



IGU/LULC Northeast East Asia Conference-2009

# Mapping the Heilongjiang Basin LULC with MODIS 250 m NDVI time series data



**Kaishan Song, Zongming Wang, Bai Zhang**  
Northeast Institute of Geography and Agricultural Ecology, CAS  
September 7, 2009, Vladivostok



# Outlines

**1**

**Introduction**

**2**

**Study Area Descriptions**

**3**

**Images Data and Methodology**

**4**

**Classification Process**

**5**

**Result and Analysis**

**6**

**Discussions**



# Introduction

- ❖ Land use/land cover (LULC) data sets are primary inputs for environmental modeling and monitoring, natural resources management, policy makings, and global climate change analysis;
- ❖ The purposes of the research:
  - To investigate the feasibility of using MODIS 250m NDVI time-series data to identify LULC in the Heilongjiang River Basin(Amur River Basin:AMB);
  - To develop classification methodology for updated LULC information retrieval using a nearly continuous high-quality data stream that could also be used to support phenology-based cover type classification.
  - To provide updated annual step LULC information for environmental modelling and natural resources management.



# Study Area Descriptions

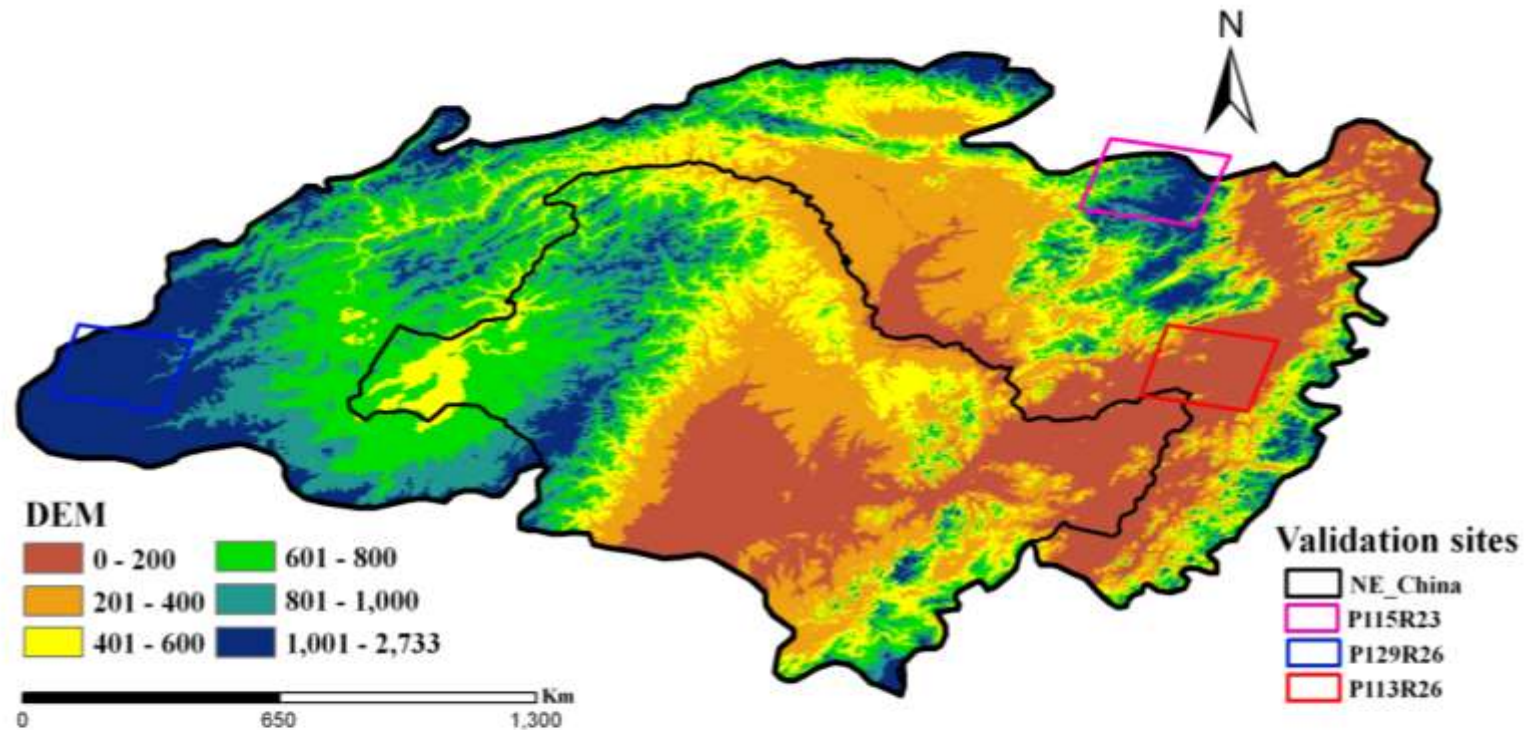


The study area of Heilongjiang River Basin and its boundary





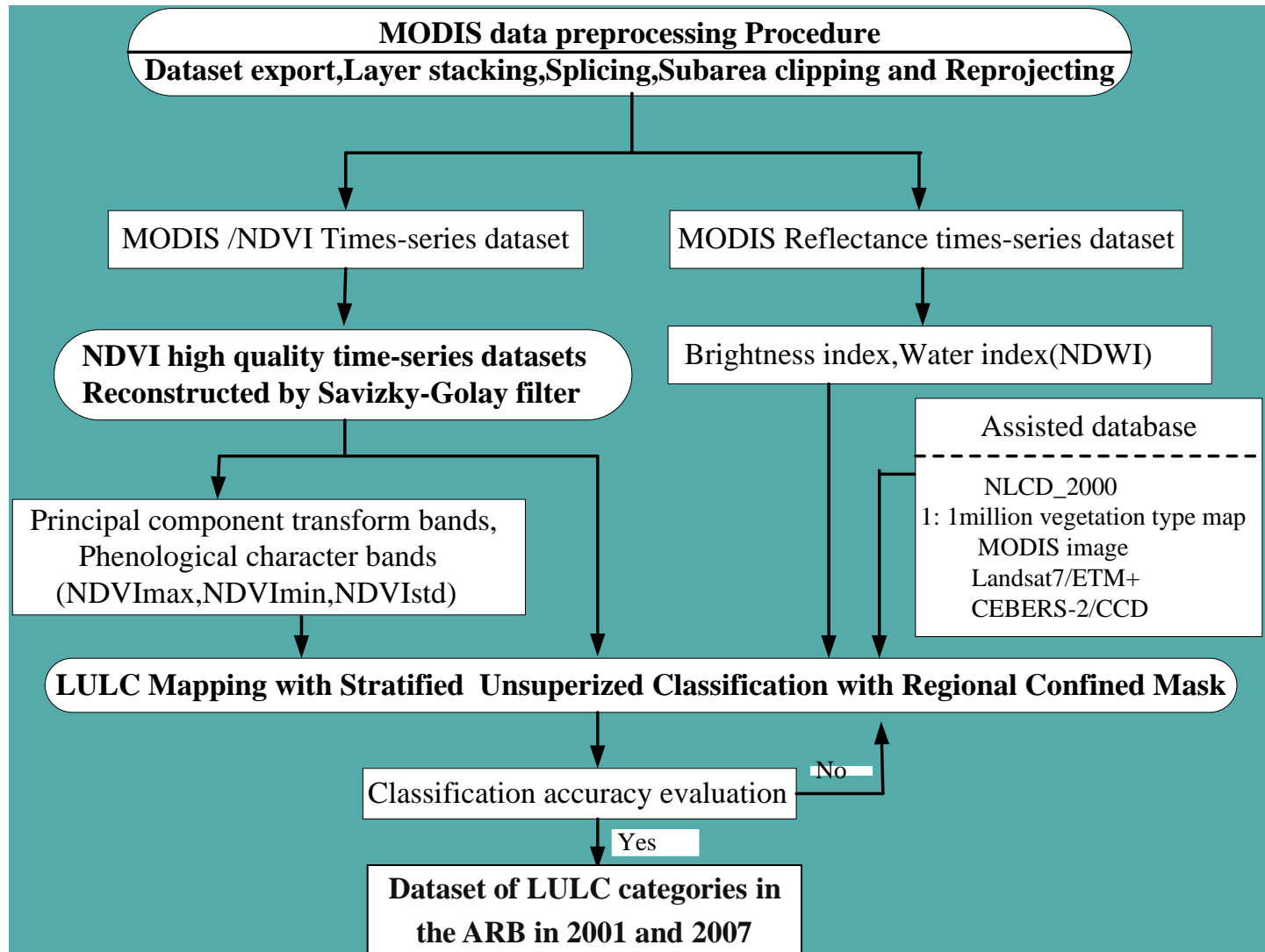
# Study Area Descriptions



The area of Heilongjiang River (Amur River) Basin regions is about 2,095,000 km<sup>2</sup>; Land cover: boreal evergreen conifer-deciduous broad leaf mixed forests, deciduous broad leaf forests, and shrublands, typical steppes and desert steppes; The study area is characterized by a temperate continental monsoon climate.



# Images Data and Methodology





# Images Data and Methodology

- ❖ MOD13Q1: 16-day composite MODIS NDVI Product , 250m;
- ❖ Tiles:H25v03,H26v03,H25v04,H26v04,H27v04,H27v05;
- ❖ Time: January 2001 to Decemeber 2001 (Jan to Dec, 2007);
- ❖ Duration Series: twenty-three 16-day composites for each tile;
- ❖ Source: USGS EROS Data Center (<http://edc.usgs.gov/>)
- ❖ Landsat TM/ETM+ imagery data in year 1999-2002 for evaluation and AOI selection to assess LULC accuracy from Modis 250 NDVI time series data of the concurrent year.



# Images Data and Methodology

## ❖ NDVI time series generation

NDVI time series contain noise, caused by cloud contamination, atmospheric variability, and bi-directional effects.

