

Models to carry out inference

vs.

Models to mimic (spatio-temporal) systems

5/5/15

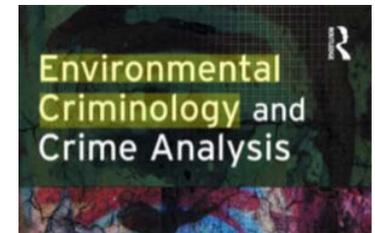
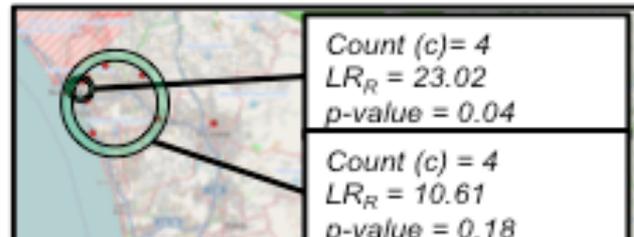
Where is a crime source?

- Representation choices beyond Linear Algebra
- Environmental Criminology
 - Routine Activities Theory, Crime Pattern Theory, Doughnut Hole pattern
- Formulation: **rings**, where **inside** density is significantly higher than **outside** ...

Input



– Output: Ring Shaped Hotspot Detection (RHD)



Mathematics	Concepts	Relationships
Sets	Set Theory	Member, set-union, set-difference, ...
Vector Space	Linear Algebra	Matrix & vector operations
Euclidean Spaces	Geometry	Circle, Ring , Polygon, Line_String, Convex hull, ...
Boundaries, Graphs, Spatial Graphs	Topology, Graph Theory, Spatial graphs, ...	Interior, boundary, Neighbor, inside , surrounds , ..., Nodes, edges, paths, trees, ... Path with turns, dynamic segmentation, ...



