Automatic Classification and Tracking of Solar Features

David Stenning[1] Vinay Kashyap[2] Thomas Lee[3]
David van Dyk[4] C. Alex Young[5]

[1]Department of Statistics, University of California, Irvine

[2]Smithsonian Astrophysical Observatory

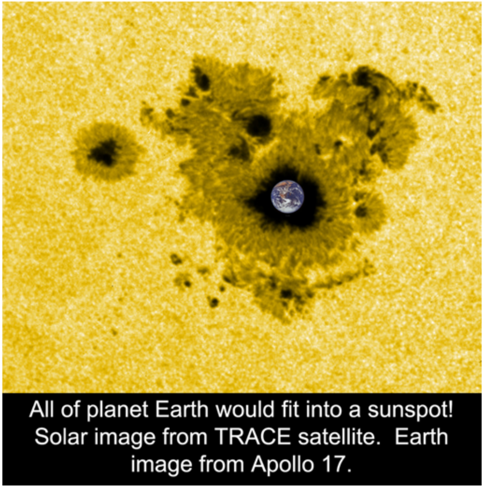
[3]Department of Statistics, University of California, Davis

[4]Statistics Section, Imperial College London

[5]NASA Goddard Space Flight Center

Outline

Sunspots

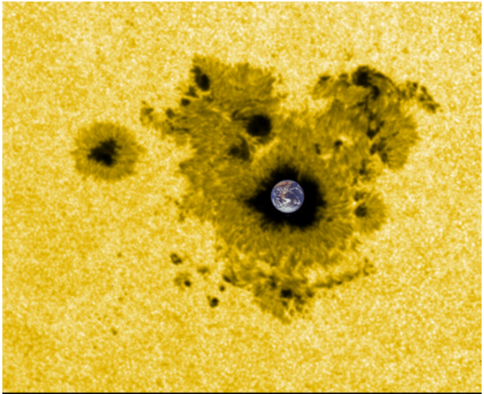


All of planet Earth would fit into a sunspot! Solar image from TRACE satellite. Earth image from Apollo 17.

Image Credit: NASA

- *Sunspots* appear as dark regions on the photosphere.
- They are formed when intense magnetic fields inhibit convection, cooling the associated surface area.
- These areas appear as dark spots when viewed in optical (white) light.
- (show sunspot_evolve movie)

Sunspots

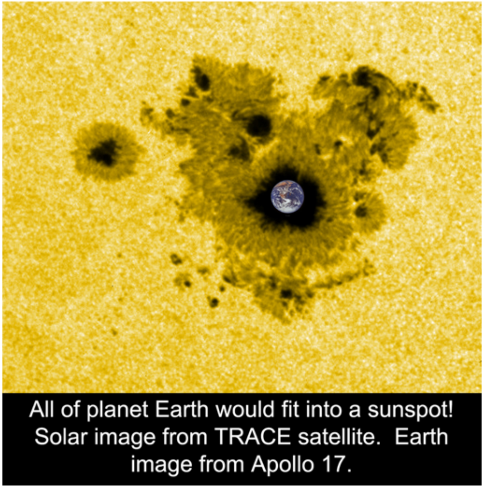


All of planet Earth would fit into a sunspot! Solar image from TRACE satellite. Earth image from Apollo 17.

Image Credit: NASA

- *Sunspots* appear as dark regions on the photosphere.
- They are formed when intense magnetic fields inhibit convection, cooling the associated surface area.
- These areas appear as dark spots when viewed in optical (white) light.
- (show sunspot_evolve movie)

Sunspots

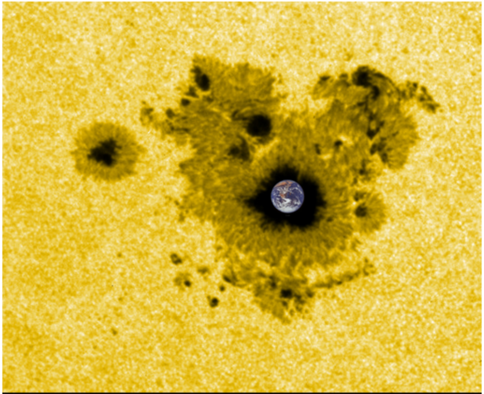


All of planet Earth would fit into a sunspot! Solar image from TRACE satellite. Earth image from Apollo 17.

Image Credit: NASA

- *Sunspots* appear as dark regions on the photosphere.
- They are formed when intense magnetic fields inhibit convection, cooling the associated surface area.
- These areas appear as dark spots when viewed in optical (white) light.
- (show sunspot_evolve movie)

Sunspots

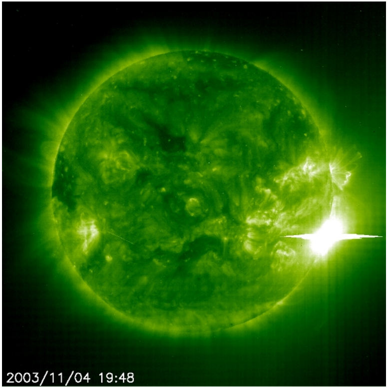


All of planet Earth would fit into a sunspot! Solar image from TRACE satellite. Earth image from Apollo 17.

