land cover mapping with FY3C VIRR data

Yuanyuan Wang
wangyuany@cma.gov.cn

National Satellite Meteorological Center (NSMC)
China Meteorological Administration (CMA)

AOMSUC-10, Melbourne, Australia
Dec. 4-6, 2019
Outline

1. Background
2. FY3C land cover product algorithm
3. Validation of FY3C land cover product
4. Summaries and future plans
1. Background

- Land cover is an **Essential Climate Variable** needed in many studies.

  Modelling studies
  - Land surface parameterization for GCMs.
  - Biogeochemical cycles.
  - Hydrological process.

  Carbon and ecosystem studies
  - Carbon stock, fluxes.
  - Biodiversity.

  Feed to other land surface product
  - LST, LAI, albedo, etc.
1. Background

Existing global land cover products

<table>
<thead>
<tr>
<th>Product</th>
<th>Sensor</th>
<th>Resolution</th>
<th>Input data</th>
<th>Legend</th>
<th>Classification method</th>
<th>Accuracy</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODIS LC</td>
<td>MODIS</td>
<td>500 m</td>
<td>Modis surface reflectance (channels 1–7), EVI, LST</td>
<td>IGBP (17 classes)</td>
<td>Supervised classification</td>
<td>75%</td>
<td>Friedl et al., 2010</td>
</tr>
<tr>
<td>GlobCover</td>
<td>MERIS</td>
<td>300 m</td>
<td>Bi-monthly reflectance composites 15 channels</td>
<td>FAO LCCS (22 classes)</td>
<td>Unsupervised/Supervised Clustering</td>
<td>73%</td>
<td>Defourny et al., 2007</td>
</tr>
<tr>
<td>VIIRS</td>
<td>VIIRS</td>
<td>1000 m</td>
<td>Global annual metrics</td>
<td>IGBP (17 classes)</td>
<td>Supervised classification</td>
<td>78%</td>
<td>Zhan et al., 2018</td>
</tr>
<tr>
<td>GLC2000</td>
<td>SPOT-VEGETATION</td>
<td>1000 m</td>
<td>Daily mosaic of 4 spectral channels</td>
<td>FAO LCCS (23 classes)</td>
<td>Generally Unsupervised, depending on the partner</td>
<td>68.6%</td>
<td>Mayaux et al., 2006</td>
</tr>
<tr>
<td>CCI</td>
<td>MERIS</td>
<td>300 m</td>
<td>7-day composite reflectance</td>
<td>FAO LCCS (22 classes)</td>
<td>Unsupervised/Supervised classification</td>
<td>73.2%</td>
<td>Defourny et al. (2016)</td>
</tr>
<tr>
<td>GLCNMO</td>
<td>MODIS</td>
<td>500 m</td>
<td>16-day NBAR (1–7), NDVI, DMSP-OLS, LCCS (20 classes)</td>
<td>Supervised classification</td>
<td></td>
<td>77.9%</td>
<td>Tateishi et al. (2014)</td>
</tr>
</tbody>
</table>
Outline

1. Background
2. FY3C land cover product algorithm
3. Validation of FY3C land cover product
4. Summaries and future plans
2. FY3C land cover product

- Main properties of FY3C land cover
  
  Legend: **IGBP 17 classes**
  
  Data input: **10-day compositing VIRR/NDVI products of one year.**
  
  NDVI products provide not only 10-day maximum NDVI but also the corresponding channel information (red, nir, swir, MIR, TIR1, TIR2).