Taxonomy of Models for Inference

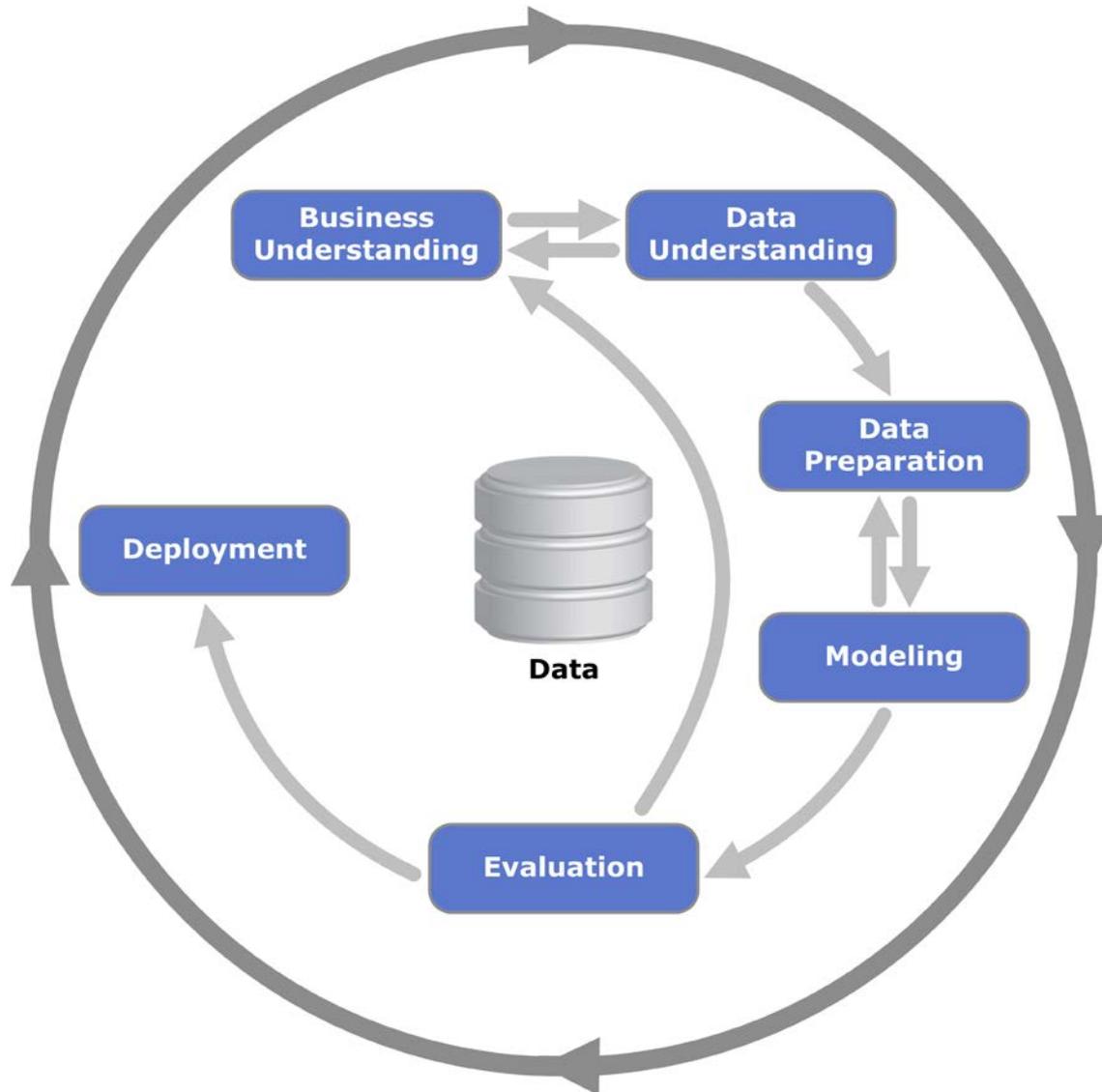
Models	Manual (Paper, Pencil, Slide-rules, log-tables, ...)	Computer-assisted (HPCC, cyber-infrastructure, data-intensive, big-data)
Process based	Differential Equations (D.E.), Algebraic equations, ...	Computational Simulations using D.E.s, Agent-based models, etc.
Information	Conceptual Data Model: Entity Relationship, UML, Semantic Web, ...	Abstract Data Types (Algebras): Sets, Vectors, Graphs, Points/Lines/Polygons, Relational Algebra, XML, ... Data Mining: frequent patterns, clustering, decision tree learning, ...
Empirical	Statistical: Regression, Correlations, Bayesian, ...	Machine Learning, Computational Statistics: Lasso, MCMC, kernel density estimation, ... Exploratory Data Analysis: data visualization, visual analytics, ... Spatio-temporal: Spatial statistics, Remote Sensing, GIS, satScan, change-point detection, ... Social Networks: ...

