ESA - MOST Dragon 2 Programme
2011 DRAGON 2 SYMPOSIUM
中国科技部-欧洲空间局合作 "龙计划"二期
"龙计划"二期2011年学术研讨会

Comparison of different algorithms for urban land-use classification

Congcong Li[1], Peng Gong[2], Lei Wang[3], Jie Wang[3]

1. Beijing Normal University, CHINA;
2. Center for Earth System Science/Institute for Global Change studies, Tsinghua University, CHINA;
3. Jointly sponsored by Institute of Remote Sensing Applications, CHINA

Outline

- Introduction
- Material and Methods
- Training Sites
- Classification Results
- Integration of GLCM-features
- Classification Results (GLCM)
- Conclusions

Introduction

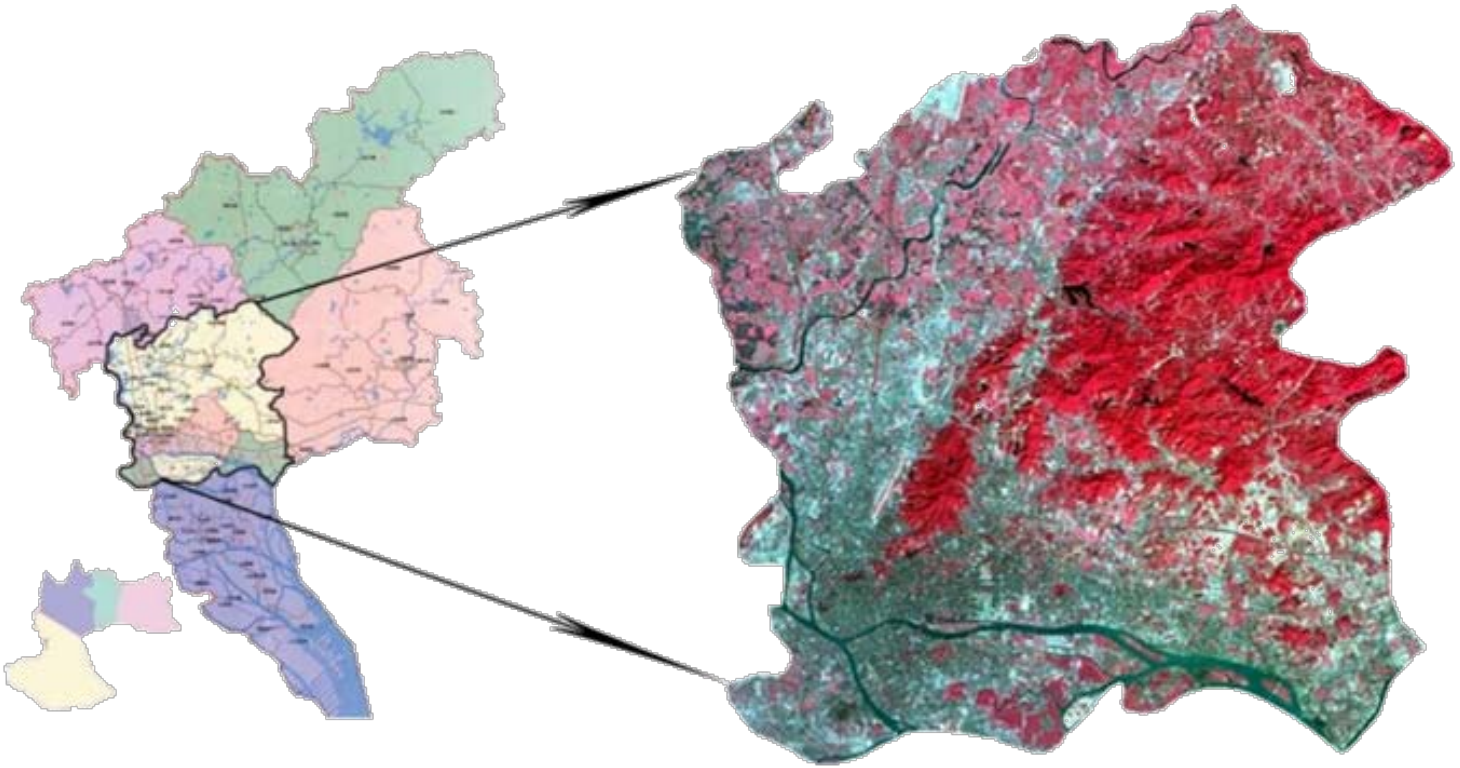
Land cover/land use classification is the basis for a large number of environmental planning and management activities.

Although many different classifiers are emerging, a thorough comparison with the same set of data and training samples is rare.

With the development of China's satellite technology, BJ-1, HJ-1 and CBERS data with 20-30m resolution became available.

A comparison of multiple Chinese EO data using different classification algorithms.

Material and Methods

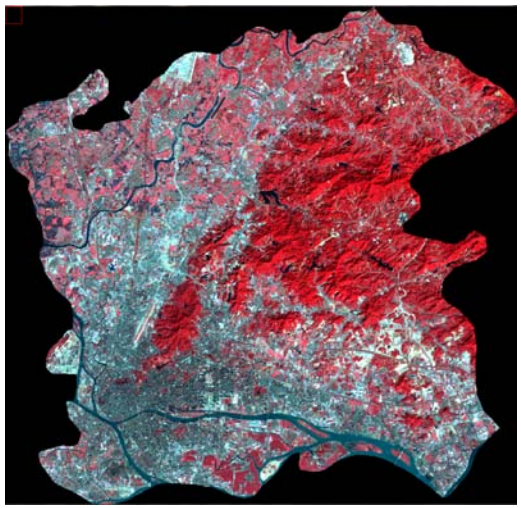


The study area is the old Guangzhou city area (23° 2’-23° 25’N, 113° 8’-113° 35’E). The area is composed of variety of landforms ranging from plains to hills and mountains. 8 land use types could be identified for this study.

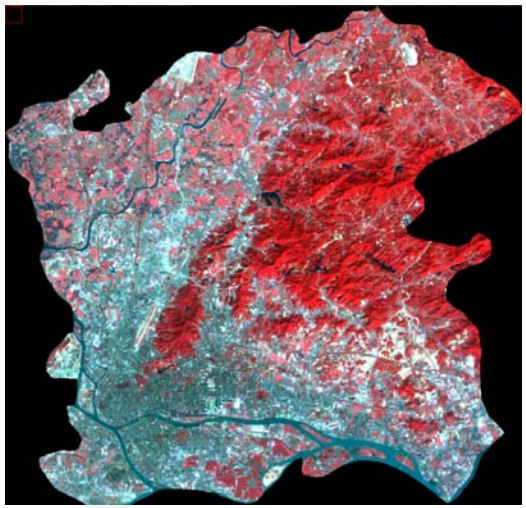
Material and Methods

	Cbers-02B	Beijing-1	HJ-1A	TM
Acquisition date	2009-1-5	2009-1-14	2009-1-2	2009-1-2
Description	B1:0.45-0.52(μm) B2:0.52-0.59(μm) B3:0.63-0.69(μm) B4:0.77-0.89(μm) B5:0.51-0.73(μm)	B1:0.52-0.62(μm) B2:0.63-0.69(μm) B3:0.76-0.90(μm)	B1:0.43-0.52(μm) B2:0.52-0.60(μm) B3:0.63-0.69(μm) B4:0.76-0.90(μm)	B1:0.45-0.52(μm) B2:0.52-0.60(μm) B3:0.63-0.69(μm) B4:0.76-0.90(μm) B5:1.55-1.75(μm) B6:10.4-12.5(μm) B7:2.08-2.35(μm)
Sun Elevation	40.827801°	35.1605°	42.573°	36.5679°
Sun Azimuth	156.694494°	141.494°	158.361°	147.2278°
Spatial resolution	19.5m	30.0m	32.0m	30.0m

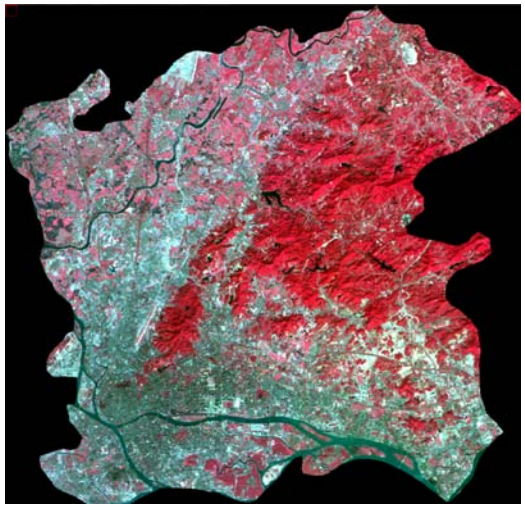
Material and Methods



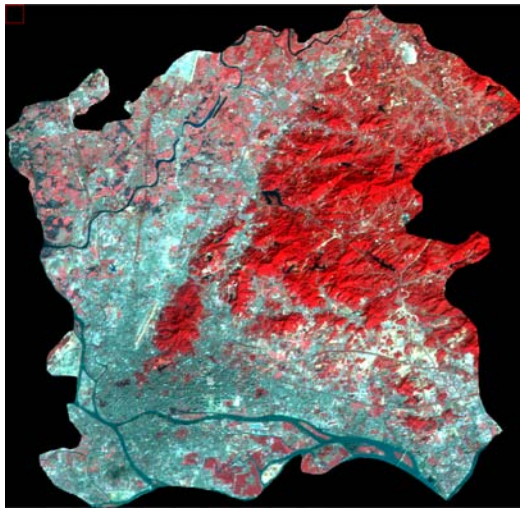
TM



HJ-1



CBERS-02



Beijing-1

Material and Methods

Land-use types	Description
Residential area	Residential areas where trees, lawns, driveways and roof tops dominate/regular, homogeneous patterns
Industrial/commercial	Roof tops of large building with little vegetation
Idle land	No vigorous growth of vegetation
Farmland	Fields with vegetation or crops
Urban forest	Large area of trees
Cleared land/Land under construction	Denuded of vegetation and top soil showing evenly high reflectance in every band
Orchard	Large area of planted fruit trees
Water	Several relatively large water surfaces such as reservoirs and ponds

Material and Methods

Algorithms tested	Abbreviation
1. Maximum likelihood	ML
2. Radial basis function network	RBFN
3. Logistic	L
4. K-nearest neighbours	KNN
5. Support vector machine	SVM
6. C4.5	C4.5
7. CART	CART
8. QUEST	QUEST
9. Logistic model trees	LMT
10. Random Forest	RF
11. Bagging	Bagging
12. AdaboostM1	Adaboost
13. Vote	Vote

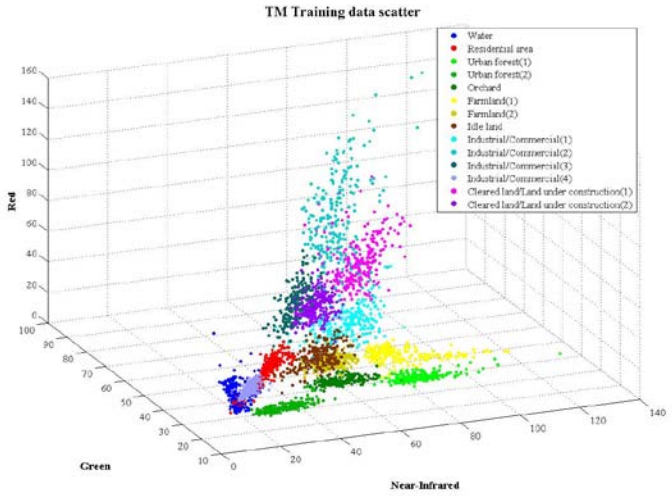
Training Sites

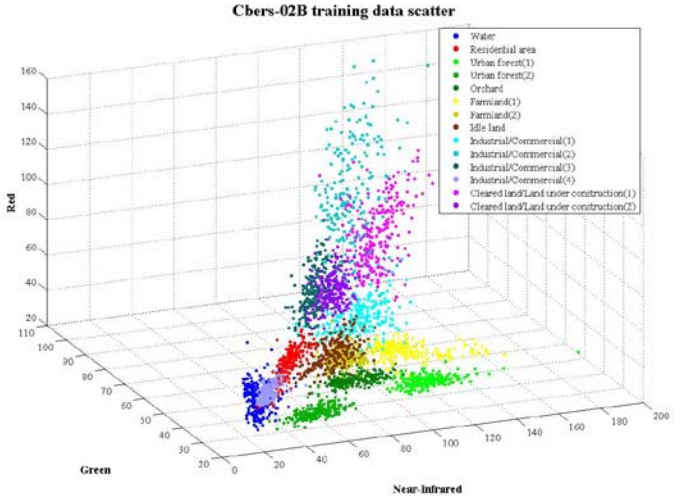
The selection of **training sites** is critical for land use classification.

