

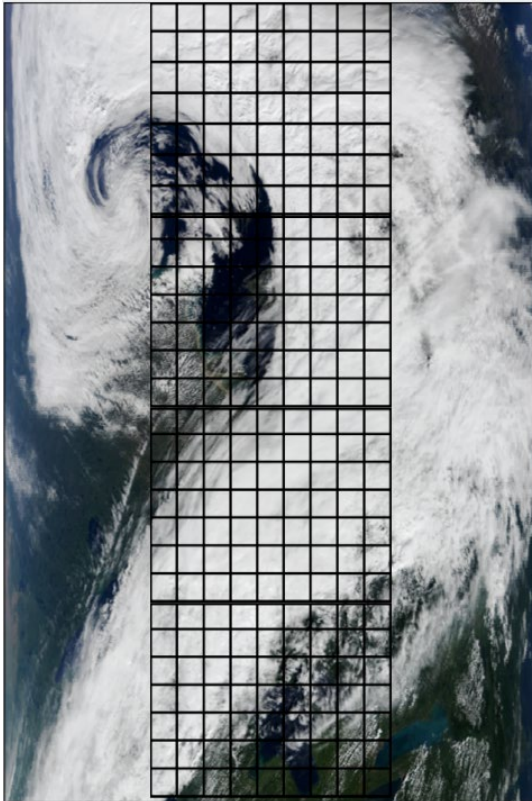
MODIS Cloud Mask Machine Learning Data

File naming scheme

- MODIS_MLDATA_Shape_64x64_2018300.1430_.hdf
 - o 2018300.1430 = YYYYDDD.HHMM
 - YYYY= year/DDD= Julian day/HHMM hour and minute in UTC time.

Inside the file

- Each file contains n images which are a non-overlapping subset of the source image



- o ClassificationAccuracy
 - -1 not classified
 - 1 good training data candidate
 - 0 bad training data candidate
- o FeatureLabels
- o 43 features per pixel. Labels the features on axis 0 of ImageFeatures
- o ImageClassification
 - “Ground truth” from TERRA MODIS cloud mask algorithm
- o ImageFeatures
 - 43x64x64 size array containing 43 features at each pixel in 64x64 pixel image
 - Axis 0: features
 - Axis 1 and 2: image length and width

How to read the file in python

Documentation link: <http://docs.h5py.org/en/stable/build.html>

- "Conda install h5py" in terminal to install package
- import h5py
- hf = h5py.File("MODIS_MLDATA_Shape_64x64_2018300.1430_.hdf", "r")
 - o saves file object into hf variable
- Read in data as numpy arrays
 - o ClassificationAccuracy = hf["Image_000/ClassificationAccuracy"][(0)]
 - o FeatureLabels = hf["Image_000/FeatureLabels"][(0)]
 - o ImageClassification = hf["Image_000/ImageClassification"][(0)]
 - o ImageFeatures_ = hf["Image_000/ImageFeatures"][(0)]