WHY?
WHAT?

WHEN?
WHERE?

WHO?

Empirical Models: Traditional Work-flow (CRISP-DM)

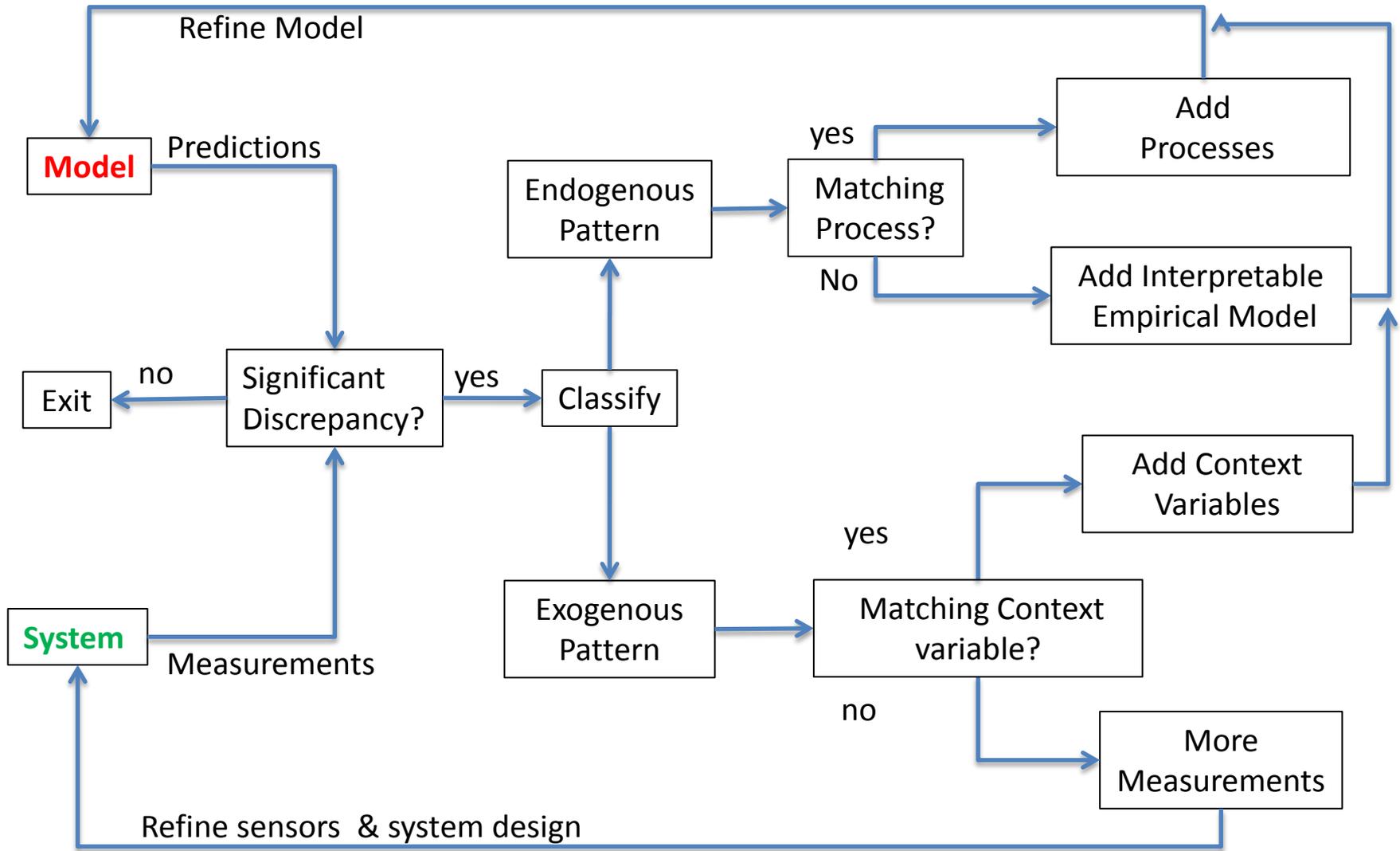


Empirical Models: Tradition Workflow

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
<p>Determine Business Objectives <i>Background</i> <i>Business Objectives</i> <i>Business Success Criteria</i></p>	<p>Collect Initial Data <i>Initial Data Collection Report</i></p> <p>Describe Data <i>Data Description Report</i></p>	<p>Select Data <i>Rationale for Inclusion/Exclusion</i></p>	<p>Select Modeling Techniques <i>Modeling Technique</i> <i>Modeling Assumptions</i></p>	<p>Evaluate Results <i>Assessment of Data Mining Results w.r.t. Business Success Criteria</i> <i>Approved Models</i></p>	<p>Plan Deployment <i>Deployment Plan</i></p> <p>Plan Monitoring and Maintenance <i>Monitoring and Maintenance Plan</i></p>
<p>Assess Situation <i>Inventory of Resources</i> <i>Requirements, Assumptions, and Constraints</i> <i>Risks and Contingencies</i> <i>Terminology</i> <i>Costs and Benefits</i></p>	<p>Explore Data <i>Data Exploration Report</i></p> <p>Verify Data Quality <i>Data Quality Report</i></p>	<p>Construct Data <i>Derived Attributes</i> <i>Generated Records</i></p>	<p>Generate Test Design <i>Test Design</i></p>	<p>Review Process <i>Review of Process</i></p>	<p>Produce Final Report <i>Final Report</i> <i>Final Presentation</i></p>
<p>Determine Data Mining Goals <i>Data Mining Goals</i> <i>Data Mining Success Criteria</i></p>		<p>Integrate Data <i>Merged Data</i></p>	<p>Build Model <i>Parameter Settings</i> <i>Models</i> <i>Model Descriptions</i></p>	<p>Determine Next Steps <i>List of Possible Actions</i> <i>Decision</i></p>	<p>Review Project <i>Experience</i> <i>Documentation</i></p>
<p>Produce Project Plan <i>Project Plan</i> <i>Initial Assessment of Tools and Techniques</i></p>		<p>Format Data <i>Reformatted Data</i></p> <p><i>Dataset</i> <i>Dataset Description</i></p>	<p>Assess Model <i>Model Assessment</i> <i>Revised Parameter Settings</i></p>		

