Solar Phys (2010) 262: 277–298
DOI 10.1007/s11207-009-9490-y

Statistical Feature Recognition for Multidimensional Solar Imagery

Michael Turmon · Harrison P. Jones · Olena V. Malanushenko · Judit M. Pap

Received: 16 May 2009 / Accepted: 2 December 2009 / Published online: 11 March 2010

Abstract A maximum *a posteriori* (MAP) technique is developed to identify solar features in cotemporal and cospatial images of line-of-sight magnetic flux, continuum intensity, and equivalent width observed with the NASA/National Solar Observatory (NSO) Spectromagnetograph (SPM). The technique facilitates human understanding of patterns in large data sets and enables systematic studies of feature characteristics for comparison with models and observations of long-term solar activity and variability. The method uses Bayes' rule to compute the posterior probability of any feature segmentation of a trio of observed images from per-pixel, class-conditional probabilities derived from independently-segmented training images. Simulated annealing is used to find the most likely segmentation. New algorithms for computing class-conditional probabilities from three-dimensional Gaussian mixture models and interpolated histogram densities are described and compared. A new extension to the spatial smoothing in the Bayesian prior model is introduced, which can incorporate a spatial dependence such as center-to-limb variation. How the spatial scale of training segmentations affects the results is discussed, and a new method for statistical separation of quiet Sun and quiet network is presented.

Solar Image Processing and Analysis
Guest Editors: J. Ireland and C.A. Young.

M. Turmon (✉)
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA
e-mail: turmon@jpl.nasa.gov

H.P. Jones
National Solar Observatory, Tucson, AZ 85719, USA
e-mail: hjones@nso.edu

O.V. Malanushenko
Apache Point Observatory, New Mexico State University, Sunspot, NM 88349, USA
e-mail: elenam@apo.nmsu.edu

J.M. Pap
University of Maryland, College Park, MD 20742, USA
e-mail: judit.m.pap@nasa.gov

Keywords Active regions, structure · Sunspots, statistics · Magnetic fields, photosphere · Solar irradiance

1. Introduction

Solar magnetograms provide essential information for many aspects of solar physics, and a rich literature documents an extended and continuing effort to organize and understand these complex data. Well-defined, stable methods for marking features on full-disk solar images enable systematic documentation of size, location, and physical characteristics of such entities as sunspots, active regions, enhanced network, and quiet network for comparison with both observed measures and theoretical models of solar activity and long-term variability. This paper presents a statistical technique for solar feature recognition that we plan to apply to large data sets to address many of these issues.

A highly successful and important recent example of feature-based solar research is its application to understanding variations in solar irradiance, both bolometric and at various wavelengths. Even small changes in the total solar irradiance give information about internal-energy transport processes, while analyses of spectral-irradiance observations from the ultraviolet to the infrared help to understand the changes taking place in the photosphere and chromosphere. The terrestrial implications of solar irradiance variability are equally important. Since the Sun's radiative output establishes the Earth's thermal environment, knowing the source and nature of its variability is essential to understanding and predicting the interactions in the Sun – Earth system. Assessing the impact of human activities depends critically on such knowledge to quantify natural background variability. Sunspot – facular models (Foukal and Lean, 1988; Krivova *et al.*, 2003; Wenzler *et al.*, 2006) explain most of the variation in total solar irradiance for periods up to several years. On the other hand, the models are less successful for decadal and longer time periods (de Toma *et al.*, 2001, 2004; Pap *et al.*, 2002; Jones *et al.*, 2003), especially when spanning more than one sunspot cycle, which suggests that understanding the causes of irradiance variation and/or the observational data is incomplete. Accurate and consistent classification of sunspots, faculae, and other features, across time scales of years to decades, is necessary to separate these issues.

These questions can be addressed in the original gridded coordinates. Global variables such as irradiance can be modeled through direct summation over high-resolution simulations or data (Preminger, Walton, and Chapman, 2002), and spatial structures can be represented by continuous changes of basis such as Fourier and wavelet transforms, and empirical orthogonal functions (Bloomfield *et al.*, 2004; Cadavid, Lawrence, and Ruzmaikin, 2008). We agree that such grid-level, continuous-domain representations are necessary and informative, but we contend that, as in the example of solar irradiance, quantifying the behavior of various structures is a critical and complementary component of understanding both observations and numerical simulations.

In addition to quantifying visual impressions, feature segmentation reveals characteristics and suggests approaches to research that are masked by the overwhelming volume of the original imagery. For example, reduction from full-disk images to active regions has saved a factor of over 100 in data volume in *Michelson Doppler Imager* (MDI) imagery from NASA's *Solar and Heliospheric Observatory* (SOHO). High-resolution and high-cadence missions like *Hinode* and *Solar Dynamics Observatory* (SDO) will enlarge the existing gap between human apprehension and solar imagery. For example, SDO plans to generate over two petabytes of data over its planned five-year mission, and a significant element of its science strategy is to maintain a catalog of identified features and events called the

Heliophysics Knowledgebase (Schrijver *et al.*, 2008). Additionally, some computationally-intensive inversions, such as are required for resolving the azimuthal ambiguity in vector magnetic fields (Crouch, Barnes, and Leka, 2009), are not presently feasible at these data rates, and attention may need to be restricted to the most salient regions. Feature segmentation can also be used as the initial processing step in a system for grouping active regions spatially, and tracking them automatically through image sequences. The resulting active regions and tracks can be used in image catalogs and for data selection and subsetting.

In this paper, we concentrate on those features that can be identified in full-disk magnetograms and other images derived from spectral lines formed primarily in the photosphere. These include sunspots (umbra and penumbra), faculae, plage, active regions, enhanced network, network, and quiet Sun. Unfortunately, although solar physicists are familiar with a wide variety of characteristic structures, a uniform lexicon and quantitative definitions are lacking; for individual studies the terminology and operational definitions are tailored to the subject at hand. Faculae, plage, and active regions, for example, probably refer to different observational manifestations of the same physical structures, but, depending on subjective judgement, can yield rather different labellings. Jones *et al.* (2008) explore some of these issues by comparing several different methods for identifying features important for studies of solar irradiance variation. Even for apparently well-defined structures such as sunspots and active regions, substantial differences are found among techniques.

Many feature-recognition methods for photospheric and chromospheric structures have been described in the literature (Worden, White, and Woods, 1998; Gyori, 1998; Harvey and White, 1999; Ermolli, Berrilli, and Florio, 2000; Preminger, Walton, and Chapman, 2001; Gyori *et al.*, 2004; McAteer *et al.*, 2005; Zharkova *et al.*, 2005; Ermolli *et al.*, 2007; Curto, Blanca, and Martínez, 2008). Typically, observationally-established thresholds, possibly followed by region-growing or shrinking, have been used to distinguish feature classes. The thresholds may be applied to normalized images, or to derived images computed by spatial filtering including differentiation for edge enhancement and convolutions with Gaussian kernels. But fundamentally, rules based on thresholds make decisions directly in the image space.

The method used here, extending earlier work (Turmon, Pap, and Mukhtar, 2002, hereafter TPM2002), makes these decisions using relative probabilities. Transformation into probabilities allows principled combination of multiple information sources, including *i*) the relative importance of different image sources; and *ii*) the relative importance of per-pixel observations *versus* local structure. An example of the former case is distinguishing enhanced network from active regions: the classifier must weigh magnetic-field evidence *versus* intensity evidence. Of course, the distributions of per-pixel observations for various structures overlap considerably. The computation of probabilities allows the evidence provided by local structure to influence boundary cases where the per-pixel observations are inconclusive. These spatially-linked decisions are made in a mathematical optimization framework that is implied by the problem structure.

The contents of the paper are as follows: In Section 2, we develop the formalism linking observed data to image labels, explain the parts of the formalism, and discuss the Bayesian inversion algorithm used to derive a labeling from images. In particular, Section 2.1 examines two important generalizations of the prior model introduced by TPM2002, one of which can be used to accommodate spherical geometry. Also, in Section 2.2 we develop techniques for computing conditional probabilities both by extending to three dimensions the Gaussian mixture models presented by TPM2002, and by introducing a new method, interpolated histograms, for probability-density modeling in this context. Section 3 describes the data we analyze: magnetic-field, intensity, and equivalent-width images taken by the

NASA/NSO *Spectromagnetograph* (SPM) from 1992 – 2003. Section 4 makes specific the generalized prior model introduced in Section 2.1, using SPM data to illustrate a new prior having a spatially-varying neighborhood structure which accounts for projection of the Sun onto the image plane. Finally, in Section 5, we apply our technique to a training set based on the segmentations described by Harvey and White (1999), hereafter HW1999. We compare the results of the Gaussian mixture and interpolated histogram methods for conditional probability modeling that were introduced in Section 2.2. Furthermore, this section shows results from the separation of quiet Sun from quiet network. The main new contributions of this paper are the introduction and demonstration of the generalized prior model, the introduction and demonstration of the interpolated histogram technique, and the extension of the enhanced TPM2002 methodology to generate labellings of the SPM data.

