

Automatic Classification of Sunspot Groups Using SOHO/MDI Magnetogram and White Light Images

David Stenning

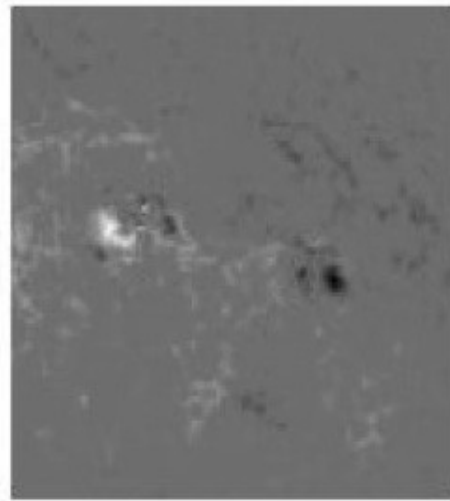
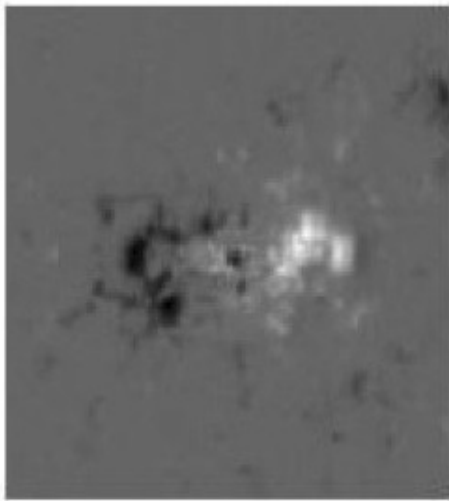
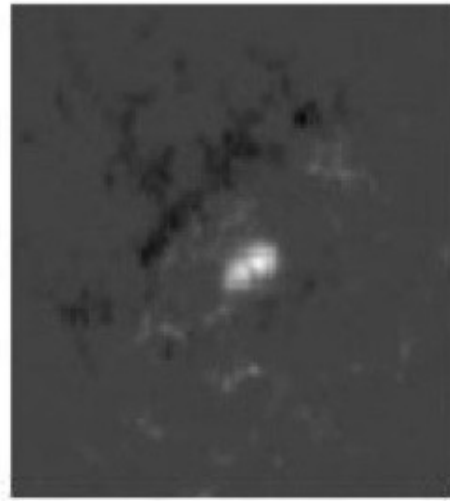
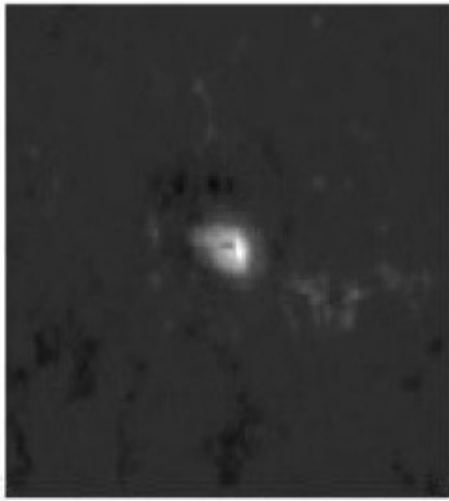
March 29, 2011

Introduction and Motivation

- Although solar data is being generated at an unprecedented rate, the majority of sunspot classification is done manually by experts.
- In addition to being subject to observer bias, manual classification is labor-intensive and time-consuming.
- Therefore, a fast and automatic classification procedure is highly desirable.

Mount Wilson Classification

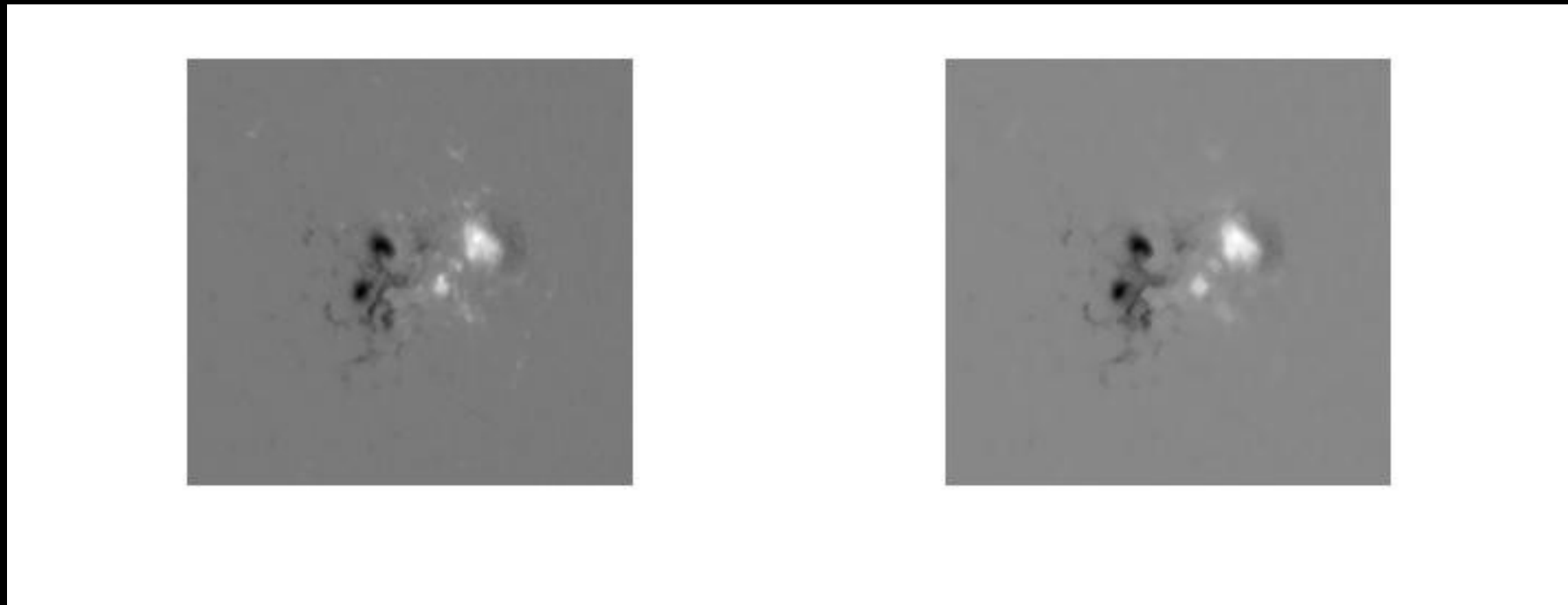
- Using the Mount Wilson scheme, sunspot groups can be classified into the following four broad categories (from Ireland et al. 2008):
 - *Alpha*: a single dominant spot often linked with a plage of opposite magnetic polarity
 - *Beta*: a pair of dominant spots of opposite polarity
 - *Beta-Gamma*: bipolar groups with more than one clear North-South polarity inversion line
 - *Beta-Gamma-Delta*: bipolar groups with umbrae of opposite polarity together in a single penumbra



Top : alpha sunspot (left); beta sunspot (right) *Bottom*: beta-gamma sunspot (left); beta-gamma-delta sunspot (right)

Extracting Features From Magnetograms

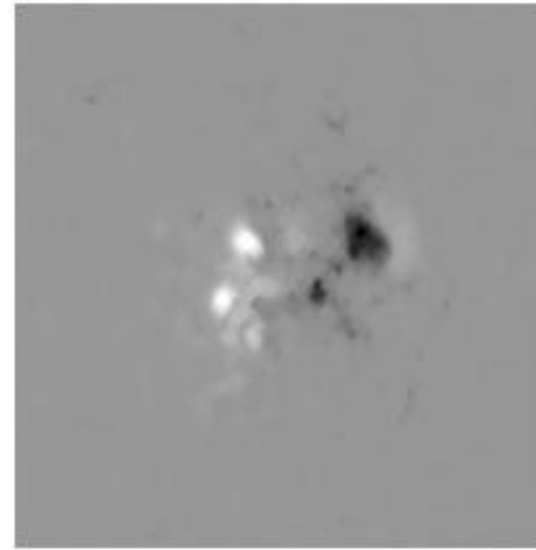
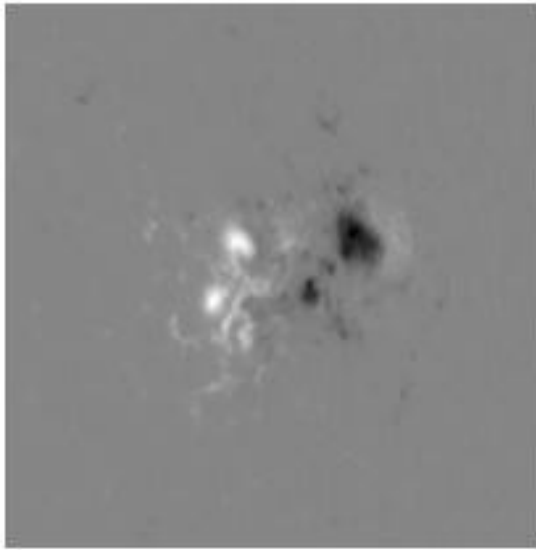
- As a first step, magnetograms are “cleaned” by a morphological opening operation (see Soille, 2003) with a spherical structuring element of radius 2. This is to smooth the “white” areas in the magnetograms.