Image Credit: NASA

- *Sunspots* appear as dark regions on the photosphere.
- They are formed when intense magnetic fields inhibit convection, cooling the associated surface area.
- These areas appear as dark spots when viewed in optical (white) light.
- (show sunspot_evolve movie)

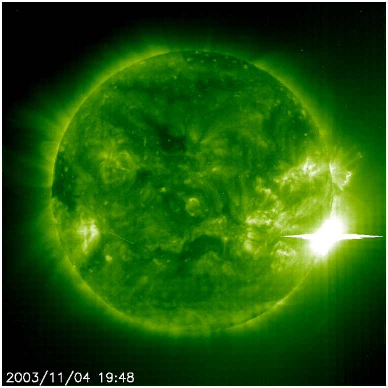
Solar Flares



A powerful solar flare captured by the Extreme Ultraviolet Imager in the 195A emmision lineaboard the SOHO spacecraft. **Image Credit:** SOHO, ESA & NASA

- Sunspots are associated with highly energetic *solar flares*.
- A *solar flare* is a sudden burst of radiation caused by a large release of (magnetic) energy.
- Flares appears as bright areas on the sun and they eject charged particles into space.
- (Show solar_flare movie)

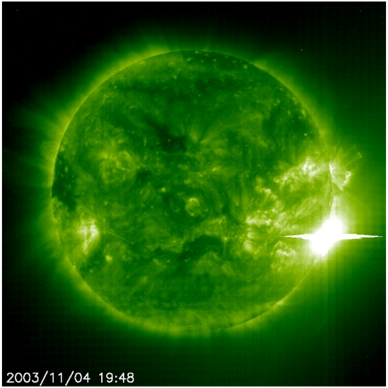
Solar Flares



A powerful solar flare captured by the Extreme Ultraviolet Imager in the 195A emmision lineaboard the SOHO spacecraft. **Image Credit:** SOHO, ESA & NASA

- Sunspots are associated with highly energetic *solar flares*.
- A *solar flare* is a sudden burst of radiation caused by a large release of (magnetic) energy.
- Flares appears as bright areas on the sun and they eject charged particles into space.
- (Show solar_flare movie)

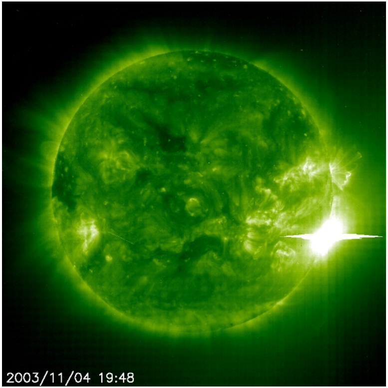
Solar Flares



A powerful solar flare captured by the Extreme Ultraviolet Imager in the 195A emmision lineaboard the SOHO spacecraft. **Image Credit:** SOHO, ESA & NASA

- Sunspots are associated with highly energetic *solar flares*.
- A *solar flare* is a sudden burst of radiation caused by a large release of (magnetic) energy.
- Flares appears as bright areas on the sun and they eject charged particles into space.
- (Show solar_flare movie)

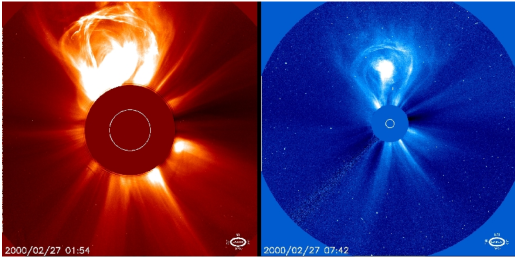
Solar Flares



A powerful solar flare captured by the Extreme Ultraviolet Imager in the 195A emmision lineaboard the SOHO spacecraft. **Image Credit:** SOHO, ESA & NASA

- Sunspots are associated with highly energetic *solar flares*.
- A *solar flare* is a sudden burst of radiation caused by a large release of (magnetic) energy.
- Flares appears as bright areas on the sun and they eject charged particles into space.
- (Show solar_flare movie)

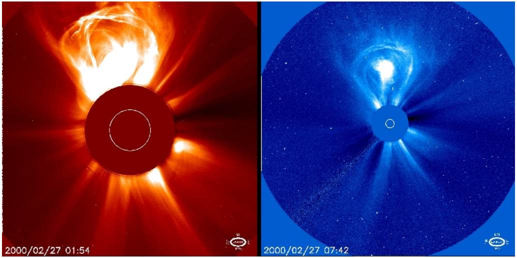
Coronal Mass Ejections



A CME captured by SOHO LASCO C2 and C3. **Image Credit:** SOHO, ESA & NASA

- Sunspots are also related to massive bursts of solar material known as *coronal mass ejections* (CMEs).
- A large CME can eject billions of tons of particles at high velocities.
- CMEs are sometimes associated with solar flares, but can occur independently.
- (show cme, cme2 and flare_and_cme movies)

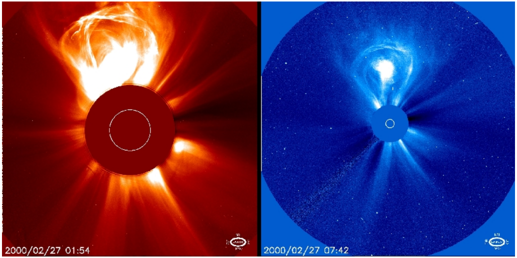
Coronal Mass Ejections



A CME captured by SOHO LASCO C2 and C3. **Image Credit:** SOHO, ESA & NASA

- Sunspots are also related to massive bursts of solar material known as *coronal mass ejections* (CMEs).
- A large CME can eject billions of tons of particles at high velocities.
- CMEs are sometimes associated with solar flares, but can occur independently.
- (show cme, cme2 and flare_and_cme movies)

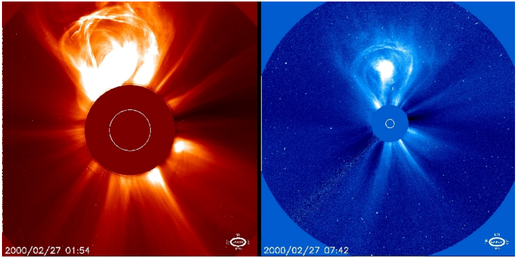
Coronal Mass Ejections



A CME captured by SOHO LASCO C2 and C3. **Image Credit:** SOHO, ESA & NASA

- Sunspots are also related to massive bursts of solar material known as *coronal mass ejections* (CMEs).
- A large CME can eject billions of tons of particles at high velocities.
- CMEs are sometimes associated with solar flares, but can occur independently.
- (show cme, cme2 and flare_and_cme movies)

