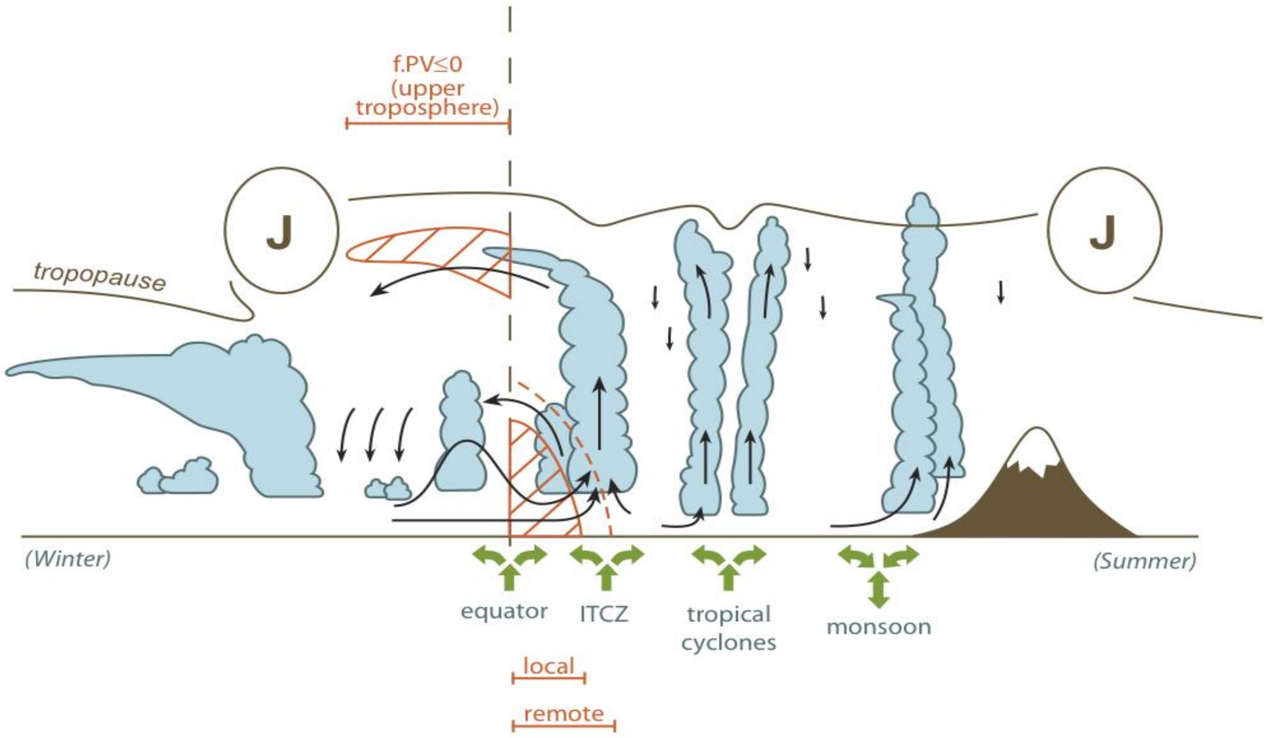
A Data Mining Algorithm for Satellite Data: Application for Double ITCZs

Andrew Geiss[1,2] and Gad Levy[1]

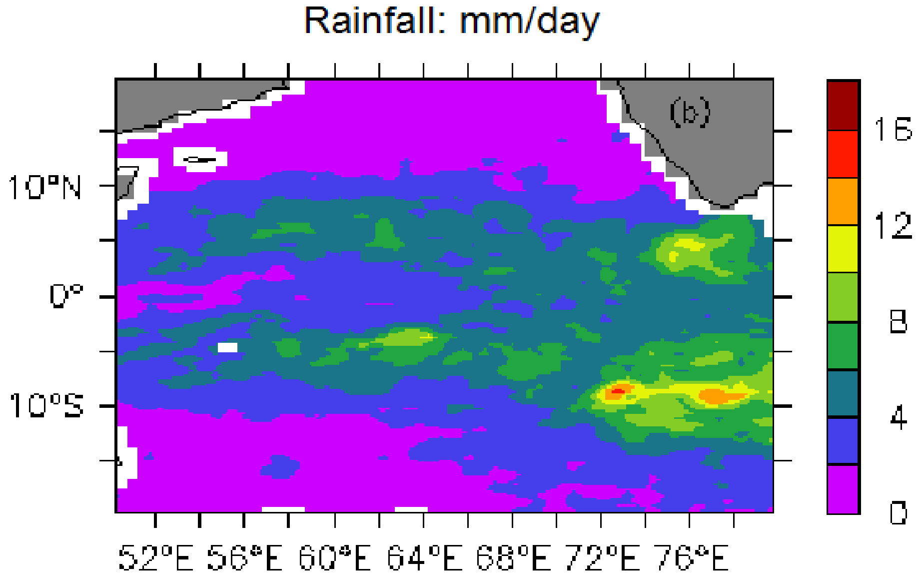
ATM- 0741832 of the US National Science Foundation

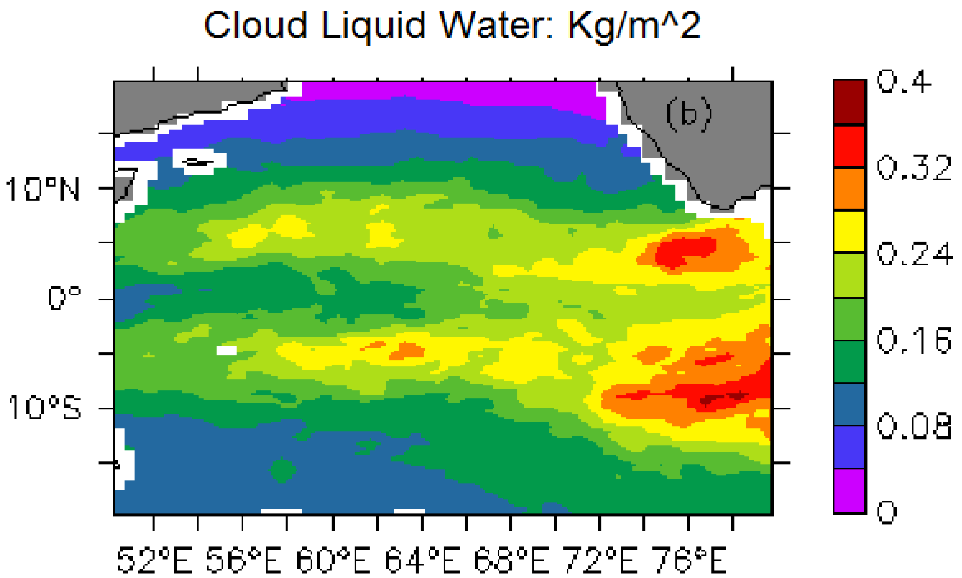
1: NorthWest Research Associates, 2: University of Washington