Left: Original Magnetogram

Right: “Opened” magnetogram

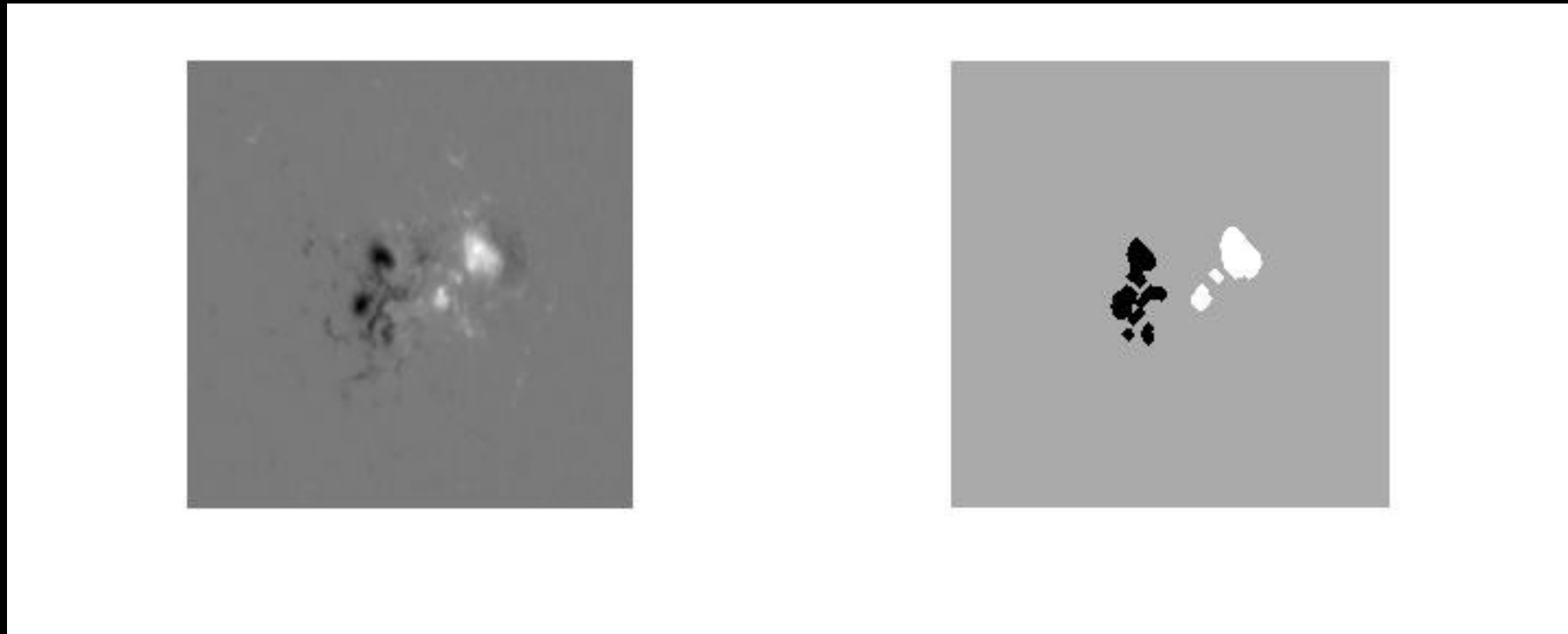
This is repeated on the inverse magnetogram image to smooth the black regions.



Left: Inverse Magnetogram

Right: “Opened” Inverse Magnetogram

Using the cleaned images, sunspot active-region “seeds” are obtained by thresholding. In this operation, the threshold value is set to be 2.5 sample standard deviations above the sample mean (calculated from all pixel values in the image). All pixels above this value are marked as belonging to the active-region.



Left: Original Magnetogram

Right: Active-region seeds

- Using these seeds, we can extract features:

Ratio: The ratio of active-region seeds.

$A(w)$:

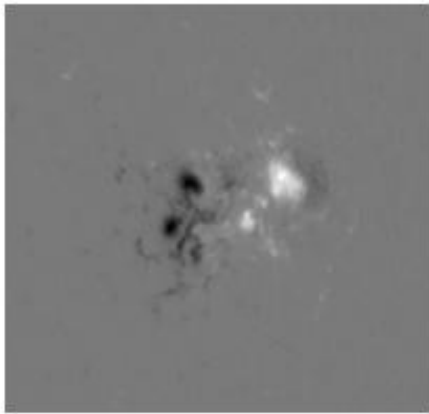
- Let W denote the set of all white active-region seeds.
- Compute the center of mass, c , of W .
- For each w in W , walk from w to c in a straight line counting (i) the total number of pixels traveled and (ii) how many pixels are in W (i.e. are part of the white active region). Denote (i) by $L(w)$ and (ii) by $l(w)$. Then $A(w)$ is given by

$$A(W) = \frac{1}{|W|} \sum_{w \in W} \frac{l(w)}{L(w)},$$

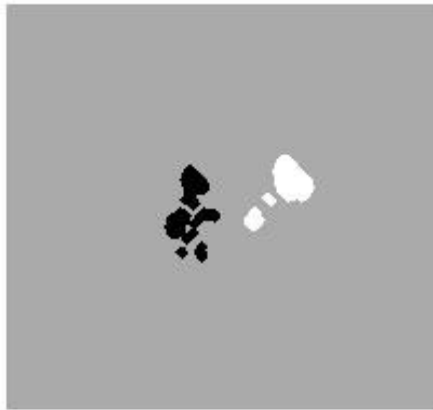
where $|W|$ is the number of pixels in W .

$B(w)$: Computed the same as $A(w)$ but with the black active-region seeds.

Since we are interested in the separation of the two polarities, we use seeded region growing (see Adams and Bischof, 1994) to obtain the separating line.



Magnetogram



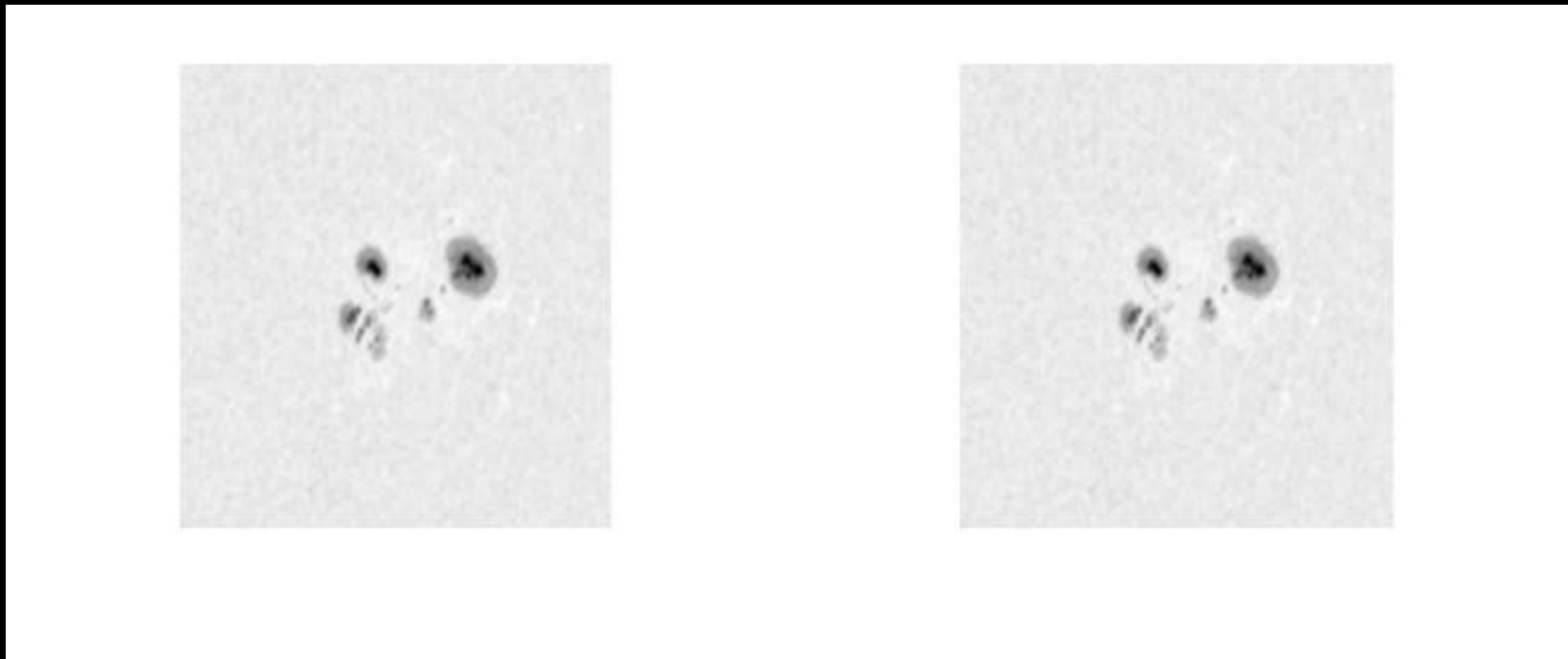
Active-Region Seeds



Binary Regions

The curvature of the separating line is computed by averaging second derivatives on the line designated by where the white and black regions meet. This is the the third feature, denoted by **curve**.