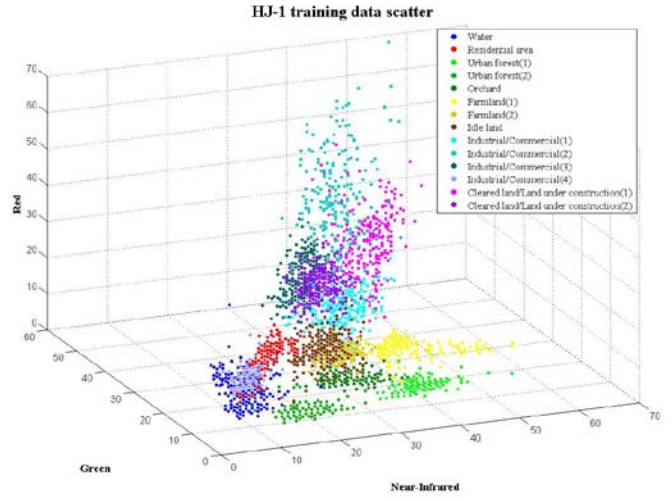
In this research different sizes of training sites were used:

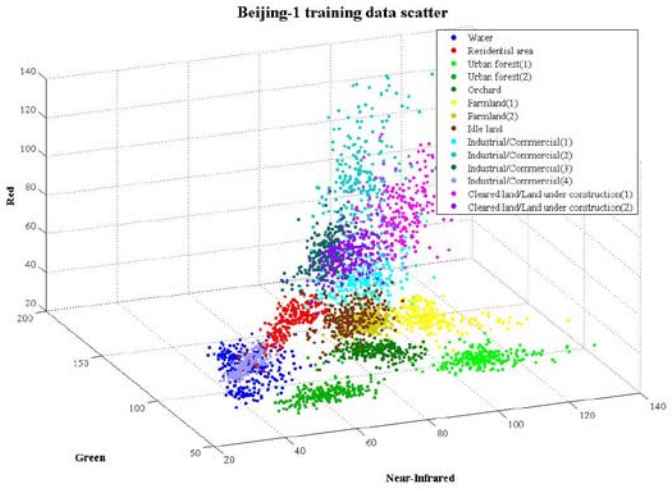
- For TM, HJ-1, and Beijing-1 data, **240 pixels** per sub-classes (14 sub-classes) were used for classification. Then the number of training pixels were subsequently randomly reduced to 220, 200, 180, 160 and 140 pixels.
- For Cbers-02B data, **260 pixels** per sub-classes (14 sub-classes) were selected. Due to it's relatively spoken higher resolution, more pixels can be selected for the training sites.

Training Sites

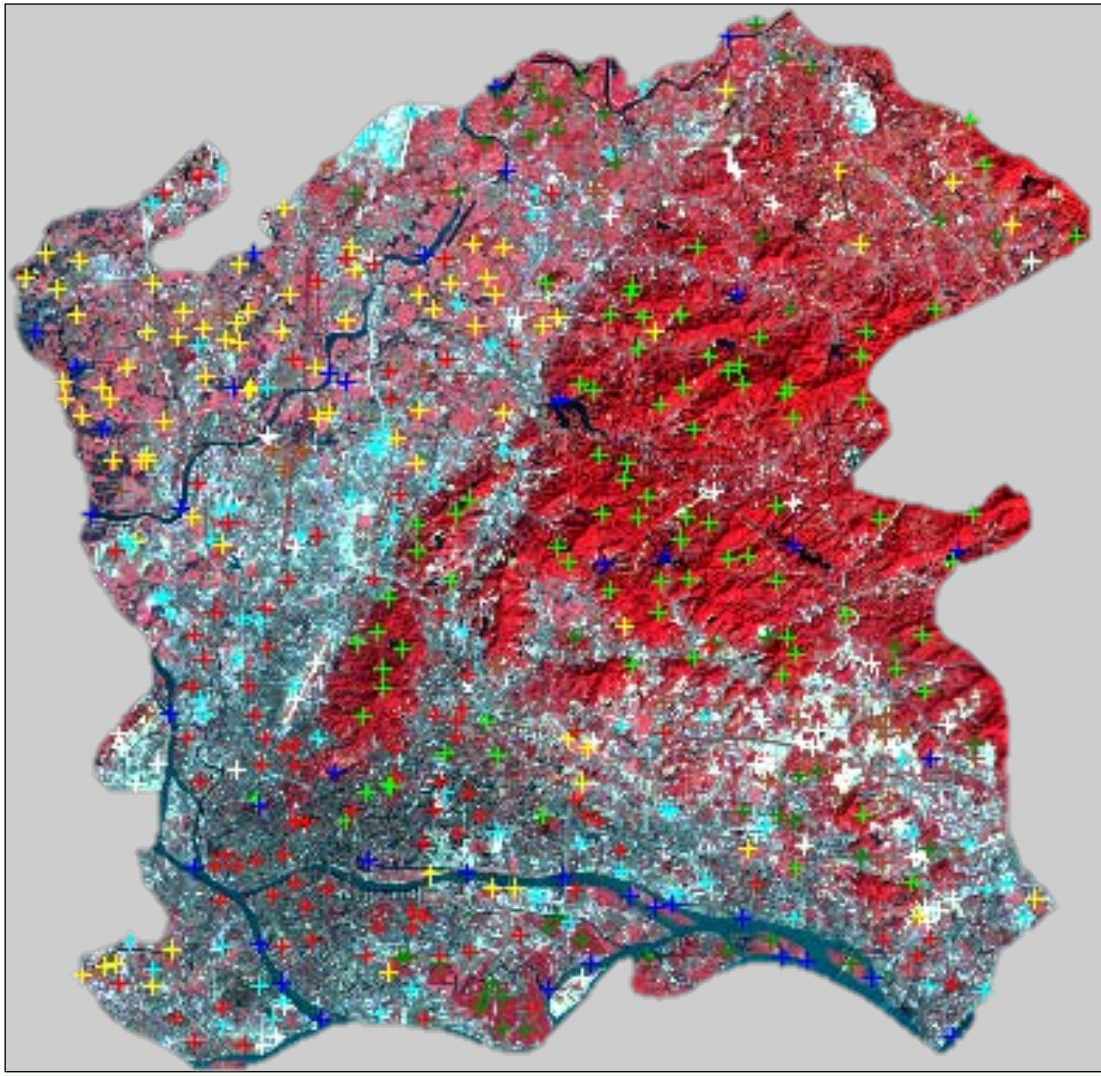








Training Sites

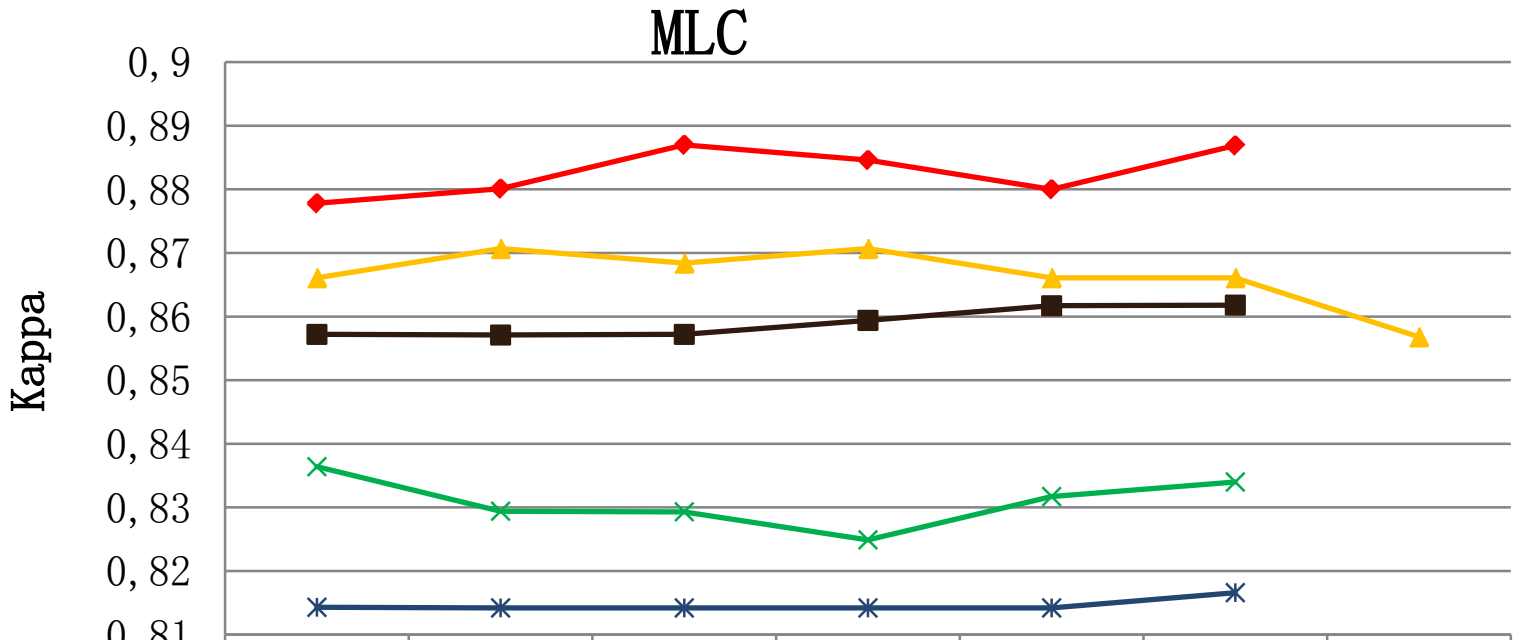


500 pixels were generated as test data, 138 of which were identified in-situ (April/December 2009 and June 2010). Each land use type comprises more than 40 test pixels.

