Some Remarks on the Data Analysis Problems in Solar Two-dimensional Spectroscopy

G. Cauzzi\textsuperscript{1}, L.A. Smaldone\textsuperscript{1}, D. Bonaccini\textsuperscript{2,3} and R. Falciani\textsuperscript{4}

\textsuperscript{1} Dipartimento di Scienze Fisiche dell'Università di Napoli, 80125 Napoli, Italy
\textsuperscript{2} National Solar Observatory/Sacramento Peak, Sunspot, NM, USA
\textsuperscript{3} Osservatorio Astrofisico di Arcetri, 50125 Firenze, Italy
\textsuperscript{4} Istituto di Astronomia dell'Università di Firenze, 50125 Firenze, Italy

Abstract. Using the solar bidimensional spectroscopy, one of the main physical limitations in the data analysis is the huge data's quantity. The only way to perform the analysis, for all the points in the FOV, is to extract some parameters from the line profiles. Based on our experience in the field, we suggest what are the most useful parameters, i.e. most physically significant and easy to compute. In order to find correlations among the different parameters and clustering of points on the FOV, useful methods are based on the Multivariate Analysis. We show some applications of this statistical technique, rarely used in solar physics.

1. Introduction

It is certainly recognized the importance of bidimensional spectroscopic observations for a fully satisfactory knowledge of the physical conditions existing in the solar fine structures, of their mutual interactions and effects on the global properties of the star "The Sun". We can refer to the proceedings of recent scientific meetings (among others, to Muller (ed.), 1985; Schmidt (ed.), 1985; Deinzer,Knoelker and Voigt (eds.), 1985; Athay and Spicer (eds.), 1987; to the set of papers presented by the SOUP group) for a wider survey of the scientific motivations asking for bidimensional spectroscopy observations. We can recall here, very briefly, that classical broad-band or "monochromatic" images, even if obtained at very high spatial resolution, can offer only a "section" (in some cases very important for morphological studies) of the true spatial distribution of the observed features, but practically don't allow the determination of any thermodynamic parameter. On the contrary, the classical spectroscopic measurements can supply a complete set of various physical parameters and their height dependence, but referring only to the "points" intersected by the spectrograph slit position on the solar image.

It is now very clear that, especially for very fine, small scale structures (i.e., with lateral dimensions of the order of the geometrical depth of the atmosphere), the lateral energy transfer may deeply influence the physical equilibrium situation of each fine structure. Many plasma instabilities can also occur as a consequence of mutual "lateral" interactions. It is then evident that the bidimensional determination of various physical parameters is absolutely needed for an efficient diagnostics of small scale structures.

Two different techniques have been basically used to perform bidimensional spectroscopic observations, namely the multichannel subtractive double-pass spectrography (MSDP) technique (Mein, 1977) and the narrow passband filter technique (NPF) (Caccin et al., 1983). Both methods present limitations, but, in our opinion, the NPF method offers performances of higher quality as far as the following items are concerned:

- wider allowed FOV;
Fig.1: Monochromatic bidimensional images of an active solar region acquired in the continuum and in the center of FeII-5991.

- easier and more flexible changes of the wavelength peak and, in some case, passband halfwidth, without any hardware modifications of the instrument set-up;
- the acquired images can be immediately used for morphological studies and comparison purposes, without software elaborations, which are, in turn, rather laborious and time consuming in the case of the MSDP data.

The main limitation of the NPF method was represented by the limited spectral resolving power (of the order of 2.5x10⁴ at Hα), which drastically reduced the effective use of the only available instrument for this type of scientific observations (namely, the Universal Birefringent Filter (UBF) of the NSO/Sacramento Peak) to analyze the widest Fraunhofer lines (Hα, Na-D, Mg-b, Hβ and so on).

Recent very sophisticated filter systems have been developed in solar physics: 1) the SOUP filter (Title et al., 1986); 2) the tunable Fabry-Perot interferometer in series to the NSO/SP-UBF (the 20 mÅ Filter) (Bonaccini, 1988). These instruments offer a spectral resolving power one order of magnitude better than the UBF, viz. at the level of good photoelectric slit spectroscopy. Practically all the lines present in the visible range become suitable to be used and, consequently, the diagnostics capabilities of the bidimensional spectroscopic observations emormously increased. It is, then, the right time to address scientific efforts on a deep analysis of the performances and on the limitations offered by this diagnostic method for small scale, fine scale structures.

