Building semantic knowledge organization systems for interdisciplinary research

----An example of wetland remote sensing
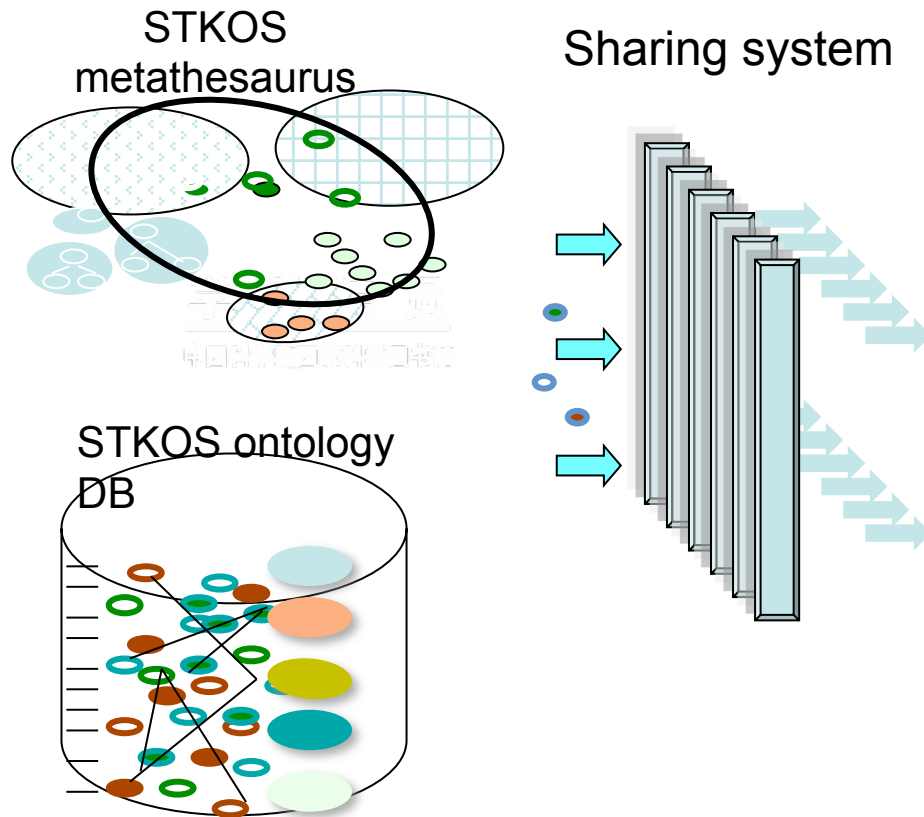
Asian NKOS workshop, Seoul,Dec,9th,2015

Outline

- **1. Introduction: background, aim**

- **2. Experimental design: workflow**

- **3. Results and discussions**

- **4. Conclusions and future work**

中国科学院国家科学图书馆
NATIONAL SCIENCE LIBRARY,CAS

● **Background**

STKOS metathesaurus

Sharing system

STKOS ontology DB

?

Interdisciplinary and emerging field：
- • brain Science
- • quantum communications
- • wetland

