Focal-Test-Based Spatial Decision Tree Learning: A Summary of Results

Zhe Jiang, Shashi Shekhar, Xun Zhou
Joseph Knight, Jennifer Corcoran

Outline

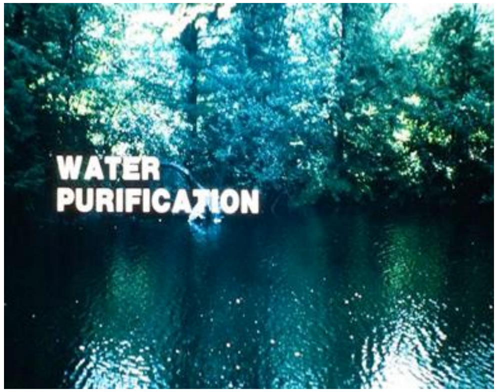
- Motivation
- Problem statement
- Challenges
- Related work
- Proposed approach
- Evaluation
- Conclusion

Motivation

Wetland mapping

Accurately identify wetland areas on land surface via aerial photos.

water purification

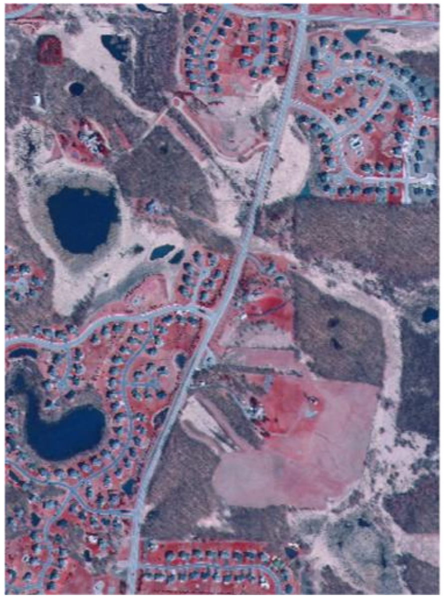


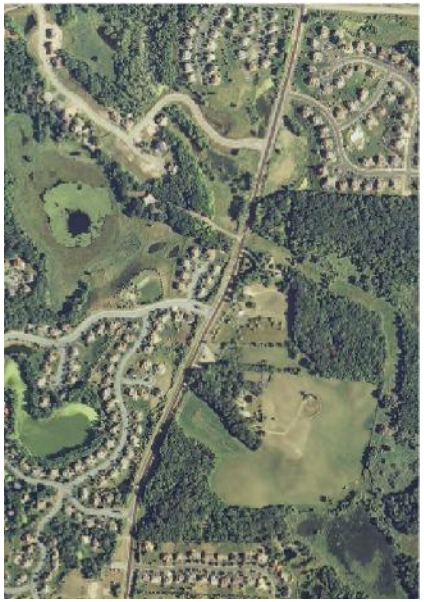
wildlife habitats

carbon cycle, methane

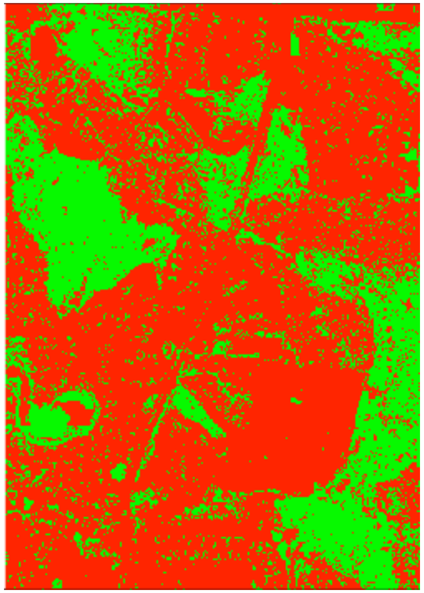
flood control

Motivation


aerial photo 2005


aerial photo 2008


ground truth classes
green: wetland
red: dryland


decision tree prediction

- Decision tree produces lots of salt-and-pepper noise.
- Poor appearance accuracy for decision making agencies.

