Distinguishing Cloud and Haze by Satellite Imagery Using Fuzzy Clustering Method

Yizhou Fan*, Fuding Xie*, Zhanqing Li

Earth System Science Interdisciplinary Center, University of Maryland, College Park

*Visiting scholar
12/1/2016
Background

About haze episode

What is haze

- Haze is pollution phenomenon caused by suspend fine particles
- Deterioration of visibility, influencing transportation
- PM2.5 particles harming human health

Haze in China

- Continued increasing in last 50 years
- Rapidly increasing in 21th century
- High attention paid by government environment division

Annual haze days in China during 1960s and 2000s

- Continued increasing in last 50 years
- Rapidly increasing in 21th century
- High attention paid by government environment division
Background

Haze and cloud in satellite imagery

Conventional cloud detection algorithms

- Infrared brightness temperature thresholds and difference tests
  \[ BT_{11} < 270K \]
  \[ BT_{8.6} - BT_{11} > -3.7 - 1.6\ln(PW) \]
  \[ BT_{11} - BT_{12} < -0.51 + 0.49\ln(PW) \]

- Non-cloud obstruction flag and suspend dust flag
  \[ BT_{11} - BT_{12} > -1K \]
  \[ BT_{3.75} < 350K \]
  \[ R_{2.1} > 0.20 \]

- Near infrared cirrus test
  \[ R_{1.38} > 0.035 \]

- Spatial variability tests (new for MODIS collection-6 aerosol products)
  \[ \rho_{0.47} > 0.4 \]
  \[ \sigma_{0.47}^* > 0.0025, \sigma_{0.47} > 0.0075 \]
  \[ \rho_{1.38} > 0.025, \sigma_{1.38} > 0.003 \]
Background

Space & possibilities of optimization

- Uses too many fixed threshold tests.
  - Thresholds need to be clear-cut and explicit, and need to be updated according to new observation findings.
  - Does not lead to uniquely good result in different scenarios.
- Hardly uses information other than pixel reflection.
  - Ignores many important characteristics that help distinguishing cloud and haze.
- This leads our goal being
  - Decreasing the number of fix threshold tests.
  - Decreasing the use of unnecessary bands, while considering more features of cloud and haze.
  - Confirming good results in different local cloud-haze interacting scenes.
Fuzzy clustering method

- **Machine learning** gives computers the ability to learn without being *explicitly* programmed.
- **Clustering** involves assigning data points to clusters, or homogeneous classes.
- In **fuzzy** clustering, each data point can belong to more than one cluster.
Fuzzy c-means clustering

- In **Centroid-based** clustering method, clusters are represented by a central vector.
- FCM is based on the minimization of objective function,
  \[ J_m = \sum_{i=1}^{D} \sum_{j=1}^{N} \mu_{ij}^m ||x_i - c_j||^2 \]
- Possibility, or the extent to which each data point belongs to each cluster is measured by **membership matrix**. Its element \( \mu_{ij} \) means the degree of membership of \( i \)th point in \( j \)th cluster.
- \( m \) controls the degree of **fuzzy overlap**, which refers how fuzzy the boundaries between clusters are.
Fuzzy c-means clustering

- FCM can be performed in following iterative way.
  1. Randomly initialize the cluster membership values, $\mu_{ij}$.
  2. Calculate the cluster centers:
     \[
     c_j = \frac{\sum_{i=1}^{D} \mu_{ij}^m x_i}{\sum_{i=1}^{D} \mu_{ij}^m}.
     \]
  3. Update $\mu_{ij}$ according to the following:
     \[
     \mu_{ij} = \frac{1}{\sum_{k=1}^{N} \left( \frac{\| x_i - c_j \|}{\| x_i - c_k \|} \right)^{2/(m-1)}}
     \]
  4. Calculate the objective function, $J_m$.
  5. Repeat steps 2–4.
Method

Characteristics of cloud and haze

• A cloud is...
  • White
  • Opaque in visual and infrared bands
  • In high altitude
  • Rich in spatial texture, spatially rough
  • Often moving quickly
• A haze is...
  • Grey, brown, yellow
  • Semi-transparent in visual bands, transparent in infrared bands.
  • In low altitude, limited to terrain features
  • Less in spatially texture, spatially smooth
  • Often moving slowly
Characteristics of cloud and haze

Method

- Each point is clustered according to
  - Thermal infrared bands (860nm, 1240nm).
  - Spatial texture in bands above.
  - Terrain feature underneath.
  - Temporal image series.

If \( p \) features are used, each point is seen as a point in a \( p \)-dimensional space, such that
\[
\mathbf{x}_i = (R_{1i}, R_{2i}, \ldots R_{pi})
\]

Euclidean distance between 2 \( p \)-dimensional points is
\[
\| \mathbf{x}_i - \mathbf{x}_j \| = \sqrt{\sum_{k=1}^{p} (R_{ki} - R_{kj})^2}
\]
Result

Cloud boundary & mask comparison

RGB scene
Cloud mask by Deep Blue
Cloud contour by Deep Blue

Histogram of cloud fuzzy partition
Cloud fuzzy partition
Cloud contour by FCM, threshold = 0.8
Result

Cloud mask & surface data comparison

Cloud mask by Deep Blue
Red dots = Surface stations with RH<80% and VIS<10km. Blue dots = other stations

Cloud fuzzy partition
Red dots = Surface stations with RH<80% and VIS<10km. Blue dots = other stations
Cloud mask & surface data comparison
Result

Cloud mask & pm2.5 data comparison
AOD data synergy and haze region
Conclusion and future work

• Fuzzy clustering method can learn from characteristics of haze region, hence distinguish haze and cloud from satellite imagery.
• Distinguishing results are plausible comparing to station & satellite auxiliary data.
• Haze region can be retrieved, filling the blank area of the nullification of AOD retrieval algorithms.

• Introduce more features to classify cloud and haze.
• Develop a systematic quantitative method for result validation.
Thank you