## ❖ How to minimize NDVI time series data noise?

Earlier: MVC (Maximum Value Composite)

In the last decade: CV-MVC, BISE, Fourier-Based Fitting methods and asymmetric function fitting methods, etc;

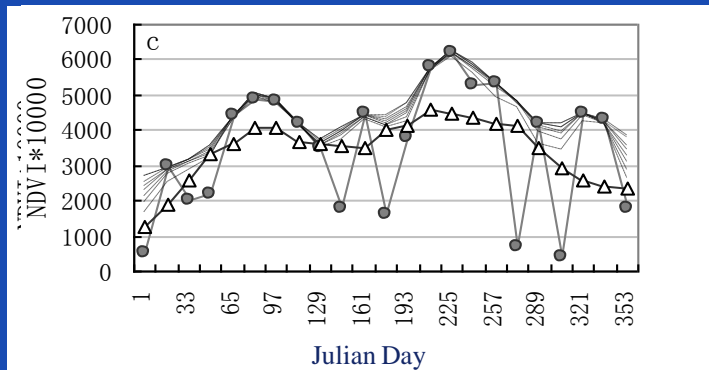
In this research: A simple but robust method based on the Savitzky–Golay filter (Chen et al, 2004), to more efficiently reduce contamination in the NDVI data.



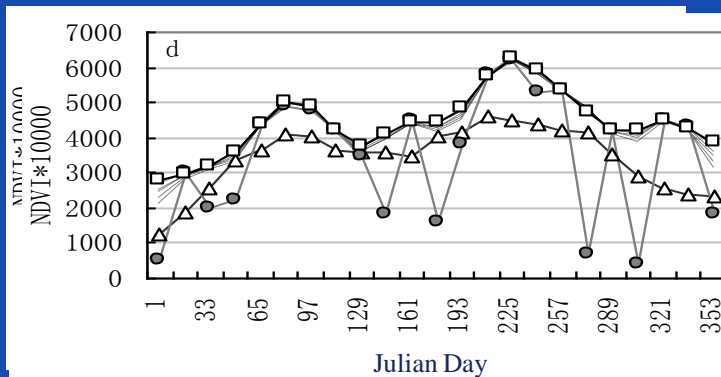


# Images Data Processing

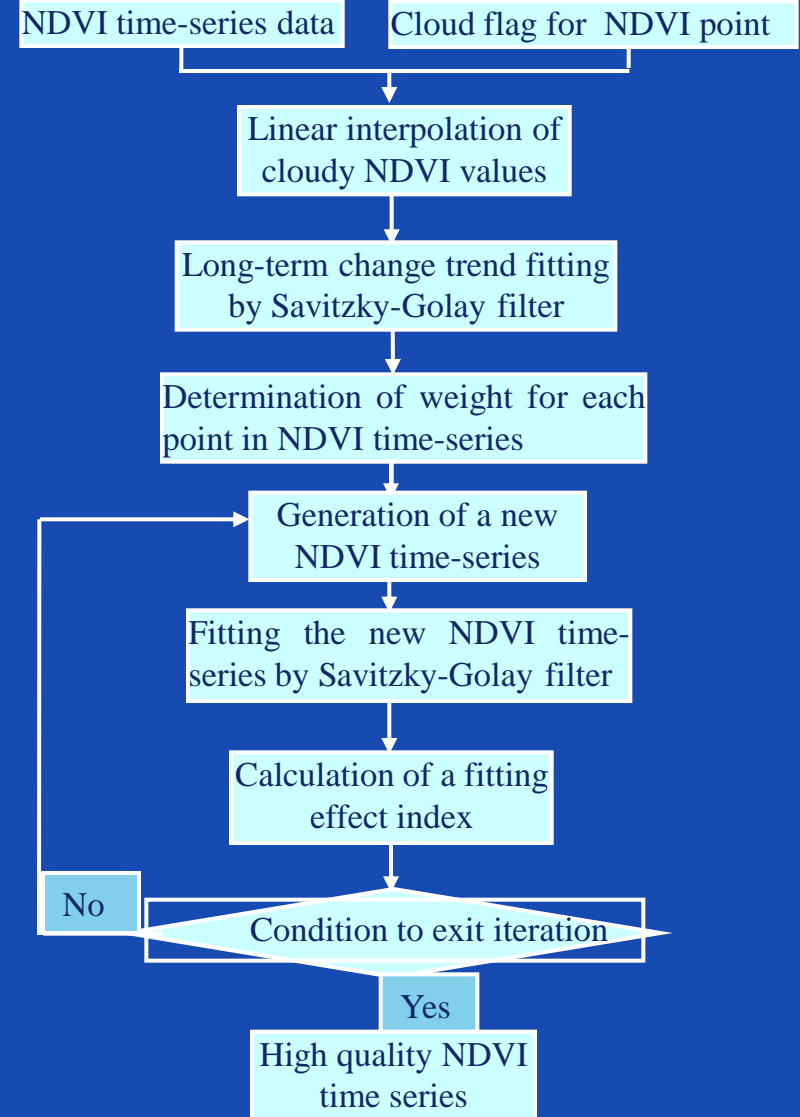
## ➤ NDVI time series generation



**The Process of S-G filtering after several iteration**



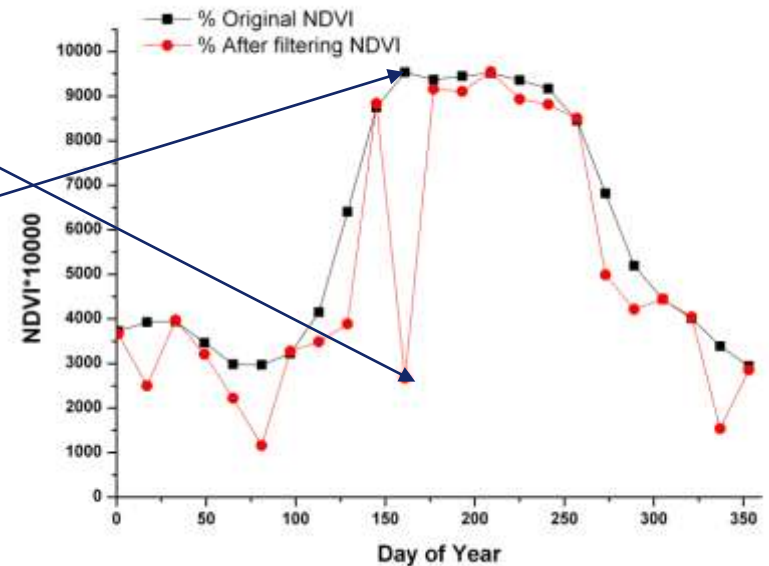
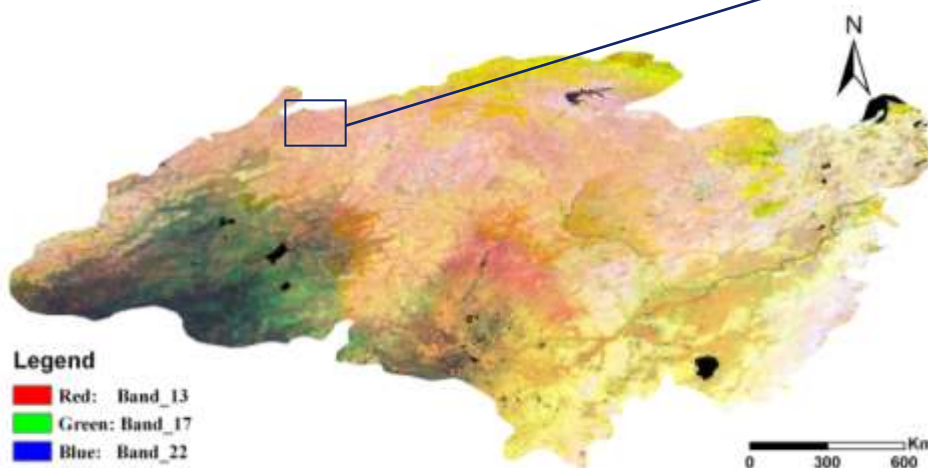
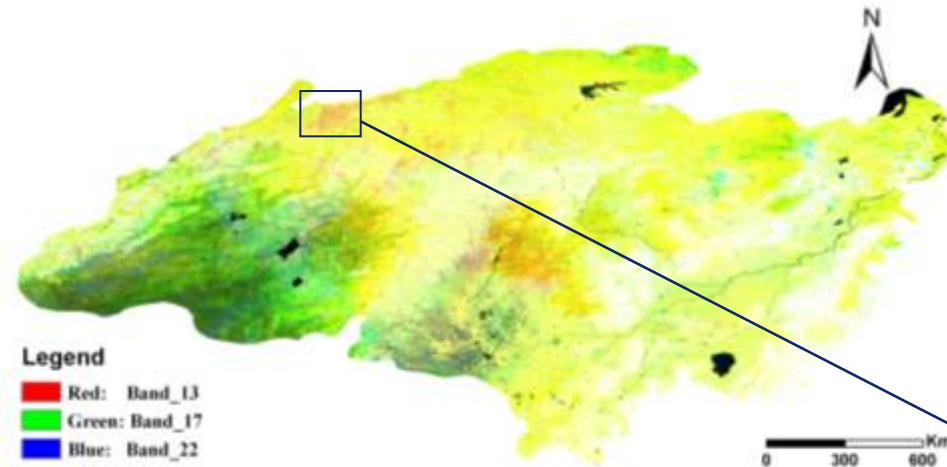
**The qualified NDVI time series data reconstructed with S-G filtering**