  Spatial resolution: **1 km for FY3C; 250 m for FY3D**
  
  Classification algorithm: **Boosting decision tree algorithm (supervised learning)**

  Accuracy expected: **~70%**
Flowchart of FY3 Land cover algorithm

- NDVI 10-day compositing products of one year (including NDVI and corresponding information for 6 channels: red, NIR, SWIR, MIR, TIR1, TIR2)

  - Quality control
    - Features extraction

  - Ancillary data (land, water mask)
    - Features extracted

  - Training data
    - Boosting C4.5
      - Training
        - Trained classifier
          - Classification results

    - Special classes processing (permanent snow/ice, urban)
      - Post-processing
        - Adjustment with geographic rules
Features extraction from NDVI time series

• Why extract features?
  (1) Increase data quality
  *Original NDVI time-series is noisy due to various reasons (cloud/shadow contamination, sensor mal-function...*)

  (2) Increase separability between classes and decrease scatter within class;
  *Original NDVI time-series is not optimum with regard to land cover type separability. Extracted metrics can better reflect vegetation phenology and other surface characteristics.*

  (3) Decrease dimensionality and redundancy
  *Original data is 36 periods * 7 features (NDVI, red, nir, swir, MIR, TIR1, TIR2). High dimensionality leads to low efficiency and probably low accuracy.*
Features extraction from NDVI time series

• What features are extracted for vegetation?

(1) Features when vegetation reaches its highest vigor (n=7).

Maximum NDVI and the corresponding channel information (NDVI, red, nir, swir, MIR, TIR1, TIR2).

(2) Features from growing season (n=9). The growing season is defined as the period showing the 24 highest 10-day compositing NDVI values.

Minimum, mean, amplitude of NDVI, Red, NIR and SWIR.

(3) Features from senescent season of tropical area (n=7). The senescent season is defined as the period showing the 6 highest 10-day compositing TIR1 (10.4 um) values.

Mean values of NDVI, red, NIR, SWIR, MIR, TIR1, TIR2
Features extraction from NDVI time series

• What features are extracted for snow and barren land?

(1) To classify permanent snow/ice, we extract median reflectance of red, nir, and swir of a year and calculate NDSI (Normalized Difference Snow Index)

\[ NDSI = \frac{(red - swir)}{(red + swir)} \]

(2) To classify barren land cover type, we extract red, nir, and swir when TIR1 (10.4 um) is the maximum, and calculate NDBI (Normalized Difference Build-up Index)

\[ NDBI = \frac{(swir - nir)}{(swir + nir)} \]
False color image when vegetation reaches its highest vigor in 2008
(R: swir, G: nir, B: red)
False color image for the senescent season in 2008.

(R: swir, G: nir, B: red)
False color image of medium reflectance in 2008

(R: swir, G: nir, B: red)
Minimum NDVI of growing season in 2008.
Maximum NDVI of growing season in 2008.
Flowchart of FY3 Land cover algorithm

- NDVI 10-day compositing products of one year (including NDVI and corresponding information for 6 channels: red, NIR, SWIR, MIR, TIR1, TIR2)
- Quality control
- Features extraction
- Ancillary data (land water mask)
- Features extracted
- Training data
- Boosting C4.5 Training
- Trained classifier
- Classification results

Special classes processing (permanent snow/ice, urban)
Post-processing
Adjustment with geographic rules
Training data collection

- Representativeness and credibility of training samples are very important to supervised classification results.
- Sample collection method:
  
  *Visual interpretation.*
  *Referencing existing land cover products (MODIS, GlobCover).*
  *Referencing high spatial resolution imagery.*
Number of training sample collected for each class.
Flowchart of FY3 Land cover algorithm

- NDVI 10-day compositing products of one year (including NDVI and corresponding information for 6 channels: red, NIR, SWIR, MIR, TIR1, TIR2)
- Ancillary data (land water mask)
- Features extracted
- Quality control
- Features extraction
- Training data
- Special classes processing (permanent snow/ice, urban)
- Adjustment with geographic rules
- Classification results
- Post-processing
- Trained classifier
- Boosting C4.5
- Training
Boosting Decision Tree algorithm

Merits:

- High accuracy
- Fast speed
- Non-linear classifier
- Non-parametric
- Can process both numeric and categorical variables
- Robust to missing values
- Output confidence information......
Illustration of Decision Tree algorithm

Input data is recursively partitioned into subspaces by maximizing the gain criterion.
Illustration of Boosting Decision Tree algorithm

Training several decision trees in a sequential order.

