Title: Blind spectral unmixing and ion lines clustering of low resolution spectra based on non-negative matrix decomposition.

Authors: T. Odstrcil[1,3],. M. Odstrcil[1,2], O. Grover[1], V. Svoboda[1], GOLEM Team

[1]Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Prague, Czech Republic
[2]University of Southampton, Southampton, United Kingdom
[3] Max Planck Institut für Plasmaphysik, EURATOM Association, Garching, Germany

Email: odstrtom@fjfi.cvut.cz

Keywords: Blind spectral unmixing, sparse nonnegative matrix factorization, spectral data analysis

The application of wide-range spectrometers in plasma spectrometry is limited because their resolution is not sufficient for resolving and clear identification of impurity lines. Moreover, common algorithms based on successive peak identification, fitting and subtraction fails when majority of the spectral lines are overlapped or covered by more intensive surrounding lines.

Therefore, blind spectra unmixing technique based on sparse non-negative matrix factorization (NMF) was used. The spectral lines were decomposed into clusters of lines of single ions, so called non-negative endmembers and the radiated power fraction was estimated for each ion. In contrast to other methods of decomposition, as the principal component analysis (PCA), the sparse NMF takes into account physically relevant a-priory knowledge of the problem i.e. non-negativity and sparsity.

The NMF optimization task is defined as $\min\|A-WH\|_2^2+\beta S(W)$ with constrains $W_{ij}\geq 0, H_{ij}\geq 0, \frac{1}{m}\sum_j H_{ij}=1$, where columns of the matrix $A\in R^{m,n}$ contain *n* measured spectra of the length *m*, columns of $W\in R^{m,r}$ are *r* sparse endmembers and finally $H\in R^{r,n}$ denotes relative normalized intensities. The number of endmembers *r* is determined by the physical model.

Our spectroscopy data were obtained from the small tokamak GOLEM [1] working with hydrogen and helium plasma. The most common impurities are C, O, N and He, however, because the core temperature does not exceed 70 eV, only ionization states of up to 3^+ were observed. The spectra were acquired by an Ocean Optics HR2000+ spectrometer with a range of 190-1100 nm (UV, VIS, NIR) and a resolution of 1 – 2 nm.

High correlation of endmembers, their sparsity about 95% in combination with large range of the Signal-to-Noise ratio from ~ 5 for majority of lines up to ~ 100 ,makes decomposition challenging for most of the common NMF algorithms. Hence an alternative NMF method based on Powell algorithm was developed and compared to standard methods.

The sensitivity and reliability of the obtained relative intensities was also improved by adaptive compensating for subpixel wavelength drift and variation of the spectrometer instrumental function. Subsequently, the observed line intensities were used to estimate a robust physical model describing the observed spectral line intensities by variation of the electron temperature and relative impurity densities.

[1] Svoboda, V. et al. Multi-mode Remote Participation on the GOLEM Tokamak. Fusion Engineering and Design, 86(6-8):1310–1314, 2011.