**Flowchart of the Savitzky-Golay filter**



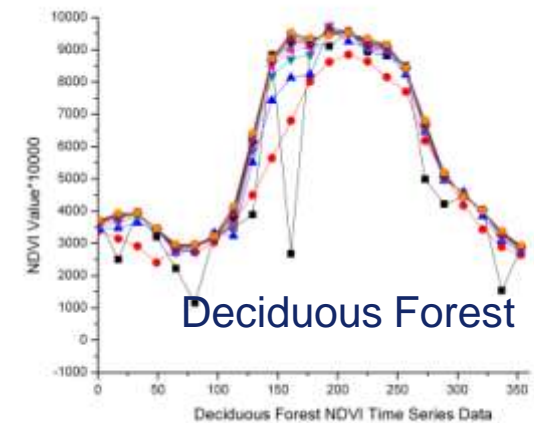
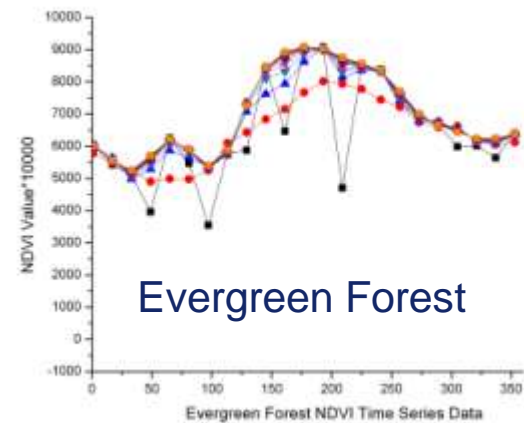
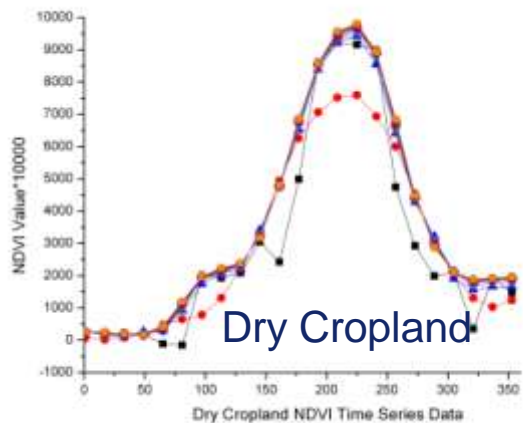
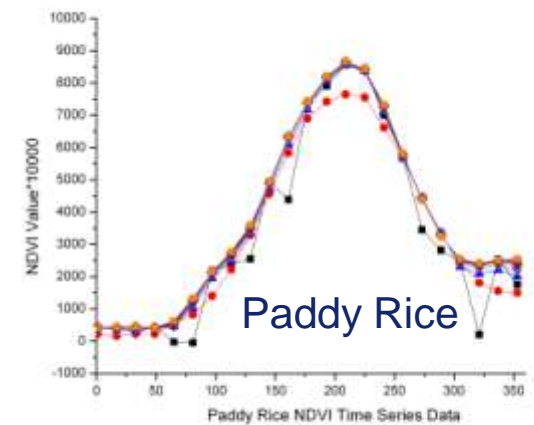
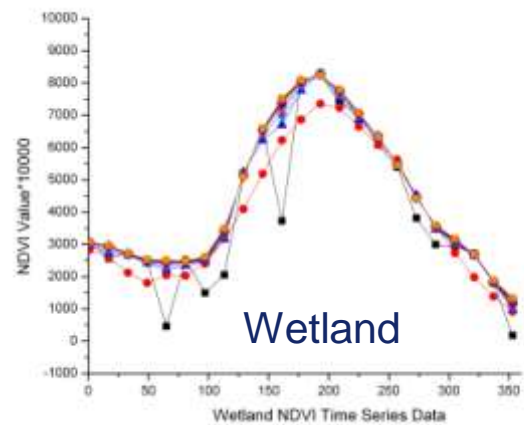
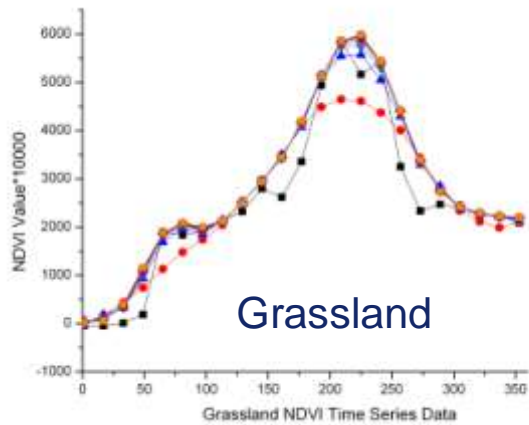
# Images Data Processing



Aftering filtering, the NDVI composite image looks much better, and the quality of time series data are improved greatly



# Classification Methodology





# Classification Methodology

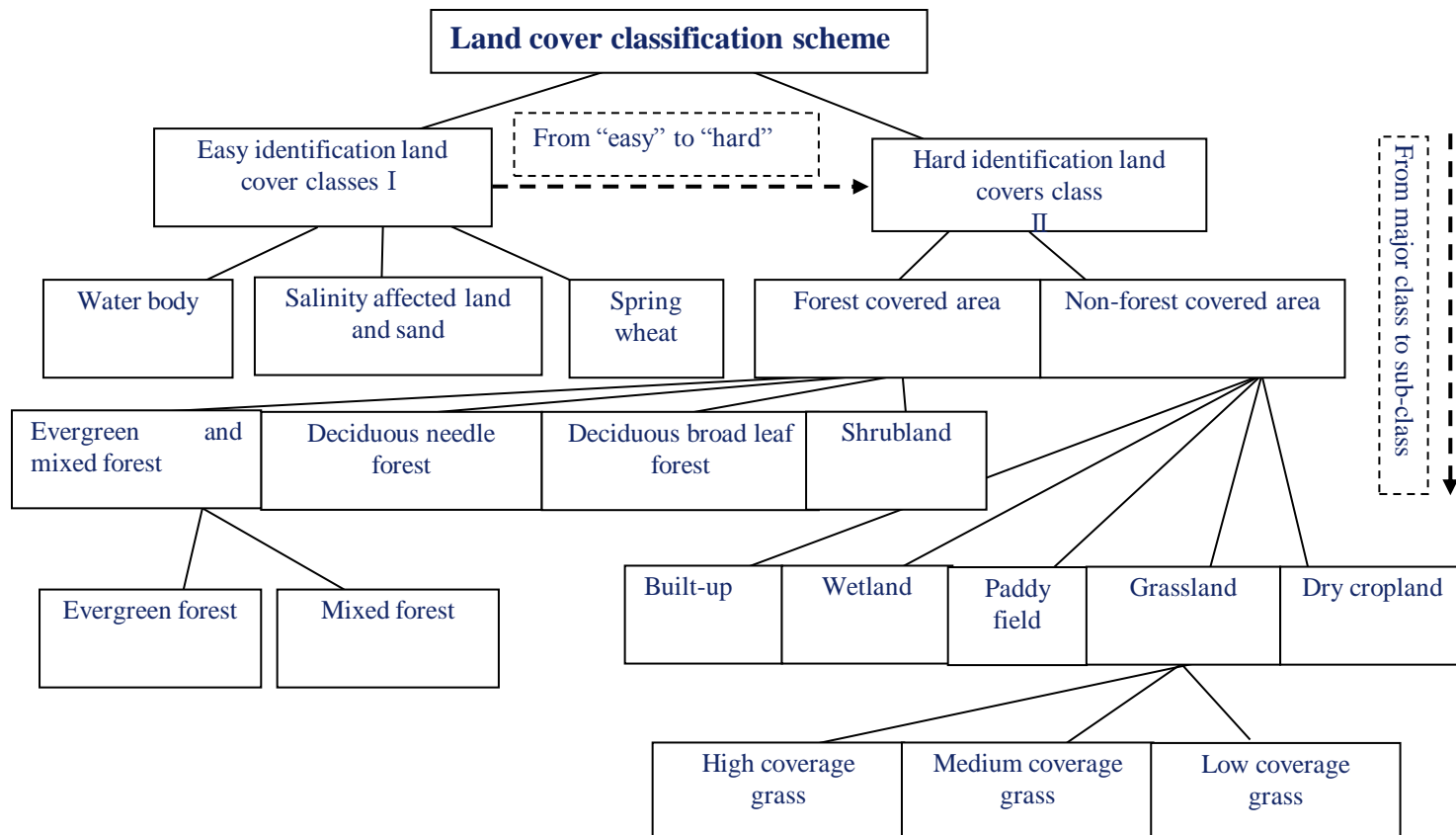
Modis-NDVI-250 class	Class description in this study
Evergreen needle leaf forest	Dominated by evergreen needle leaf trees with a percent canopy cover greater than 60%.
Deciduous needle leaf forest	Dominated by with deciduous needle leaf tree with a percent canopy cover greater than 50% and height exceeding 2 m.
Deciduous broad leaf forest	Dominated by with deciduous broad leaf tree with a percent canopy cover greater than 60% and height exceeding 2 m.
Mixed forest	Dominated by mixed trees types with a percent canopy cover greater than 50% and height exceeding to 2 m.
Shrubland	Vegetation cover (woody or herbaceous) of 10–30%.
Dry cropland	Land dedicated to the production of rain fed crops in the summer.
Paddy rice	Land dedicated to the production of rice with summer irrigation practice.
Wheat	Cropping land with spring wheat planting practice
Wetland	Lands with permanent mixture of water and herbaceous or woody vegetation.
Water body	Permanent water bodies, including lake, river and tanks
Built-up	Covered by buildings or other man-made structures
Salinity/sandy land	Land affected by salinity problem or covered by sand with sparse desert steppe.
High density grassland	Lands with herbaceous types of cover. with a percent canopy cover greater than 60%
medium density grassland	Lands with herbaceous types of cover. with a percent canopy cover greater than 30-60%
Low density grassland	Lands with herbaceous types of cover. with a percent canopy cover less than 30%