The first tree is trained by assigning equal weights to all instances in the training data.

All the subsequent trees are trained for reweighted data (the weights of correctly classified instances are increased, and the weights for misclassified instances are decreased).
Flowchart of FY3 Land cover algorithm

- NDVI 10-day compositing products of one year (including NDVI and corresponding information for 6 channels: red, NIR, SWIR, MIR, TIR1, TIR2)
- Ancillary data (land water mask)
- Quality control
  - Features extraction
- Features extracted
- Training data
- Boosting C4.5 Training
- Trained classifier
- Classification results
- Post-processing
- Special classes processing (permanent snow/ice, urban)
- Adjustment with geographic rules
Post classification

• Optimizing urban classification results by using density of constructed impervious surfaces derived from nighttime lights imagery.

  Urban class has a high commission error. A pixel identified as urban must have a high fraction impervious surfaces, otherwise, it is labeled as the class predicted by the classifier trained without urban training samples.

• Regional relabeling according to geographic rules. About 1.12% pixels are relabeled.

  Rule examples:
  - cropland above boreal zone are mapped to grassland;
  - evergreen broadleaf forest in temperate latitudes are mapped to evergreen needleleaf forest.

  ......
An example of reducing commission error of urban class by using impervious surface density data (west coast of North America)
Flowchart of FY3 Land cover algorithm

NDVI 10-day compositing products of one year (including NDVI and corresponding information for 6 channels: red, NIR, SWIR, MIR, TIR1, TIR2)

Ancillary data (land water mask)

Features extracted

Quality control Features extraction

Training data

Boosting C4.5 Training

Trained classifier

Classification results

Special classes processing (permanent snow/ice, urban)

Post-processing

Adjustment with geographic rules
FY3C land cover product (2018)
Outline

1. Background
2. FY3C land cover product algorithm
3. Validation of FY3C land cover product
4. Summaries and future plans
Compare with MODIS land cover

Similar global pattern with obvious differences in boreal forest region and semi-arid region in north America, the Sahel region, etc.
Compare with VIIRS land cover

**VIIRS land cover (2017)**

**FY3C land cover (2018)**

*Similar global pattern with obvious differences in the Sahel region, Northeast Asia, western Australia, etc.*
Compare with validation samples

- Based on validation samples (n=10928) provided by the GOFC-GOLD reference data portal, the overall accuracy and kappa coefficient of FY3A land cover product is determined to be **69.31%** and **66.09%**, respectively.
(1) FY has higher producer’s accuracies for Evergreen Needleleaf forest, Deciduous needleleaf forest, cropland/natural mosaic.

(2) MODIS has higher user’s accuracies for Deciduous needleleaf forest, dense shrub.

(3) For other classes, the two products have similar accuracies.
Outline

1. Background
2. FY3C land cover product algorithm
3. Classification results comparison and validation
4. Summaries and future plans
Summaries

◆ Land cover data of good quality can be derived from FY3C VIRR/NDVI data.

◆ For some land cover types, there is still a large room for improvements in classification. *(dense shrub, Woody Savanna, Savanna, cropland/natural mosaic, cropland)*
Future plans

◆ Improve classification results further by extracting more discriminative features and fusing ancillary data.

◆ Enlarge existing training dataset by referencing regional classification map with high spatial resolution (~100 m).

◆ Use Markov models to stabilize classification results across years.

◆ Adopt LCCS classification system which has been adopted by MODIS V6.0 land cover product.

◆ Produce classification map based on 250 m FY3D/MERSI data.
False color image of mean reflectance for growing season in 2019.  
(FY3D/MERSI, 250 m)
True color image when vegetation reaches its highest vigor in 2019 (FY3D/MERSI, 250 m)
Make the data better and easier to use!