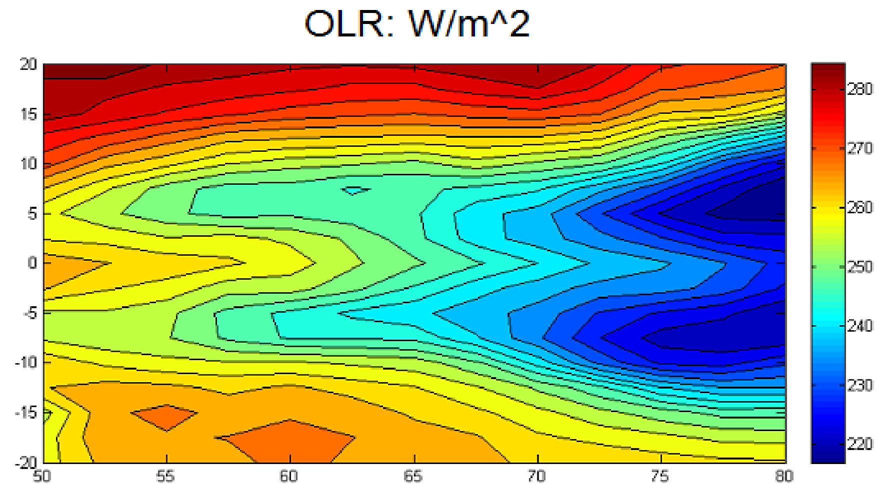
AMS 2011 – Data Mining Algorithm for Satellite Data: Double ITCZs



Climatological 18-year Mean: Nov 16-30 1988-2005

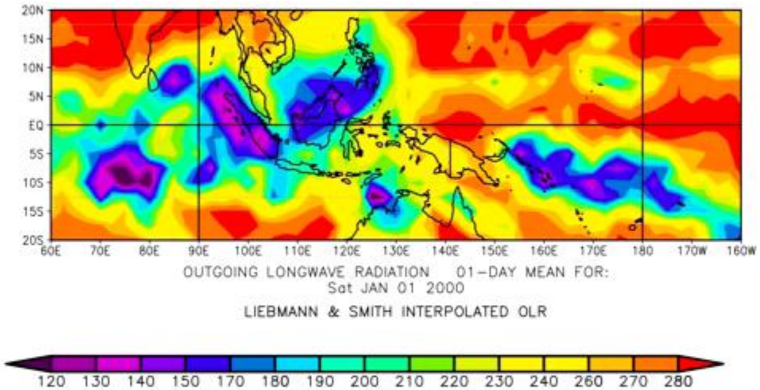




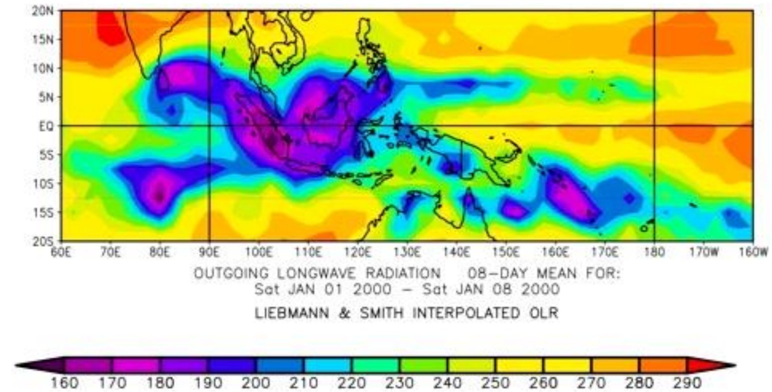


AMS 2011 – Data Mining Algorithm for Satellite Data: Double ITCZs

Temporal Mean:

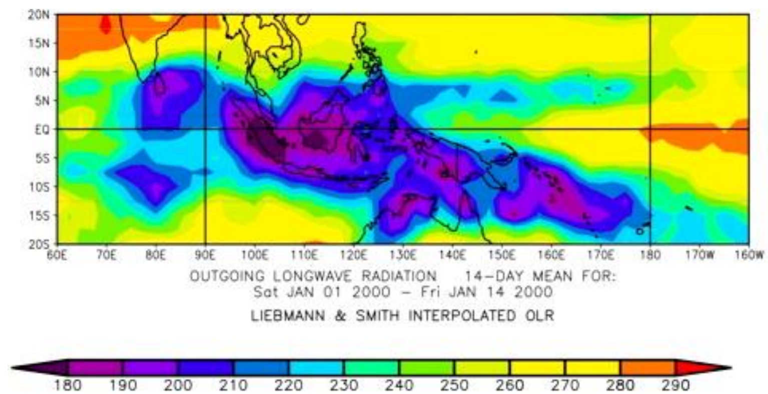


1-Day OLR



5-Day mean

14-Day mean



Goal:

-Develop an automated algorithm for detecting DITCZ manifestations capable of:

- -Searching at any location in an OLR image
- -Direct application to other data sets
- -Direct application to other regions
- -Being User friendly

Motivation:

-Relation of DITCZs with annual monsoon

-Tendency of GCM to over-predict DITCZs

-Potential relation of DITZCs to symmetric instability

-Application of algorithm to DITCZs in other ocean basins

-Potential application of feature detection to other fields?

Algorithm Process:

1) User generates training set by classifying a number of images

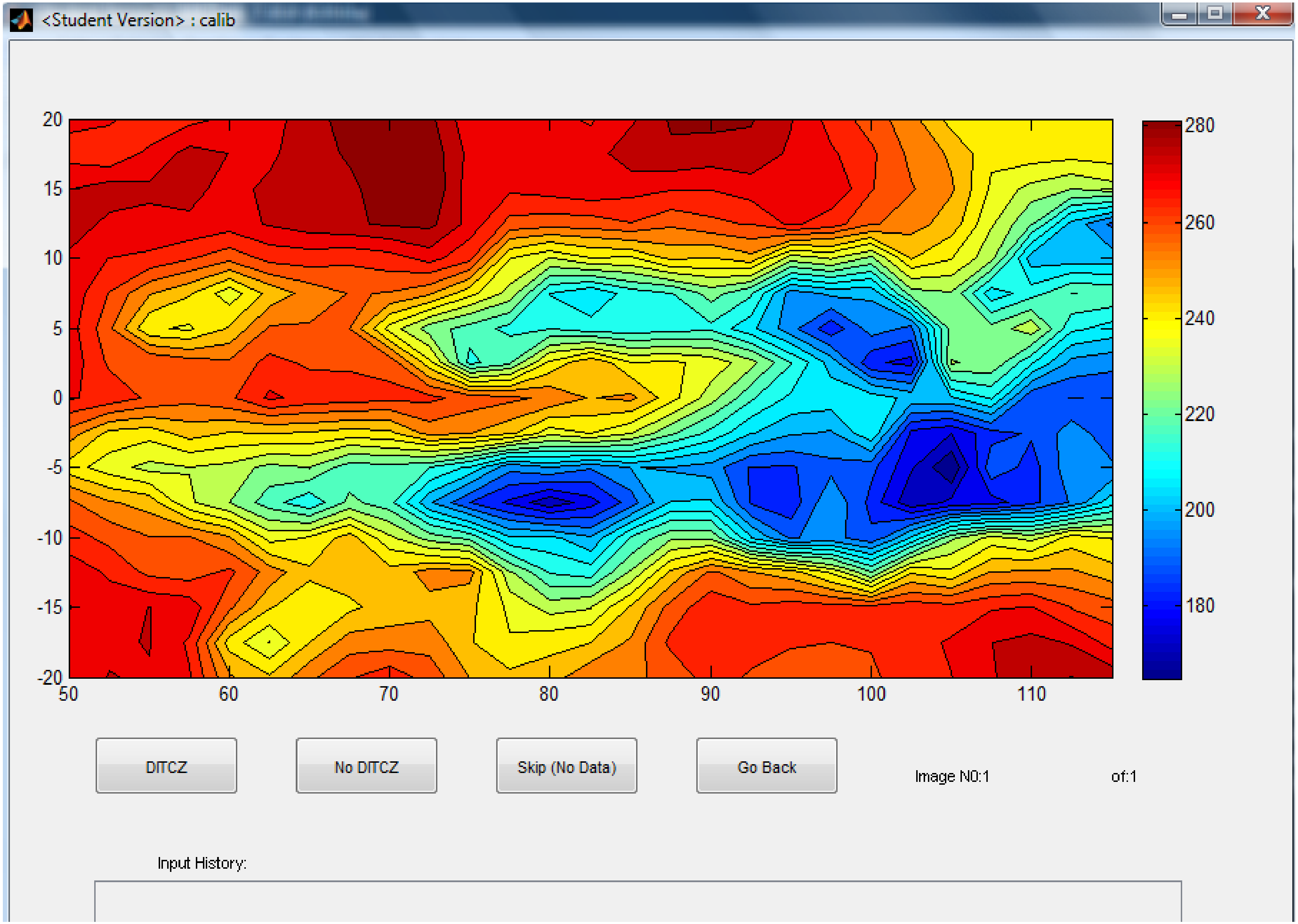
2) Images are decomposed via vertical wavelet transform and resulting coefficient energies are saved.

3) Singular value decomposition is applied to coefficient energy data

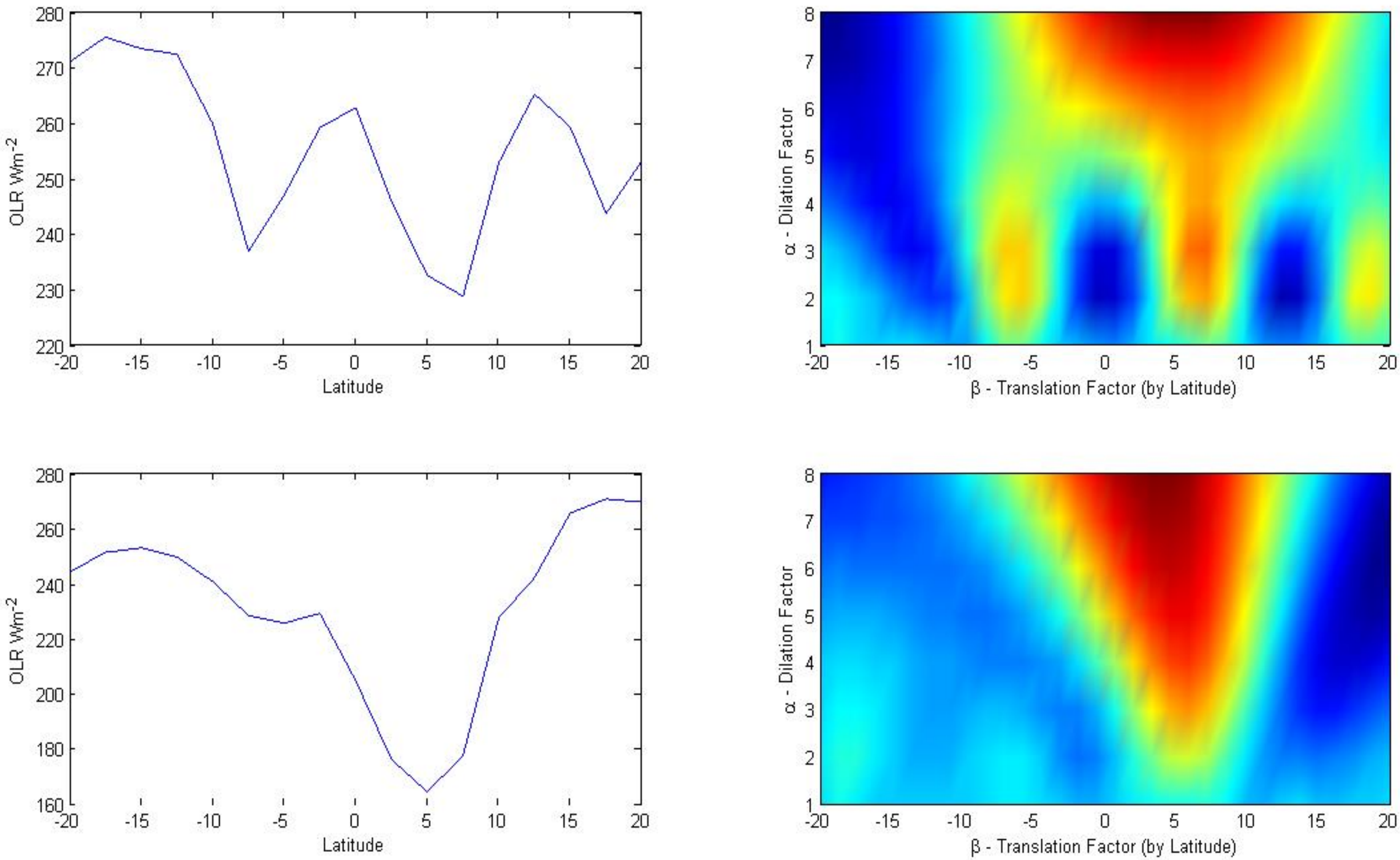
4) Linear discrimination analysis used to differentiate between images based on highest SVD modes