# Classification Methodology

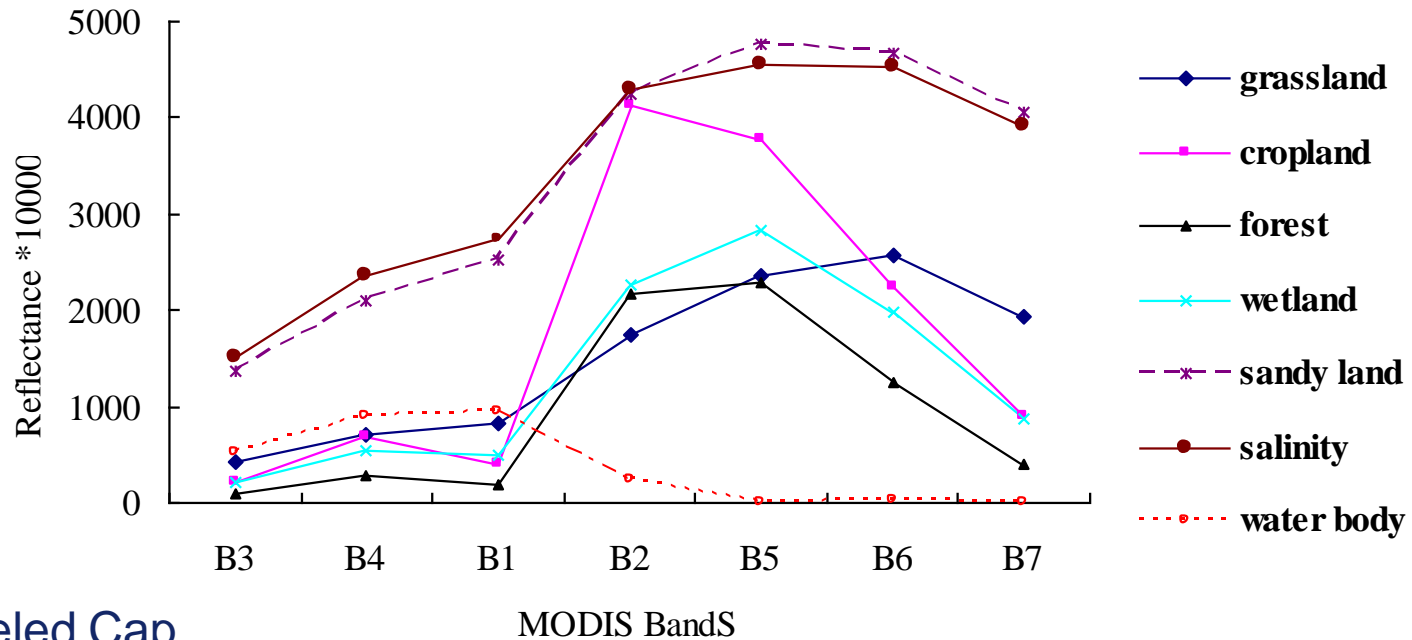
**Hybrid Classification**  *Unsupervised classification with ISODATA*  
*Stratified Hierarchy Decision Tree Approach*





# Classification Methodology

## ◆ Water body and Barren Retrieval



Tasseled Cap

$$B\_brightness = 0.396*B1 + 0.472*B2 + 0.335*B3 + 0.383*B4 + 0.394*B5 + 0.343*B6 + 0.296*B7$$

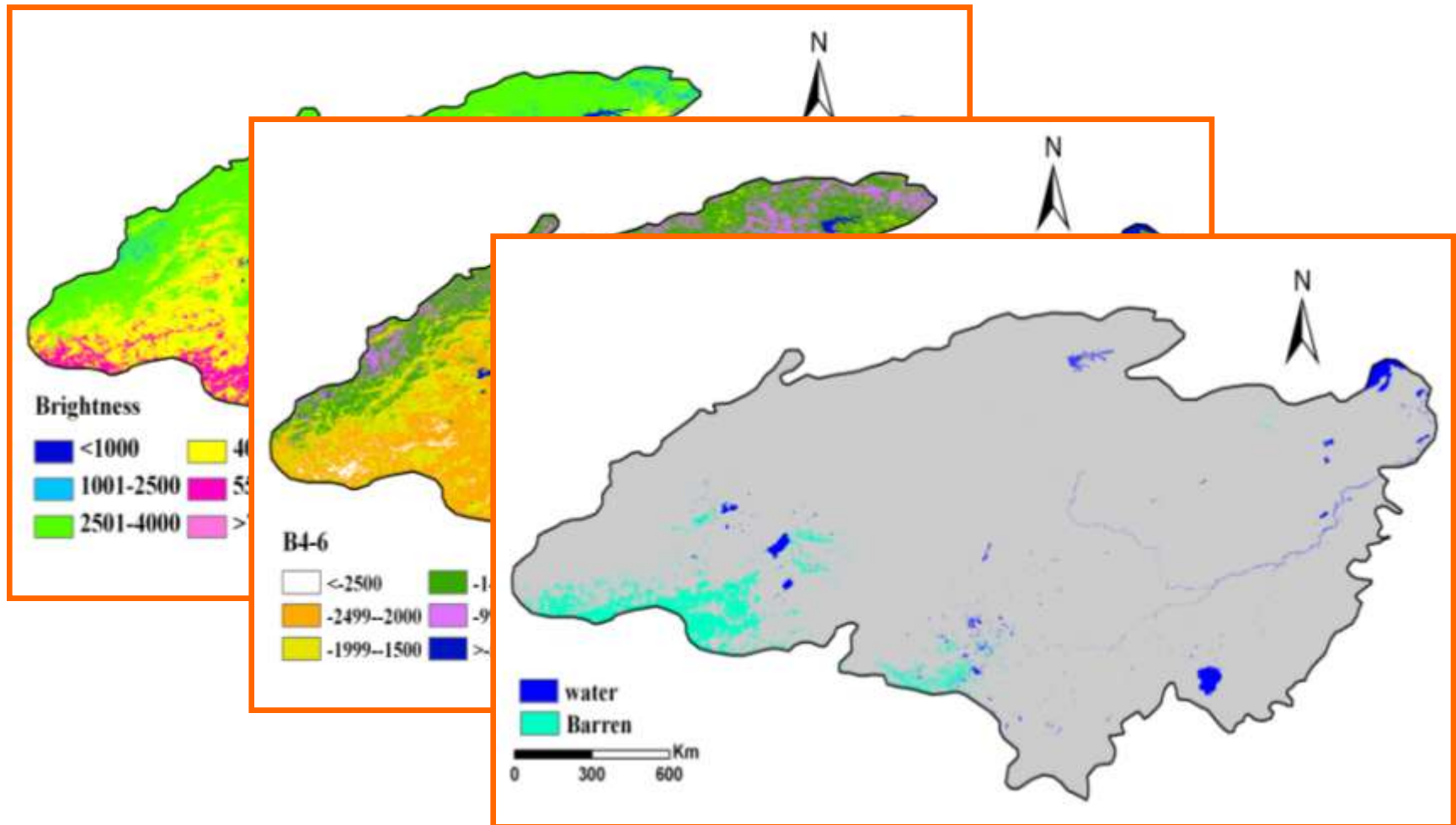
Water body:  $(B4 - B6) > -500$  and  $NDVI_{max} < 6000$

Salinity affected and sandy land:  $B\_brightness > 6500$  and  $NDVI_{max} < 6000$



# Classification Methodology

## ◆ Water body and barren extraction method

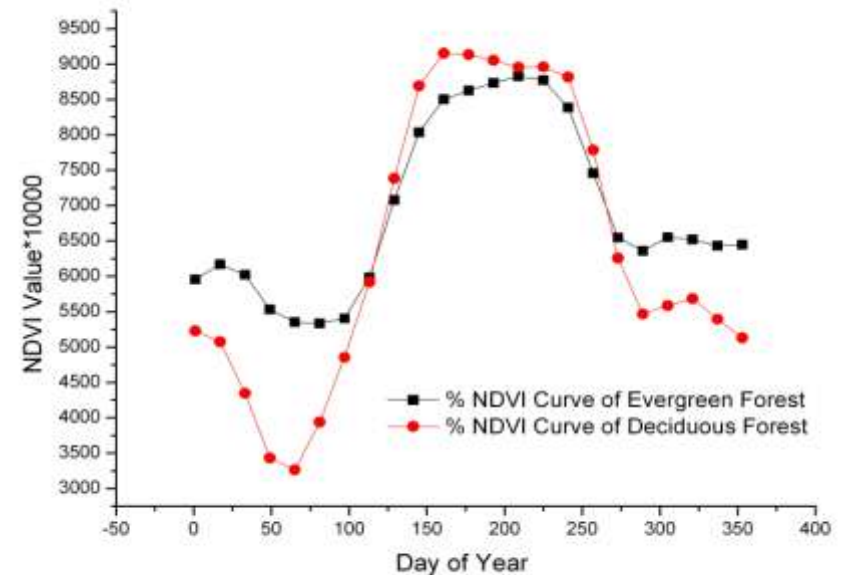
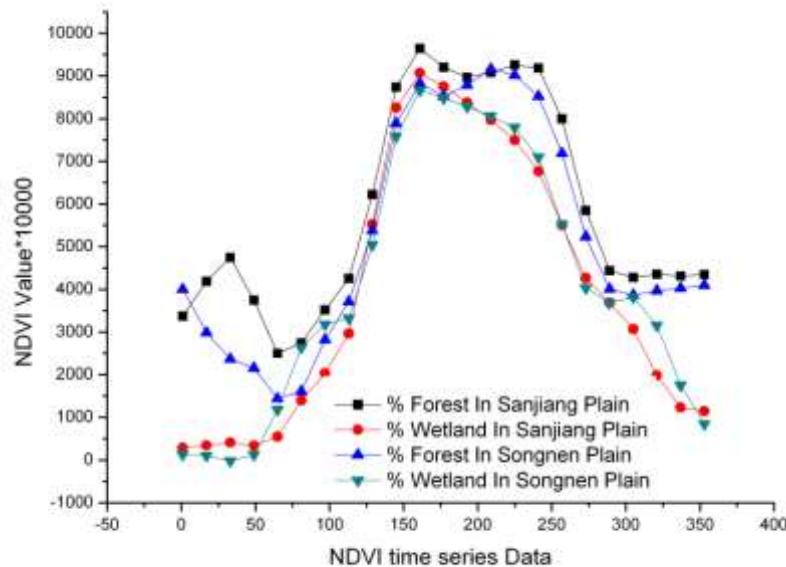




# Classification Methodology

## ◆ Forest Extraction Methodology

- Principal Component Analysis
- ISODATA clustering method
- Methods for subclass forest classification



