AUTOMATICALLY FINDING SOLAR ACTIVE REGIONS USING SOHO/MDI PHOTOGRAMS AND MAGNETOGRAMS

M. Turmon[1], **J.M. Pap**[2], **S. Mukhtar**[1]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, 91109, U.S.A.
[2]Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, U.S.A.

ABSTRACT

This paper presents a new application of statistically-based methods for automatically identifying the various surface structures on the Sun which may contribute to observed changes in solar total and spectral irradiance. These structures are divided for our purposes into three types: quiet sun, faculae, and active regions. The procedures described here find labelings from synchronized full-disk high-resolution magnetic-field and light-intensity maps from the Michelson Doppler Imager experiment on SoHO. Statistical models characterizing the properties of the region types are found from expert identification of a sample set of regions. Once these parameters are fixed, the inference procedure becomes to maximize the probability of an image-labeling given the observed data. This allows objective and automated classification of a large set of images.

Key words: solar irradiance; magnetograms; photograms; image processing; labeling; segmentation

1. INTRODUCTION

Observations of solar irradiance (both bolometric and at various wavelengths) have demonstrated that the solar energy flux changes over a wide range of periodicities: from minutes to the 11-year solar cycle (Willson and Hudson, 1991; Fröhlich, 1994). It has been shown that the long-term irradiance changes are mainly associated with bright magnetic features, including faculae and the magnetic network (Foukal and Lean, 1988). Former analyses have indicated that the short-term irradiance variations are directly related to the evolution of active regions via the combined effect of dark sunspots and bright faculae (Hudson *et al.*, 1982; Chapman, 1987; Fröhlich and Pap, 1989). However, the physical mechanisms underlying irradiance changes are not fully understood, and the SoHO observations indicate that a physical origin of solar irradiance variability, based solely on solar magnetic activity, cannot explain all the aspects of the observed irradiance changes (Fröhlich *et al.*, 1997; Wehrli *et al.*, 1997). It is not clear whether these discrepancies are related to measurement uncertainties in irradiance and its magnetic surrogates, or to additional global effects, such as temperature changes (Kuhn *et al.*, 1988), radius changes (Delache *et al.*, 1986; Ulrich and Bertello, 1995), large scale mixing flows, or large scale convective cells (Ribes *et al.*, 1985; Fox and Sofia, 1994).

Nonetheless, as the nearest star, the Sun is the only star where we can observe and identify the variety of structures and processes which do lead to irradiance variations. High-resolution observations conducted from the ground and various space missions have shown that the surface of the Sun and its outer atmosphere are highly dynamic on almost all spatial scales. To understand the physical causes of the changes observed in solar total and spectral irradiances, it is necessary to study the spatial and temporal characteristics and evolution of the solar magnetic fields and related thermal structures in the various layers of the solar atmosphere.

The main goal of this work is therefore to identify and characterize the various surface structures on the Sun which contribute to the changes in solar total and spectral irradiances. While our earlier work used only magnetic-field observations (Pap *et al.*, 1997), the present study uses both high-resolution (1024^2 pixels) magnetograms, taken approximately every 96 minutes, and quasi-white-light photograms, which are taken roughly once every 360 minutes. For our purposes, it is important to distinguish further between the 1-minute and 5-minute magnetogram integration times used by MDI, as these produce images having different noise characteristics. All these images are taken by MDI with a CCD camera near the Ni I 676.8 nm absorption line originating in the mid-photosphere (Scherrer *et al.*, 1995).

Adding photograms necessitates some pre-processing steps which do not arise with magnetograms alone. In particular, the photograms must be corrected for limb-darkening effects and diminishing instrument sensitivity. Also, both image modalities must be placed on a common time frame. Finally, the two-dimensional observable must be related to the object classes, and this is the heart of the procedures described here.

