

**Monitoring the Earth from above:  
Mining satellite data to map urbanization  
at local to global scales**

AI in the Wild - Artificial Intelligence Seminar Series  
Department of Computer Sciences

Professor Annemarie Schneider  
Center for Sustainability and the Global Environment, University of Wisconsin-Madison  
March 6, 2012

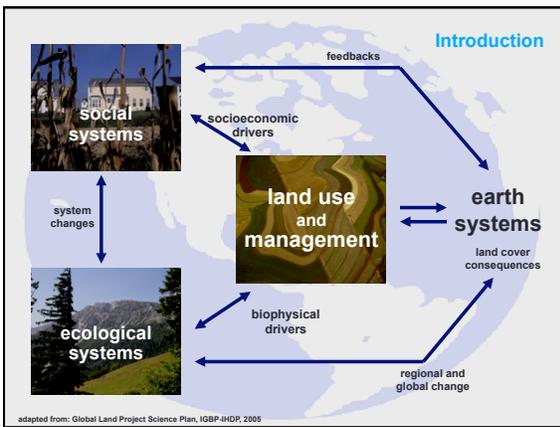
### Introduction

**Human dimensions of global change**

- How are humans changing the Earth?
- What are the larger patterns and trends occurring across places?
- What are the drivers and implications of this change?

New attention to urban areas in land use-land cover research

- Can urban development strategies be aligned with climate change adaptation?
- How can urban planning tools be used to develop more resilient cities?



### Introduction

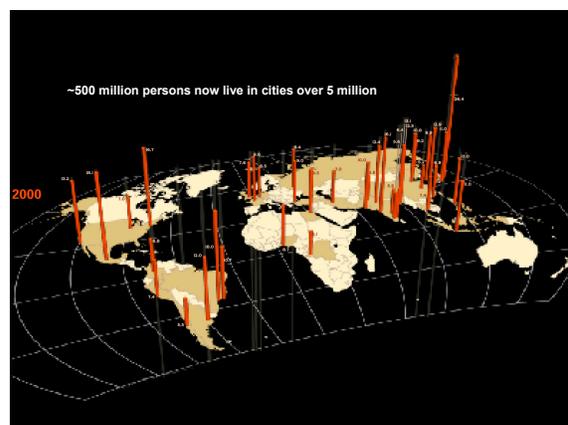
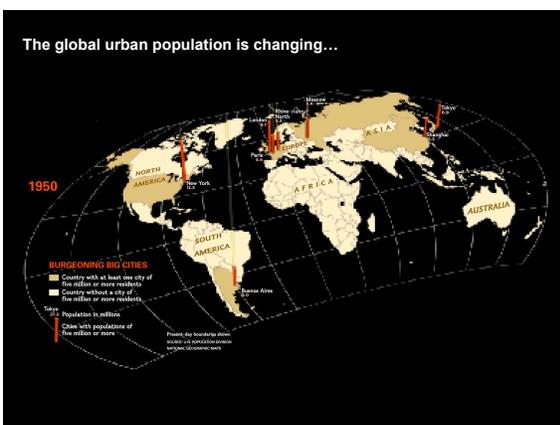
Why should we care about urban areas?

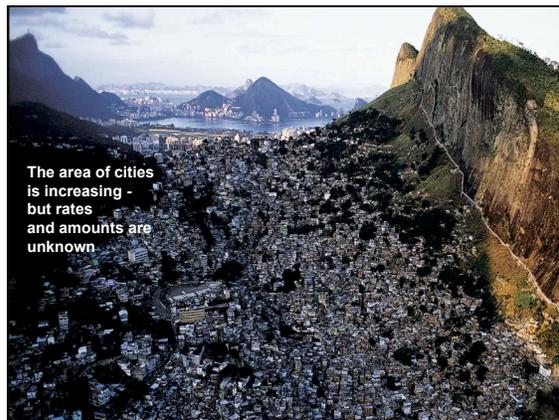
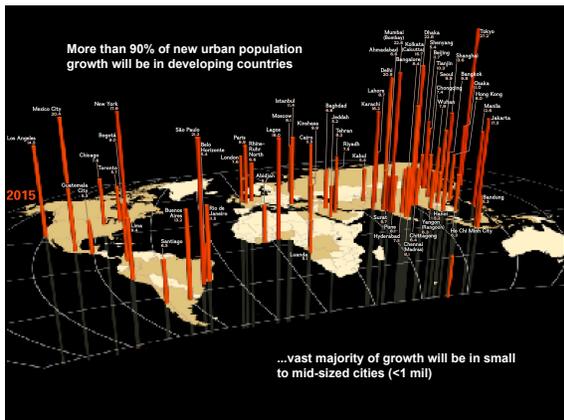
Urban population is increasing