NDVI161-NDVI241>200 and NDVI353<5000 and DEM<400

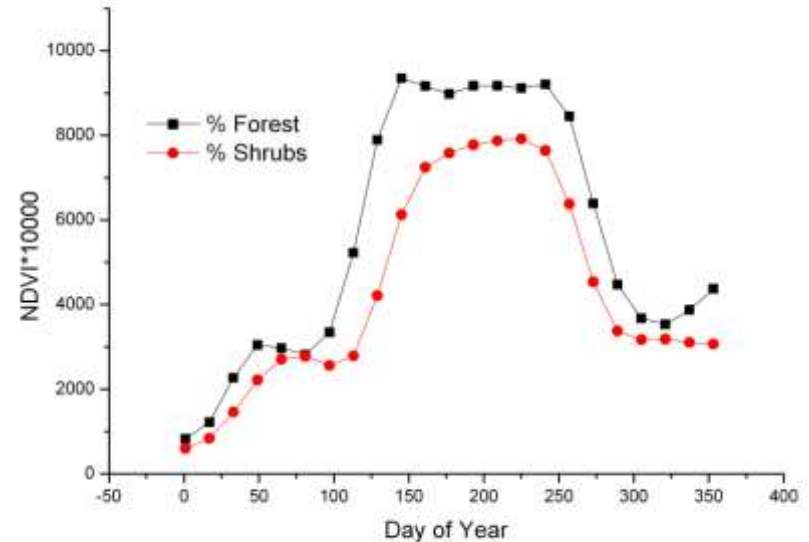
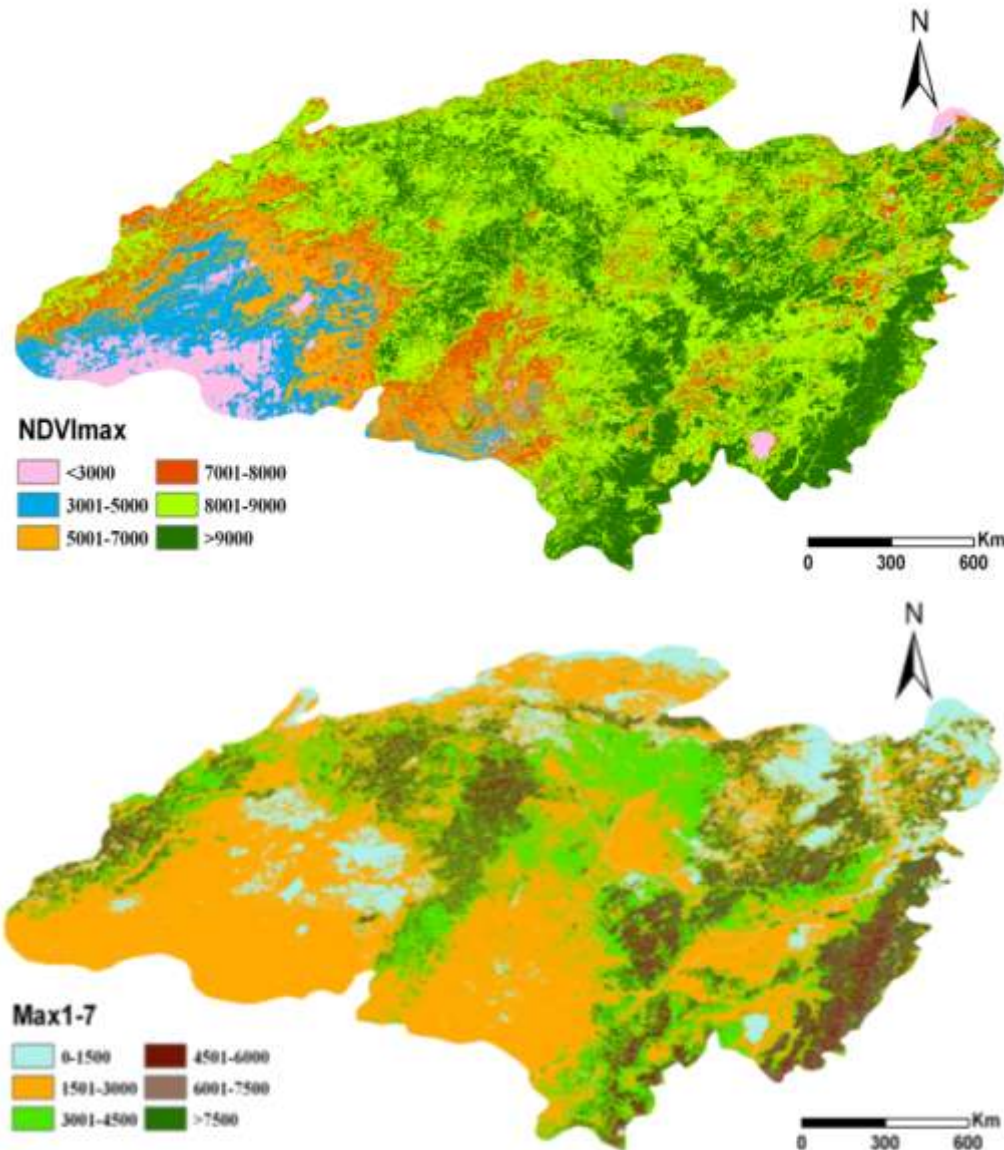
NDVI1-133\_min>3000 and NDVI289-353\_min>4000

Evergreen forest =  $\begin{cases} DEM > 700m \dots \text{Xiaoxing'an Mountain} \\ DEM > 900m \dots \text{Zhangguangcai Mountain} \\ DEM > 1100m \dots \text{Changbai Mountain} \end{cases}$





# Classification Methodology



Vegetation map of study region with scale of 1:1000000 was used as reference for shrubland extraction, and the scrubland area can be optimally extracted when threshold value was set to  $NDVImax < 8300$