2. Method

Before going on to detail the classification procedures we use, we pause to introduce some notation. The generality of the mathematical setting can be off-putting, so we illustrate the notation for the SPM images (magnetic field, intensity, and equivalent width) analyzed in this paper. The classification algorithm takes these three images, which are strictly cospatial and cotemporal, and produces a labeling, or image-sized classification mask, with one label for each input pixel.

All images are indexed by spatial location $s = [s_1\, s_2]$; the set of all such *sites* in the image is $\mathcal{S}$. SPM images are 1788 pixels square, so $\mathcal{S}$ consists of 1788^2 sites. A single pixel's observable *feature vector* is $y = [y_1 \cdots y_d] \in R^d$; for us, $d = 3$. When discussing the SPM imagery in an algorithmic context, the magnetic-field observable is taken to be y_1, intensity contrast y_2, and equivalent-width contrast y_3. For observational and physical discussion, we refer more descriptively to B for magnetic field, I for intensity, and W for equivalent width. The observed feature vectors $y(s)$ are grouped into an image $\mathbf{y}$, indexed by $s \in \mathcal{S}$, which can be thought of as a matrix of feature vectors. The decomposition of an image into K classes is captured by defining a *labeling* $\mathbf{x}$, also indexed by s, where each $x(s)$ is an integer in $\{1, \ldots, K\}$: a matrix of labels. The labeling algorithm deduces the labeling $\mathbf{x}$ from the trio of images $\mathbf{y}$.

This notation assumes that all observations are made as rectangular images, leading to sites that are embedded on the image plane, and labellings that correspond to rectangular images. In fact, the two-dimensional geometry is not critical to the formalism described below, and the same inference procedure and spatial models have been used in spherical and three-dimensional settings.

Following well-established statistical practice (Geman and Geman, 1984; Ripley, 1988; Turmon and Pap, 1997), we treat labeling in a Bayesian framework as inference of the underlying pixel classes (symbolic variables represented by small integers) based on the observed (vector-valued) pixel characteristics. The idea guiding the model is that there is a family of K physical processes, and at any site $s \in \mathcal{S}$, exactly one process is dominant, say $x(s)$. Typically, the dominance of a given process has a spatial coherence, so that a labeling $\mathbf{x}$ has clumps of identical labels, which represent the object-level structures of Section 1. The full labeling $\mathbf{x}$ is linked probabilistically to the observed Eulerian features $\mathbf{y}$.

The posterior probability of labels given the observed data is central in the Bayesian framework. Here, by Bayes' rule, this probability is

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

The constant of proportionality is unimportant because we are only interested in the behavior of the posterior as the labeling $\mathbf{x}$ is varied. Our goal is to find the optimal labeling. The well-known maximum *a posteriori* (MAP) rule tells us to maximize Equation (1), or, equivalently, its logarithm:

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

This maximization can be interpreted as trading off fidelity of the labeling to the data (the first term) *versus* spatial coherence of the labeling itself (second term).

To use the MAP rule, we must specify the prior, $P(\mathbf{x})$, and the likelihood, or class-conditional probabilities, $P(\mathbf{y} \,|\, \mathbf{x})$. The key structural assumption of the model is that $y(s)$ depends on $\mathbf{x}$ only through $x(s)$, the dominant process at s. Formally, we say that $P(y(s) \,|\, \mathbf{x}) = P(y(s) \,|\, x(s))$, which implies

$$P(\mathbf{y} \,|\, \mathbf{x}) = \prod\nolimits_{s \in \mathcal{S}} P\big(y(s) \,|\, x(s)\big). \tag{3}$$

So, to specify the likelihood in Equation (2), we only need to find the K class-conditional distributions

$$P(y \,|\, x = k), \quad \text{for } k = 1, \ldots, K. \tag{4}$$

For example, to classify an image into quiet-Sun, network, active-region, and sunspot types, four distributions would be needed, each reflecting the typical scatter of the three-dimensional $[B\ I\ W]$ feature vector within one type of region. In Section 2.2 we discuss methods for estimation of these distributions, but first we describe our choice of prior.

2.1. Prior Models

Prior models $P(\mathbf{x})$ may be specified in many ways, always with the fundamental goal of coupling activity classes at nearby sites. We advocate using a diffuse or *weakly informative* prior (Gelman *et al.*, 2008) because we believe that the data term, not the prior, should have the strongest influence on the labeling. In particular, the prior we use only encodes spatial coherence, and does not attempt to model the spatial scale of the network, or to model sunspot shapes (as did Turmon, Mukhtar, and Pap, 1997). We use a metric Markov random field (metric MRF, described by Boykov, Veksler, and Zabih, 2001) which generalizes the Potts random field used in TPM2002. Priors that fall in the metric MRF class, which were originally made accessible by the seminal paper of Geman and Geman (1984), have by now been extremely widely used in image segmentation for scientific, medical (Pham, Xu, and Prince, 2000; Zhang, Brady, and Smith, 2001) and computer vision (Li, 2009) applications.

The metric MRF prior on a discrete labeling $\mathbf{x}$ is

$$P(\mathbf{x}) = Z^{-1} \exp\left[-\sum_{s' \sim s} \beta(s, s')\, \delta\big(x(s), x(s')\big) \right], \tag{5}$$

where the constant Z (the *partition function*) depends on β and δ, and Z serves to normalize the probability mass function. Note that the maximizer in Equation (2) does not depend on Z. The sum extends over pairs of neighboring sites $s, s' \in \mathcal{S}$. On our square grid, sites are neighbors if they adjoin vertically, horizontally, or diagonally, so each site s has eight neighbors, denoted $N(s)$.

Intuitively, Equation (5) applies a penalty $\beta(s, s')\, \delta(x(s), x(s')) \geq 0$ to neighboring pixels. The penalty is zero when the labels agree and increases corresponding to the significance

of the disagreement. In this sense, the exponent is a penalty on *rough* labellings, or labellings with many significant disagreements between adjacent pixels. The roughness penalty consists of two factors, a spatially-varying weighting function β, which measures the closeness of sites, and a label-dependent penalty δ, which gives the cost of a disagreement between adjacent labels.

We use the first part of the penalty, the *site coupling* $\beta(s, s') \geq 0$, to account for differing distances between neighboring pixels. This becomes important near the limb, as will be demonstrated in Section 4.2. By choosing β properly, labels near the limb can have considerable coupling in the tangential direction (constant μ), but loose coupling in the radial direction. We have successfully used

$$\beta(s, s') = \beta_0 \exp\bigl(-\gamma \operatorname{dist}(s, s')\bigr) \tag{6}$$

where $\operatorname{dist}(s, s')$ is the great-circle distance, on a solar sphere of unit radius, between the 3D locations corresponding to pixels s and s', and γ is a dimensionless scale parameter. This distance can be readily computed using image metadata: disk center and radius, and ellipticity if known. If the solar radius is R pixels, the smallest value of $\operatorname{dist}(s, s')$ of approximately $1/R$ occurs at disk center, and the corresponding $\beta(s, s') = \beta_0 \mathrm{e}^{-\gamma/R}$. Thus, to avoid saturating the exponential, one must choose $\gamma \ll R$; we have successfully used $\gamma = 150$ for the typical SPM radius of about 840 pixels, which means $\beta(s, s')$ is near β_0 across a broad central area of the disk. Considerations related to preservation of edge structures outlined in TPM2002 imply $\beta_0 < 1.10$ is advisable; we typically use a fraction of this value. For examples of how the site coupling works, see Section 4.2.

If β is lowered towards zero, the exponent of Equation (5) also approaches zero, and the prior distribution becomes more nearly uniform. Finally, with $\beta(s, s') \equiv 0$, $P(\mathbf{x})$ is constant, and labels are spatially uncoupled. Alternatively, according to the interpretation of the exponent as a penalty on rough labellings, we could observe that the roughness penalty is curtailed as β drops. We emphasize that any prior model such as Equation (5) with $\beta > 0$ couples the labels at neighboring pixels, and so the Bayesian inference procedure does not correspond to a per-pixel decision rule. Even if $\beta \equiv 0$, the decision region as a function of the image data may be complex depending on the per-class distributions.