Legend

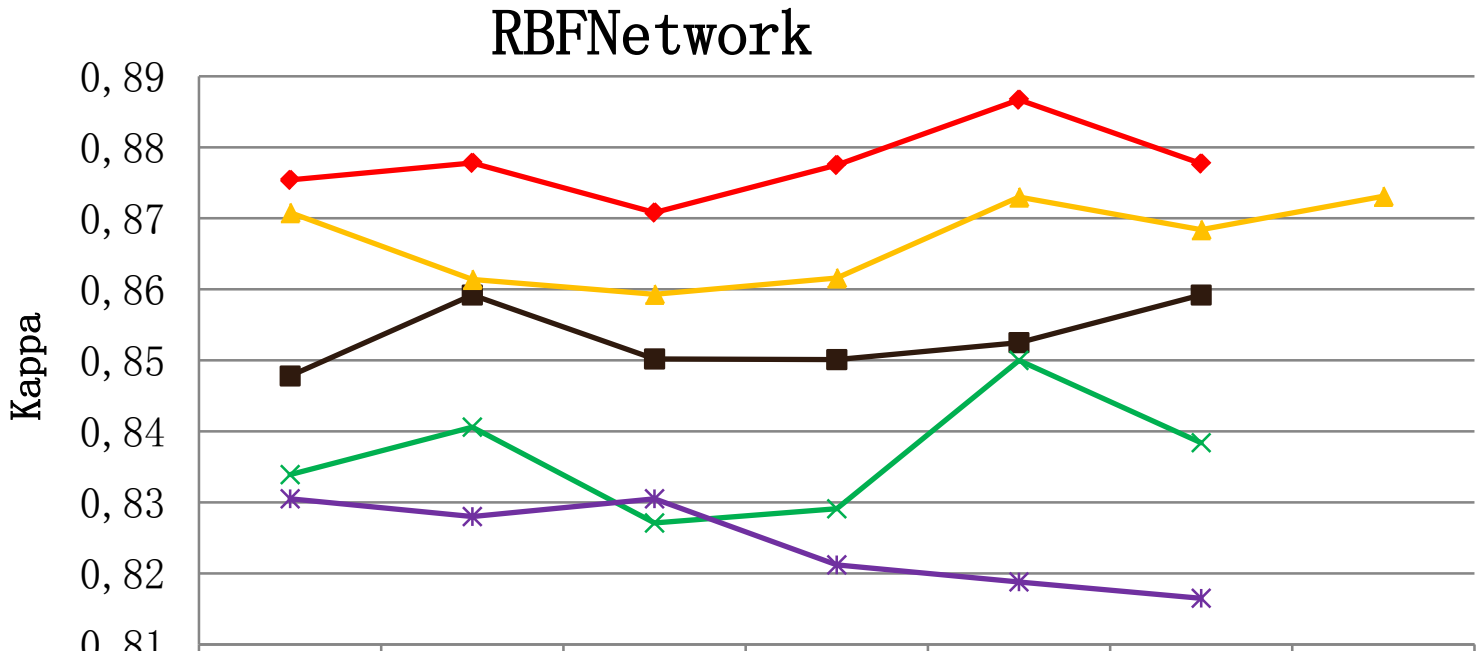
- Water
- Urban residential
- Urban forest
- Orchard
- farmland
- Industrial/Commerical
- Clear land/Land under construction
- Idle land

Classification Results



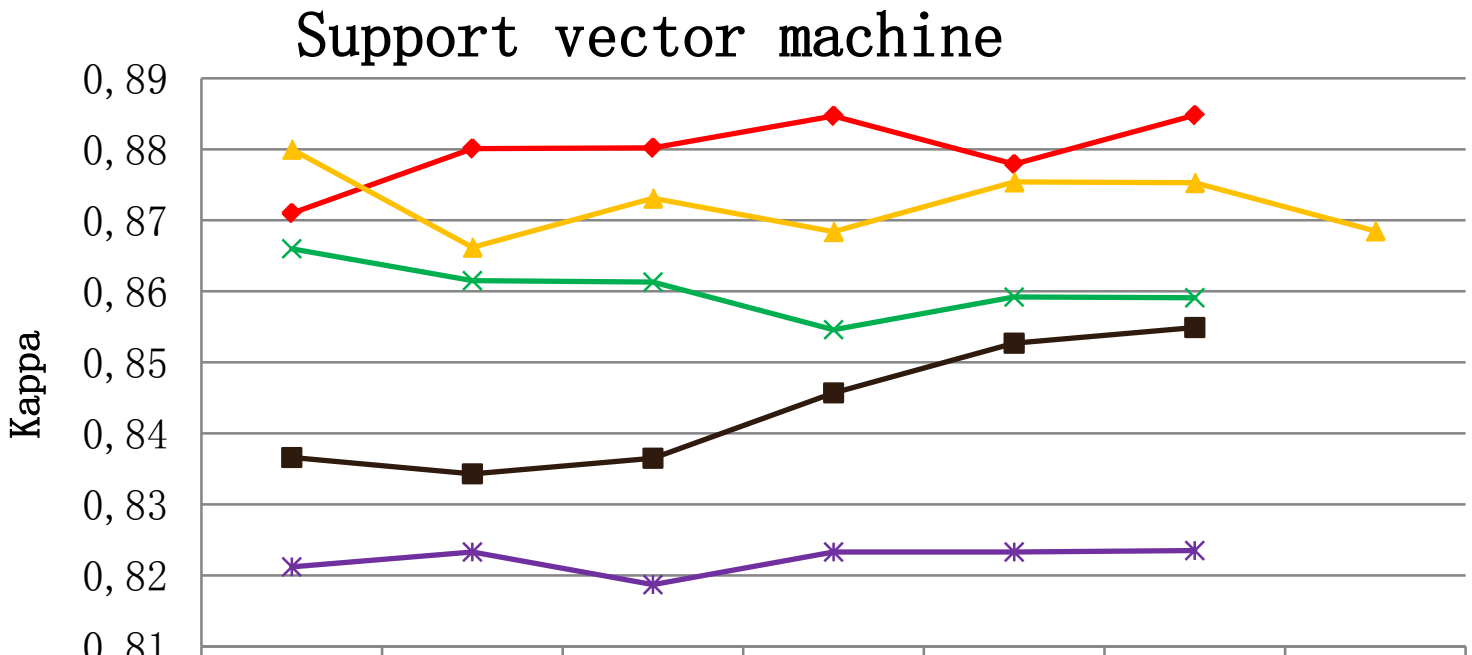
	140roi	160roi	180roi	200roi	220roi	240roi	260roi
TM(6个波段)	0,8778	0,8801	0,887	0,8846	0,88	0,8869	
TM(4个波段)	0,8572	0,8571	0,8572	0,8594	0,8617	0,8618	
Cbers-02B	0,8661	0,8707	0,8684	0,8707	0,8661	0,8661	0,8568
HJ-1A	0,8364	0,8294	0,8293	0,8249	0,8317	0,834	
Beijing-1	0,8143	0,8142	0,8142	0,8142	0,8142	0,8166	

Classification Results



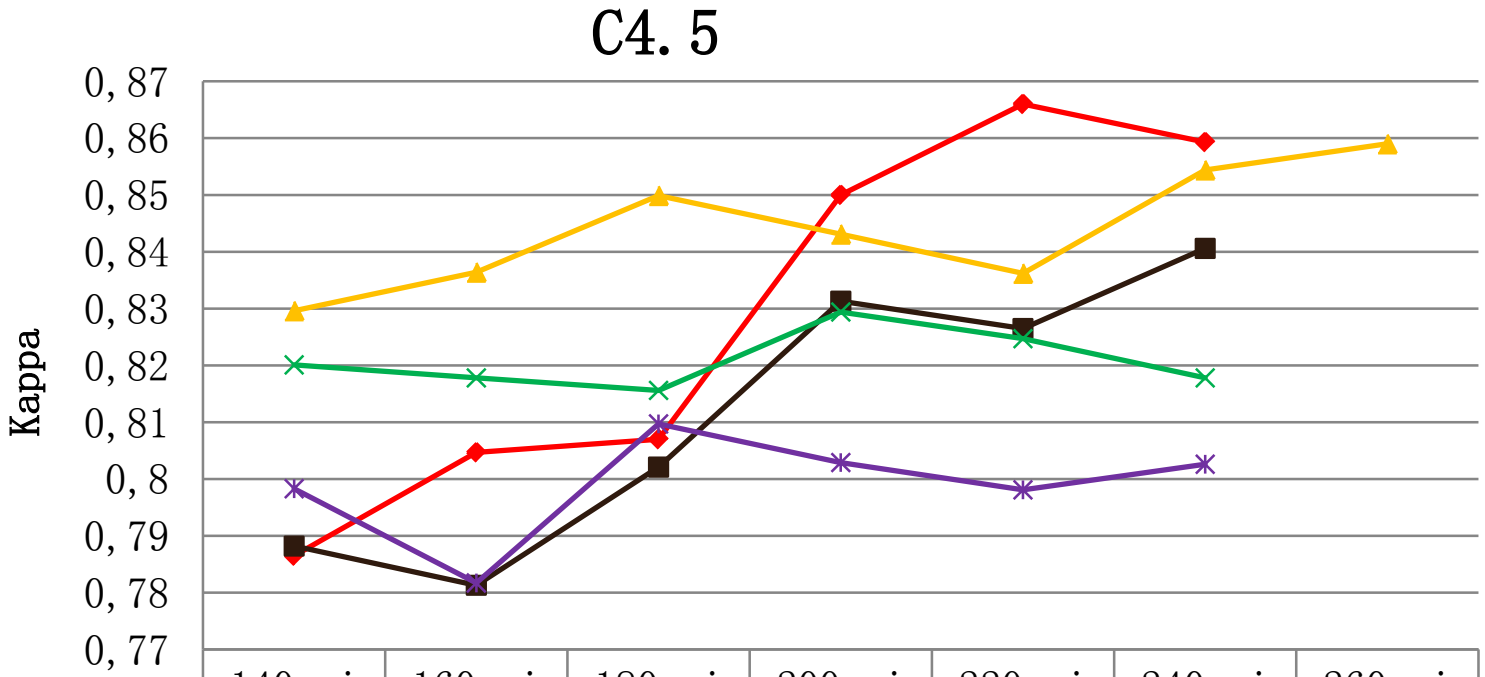
	140roi	160roi	180roi	200roi	220roi	240roi	260roi
TM(6个波段)	0,8754	0,8778	0,8708	0,8775	0,8867	0,8777	
TM(4个波段)	0,8478	0,8592	0,8502	0,8501	0,8525	0,8592	
Cbers-02B	0,8708	0,8614	0,8593	0,8616	0,873	0,8684	0,8731
HJ-1A	0,8339	0,8406	0,8271	0,8291	0,85	0,8384	
Beijing-1	0,8305	0,828	0,8305	0,8212	0,8188	0,8165	