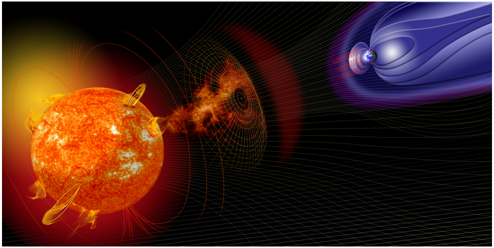
Coronal Mass Ejections



A CME captured by SOHO LASCO C2 and C3. **Image Credit:** SOHO, ESA & NASA

- Sunspots are also related to massive bursts of solar material known as *coronal mass ejections* (CMEs).
- A large CME can eject billions of tons of particles at high velocities.
- CMEs are sometimes associated with solar flares, but can occur independently.
- (show cme, cme2 and flare_and_cme movies)

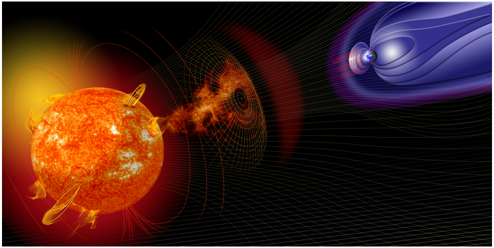
Space Weather Effects



Artist illustration of events on the sun changing the conditions in Near-Earth space. **Image Credit:** NASA

- The highly energetic particles released by solar flares and CMEs can impact the Earth's magnetosphere.
- These impacts cause radio interference and can damage satellites and electric power transmission.
- Strong solar flares and CMEs have been known to cause power failures and blackouts on Earth.

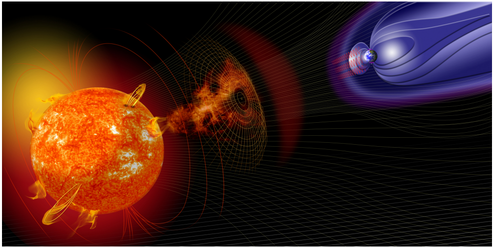
Space Weather Effects



Artist illustration of events on the sun changing the conditions in Near-Earth space. **Image Credit:** NASA

- The highly energetic particles released by solar flares and CMEs can impact the Earth's magnetosphere.
- These impacts cause radio interference and can damage satellites and electric power transmission.
- Strong solar flares and CMEs have been known to cause power failures and blackouts on Earth.

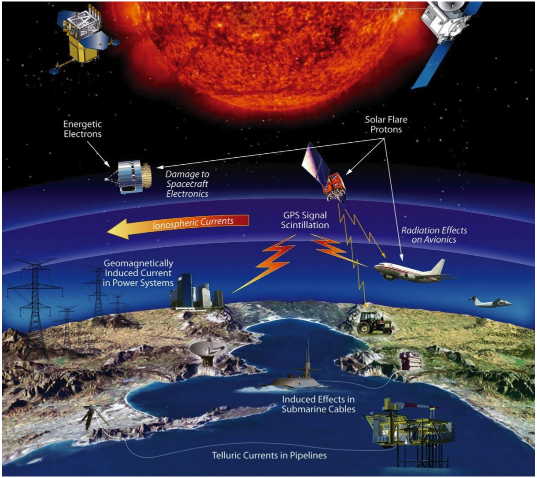
Space Weather Effects



Artist illustration of events on the sun changing the conditions in Near-Earth space. **Image Credit:** NASA

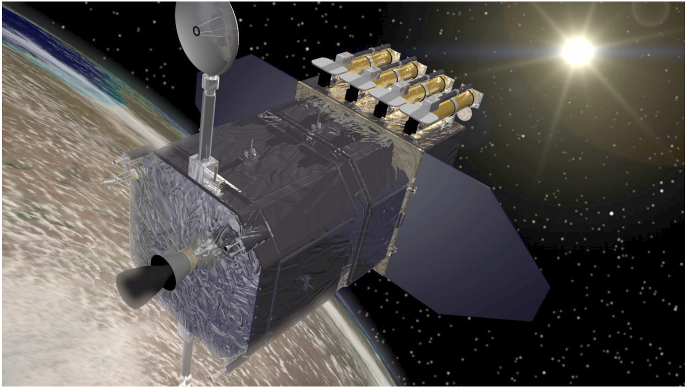
- The highly energetic particles released by solar flares and CMEs can impact the Earth's magnetosphere.
- These impacts cause radio interference and can damage satellites and electric power transmission.
- Strong solar flares and CMEs have been known to cause power failures and blackouts on Earth.

Space Weather Effects



Technological infrastructure affected by space weather events. **Image Credit:** NASA

Solar Dynamics Observatory



The Solar Dynamics Observatory (SDO) was launched on February 11, 2010 with the goal of better understanding the Sun's influence on Earth and near-Earth space. **Image Credit:** NASA

Data Volume

- Compared to older generation observatories, SDO generates an enormous volume of solar data with a continuous science data downlink rate of 130 Megabits per second.
- This is enough data to fill a typical CD every 36 seconds.
- Manual tracking and classification of sunspots—the method currently in practice by experts—is impossible in real-time with massive data streams.
- The data is also of higher quality, necessitating the development of sophisticated, robust, and automatic analysis procedures.

Data Volume

- Compared to older generation observatories, SDO generates an enormous volume of solar data with a continuous science data downlink rate of 130 Megabits per second.
- This is enough data to fill a typical CD every 36 seconds.
- Manual tracking and classification of sunspots—the method currently in practice by experts—is impossible in real-time with massive data streams.
- The data is also of higher quality, necessitating the development of sophisticated, robust, and automatic analysis procedures.