over 50% of the population is urban in 2000, compared to 30% urban in 1950 (UN 2008)

**Current and projected global urban population growth 1950-2030**

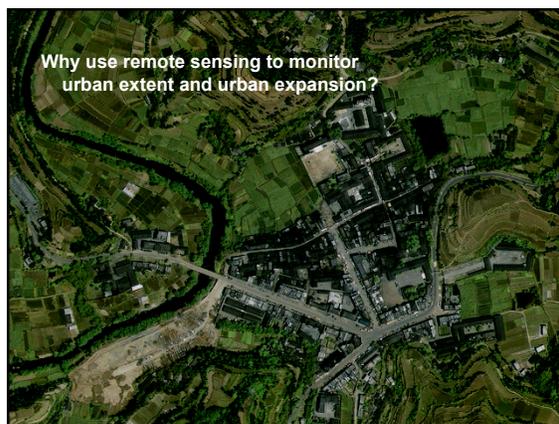
The chart shows a significant increase in urban population from approximately 1 billion in 1950 to over 4 billion in 2030. The growth is projected to be most rapid in Asia and Africa.





Monitoring urban expansion can improve understanding of the urban influence on the biosphere...

- microclimate, regional climate
- phenology, net primary production
- resource use, waste sinks
- loss, fragmentation of natural ecosystems
- air, soil and water pollution
- increased water use and runoff
- reduced biodiversity
- introduction of non-native species



**Introduction**

What is remote sensing?

- Measurement, analysis of an object without being in direct contact with it
- Science of deriving information about the earth's land and water areas from *images* acquired at a distance
- Relies on measurement of *electromagnetic energy* reflected or emitted from features

**Introduction**

What is remote sensing?

- Science of deriving information about the Earth from *satellite images* and aerial photographs

Why use remote sensing?

- Synoptic view
- Repeat coverage
- Space matters - ability to map trends in space and time
- Faster, cheaper than ground surveys
- Access to unmapped regions, difficult terrain
- Allows routine monitoring of change
- Cross-cutting scales of information

### The issue of scale

(1) Neighborhood scale  
Very high resolution (4 m)  
High level of detail,  
small scene footprint

(2) Local, city-scale  
Medium resolution imagery (30 m)

(3) Regional to global scale  
Coarse resolution imagery (500-1000 m)  
Sacrifice detail to gain spatial coverage

### Defining urban areas

The fundamental question -  
how do we distinguish  
urban from rural land?

- Population size threshold, population density
- Administrative boundaries
- Economic definition - presence of industrial activities, services; proportion of labor force in non-agricultural work
- Functional definition - commuter-shed
- Contiguity of built-up land

Nighttime lights data -  
Maps of light emissions that vary by level of economic development

### Lessons in defining urban areas

**Urbanization** as a physical manifestation on the ground...

transformation of land use to an *urban pattern* of organization

land use change from rural to urban, with replacement of pervious surfaces with impervious materials (asphalt, concrete, etc.)

development of an area into residential, industrial and commercial buildings and structures

### Defining urban areas for remote sensing applications

With satellite data,  
we can map **land cover** -

- Biophysical state of the Earth's surface
- Quantity, type of earth materials, surface vegetation, water
- Examples: forest, grassland, wetland, concrete

... but it is difficult to map **land use** -

- Manner in which land attributes are manipulated or managed
- Examples: agriculture, wildlife preserve, urbanized area

Urban areas are comprised of many different land cover types

### Outline

Introduction

Why use remote sensing?