Basic Concepts

- Spatial neighborhood:
  - range of dependency
  - represented as W-matrix

1	1	1	1	1
1	3	3	3	1
1	3	3	3	1
1	3	3	3	1
1	1	1	1	1

f, neighborhood

1	1	1	1	1
1	-1	-1	-1	1
1	-1	-1	-1	1
1	-1	-1	-1	1
1	1	1	1	1

I(f ≤ 2)

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{pmatrix}$$

w-matrix of 9 pixels 3*3

- Salt-and-pepper noise:
  - in gray scale image: heavy tail impulse noise, a few pixels noisy but extreme (min, max);
  - in raster classification: pixels with classes different from neighbors
- Spatial autocorrelation:
  - degree of dependency of attribute
  - global and local versions
  - I(f ≤ 2), $\Gamma^3(5) = 1$, $\Gamma^5(5) = -0.33$

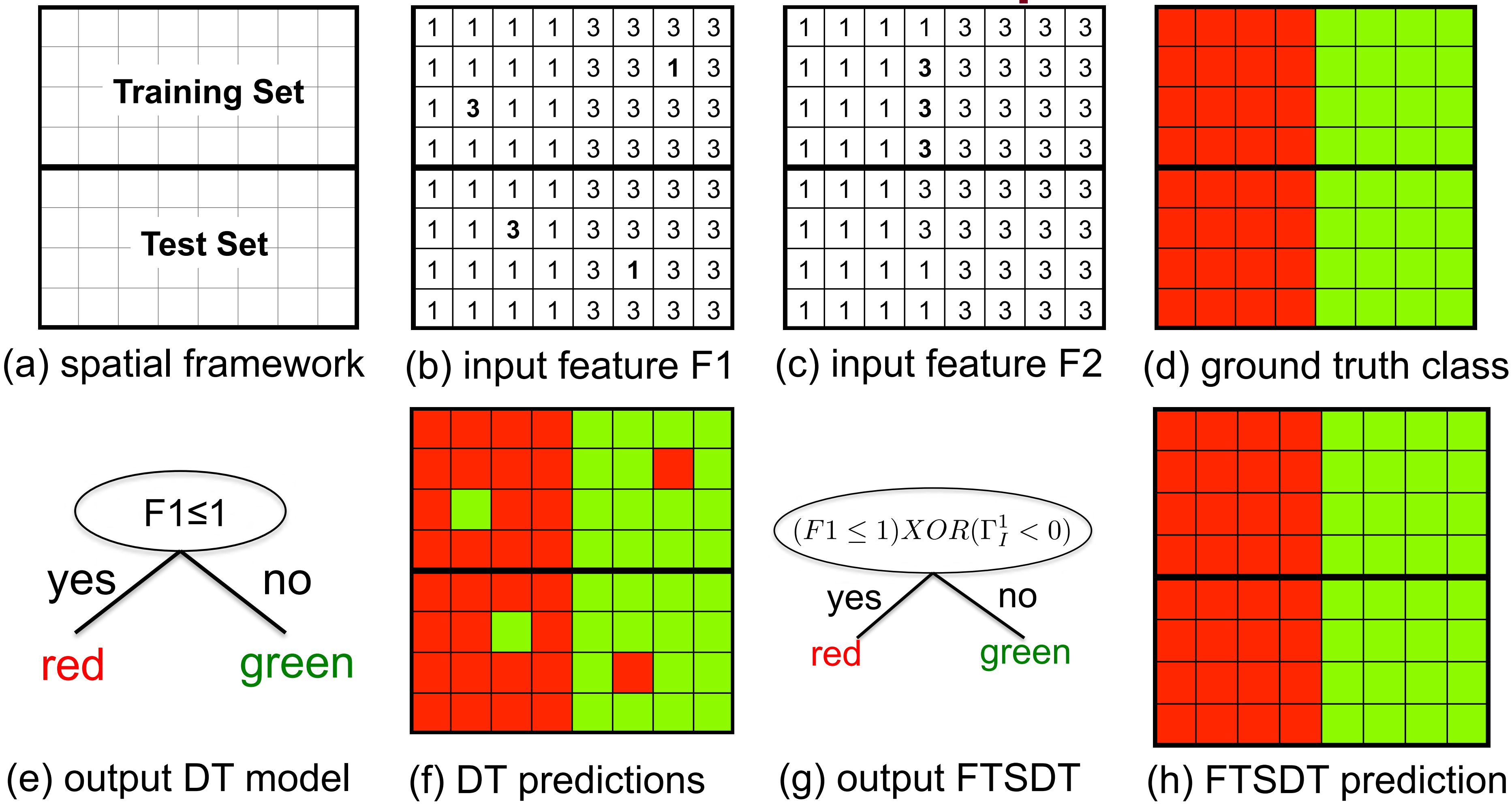
$$\Gamma_I^S = \frac{\sum_{i,j} W_{i,j}^S I(i) I(j)}{\sum_{i,j} W_{i,j}^S}$$

