Segregating Solar Features by Temperature

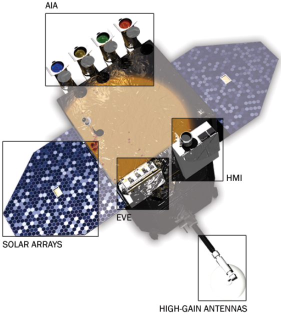
Nathan Stein

California-Boston-Smithsonian AstroStatistics Collaboration

February 8, 2011

Big Picture

- Solar Dynamics Observatory generates up to 1.4 terabytes/day
- Atmospheric Imaging Assembly: four-telescope array on the SDO satellite
- High-resolution (4096×4096) images of the corona in 7 extreme ultraviolet filters every 10 seconds
- More than anyone could examine by eye
- Need fast methods for processing data



Statistical Model

- Observe photon counts $\{Y_{ib}\}$ in pixel i through filter b
- γ_i = unknown amount of plasma in pixel i
- θ_{ij} = proportion of plasma in temperature bin j ($\sum_j \theta_{ij} = 1$)
- $\gamma_i \theta_{ij} = \mathrm{DEM}(\log T_j)$
- τ_b = known exposure time
- Λ_{bj} = known response function
- Likelihood:

$$Y_{ib} \sim \mathrm{Poisson}\left(\gamma_i \ \tau_b \sum_j \Lambda_{bj} \ \theta_{ij}\right)$$

- **Goal:** identify regions with similar θ_i

Statistical Model

- γ_i is a nuisance parameter
- $\sum_b Y_{ib} = N_i \sim \text{Pois}\left(\gamma_i \sum_j M_j \theta_{ij}\right)$, with $M_j = \sum_b \tau_b \Lambda_{bj}$
- Distribution of N_i depends on γ_i and θ_i, whereas distribution of

$$\left(\frac{Y_{i1}}{N_i}, \ldots, \frac{Y_{iB}}{N_i}\right) \tag{1}$$

  (conditional on N_i) only depends on θ_i
- Cluster pixels with similar proportions, ignore totals N_i

Clustering probability vectors

- How to cluster vectors of probabilities or proportions?
- Squared Hellinger distance between $\boldsymbol{p}$ and $\boldsymbol{q}$ is

$$\begin{aligned} d_H^2(\boldsymbol{p}, \boldsymbol{q}) &= \frac{1}{2} \sum_b (\sqrt{p_b} - \sqrt{q_b})^2 \\ &= 1 - \sum_b \sqrt{p_b q_b} \end{aligned}$$

- Modify k-means clustering to use Hellinger distance:
  - "h-means clustering"

Clustering probability vectors

- Observations $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$
- Cluster assignments $c_1, \ldots, c_n$
- Cluster centers $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k$

1. Given cluster centers, set

$$c_i = \arg\min_j \; d_H^2(\boldsymbol{p}_i, \boldsymbol{q}_j)$$

2. Given cluster assignments, set

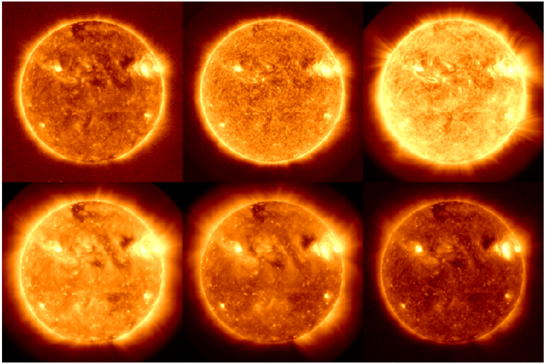
$$\boldsymbol{q}_j = \arg\min_{\boldsymbol{q}} \sum_{i:c_i=j} d_H^2(\boldsymbol{p}_i, \boldsymbol{q}) \tag{2}$$

- The minimization in (2) has an analytic solution:

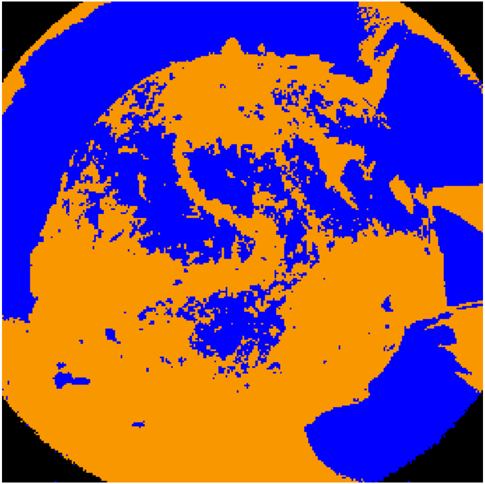
$$q_{jb} = \frac{\left(\sum_{i:c_i=j} \sqrt{p_{ib}}\right)^2}{\sum_{b'} \left(\sum_{i:c_i=j} \sqrt{p_{ib'}}\right)^2}$$

Clustering AIA data

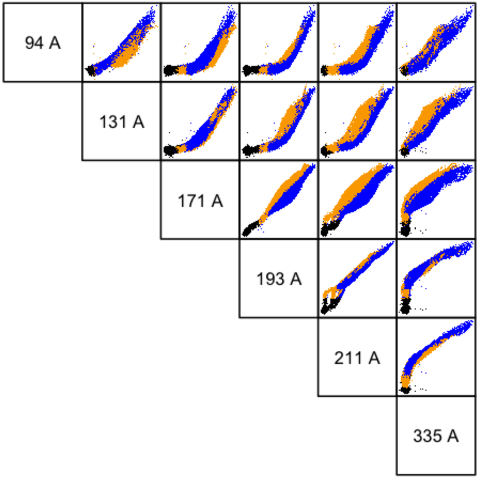
- Cluster the vectors $(y_{i1}/n_i, \ldots, y_{iB}/n_i)$ for $i = 1, \ldots, n$
- For illustration, examine a coarsened (256×256) set of images, with 3 clusters