Monitoring urban areas at a local scale

- Urban expansion in 40 metropolitan areas
- Classification, change detection methods

Monitoring urban areas globally

- Data mining multi-temporal global imagery
- Approaches using data fusion

Monitoring rapidly expanding urban regions in China

- Multi-temporal change detection using dense data stacks

Challenges and opportunities

Future directions

Built-up areas of Western Madrid, Spain

### Comparative urban analysis at a local scale

What do we know about global patterns of urban growth?

- Monitoring change -  
What is the rate of land conversion to urban uses that has occurred from 1990 to 2000 in cities across the globe?
- Assessing pattern -  
How have the shape, form and patterns of cities changed during the last decade?
- Quantifying urban sprawl -  
How do rates and forms of urban growth around the world compare to urban sprawl in the U.S.?
- Understanding drivers -  
What factors explain the widely divergent rates and shapes of urban growth in cities around the globe?

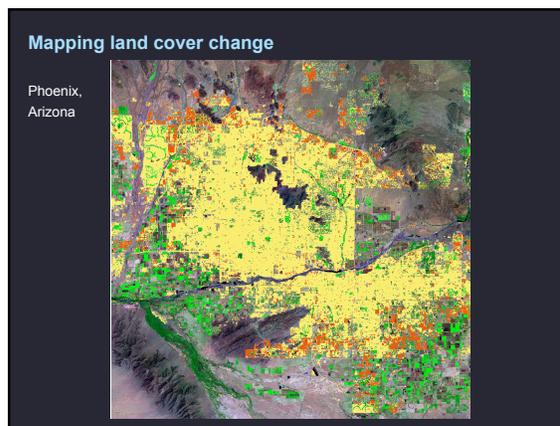
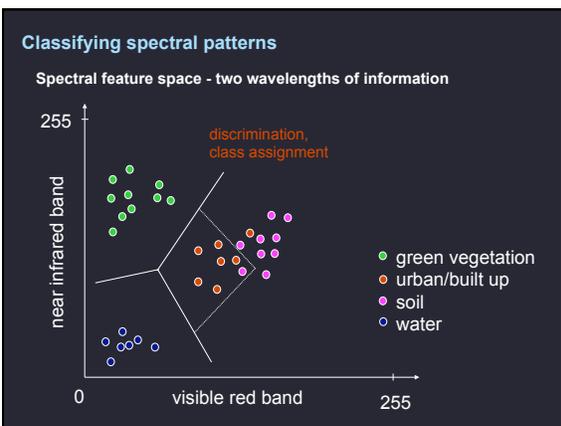
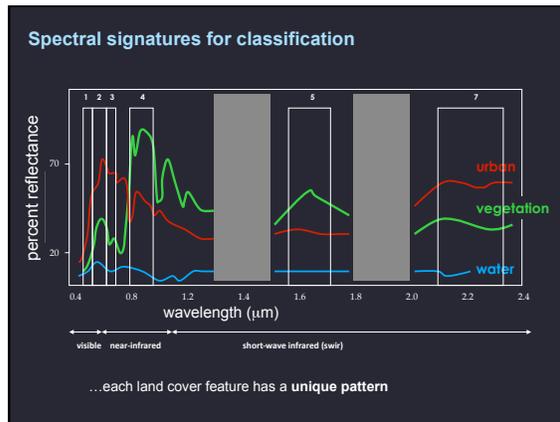
Urban sprawl typical of U.S. cities

### Comparative urban analysis - methods

Classification of remotely sensed imagery

- Assign labels to pixels
- Reduce information content to make manageable picture of landscape
- Pattern recognition based on **spectral signatures**
- Continuous vs. discrete label
- Methods?

desired result      classifier result



### Comparative urban analysis - methods

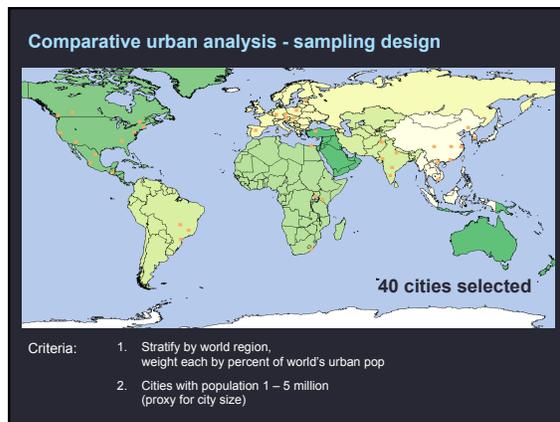
Choice of remote sensing method:

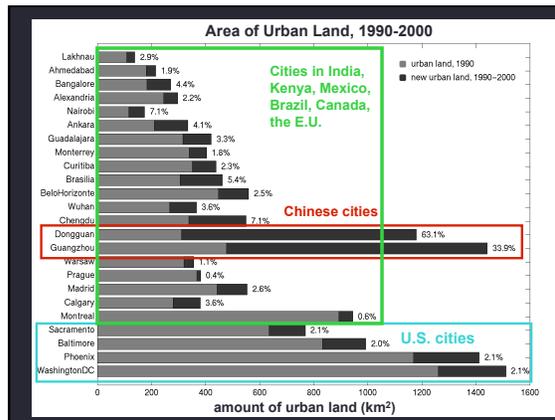
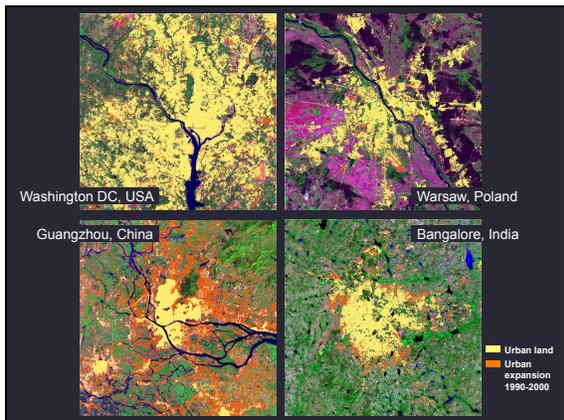
- Unsupervised multi-date change detection (k-means)
- Simple class structure: urban 1990, urban change 1990-2000, non-urban areas
- accuracy assessment

In addition...

- Quantify physical trends using spatial metrics
- Assess demographic, economic trends using census data at a disaggregated scale

Field work to understand peri-urban growth near Chinese cities

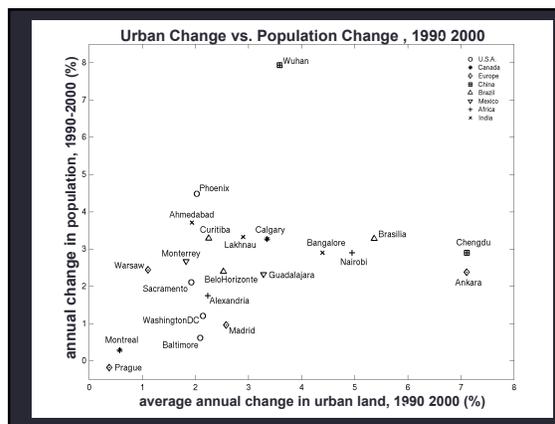




**Comparative urban analysis - results**

Taxonomy of city types, mid-sized metropolitan areas

group	high rate of change?	dispersed? (location)	fragmented? (pattern)	high population density?	example
expansive-growth	no	yes	no	no	Washington DC, Baltimore, Phoenix
frantic-growth	yes	yes	yes	yes	Guangzhou, Chengdu, Dongguan
high-growth	yes	no	yes	yes	Bangalore, Brasilia, Ankara
low-growth	no	no	no	yes	Guadalajara, Curitiba, Prague



**How are residential, industrial, and commercial areas changing across cities?**

What are the methodological requirements for mapping land use?