5) Post-processing applied to image to retrieve information about detected DITCZ

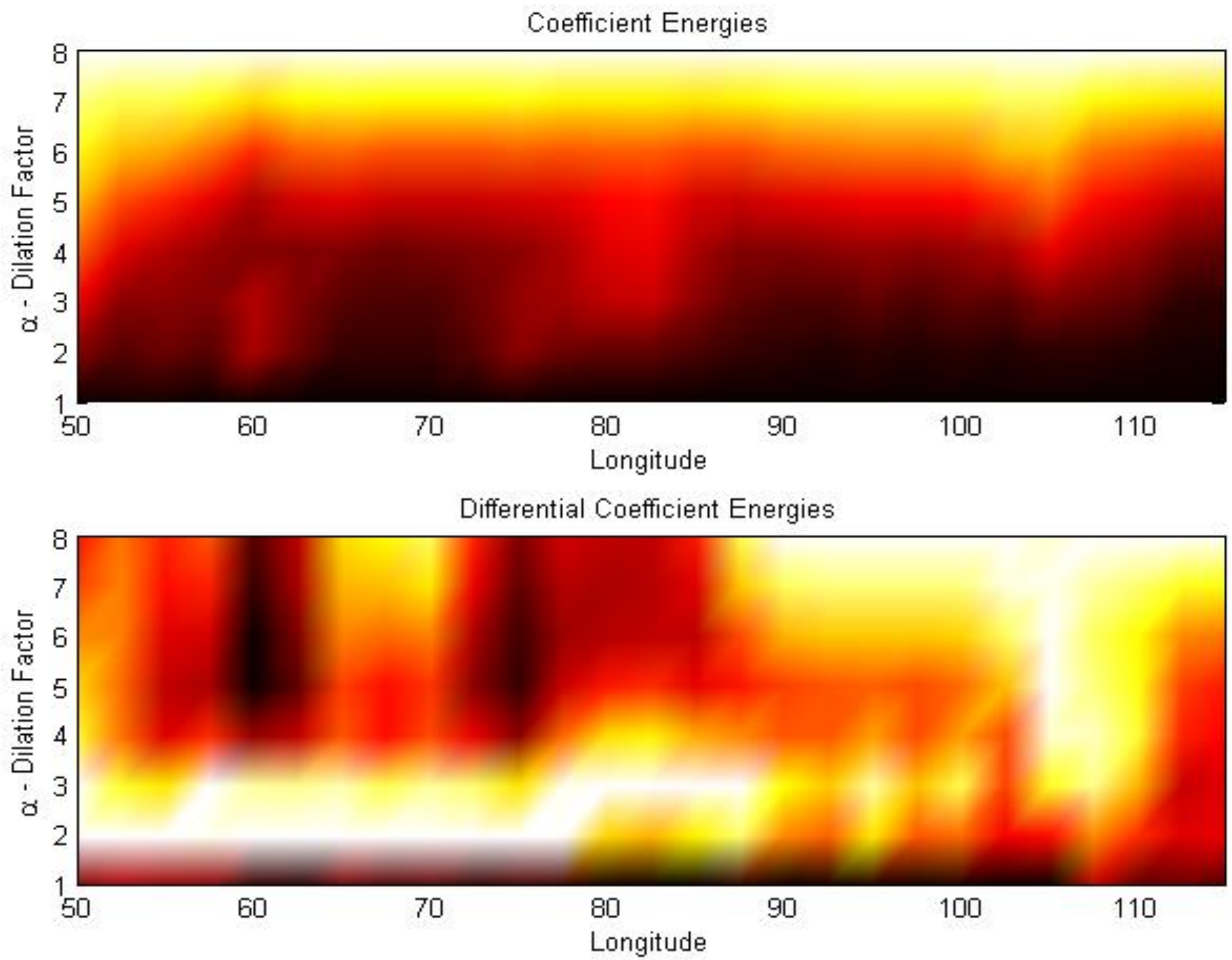
Algorithm Example (1/5): Training Set



Algorithm Example (2/5): Wavelet Decomposition

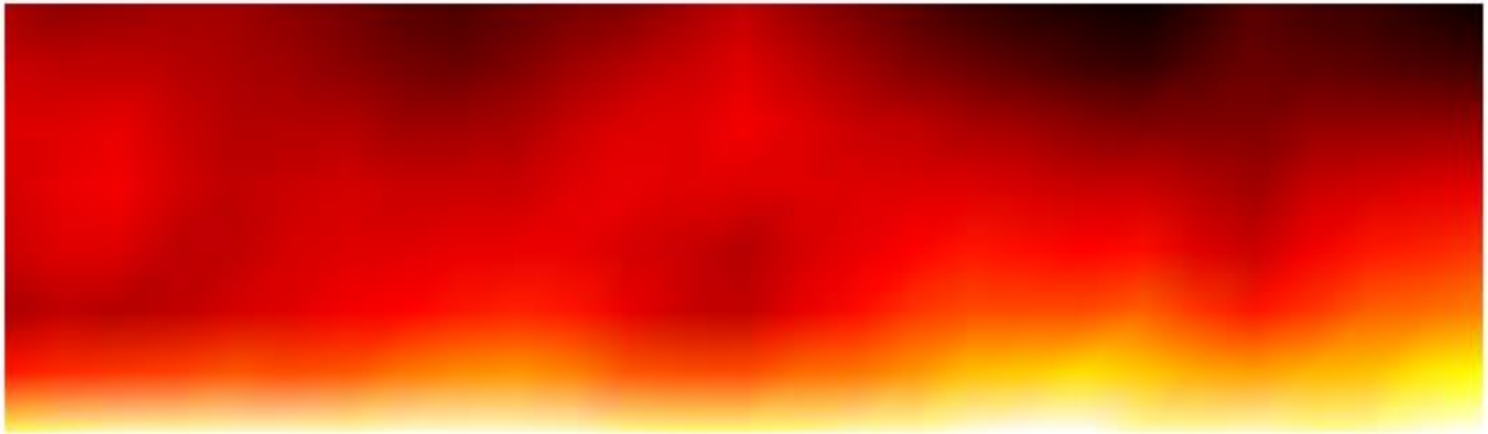