To locate the causes of solar irradiance variations, about 10^4 solar images must be analyzed. These images show a wide variety of solar structures: active regions (sunspots and faculae), remnants of active

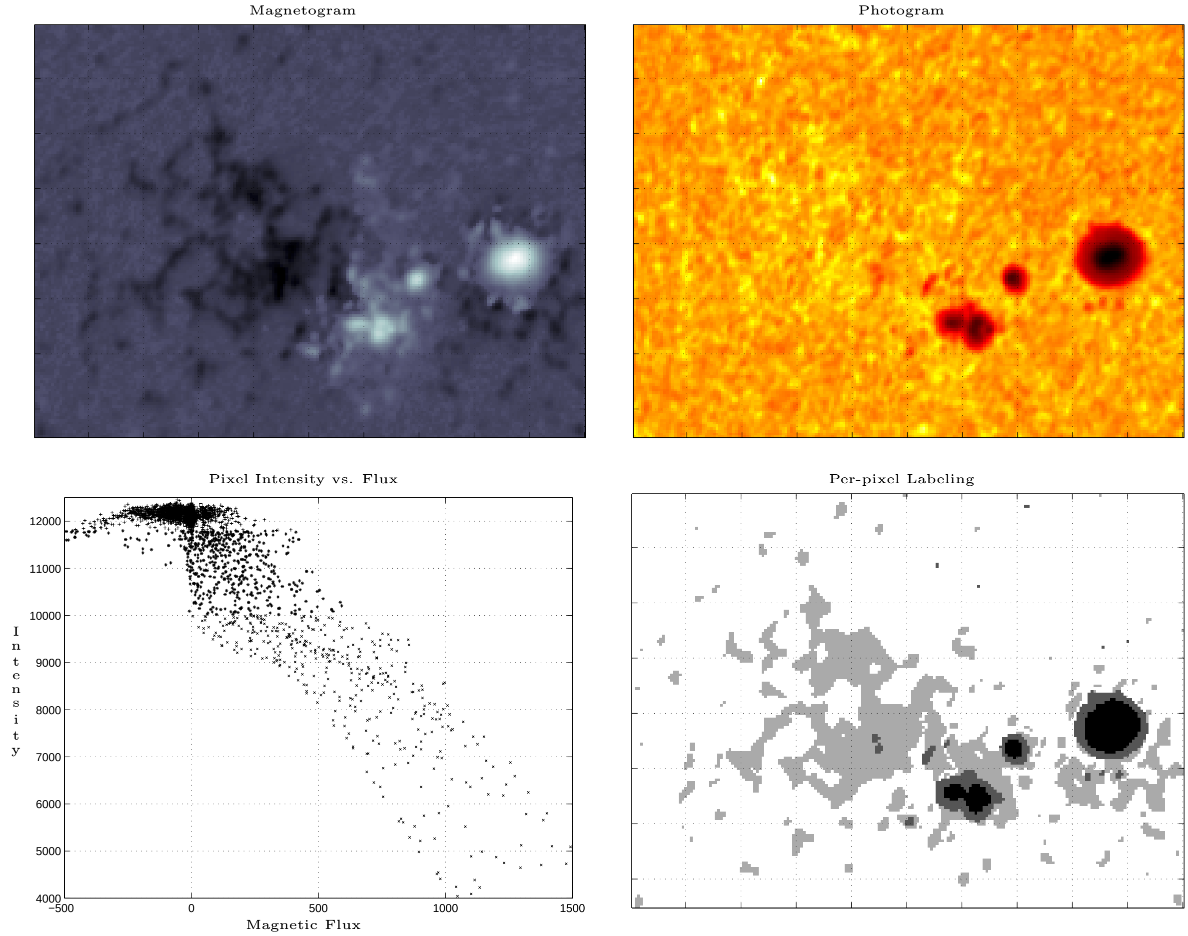

Figure 1. Top, near-simultaneous magnetogram and photogram, with some contrast enhancement; bottom, corresponding scatter plot and rough per-pixel labeling.

regions, and the active network and the relatively quiet network which are distributed as cell-like structures over the solar disk. For the purposes of this analysis, we have concentrated on four specific structures: sunspot umbra, sunspot penumbra, faculae, and background.