$$\Gamma_I^S(i) = \frac{\sum_{j} W_{i,j}^S I(i) I(j)}{\sum_{j} W_{i,j}^S}$$

Problem Statement

- Given:
  - a raster spatial framework S
  - a spatial neighborhood definition N, and its maximum size s_{max}
  - training and test samples
- Find:
  - a decision tree model from training set
- Objective:
  - minimize classification error and salt-and-pepper noise
- Constraint:
  - training samples form contiguous patches in S
  - spatial autocorrelation exists in class labels

Problem Example



(a) spatial framework

(b) input feature F1

1	1	1	1	3	3	3	3
1	1	1	1	3	3	**1**	3
1	**3**	1	1	3	3	3	3
1	1	1	1	3	3	3	3
1	1	1	1	3	3	3	3
1	1	**3**	1	3	3	3	3
1	1	1	1	3	**1**	3	3
1	1	1	1	3	3	3	3

(c) input feature F2

1	1	1	1	3	3	3	3
1	1	1	**3**	3	3	3	3
1	1	1	**3**	3	3	3	3
1	1	1	**3**	3	3	3	3
1	1	1	3	3	3	3	3
1	1	1	3	3	3	3	3
1	1	1	1	3	3	3	3
1	1	1	1	3	3	3	3

(d) ground truth class

(e) output DT model

(f) DT predictions

(g) output FTSDT

(h) FTSDT prediction

Challenges

- Spatial autocorrelation exists in class labels
  - spatial samples violate i.i.d. assumption
  - testing only local feature leads to salt-and-pepper noise
- Exponential number of candidate trees
  - computationally challenging
  - optimal solution is NP-hard

Related Work & Limitations

Decision Tree

Local-Test-Based Decision Tree

Focal-Test-Based Spatial Decision Tree

Proposed Approach

Traditional Non-spatial Tree (ID3, C4.5, CART)

Spatial Entropy or Information Gain (Jiang 2012, Li 2006, Stojanova 2011, 2012)

- ignoring spatial autocorrelation
- i.i.d. assumption
- salt-and-pepper noise

- measuring spatial autocorrelation in feature selection
- tree nodes testing local feature information
- salt-and-pepper noise when all candidate tests have poor autocorrelation

Contributions

- A novel focal-test-based spatial decision tree (FTSDT)
- Algorithm for FTSDT training and prediction
- Experimental evaluations on real world remote sensing dataset
- A case study with domain interpretations

Proposed Approach: Key Concepts

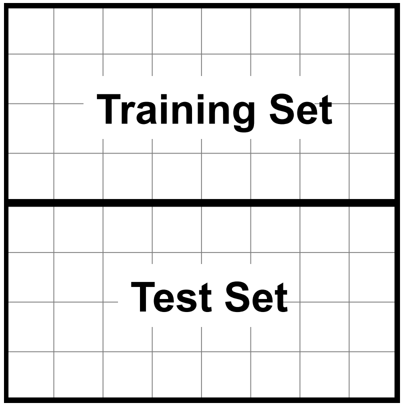
- Local test:
  - test local feature properties, e.g., F ≤ δ
- Focal test:
  - tree node testing not only feature, but also local spatial autocorrelation information
  - for example, $(f \leq \delta) \oplus (\Gamma_I^S < 0)$
- Focal-test-based spatial decision tree (FTSDT)
  - tree node with focal tests
  - sample tree traversal direction based on not only local but also focal properties