Algorithm Example (2/5): Coefficient Energies

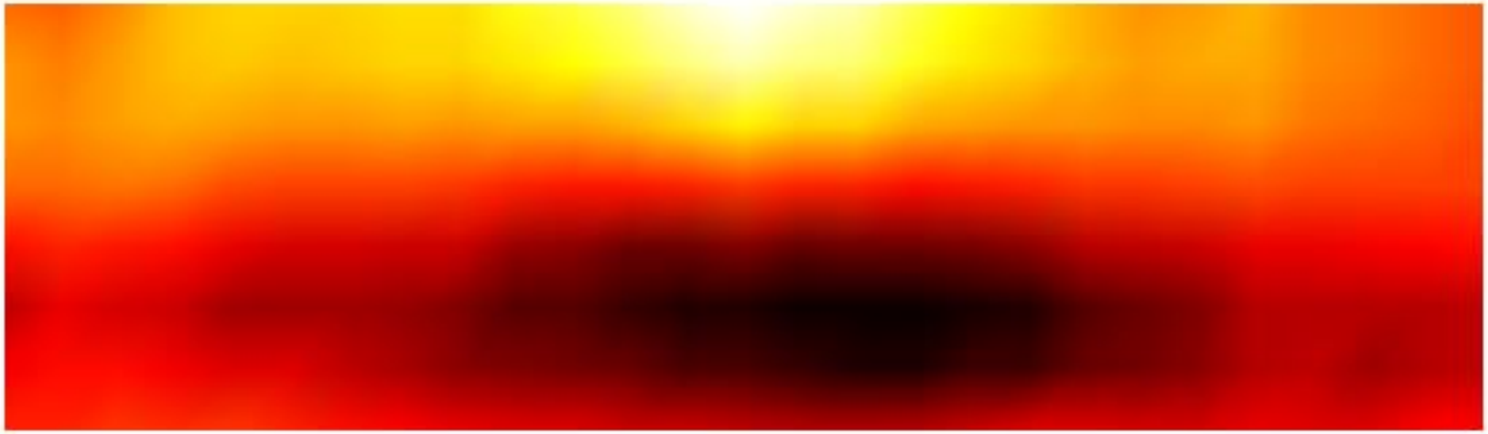


Algorithm Example (3/5): SVD, Primary Modes

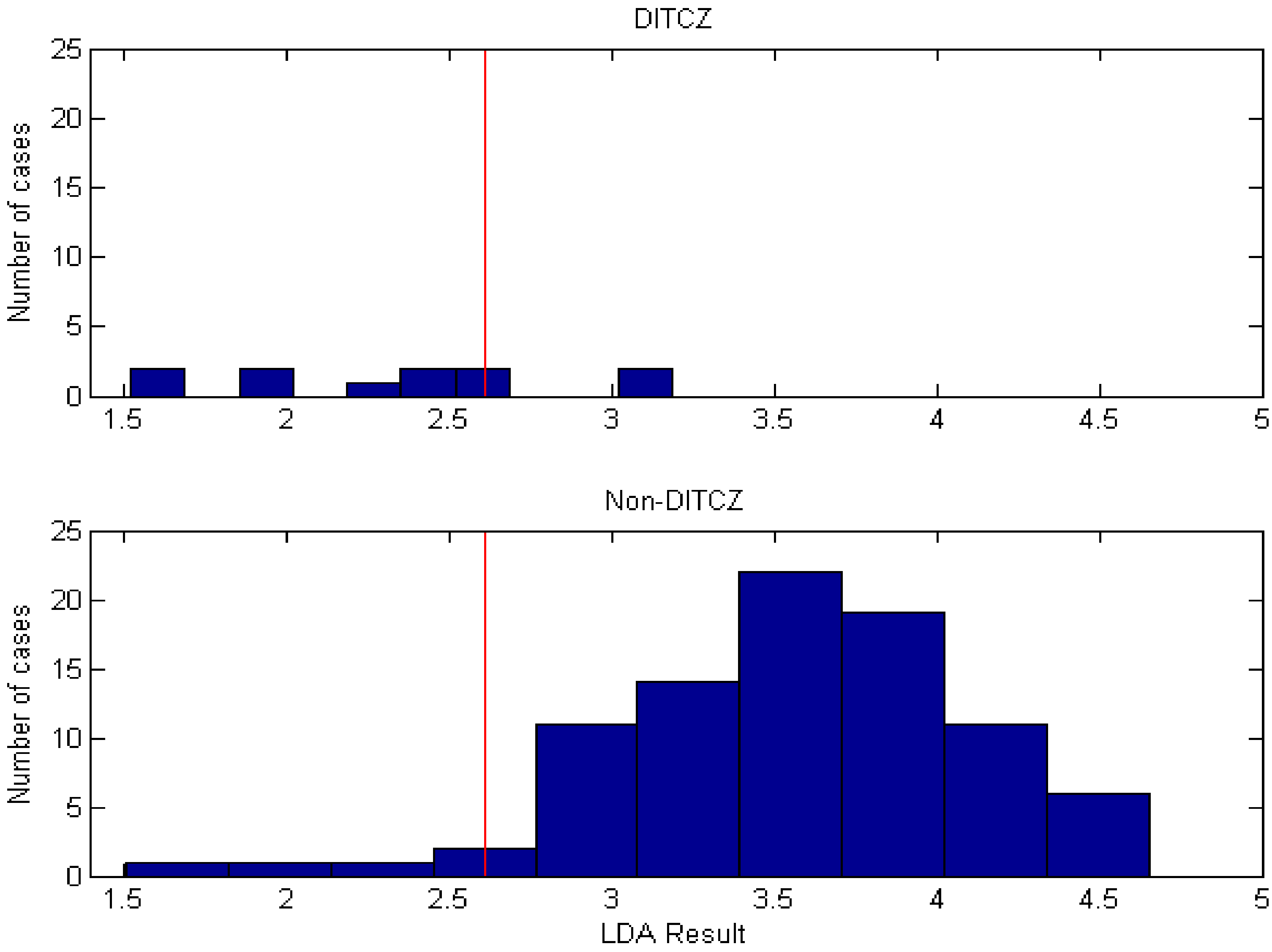
SVD Mode 1



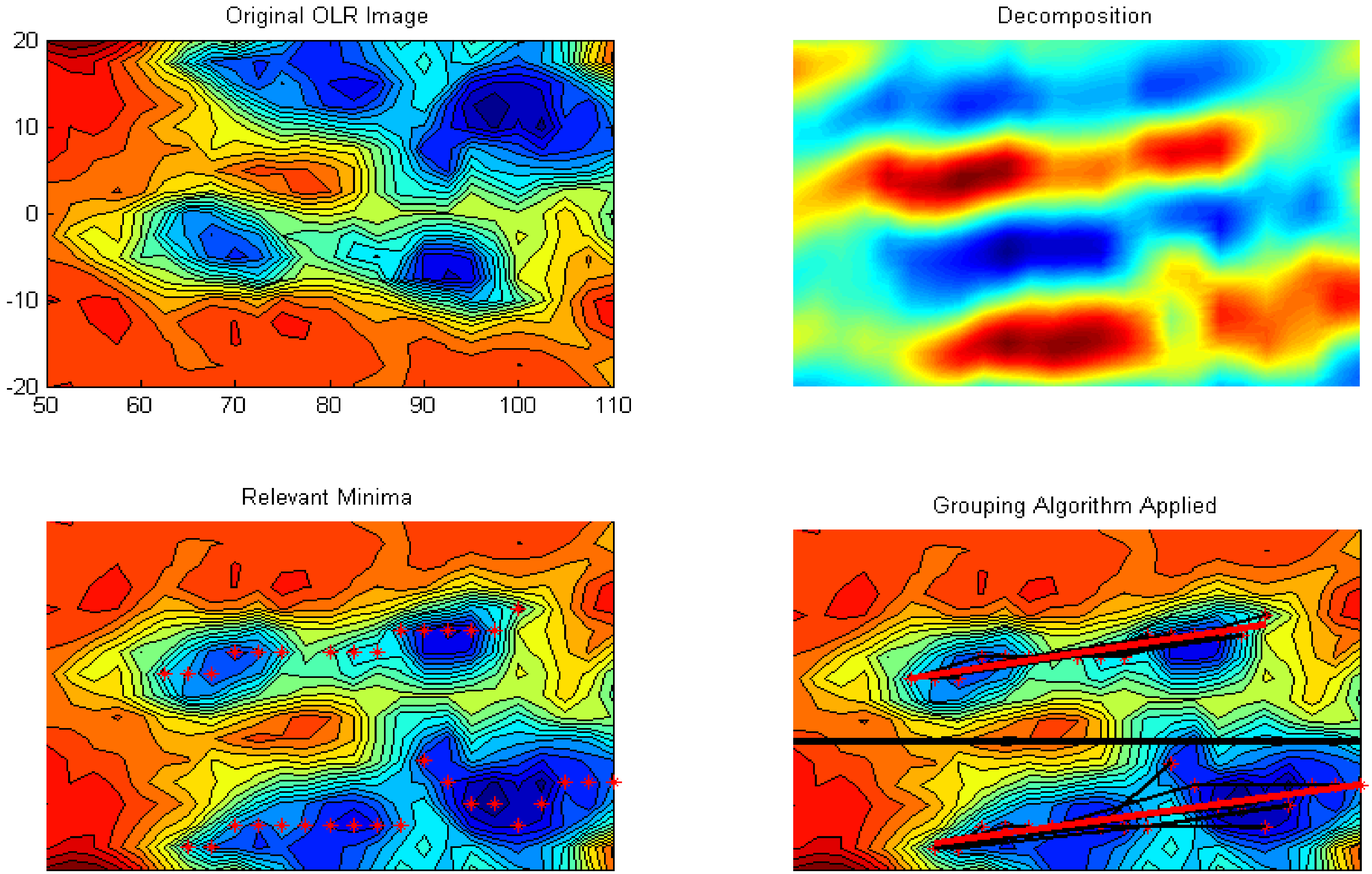