2. IMAGE ANALYSIS TECHNIQUE

Figure 1 illustrates the potential of the MDI data for extracting these structures from solar images. The upper panels show, respectively, details from a one-minute magnetogram and a flat-fielded photogram which happened to be taken 6.0 minutes apart by MDI at 4:15 UTC on 1 August 1996. (Six minutes implies a relative displacement of less than half a pixel at disc center.) The corresponding scatter-plot of observed values is shown in the third panel. As this plot shows, magnetic fields of about ±200 Gauss (we use MDI nominal units throughout) may produce either sunspots or faculae, and so a two-parameter observation aids in separating these features reliably. In fact the umbra and penumbra can be separated at the knee of the two-dimensional scatterplot. Adding the second dimension of intensity permits a lower error rate in distinguishing these classes. The crude labeling in the last panel corresponds to a simple thresholding of this bivariate data, ignoring all spatial relationships. (For this reason, we do not advocate per-pixel thresholding as a labeling technique; however, doing so shows the potential of the data.) The coherence of the resolved structures is apparent.

This conveniently synchronized image-pair demonstrates the potential of such coupled intensity and flux measurements to allow inference of structural properties. Doing this requires a uniform, automated technique whose parameters have been determined objectively to the greatest extent possible. In what follows, we describe the setup which we have developed for this purpose.

2.1. Preprocessing and Synchronization

It remains to formalize and automate the procedure illustrated above — in particular, to generalize to the case of non-synchronous magnetogram and photograms. Figure 2 illustrates the method:

1. Flat-field each photogram
2. Interpolate a photogram to the magnetogram observation time
3. Infer a labeling from the magnetogram and photogram

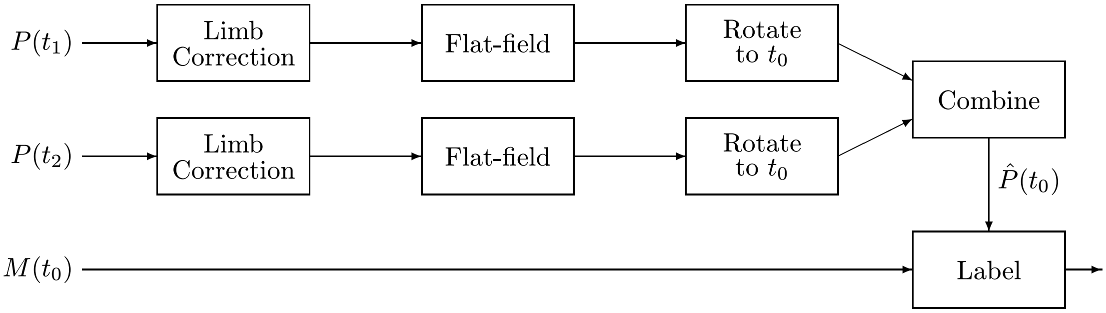

Figure 2. Data flow used to preprocess photograms and magnetograms in preparation for feature identification

Here we cover the first two steps, and subsequently we describe the third step which is the heart of the decomposition scheme.

Having a magnetogram at a given time t_0, the closest photograms at times before and after t_0 are found. Center-to-limb variation is removed via an adaptation of the procedure of Steinegger *et al.* (1996), yielding a correction factor

$$0.3728 + 0.9605\mu - 0.5135\mu^2 + 0.1802\mu^3$$

which agrees with that of Bogart *et al.* (1998) to within 0.004 for $\mu \geq 0.2$. A further image-wide correction factor is necessary to compensate for filter degradation. This factor is linear in time, albeit with breaks in several points at which the instrument was recalibrated. See Bogart *et al.* (1998) for a more detailed description of these effects. Next, a small residual flat-field (accounting for roughly 0.4% spatial variation) is applied. The result is two spatially and temporally uniform photograms.

Ensuring temporal homogeneity of these preprocessing steps is paramount due to their expected use in irradiance study. Both the limb correction and the residual flat-field correction are held constant over the entire image sequence, except for the recalibration breaks mentioned above. Consequently, the only continuously-varying temporal adjustment that is applied is the one for filter degradation, which is well-understood. Temporal stability is crucial to eliminating side-effects from the final region-maps.