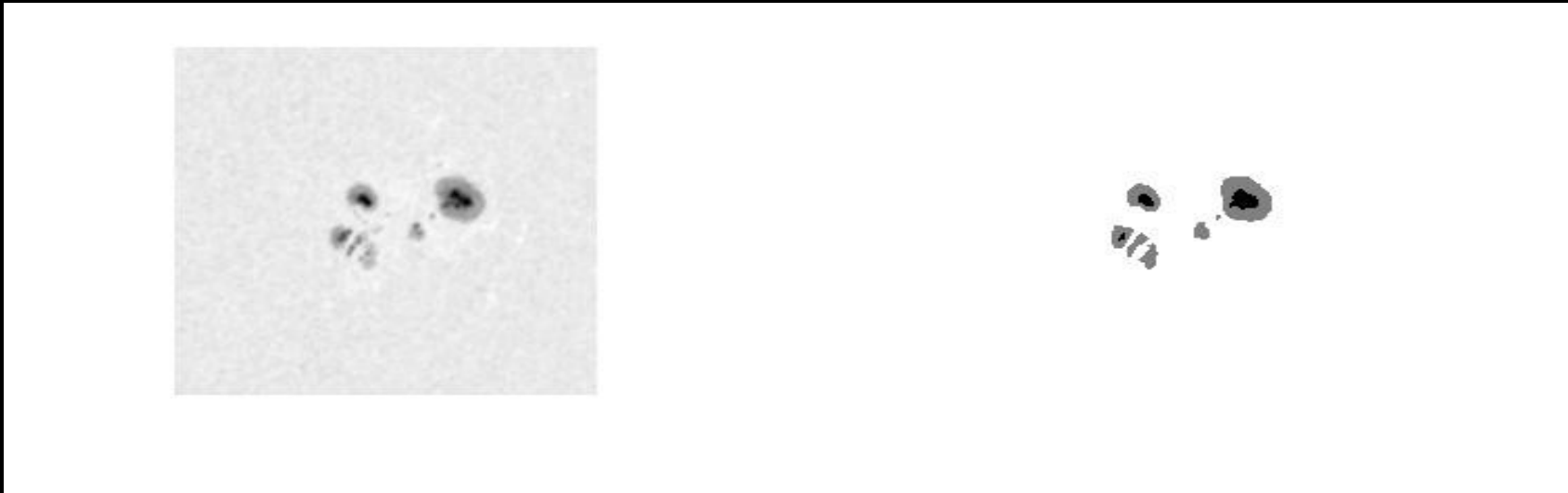
- To detect delta spots, we need to be able to compare polarities of umbrae within a single penumbra in sunspots.
- To do this, we need to utilize the white-light (intensity-gram) images.
- Our first step, as with the magnetograms, is to clean the image with a morphological opening:



Left: White-Light Image

Right: Cleaned White-Light Image

- Using the cleaned image, a two-stage thresholding operation is applied. First, the entire image is thresholded to obtain pixels that are part of the sunspot. Then, thresholding is applied on only the sunspot pixels to identify pixels that are part of an umbra.



Left: Original WL Image

Right: Penumbra and Umbra Regions

- We can then examine each closed penumbra region to determine if there are umbrae of different polarities within a single penumbra. The outcome is denoted by the categorical variable **delta** (1 if true, 0 otherwise).

Features

A(W): “Scatter” of white active-region pixels

B(W): “Scatter” of black active-region pixels

Ratio: The ratio of white-to-black or black-to-white active-region pixels (depending on which active-region has more pixels associated with it).

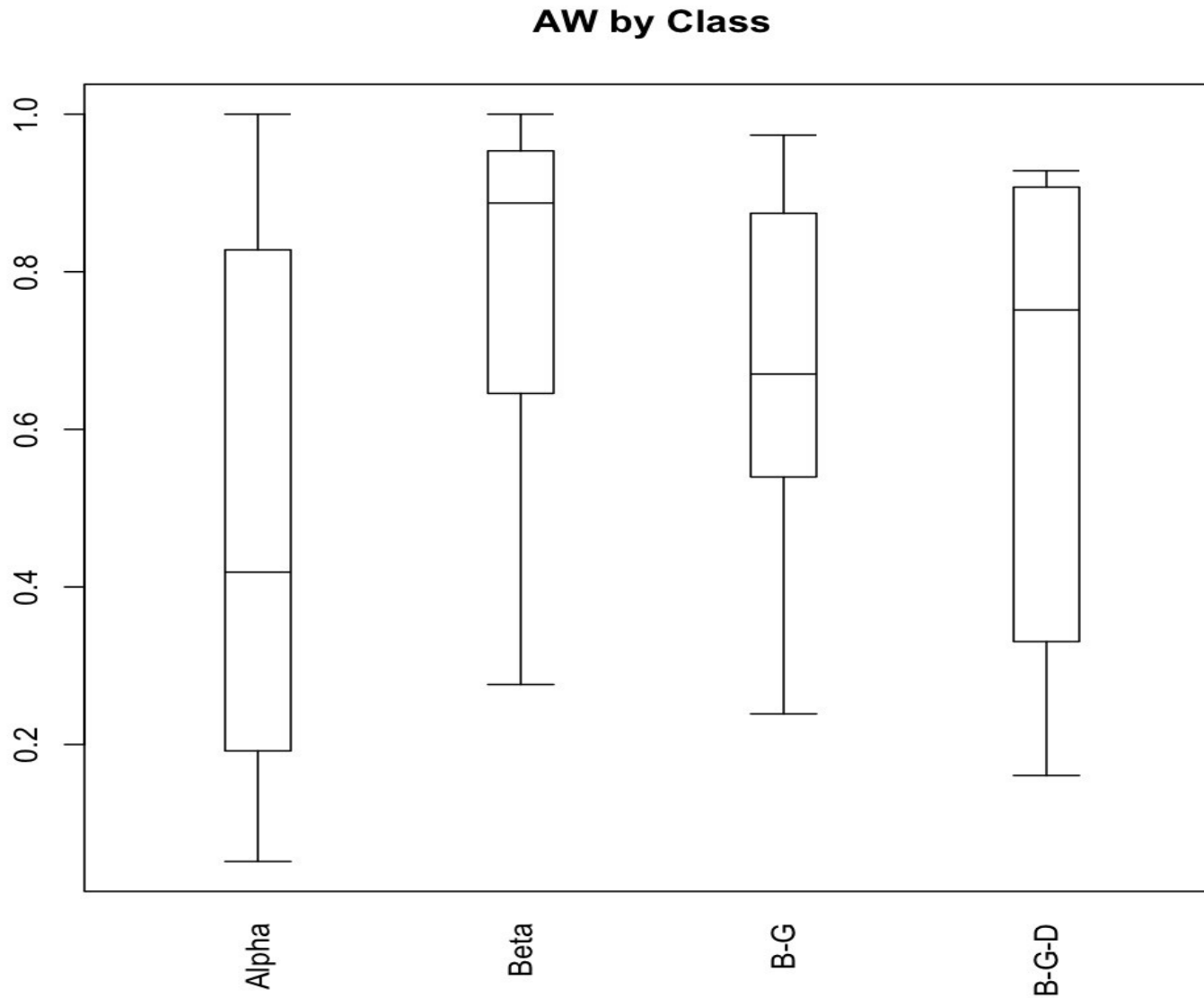
Curve: The curvature of the separating line

Delta: Indicator as to whether there are umbrae with different polarities within one penumbra.