Proposed Approach: An Algorithm

- **FTSDT-Learner** (samples, W-matrix, min_size, max_neighbor_size)
- **Output**: root node of an FTSDT
  - **If** # of samples < min_size or same class
    - **then** return a Leaf node.
  - **For each** candidate feature f
    - **For each** candidate threshold δ
      - **For each** neighborhood size s
        - » Split samples by focal test $(f \leq \delta) \oplus (\Gamma_I^S < 0)$
        - » Compute information gain (IG)
  - Find f, δ, s that maximizes IG
  - Create an internal node with test: $(f \leq \delta) \oplus (\Gamma_I^S < 0)$
  - Split samples and **recursively** call function on each subset

Execution Trace: Inputs

Same input as previous problem example



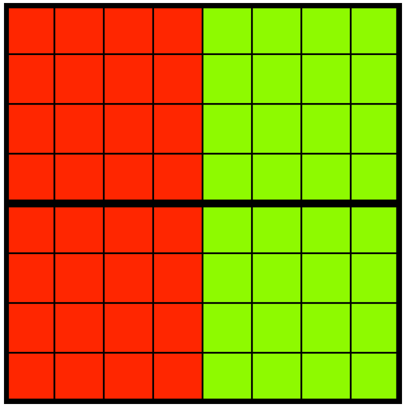
(a) spatial framework

1	1	1	1	3	3	3	3
1	1	1	1	3	3	**1**	3
1	**3**	1	1	3	3	3	3
1	1	1	1	3	3	3	3
1	1	1	1	3	3	3	3
1	1	**3**	1	3	3	3	3
1	1	1	1	3	**1**	3	3
1	1	1	1	3	3	3	3

(b) input feature F1

1	1	1	1	3	3	3	3
1	1	1	**3**	3	3	3	3
1	1	1	**3**	3	3	3	3
1	1	1	**3**	3	3	3	3
1	1	1	3	3	3	3	3
1	1	1	3	3	3	3	3
1	1	1	1	3	3	3	3
1	1	1	1	3	3	3	3

(c) input feature F2



(d) ground truth class

Trace: Local & Focal Tests on F2

1	1	1	1	3	3	3	3
1	1	1	**3**	3	3	3	3
1	1	1	**3**	3	3	3	3
1	1	1	**3**	3	3	3	3

(a) feature F2

1	1	1	1	-1	-1	-1	-1
1	1	1	**-1**	-1	-1	-1	-1
1	1	1	**-1**	-1	-1	-1	-1
1	1	1	**-1**	-1	-1	-1	-1

(b) indicator I(F2≤1)

T	T	T	T	F	F	F	F
T	T	T	**F**	F	F	F	F
T	T	T	**F**	F	F	F	F
T	T	T	**F**	F	F	F	F

(c) local test result: F2≤1

1	1	1	0.2	0.6	1	1	1
1	1	0	0	0.8	1	1	1
1	1	0.3	0.3	1	1	1	1
1	1	0.2	0.2	1	1	1	1

(d) focal function Γ_I

F	F	F	F	F	F	F	F
F	F	F	F	F	F	F	F
F	F	F	F	F	F	F	F
F	F	F	F	F	F	F	F

(e) focal test: $\Gamma_I<0$

T	T	T	T	F	F	F	F
T	T	T	**F**	F	F	F	F
T	T	T	**F**	F	F	F	F
T	T	T	**F**	F	F	F	F

(f) final focal test: F2≤1 xor $\Gamma_I<0$

Information gain: (1): IG(F2 ≤ 1) = 0.63; (2): IG(F1 ≤ 1 xor Γ_I < 0) = 0.63

Trace: Local & Focal Tests on F1

1	1	1	1	3	3	3	3
1	1	1	1	3	3	**1**	3
1	**3**	1	1	3	3	3	3
1	1	1	1	3	3	3	3

(a) feature F1

1	1	1	1	-1	-1	-1	-1
1	1	1	1	-1	-1	**1**	-1
1	**-1**	1	1	-1	-1	-1	-1
1	1	1	1	-1	-1	-1	-1