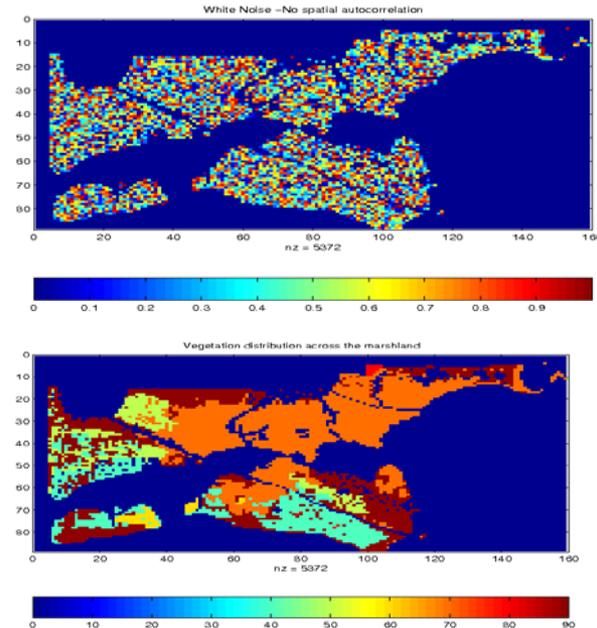
Figure 3: Generic tasks (bold) and outputs (italic) of the CRISP-DM reference model

Process-Empirical Models to mimic Systems: Workflow



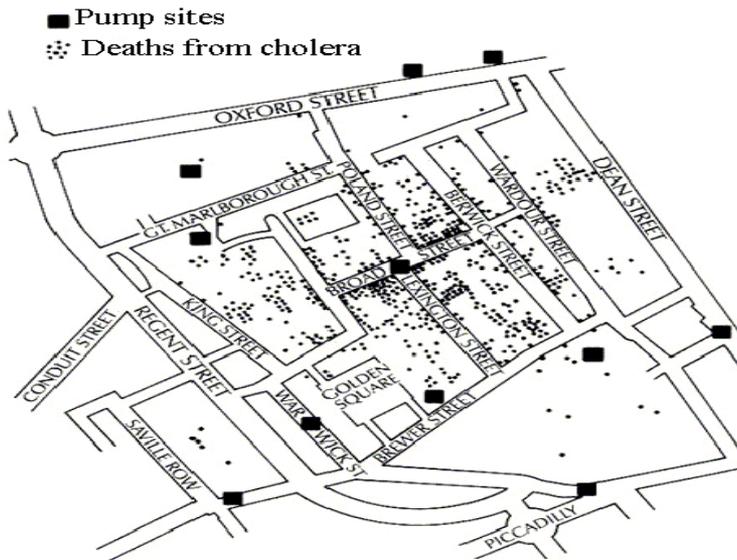
Empirical Models to mimic Systems

- Facilitate interpretation by system modelers
 - Reduce violations of system properties,
 - i.i.d. assumption in data science
 - vs. continuity of physical systems
 - Explicit modeling of domain properties
 - continuity inside regularizer
 - as spatial auto-regression term
- Relate to the goals of model building
 - Future system behavior – reduce prediction variance
 - Understand a system – reduce bias in parameters



Exogenous Patterns

- Interacting external systems,
 - e.g., food, energy, water ...
- Geographic,
 - e. g., location, time, ...
- Social,
 - e.g., social networks, organization
- ...



Rising economies

Gove

ate nge

NOTICE.

PREVENTIVES OF

CHOLERA!

Published by order of the Sanitary Committee, under the sanction of the Medical Council.

BE TEMPERATE IN EATING & DRINKING!

Avoid Raw Vegetables and Unripe Fruit!
Abstain from **COLD WATER**, when heated, and above all from *Ardent Spirits*, and if habit have rendered them indispensable, take much less than usual.

SLEEP AND CLOTHE WARM!

DO NOT SLEEP OR SIT IN A DRAUGHT OF AIR.
Avoid getting Wet!