At this point, standard formulas for the latitude-dependent angular velocity of active regions (e.g., Zappalà and Zuccarello 1991), applicable across time scales on the order of several hours, are used to synthesize a photogram at t_0 from the bracketing photograms. (A 240-minute rotation, a typical outer limit for these data, corresponds to a movement of about 18 pixels at disk center. An error in sunspot angular velocity of 0.5°/day over 240 minutes corresponds to a displacement of about half a pixel at disk center.) Given t_0, this registration step is done automatically using the appropriate MDI FITS header information. The synthesized photogram and the original magnetogram, now in a common coordinate system, are used together to infer a labeling, or assignment of region-type across the image.

Figure 3 shows sample results from the interpolation process. Shown at top are details from a five-minute magnetogram (taken 3:15 UTC on 6 August 1996), and a photogram synthesized from bracketing photograms at 14:57 UTC on 5 August, and 7:51 UTC on 6 August. This implies a minimum rotation of 276 minutes, which is a rather large gap for the MDI cadence (about 14 magnetograms/day). The scatter plots for two subregions are in the bottom panels. The bipolar spot shows clearly as having lower intensity. Most encouraging, the brightness enhancement of the faculae, amounting to only about 2% of the nominal intensity level of 12150 units, is clearly visible, increasing with magnetic flux strength. Such results are consistent with a well-calibrated limb-compensation and flat-field mechanism.

2.2. Image Decomposition

The final step, inferring the labeling, is more complex because doing so involves scientific judgment. In the system we propose, this judgment is isolated in a falsifiable statistical model whose parameters are chosen according to a scientist-provided labeling. This practice isolates the problem-specific elements of the decision procedure into a concise and testable model which naturally accounts for uncertainty and noise in the observables. These characteristics distinguish the approach described below from strictly “algorithm-based” procedures in which the only way to describe the labeling method is via the computer code that is ultimately used, and for which reckoning with the uncertainty inherent in noisy data and uncertain labelings takes place away from its natural probabilistic setting.

Because of the abundance of prior information, we adopt a Bayesian view which allows us to incorporate expert knowledge of this data in the consistent framework referred to above. Following well-established statistical practice (Geman and Geman, 1984; Ripley, 1988; Turmon and Pap, 1997), we describe the labeling step in this Bayesian framework as inference of the underlying pixel classes (symbolic variables represented by small integers) based on the observed (vector-valued) pixel characteristics.