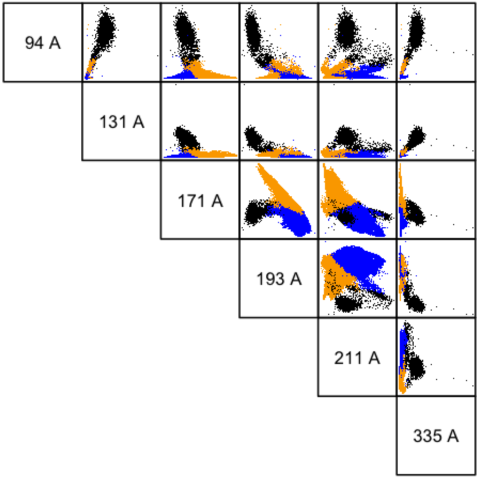
Clustering AIA data



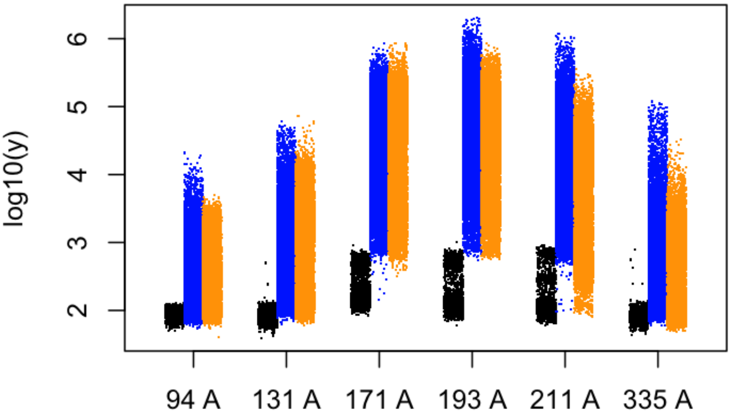
Clusters in $\log Y$ space



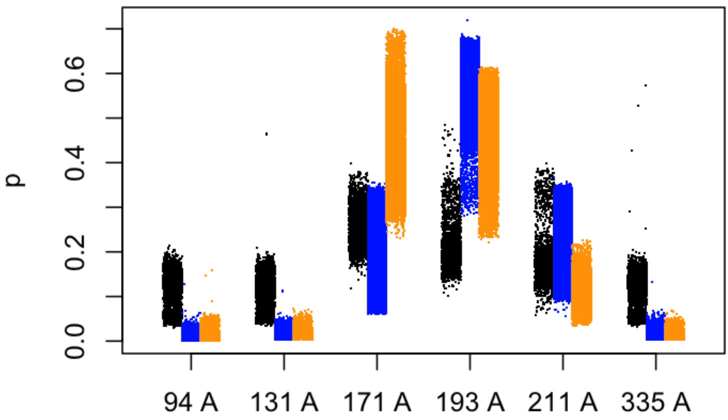
Clusters in Y/n space



Distribution of pixels in each cluster

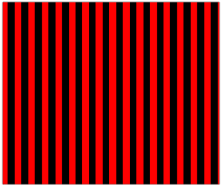


Distribution of pixels in each cluster

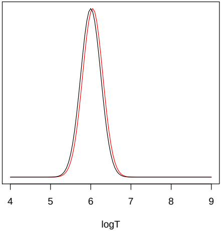


Simulated Data

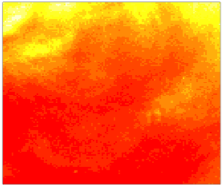
Simulated Temperature Map



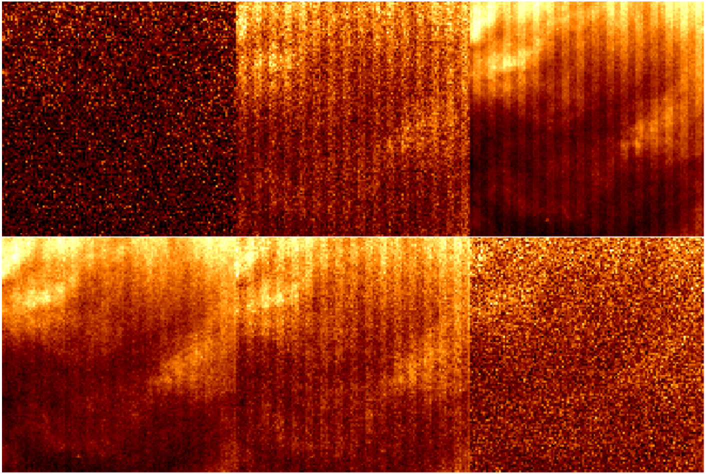
Simulated Temperature Distributions

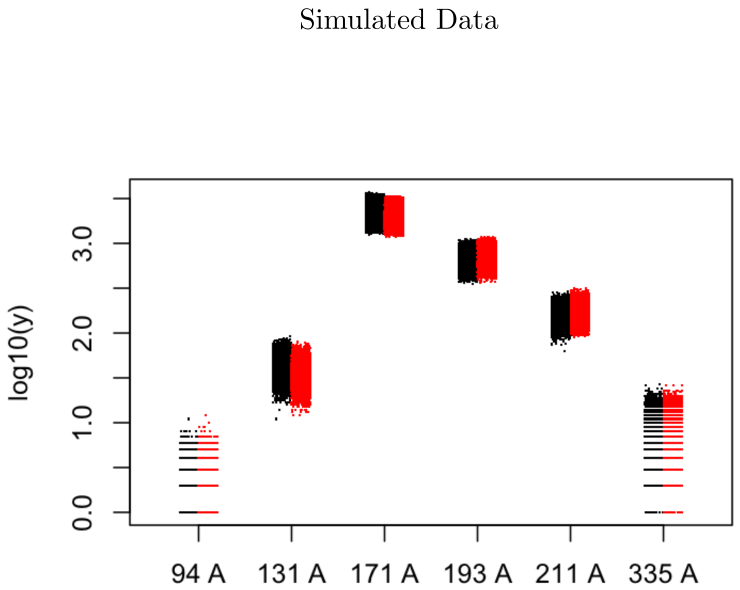


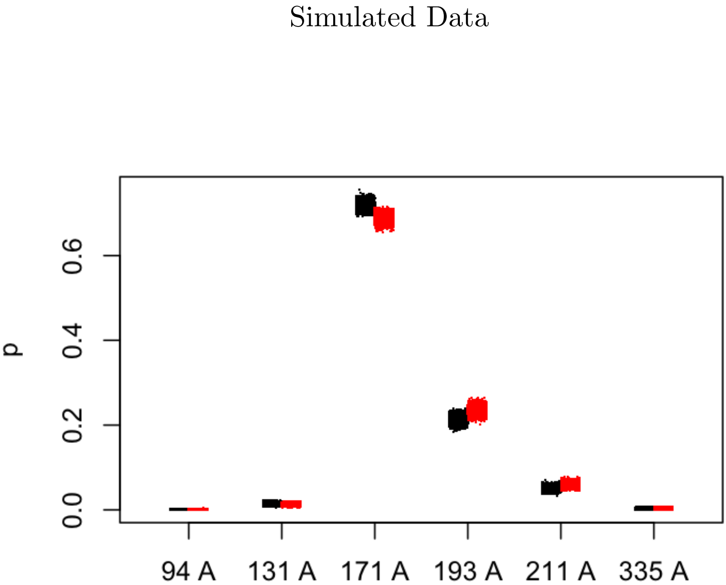
Simulated Map of Total Plasma