Classification Results



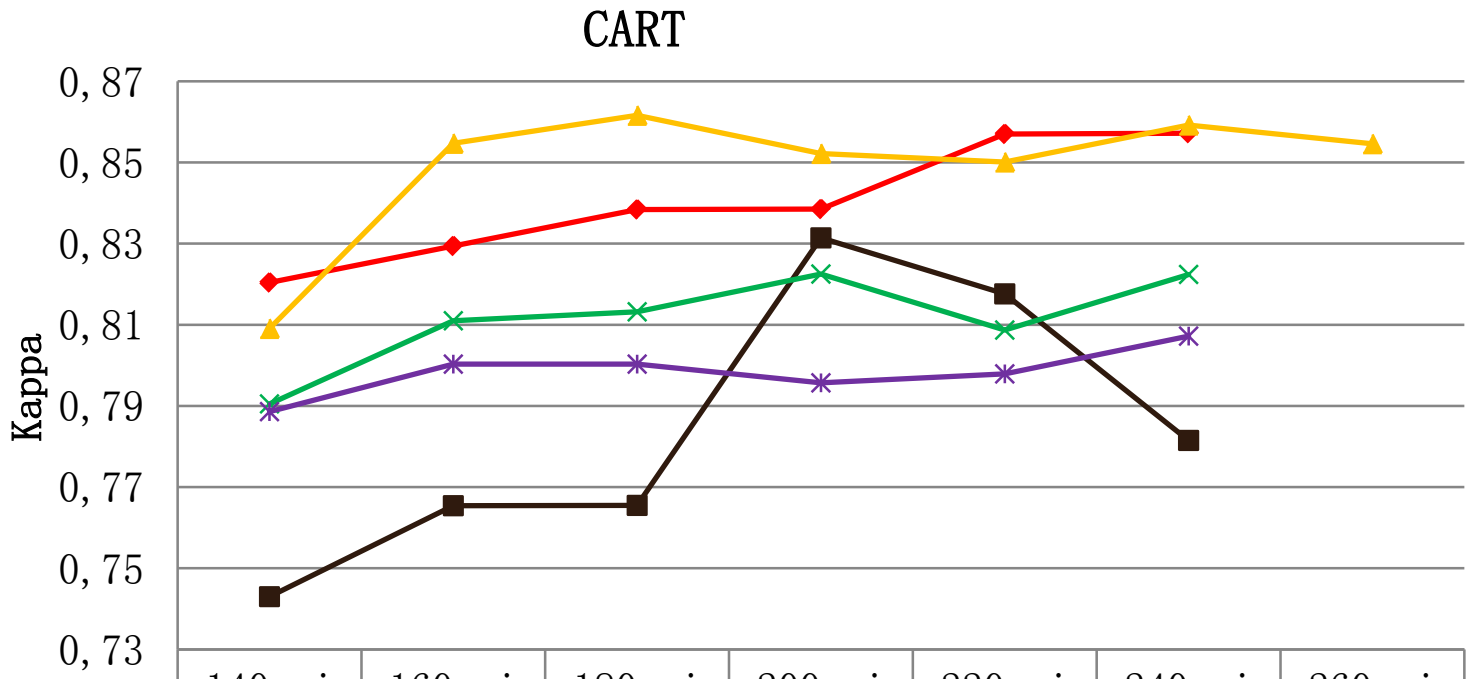
	140roi	160roi	180roi	200roi	220roi	240roi	260roi
TM(6个波段)	0,871	0,8801	0,8802	0,8847	0,8779	0,8848	
TM(4个波段)	0,8366	0,8343	0,8365	0,8457	0,8527	0,8549	
Cbers-02B	0,88	0,8662	0,8731	0,8684	0,8754	0,8753	0,8685
HJ-1A	0,866	0,8615	0,8613	0,8546	0,8592	0,8591	
Beijing-1	0,8212	0,8233	0,8187	0,8233	0,8233	0,8235	

Classification Results



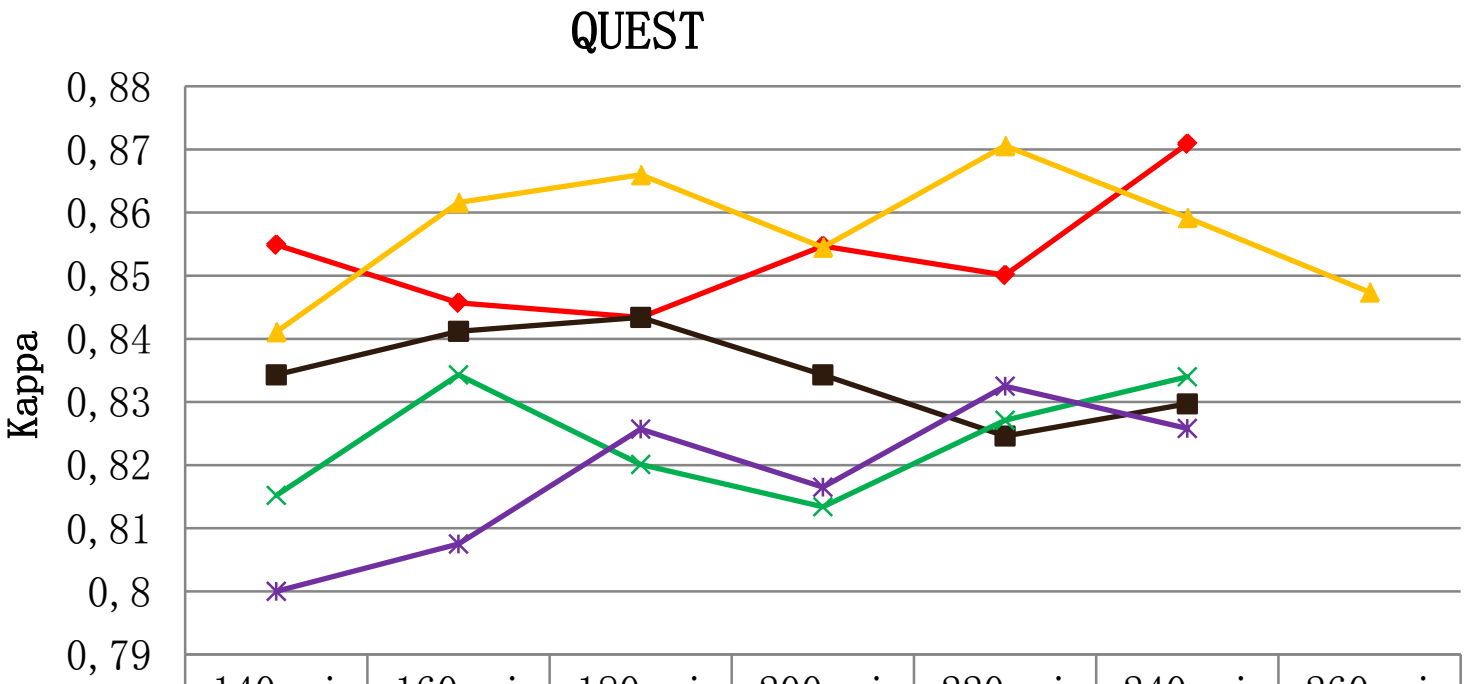
	140roi	160roi	180roi	200roi	220roi	240roi	260roi
TM(6个波段)	0,7865	0,8047	0,807	0,85	0,866	0,8593	
TM(4个波段)	0,7882	0,7813	0,8021	0,8313	0,8265	0,8406	
Cbers-02B	0,8296	0,8364	0,8499	0,8431	0,8362	0,8544	0,859
HJ-1A	0,8201	0,8178	0,8156	0,8294	0,8247	0,8178	
Beijing-1	0,7983	0,7817	0,8097	0,8029	0,7981	0,8026	

Classification Results



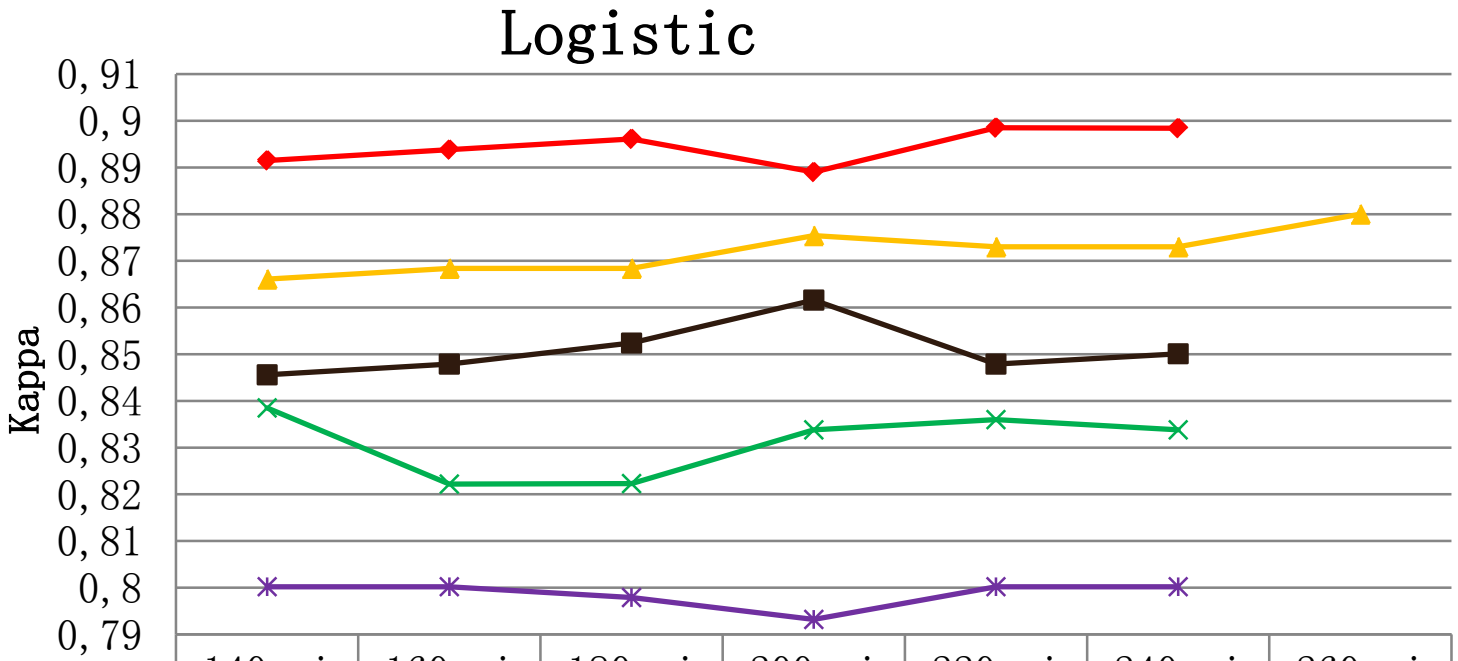
	140roi	160roi	180roi	200roi	220roi	240roi	260roi
TM(6个波段)	0, 8204	0, 8294	0, 8384	0, 8385	0, 857	0, 8572	
TM(4个波段)	0, 743	0, 7654	0, 7655	0, 8314	0, 8176	0, 7815	
Cbers-02B	0, 8091	0, 8547	0, 8616	0, 8522	0, 8501	0, 8592	0, 8546
HJ-1A	0, 7905	0, 811	0, 8132	0, 8225	0, 8087	0, 8224	
Beijing-1	0, 7886	0, 8003	0, 8003	0, 7957	0, 7979	0, 8072	

