Cluster analysis of spectroscopic line profiles and EUV emission in RMHD simulations and observations of the solar atmosphere

Viacheslav Sadykov1,2,3, Irina Kitishvili2,3, Alexander Kosovichev1,2
1BAEIR, NASA ARC, NJIT

Spatially-observed emissions from the IRIS, SDO/AIA, and other space mission and ground-based telescopes, coupled with realistic 3D RMHD simulations, are powerful artifacts for analysis of the role of physical processes in the solar atmosphere. To better understand the dynamical and thermodynamic properties in the simulation data and their connection to observations, it is essential to determine similarities in the behavior of the synthesized and observed emission. However, the complexity of observational data and physical processes makes comparison of observations and modeling results difficult. In this work, we show the initial results of application of K-Means clustering (unsupervised machine learning algorithm) to two different problems: recognition of the typical spectroscopic line profiles observed by IRIS during solar flares and their typical dynamic behavior; recognition of shocks and heating events in synthetic AIA emission data obtained from StellarBox quiet-Sun simulations.

The average silhouette width is used for the K-Means algorithm, allowing in different ways to obtain optimal numbers of clusters. We discuss application of the emission clustering to visualizations of the computational volume, understanding its evolutionary trends and behavior patterns, and inversion (reconstruction) of physical properties of the solar atmosphere from synthesizes emission data.

Description of data processing and clustering algorithms

- **Statistical moments of line profiles.** The zeroth moment will represent the maximum or the integrated intensity, second (line width), third (line asymmetry), and higher statistical moments can be computed as:

\[ S_k = \int (a - \bar{a})^k d\lambda / \int d\lambda, \quad (k = 1, 2, 3) \]

- **K-Means clustering.** The K-Means takes the number of clusters as an input parameter, and initially seeds the cluster centers randomly among data points. After this, K-Means assigns the points to belong to the nearest cluster center (i.e. labels them), recomputes the cluster centers as the means among the points of the same labels, and repeats the procedure until there are no changes in the labels of the points.

- **Average silhouette width.** The silhouette is defined for a data point as \( s(i) = (b(i) - a(i)) / \max\{a(i), b(i)\} \)

where \( a(i) \) is the average distance from point \( i \) to points of the same cluster, \( b(i) \) is the average distance from the point \( i \) to points of another closest cluster. The average \( s(i) \) across the points indicates how well the points lie within their clusters.

- The optimal number of clusters can be estimated by maximization of the \( s(i) \).

- When \( s(i) \leq 0 \), the points no longer “belong” to their clusters.

**Identifying typical response of the upper chromosphere and lower transition region for the flare heating from IRIS observations**

- The Interface Region Imaging Spectrograph (IRIS, De Pontieu et al. 2014) has observed hundreds of flares of \( \geq C1.0 \) activity. However, statistical studies of atmospheric response to the flare heating by IRIS are hard to perform because of the complexity of imaging spectroscopy data their high-dimensional space, large data volumes, optically-thick nature of the lines.

- Finding compact illustrative representation of the atmospheric response to the flare heating using unsupervised machine learning clustering techniques can simplify the analysis of large observational datasets and increase their understanding.

- An example of clustering of C II 1334.5 Å line profiles for the M1.0 flare of June 12, 2014, is presented in Figure 1. The mean (Doppler shift) of each representative point indicates that most of the southern part of the flare region exhibits redshifts of the C II line profiles, while the northern part does not show any strong Doppler shift with respect to unperturbed gray line profile. Line clustering was previously used by Sainz Dalda et al. (2018) and Sainz Dalda et al. (2019).

- Figure 2 illustrates the typical evolution of Mg II k 2796 Å line profiles during the M1.8 class solar flare of February 13, 2014. The K-Means clustering was performed simultaneously for the line intensity, Doppler shift, and line width evolution, with equal contribution from each considered statistical moment. The red, blue, and black clusters are of special interest: while red cluster behaves as typically expected during “explosive” chromospheric evaporation, blue and black clusters revealed slight redshift followed by a strong blueshift of the spectral lines.

**Recognitions of shock and heating events from synthetic AIA emission**

- “StellarBox” code solves the fully compressible MHD equations with radiative transfer solve by ray-tracing and opacity bounding techniques, and large-eddy simulation (LES) treatment of subgrid turbulence transport (Wu et al. 2015). The current version of the code supports support to extend the computational domain to corona and deeper convective layers and in horizontal directions.

- The computational domain of 12.8 x 12.8 x 15.2 Mm includes a 10-Mm layer from the photosphere to the low corona. The grid size is 200x200 in the horizontal directions; a variable grid spacing of similar size is used in the vertical direction. The lateral boundary conditions are periodic. For initial conditions of the chromosphere and corona, the model by Vernetta et al. (1981) has been used. The 176 simulation snapshots delivered with 2s temporal cadence are analyzed.

- The line-averaged AIA emission is computed for each snapshot for each column separately, using SDO/AIA temperature response functions available from SSW IDL. Strong impacts ("shocks" hereafter) are observed in AIA running difference images.

- K-Means clustering is performed for sparse selection of columns and snapshots for all AIA channels together. The contribution of each channels was normalized. Seven clusters are used in the result explained below.

- **Preliminary result:** one of the clusters (cyan) correlates well with the shock signatures. The corresponding differential emission measure profile (DEM, cyan) shows the peak at +1MK and contribution from +80fMK plasma.

**How many clusters to select? Inverse problem POV**

- The number of clusters depends on the problem type: some inverse problems (e.g. recognition of tiny features in line profiles) may require selection of more clusters than dictated by the maximization of the average silhouette width.

- In the example above, synthetic AIA emission is a function of the Differential Emission Measure (DEM, Cheung et al. 2015) of the computational domain. Reconstruction of the DEM from AIA emission is an example of ill-posed inverse problem.

- Figure 3 illustrates that, of more than 7 clusters are selected in AIA emission space, the corresponding DEMs of these clusters strongly mix with each other. The average synthetic computed in DEM (generating parameter) space becomes negative, i.e. the points no longer correspond to their clusters in that space.

- In general, the combination of unsupervised clustering in image space and unsupervised clustering in parameter space is well known as a broad and deep research.

**Future plans and ideas.**

- Recognition of typical line profiles and dynamical responses of the atmosphere to flare heating from solar and ground-based data in large-scale flare events.

- Correlation of appearance of certain line profile shapes and dynamical behavior with properties of hard X-rays (from RHESSI and Sun-19) and soft X-rays (from GOES).

- Computation of IRIS line profiles (Mg II, C II, Si IV) for the considered StellarBox run using RHESSI radiative transfer code. Testing the clustering algorithm on the synthesized emission reduced to IRIS and SDO/AIA line profiles.

- Development of diagnostics tool for recognition of shocks, strong flows and heating events from SDO/AIA data and IRIS data based on cluster analysis discoveries.

**References.**