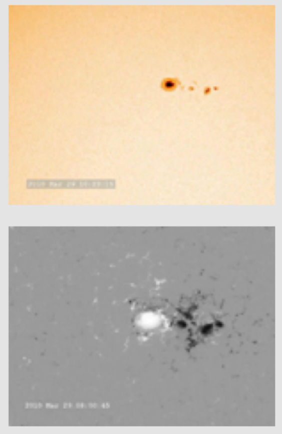
Data Volume

- Compared to older generation observatories, SDO generates an enormous volume of solar data with a continuous science data downlink rate of 130 Megabits per second.
- This is enough data to fill a typical CD every 36 seconds.
- Manual tracking and classification of sunspots—the method currently in practice by experts—is impossible in real-time with massive data streams.
- The data is also of higher quality, necessitating the development of sophisticated, robust, and automatic analysis procedures.

Data Volume

- Compared to older generation observatories, SDO generates an enormous volume of solar data with a continuous science data downlink rate of 130 Megabits per second.
- This is enough data to fill a typical CD every 36 seconds.
- Manual tracking and classification of sunspots—the method currently in practice by experts—is impossible in real-time with massive data streams.
- The data is also of higher quality, necessitating the development of sophisticated, robust, and automatic analysis procedures.

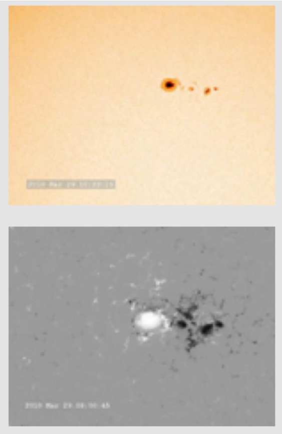
White-light Images and Magnetograms



White-light image (top) and magnetogram (bottom) captured by SDO **Image Credit:** SDO & NASA

- Sunspots are primarily visible in white-light images.
- Magnetogram images provide a representation of the Sun's magnetic fields, and regions of high magnetic flux are associated with sunspots.

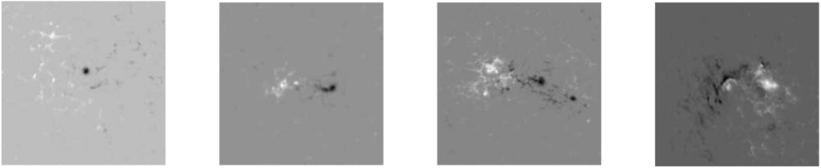
White-light Images and Magnetograms



White-light image (top) and magnetogram (bottom) captured by SDO **Image Credit:** SDO & NASA

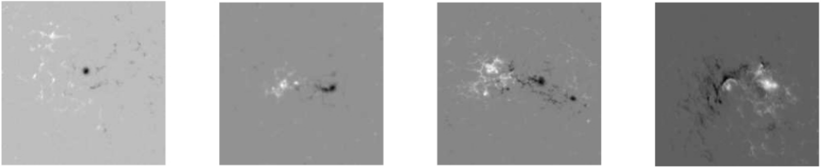
- Sunspots are primarily visible in white-light images.
- Magnetogram images provide a representation of the Sun's magnetic fields, and regions of high magnetic flux are associated with sunspots.

Project Goals



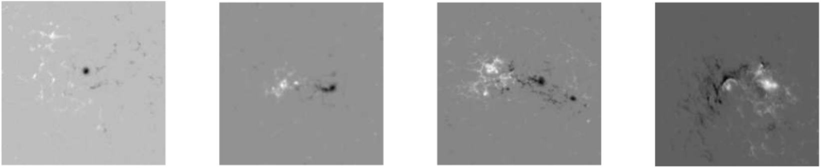
- Our goal is automated classification, tracking and prediction of energetic solar activity and effects on near-Earth space.
- First step: classification of sunspots.

Project Goals



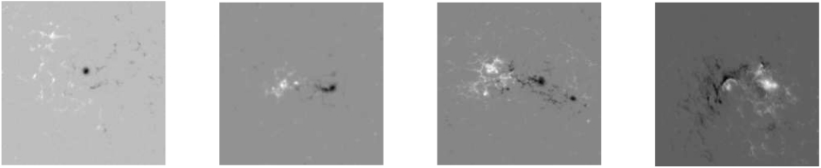
- Our goal is automated classification, tracking and prediction of energetic solar activity and effects on near-Earth space.
- First step: classification of sunspots.

Mount Wilson Classification



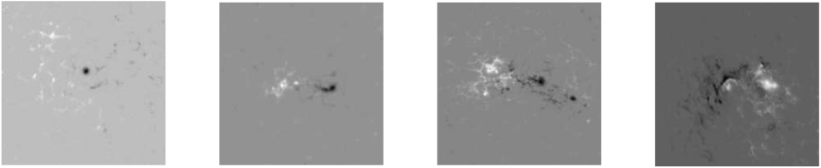
- The *Mount Wilson* scheme puts sunspot groups into four broad classes— α, β, $\beta\gamma$, and $\beta\gamma\delta$ based on the complexity of magnetic flux distribution.
- It is ideal for manual classification since it is based on a few simple rules.
- It has some power for predicting activity in the solar corona and, in particular, the $\beta\gamma\delta$ class is known to be associated with flare activity.

Mount Wilson Classification



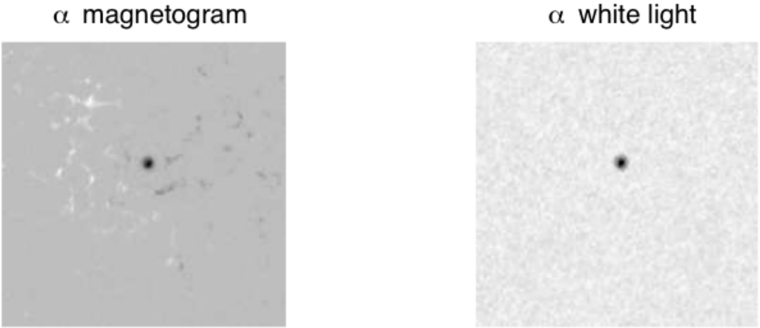
- The *Mount Wilson* scheme puts sunspot groups into four broad classes— α, β, $\beta\gamma$, and $\beta\gamma\delta$ based on the complexity of magnetic flux distribution.
- It is ideal for manual classification since it is based on a few simple rules.
- It has some power for predicting activity in the solar corona and, in particular, the $\beta\gamma\delta$ class is known to be associated with flare activity.