The *class interaction metric* $\delta(x(s), x(s')) \geq 0$, at its simplest, equals one if $x(s) \neq x(s')$, and zero otherwise. In general, δ is a symmetric, $K \times K$ table that can provide stronger penalties for more significant label disagreements: a network/quiet-Sun disagreement may be less significant than sunspot/quiet Sun. For example, compare Equations (17) and (18) in Section 4.1, where the impact of δ on labellings is explored. The *metric* MRF models that we use have the physically-realistic requirement that δ must be a metric on the K labels. That is, $\delta(x, x') \geq 0$, with equality if and only if $x = x'$, plus the triangle inequality, $\delta(x, x') + \delta(x', x'') \geq \delta(x, x'')$. The former property means that all label disagreements between neighboring sites are penalized by the metric, while agreements are not penalized. (But note that there can still be no spatial penalty if $\beta = 0$.) The triangle property is equivalent to the statement that the penalty for changing x to x', and then from x' to x'', is no smaller than changing x directly to x''.

Unless $\beta \equiv 0$, the functional form of Equation (5) couples label decisions at a site s to its neighbors $N(s)$, and so on across the entire image. This makes the maximization in Equation (2) a difficult problem, because the space of admissible labellings $\mathbf{x}$ is vast. Several methods of solving this problem are available. We use the simulated annealing and sampling algorithm detailed in TPM2002, which for solar imaging applications, compares well in terms of solution quality and computation time with max-cut and belief propagation

methods. We used 1000 image sweeps, or epochs, and a geometric annealing schedule with a temperature at epoch r of $T_r = 1.5 \times 0.995^r$ to converge on a labeling.

2.2. Likelihood Models

We describe below two methods for estimating the K class-conditional probability densities $P(y \mid x = k)$. To estimate a particular density, both methods use a training data set $Y = \{y^{(1)}, \ldots, y^{(N)}\}$ consisting of N SPM pixels selected from the class of interest. As a rough guide, we have used training sets consisting of thousands to millions of pixels of each class. Samples from some classes, such as quiet Sun, are particularly easy to obtain.

In TPM2002, the density is modeled via a superposition of Gaussian basis functions, an approach we abbreviate GMM (Gaussian mixture model). An alternative estimation method introduced below interpolates the density directly from histograms abbreviated IHM (interpolated histogram model). For background on histogram-based density estimation, see Scott (1992). GMM has the advantages of positivity and smoothness, but parameter selection is relatively complex, and it may require prohibitively many parameters to accurately fit highly structured densities. IHM is more directly based on observations and involves only mild assumptions reflected in the functional form used for interpolation. However, it is more sensitive to noise and requires special attention to binning. Even for moderate d, naively-binned histograms will require impossibly large storage and will be sparsely filled even with very large training sets. Some of these tradeoffs (smoothness and robustness to noise for GMM *versus* model flexibility for IHM) are simply reflections of the well-known *bias/variance dilemma*: sparsely-parameterized estimators like GMM tend to have lower variance at the expense of higher bias (Geman, Bienenstock, and Doursat, 1992). Practical solutions to the potential problems for both methods have been evolved, and the two algorithms serve as cross checks of the overall procedure. In practice, they so far have given very similar segmentation results with comparable computing effort.

2.2.1. Gaussian Mixture Models

The simplest per-class distribution is multidimensional normal, modeling each class with a Gaussian distribution centered about μ with covariance matrix $\mathbf{\Sigma}$:

$$N(y;\, \mu, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{d/2}(\det \Sigma)^{1/2}} \exp\left[-\frac{1}{2}(y-\mu)^{\mathrm{T}}\mathbf{\Sigma}^{-1}(y-\mu)\right]. \tag{7}$$

As discussed in TPM2002, this simple functional form is inadequate for our purposes. For example, scatter plots of feature vectors $y(s)$ sampled from within sunspots show considerable structure, with two lobes corresponding to positive and negative polarity regions.

To get around this problem, we generalize Equation (7) to the class of finite normal mixtures (McLachlan and Peel, 2000). These densities, parameterized by weights, means, and covariances, are additive combinations

$$P\left(y; \{\lambda_j, \mu_j, \Sigma_j\}_{j=1}^{J}\right) = \sum\nolimits_{j=1}^{J} \lambda_j N(y;\, \mu_j, \Sigma_j), \tag{8}$$

where the function N is defined in Equation (7) above. Here, the positive weights λ_j add to one, the constituent mean vectors μ_j are arbitrary, and the $d \times d$ covariance matrices Σ_j are symmetric positive-definite. By letting J increase, the underlying distribution is fit more exactly; indeed, any continuous density may be modeled arbitrarily closely with Gaussian

basis functions. Mixtures have been used to generalize the relatively restrictive distribution of Equation (7) to complex distributions typical of practical applications by other authors (Hastie and Tibshirani, 1996).

Separately for each region type, the free parameters $\{\lambda_j, \mu_j, \Sigma_j\}_{j=1}^{J}$ are chosen by maximum-likelihood using the corresponding training set Y. That is, we maximize the training set log-likelihood

$$\log P(Y) = \sum\nolimits_{n=1}^{N} \log P\left(y^{(n)}; \{\lambda_j, \mu_j, \Sigma_j\}_{j=1}^{J}\right). \tag{9}$$

When $J > 1$, there is no longer a closed-form solution for these parameters so we use a well-known numerical technique called the Expectation-Maximization (EM) algorithm (McLachlan and Peel, 2000, Section 3.2). For $N = 20\,000$, $d = 3$, and $K = 10$, optimization takes about 60 seconds using a fast computer.

Furthermore, we have modified the EM maximization to account for the symmetry constraint on the magnetic field: all distributions must be invariant with respect to a reversal in the polarity of the magnetic field (Turmon, 2004). This implies that the Gaussian basis functions can either occur in symmetric pairs of opposite magnetic polarity, or in singletons which in themselves are symmetric.

2.2.2. Interpolated Histogram Models

This section develops a histogram-based density estimate $\rho(y)$ for the true per-class density $p(y)$. Begin by partitioning the observation space into d-dimensional voxels $\mathcal{V}_j$ of volume V_j. The number of pixels of the N-length training set expected to fall in bin j is

$$H(j) = N \int_{\mathcal{V}_j} p(y)\,\mathrm{d}y \approx N V_j\, p\left(y_c(j)\right), \tag{10}$$

where $y_c(j)$ is the center of $\mathcal{V}_j$, and the approximation is best when p is smooth. In practice, we have discrete bin counts $\eta(j)$, which approximate the expected counts $H(j)$ above, leading us to the estimate

$$\rho\left(y_c(j)\right) = \eta(j) \,/\, N V_j. \tag{11}$$

These points serve as the knots for the interpolation scheme described below.

Note that Equation (10) characterizes the true, but unknown, density and histogram, whereas the estimate in Equation (11) uses observed histogram values. The error in the final estimate $\rho(y)$ is caused by systematic errors (bias due to mismatch between the interpolated function and the true density) and random errors (variance due to limited training set size). The estimate needs to be most accurate near the peaks of the distribution, where the observed histogram counts are large, diminishing the impact of random errors. That is, in classifying images, the density estimate may be evaluated in its high-variance tails, but it is most likely that another class, the one actually responsible for that pixel, will have vastly higher probability, obviating the need for a precise probability estimate.

Regarding systematic errors, it is well known (Scott, 1992, Section 4.1) that the *frequency polygon* produced by linearly interpolating between bin centers has significantly lower bias than the histogram itself, which is constant over $\mathcal{V}_j$. As described below, such systematic errors can be further mitigated by taking into account the known power-law structure of $p(y)$ and its interaction with the underlying linear interpolation between bin centers.

Direct interpolation of histograms has many practical difficulties. The domains of the independent variables tend to be large while the histograms vary rapidly in small sub-intervals.

Linear interpolation is often inaccurate, while higher-order interpolations, including splines, lead to oscillations and negative densities. To avoid these difficulties, we follow Jones *et al.* (2000, 2003) and Jones and Ceja (2001) and tailor our interpolation algorithm to magnetograms whose histograms vary asymptotically as power laws. These authors have empirically developed the following quasi log – log transformation of variables which allows simple basis functions to adequately represent observed SPM histograms. Specifically, for $1 \le i \le d$, we compute coordinate-wise sample means and variances $[\mu_i$ and $\sigma_i^2]$, define the scaled independent variables

$$\tilde{y}_i = \operatorname{sign}(y_i - \mu_i) \log\bigl(1 + |y_i - \mu_i|/\sigma_i\bigr), \tag{12}$$

and bin the histograms in equal increments of $\tilde{y}$ rather than in original units (Simonoff, 1998, Section 2.3). Thus the volume elements $\mathcal{V}_j$ in the original y units are variably-sized. Convergence of histograms computed from such data-driven binning schemes to the underlying density is shown by Lugosi and Nobel (1996). In particular, it is not necessary that each voxel contain a minimum number of pixels.