(b) indicator I(F1≤1)

T	T	T	T	F	F	F	F
T	T	T	T	F	F	**T**	F
T	**F**	T	T	F	F	F	F
T	T	T	T	F	F	F	F

(c) local test result: F1≤1

1	1	1	0.2	0.2	0.6	0.6	0.3
0.6	0.8	0.8	0.3	0.3	0.8	**-1**	0.6
0.6	**-1**	0.8	0.3	0.3	0.8	0.8	0.6
0.3	0.6	0.6	0.2	0.2	1	1	1

(d) focal function Γ_I

F	F	F	F	F	F	F	F
F	F	F	F	F	F	**T**	F
F	**T**	F	F	F	F	F	F
F	F	F	F	F	F	F	F

(e) focal test: $\Gamma_I<0$

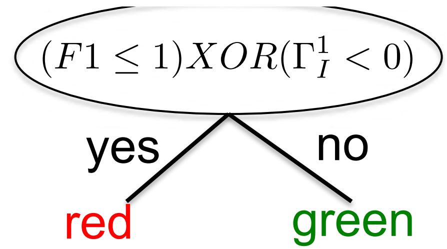
T	T	T	T	F	F	F	F
T	T	T	T	F	F	F	F
T	T	T	T	F	F	F	F
T	T	T	T	F	F	F	F

(f) final focal test: F1≤1 xor $\Gamma_I<0$

Information gain: (1): IG(F1 ≤ 1) = 0.66; (2): IG(F1 ≤ 1 xor Γ_I < 0) = 1

Trace: Output FTSDT & Predictions

candidate	F1 local	F1 focal	F2 local	F2 focal
info. gain	0.66	1	0.63	0.63



(a): output FTSDT

1	1	1	1	3	3	3	3
1	**3**	1	1	3	3	3	3
1	1	1	1	3	**1**	3	3
1	1	1	1	3	3	3	3

(b): F1 in test area

1	1	1	1	-1	-1	-1	-1
1	**-1**	1	1	-1	-1	-1	-1
1	1	1	1	-1	**1**	-1	-1
1	1	1	1	-1	-1	-1	-1

(c): indicator I(F1≤1)

1	1	1	0.2	0.2	0.6	0.6	0.3
0.6	**-1**	0.8	0.3	0.3	0.8	0.8	0.6
0.6	0.8	0.8	0.3	0.3	**-1**	0.8	0.6
0.3	0.6	0.6	0.2	0.2	1	1	1