SVD Mode 2



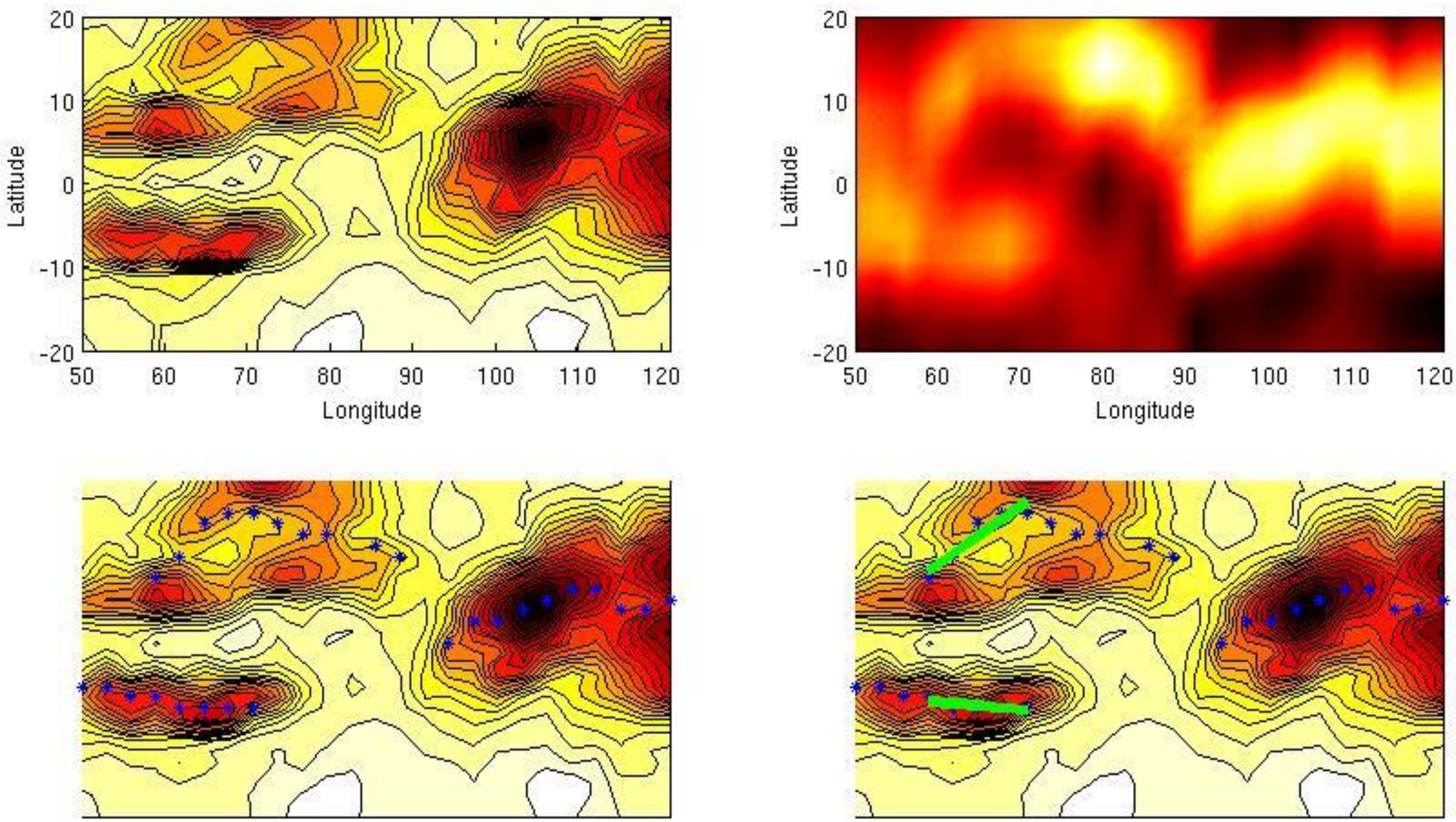
Algorithm Example (4/5): Linear Discrimination Analysis



Algorithm Example (5/5): Post-Processing



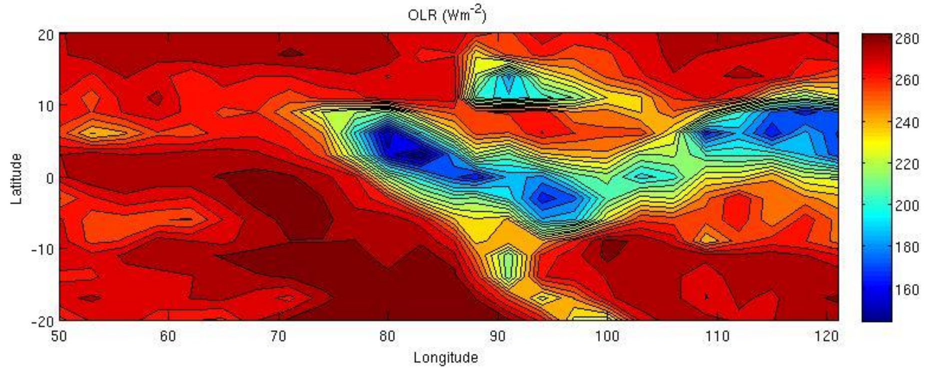
Algorithm Example (5/5): Post-Processing