How can we monitor land use types rapidly and accurately using high resolution remote sensing data?

region-based spectral means    texture    classified map

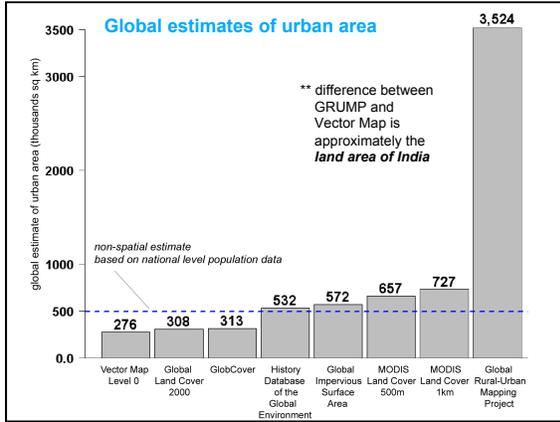
**Global maps of urban extent**

How much of the planet is covered by urban land?

How are the world's cities distributed across continents, regions, and countries?

Move towards regional, global analysis

- Global climate, carbon models
- Environmental indicators
- Social science applications



### Why such divergent estimates?

Differences in data inputs

- Daytime satellite observations → physical presence of built-up land
- Census data → population distribution
- Nighttime lights → income level

Differences in data resolution

- Early datasets: 1 - 2.2 km pixel
- New satellite data: 300-500 m pixel

Different definitions of what is 'urban'

- populated places → Digital Chart of the World
- artificial surfaces and associated areas → Global Land Cover 2000
- urban and built-up areas → MODIS Land Cover (IGBP)
- urban extent → GRUMP (CIESIN)

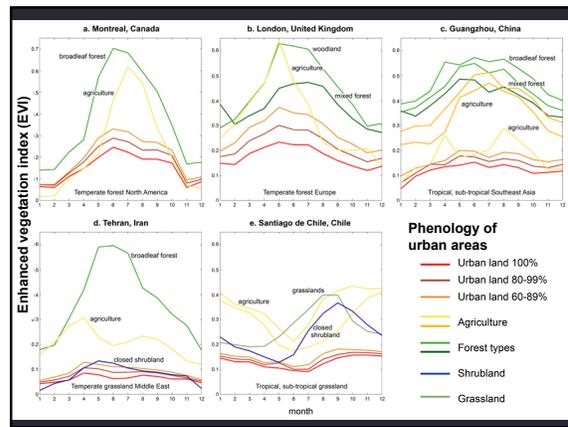
### Methods for mapping global urban extent

Global land cover classification

- One year of 8-day MODIS 500 m data
  - 7 bands nadir BRDF-adjusted data (0.4-2.2 μm), monthly composites plus EVI, annual metrics (min, max, mean of each band)
- Supervised decision tree classification
  - Recursive partitioning of training data into successively more homogeneous subsets
  - Entropy