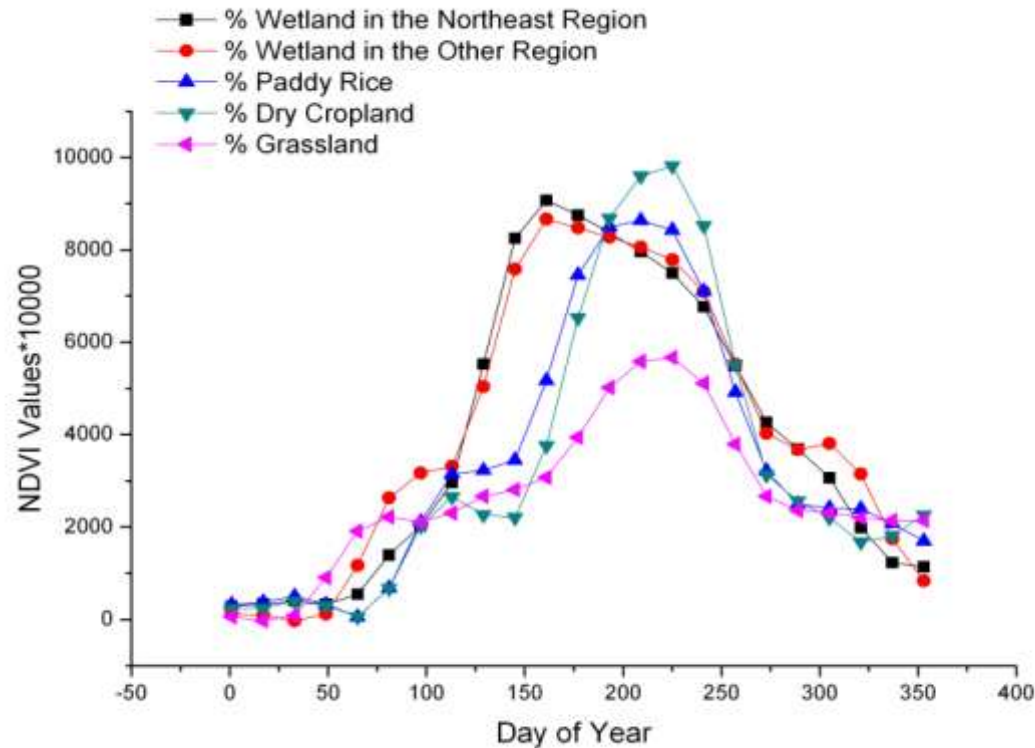


# Classification Methodology

## ● Wetland extraction method

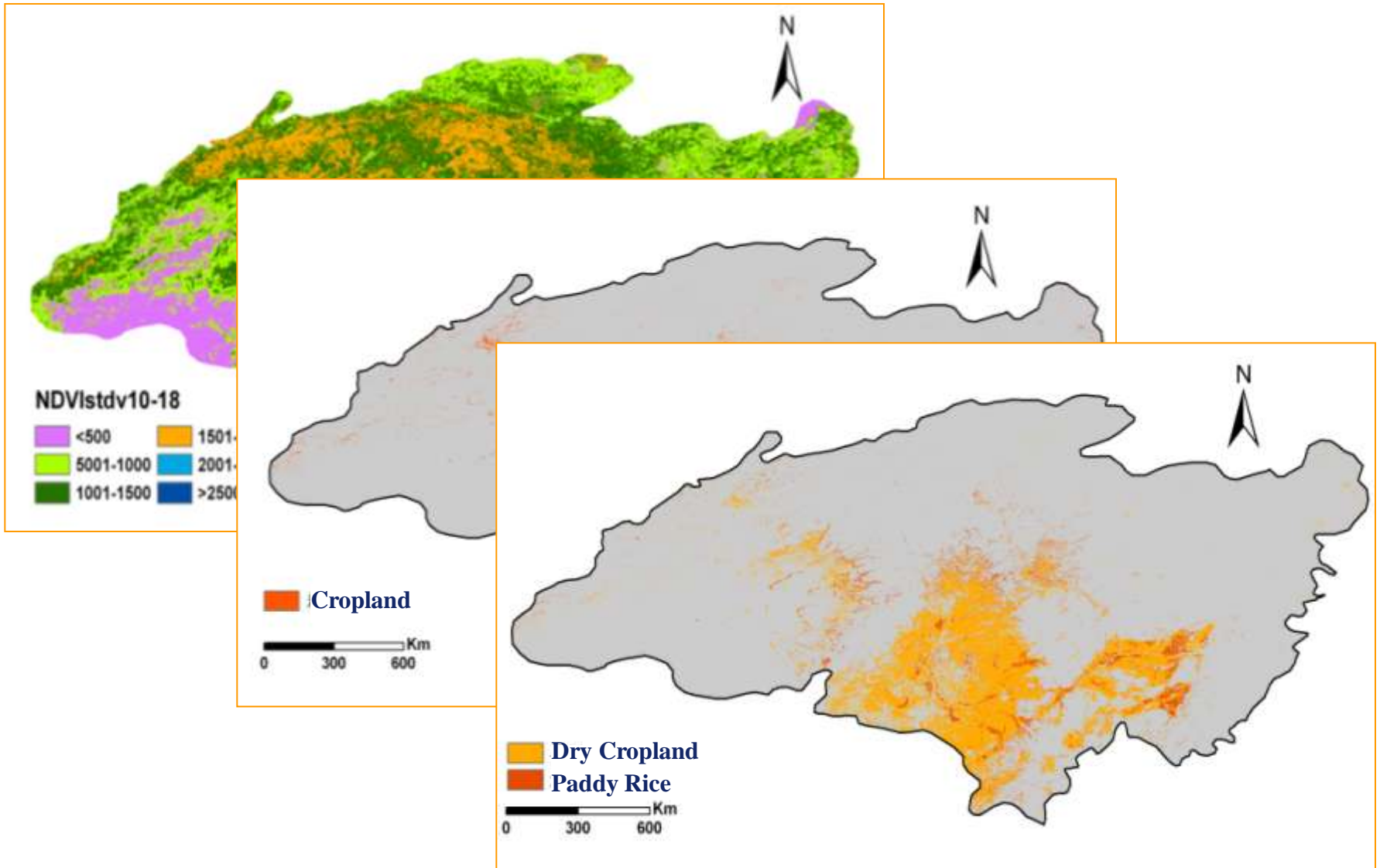
Wetland in Northeast Region of HLJ:  $NDVI_{161-241} > 0$  and  $NDVI_{161} > 5000$ ;

Wetland in the other regions:  $NDVI_{161-241} > -500$  and  $NDVI_{161} > 5000$ .



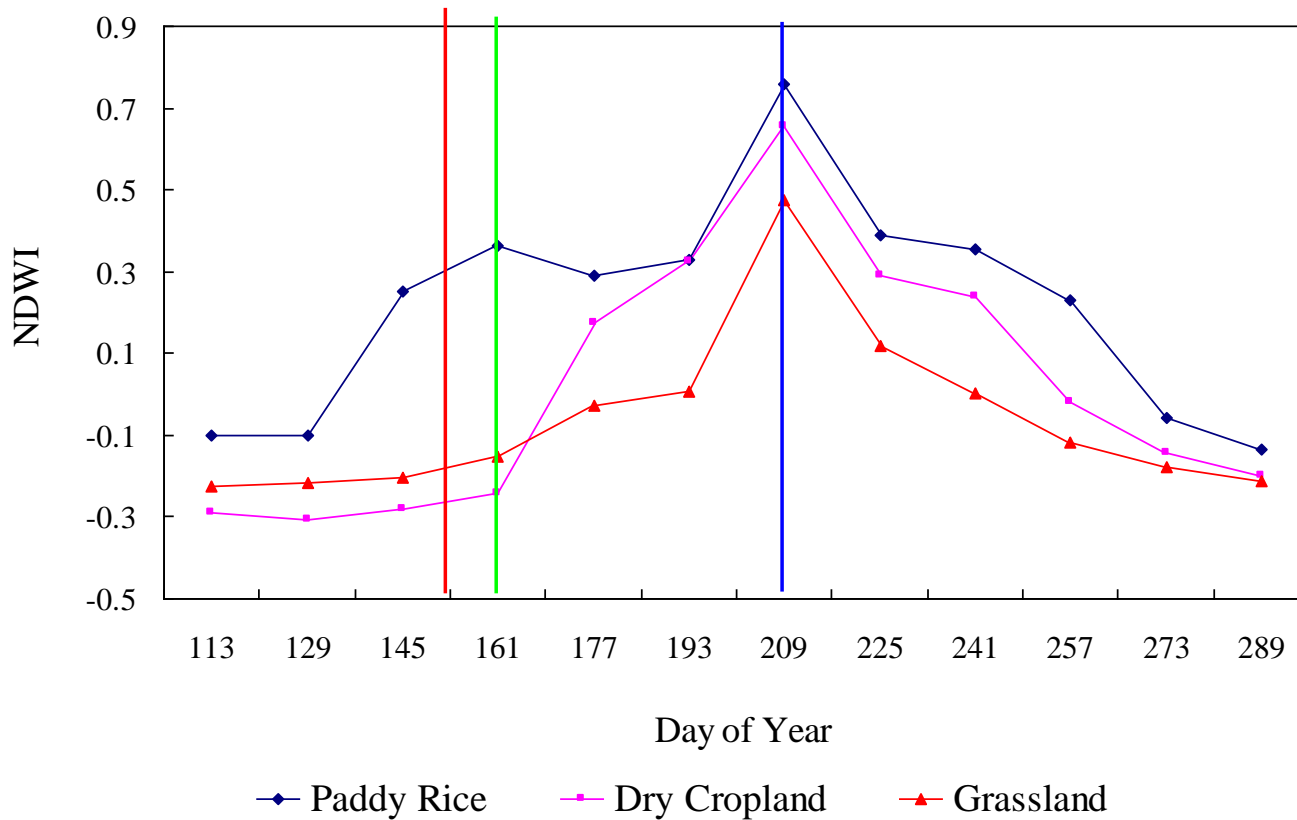


# Classification Methodology





# Classification Methodology



mid May (DOY: 145) to mid June (161) when NDWI of paddy field ranges from 0.2 to 0.4 whilst dry cropland and grassland are below 0.

$$NDWI = (B2 - B6) / (B2 + B6) \quad NDWI_{145} > 0 \text{ and } NDWI_{161} > 0.05$$





# Classification Methodology

## ● Methodology for grassland coverage grading

$$f = (NDVI - NDVI_{\min}) / (NDVI_{\max} - NDVI_{\min})$$

$$\overline{f} = (\overline{NDVI} - \overline{NDVI}_{\min}) / (\overline{NDVI}_{\max} - \overline{NDVI}_{\min})$$

$\overline{f}$  indicate the average vegetation coverage fraction

$\overline{NDVI}$  is the average vegetation index value during a year

$\overline{NDVI}_{\min}$  and  $\overline{NDVI}_{\max}$  indicate the maximum and minimum average  $\overline{NDVI}$

## Classification Criteria:

$\overline{f}$  is equal to 0.6, and  $\overline{f}_{std}$  is equal to 0.142.

Low coverage grassland:  $\overline{f} \leq (\overline{f}_{mean} - \overline{f}_{std})$

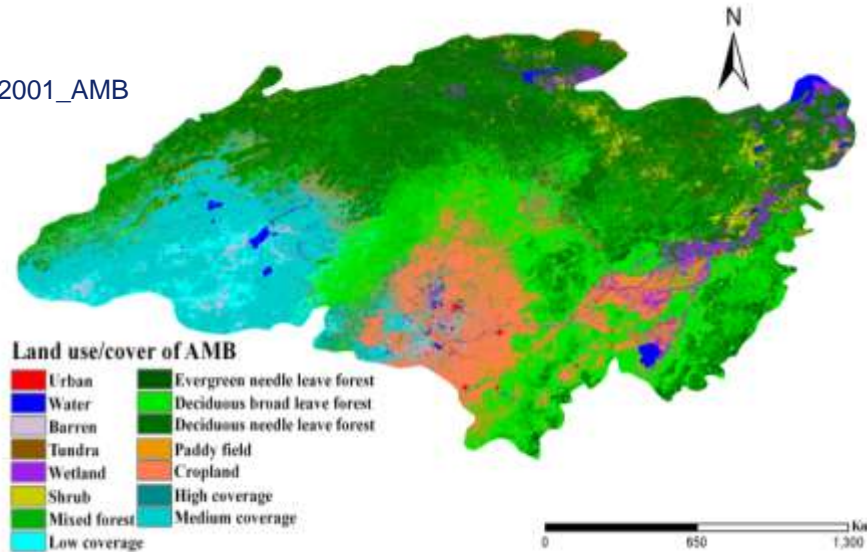
Medium coverage grassland:  $(\overline{f}_{mean} - \overline{f}_{std}) \leq \overline{f} \leq (\overline{f}_{mean} + \overline{f}_{std})$

High coverage grassland:  $\overline{f} \geq (\overline{f}_{mean} + \overline{f}_{std})$