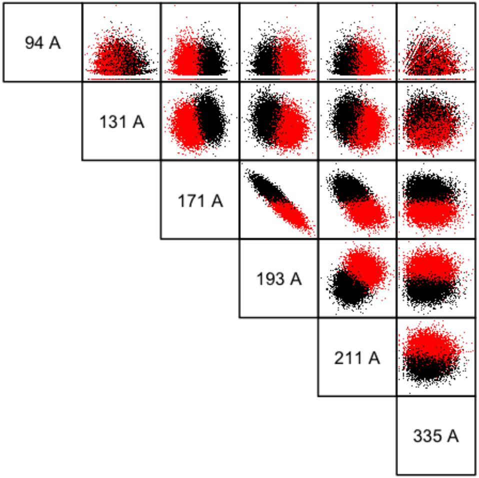
Simulated Data



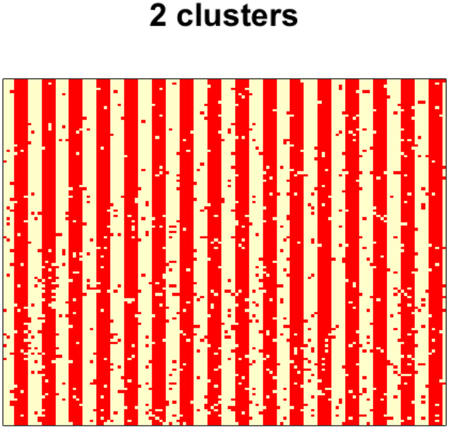

Simulated Data
log10(y)
0.0
1.0
2.0
3.0
94 A
131 A
171 A
193 A
211 A
335 A


Simulated Data
p
0.0
0.2
0.4
0.6
94 A
131 A
171 A
193 A
211 A
335 A

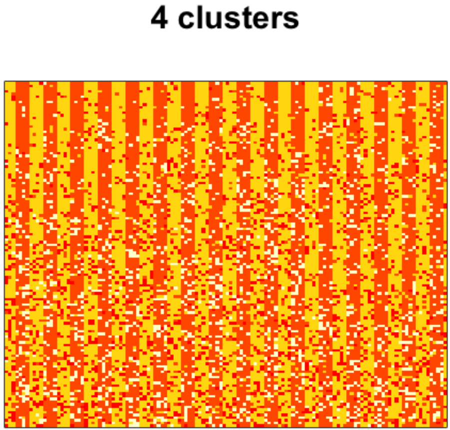
Simulated Data



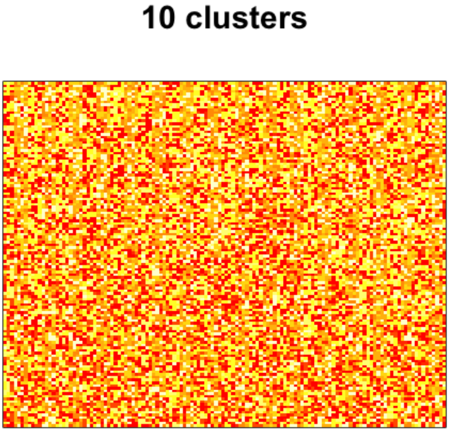
Simulated Data: Results



Simulated Data: Results



Simulated Data: Results

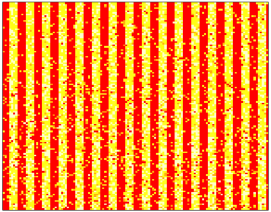


Simulated Data: Results

- How to make images of results more meaningful? (Cluster label is arbitrary number)
- Assign a level l_j to each cluster
- Many choices for quantitatively meaningful l_j's