$$E(S) = - \sum_{i=1}^c p_i \log_2(p_i)$$

Given set of cases S, with c possible classes, each with proportion p

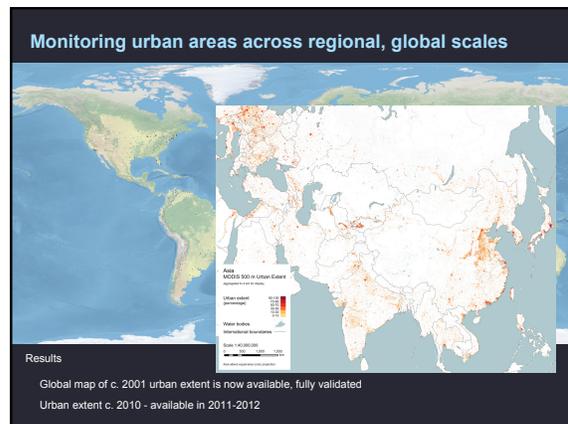


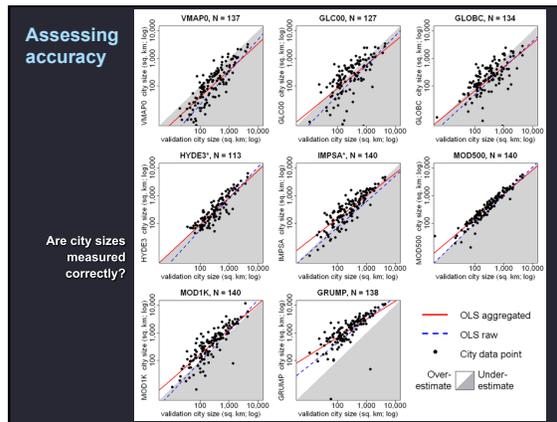
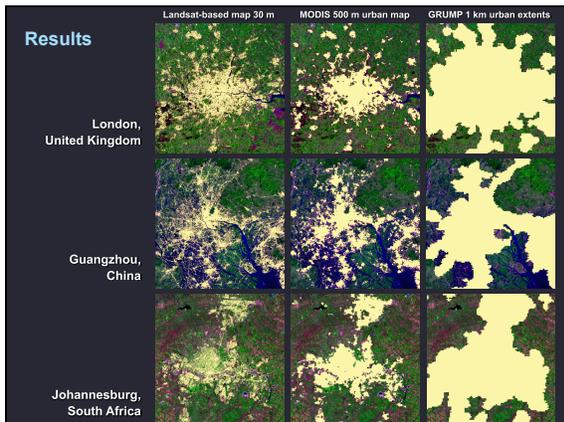
### Methods for mapping global urban extent

Global set of training data

- Parameterized from Landsat 7 ETM+, Google Earth

- Evergreen needleleaf forest
- Evergreen broadleaf forest
- Deciduous needleleaf forest
- Deciduous broadleaf forest
- Mixed forest
- Closed shrubland
- Open shrubland
- Wetlands
- Savanna
- Grassland
- Barren/rocky wetlands
- Croplands
- Urban areas
- Cropland vegetation mosaic
- Open ice
- Barren
- Water





### Monitoring regional urbanization

Understanding urbanization in China and the Monsoon Asia region

- Monitoring urban systems and land patterns regionally-globally using satellite data
- Local case-study analysis of geographically comprehensive sample of cities
- Predictive modeling, forecasting of dynamic socio-economic forces and land-based outcomes

Collaborators  
Kurt Paulsen, Jennifer Alix-Garcia (Wisconsin)  
Jianfa Shen (Chinese Univ. of Hong Kong)

### Monitoring regional urbanization in China

Tremendous opportunity to shape the built environment

~85% of urban development on the ground by 2050 in China and India will be built between now and then

### Urbanization and urban expansion in China

Policy reforms

- 1978 economic, land reforms: decentralization, land use rights, liberalization of household registration system (*hukou*) and work unit (*danwei*)
- 1990s great western development program
- 2001 new emphasis on villages

Impacts?

- Rapid rural-urban migration
- Rapid land use change
- Agricultural expansion, intensification
- GDP 1978-2008: 8-14%
- Increase in income - vehicles, housing, diet

### Monitoring regional urbanization in China

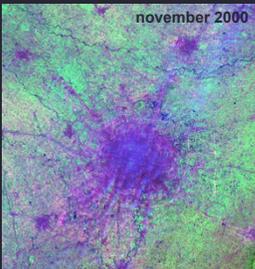
Sample of urbanizing regions...

Case study methods

- remote sensing - multi-date change detection
- assess trajectory of multiple time points
- monitor peri-urban, village development
- spatial analysis

### Urbanization and urban expansion in China - methods

Numerous change detection methods, 30 year history

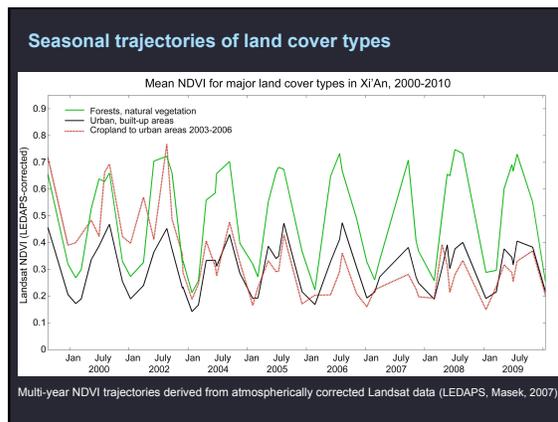


Common problems

- Complexity of landscape in SE Asia
- Confusion between new urban and bare agriculture plots
- Agricultural, vegetation variability