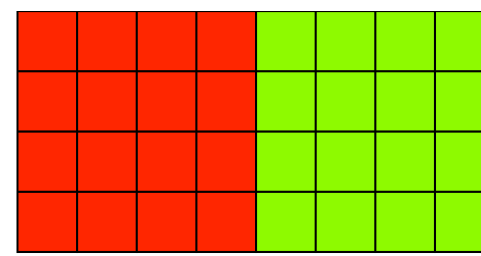
(d): focal function Γ_I

F	F	F	F	F	F	F	F
F	**T**	F	F	F	F	F	F
F	F	F	F	F	**T**	F	F
F	F	F	F	F	F	F	F

(e): focal test Γ_I<0

T	T	T	T	F	F	F	F
T	T	T	T	F	F	F	F
T	T	T	T	F	F	F	F
T	T	T	T	F	F	F	F

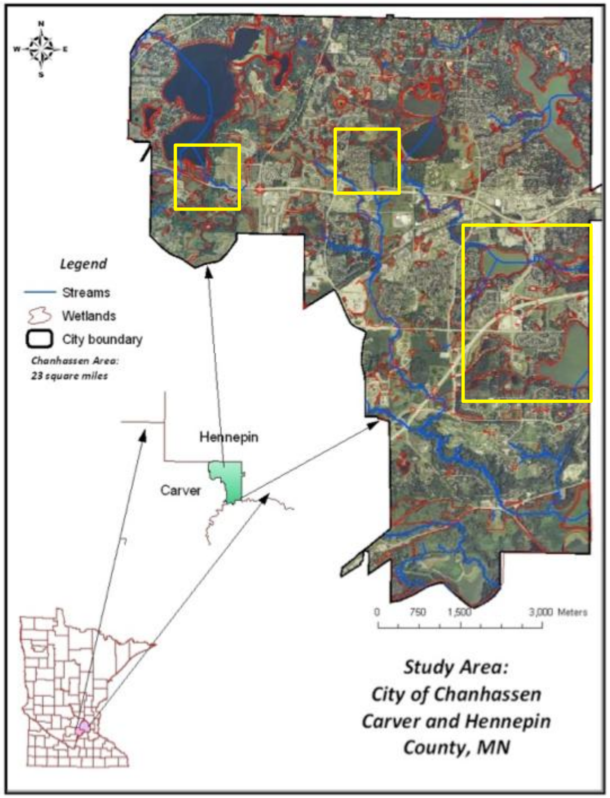
(f):final focal test F1≤1 xor Γ_I<0



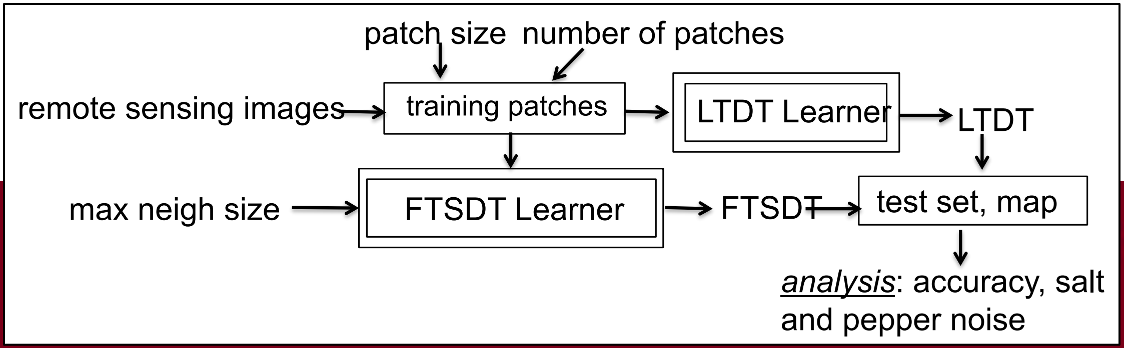
(g): final prediction on test area

Evaluation: Experiment Design

- Questions to answer:
  - FTSDT v.s. DT classification accuracy
  - FTSDT v.s. DT salt-and-pepper noise
  - FTSDT sensitive to parameters
- Dataset:
  - features: remote sensing imagery (RGB, NIR, NDVI) in 2003, 2005, 2008 (3m*3m)
  - classes: wetland, dryland
  - training set: contiguous circular clusters
  - test set: remaining pixels of a scene