Specifically, let pixel sites $s = (s_1\, s_2)$ exist on an

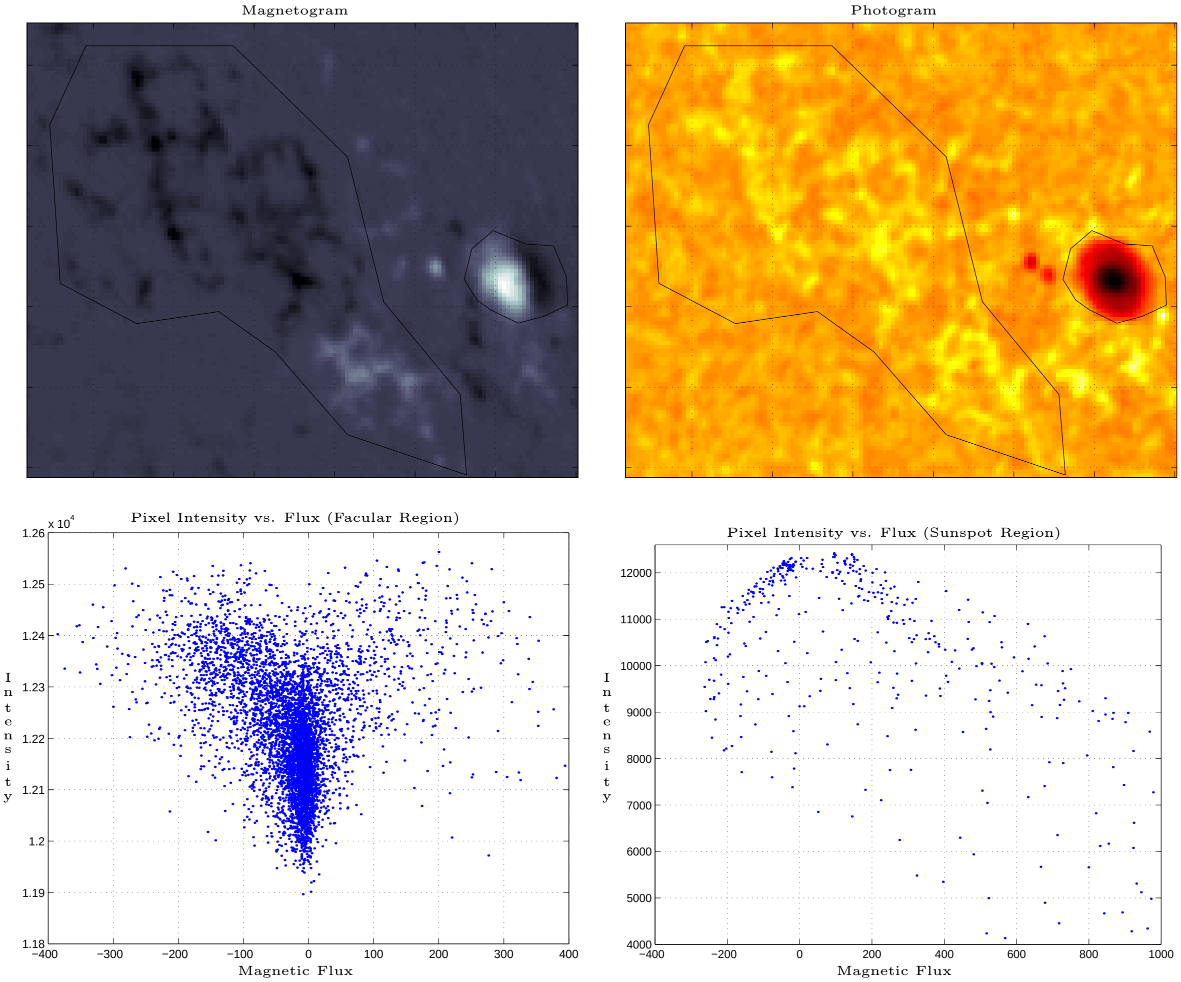

Figure 3. Synchronized magnetogram and photogram, with scatter plots of sites within the two indicated regions.

image domain N. Define a matrix of discrete class labels $\mathbf{x} = \{x_s\}_{s \in N}$. The observed characteristics $\vec{y}_s$ are similarly combined into a three-dimensional array $\mathbf{y}$. The posterior probability of labels given data is

$$P(\mathbf{x} \,|\, \mathbf{y}) = P(\mathbf{y} \,|\, \mathbf{x}) P(\mathbf{x}) / P(\mathbf{y}) \propto P(\mathbf{y} \,|\, \mathbf{x}) P(\mathbf{x}) \ . \quad (1)$$

For the estimation of labels, the maximum a posteriori (MAP) decision rule maximizes this probability, or, equivalently, its logarithm:

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \log P(\mathbf{y} \,|\, \mathbf{x}) + \log P(\mathbf{x}) \ . \quad (2)$$

The first term is the familiar likelihood function, forcing fidelity to the data. The second is the prior probability of a given labeling, which penalizes physically unrealistic labelings. For illustration, consider the simplest framework, in which data are generated from labels according to a Gaussian distribution with per-class mean vectors $\vec{\mu}_k$ resulting in the objective function

$$-\frac{\lambda}{2} \sum_{s \in N} \|\vec{y}_s - \vec{\mu}_{x_s}\|^2 + \log P(\mathbf{x}) \quad (3)$$

where $\|\cdot\|$ is Euclidean norm, so $\|\vec{v}\|^2 = v_1^2 + v_2^2$. This can be interpreted as a Lagrangian form which dictates that we maximize fidelity to the data subject to a constraint on the physical reasonableness of the labeling which is captured by the prior $P(\mathbf{x})$.