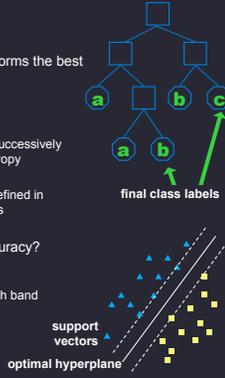
Choice of method:

- Supervised multi-date classification
- Multi-temporal, **multi-seasonal** approach
- Dense time stacks of Landsat data
- Training examples: **stable** and **changed** classes
- Accuracy assessment



### Urbanization in China - methods

- Which supervised classification algorithm performs the best given complex, dense temporal stacks?
  - Traditional maximum likelihood classifier
  - Boosted decision trees (C4.5)
    - recursive partitioning of training data into successively more homogeneous subsets based on entropy
  - Support vector machines (libSVM)
    - optimal boundaries between classes are defined in feature space using optimization algorithms
- Can feature selection lead to greater map accuracy?
  - 40-55 scenes, uncorrected Landsat data
  - Additional inputs: NDVI, min, max, mean of each band
- How well do multi-temporal approaches work in peri-urban environments given the small size of settlements?
  - support vectors
  - optimal hyperplane

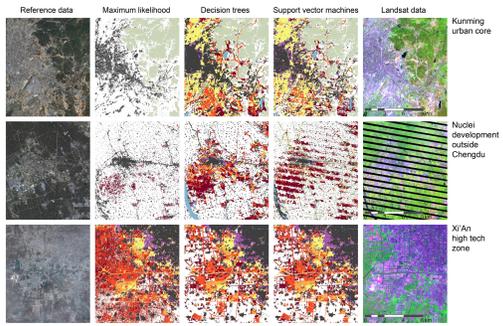


### Results

Accuracy assessment

- Cross-validation approach using tenfold 80/20 splits
- Three algorithms tested
- Decision trees and SVM outperform traditional max likelihood
- Additional data inputs caused small increase in accuracy

	Chengdu	XiAn	Kunming
<b>Overall accuracy results</b>			
<b>a. Maximum likelihood</b>			
all landsat bands	55.3 (1.6)	75.1 (0.6)	68.6 (1.1)
all landsat, all ndvi	53.4 (1.5)	74.9 (1.0)	68.2 (1.7)
all landsat, all metrics	48.9 (2.1)	73.4 (1.5)	67.7 (1.2)
all landsat, ndvi, metrics	51.4 (1.5)	74.5 (0.9)	68.7 (0.9)
all landsat, no sic-off	67.9 (1.4)	<b>80.4 (0.6)</b>	63.6 (2.4)
all ndvi	71.8 (1.3)	84.1 (0.7)	76.8 (0.7)
all ndvi, all metrics	<b>74.6 (1.6)</b>	85.6 (1.2)	71.3 (1.7)
<b>b. Decision trees</b>			
all landsat bands	85.5 (0.7)	93.0 (0.6)	91.1 (0.9)
all landsat, all ndvi	88.1 (0.9)	93.6 (0.7)	91.7 (1.0)
all landsat, all metrics	88.0 (0.6)	93.7 (0.7)	91.9 (0.8)
all landsat, ndvi, metrics	<b>89.8 (0.6)</b>	<b>93.8 (0.6)</b>	92.9 (0.4)
all landsat, no sic-off	75.5 (1.4)	91.8 (0.3)	85.7 (0.7)
all ndvi	77.5 (1.1)	91.0 (0.9)	85.7 (0.7)
all ndvi, all metrics	82.6 (0.8)	93.2 (0.5)	91.2 (0.7)
<b>c. Support vector machines</b>			
all landsat bands	90.0 (1.3)	93.9 (0.8)	92.5 (0.6)
all landsat, all ndvi	89.3 (0.7)	<b>94.1 (0.5)</b>	92.8 (0.8)
all landsat, all metrics	<b>91.0 (0.7)</b>	93.8 (0.6)	92.9 (0.7)
all landsat, ndvi, metrics	90.9 (0.8)	94.1 (0.7)	93.0 (0.8)
all landsat, no sic-off	79.5 (1.0)	91.5 (0.5)	87.8 (0.4)
all ndvi	75.1 (1.1)	90.2 (0.9)	84.4 (1.2)
all ndvi, all metrics	84.0 (0.9)	93.4 (0.7)	90.9 (0.5)