Similarly, we define a scaled dependent variable in units corresponding to log-counts

$$\tilde{\rho} = \log(1 + N\bar{V}\rho) \tag{13}$$

where $\bar{V}$ is the average bin volume. In our construction, $\tilde{\rho}$ will be continuous, including across bin boundaries, and ρ will inherit that continuity. The scaled variables $(\tilde{y}, \tilde{\rho})$ tend to linear behavior far away from the means while avoiding singular behavior for empty histogram bins. For each observed quantity, we choose our unit-free scaled bin intervals to fall at intervals of 0.25 over the range from minimum to maximum in the training set of images. The quasi-logarithmic character of the scaling for the observed quantities produces rather coarse bins away from the means. Consequently, the bin counts remain moderately large over a wide range of inputs, and empty bins are rare and distant from the means.

Combining the last equation with Equation (11) shows the knots in $\tilde{\rho}$ are at

$$\tilde{\rho}\bigl(\tilde{y}_c(j)\bigr) = \log\bigl(1 + N\bar{V}\rho\bigl(y_c(j)\bigr)\bigr) = \log\bigl(1 + \eta(j)\bar{V}/V_j\bigr). \tag{14}$$

We interpolate $\tilde{\rho}$ at arbitrary $\tilde{y}$ from a tensor product of one-dimensional interpolating basis functions with tie points as above. In each dimension, the interpolation is cubic in the interval containing the mean, adequately representing Gaussian behavior typical of noise-dominated distributions near the means, and is piecewise linear elsewhere, adequately representing power-law behavior. (Shorn of the bin-center enhancement, this multivariate linear interpolation scheme is known as the *linear blend frequency polygon* (Scott, 1992, Section 4.2). For more on the bin-center enhancement in the context of frequency polygons, see Scott (1992, page 98).) Finally, the desired class-conditional probability density is computed by inversion of Equation (13):

$$\rho = \bigl(\exp(\tilde{\rho}) - 1\bigr) / N\bar{V}. \tag{15}$$

Once the free parameters of the IHM model (bin centers and scales $[\mu_i, \sigma_i^2, 1 \le i \le d]$, and bin counts $[\eta(j)$ for each $\mathcal{V}_j]$) have been chosen, the interpolated density may be computed for each activity class $1 \le k \le K$. This interpolated density is precisely analogous to the GMM of Equation (8), and either is directly usable as a model for $P(y(s) \mid x(s) = k)$ in Equation (3). In Sections 5.2 and 5.3 we compare these two density estimates using SPM data.

3. Observations

From April 1992 to September 2003, strictly cotemporal and cospatial full-disk images were obtained with the NASA/NSO *Spectromagnetograph* (SPM) and consisted of line-of-sight (LOS) magnetic flux, LOS velocity, continuum intensity, equivalent width, and central line depth (Jones *et al.*, 1992). The SPM was retired on 20 September 2003 to make room for a vastly improved instrument, the *Synoptic Optical Long-term Investigations of the Sun* (SOLIS) *Vector Spectromagnetograph* (VSM) which is now operational on Kitt Peak and which both continues and expands the SPM observations to include vector polarimetry. Since the photospheric magnetic field is both a fundamental indicator of the Sun's global state and a dominant driver of solar and heliospheric activity, this record will provide important data for understanding and predicting solar behavior and its heliospheric and terrestrial consequences.

For the feature classifications described here, we have used three observables: magnetic flux, intensity contrast in the continuum, and equivalent-width contrast. While velocity might be useful in outlining network near the limb, it is also mixed, through the image scanning process, with photospheric oscillations. Line depth would be at least as useful as equivalent width, but a problem with the on-line analysis software for SPM full-disk observations severely limits the number of days with reliable line depths over the entire field of view. We have also experimented with factor analysis (see, for example, Gorsuch, 1983) to derive linear combinations of observed quantities but do not discuss these preliminary results in this paper.

To provide the observational inputs for the statistical method, mean center-to-limb variations for velocity, continuum intensity, equivalent width, and line depth are extracted as described by Jones *et al.* (2000). For example, contrast for continuum intensity is computed by

$$I(s) = \left[I_{\text{obs}}(s) - I_{\text{cl}}\big(\mu(s)\big)\right] / I_{\text{cl}}\big(\mu(s)\big) \tag{16}$$

where $I_{\text{cl}}(\mu)$ is the mean center-to-limb variation as a function of the cosine of the heliocentric angle and I_{obs} is the observed intensity. Small systematic variations introduced by the least-squares fitting procedures to derive the center-to-limb variations are removed from the contrasts as computed in Equation (16) with a median filter applied in equal-area annuli. Contrasts for equivalent width and line depth are similarly computed. Figure 1 shows an example, taken on 06 November 1998, of the SPM contrast images used in this analysis.

4. Tailoring the Prior Model

As remarked in Section 2.1, the prior model in Equation (5) has the ability to incorporate some problem-specific characteristics. In this section we demonstrate how this works. As we shall see in the following section, the likelihood term $[P(y \mid x)]$ is relatively easy to fit from image feature-vectors. But the free parameters of the prior are hard to choose in a purely data-driven way. For example, the clearest path to estimate the smoothness parameter β would be to use maximum-likelihood, but this is complicated by the intractability of the partition function Z as a function of β (see Equation (5)).

Fortunately, our prior has relatively few free parameters, and experimentation with labellings is sufficient to choose them. To do so, we use an SPM image set taken on 13 December 1998, which has good seeing and contains many activity classes. In this section, we arrive at density models (likelihoods) $p(y \mid x)$ in an *ad hoc* way, because our sole purpose is to demonstrate the effect of the prior. Three classes are used, corresponding to quiet-Sun, sunspot, and a network/active network class.

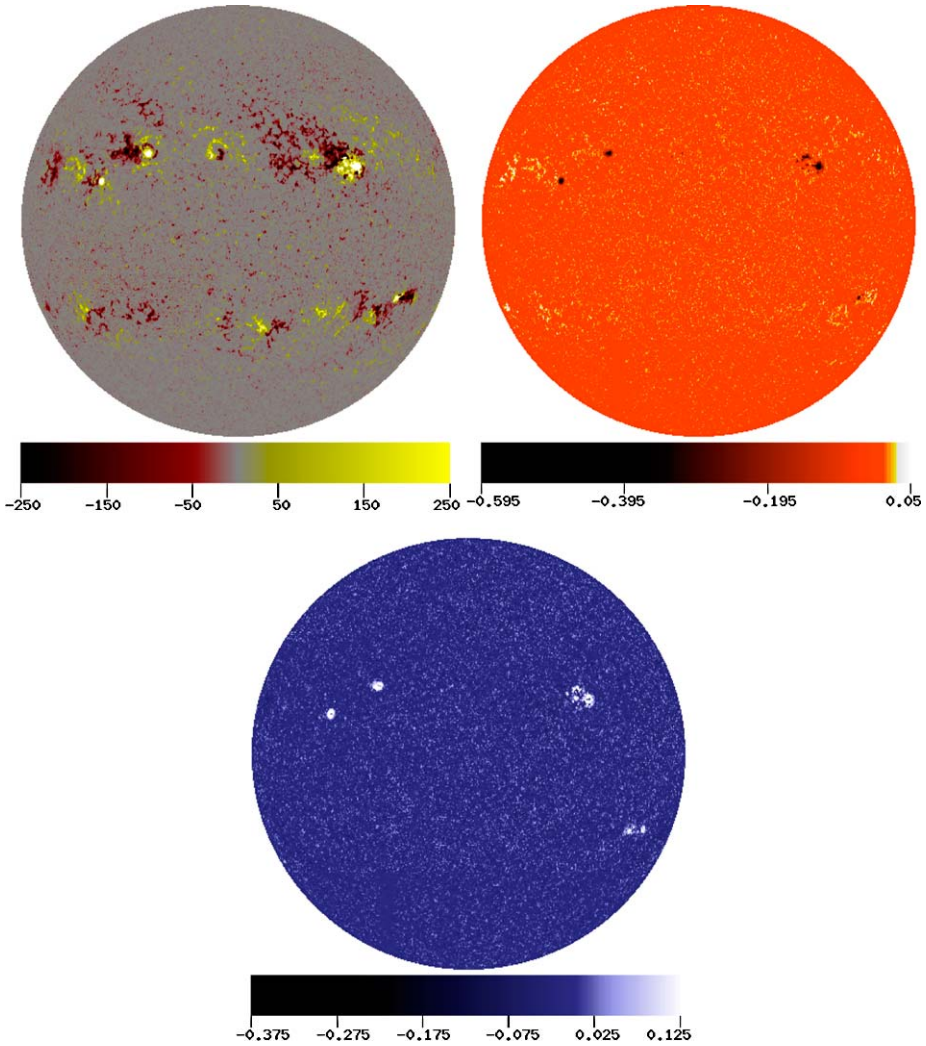

Figure 1 SPM observations on 06 November 1998. Upper left: magnetogram (LOS magnetic flux, Gauss); upper right: intensity contrast; lower: equivalent-width contrast.