How to use

1. Introduction

- **Aims**

  develop the methods and workflows in building semantic knowledge organization system (KOS), in support of interdisciplinary scientific research
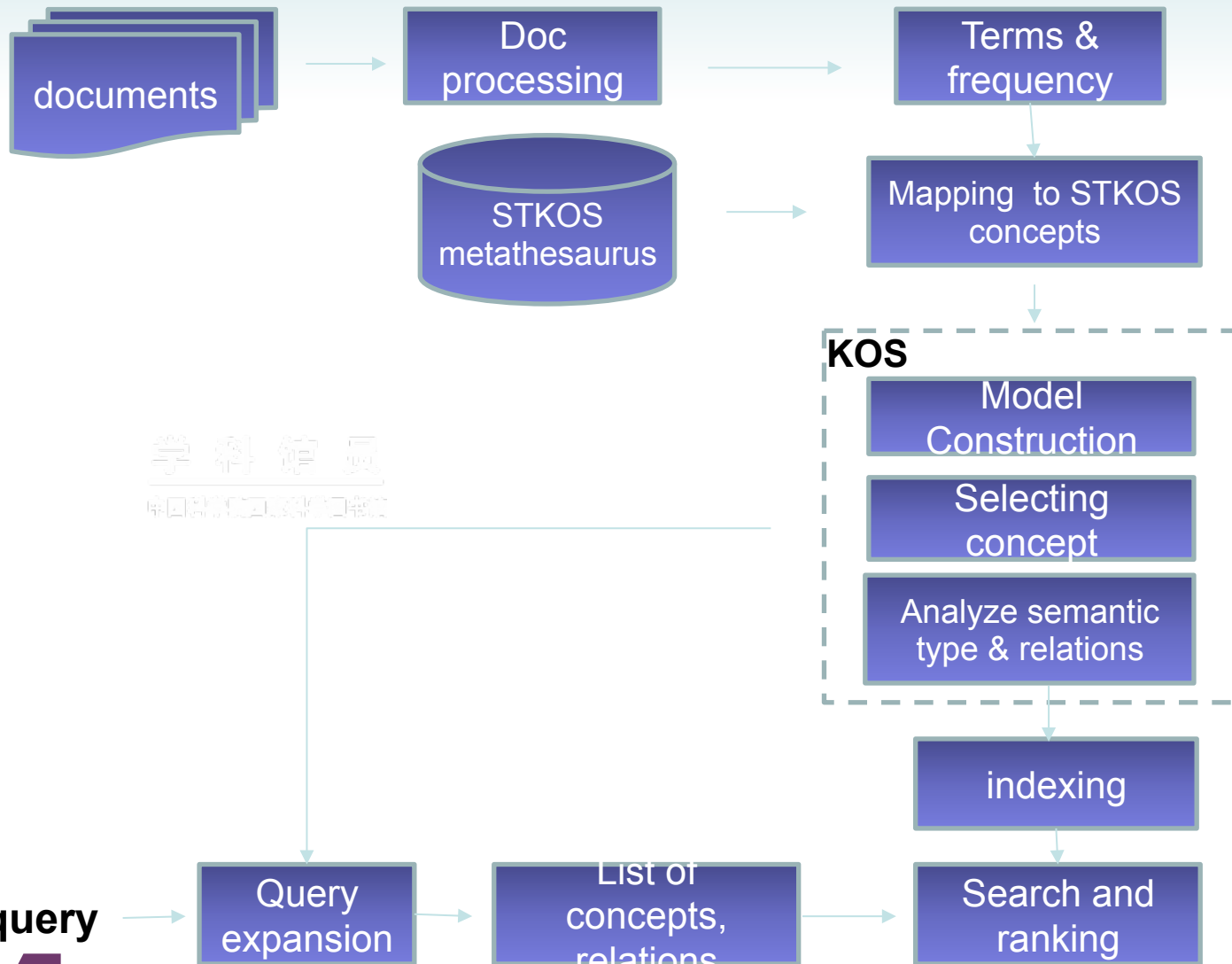
- **Experimental Field**

  - Wetland

    - remote sensing
    - Biosphere
    - Atmosphere
    - Lithosphere
    - ……

2. Workflow

```
documents  →  Doc processing  →  Terms & frequency
                                          ↓
STKOS metathesaurus  →  Mapping to STKOS concepts
                                          ↓
                          ┌─────────────────────────────┐
                          │ KOS                          │
                          │   Model Construction         │
                          │   Selecting concept          │
                          │   Analyze semantic type &    │
                          │   relations                  │
                          └─────────────────────────────┘
                                          ↓
                                      indexing
                                          ↓
query  →  Query expansion  →  List of concepts, relations  →  Search and ranking
```