Prior models $P(\mathbf{x})$ may be specified in many ways. We have used the Markov field smoothness priors

$$P(\mathbf{x}) = Z^{-1} \exp\big[-\beta \sum_{s' \sim s} 1(x_{s'} \neq x_s)\big] \quad (4)$$

where $1(\cdot)$ is one if the indicated proposition is true, and $\beta \geq 0$. The constant Z is chosen to normalize the probability mass function, and the sum extends over 'neighboring' sites in N. On our rectangular grid, sites are neighbors if they adjoin vertically, horizontally, or diagonally. As β drops, rougher labelings are penalized less, and the uniform distribution is obtained at $\beta = 0$.

To make this formalism more concrete, examination of quiet-sun areas in the MDI images reveals an excellent fit of $\vec{y}_s$ to a bivariate normal distribution. (These quiet-sun areas are seen in the core of the lower-left plot of Figure 3.) The mean value of the magnetic-field component $y_{s,1}$ is zero Gauss, while that of the photograms is about 12 150 units. The spatial RMS of the 5-minute magnetograms is roughly 15 Gauss, while that of the photograms is about 100 units. The distribution of $\vec{y}_s$ under quiet-sun conditions is therefore

$$P(\vec{y}_s \,|\, x_s = 1) = N\Big(\begin{bmatrix} 0 \\ 12\,150 \end{bmatrix}, \begin{bmatrix} 20^2 & 0 \\ 0 & 100^2 \end{bmatrix}\Big) \quad (5)$$

where N is the bivariate normal distribution having the indicated mean and covariance matrix.

The distributions for the other region types are determined similarly, as indicated in Figure 3. That is, areas in the image which represent (say) sunspots are outlined, and a distribution is fitted to the resulting scatter plot. To statistically describe the characteristics of these more complex features, a more flexible distribution than a simple Gaussian can be chosen from the wider class of finite mixture distributions (Titterington et al., 1985) and fit by maximum-likelihood (McLachlan and Krishnan, 1997). Once the distributions $P(\vec{y}_s \,|\, x_s)$ for $x_s = 2, 3$ (i.e., faculae and sunspot) have been found and verified, $P(\mathbf{y} \,|\, \mathbf{x})$ is fully specified and the labeling is implicitly defined via (2). In practice, numerical methods adapted from statistical physics (Metropolis *et al.*, 1953) are employed to find the maximizer. We emphasize that the prior model (4) with $\beta > 0$ couples the labels at neighboring pixels, and so the Bayesian inference procedure does not correspond to a per-pixel threshold rule.

While we have not done so here, in a full "pattern-theoretic" approach (Grenander and Miller, 1994), this label information may be linked into larger, more abstract structures describing individual active regions and facular groups. Such structural models have been used, for example, to infer galactic shapes from digitized images (Ripley and Sutherland, 1990) and to describe chromospheric plages (Turmon and Mukhtar, 1997). The plage models in particular may be readily adapted to describe photospheric faculae via an automatically-determined bounding polygon.

3. DISCUSSION

Determining structural information about solar phenomena is one way to understand and refine the mechanisms of solar irradiance variability. We have demonstrated that the MDI photograms and magnetograms, used together, have the potential to identify these solar structures accurately. The first step in using these image sources together is a well-calibrated, temporally stationary, preprocessing scheme for putting flattened photograms and magnetograms in the same reference frame. Once this is done, a formalism for automated image segmentation, based on contemporary image segmentation techniques, is applied to the normalized data. The parameters in this model are fitted from scientist-provided image labelings, and their accuracy may be checked by standard statistical methods. Currently, the preprocessing stage is completed and we are determining models for the sunspot and facular region-types.

One item not stressed here is the model flexibility allowed in defining $P(\vec{y}_s \,|\, x_s)$. This distribution, relating the region-types to the observables, may be let depend on the spatial location of a site. This may be an advantage for specifying the characteristics of faculae, which have a spatially-varying contrast. We have already made attempts to establish an empirical relation between the intensity and magnetic flux of different features as a function of their position on the solar disk. Whether or not all such second-order adjustments are incorporated in the statistical model, identification of the various regions will facilitate finding the so-called empirical calibration curves between the MDI magnetic field values and the continuum intensity images as a function of the evolution of the active region involved. It is anticipated that having labelings readily at hand will allow other such "per-region" quantities to be computed.