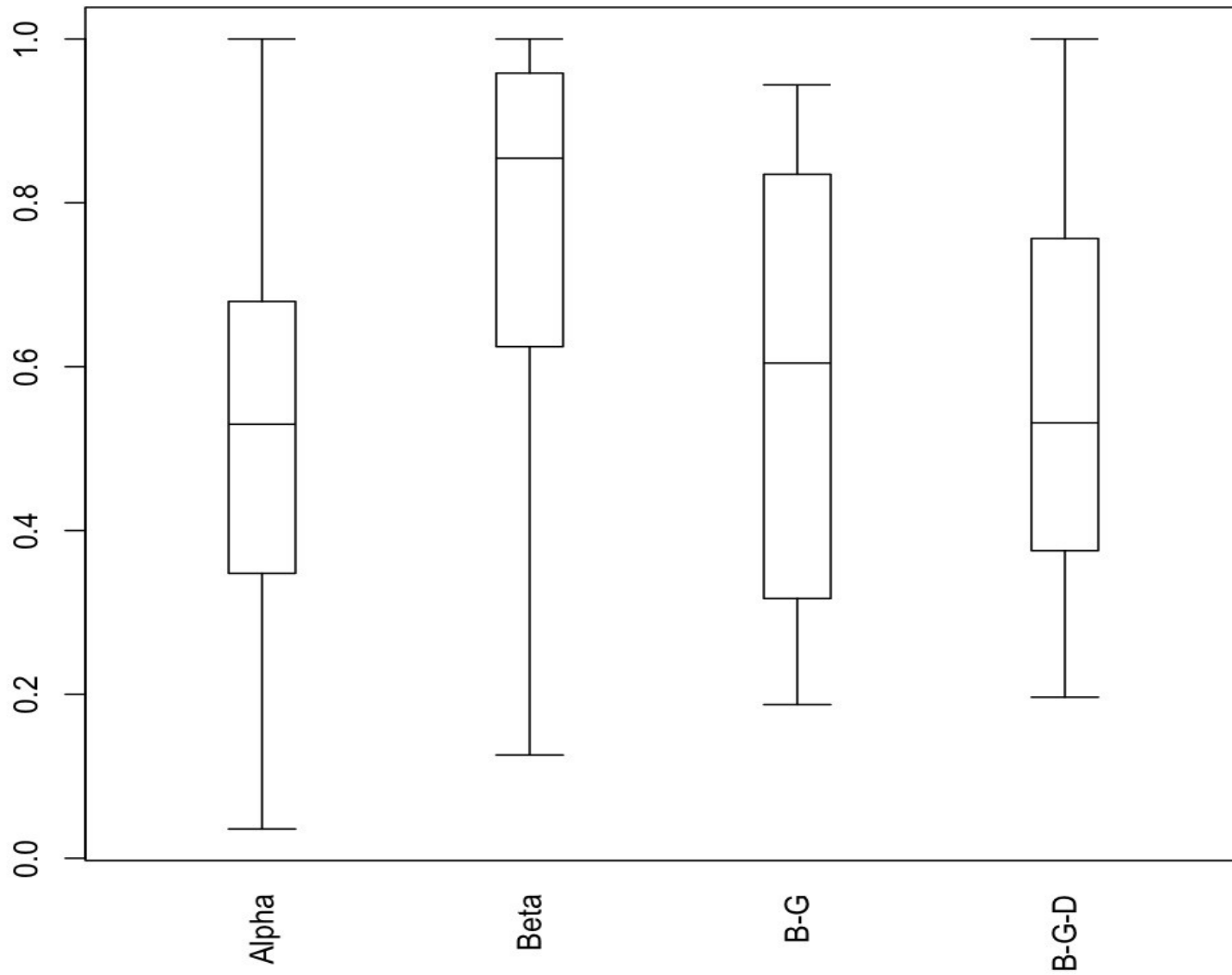
Using “Good” Examples

- To test the features we have created, 128 sunspot groups were selected to be “good” examples of the 4 classes. They are broken down as follows:
 - 28 Alpha
 - 66 Beta
 - 22 Beta-Gamma
 - 12 Beta-Gamma-Delta