City of Chanhassen, MN



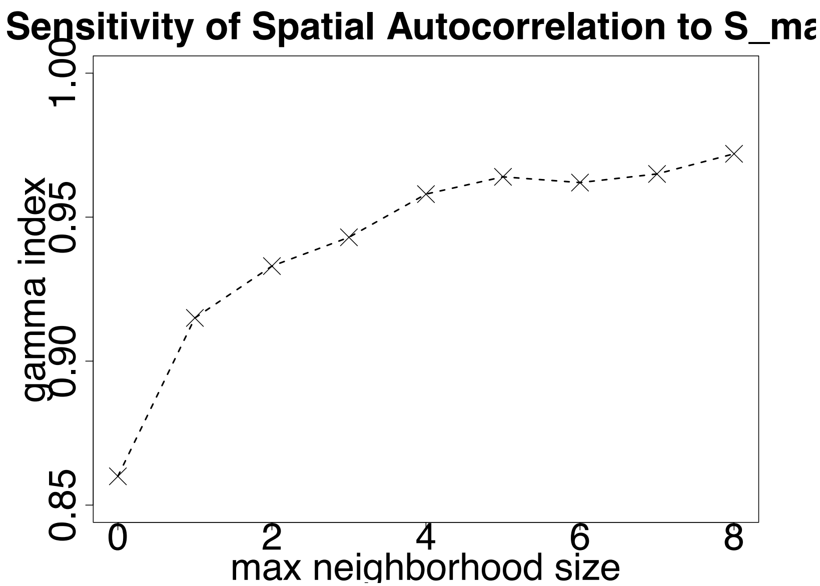
Evaluation: Experiment Results

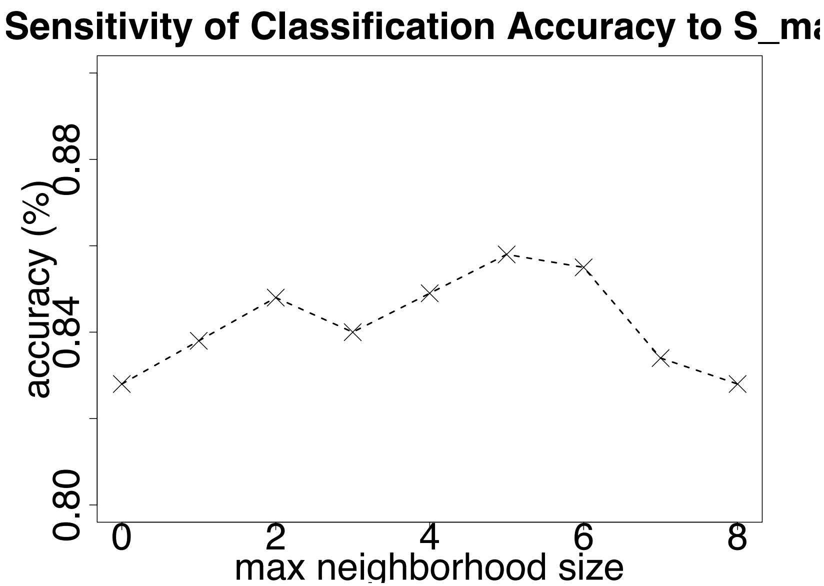
Model	Confusion Matrix		Prec.	Recall	Acc.	Node #	Gamma
DT	383,044	**29,749**	0.84	0.69	**0.84**	21	**0.92**
	71,134	161,192					
FTSDT	384,631	**28,162**	0.87	0.80	**0.88**	21	**0.98**
	46,673	185,653					

Model	Confusion Matrix		Prec.	Recall	Acc.	Node #	Gamma
DT	99,732	**18,234**	0.61	0.77	**0.83**	55	**0.86**
	8,837	28,935					
FTSDT	103,913	**14,044**	0.68	0.79	**0.86**	35	**0.96**
	8,008	29,764					

Evaluation: Parameter Sensitivity

Setup: minimum node size: 50; training samples: 2nd scene;