Mount Wilson Classification



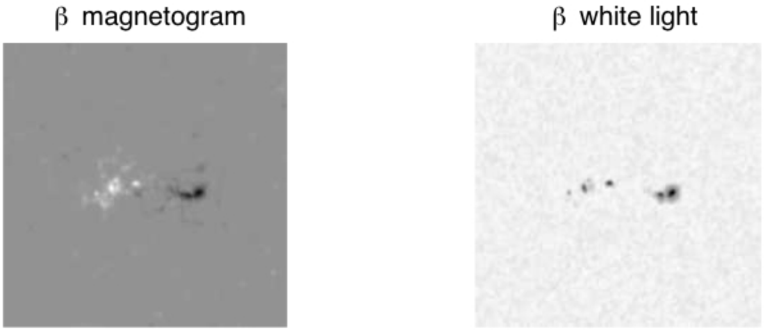
- The *Mount Wilson* scheme puts sunspot groups into four broad classes— α, β, $\beta\gamma$, and $\beta\gamma\delta$ based on the complexity of magnetic flux distribution.
- It is ideal for manual classification since it is based on a few simple rules.
- It has some power for predicting activity in the solar corona and, in particular, the $\beta\gamma\delta$ class is known to be associated with flare activity.

α Class



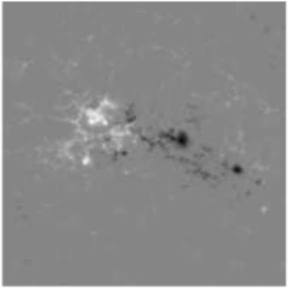
α : a single unipolar spot which may be linked with plage of opposite magnetic polarity.

β Class

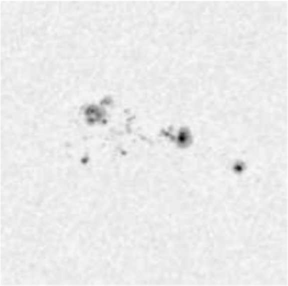


β: a pair of spots with opposite magnetic polarity (bipolar), but with a simple and distinct spatial division between the polarities.

$\beta\gamma$ Class



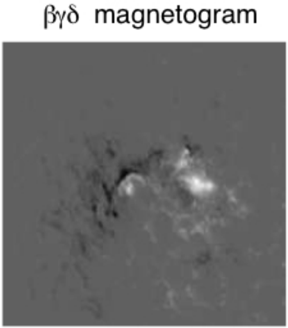
βγ magnetogram

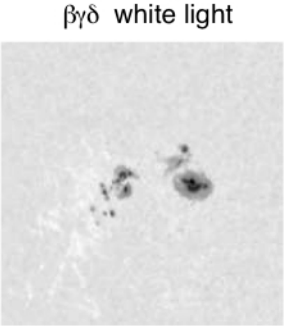


βγ white light

$\beta\gamma$: a bipolar group sufficiently complex that a straight line cannot divide the two polarities.

$\beta\gamma\delta$ Class





$\beta\gamma\delta$: a bipolar $\beta\gamma$ group with umbrae of opposite polarity inside a single penumbra.

Drawbacks of Manual Classification

- The discrete classification is artificial.
- The morphology of active regions is continuous and sunspot groups evolve from one class to another in short periods of time. (show sunspot_evolve movie)
- The classification of particular sunspot groups is often ambiguous and subject to human observer bias.
- Multiple sunspot groups/active regions may mix, particularly when the solar cycle is at a peak.

Drawbacks of Manual Classification

- The discrete classification is artificial.
- The morphology of active regions is continuous and sunspot groups evolve from one class to another in short periods of time. (show sunspot_evolve movie)
- The classification of particular sunspot groups is often ambiguous and subject to human observer bias.
- Multiple sunspot groups/active regions may mix, particularly when the solar cycle is at a peak.

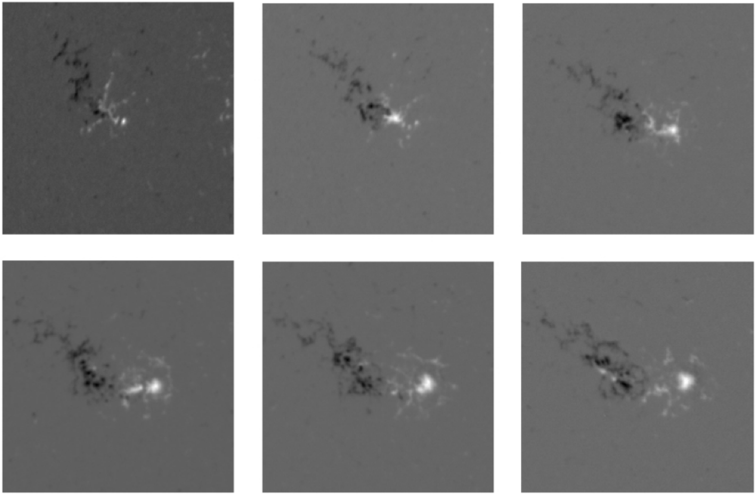
Drawbacks of Manual Classification

- The discrete classification is artificial.
- The morphology of active regions is continuous and sunspot groups evolve from one class to another in short periods of time. (show sunspot_evolve movie)
- The classification of particular sunspot groups is often ambiguous and subject to human observer bias.
- Multiple sunspot groups/active regions may mix, particularly when the solar cycle is at a peak.