Attend immediately to all disorders of the **Bowels.**

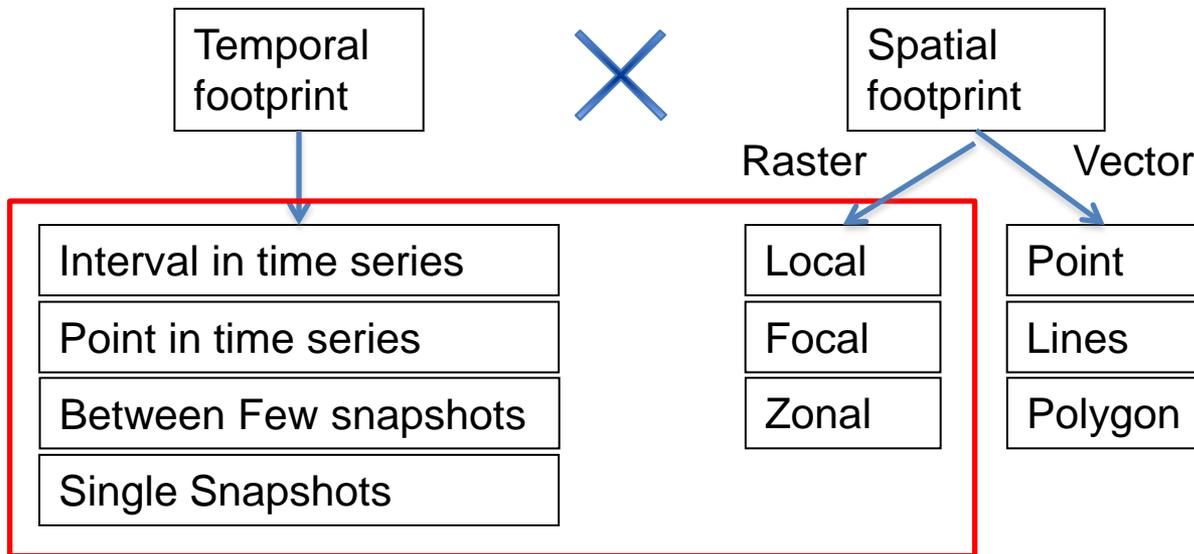
TAKE NO MEDICINE WITHOUT ADVICE.

Medicine and Medical Advice can be had by the poor, at all hours of the day and night, by applying at the Station House in each Ward.

CALEB S. WOODHULL, Mayor
JAMES KELLY, Chairman of Sanitary Committee.

Context: Spatiotemporal (ST) Models

- “Where” and “when” a phenomena occurs?
- A taxonomy of ST footprint:



WHAT?

WHEN?

WHERE?

WHO?