4.1. Class Interaction

Using these three models, a per-pixel segmentation can be found as in the top right panel of Figure 2. In our notation, this corresponds to $\beta \equiv 0$: a uniform prior on all labellings, and no spatial interactions. The corresponding magnetogram, with enhanced contrast, is shown on the top left of Figure 2 for reference. On close inspection of the labeling for $\beta \equiv 0$, it is evident that there are many small, isolated pixel-groups of all classes. This is especially salient for the network class, and it is presumably due to several effects, such as CCD noise affecting quiet pixels which happen to be near the class boundary, and small local fluctuations in the quiet Sun that cause the threshold to be crossed. This noise is typical of solar-activity classification algorithms in the photosphere and chromosphere.

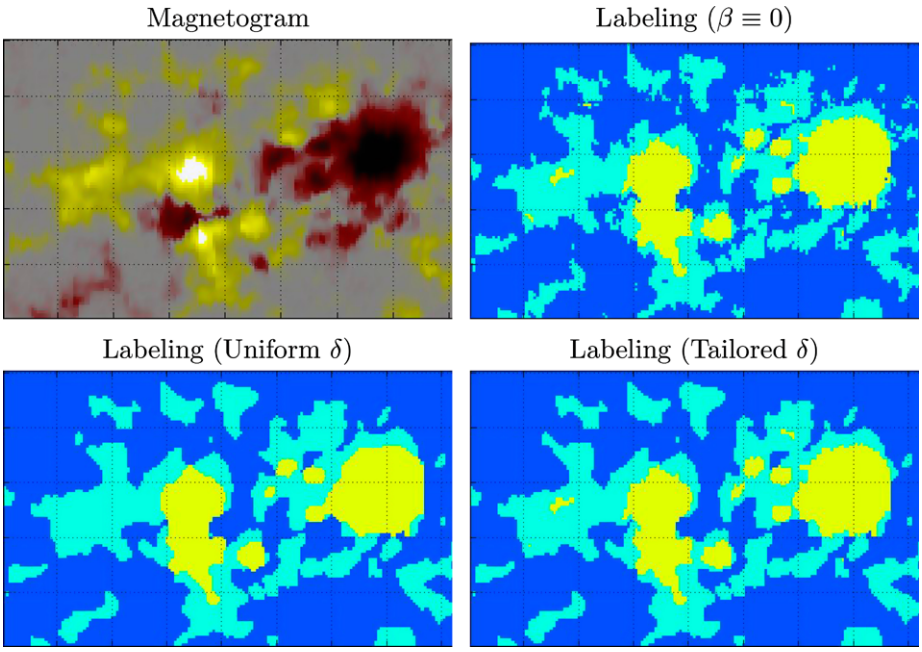

Figure 2 Demonstration of image smoothing using three-class segmentation of NSO data on 13 December 1998, 17:14 UTC. The sunspot is NOAA 8408, seen near disk center. At top left, a pseudocolor magnetogram. At top right, an unsmoothed ($\beta \equiv 0$) segmentation. The lower panels show two $\beta \equiv 0.2$ labellings: left, a uniform class-boundary penalty ($\delta = 1$ for the off-diagonals); right, a tailored penalty which falls more lightly on facula – spot neighbors.

The spatial interactions of the MRF serve to suppress such noise. The lower-left panel of Figure 2 shows an MRF segmentation with the same likelihood, but $\beta \equiv 0.2$, a low level of smoothing, and a uniform penalty

$$\delta(x, x') = D_1 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}. \tag{17}$$

The columns and rows of this matrix are in the order: quiet Sun, network, sunspot. This penalty results in the change of 682 pixels (4.3%) of the 100×160 pixel classification, and eliminates the small clumps of network.

The remaining panel of Figure 2 illustrates how the class interaction metric tailors the smoothing effect. In a given application, we may wish to suppress small clumps of network superimposed on quiet Sun, while leaving the network/sunspot boundary intact. The classification shown at the lower right also used $\beta \equiv 0.2$ but

$$\delta(x, x') = D_2 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0.1 \\ 1 & 0.1 & 0 \end{bmatrix}. \tag{18}$$

Network/sunspot neighbors now receive a much smaller penalty. (Note that the metric condition on δ prevents lowering the penalty to zero. This would, of course, have the effect

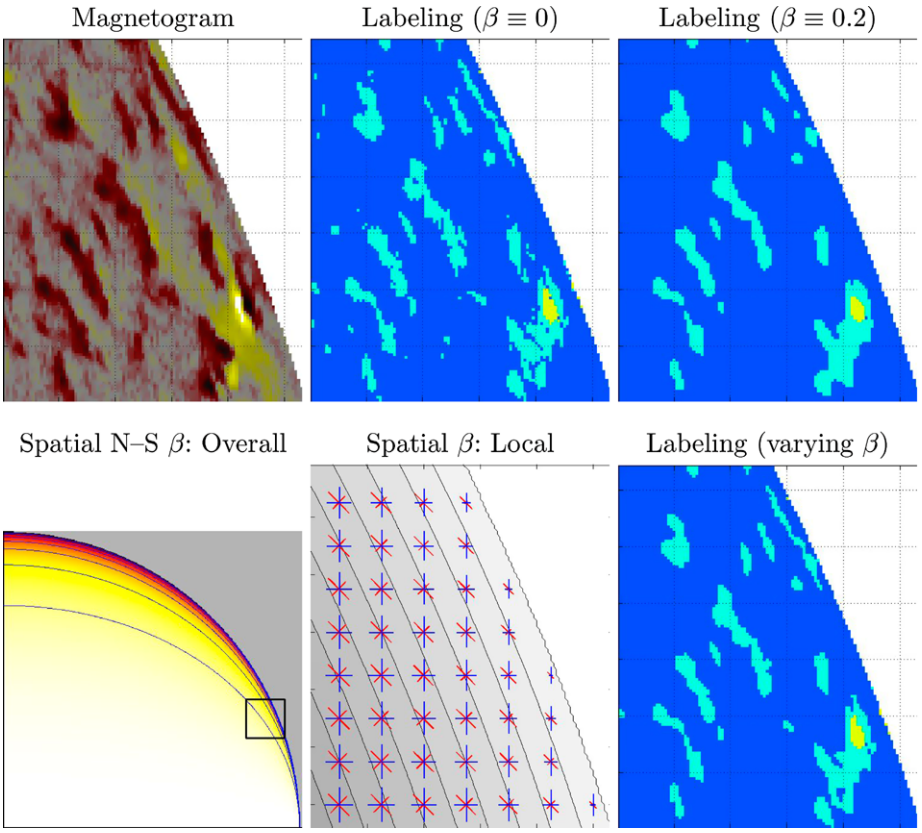

Figure 3 Demonstration of image smoothing using three-class segmentation of NSO data on 13 December 1998, 17:14 UTC. At top left, a pseudocolor magnetogram; the sunspot is NOAA 8403, at the west limb. Top center: a pixel-wise image classification ($\beta \equiv 0$), and top right: classification with $\beta \equiv 0.2$. The lower panels show the Bayesian classification with a spatially-varying β ($\gamma = 150$). Left: map of the North – South component of β in one quadrant of the disk. This is $\beta(s, s')$, plotted over image index s, where s' is the northward neighbor of the corresponding s. Contours are placed at equal intervals of β. The black box in the left panel shows the area of all detail images. Lower center: the local map of β (short blue lines, North – South and East – West coupling; red lines, diagonal coupling), superimposed on a contour map of constant distance to disk center. Lower right: labeling with spatially-varying β using $\gamma = 150$.

of merging the two classes.) Only 536 (3.3%) of the pixels are now changed from the $\beta \equiv 0$ labeling, and only 24 network or sunspot labels were changed, as opposed to 155 in the uniform-penalty labeling. This is readily appreciated by comparing the sunspot/facula boundary in the three labellings of Figure 2. The lower-right labeling combines the fine spatial detail of the sunspot/facula boundary in the upper-right labeling with the clear network feature of the lower-left labeling.

4.2. Spatial Coupling

For full-disk solar images, we have found the spatially-varying weighting $\beta(s, s')$ to be critical in preserving structure at the limb. Figure 3 illustrates this. Again the magnetogram is at top left, showing one sunspot (NOAA 8403) about 11 pixels from the west limb ($R = 846.9$ pixels, $\mu \approx 0.16$), as well as some foreshortened network features. The

unsmoothed segmentation (upper right) shows artifacts similar to Figure 2. However, the isotropic $\beta = 0.2$ segmentation (lower left), while suppressing single-pixel artifacts, also partially eliminates the most slender network features. The labeling with a spatially-varying coupling ($\gamma = 150$ in Equation (6)), lower right, does not suppress foreshortened features. Note especially the full retention of the entire two-pixel wide network feature about four pixels from the limb ($\mu \approx 0.09$). This labeling shows the ability of the distance-weighted smoothing criterion to enhance the visible network structures.