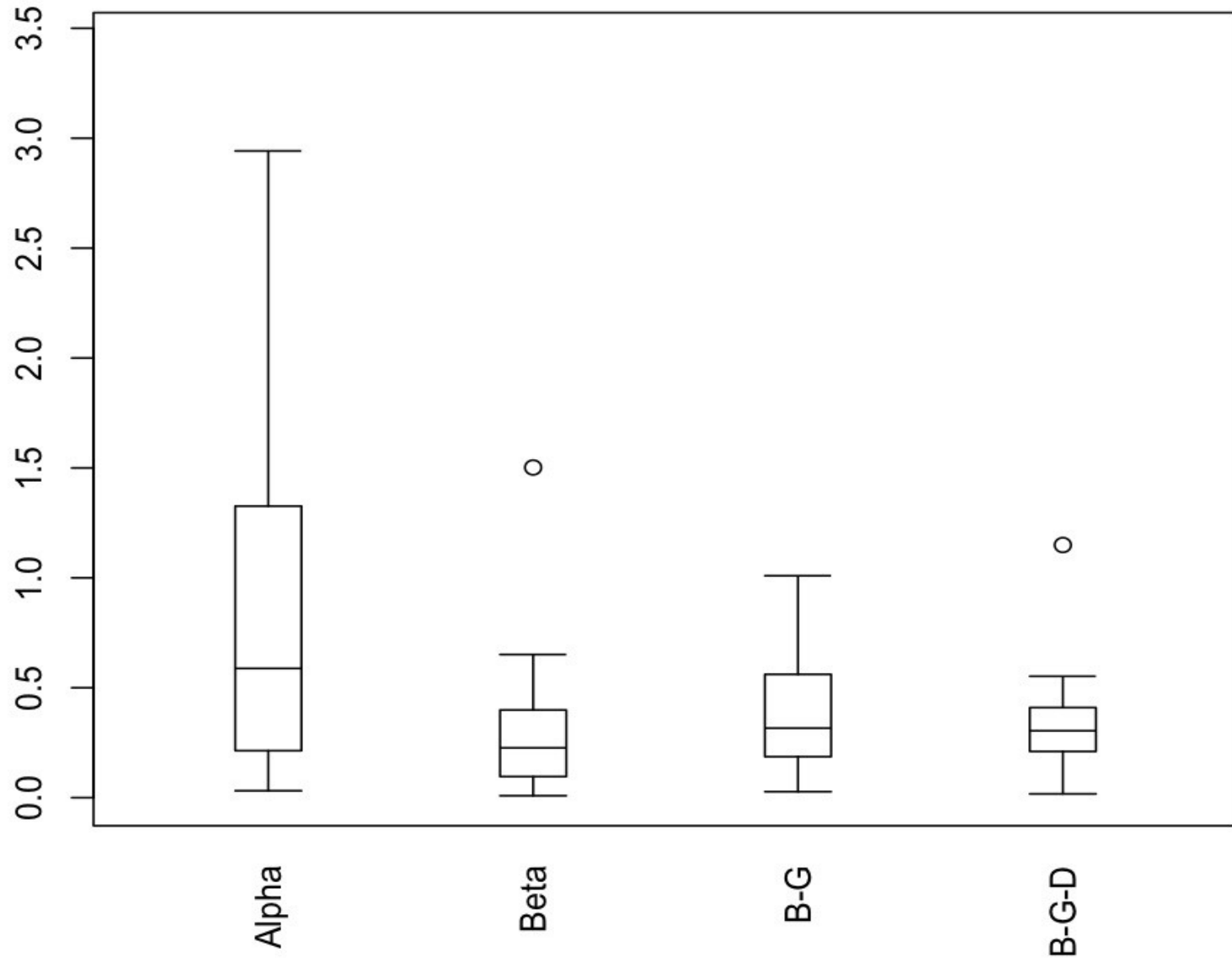
Exploratory Data Analysis



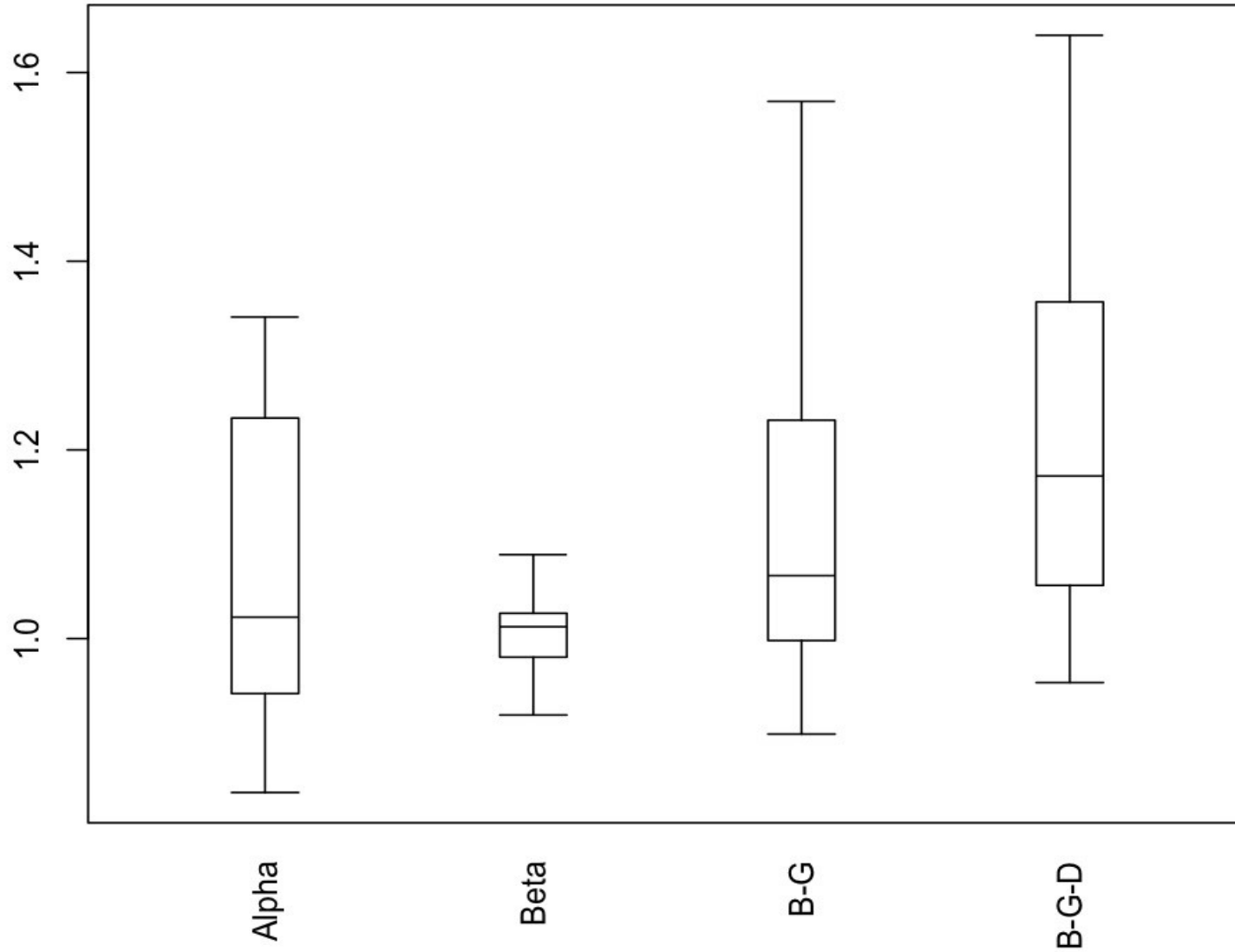
BW by Class



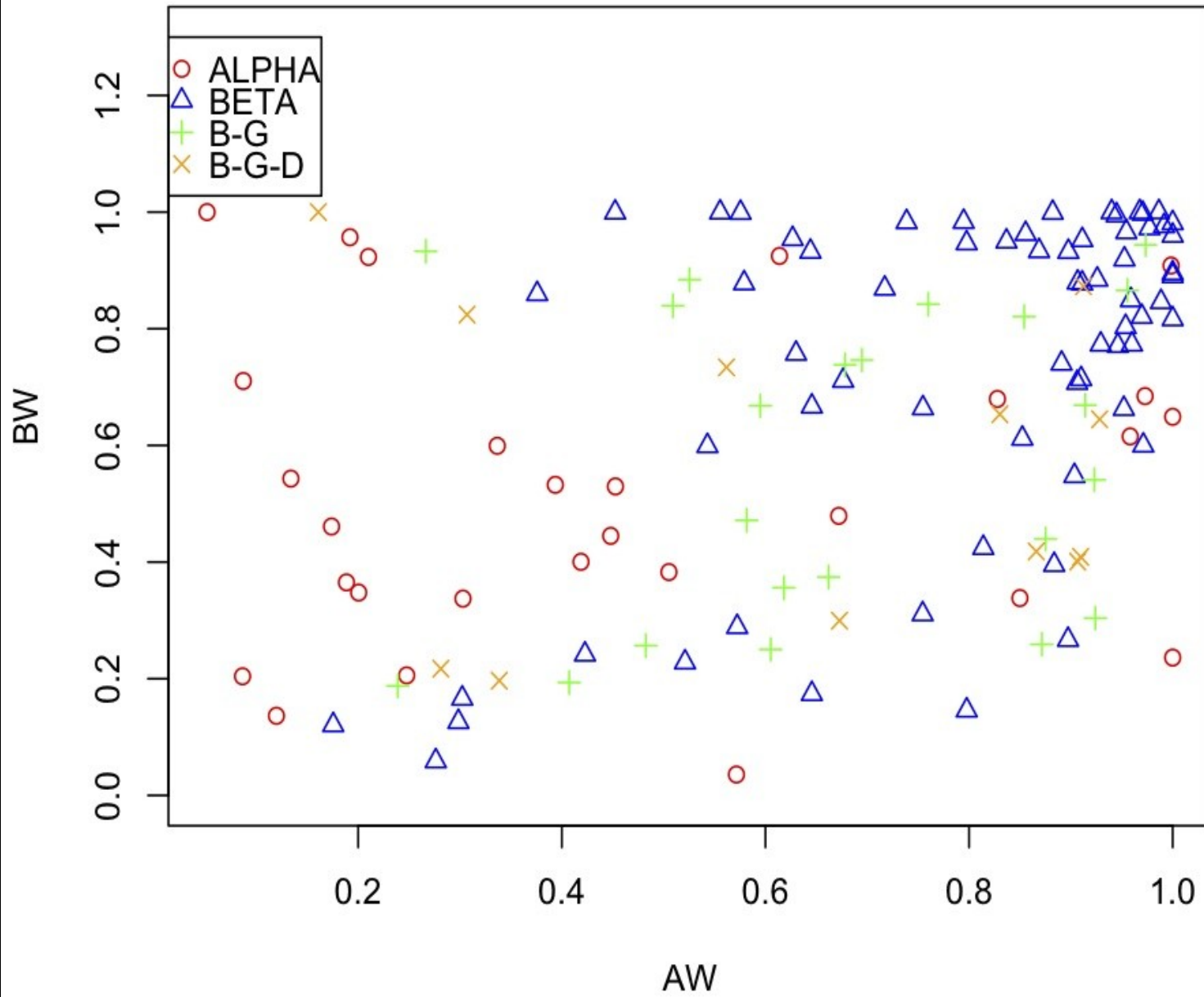
log(Ratio) by Class