- For example:

$$l_j = \frac{\sum_{i:c_i=j} y_{i,171}}{\sum_{i:c_i=j} n_i}$$

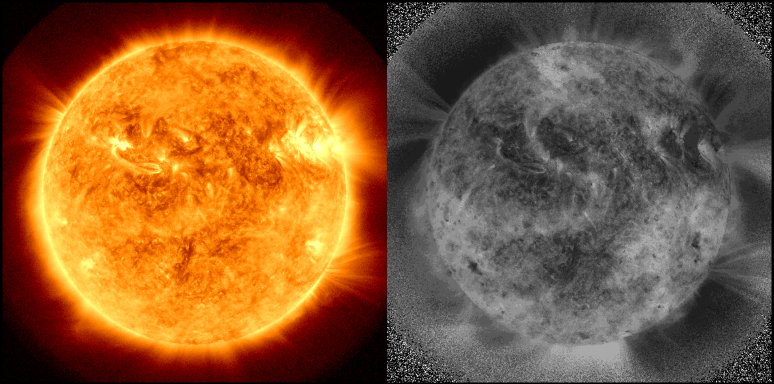
10 clusters



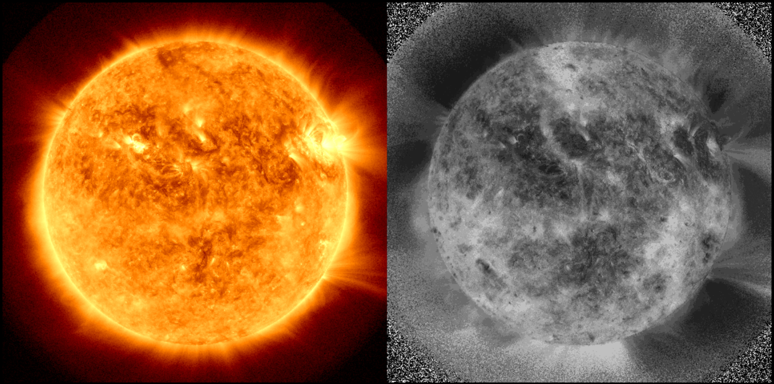
Full Resolution Images

- 4096×4096 pixels $\times$ 6 bands
- $k = 64 + 1$ clusters (1 extra for pixels with zero counts in all 6 images)
- Two observations:
  1. October 2, 2010, 05:57
  2. October 2, 2010, 18:43

05:57



18:43



Next Steps

- Model-based clustering to compare multiple images, to identify regions that are thermally similar in different observations
- DEM reconstruction: what is the state of the art?

Acknowledgements

Vinay Kashyap

Xiao-Li Meng