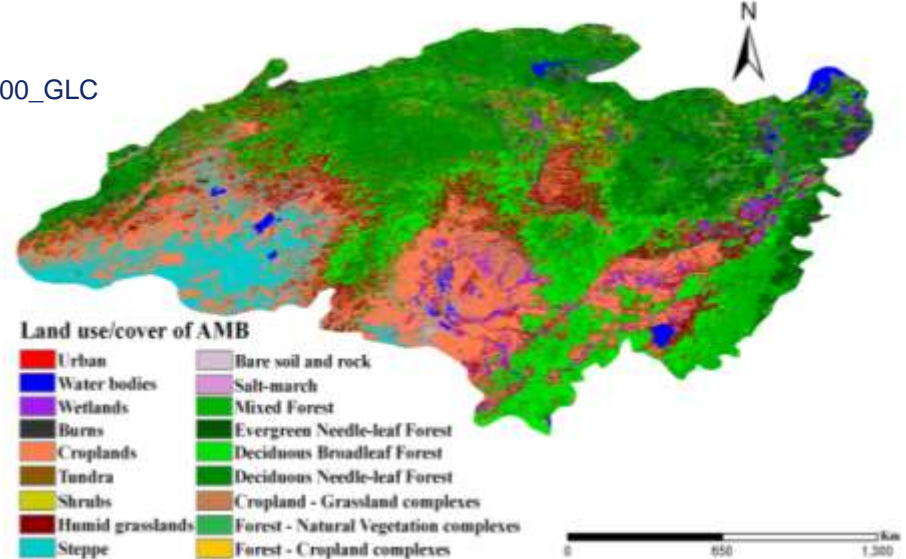


# Classification Result

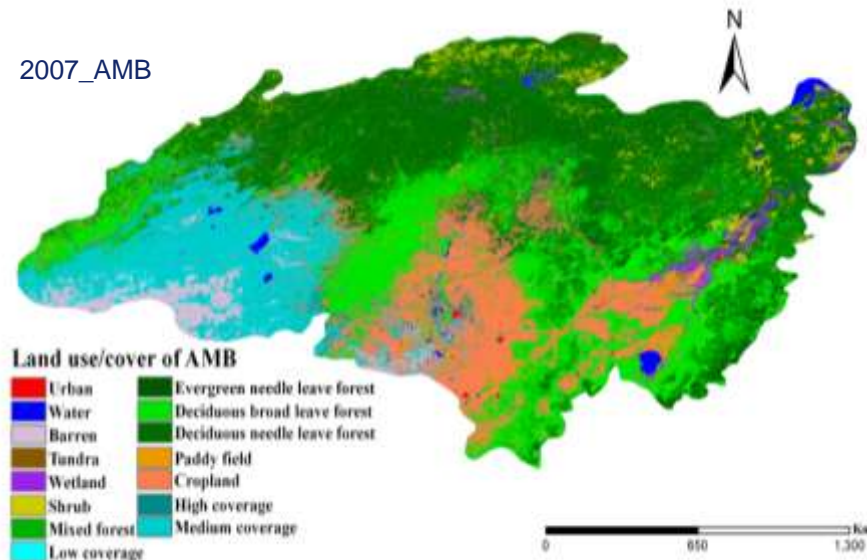
2001\_AMB



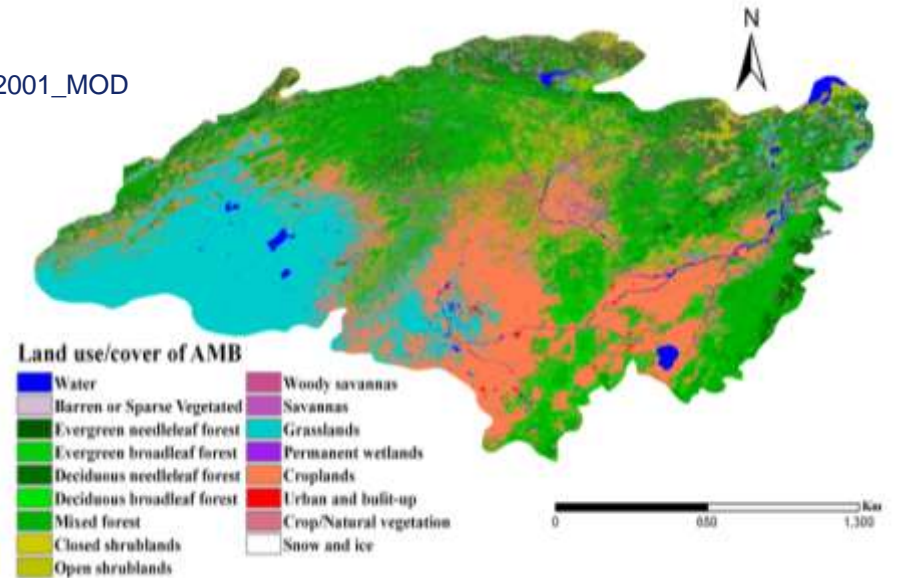
2000\_GLC



2007\_AMB



2001\_MOD





# Accuracy Assessment

## ● Pixel Scale Assessment: Chinese Part

Table.1 The confusion matrix for ARB\_2001 data sets measured by NLCD-2000

		NLCD_2000 Result Ground Truth (Pixels)							
Classes		Crop	Forest	Grass	Water	Built-up	Baren	Wetland	Total
ARB_2001 Result	Crop	3676	468	540	48	235	60	197	48.14
	<i>f</i>	48.84							

Table.2 The confusion matrix for GLC\_2000 data sets measured by NLCD-2000

		NLCD_2000 Data Set as Ground Truth (Pixels)							
Classes		Crop	Forest	Grass	Water	Built-up	Baren	Wetland	Total
ARB_2001 Result	Crop	3858	574	879	52	293	67	185	5908
	<i>f</i>	48.14							

Table.3 The confusion matrix for MOD12Q1\_2001 data sets measured by NLCD-2000

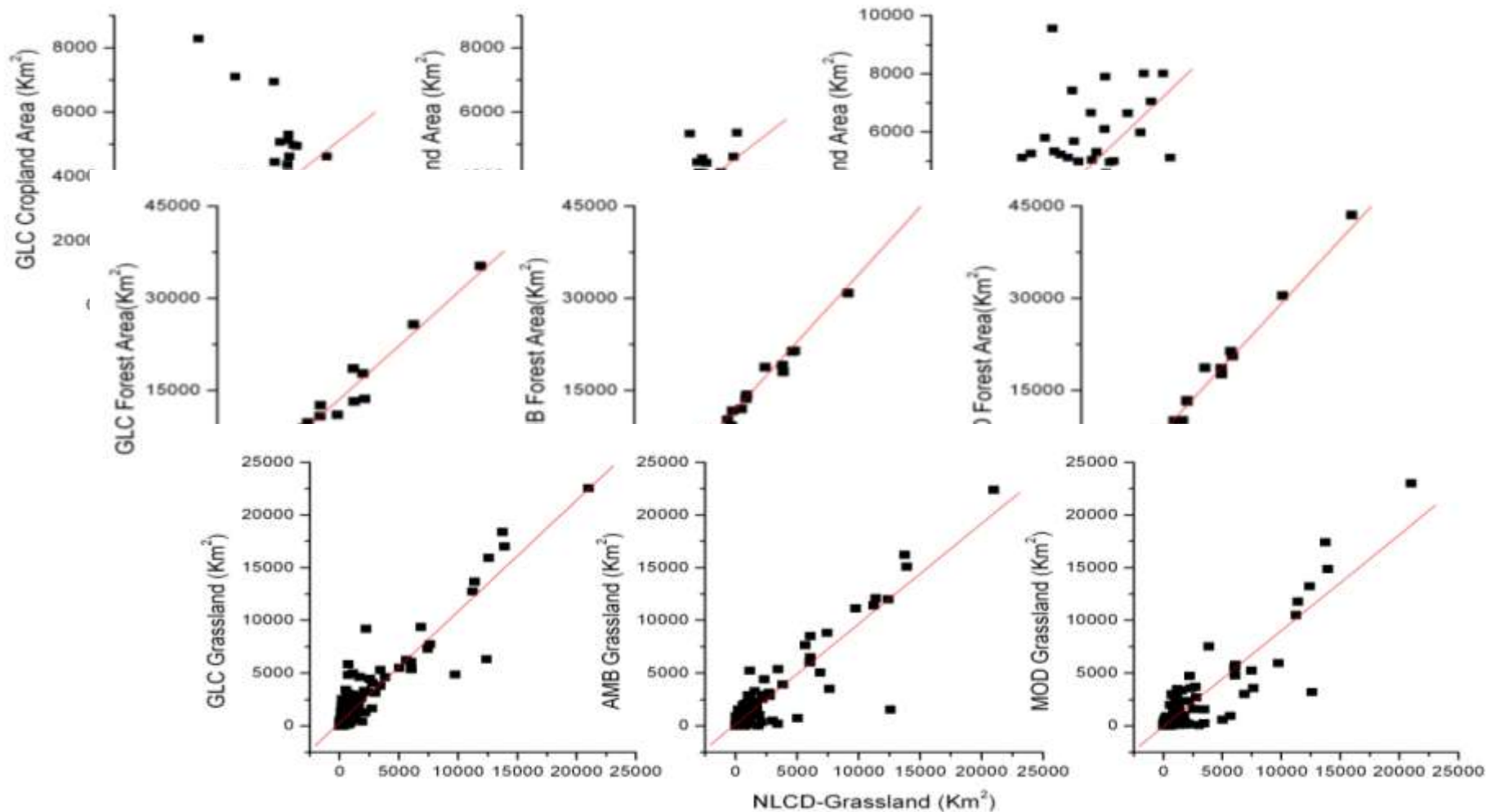
		NLCD_2000 Data Set as Ground Truth (Pixels)							
Classes		Crop	Forest	Grass	Water	Built-up	Baren	Wetland	Total
ARB_2001 Result	Crop	4688	1852	1221	84	348	188	358	8739
	<i>f</i>	46.78							
ARB_2001 Result	Forest	483	5538	937	31	40	62	140	7231
	<i>f</i>	56.94							
ARB_2001 Result	Grassland	695	561	1804	66	45	200	176	3547
	<i>f</i>	31.24							
ARB_2001 Result	Water	43	31	38	86	7	1	6	212
	<i>f</i>	21.66							
ARB_2001 Result	Built-up	53	27	8	3	39	0	1	131
	<i>f</i>	6.83							
ARB_2001 Result	Baren	7	24	19	1	0	22	3	76
	<i>f</i>	4.17							
ARB_2001 Result	Wetland	1	0	5	0	0	0	0	6
	<i>f</i>	0.00							
Total		5970	8033	4032	271	479	473	684	19942
		-							

Overall accuracy = (12177/19942)=61.06% ; Kappa Coefficient = 0.4327



# Accuracy Assessment

## ● County Scale Assessment: Chinese Part







# Accuracy Assessment

## ● Pixel Scale Assessment: Other Part

Landsat7-ETM+ Acquisition date	LULC types	Producer's Accuracy	User's Accuracy	Overall Accuracy	Kappa Coefficient
P113 R26 Sep 5, 2002	Forest	65.97	90.19	0.801	0.647
	Shrubland	98.8	82.01		
	Residential	82.61	69.33		
	Water body	53.81	58.18		
P129 R26 Sep 5, 2002	Forest	71.57	74.25	0.743	0.664
	Grassland	79.4	77.09		
	Shrubland	68.2	71.6		
P115 R23 Aug 20, 2001	Forest	78.7	73.4	0.821	0.712
	Tundra	89.4	92.09		
	Wetland	74.5	63.71		