Legend: Land cover maps

- Croplands
- Forests
- Rivers, lakes, water bodies
- Urban and built-up areas
- Cropland to urban 1988-1995
- Cropland to urban 1995-2000
- Cropland to urban 2000-2003
- Cropland to urban 2003-2006
- Cropland to urban 2006-2009

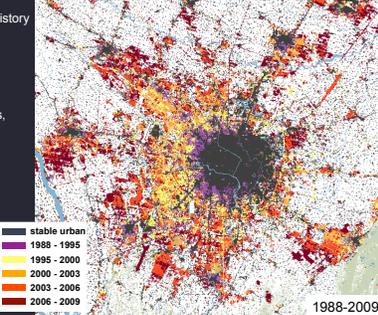
### Chengdu - impacts of the 'Go West' program

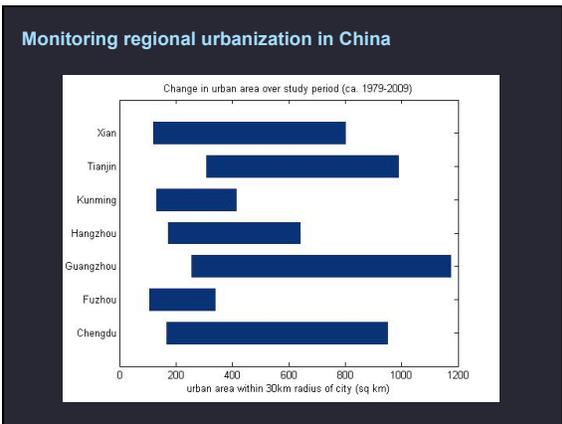
Chengdu, Sichuan Province

- Fertile plain, 2500 yr history
- Industrial center in 1950s, 1960s
- City targeted early for investment: high tech zones, roads, airport

Loss of cropland to urban expansion: > 988 km<sup>2</sup>

400% increase in urban land 1978-2009





### Conclusions - monitoring urban areas with satellite data

Opportunities?

- Synoptic view, repeat coverage
- Comparative analysis is possible
- Routine monitoring of change
- Cross-cutting scales of information
- Scenarios can assist policy development
- Maps can be used across a range of disciplines
- Tremendous opportunity to exploit machine learning algorithms

Challenges?

- Definitions of urban land vary
- Difficulty of characterizing 'mixed' signal within urban areas
- Trade-off between spatial detail and areal coverage
- Time needed to apply semi-automated image processing methods
- Map validation
- Data availability
- Needs of the user community?

### Future research

#### Monitoring land cover and land cover change

- How can data from new sensors and new sources help map urban land cover locally and globally?
- How do we derive **land surface properties** from remotely sensed data that are useful to regional-global environmental models (e.g. urban densities, building heights, surface materials)?
- How do we explain and predict concurrent changes in urban land and population density occurring across the globe?

### Future research

#### Understanding the impacts of urbanization and land cover change

- What are the **environmental impacts** of rapid urban expansion on our most productive agricultural lands at local, regional, and global scales? How will these changes affect food security?
- How can we develop a better conceptual model of how urban land and urban change should be represented within an earth system science framework?
- How can we integrate the **biophysical processes** with the **social aspects** of cities at different temporal and spatial scales to better model global environmental change?
- Is urban growth and urban expansion sustainable, and if so, what would be the form and rate of this type of development?

### Acknowledgments

Jennifer Alix-Garcia (Wisconsin)  
 Mark Friedl (Boston University)  
 Kurt Paulsen (Wisconsin)  
 Karen Seto (Yale)  
 Jianfa Shen (Chinese Univ. Hong Kong)  
 Curtis Woodcock (Boston University)

Graduate students:  
 Chaoyi Chang (Wisconsin)  
 Carly Mertes (Wisconsin)  
 David Potere (Princeton University)  
 Zhiwei Ye (Wisconsin)  
 Na Zhao (Wisconsin)

Global urban datasets and publications available at: <http://www.sage.wisc.edu> or email [aschneider4@wisc.edu](mailto:aschneider4@wisc.edu)

# Thank you