lambda_0(B))\right]\right)$$

Let $(K, B^{t-1}, \delta^{t-1}) = (\hat{t}^*, \widehat{B^*}, \hat{\delta})$.

Previous data only


Surveillance
True ChPt=16
Spatial Change Region
ChPt search
k
t

Other Considerations

- Setting Alarm decision rules
  - Threshold controls false alarm rate and detection delay
  - Adaptive procedures have good bounds
- Performance Measures
  - Bayes Risk, Neyman Pearson, Minimax
  - E[detection delay], E[False Alarm Rate]
  - How much of B^* must be captured?
- Also for marked PP's (space-time case-control, competing animal habitat, etc)
- Comparison to space-time scan statistic (SatScan)

-¿QUESTIONS?-

Michael D. Porter

porter@stat.ncsu.edu

Department of Statistics
North Carolina State University