ACKNOWLEDGEMENTS

SoHO is a mission of international cooperation between ESA and NASA. The authors gratefully acknowledge the effort of the MDI team. In particular, Rick Bogart helped us to understand the physical origin of several features of the MDI images. Jeneen Sommers explained the fine points of the MDI FITS data files and their format. The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology and University of California, Los Angeles under a contract with NASA. The research was supported by a grant of the SoHO Guest Investigator Program.

REFERENCES

Bogart, R. S., Bush, R. I., Wolfson, C. J., 1998, "MDI Calibration Notes", this volume

Chapman, G.A., 1987, Ann. Rev. Astron. Astrophys., 25, 633

Delache, Ph., Lacrare, F., Sadsaoud, H., 1986, Advances in Helio- and Asteroseismology (eds. J. Christensen-Dalsgaard, S. Frandsen), IAU Press, 223

Foukal, P., Lean, J., 1988, ApJ., 328, 347

Fox, P., Sofia, S., 1994, The Sun as a Variable Star (eds. J.M. Pap, C. Fröhlich, H.S. Hudson, and S.K. Solanki), Cambridge Univ., 280

Fröhlich, C., Pap, J., 1989, Astron. Astrophys., 220, 272.

Fröhlich, C., Pap, J., Hudson, H.S, 1994, Solar Physics, 152, 111

Fröhlich et al. 1997, Solar Physics 170, 1

Geman, S., Geman, D., 1984, IEEE Trans. Patt, Anal. Mach. Int., 271

Grenander, U., Miller, M.I., 1994, J. Roy. Stat. Soc. Ser. B, 549

Hudson, H.S., Silva, S., Woodard, M., Willson, R.C., 1982, Solar Physics, 76, 211

Kuhn, J., Libbrecht, K.G., Dicke, R.H., 1988, Science, 242, 908

McLachlan, G.J., Krishnan, T., 1997, The EM Algorithm and Extensions, Wiley

Metropolis N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953, J. Chem. Phys., 1087

Pap, J., Turmon, M, Mukhtar, S., Bogart, R., Ulrich, R., Fröhlich, C., and Wehrli, Ch.: 1997, Correlated Phenomena at the Sun, in the Heliosphere and in Geospace, (eds. A. Wilson and B. Fleck), ESA-SP-415, 477

Ribes, E., Mein, P., Mangeney, A., 1985, Nature, 318, 170

Ripley, B.D., 1988, Statistical Inference for Spatial Processes, Cambridge Univ.

Ripley, B.D., Sutherland, A.I., 1990, Phil. Trans. R. Soc. Lond. A, 477

Scherrer et al. 1995, Solar Physics, 162, 129

M. Steinegger, P. N. Brandt, H. F. Haupt, 1996, Astron. Astrophys., 310, 635

Titterington, D.M., Smith, A.F.M., Makov, U.E., 1985, Statistical Analysis of Finite Mixture Distributions, Wiley

Turmon, M., Pap., J.: 1997, Statistical Challanges in Modern Astronomy II (eds. G. J. Babu and E. D. Feigelson), Springer-Verlag, 408.

Turmon, M., Mukhtar, S.: 1997, Proc. IEEE Intl. Conf. Image Proc., III-320

Ulrich, R., Bertello, L., 1995, Nature, 377, 214

Wehrli, Ch, Appourchoux, Th., Crommelynck, D., Finsterle, W., Pap, J.: 1997, Sounding Solar and Stellar Interiors, (eds. J. Provost, F.X. Schmeider, G. Berthomieu), Kluwer Academic Publishers, in press

Willson, R.C., Hudson, H.S.: 1991, Nature, 351, 42.

Zappalá, R.A., Zuccarello, F., 1991, Astron. Astrophys., 242, 480