Classification Results



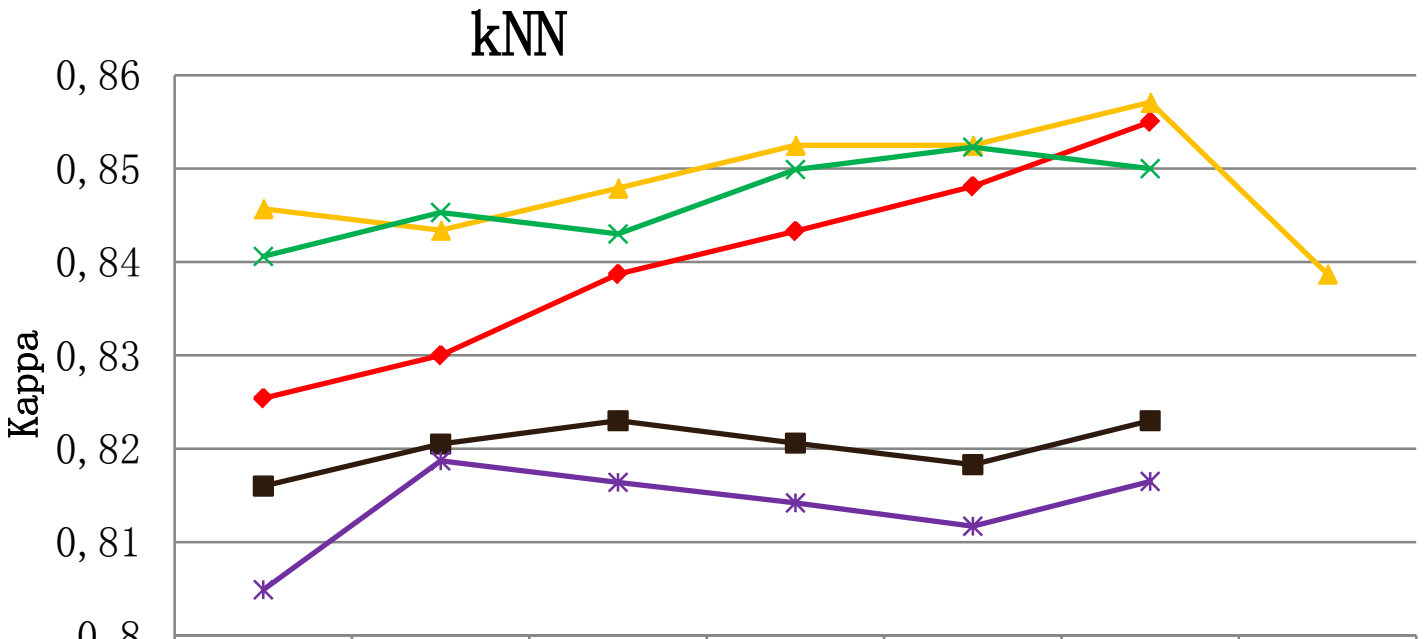
	140roi	160roi	180roi	200roi	220roi	240roi	260roi
TM(6个波段)	0, 8549	0, 8457	0, 8434	0, 8547	0, 8501	0, 8709	
TM(4个波段)	0, 8343	0, 8412	0, 8434	0, 8343	0, 8246	0, 8297	
Cbers-02B	0, 8411	0, 8616	0, 866	0, 8545	0, 8706	0, 8592	0, 8474
HJ-1A	0, 8152	0, 8343	0, 8201	0, 8134	0, 8271	0, 834	
Beijing-1	0, 8	0, 8075	0, 8257	0, 8165	0, 8325	0, 8258	

Classification Results



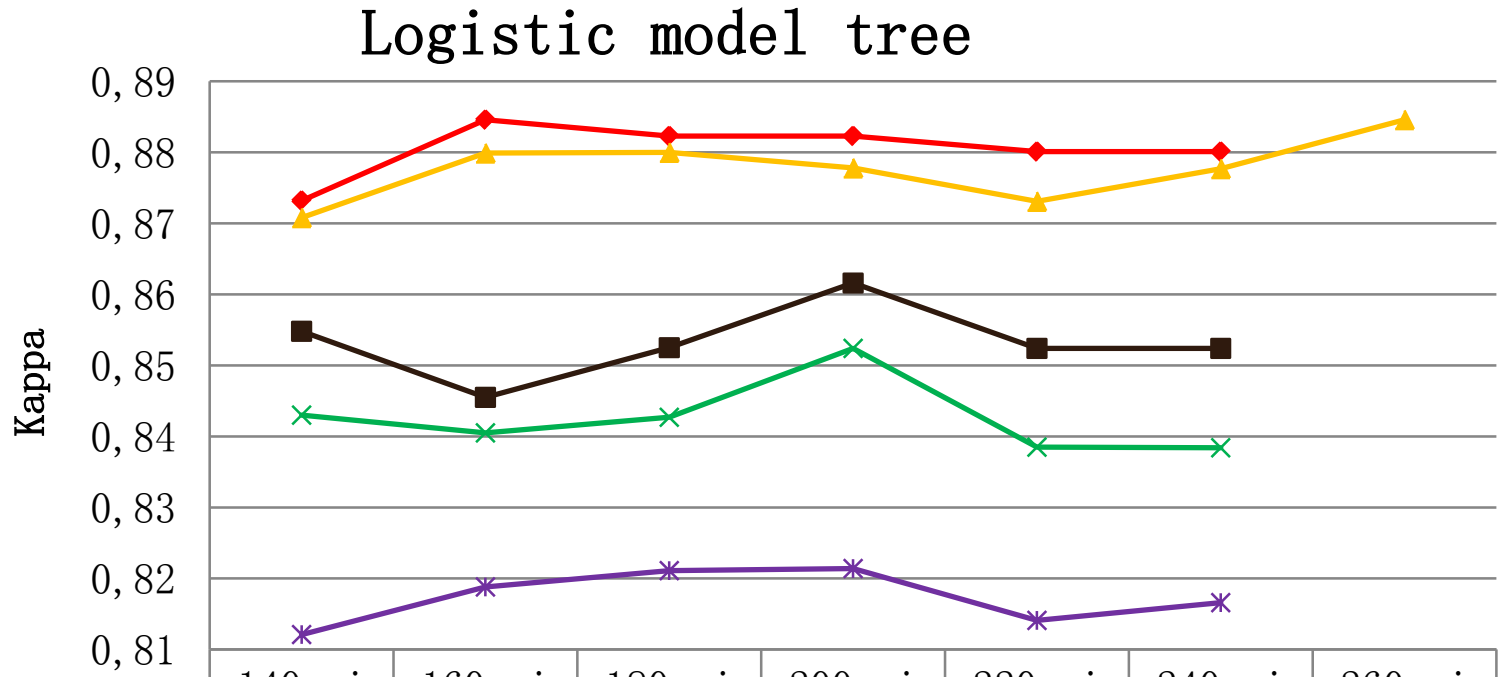
	140roi	160roi	180roi	200roi	220roi	240roi	260roi
TM(6个波段)	0,8915	0,8938	0,8961	0,889	0,8985	0,8984	
TM(4个波段)	0,8456	0,8479	0,8524	0,8616	0,8479	0,8501	
Cbers-02B	0,8661	0,8684	0,8684	0,8754	0,873	0,873	0,88
HJ-1A	0,8385	0,8222	0,8223	0,8338	0,836	0,8338	
Beijing-1	0,8002	0,8002	0,7979	0,7932	0,8002	0,8002	

Classification Results



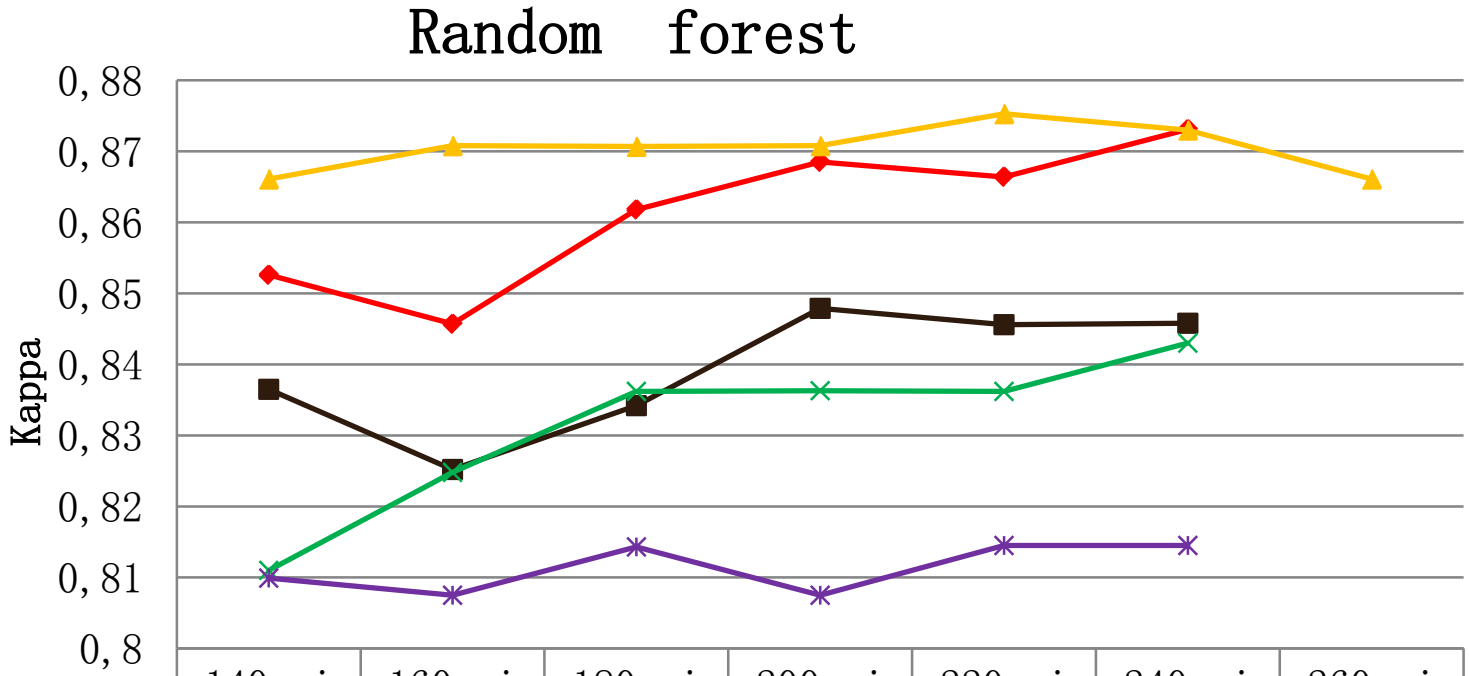
	140roi	160roi	180roi	200roi	220roi	240roi	260roi
TM(6个波段)	0, 8254	0, 83	0, 8387	0, 8433	0, 8481	0, 855	
TM(4个波段)	0, 816	0, 8205	0, 823	0, 8206	0, 8183	0, 823	
Cbers-02B	0, 8457	0, 8434	0, 8479	0, 8525	0, 8525	0, 8571	0, 8387
HJ-1A	0, 8406	0, 8453	0, 843	0, 8499	0, 8523	0, 85	
Beijing-1	0, 8049	0, 8187	0, 8164	0, 8142	0, 8117	0, 8165	

Classification Results



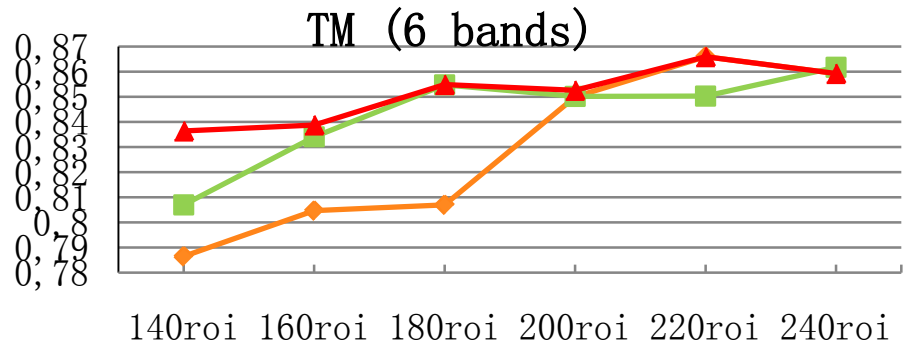
	140roi	160roi	180roi	200roi	220roi	240roi	260roi
TM(6个波段)	0, 8732	0, 8846	0, 8823	0, 8823	0, 8801	0, 8801	
TM(4个波段)	0, 8548	0, 8455	0, 8525	0, 8616	0, 8524	0, 8524	
Cbers-02B	0, 8708	0, 8799	0, 88	0, 8778	0, 8731	0, 8777	0, 8846
HJ-1A	0, 843	0, 8405	0, 8427	0, 8524	0, 8385	0, 8384	
Beijing-1	0, 8121	0, 8188	0, 8211	0, 8214	0, 8141	0, 8166	