学 科 馆 员

中国科学院国家科学图书馆
NATIONAL SCIENCE LIBRARY,CAS

2. Workflow

- **Data source**
  - Terms from current available KOS
    - STKOS metathesaurus: 199 KOS (150 thesauri, 3 classifications, 37 glossaries, etc.)
    - *Remote sensing lexicon*, Science publishing, 1990
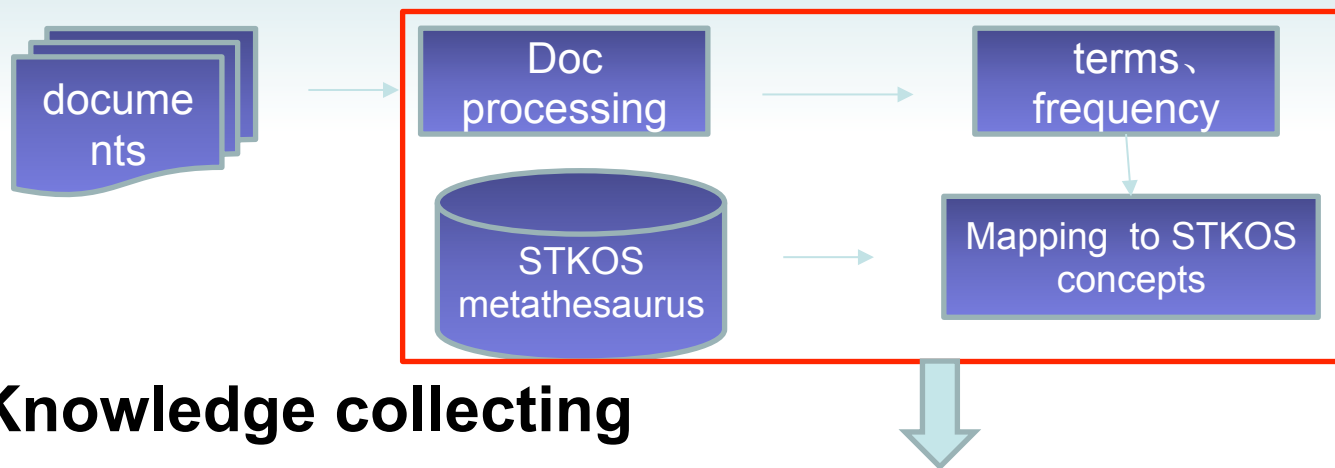 - Terms extracted from literature
    - Web of Science Database(1900-2014), Topic field = "remote sensing & wetland", bibliographic records

- **Methods**
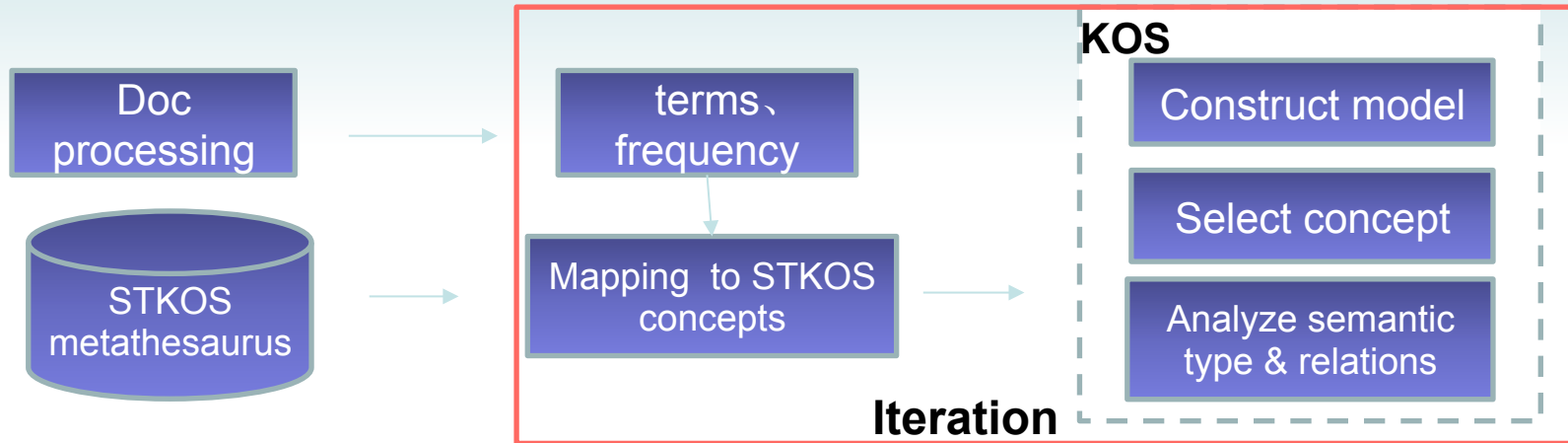 - Semantic matching
  - Domain experts participation

