Data Mining for Climate Change Model Intercomparison and Phenoregions

Forrest M. Hoffman†, Jitendra Kumar†, and William W. Hargrove‡

†Oak Ridge National Laboratory and ‡USDA Forest Service, Eastern Forest Environmental Threat Assessment Center (EFETAC)

National Phenology Network (NPN) Seasonal Timing Working Group Meeting

August 12–13, 2013
Data Mining for Climate Change Model Intercomparison

Hoffman et al. (2005)
Global LAI for 47 CMIP5 Simulations Compared to MODIS

CMIP5 Model Monthly Mean Leaf Area Index (2000−2009)

(Hoffman et al., in prep.)
Zonal LAI for 47 CMIP5 Simulations Compared to MODIS

<table>
<thead>
<tr>
<th>CMIP5 Model DJF Zonal Mean Leaf Area Index (2000–2009)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latitude (degrees)</td>
</tr>
<tr>
<td>Leaf Area Index (m² m⁻²)</td>
</tr>
</tbody>
</table>

(Hoffman et al., in prep.)
Zonal LAI for 47 CMIP5 Simulations Compared to MODIS

CMIP5 Model JJA Zonal Mean Leaf Area Index (2000–2009)

Leaf Area Index (m² m⁻²)

Latitude (degrees)

(Hoffman et al., in prep.)
Hoffman and Hargrove previously used $k$-means clustering to detect brine scars from hyperspectral data (Hoffman, 2004) and to classify phenologies from monthly climatology and 17 years of 8 km NDVI from AVHRR (White et al., 2005).

This data mining approach, using high performance computing, was applied to the entire body of the high resolution MODIS NDVI record for the continental U.S.

> 80B NDVI values, consisting of $\sim 146.4$M cells for the CONUS at 250 m resolution with 46 maps per year for 12 years (2000–2011), analyzed using $k$-means clustering.

The annual traces of NDVI for every year and map cell are combined into one 323 GB single-precision binary data set of 46-dimensional observation vectors.

Clustering yields 12 maps in which each cell is classified into one of $k$ phenoclasses, and phenoregions form representative prototype annual NDVI traces.
50 Phenoregions for year 2011 (Random Colors)
50 Phenoregion Prototypes (Random Colors)

.cluster

Cluster 31 Cluster 40 Cluster 28 Cluster 25 Cluster 10 Cluster 3 Cluster 2 Cluster 7 Cluster 22 Cluster 6
Cluster 4 Cluster 42 Cluster 5 Cluster 13 Cluster 41 Cluster 39 Cluster 15 Cluster 48 Cluster 21 Cluster 30
Cluster 11 Cluster 49 Cluster 46 Cluster 32 Cluster 45 Cluster 8 Cluster 47 Cluster 16 Cluster 26 Cluster 38
Cluster 27 Cluster 35 Cluster 29 Cluster 14 Cluster 20 Cluster 50 Cluster 37 Cluster 33 Cluster 12 Cluster 18
Cluster 36 Cluster 34 Cluster 24 Cluster 44 Cluster 1 Cluster 23 Cluster 43 Cluster 19 Cluster 9 Cluster 17

day of year
50 Phenoregions Mode (Random Colors)
50 Phenoregions Max Mode (Random Colors)
50 Phenoregions Max Mode (Similarity Colors)
National Phenological Ecoregions (2000–2011)

William W. Hargrove, Forrest M. Hoffman, Jiendra Kumar, Joseph P. Spruce, and Richard T. Mills
January 14, 2013

Jump to 50 National Phenoregions
Jump to 100 National Phenoregions
Jump to 200 National Phenoregions
Jump to 500 National Phenoregions
Jump to 1000 National Phenoregions
Jump to 5000 National Phenoregions

50 Most-Different National Phenological Ecoregions (2000–2011)
This research was sponsored by the U.S. Department of Energy’s Biological and Environmental Research (BER) program and the U.S. Department of Agriculture Forest Service, Eastern Forest Environmental Threat Assessment Center (EFETAC). This research used resources of the National Center for Computational Science at Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.
F. M. Hoffman. Analysis of reflected spectral signatures and detection of geophysical
disturbance using hyperspectral imagery. Master’s thesis, University of Tennessee,
Department of Physics and Astronomy, Knoxville, Tennessee, USA, Nov. 2004.

F. M. Hoffman, W. W. Hargrove, D. J. Erickson, and R. J. Oglesby. Using clustered
climatic regimes to analyze and compare predictions from fully coupled general