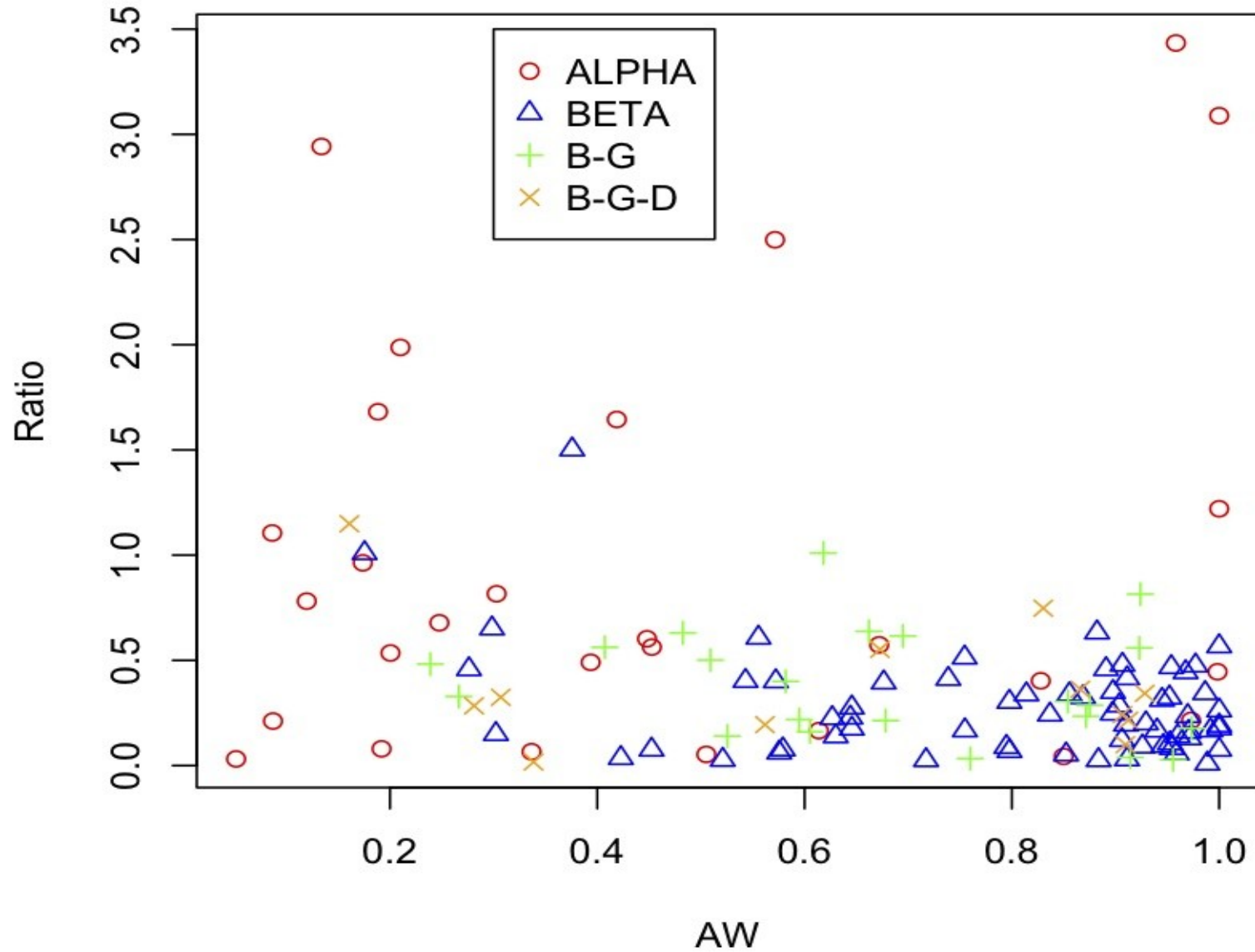
Curvature by Class



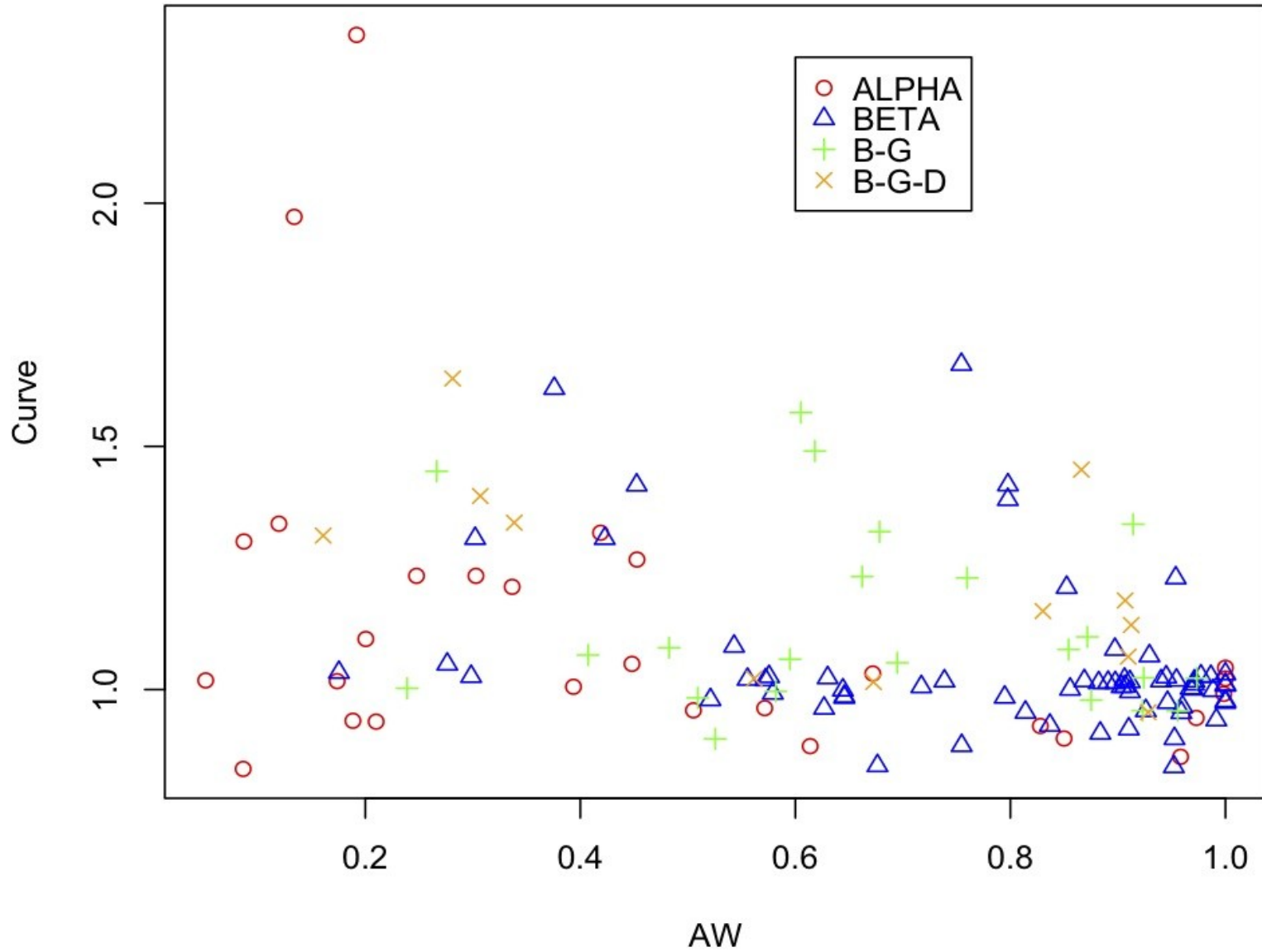
AW vs. BW



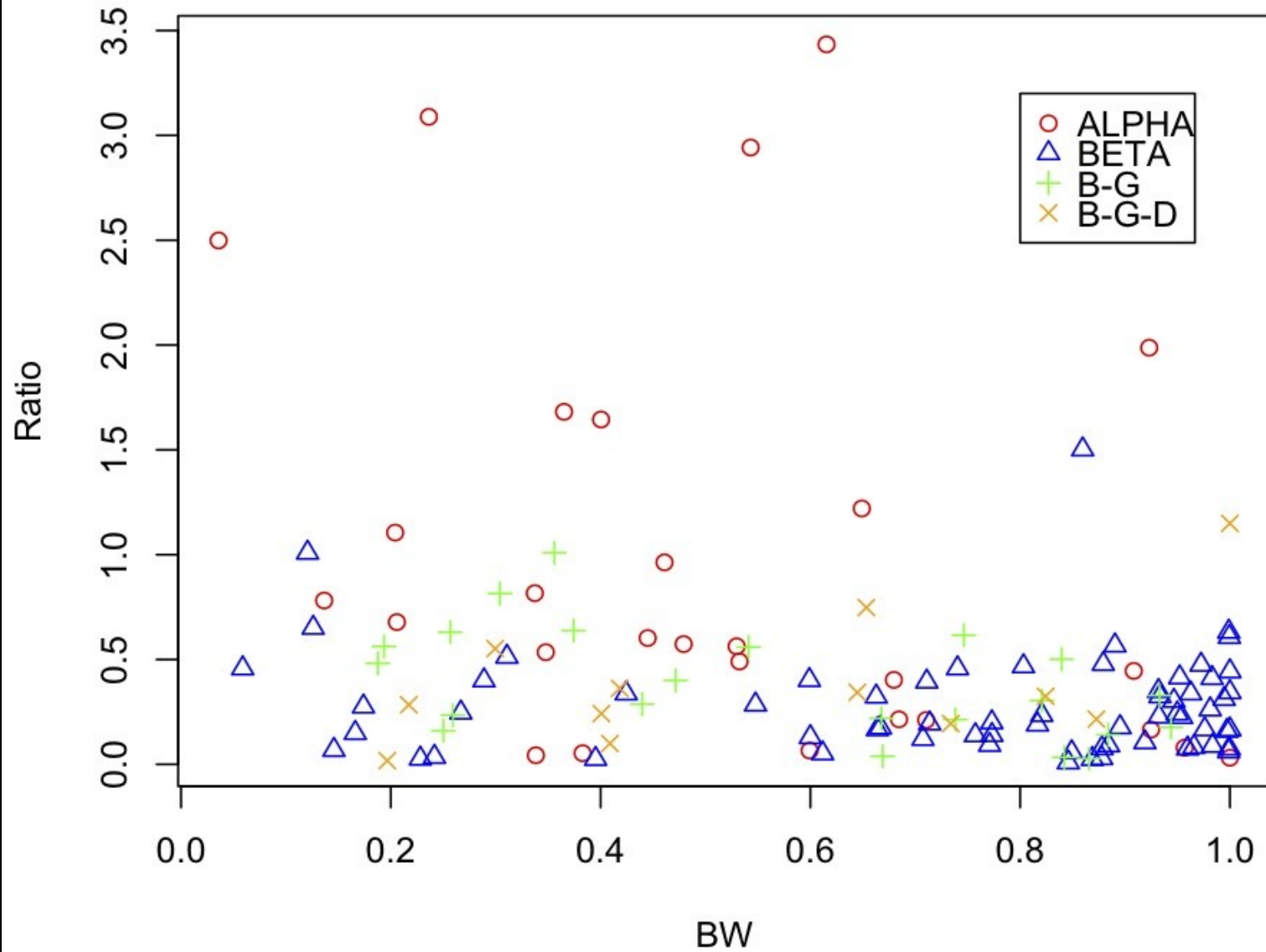
AW vs. log(Ratio (2))