5. Computation of Conditional Probabilities and Image Masks

5.1. Training Data

For the initial development of our class-conditional algorithms, we selected as training data 38 of the five-feature (quiet network, enhanced network, active region, decaying region, and sunspot) HW1999 segmentations of NSO/KP magnetograms from June 1996 through December 1998. Observations from this period were selected for which the seeing identified by the NSO observer was categorized as "good" or better, which were complete and free from obvious defects, and for which the time of observation was within 30 minutes of a valid full-disk MDI magnetogram. The time period covers much of the rapid rise of cycle 23 so that many different instances of solar activity are represented. Among the many possible segmentations available for training the statistical model, these have the distinct advantage of being derived from the SPM data so that no cross-registration of images or features is required.

For each of the training images, lists (feature vectors) of observed quantities for every class were extracted from the reduced observations by selecting only those pixels within the Harvey – White segmentation masks. Three-dimensional histograms and Gaussian mixture-models were then computed for each of the HW1999 feature classes from the accumulated feature vectors for all images in the training set. Class-conditional probabilities were calculated using both mixture models and histogram interpolation, and new feature masks for each training image were computed using our statistical method.

5.2. Gaussian Mixture Models

An example mixture model for active region is shown in Figure 4. This Gaussian mixture was computed for the active-region class as identified by HW1999 on the basis of $N = 73\,000$ feature vectors. A good fit was obtained with $J = 10$ components having a total of 61 free parameters after accounting for symmetry with respect to B. Because we are restricted to showing 1D or 2D plots, we show marginal distributions of one or two image types, while integrating out the other image types. Because the basis functions are Gaussian, integration can be done analytically and results in trivial expressions for the 1D and 2D marginal distributions. Indeed, the marginal mean and covariance of any component is just the corresponding slice of the mean vector and covariance matrix.

There is good agreement of the 1D histograms with the fitted mixture (thick blue lines in the diagonal plots). The effect of the symmetry constraint is clear in the first column and row of plots, which show symmetry with respect to flips in the sign of B, and the structure of symmetric mixture models. Finally, the 2D marginal distributions (upper plots) show some familiar features, such as the two "wings" of progressively enhanced intensity for positive and negative magnetic field in the B *versus* I plot. Also in this plot, the supergranulation

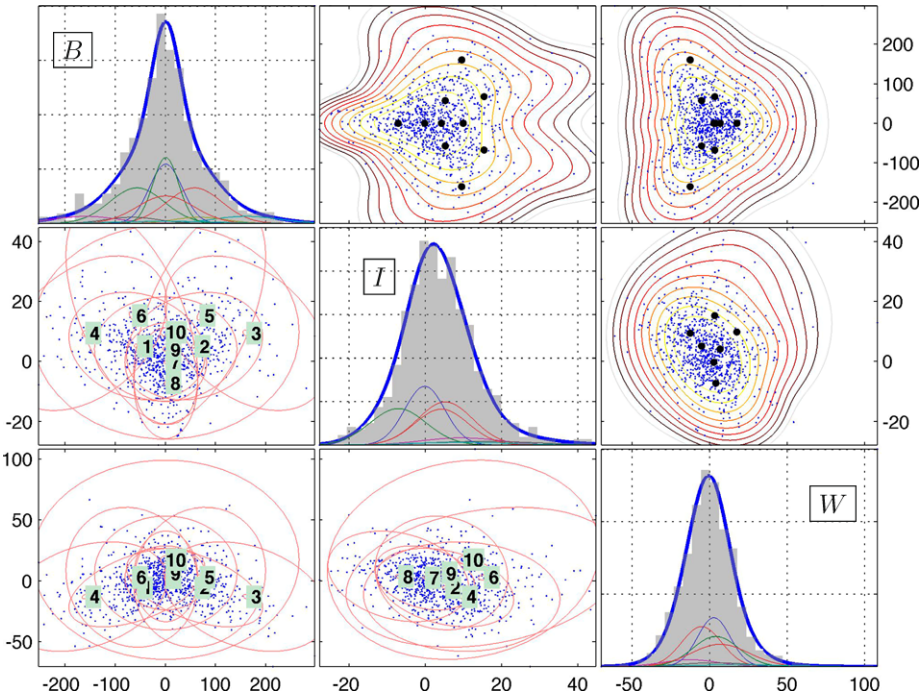

Figure 4 One and two-dimensional marginal distributions of a 3D mixture model for active-region pixels. The shaded portion of each on-diagonal plot (respectively, B, I, W) is a 1D histogram of active-region pixels for the corresponding image type. The colored lines underneath are the 1D projections of each of $J = 10$ Gaussian components. The thicker blue line is the weighted sum of all the Gaussian components, which is the same as the 1D marginal distribution of the Gaussian mixture model. The lower off-diagonal plots superimpose the concentration ellipses of the J Gaussian basis functions upon a 2D scatter plot of data. The upper off-diagonal plots, symmetric with the lower plots, show different quantities in rotated coordinates. The scatter plot of feature vectors is there superimposed on a contour plot showing the logarithm of the 2D marginal distribution of the weighted sum of basis functions.

structure manifests itself as an elongation in the density contours along I for B near zero. To aid readability, the density contours have been truncated at a factor of 1000 below the maximum density value. Below this value, the precise contours are not relevant to classification. It is clear from both the scatter plot and the fitted mixture that the contours of the density, representing pixels equally likely to be active region, are not axis-parallel.

5.3. Histogram Models

A directly comparable interpolated histogram model, based on the same active-region training data set, is shown in Figure 5. The histogram contained 16 bins for each of B, I, and W, of which 2176 were nonzero. Because the histogram is symmetric with respect to B and integrates to unity, it effectively had 1087 free parameters.

The 1D and 2D marginal distributions shown in the figure were produced integrating out the variable that is not shown. For the lower triangle of histogram-based plots, this amounts to summing over the 16 bins of the extra variable. For the upper triangle of plots showing marginal densities, this was done by interpolating a full 3D distribution on a dense grid and integrating out the variable that is not shown. The plots of 1D marginal distributions

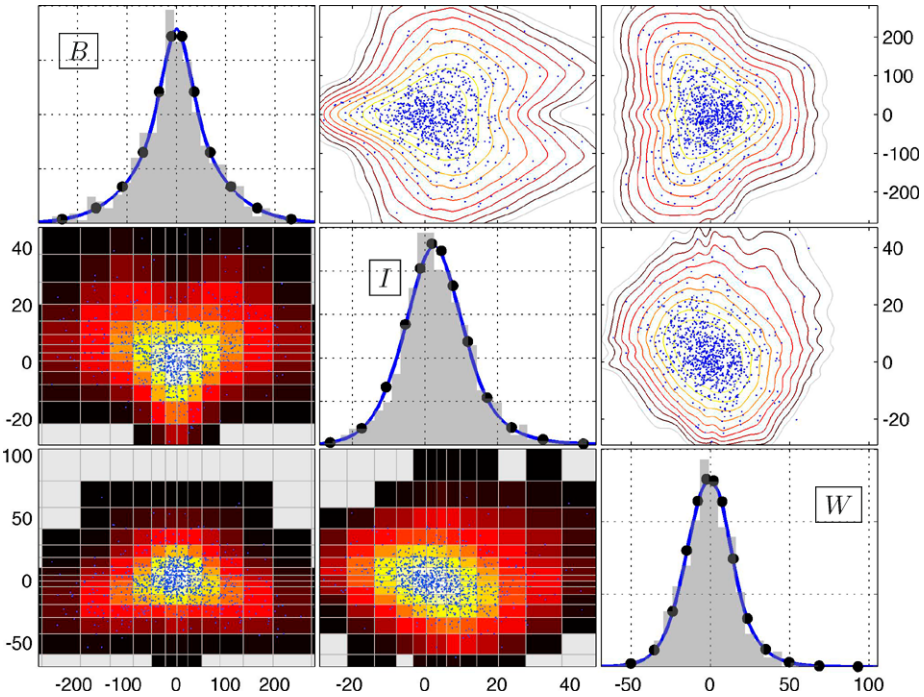

Figure 5 One and two-dimensional marginal distributions of a 3D histogram model for active-region pixels. The plot layout parallels Figure 4. The shaded portion of each on-diagonal plot (respectively, B, I, W) is an ordinary 1D histogram of active-region pixels for the corresponding image type. The thick blue line is the 1D marginal distribution of the interpolated histogram model; the black dots are histogram bin centers. The lower off-diagonal plots overlay a 2D scatter plot of data onto a patchwork corresponding to the variably-binned histogram, summing along the absent dimension. Light-gray patches indicate no data in any cell. The upper off-diagonal plots, symmetric with the lower plots, show different quantities in rotated coordinates. The scatter plot of feature vectors is superimposed on a contour plot showing the logarithm of the 2D marginal distribution of the histogram model.

along the diagonal, and the lower three plots of histogram bins, show the relative size of the histogram bins relative to the scatter of the data. The light-gray bins at the periphery of the 2D plots, and some others beyond the bounds of the plot, contained no data at all. (The gray 1D marginal histograms were computed using only a subset of the full data set and are for general comparison only.) The contour plots in the upper portion of the figure show many of the same features of the mixture densities, including the two-lobed structure and the projection due to supergranulation.