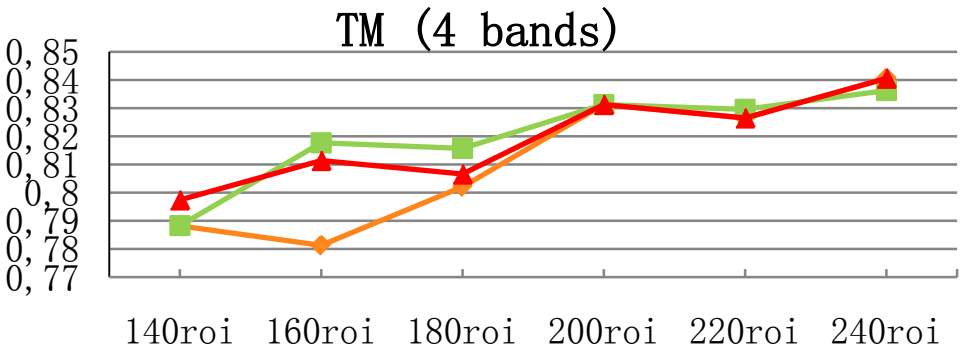
Classification Results

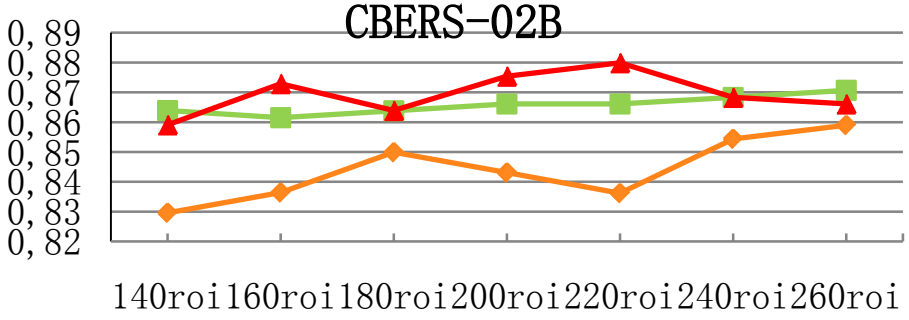


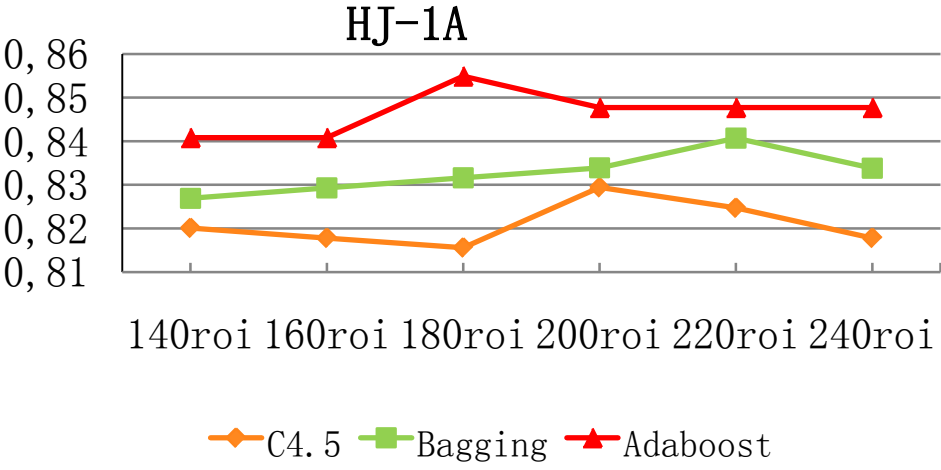
	140roi	160roi	180roi	200roi	220roi	240roi	260roi
TM(6个波段)	0, 8526	0, 8457	0, 8618	0, 8685	0, 8664	0, 8731	
TM(4个波段)	0, 8365	0, 8252	0, 8342	0, 8479	0, 8456	0, 8458	
Cbers-02B	0, 8661	0, 8708	0, 8707	0, 8708	0, 8753	0, 873	0, 8661
HJ-1A	0, 811	0, 8248	0, 8362	0, 8363	0, 8362	0, 843	
Beijing-1	0, 8099	0, 8075	0, 8143	0, 8075	0, 8145	0, 8145	