2. Line profile reconstruction and parameters to be extracted

By means of a filter like the 20 mÅ Filter we can realize a sampling of any line in the visible range, acquiring for each line a set of "monochromatic" bidimensional images like those shown in Fig. 1. The Fig. 2 shows the 20 mÅ Filter passband, the Liège Solar Atlas and the solar spectrum convoluted with the 20 mÅ Filter passband in a spectral region near the FeII- 5991 line.

Having at disposal a wavelength sampling of the line profile for the points of the FOV, it is then trivial to recover the line profile or (if the sampling theorem was not fulfilled) an estimate for all the points. If we are interested to the spectral behaviour of a limited number of selected regions on the FOV (AR’s, flares, network
bright points, etc.), they can be described by providing the line profiles only for few representative points; Fig. 3 shows examples of the FeII-5991 line profile for some points of the active region shown in Fig. 1.

When we are dealing with large spatial patterns (usually the pixel's number for the observed FOV ranges from $10^4$ to $10^5$) we cannot characterize the spectral behaviour of their points but by using some compression technique on the spectral channels. A big compression would be achieved, in this case, if we could provide only few parameters, for each pixel, representative of the global behaviour of the observed spectral profile $I(\lambda)$; of course these parameters ought also to be the most directly related to the physical conditions existing in the observed area on the sun.

Looking at the line profile as a distribution, it is known that this one can be sufficiently described by means of its different order moments. The last 80 years of stellar atmosphere physics have proved the high diagnostic capabilities of the centered moments of zeroth and second order of the line profile that are, respectively, its equivalent width and full half-width; for the astrophysical importance of the third order, the global line asymmetry, see Caccin and Marmolino (1980) and Marmolino and Severino (1981) and references therein.

The $I(\lambda)$ centered moment of the $n$-th order is defined as:

$$M_n(x, y) = \int_0^{+\infty} (1 - I(\lambda, x, y)) (\lambda - \Lambda(x, y))^n d\lambda$$

where

$$\Lambda(x, y) = \frac{\int_0^{+\infty} (1 - I(\lambda, x, y)) \lambda d\lambda}{\int_0^{+\infty} (1 - I(\lambda, x, y)) d\lambda}$$

is the center of gravity. Since they are very easy and quick to compute and support a lot of physical information, we think that the most suitable parameters for the
global description of the line profile are the centered moments with \( n \leq 3 \) and the gravity center.

In the first approximation the gravity center and the third moment are related, respectively, to the mean velocity along the line of sight and its gradient; the zeroth and second moments mainly to the stratification of the thermodynamical parameters.

Fig. 4 shows the maps ("images") of the zeroth, second and third order moments and of the gravity center, for the line FeII-5991 Å and the same FOV as of Fig. 1. With respect to the same parameters of the mean profile, the gravity center red (blue) shifts, the red (blue) asymmetry and higher (lower) zeroth and second moments are coded as white (black).

We can also extract different parameters related to specific (in the sense of "non global") description of the line profile without loss of the spatial relations among the points of the FOV. The most usual parameter is the line profile bisector at fixed residual intensity. Images of the bisectors, obtained at different residual intensity, represent, as a first approximation, a sort of tridimensional map of the velocity field of the whole FOV along the line formation region.

Another parameter of this type is the line profile width at fixed residual intensity. It has been never used till now, but we think that, as for the bisector, maps of the width at different residual intensity can provide a tridimensional map of the thermodynamical parameters behaviour.

Fig. 5 shows the images of FeII-5991 bisector at residual intensity of 80\%, 85\% and 90\% respectively (FOV as in Fig. 1). Referring to the bisector of mean profile, red shifts are coded as white; large grey areas are regions where the residual intensity selected to compute the bisector maps is out of the line profile.
Fig. 4: Maps of different order moments for the line FeII-5991 Å and the FOV shown in Fig. 1.