2. Workflow

```
documents  →  Doc processing  →  terms、frequency
                                        ↓
              STKOS          →  Mapping to STKOS
              metathesaurus      concepts
```

Knowledge collecting

- **Pre-processing / cleaning: Thomson Data Analyzer（TDA)**
  - Stop word removal, analysis，clustering

- **Semantic match to STKOS:**
  - Normalization, phrase analysis, word sense analysis, structure analysis
  - term list

- **term extraction**
  - Terms selected over a specific threshold: term frequency (TF)

  Threshold $_{term}$ =mean (TF + standard deviation TF),where high TF, medium TF,, low TF
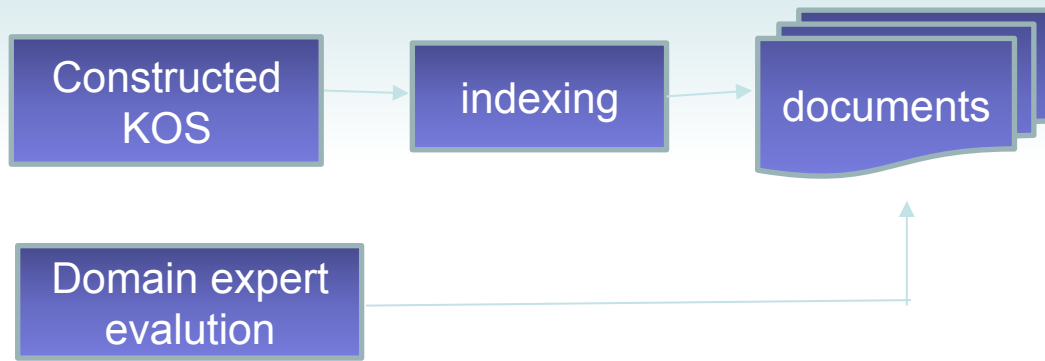
2. workflow

Construct KOS model

- **domain expert participation**
  - Reconstruct and reuse model, *Remote sensing lexicon*
  - Remove general terms, such as "classification, case study"
  - Select compound term
  - Category, Seed term, doubly-anchored patterns

2. Workflow

```
┌─────────────┐      ┌──────────┐      ┌──────────┐
│ Constructed │ ───▶ │ indexing │ ───▶ │documents │
│    KOS      │      └──────────┘      └──────────┘
└─────────────┘                              ▲
                                             │
┌─────────────┐                              │
│Domain expert│ ─────────────────────────────┘
│  evalution  │
└─────────────┘
```

- **Evaluation**
  - Annotation and index by human and computer
  - Precision, recall, how correct is the KOS learned

中国科学院国家科学图书馆
NATIONAL SCIENCE LIBRARY,CAS

3. Results and discussions

- **Match STKOS result**

frequency	term	Match STKOS	percent	Match + Standard deviation	percent
1	4271	707	14.98%	707	14.98%
2	548	178	32.48%	178	32.48%
3	196	85	43.37%	87	44.39%
4	106	57	53.77%	58	54.72%
5	65	41	63.07%	43	66.15%
6	40	21	52.5%	22	55%
7	28	19	67.86%	19	67.86%
8	25	17	68%	18	72%
9	19	14	73.68%	14	73.68%
10	11	5	45.45%	5	45.45%
11-20	47	30	63.83%	31	65.96%
21-30	9	9	100%	9	100%
31-40	8	7	87.5%	7	87.5%
41-50	6	5	83.3.%	6	100%
> 50	5	5	100%	5	100%
total	5834	1200	20.57%	1209	20.72%