Figure 6 more directly compares the $B-I$ marginal distributions as computed by GMM and IHM. The left panel shows the two distribution estimates, $\log p_{\mathrm{GMM}}(B, I \,|\, \mathrm{AR})$ and $\log p_{\mathrm{IHM}}(B, I \,|\, \mathrm{AR})$. The contour lines are placed every 0.5 units of log-probability. To give perspective on these increments, under $\beta \equiv 0.25$ and a uniform δ, a change in classification of two neighbors would have the same evidentiary weight as moving 0.5 in log-probability. (Note that our typical spatially-weighted β and class-sensitive penalty moderates the effect of neighbors further.) Very close inspection of the log-probability plot reveals some differences: misalignment of the contour lines near $B = 0$, $I = 40$ and $I = -30$, wiggles in the IHM log-probability near $B = 200$, $I = 0$, and a slightly more rounded appearance to many

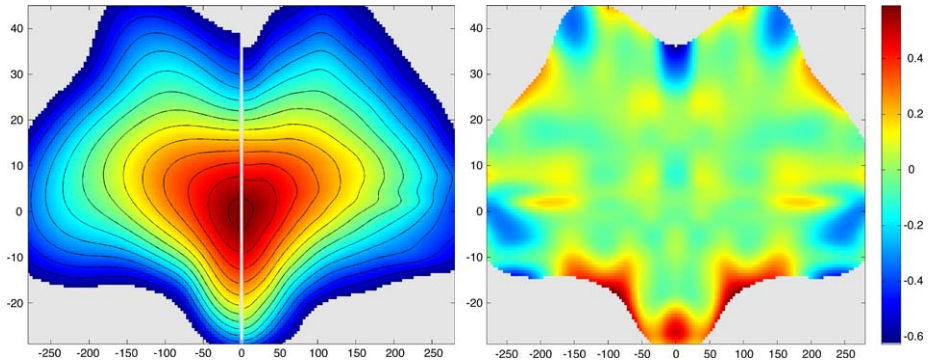

Figure 6 Comparison of GMM and IHM models for active-region pixels. All plots are for marginal distributions as a function of B (horizontal axis) and I (vertical). The left panel shows the log-probability by GMM ($B < 0$) and IHM ($B > 0$). Because of symmetry with respect to B, only half of each plot needs to be seen; they are separated by a fine gray line at $B = 0$. Black contour lines are placed every 0.5 in log-probability. The right panel shows the relative error $r(B, I)$, which is also in log units, calibrated by the colorbar (far right).

GMM contours. The right-hand panel shows the log-probability difference

$$r(B, I) = \log\big[p_{\mathrm{IHM}}(B, I \,|\, \mathrm{AR})/p_{\mathrm{GMM}}(B, I \,|\, \mathrm{AR})\big]. \tag{19}$$

Over the range of (B, I) plotted, which extends from the maximum probability density to a value 400 ($\approx e^6$) times less, the error remains less than 0.5 in log-probability. Over the central portion of the plot, the error stays significantly below 0.2, less than the typical effect of a change in classification of one neighbor. The analogous plots of marginal distributions for other pairs of observations are similar.

5.4. Feature Masks

Statistically derived masks with heavy spatial smoothing are compared to the original training segmentation for 06 November 1998 (see observations in Figure 1) in Figure 7. There is broad overall agreement between the segmentations but many differences in detail are apparent. We did not expect the statistical technique to reveal decaying regions, for example, since temporal evolution is not considered in computing class-conditional probabilities. Perhaps more interesting is the comparative lack of spatial detail revealed by the HW1999 labellings for active regions and particularly for enhanced and quiet network. Inspection of the rules for the HW1999 technique shows this to be by design: the rules use data as smoothed by broad Gaussian filters. These smoothing effects are not readily duplicated by our Bayesian prior model $[P(\mathbf{x})]$, but could be obtained by including the filtered images as additional observables. However, we feel that smaller-scale network structures are important for comparison with irradiance variation and note that HW1999 labelings do retain such spatial detail in subclasses that are not shown in the segmentation of Figure 7. We explore using this information to develop better training models in the discussion to follow.

5.5. Separation of Quiet Sun and Quiet Network

As already seen, not all of the HW1999 classes are immediately useful for our purposes of developing labellings with fine spatial scale. However, the HW1999 rules for deriving

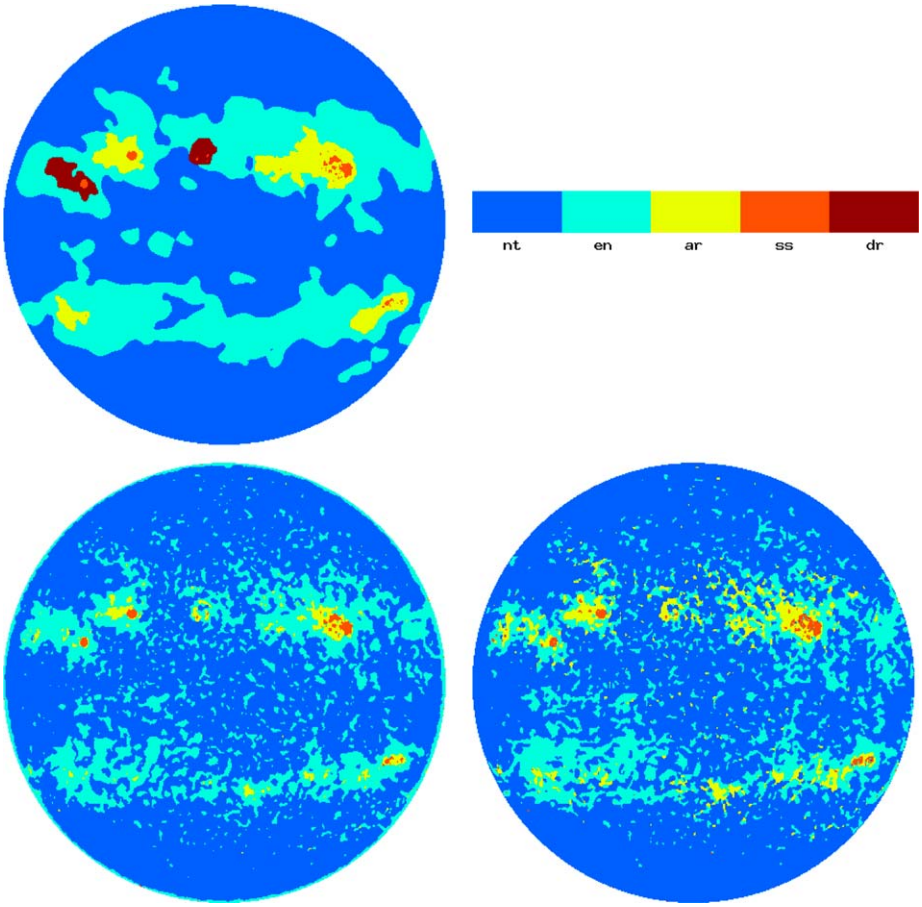

Figure 7 Image segmentations of the 06 November 1998 SPM observation for HW1999 training set (upper left). Bayesian segmentations using GMM (lower left) and IHM (lower right) are shown for comparison.

quiet-Sun, network, and sunspot labellings are specified at the pixel scale. Although not evident in the broad segmentation shown in Figure 7, these rules are encoded in more detailed masks, which are immediately usable for our purposes, as well as providing an interesting comparison.

Here we demonstrate the classifications that can be developed using the pixel-scale refinements of the HW1999 quiet and network classes. We extracted training sets of quiet and network pixels from the set of 38 labeled images described above. Typical data is shown in the left-hand panel of Figure 8. The influence of a ± 5 Gauss threshold is apparent, although there is significant crossover of quiet-Sun pixels above 5 Gauss due to neighborhood constraints. A GMM was fitted, as described above, to both the quiet and network feature vectors ($J = 12$ components were fitted from $N = 60\,000$ examples in each case). The resulting two models, $P(y \,|\, \mathrm{QS})$ and $P(y \,|\, \mathrm{NW})$, determine the relative probability of quiet Sun *versus* network.