Fig. 5: Maps of bisectors of FeII-5991 Å at different residual intensities (FOV as in Fig. 1).
3. The Multivariate Analysis use

Although the description of the physical behaviour of all the pixels present in the observed FOV by means of suitable parameters (like profile moments, etc.) allows large compression of the information, usually too many data still remain to be handled and, very often, the description of the FOV scenario (especially at chromospheric levels) in terms of few "clear and significant" parameters is very difficult for the intrinsic complexity of the solar phenomena and their interactions. Some obvious questions arise at this moment:

- **What is the minimum number of parameters to be extracted from the analyzed image needed to reliably describe the physical behaviour of all the pixels in our FOV?**

- **Could some information be deduced from other parameters (to be then considered as "real" independent parameters and, hence, possible physical discriminants of the considered structures)?**

- **Could we quickly and easily identify points with similar behaviours with respect to the extracted parameters?**

To obtain these goals we explored, to analyze our images by means of some statistical methods, the so-called "Multivariate Analysis" technique. In practice, the points of FOV are the subjects of the analysis, and the line profile parameters represent the variables (or descriptors) set on which the statistical analysis has to be made. The Multivariate Analysis techniques have been rarely used in astronomical field (cf. Murtagh and Heck, 1987), then we believe it should be useful to spend some words about the general method before showing its capabilities in the two-dimensional solar spectroscopy field.

We think that useful methods in this context are the Principal Components Analysis (P.C.A.) and the Cluster Analysis (C.A.) methods. By means of the P.C.A. method we can, when possible, reduce the number of the necessary descriptors of the physical situation and, at the same time, draw out the existence of intrinsic underlying global structures present in the FOV; with the C.A. method we can then group into separated classes all the points that can be considered sufficiently homogeneous with respect to the descriptors identified with the P.C.A. method. Finally the spatial distribution of these classes in the original image gives us the possibility of performing a sort of physical pattern recognition in the solar atmosphere represented in our FOV.

We can organize the data in a $n \times s$ matrix, $X$, where $n$ is the number of pixels of our images, and $s$ the number of original descriptors selected for a particular problem (e.g., the bisectors at $s$ different residual intensities values). We can think, therefore, to the $n$-dimensional space $S$ where each pixel (matrix line) characterize a coordinate axis: the points in this space represent the $s$ parameters (the $s$ columns of the matrix $X$ are their components); in the same way, we can think to the $s$-dimensional space $R$ where each parameter characterize a coordinate axis: now the points in this space represent the $n$ pixels (physical positions in the FOV). The whole scenario to be studied can be described by the "cloud" of these $n$ points in the $R$ space.

If two points are close each other in the $S$ space it means, intuitively, that the two corresponding parameters are strongly correlated on the points of our FOV and then provide the same information (and then both of them aren't necessary to the
physical description); similarly, close points in the $\mathbf{R}$ space are pixels with similar behaviour with respect to the parameters, i.e. with similar physical behaviour.

The P.C.A. method, applied in the $\mathbf{R}$ space, allows us to determine a new set of orthogonal (i.e. not correlated) variables, called Factorial Axes, defined as linear combinations of the old ones. The first new axis, called First Factor, is the axis that best-fits the cloud of points maximizing the sum of squares of their distances after the parameters (axes) standardization (i.e. the subtraction of its mean value and the division by its standard deviation). This fact implies the solution of the classic eigenvalue problem:

$$\mathbf{X}'\mathbf{Xu} = \mu \mathbf{u}$$

where $\mathbf{X}'$ is the transposed matrix of $\mathbf{X}$ and $\mathbf{u}$ is the eigenvector associated with the eigenvalue $\mu$ of $\mathbf{X}'\mathbf{X}$. It is possible to prove (Murtagh and Heck, 1987) that $\mu$ is the amount of variance (i.e. information contents) explained by that axis, that is

$$\text{var} (\mathbf{u}) = \mu / s$$

The following axes are chosen in order of decreasing variance explained: they define a new geometrical space, the Factorial Space, where we’ll make all our further analyses. The angles between the Factorial Axes and the old ones can be used as a key to understand the physical meaning of these Factors.

If the original descriptors are strongly correlated, few "new" Factorial Axes can explain a great fraction of total variance, i.e. we can describe the global behaviour by means of a smaller number of descriptors with respect to the original ones. This possibility, when feasible, really represents a very attractive and promising tool of describing the physical nature of the various features present in our FOV and their mutual interactions.