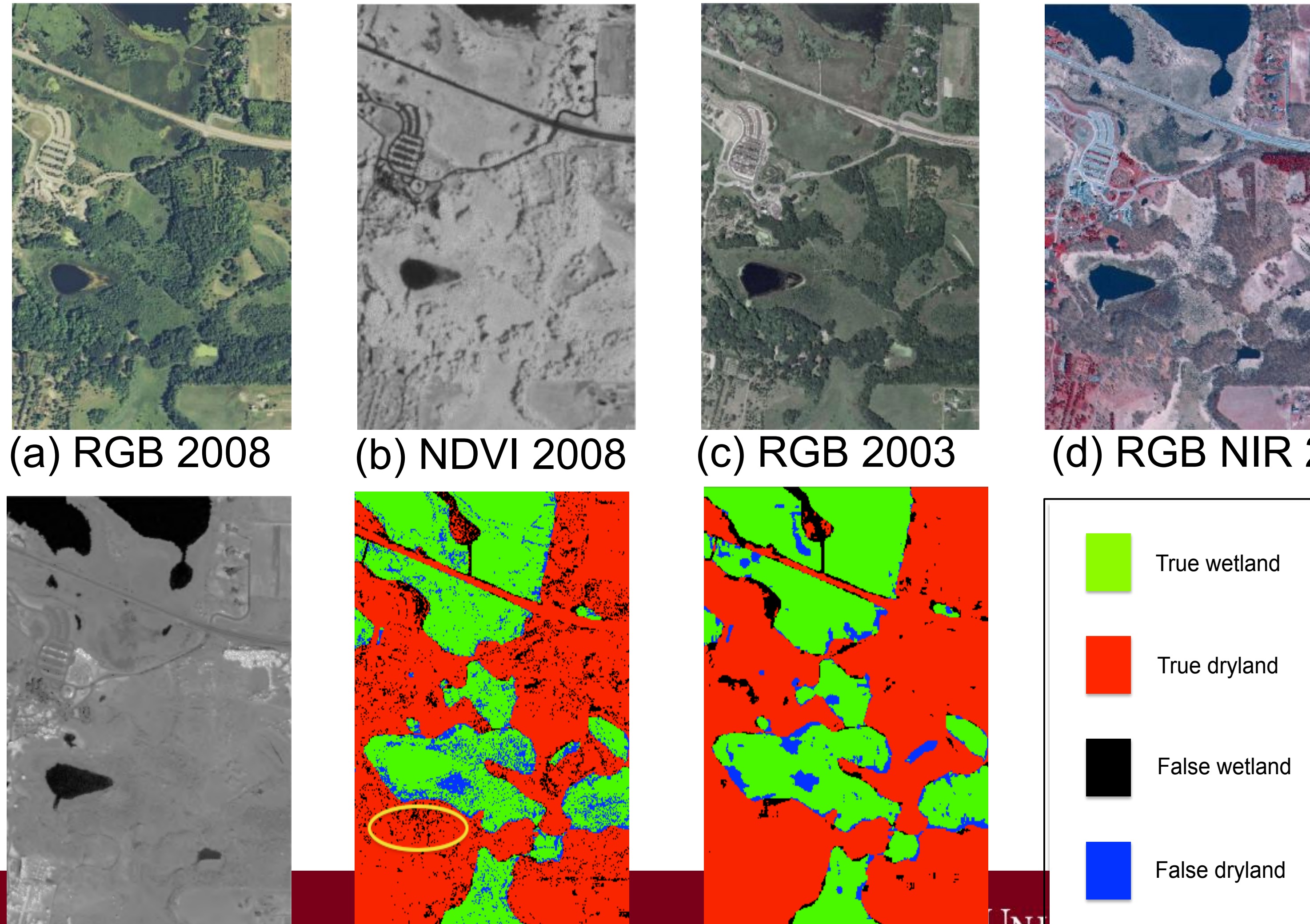




Trend:

- Spatial autocorrelation first quickly increases then gets stable
- Accuracy first increases and then decrease

Evaluation: A Case Study



(a) RGB 2008 (b) NDVI 2008 (c) RGB 2003 (d) RGB NIR 2005

(e) NDVI 2005 (f) DT prediction (g) FTSDT prediction

Conclusion & Future Work

- Proposed an FTSDT model
- Designed learning FTSDT algorithm
- Evaluated FTSDT on real world dataset
  - improve classification accuracy
  - reduce salt-and-pepper noise
- In future work, we will
  - investigate computational characteristics of FTSDT learning algorithms
  - design computational efficient algorithms
  - investigate ensembles of spatial tree