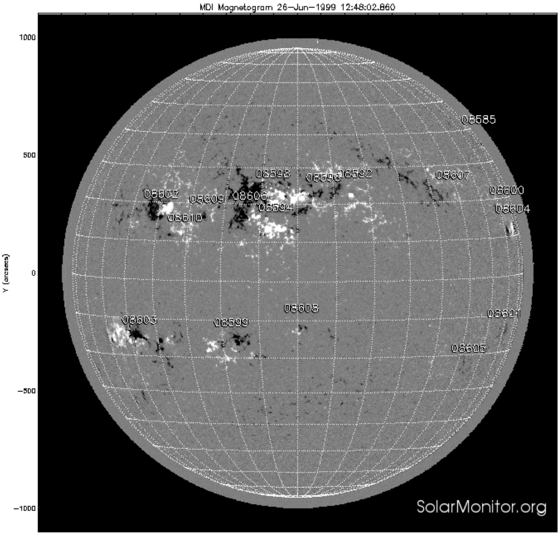
Drawbacks of Manual Classification

- The discrete classification is artificial.
- The morphology of active regions is continuous and sunspot groups evolve from one class to another in short periods of time. (show sunspot_evolve movie)
- The classification of particular sunspot groups is often ambiguous and subject to human observer bias.
- Multiple sunspot groups/active regions may mix, particularly when the solar cycle is at a peak.

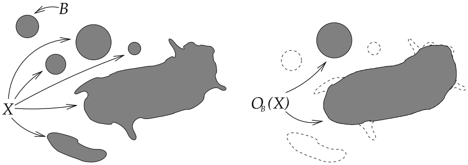
Evolving Active Region



Merging Active Regions

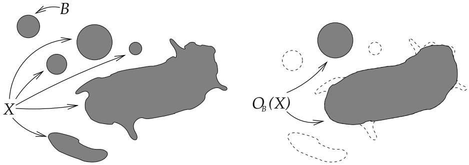


Automatic Mount Wilson Classification



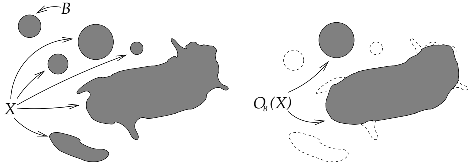
- We develop an automated version of the Mount Wilson classification scheme.
- Our approach uses science-driven feature selection—we create features that have a scientific basis and would be interpretable to a solar physicist.
- Since sunspot groups are classified according to their morphology, we use *mathematical morphology* to create numerical summaries of the magnetic flux distribution in active regions.

Automatic Mount Wilson Classification



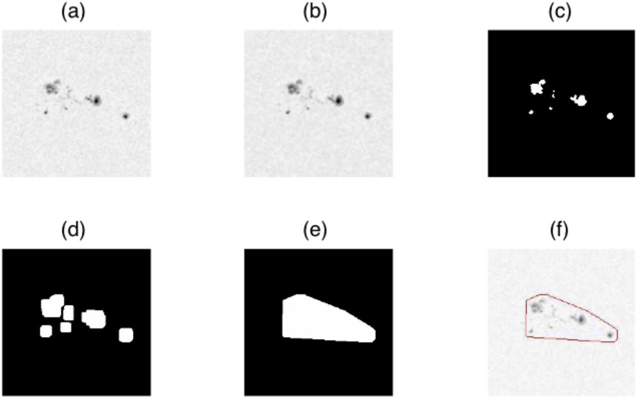
- We develop an automated version of the Mount Wilson classification scheme.
- Our approach uses science-driven feature selection—we create features that have a scientific basis and would be interpretable to a solar physicist.
- Since sunspot groups are classified according to their morphology, we use *mathematical morphology* to create numerical summaries of the magnetic flux distribution in active regions.

Automatic Mount Wilson Classification



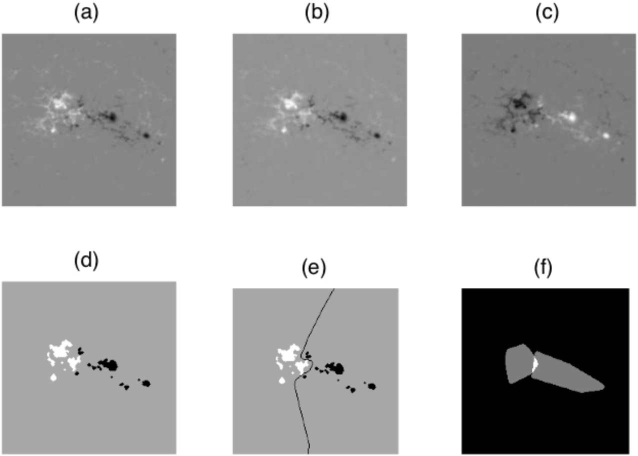
- We develop an automated version of the Mount Wilson classification scheme.
- Our approach uses science-driven feature selection—we create features that have a scientific basis and would be interpretable to a solar physicist.
- Since sunspot groups are classified according to their morphology, we use *mathematical morphology* to create numerical summaries of the magnetic flux distribution in active regions.

Locating Sunspots in White-light Data: $\beta\gamma$ Example



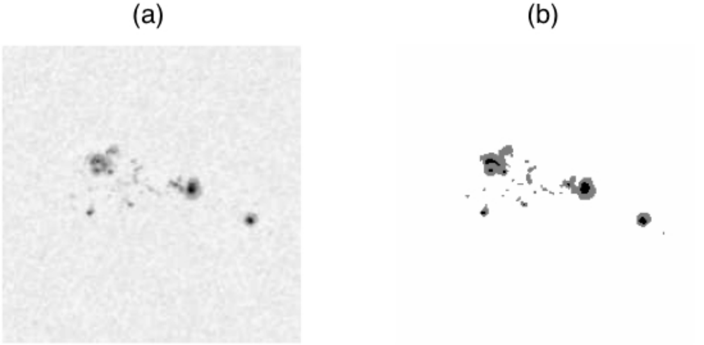
(a) The original white light image for the $\beta\gamma$ class example; (b) the morphologically closed image; (c) the sunspot pixels identified by thresholding; (d) the dilated sunspot pixels; (e) convex hull around the dilated suspot pixels; (f) sunspot region where we will look for magnetic flux in the magnetogram.