Fig. 6 shows the Hα line bisectors maps at residual intensity, respectively, of 25%, 30%, 40%, 50%, 60% and 70% for a quiet solar region (dimension 125°sec × 165°sec). The data were taken on August 27, 1985 by means of the NSO/SP UBF, near the solar disk center. The grey level code is the same as in Fig. 5.

In this case the P.C.A. analysis reveals that the 92% of the whole information is hold by the first two Factorial axes, very close to the 25% and 70% bisectors. We can qualitatively explain this fact by considering the Hα contribution functions, computed at different wavelengths along the line profile. Some Hα contribution functions, computed by Porrino S. (1988) in the VAL-C model atmosphere, are shown in Fig. 7. We can immediately deduce that a consistent fraction of the radiation near the Hα core (up to 0.3 Å from the line center, roughly where the 25% line bisectors are measured) is formed at chromospheric levels, whereas the radiation near the Hα wings (around 0.9 Å, where the 70% line bisectors are measured) is formed at photospheric levels, well below the minimum temperature region. Consequently, the two (25% and 70%) line bisectors regimes represent a marked transition in the velocity field distribution from the photospheric up to the low chromospheric (or, at least, just above the minimum temperature region) layers in the solar atmosphere.

Trying to identify clusters of points with similar physical behaviours, we can apply, in the Factorial Space, a not-hierarchical C.A. technique proposed by Hartigan, called "dynamical clouds": the idea is to select an initial partition (for the FOV pixels) and then to improve it by assigning with a suitable criterion the $n$ elements to the clusters. Therefore the problem is to find such a partition of our sample in $m$ classes representing different physical features: each class has to be the most homogeneous as possible, while the various classes have to be each other the most different as possible.
Fig. 6: Hα line bisector maps for different residual intensities, for a quiet solar region at disk center (27-Aug-1985, U.T.=15:33).
The partition is selected by means of the evaluation of a parameter, \( \tau \), related to the "intra-class inertia" (i.e. the second momentum): the greater are the homogeneity of each group (class) and the difference among the various groups, the smaller is the value of \( \tau \). Therefore the relations between \( \tau \) and the number \( m \) of classes give us useful informations about the best partition to be finally adopted.

Fig. 7 shows an example of the application of C.A. technique: the first 4 images (clockwise) are the maps of H\( \alpha \) bisectors (same FOV as in Fig. 6) at 70% of residual intensity, computed at different observing times; the last one is the resulting clustering, obtained with a C.A. technique by using as descriptors the 70% bisectors at the four observing times. The "dynamical clouds" method indicates that the best partition of the pixels distribution is represented by 6 classes. On the clusters image, we have coded as white and black the two classes containing the majority of the points. These classes simply represent "coherence patterns" of oscillation cells, as you can also see by simply inspecting the bisectors images.

The present application of the C.A. technique don't show particularly new and exciting results. It is only important, in this work, to demonstrate the large and deep diagnostics capabilities of this method of image analysis for the determination and the study of spatial structures, present in the observed FOV, connected by common physical properties. Consequently, the study of their mutual interactions and evolution might be also straightforwardly developed and in an objective way, without any personal influence of the "observer".

4. Conclusions

We want to stress out here that one of the best way to perform analysis of bidimensional spectroscopy data, for wide spatial regions, maintaining the spatial relations among the points, is by extracting maps of line profiles parameters. They can, quickly, support us a lot of physical information about the values of dynamical
Fig. 8: Bisector maps at 70% of residual intensity for Hα line, computed for different observing times. The last image (clockwise) is the clustering resulting by the application of C.A technique. The points with the same grey level belong to the same class, i.e., have the same behaviour with respect to the bisectors.
and thermodynamical parameters. The use of statistical techniques, like the Multivariate Analysis can, moreover, speed up the analyses, pointing out the really needed parameters to describe physically the observed regions, and by identifying spatial regions with similar behaviours. All the analysis process seems to be very promising for understanding the properties of the Sun and the spatial relations between the structures: we believe therefore that this data analysis techniques have to be deeply tested and more widely applied in solar physics.

Acknowledgments

All the observational material has been obtained at the NSO/SP-VTT. We are deeply indebted to the director of the Sacramento Peak Observatory and to the members of the VTT staff for the reliable help given during the observations.

References