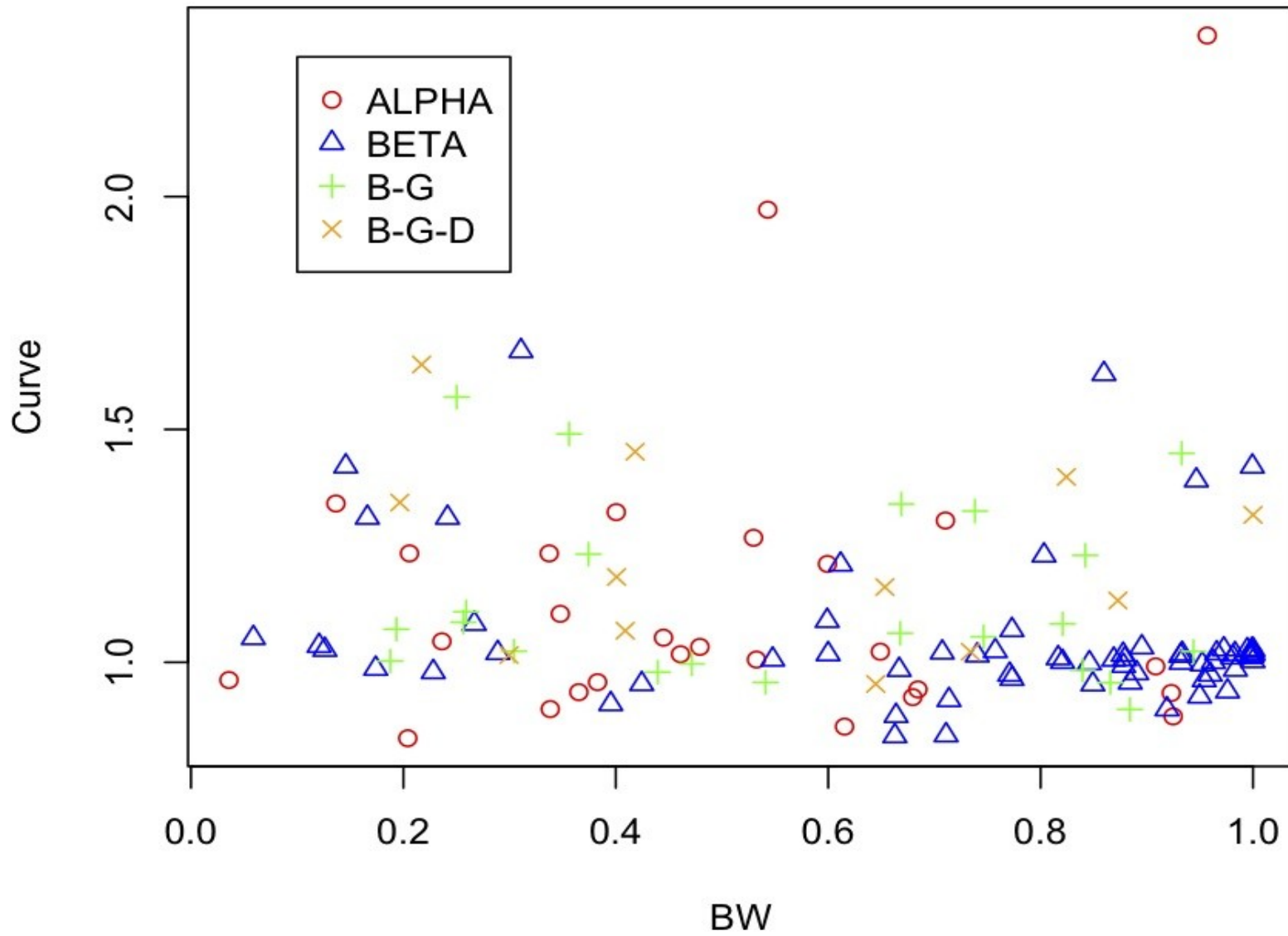
AW vs. Curve



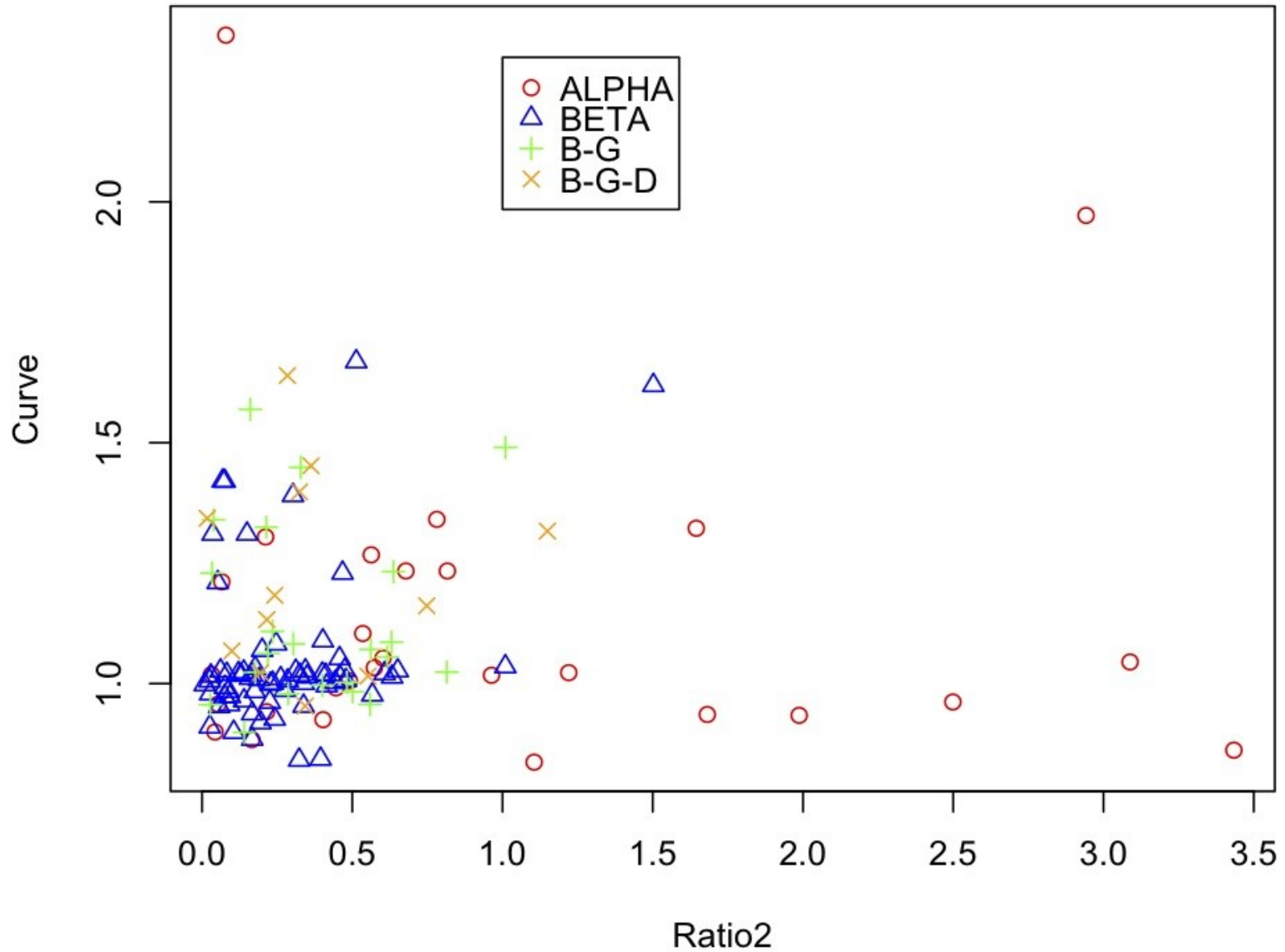
BW vs. log(Ratio (2))