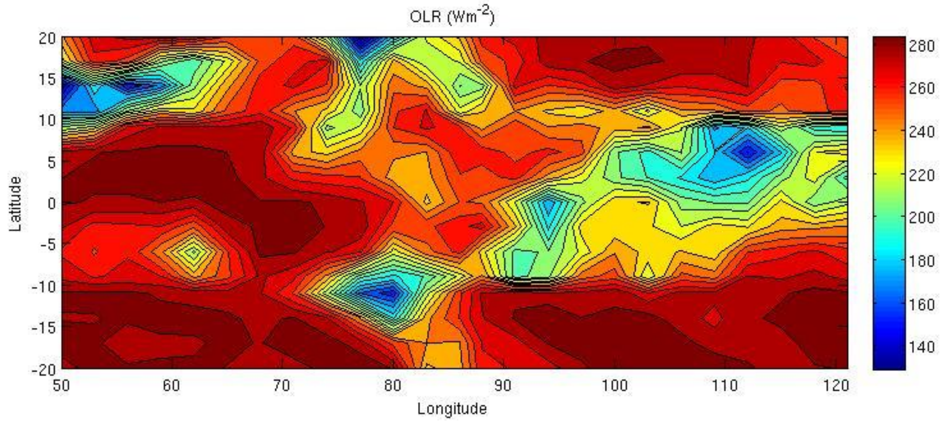
Automated DITCZ detection: Accuracy

(Out of 400 test cases:)