Threshold $_{term}$ =4

- **Source of STKOS matching result**

source of matching stkos

3. Results and discussions

KOS model of Wetland remote sensing

- **Semantic KOS of Wetland RS**
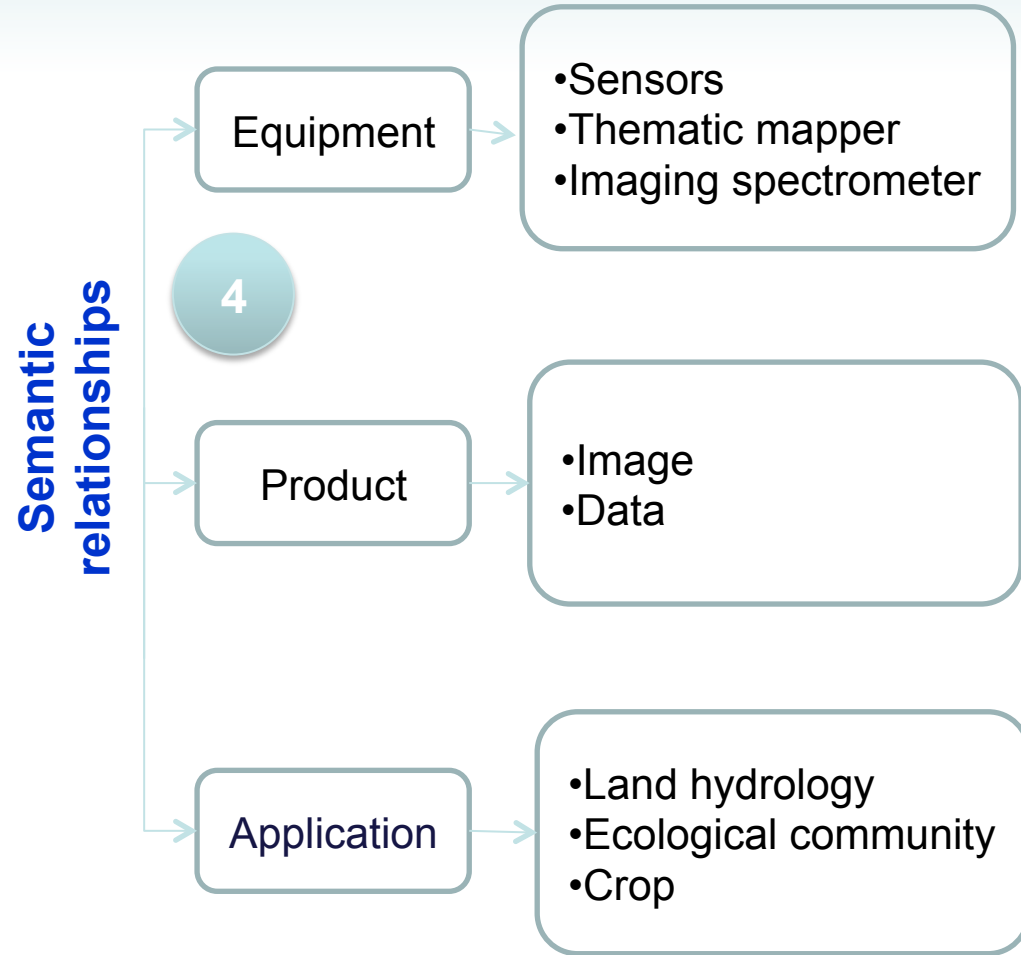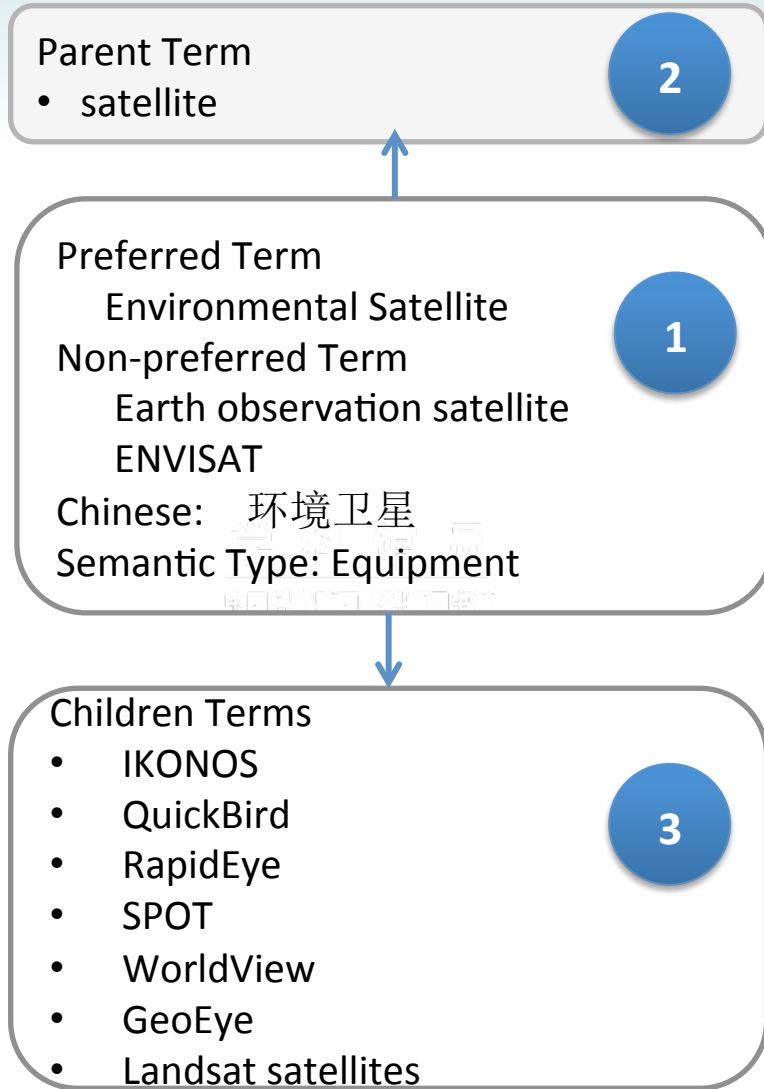 - Classification: 4 primary ,12 secondary
  - Concept：209
  - Term：409
  - Relationship：hierarchical , associative, equivalence
  - Semantic type and relation: equipment, product, concept?