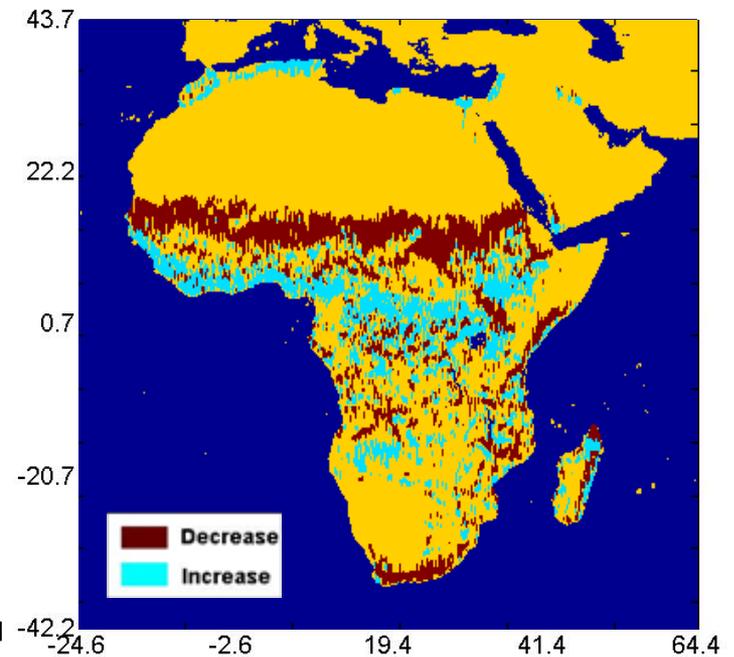
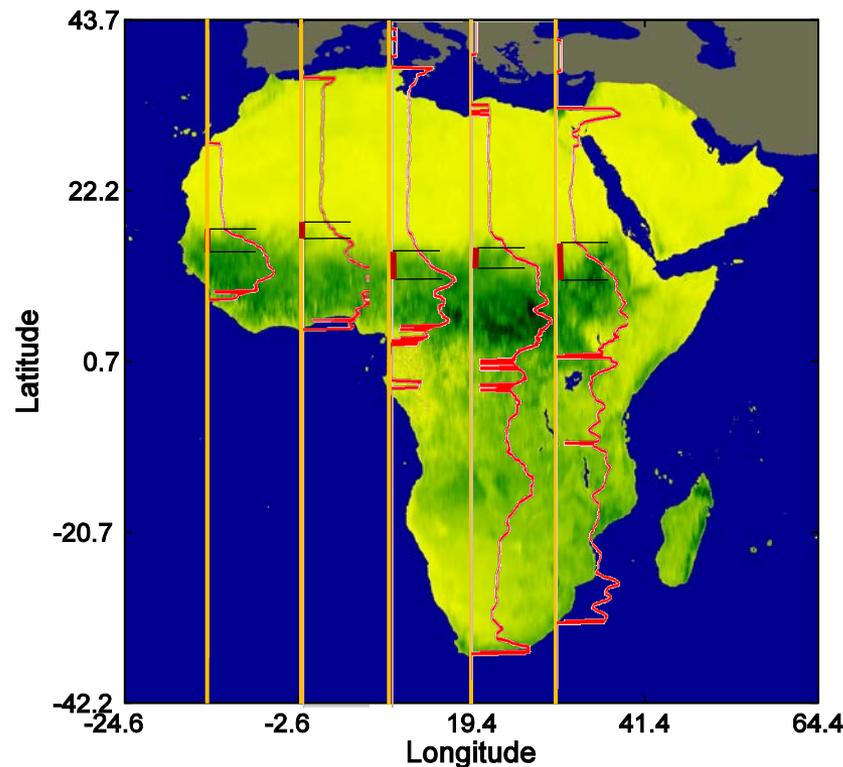
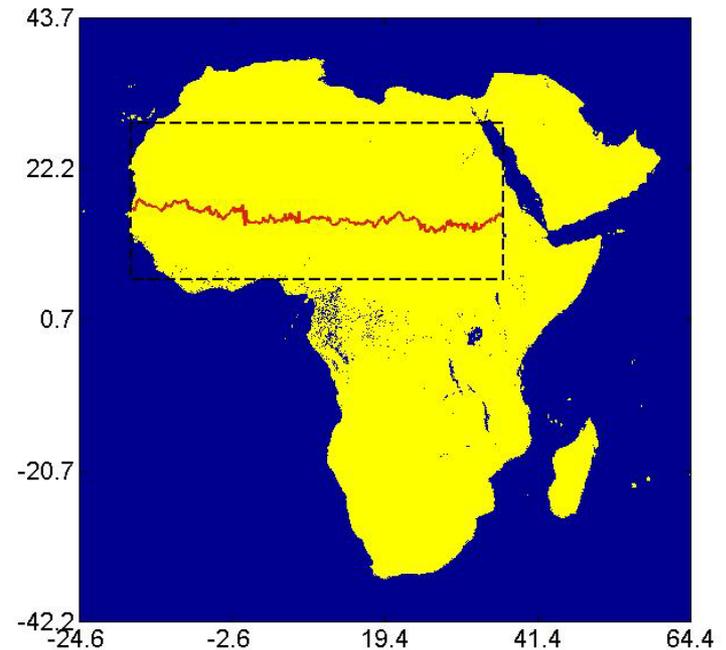
WHY?

Spatiotemporal change footprint (raster)

		Temporal				
		Snapshot	Few Snapshots	Time series (point)	Time series (interval)	Time series collections
Spatial	Local		Remote sensing change detection	Change Point Detection (e.g., CUSUM)	Change interval	
	Focal	Edge Detection (Wombling)				
	Zonal	Hotspots (e.g., Scan Statistics)			Persistent Change Window	

Where are eco-tones?

- Data: NDVI by GIMMS [4], Africa, 1981 August.
 - Resolution: 8km. Smoothed within 1x1 degree.
- Path: along each longitude (south → north)
- Interest measure: (Slope) Sameness degree $\frac{AVG\{\Delta\}}{AVG_{\geq\alpha}\{\Delta\}}$
 - Δ : unit slope
- Thresholds: $\alpha= 20\%$ percentile, $SD \geq 0.5$



Case Study: Global Data and Ecotones

- NDVI of the entire world
- Aug 1-15 1981, 0.07 degree (8km) resolution