Overall: ~89.0%

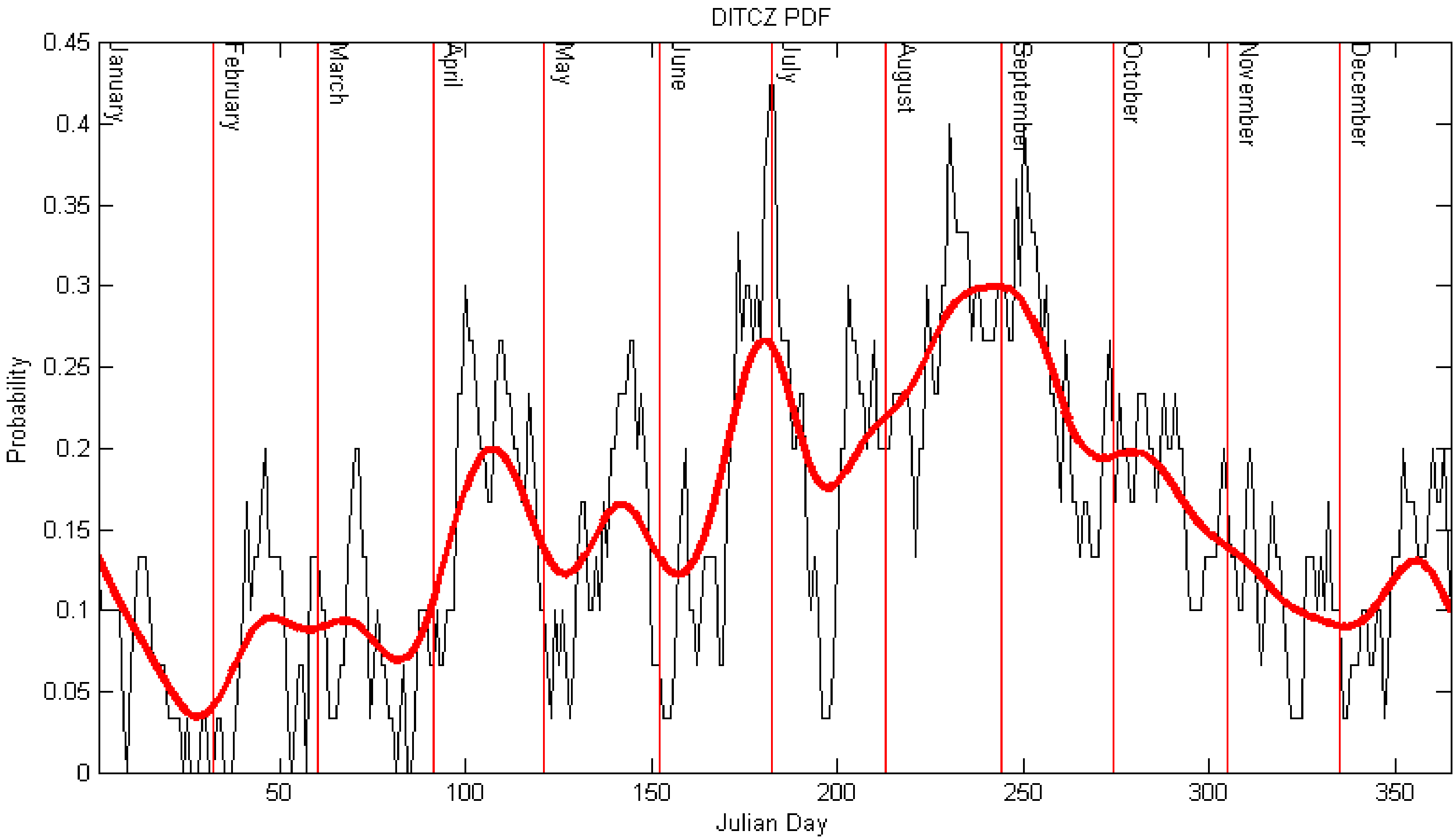


False Negatives: 7.0%



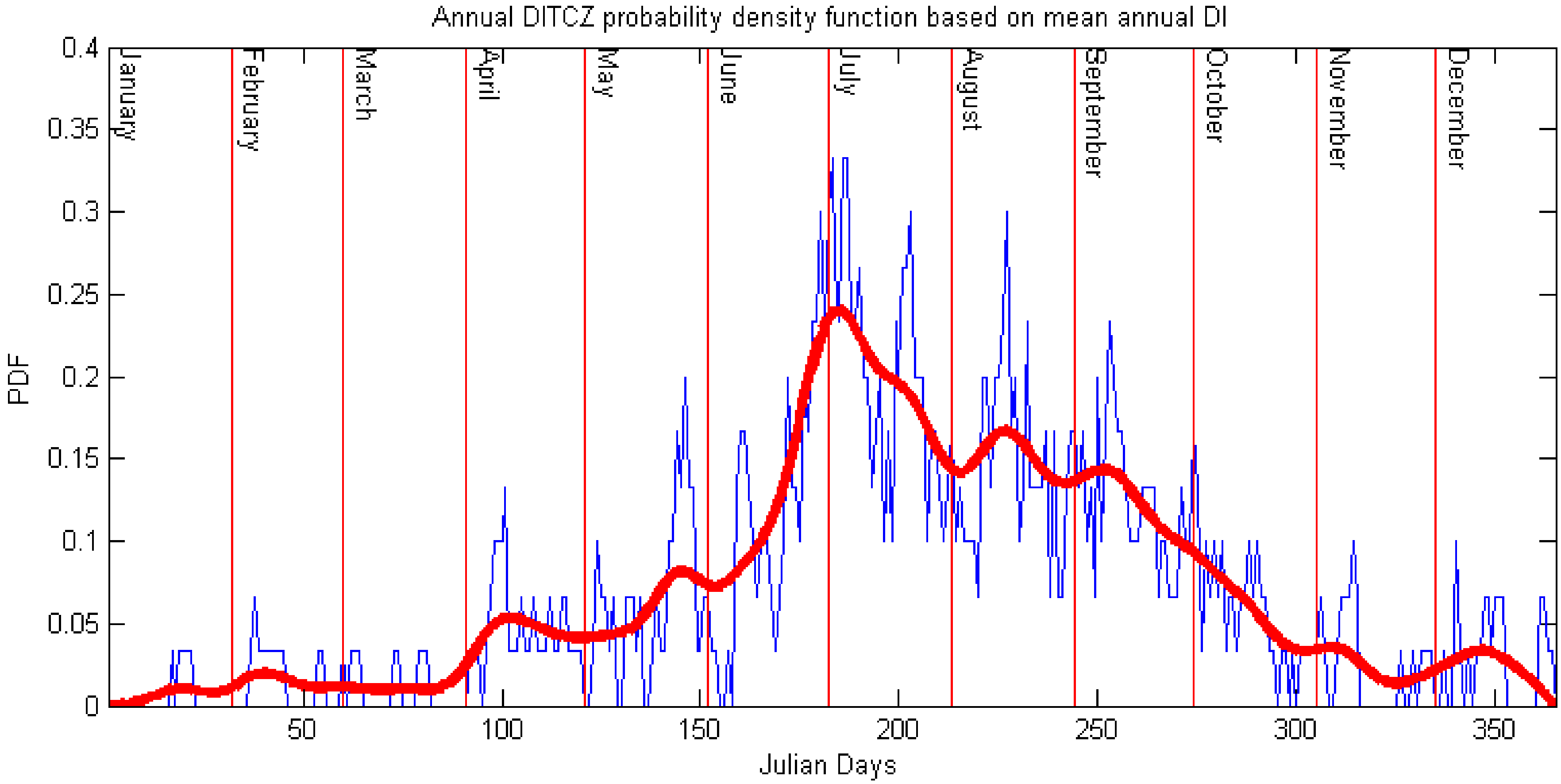
False Positives: 4.0%

Annual DITCZ Probability:



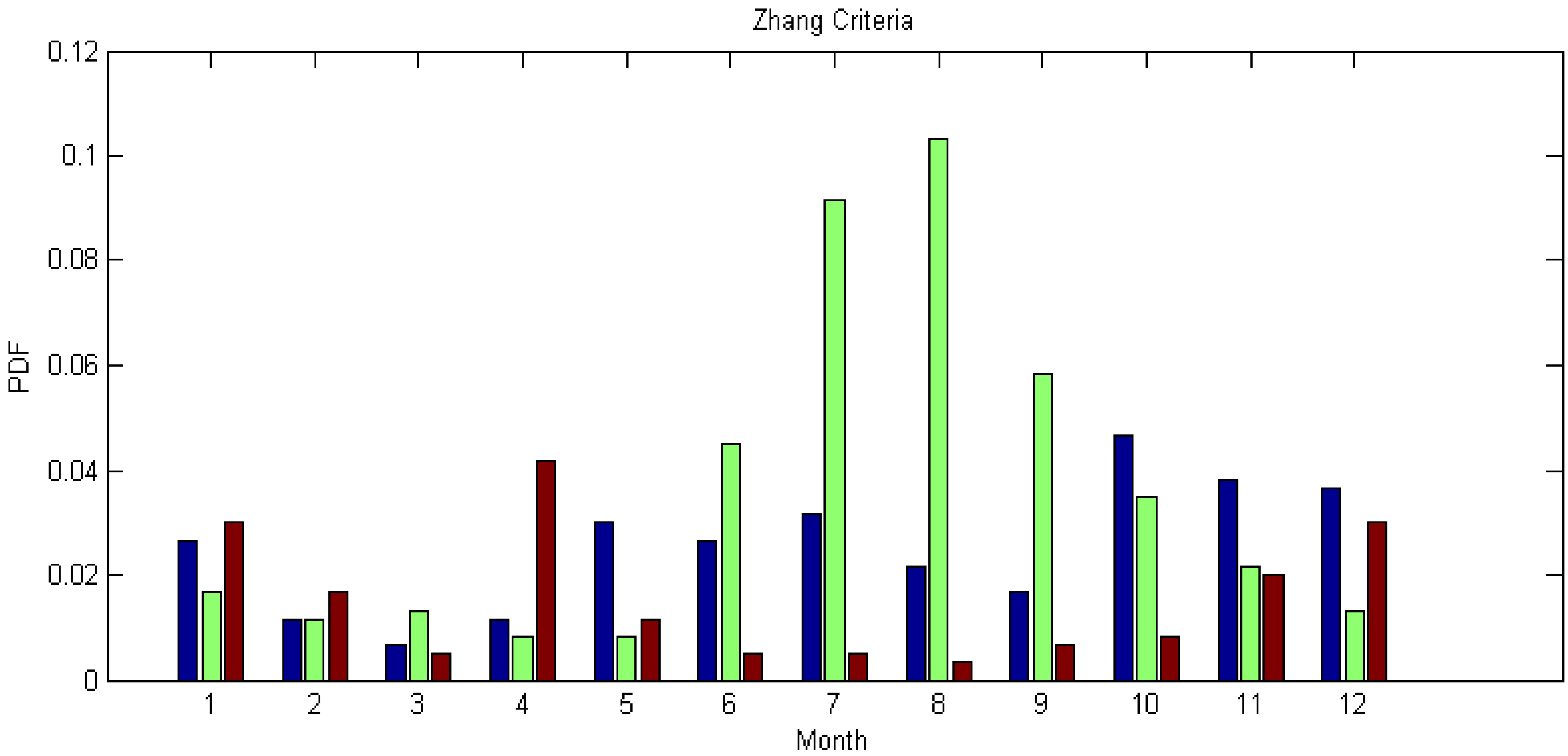
Annual signal produced by detection algorithm. OLR 10-day running mean from Jan 1st 1979 to Jan 1st 2009. Daily probability of DITCZ occurrence.

Alternative Detection Methods:



Annual signal produced by detection algorithm. OLR 10-day running mean from Jan 1st 1979 to Jan 1st 2009. Daily probability of DITCZ occurrence.

Alternative Detection Methods:



Previous algorithm output. Applied to 10-day running mean at 7° S (red), 0° (blue), and 7° N (green). Daily probability of DITCZ presence for each month.

Further Research:

-Verification by comparison to other data sets

-Comparison to surface winds and vorticity data

-Application to other ocean basins (western Pacific?)