BW vs. Curve



log(Ratio (2)) vs. Curve



Delta Spots

Number of images identified as having delta spots (**delta**=1) by class:

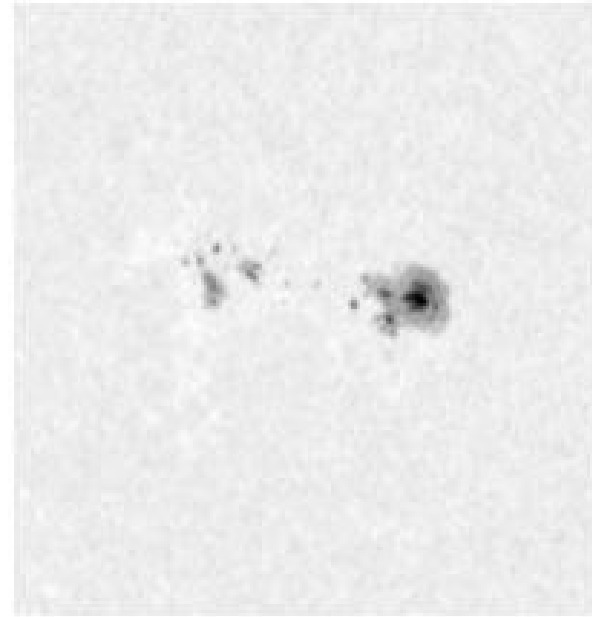
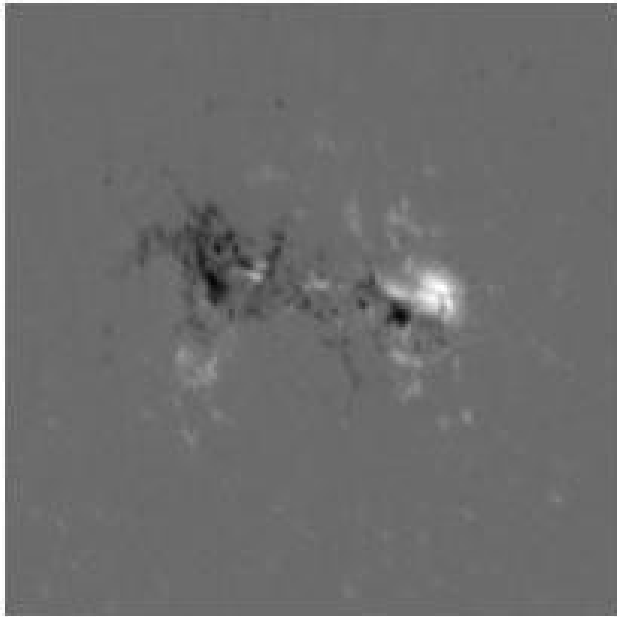
Alpha: 0 out of 28

Beta: 0 out of 66

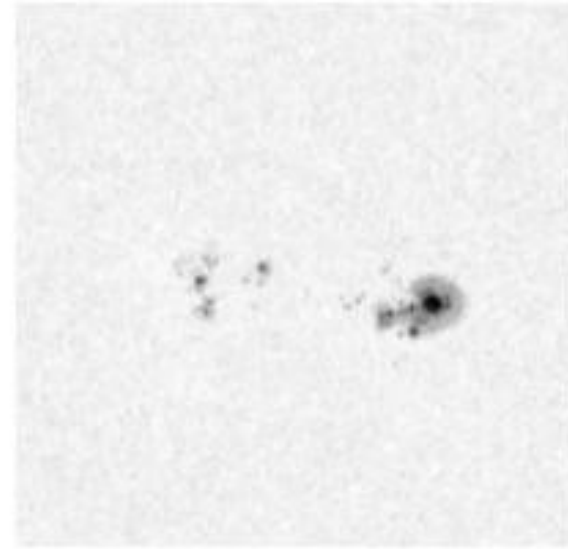
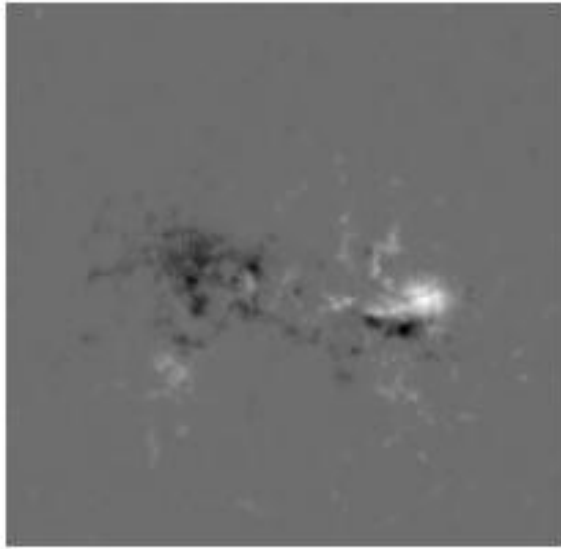
Beta-Gamma: 10 out of 22

Beta-Gamma-Delta: 7 out of 12

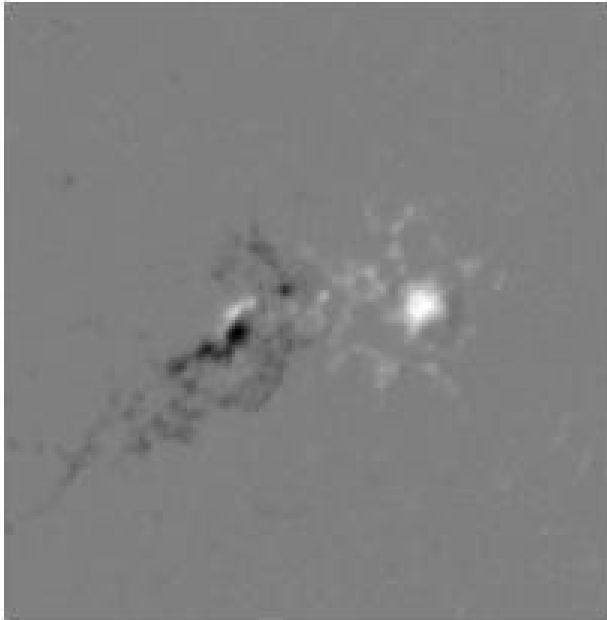
A beta-gamma active-region that was identified as having a delta spot:



The same region 24 hours later, now identified as beta-gamma-delta:



A beta-gamma-delta active-region that was not identified as having a delta spot:



Off the Shelf Classifier: Multinomial Logistic Regression

- Results from multinomial logistic regression using all available features:

		Actual			
		Alpha	Beta	B-G	B-G-D
Predicted	Alpha	17	4	3	2
	Beta	11	62	9	3
	B-G	0	0	8	6
	B-G-D	0	0	2	1

Off the Shelf Classifier: Multinomial Logistic Regression

- Results from multinomial logistic regression using all available features, including interactions and quadratic terms:

Actual		Alpha	Beta	B-G	B-G-D
Predicted					
	Alpha	19	4	3	1
	Beta	9	62	7	3
	B-G	0	0	9	3
	B-G-D	0	0	3	5

Next Steps

- Tweak features to obtain better separation between the classes
 - Examine the cases where the delta classifier is failing.
 - Since many beta-gammas are identified as having delta spots (and many beta-gamma-deltas were not identified as such), perhaps modify the delta feature to make it more continuous.
- Consider quadratic discriminant analysis and other methods to obtain more complex decision boundaries.

Acknowledgements

This work is joint with:

- V. Kashyap
- T. C. M. Lee
- D. van Dyk
- C. A. Young

References

- R. Adams and L. Bischof, “Seeded region growing,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 641–647, 1994. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=295913>
- J. Ireland, C. A. Young, R. T. J. McAteer, C. Whelan, R. J. Hewett, and P. T. Gallagher, “Multiresolution analysis of active region magnetic structure and its correlation with the mt. wilson classification and flaring activity,” *Solar Physics*, 2008. [Online]. Available: <http://arxiv.org/abs/0805.01>
- P. Soille, *Morphological Image Analysis*, P. Soille, Ed. Springer-Verlag, 2003, vol. 132.01