# Accuracy Assessement

Table.6 Land use/cover distribution in the AMB region derived from different LULC data sets

LULC types	Area in $10^4 \times \text{km}^2$			
	AMB_2001	AMB_2007	MODIS_2001	GLC_2000
Residential	0.33	0.33	0.44	0.099
Shrubland	8.52	7.90	17.81	13.50
Water body	2.23	2.25	3.36	3.07
Barren	1.85	6.59	0.46	0.34
Tundra	0.81	0.63	N/A	N/A
Wetland	7.20	3.81	0.62	5.07
Grassland	40.99	36.43	44.91	42.76
Evergreen Conifer	5.59	6.85	5.65	9.43
Mixed forest	11.98	11.09	55.86	12.95
Deciduous Conifer	61.31	64.65	9.67	48.73
Deciduous Broad	45.32	43.59	25.51	33.41
Paddy field	9.30	9.41	N/A	N/A
Dryland crop	18.22	20.42	44.28	8.95
Cultivated area	N/A	N/A	N/A	39.92
Burnt tree	N/A	N/A	N/A	4.26
Savannas	N/A	N/A	18.09	N/A



# Discussions

1

The result of this study have demonstrated the potential of time series vegetation indices (NDVI and NDWI) data, reflectance data for land cover classification in the AMB region.

2

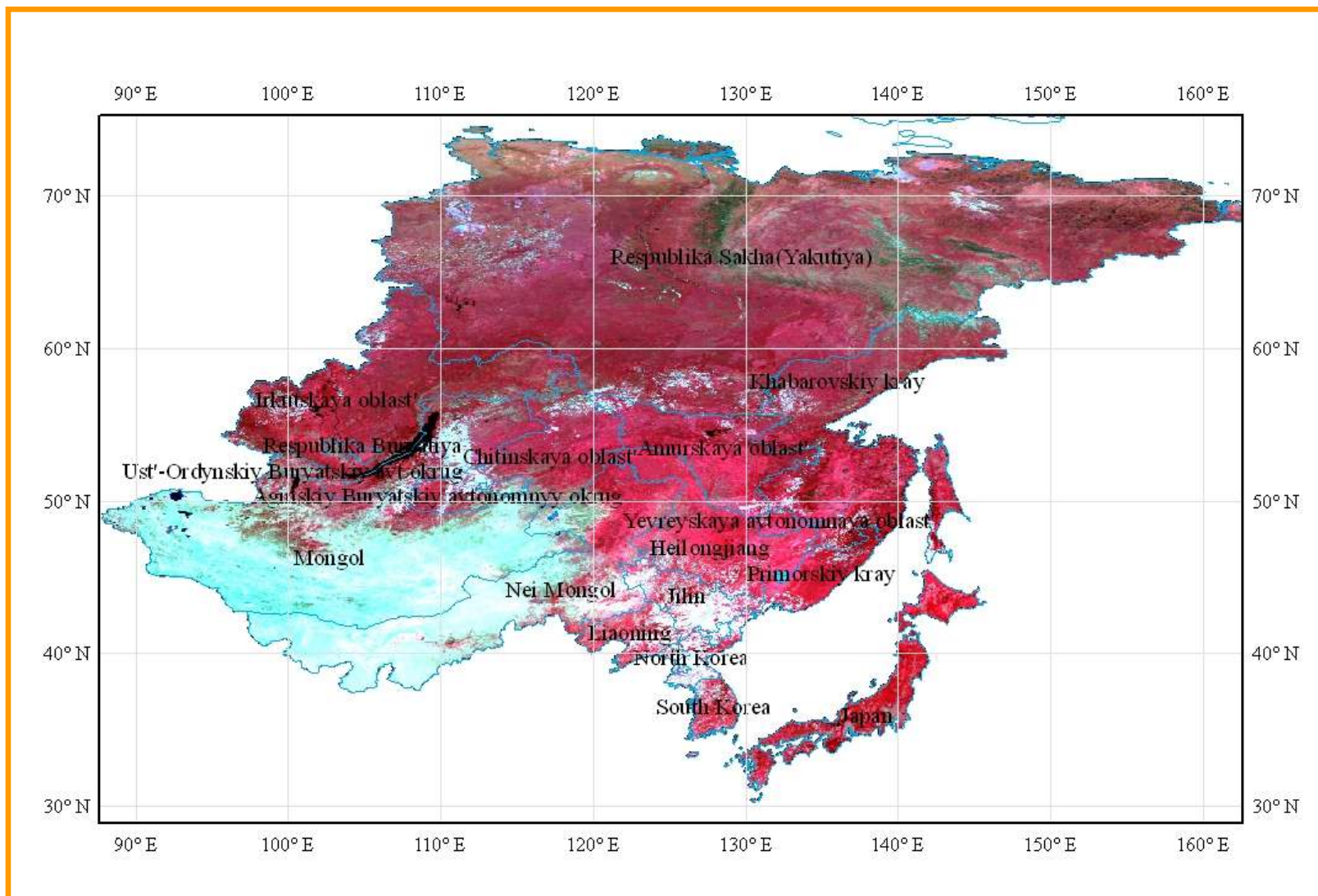
This study also highlights the potential application of MODIS reflectance time series data for some land cover recognitions, for example, sandy land and salinity affected area, which may help for more detailed spatial resolution global land cover datasets being under construction

3

Validation is still need to be done in parallel with activities for updated comparative analysis for exiting and new dataset under construction. More field works and information derived from high resolution imagery data for some LULC change hot spot area need to be generated for validation the LULC-2007.



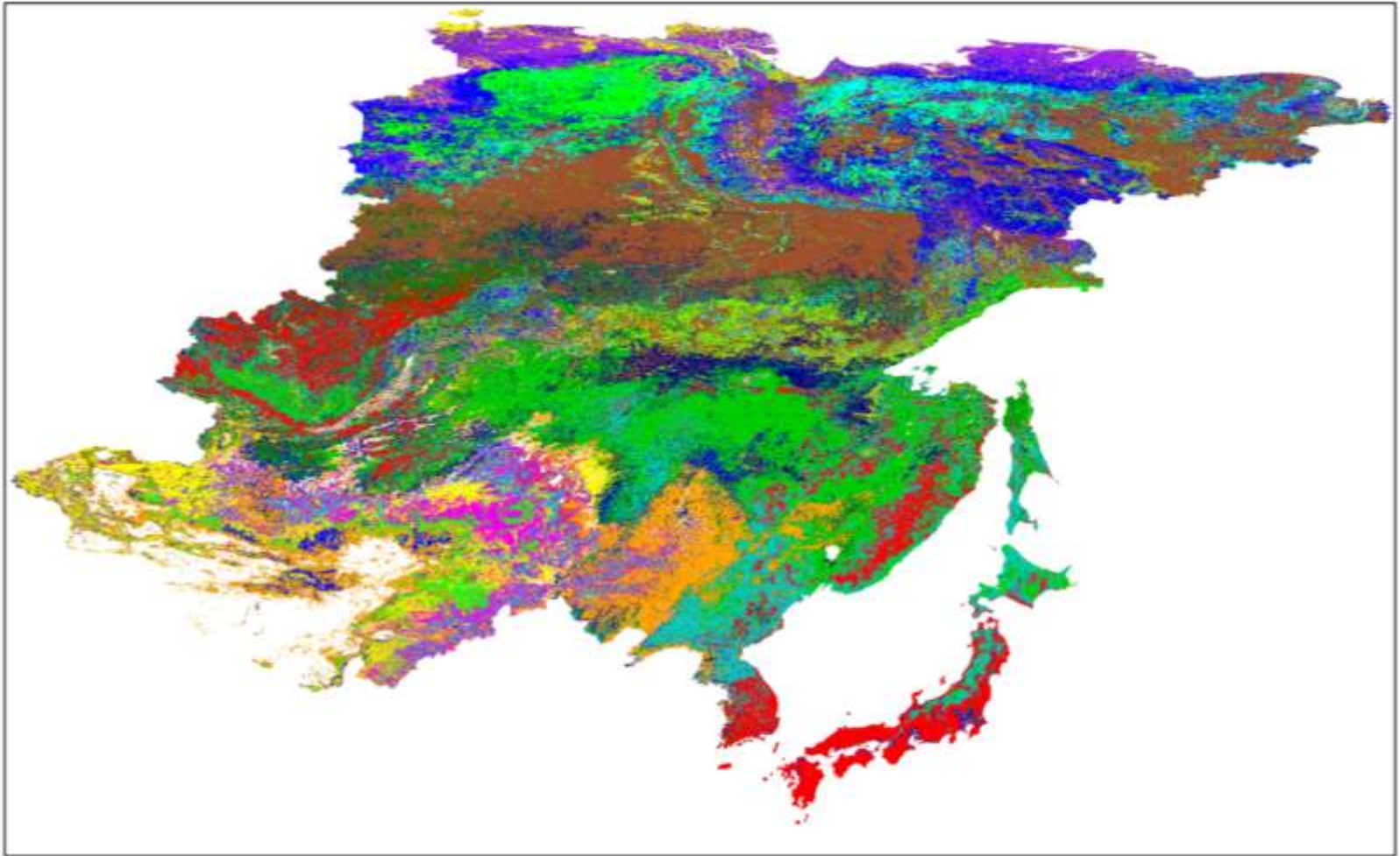
# Future Study Interest







# Future Study Interest



The Preliminary Classification Result of Northeast Asian with Modis 250m Time Series Data

A scenic landscape featuring a river in the foreground, lush green grass in the middle ground, and a range of mountains in the background under a bright blue sky with scattered white clouds. The text "Thank You for Attentions!" is overlaid in a large, bold, orange font.

# Thank You for Attentions!

Any comments and advices are welcome !