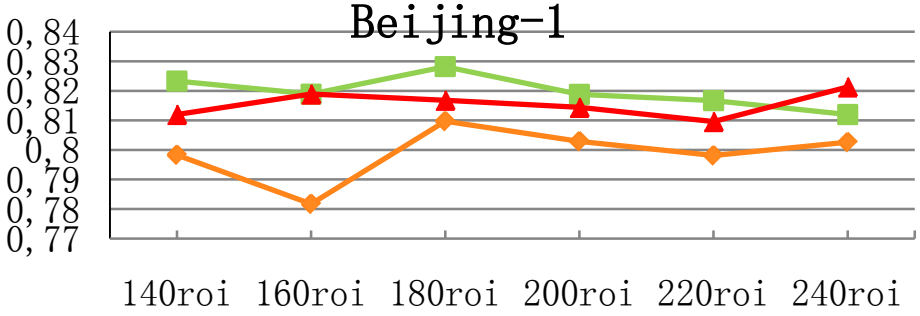
Classification Results





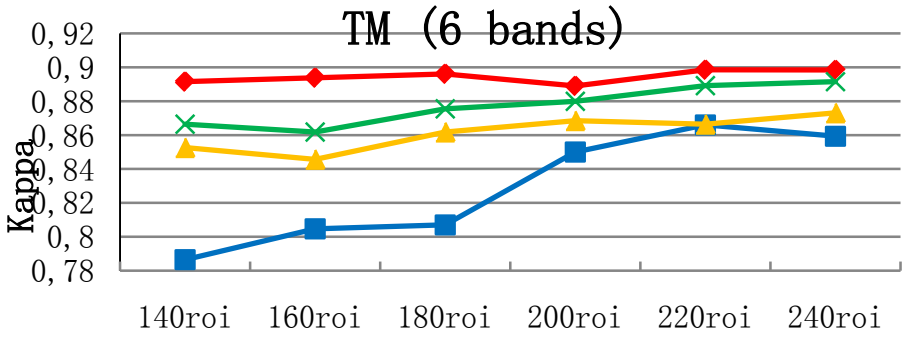


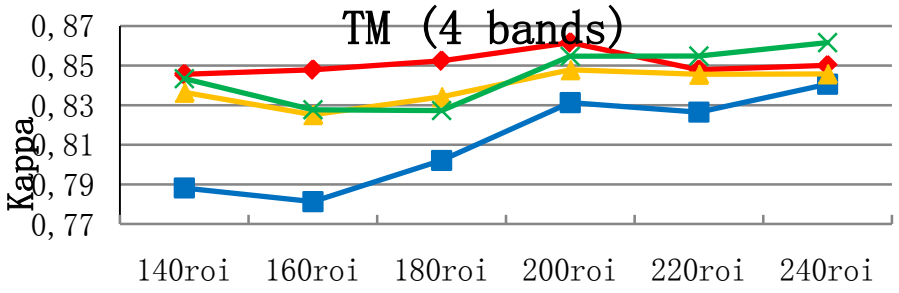


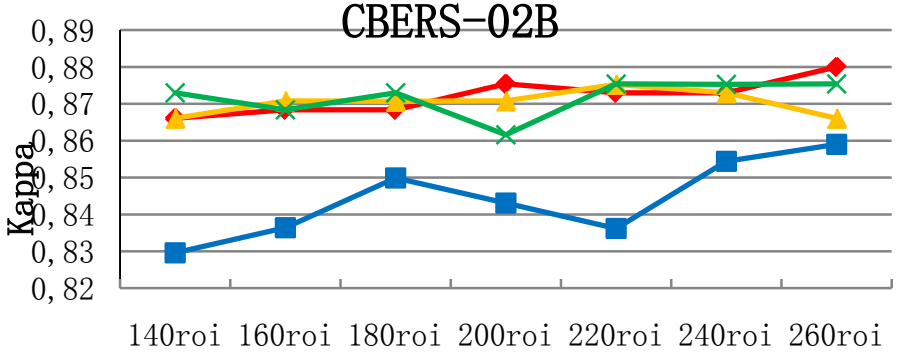


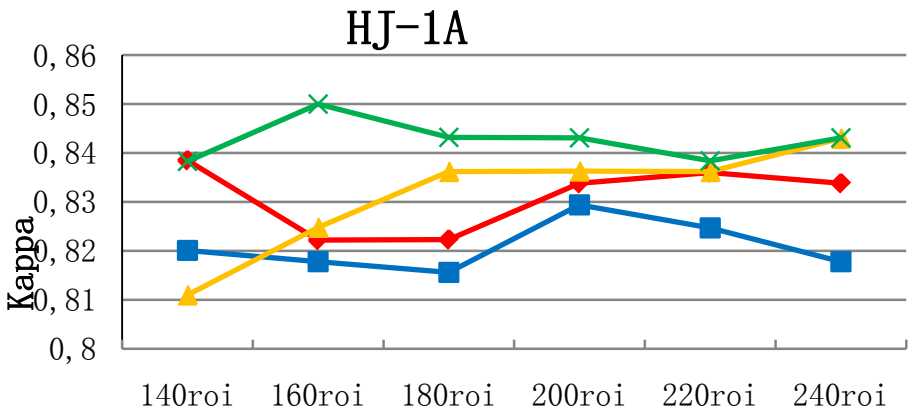
Bagging C4.5, Adaboost C4.5

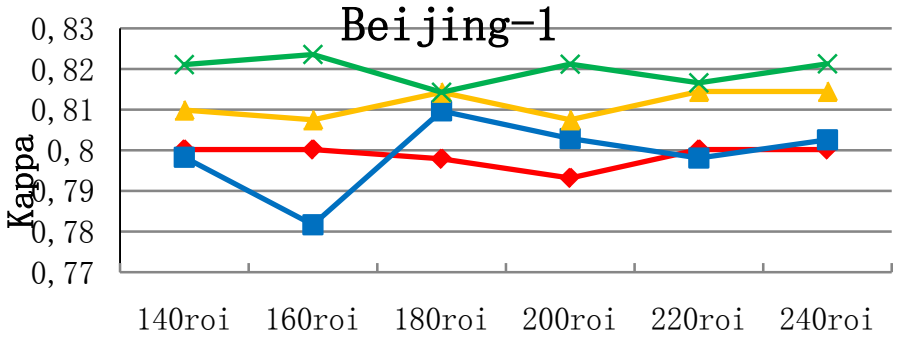
Classification Results





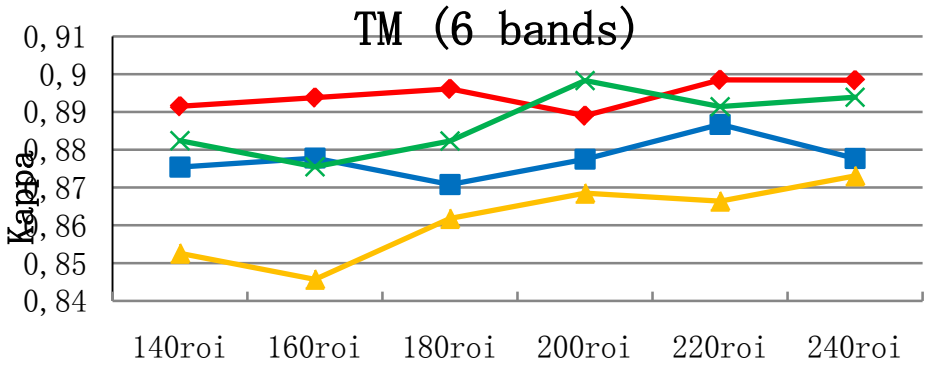


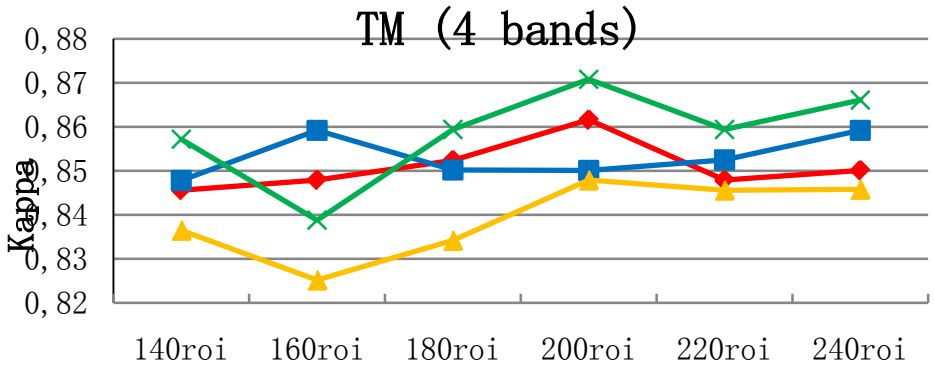


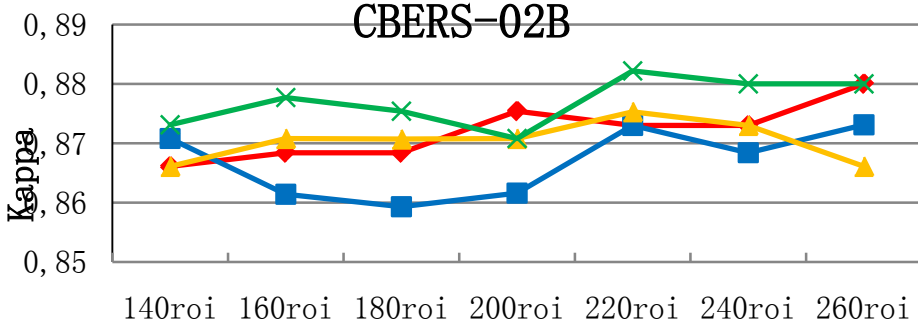


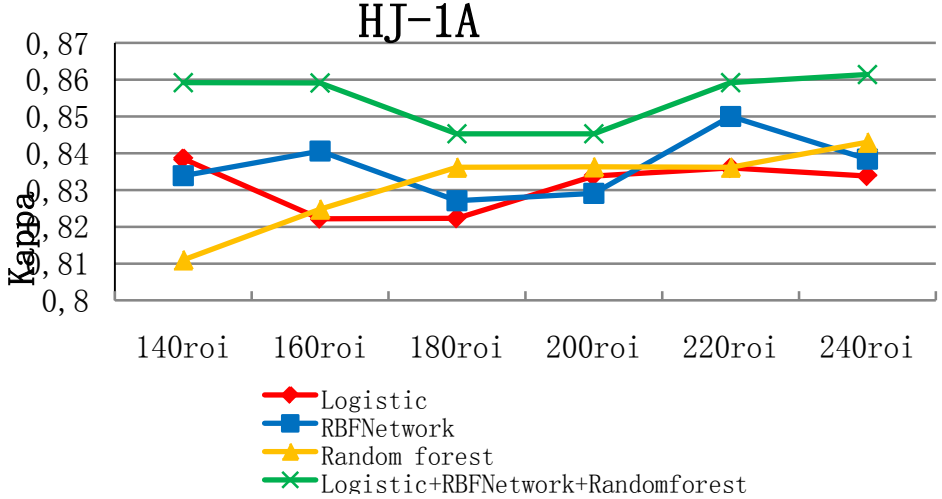
Vote (Logistic, C4.5 and Random forest)

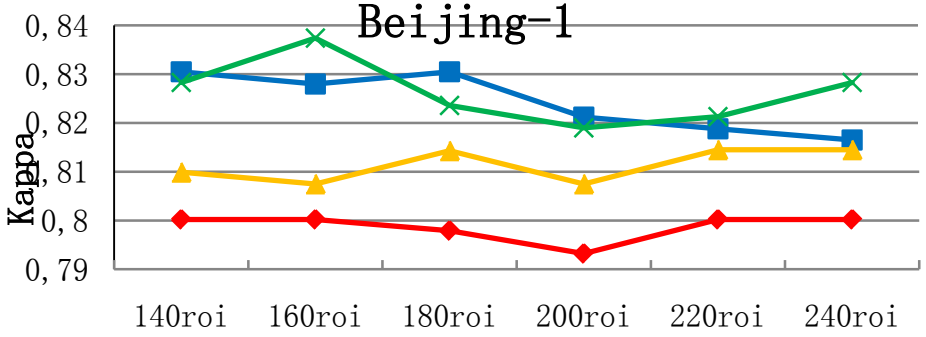
Classification Results





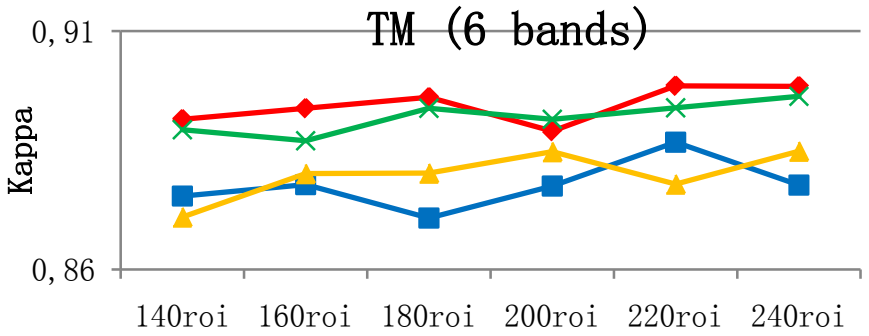


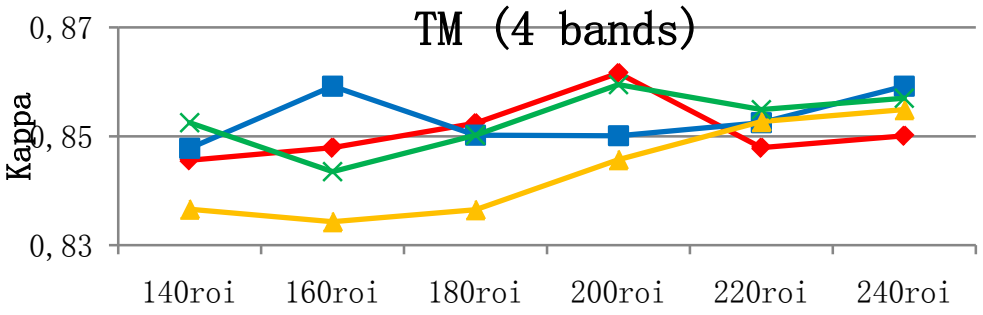


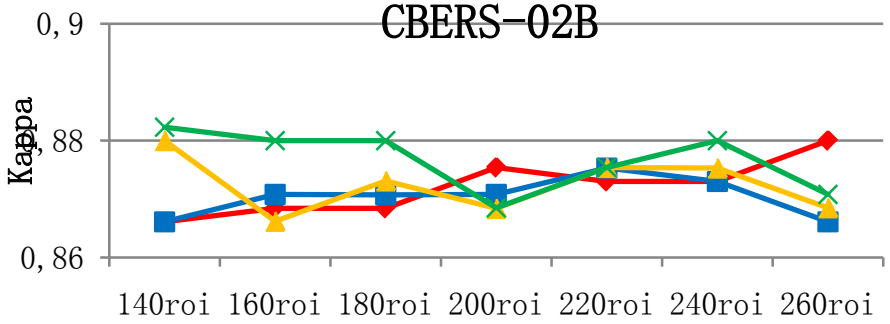


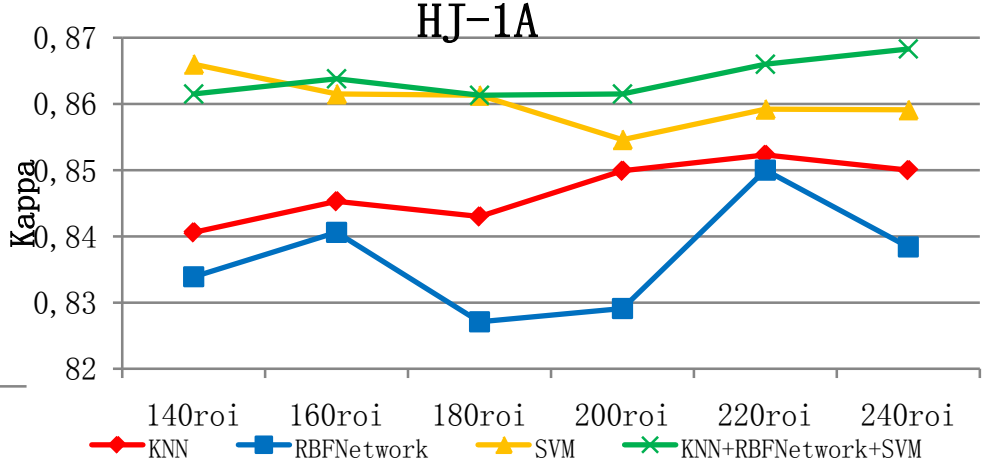
Vote (Logistic, RBFNetwork and Random forest)