Obtaining Active Region Morphology



(a) Magnetogram $\beta\gamma$ class example; (b) morphologically smoothed magnetogram; (c) morphologically smoothed inverse magnetogram; (d) active region pixels; (e) separating line obtained via seeded region growing; (f) convex hulls around opposite polarities

Finding Delta Spots



(a) original white light image; (b) smoothed and thresholded image to distinguish umbrae/penumbrae

Science-Driven Features

We use our morphological representation of sunspot groups and active regions to obtain scientifically based numerical features.

- The *ratio* of pixels of opposite polarities.
- The *amount of scattering* of the pixels for each polarity.
- Polarity inversion line *curvature*.
- Area of *mixture* for the convex hulls around each polarity region.
- The *number of delta spots* detected.

Science-Driven Features

We use our morphological representation of sunspot groups and active regions to obtain scientifically based numerical features.

- The *ratio* of pixels of opposite polarities.
- The *amount of scattering* of the pixels for each polarity.
- Polarity inversion line *curvature*.
- Area of *mixture* for the convex hulls around each polarity region.
- The *number of delta spots* detected.

Science-Driven Features

We use our morphological representation of sunspot groups and active regions to obtain scientifically based numerical features.

- The *ratio* of pixels of opposite polarities.
- The *amount of scattering* of the pixels for each polarity.
- Polarity inversion line *curvature*.
- Area of *mixture* for the convex hulls around each polarity region.
- The *number of delta spots* detected.

Science-Driven Features

We use our morphological representation of sunspot groups and active regions to obtain scientifically based numerical features.

- The *ratio* of pixels of opposite polarities.
- The *amount of scattering* of the pixels for each polarity.
- Polarity inversion line *curvature*.
- Area of *mixture* for the convex hulls around each polarity region.
- The *number of delta spots* detected.

Science-Driven Features

We use our morphological representation of sunspot groups and active regions to obtain scientifically based numerical features.

- The *ratio* of pixels of opposite polarities.
- The *amount of scattering* of the pixels for each polarity.
- Polarity inversion line *curvature*.
- Area of *mixture* for the convex hulls around each polarity region.
- The *number of delta spots* detected.

Science-Driven Features

We use our morphological representation of sunspot groups and active regions to obtain scientifically based numerical features.

- The *ratio* of pixels of opposite polarities.
- The *amount of scattering* of the pixels for each polarity.
- Polarity inversion line *curvature.*
- Area of *mixture* for the convex hulls around each polarity region.
- The *number of delta spots* detected.

Classification: Random Forests

- The features we have derived—pixel ratio, amount of scattering, separating line curvature, polarity mixture, and number of delta spots—are used in a supervised learning algorithm.
- We use random forests to do classification.
  - nonparametric and highly accurate
  - effecient on large datasets
  - can include many features
  - easily implemented using the `randomForest` package in R

Classification: Random Forests

- The features we have derived—pixel ratio, amount of scattering, separating line curvature, polarity mixture, and number of delta spots—are used in a supervised learning algorithm.
- We use random forests to do classification.
    - nonparametric and highly accurate
    - effecient on large datasets
    - can include many features
    - easily implemented using the `randomForest` package in R

Clustering: K-means

- We are interested in whether our science-driven features provide useful information when the manual classification—which is not always reliable—is ignored.
- Unsupervised learning via k-means clustering allows us to assess how our features discriminate active region morphology.
- This is the first step in developing a classification scheme based on a continuum of classes.

Clustering: K-means

- We are interested in whether our science-driven features provide useful information when the manual classification—which is not always reliable—is ignored.
- Unsupervised learning via k-means clustering allows us to assess how our features discriminate active region morphology.
- This is the first step in developing a classification scheme based on a continuum of classes.

Clustering: K-means

- We are interested in whether our science-driven features provide useful information when the manual classification—which is not always reliable—is ignored.
- Unsupervised learning via k-means clustering allows us to assess how our features discriminate active region morphology.
- This is the first step in developing a classification scheme based on a continuum of classes.

Training and Test Data

The dataset we use to validate our methodology was separated into 65% training and 35% testing with class breakdown as follows:

	Training Set	**Test Set**
α	17	10
β	40	23
$\beta\gamma$	11	6
$\beta\gamma\delta$	7	5

Random Forest Results

Using a random forest with 250 trees we obtain the following results on the test dataset:

		Actual Classification			
		α	β	$\beta\gamma$	$\beta\gamma\delta$
Predicted Classification	α	8	2		
	β	2	21	2	
	$\beta\gamma$			2	1
	$\beta\gamma\delta$			2	4

Incorrect Classification?

- Perfect classification is not necessarily the gold standard when automating a manual classification that is artificial and subjective.
- Classification “by eye” is prone to error and inconsistencies.
    - Two experts looking at the same images will not have 100% agreement.
- This makes it difficult to judge the true performance of our random forest classifier.

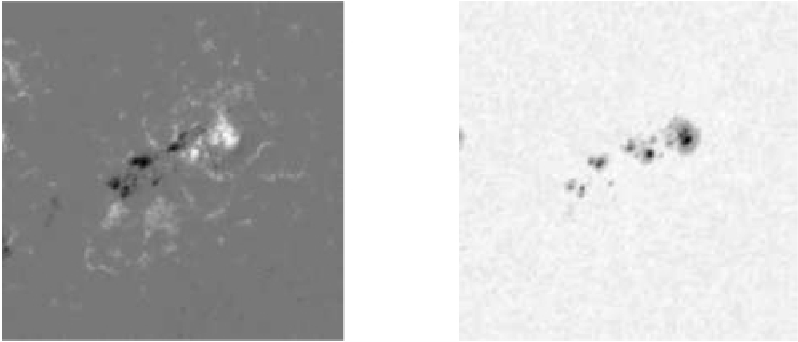
Incorrect Classification?

- Perfect classification is not necessarily the gold standard when automating a manual classification that is artificial and subjective.
- Classification "by eye" is prone to error and inconsistencies.
  - Two experts looking at the same images will not have 100% agreement.
- This makes it difficult to judge the true performance of our random forest classifier.