    Relation: Equipment A is a kind of B

中国科学院国家科学图书馆
NATIONAL SCIENCE LIBRARY,CAS

3. Results and discussions

Parent Term
- satellite

(2)

Preferred Term
Environmental Satellite
Non-preferred Term
Earth observation satellite
ENVISAT
Chinese: 环境卫星
Semantic Type: Equipment

(1)

Children Terms
- IKONOS
- QuickBird
- RapidEye
- SPOT
- WorldView
- GeoEye
- Landsat satellites

(3)

Semantic relationships

(4)

Equipment →
- Sensors
- Thematic mapper
- Imaging spectrometer

Product →
- Image
- Data

Application →
- Land hydrology
- Ecological community
- Crop

中国科学院国家科学图书馆
NATIONAL SCIENCE LIBRARY, CAS

Evaluation

Index term, index term number, index semantic type

- Does index term represent the content in wetland RS
- Need to add some terms
- Is the article belong to wetland RS?

Computer	Article Number	Evaluation by domain experts	
Index term Number		Correct	Error
0	81	31	50
1	102	79	23
2	93	83	10
3	110	97	13
4	135	126	9
5	126	124	2
6	150	148	2
>7	765	765	0

中国科
NATION

4. Conclusions and future work

- **Conclusions**
  - Methods of interdisciplinary KOS construction
  - Cover the content of wetland remote sensing
  - Use in wetland DB
  - Improve retrieval efficiency and discover

- **Future work**
  - Semantic type revise
  - continue in other field of wetland

中国科学院国家科学图书馆
NATIONAL SCIENCE LIBRARY,CAS

Our Team♪

Liu Zheng[1], Sun Tan[2], Sheng Chunlei[3], Liu Xiumin[1]♪

1 National Science Library, Chinese Academy of Sciences♪

2 Agricultural information Institute, Chinese Academy of Agricultural Sciences♪

3 Northeast institute of Geography and Agroecology, Chinese Academy of Sciences♪

Wetland experts♪

Lv Xiangguo, Song Kaishan, Wang Zongming♪

Key laboratory of wetland ecology and environment, Northeast institute of Geography and Agroecology, CAS♪

♪

Thanks

Any questions ?

学 科 馆 员

中国科学院国家科学图书馆
NATIONAL SCIENCE LIBRARY,CAS