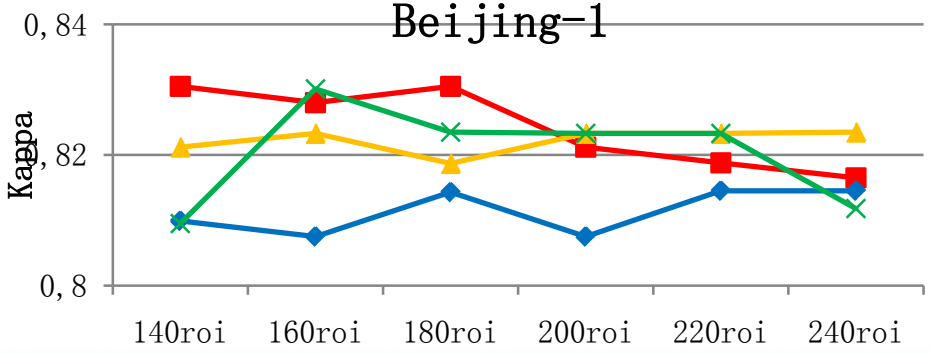
Classification Results





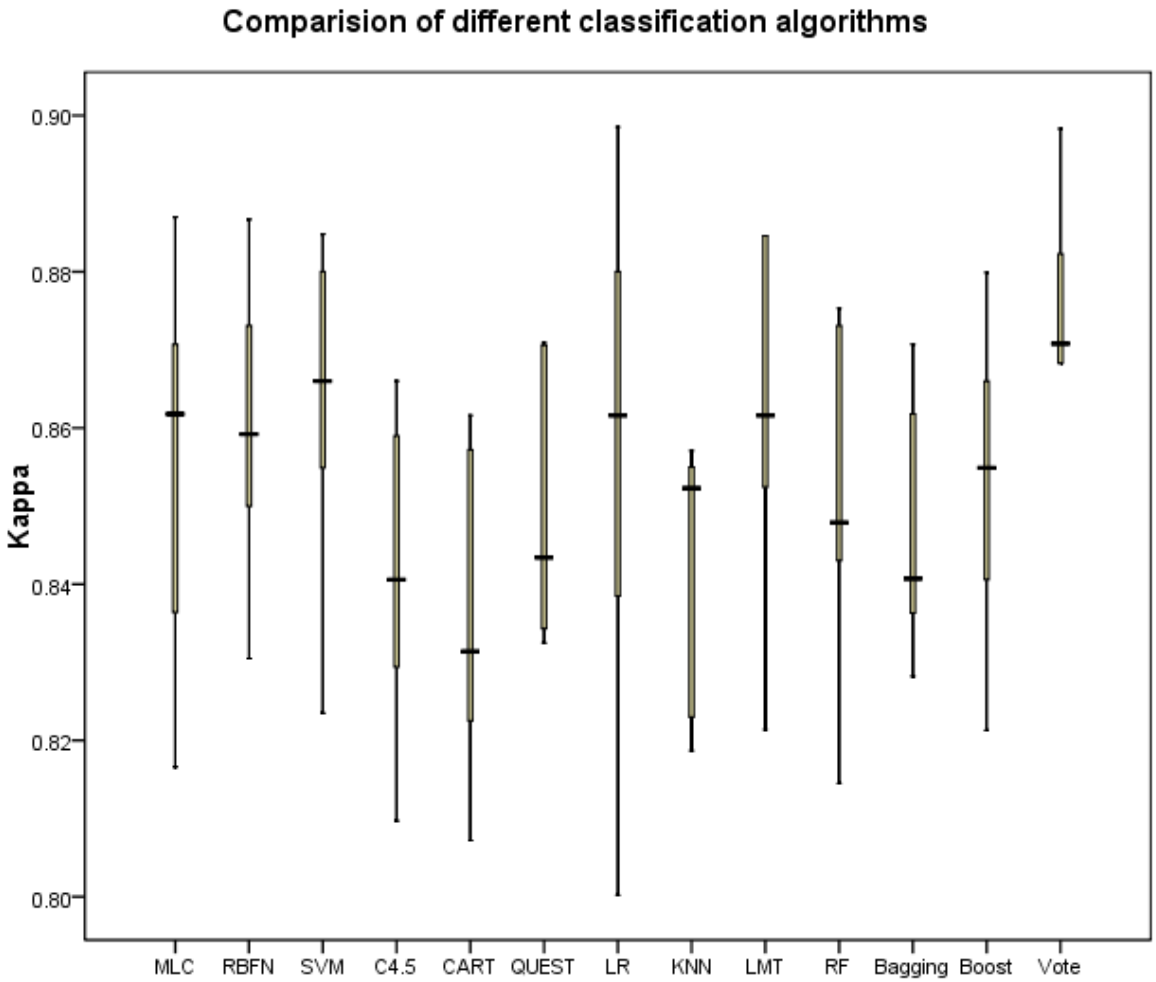




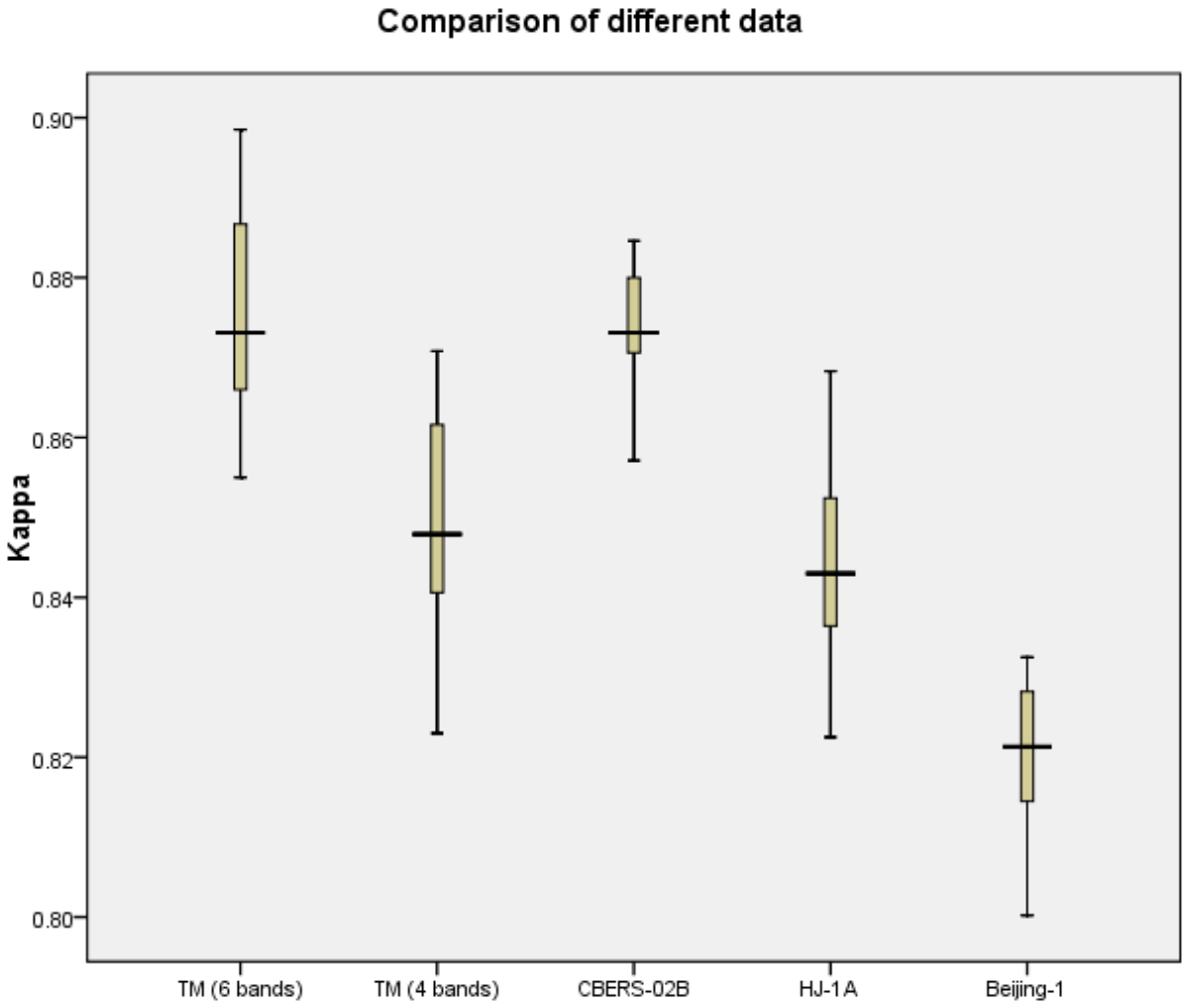


Vote (combination of different classification results)

Classification Results



Classification Results



Integration of GLCM-features

TM：

- Texture window size 7x7, angle 45°, offset 1, 16 grey levels
- 8 parameters

  mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment and correlation (except the correlation parameters of NIRand SWIR). Totally, 51 features of spectral and texture features were calculated.

HJ-1：

- Texture window size 5x5, angle 45°, offset 1, 16 grey levels
- 8 parameters

  mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment and correlation (except the correlation parameters of the red band). Totally, 35 features of spectral and texture features were calculated.

Integration of GLCM-features

Cbers-02B：

- Texture window size 9x9, angle 45°, offset 1, 16 grey levels
- 8 parameters

  mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment and correlation. Totally, 36 features of spectral and texture features were calculated.

Beijing-1：

- Texture window size 5x5, angle 45°, offset 1, 16 grey levels
- 8 parameters

  mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment and correlation. Totally, 27 features of spectral and texture features were calculated.

Classification Results (GLCM)

Classification algorithm (TM)	Kappa （spectral）	Kappa （spectral + 51 texture features）
MLC	0.887	0.848
KNN	0.853	0.848
C4.5	0.862	0.850
Logistic regression	0.899	0.933
RBFNetwork	0.878	0.864
SVM	0.885	0.924
Random forest	0.873	0.910
LMT	0.885	0.926
Bagging	0.862	0.876
Adaboost	0.857	0.896

Classification Results (GLCM)

Classification algorithm (HJ-1)	Kappa （spectral）	Kappa （spectral + 35 texture features）
MLC	0.834	0.658
KNN	0.850	0.720
C4.5	0.818	0.829
CART	0.822	0.832
Logistic regression	0.834	0.848
RBFNetwork	0.838	0.787
Random forest	0.843	0.878

Classification Results (GLCM)

Classification algorithm (Cbers-02B)	Kappa （spectral）	Kappa （spectral + 36 texture features）
MLC	0.866	0.758
C4.5	0.854	0.866
QUEST	0.859	0.894
Logistic regression	0.873	0.894
Random forest	0.873	0.924
LMT	0.878	0.919
Adaboost C4.5	0.868	0.915

Classification Results (GLCM)

Classification algorithm (Bejing-1)	Kappa （spectral）	Kappa （spectral + 27 texture features）
MLC	0.817	0.581
KNN	0.819	0.756
C4.5	0.803	0.819
CART	0.807	0.823
QUEST	0.826	0.849
Logistic Regression	0.800	0.854
RBFNetwork	0.817	0.810
SVM	0.824	0.893
Random forest	0.815	0.889
LMT	0.817	0.870
Bagging C4.5	0.812	0.856

Conclusions

1) Most algorithms are not sensitive to the size of training sample size.

2) Voting schemes, SVM, Logisitic and MLC can achieve better classification accuracies. Amongst these, MLC and logistic classifiers have proven most efficient.

3) Ensemble classifiers (such as RF and LMT) can also produce good results. Bagging and boosting of decision tree algorithms can improve weak classifiers (e.g. C4.5) especially for an insufficient amount of training samples.

4) Among the four types of satellite images, Landsat-5 can achieve the best accuracy with Logistic, MLC, RBFN, SVM and LMT classifiers. Cbers-02B can achieve the highest accuracies with LMT, SVM and Logistic classifiers. HJ-1 can achieve the best accuracy with SVM whereas Beijing-1 can achieve highest accuracy with QUEST and RBFN.

Conclusions

5） Landsat-5 data outperforms the remaining Chinese satellite data, while CBERS-02B data takes the second place. Their maximum difference is less than 0.02 kappa value. Higher resolution CBERS-02B data can be easily obtained. Therefore, it can be used as auxiliary data for land use mapping. By using only 4 Landsat-5 bands, the results are worse than the results when using 6 bands.

6）Classification accuracies are higher integrating GLCM texture features in most of the cases, except for the MLC classification, where accuracies are always lower. The kappa values could be increased by 0.028 to 0.067 for all satellites under consideration within the study.

7) SVM, Random forest, Adaboost and Logistic Regression classifiers can deal with high dimensional remote sensing data.

Thanks for your attention!
Questions?