Incorrect Classification?

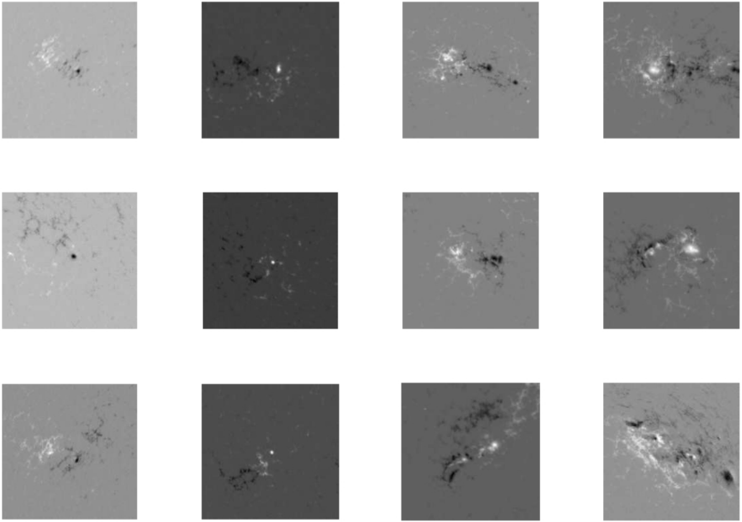
- Perfect classification is not necessarily the gold standard when automating a manual classification that is artificial and subjective.
- Classification "by eye" is prone to error and inconsistencies.
    - Two experts looking at the same images will not have 100% agreement.
- This makes it difficult to judge the true performance of our random forest classifier.

Example: $\beta\gamma$ Misclassification



This $\beta\gamma$ active region was "misclassified" as $\beta\gamma\delta$ by the random forest classifier, but the presence of a δ in the center of the active region is ambiguous.

K-means (K=4) Results by Column



Alternatives to Discrete Classification

- The distribution of magnetic flux polarity for merging active regions is not considered under discrete classification schemes.
- The added complexity from merging active regions may be predictive of flares and CMEs.
- A continuous "classification" based on the distribution of pieces of opposite polarity flux may better capture the complex and evolving nature of sunspots.
- Such a classification may also allow for better automatic prediction of energetic events.

Alternatives to Discrete Classification

- The distribution of magnetic flux polarity for merging active regions is not considered under discrete classification schemes.
- The added complexity from merging active regions may be predictive of flares and CMEs.
- A continuous “classification” based on the distribution of pieces of opposite polarity flux may better capture the complex and evolving nature of sunspots.
- Such a classification may also allow for better automatic prediction of energetic events.

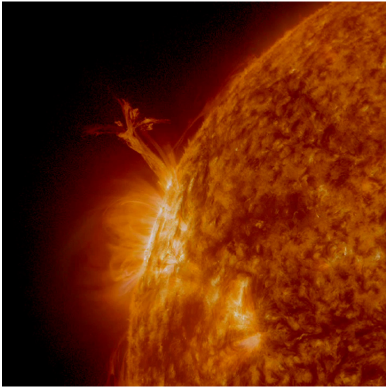
Alternatives to Discrete Classification

- The distribution of magnetic flux polarity for merging active regions is not considered under discrete classification schemes.
- The added complexity from merging active regions may be predictive of flares and CMEs.
- A continuous "classification" based on the distribution of pieces of opposite polarity flux may better capture the complex and evolving nature of sunspots.
- Such a classification may also allow for better automatic prediction of energetic events.

Alternatives to Discrete Classification

- The distribution of magnetic flux polarity for merging active regions is not considered under discrete classification schemes.
- The added complexity from merging active regions may be predictive of flares and CMEs.
- A continuous "classification" based on the distribution of pieces of opposite polarity flux may better capture the complex and evolving nature of sunspots.
- Such a classification may also allow for better automatic prediction of energetic events.

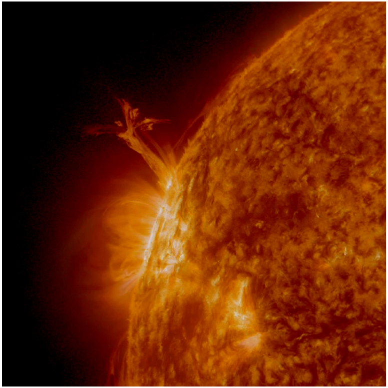
Other Interesting Phenomena



Solar Tornado. **Image Credit:** NASA

- The high cadence of new instruments—such as SDO—allow the possibility of analyzing motion.
- In particular, the rotation of sunspots in relatively short timescales may be related to solar tornados.
- (show solar_tornado movie)

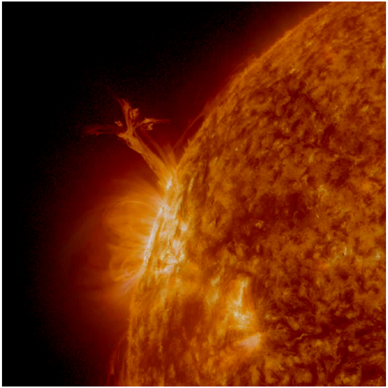
Other Interesting Phenomena



Solar Tornado. **Image Credit:** NASA

- The high cadence of new instruments—such as SDO—allow the possibility of analyzing motion.
- In particular, the rotation of sunspots in relatively short timescales may be related to solar tornados.
- (show solar_tornado movie)

Other Interesting Phenomena



Solar Tornado. **Image Credit:** NASA

- The high cadence of new instruments—such as SDO—allow the possibility of analyzing motion.
- In particular, the rotation of sunspots in relatively short timescales may be related to solar tornados.
- (show solar_tornado movie)

For Further Reading I

Ireland et al.
Multiresolution analysis of active region magnetic structure and its correlation with the mt. wilson classification and flaring activity.
Solar Physics, 2008.

Stenning et al.
Morphological Image Analysis and Its Application to Sunspot Classification.
Statistical Challenges in Modern Astronomy V, 2011.