To understand the interplay of the two classes, we plotted the "log-likelihood ratio"

$$L(B, I) = \log\bigl[p(B, I \,|\, \mathrm{NW}) / p(B, I \,|\, \mathrm{QS})\bigr] \tag{20}$$

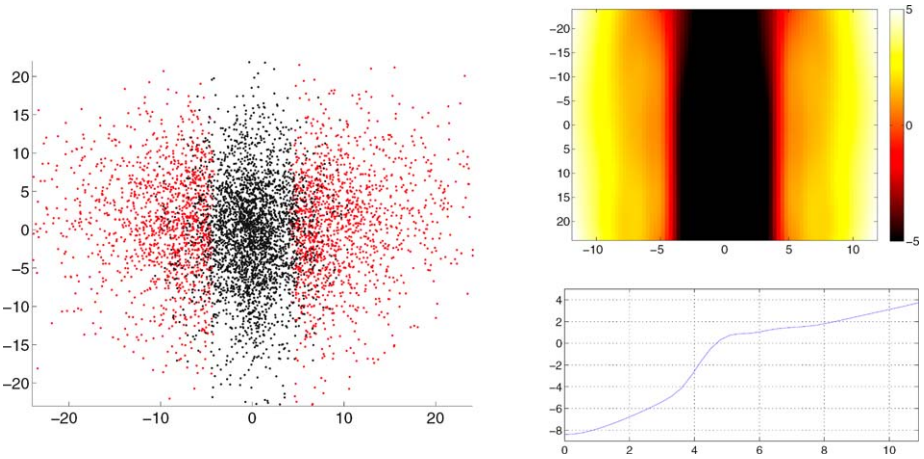

Figure 8 Classifying network and quiet Sun. Left: typical scatter plot of B (abscissa) against I showing network (red) and quiet Sun (black). Right: upper plot, relative evidence in favor of network as computed from the learned GMM, marginalized as a function of B (abscissa) and I. The log-probability scale runs from -5 to 5. Right: lower plot, relative evidence in favor of network, marginalized as a function of B alone. Because of symmetry, only positive B is shown in the lower plot.

on the right-hand panels of Figure 8. As expected from these training data, $L(B, I)$ is barely dependent on I, so we also show the log-likelihood ratio as a function of B alone. A high log-likelihood ratio is evidence for a network classification, and the contour at $L(B, I) = 0$ marks the watersheds of the two classes. What the Bayesian approach yields in this case is a quantitative measure of how far a given pixel value is from equiprobable. This evidence will be compared against the available spatial information in the optimization of Equation (2) to obtain a classification. Feature vectors falling in the red-colored area of Figure 8 are in an indeterminate zone where either classification would be supported, depending on the neighborhood composition.

Figure 9 shows the Bayesian labeling for 02 June 1996 which uses the mixture models just described. We used the image metadata to compute a spatially-weighted smoothing as in Equation (6) with $\gamma = 150$ and $\beta_0 = 0.3$. In the large image at left, several small active regions have been masked (being neither quiet Sun nor network). The overall spatial pattern of network seems to correspond well with the expected pattern of supergranulation.

The right-hand images compare the Bayesian classification with the HW1999 classification. There is very close agreement between the labellings. Across the full-disk image, 3.5% of pixels are assigned to different classes by the two methods, split about equally into the two types of disagreement (HW1999 chooses network while the Bayesian method chooses quiet Sun on 1.8% of pixels, with the reverse happening on 1.7% of pixels).

Referring to the difference image, we see that the Bayesian segmentation, for this β_0, has more small (one-pixel) network regions. This is because the rule structure of HW1999 disallows one-pixel network regions completely. Additionally, multi-pixel blocks can have each pixel below the trigger threshold of HW1999, but, taken as a block, above the effective threshold of the Bayesian method. This difference is the cause of the blue blocks in the difference image. Another difference is that the HW1999 labeling has more large, singly-connected network regions, which show up as one-pixel-wide red lines in the difference image. This is again due to the rule structure of the HW1999 algorithm. For HW1999, agreement with one neighbor has the same effect as agreement with multiple neighbors,

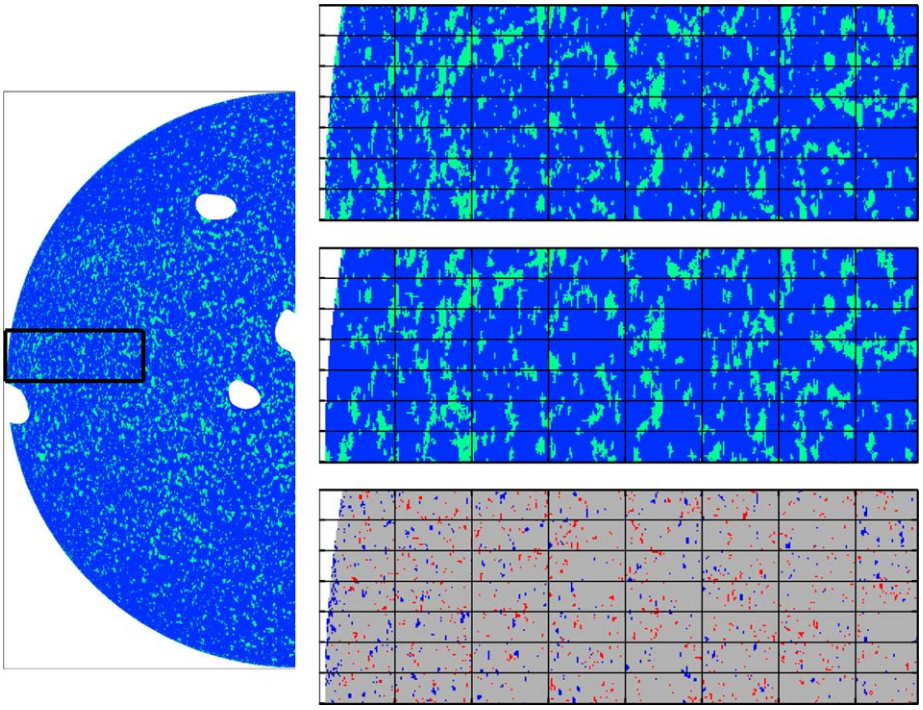

Figure 9 Labelings of quiet-Sun and network pixels based on the HW1999 labels. Left image (02 June 1996) shows quiet (blue) and network (green) labels over half the solar disk as inferred through the Bayesian segmentation algorithm and the 3D quiet-Sun and network models. Several areas around small active regions have been masked because they are neither quiet Sun nor network. The right-hand images are details of the full-disk labeling, taken from the boxed region: the Bayesian labeling (top panel), the HW1999 labeling (center panel), and the difference image (bottom panel). A grid has been superimposed to aid comparison. The difference image coding is: agreements in gray, disagreements where HW1999 determined quiet Sun in blue, disagreements where HW1999 determined network in red.

whereas the Bayesian method uses the cumulative number of agreements. For the same reason, the small shapes within the Bayesian segmentation have smoother boundaries than HW1999. We also see differences at the limb, resulting from a noisy magnetogram and the spatial decoupling by the Bayesian method there. Of course, different parameter choices could be made, which would lead to closer agreement between the two labellings.

6. Discussion

In this paper, we demonstrate the extension of the two-dimensional feature-identification algorithm of TPM2002 to the case of three observed quantities. We also develop histogram interpolation as a new alternative to Gaussian mixture models for computing class-conditional probability densities and show that the two techniques lead to very similar feature labellings. We discuss in some detail the introduction of a spatially-varying connectivity parameter and class-specific couplings, new extensions of the TPM2002 method which considerably increase modeling flexibility.

Our experimentation with training sets shows the importance of matching the spatial scale of training segmentations with the desired characteristics of the resulting la-

bellings. We discuss the separation of quiet "non-magnetized" Sun from quiet network using HW1999 labellings for "subclasses" of their image segmentations which show much more spatial structure than their summary labeling. We are developing similar techniques to separate quiet from enhanced network and enhanced network from active regions. We have developed and tested procedures (not discussed here) for computing time-series of feature characteristics from our statistical segmentations and in separate publications will compare the results with independent feature classifications from other researchers and with solar irradiance observations. Additionally, we plan to incorporate additional data, such as chromospheric observations, into our method and to extend our work to classification of SOLIS/VSM images.

Acknowledgements This research was supported by Heliophysics Supporting Research and Technology grant NNH04ZSS001N from NASA's Office of Space Science. The research described in this paper was carried out in part by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

References

Bloomfield, D.S., McAteer, R.T.J., Mathioudakis, M., Williams, D.R., Keenan, F.P.: 2004, *Astrophys. J.* **604**, 936. doi:10.1086/382062.
Boykov, Y., Veksler, O., Zabih, R.: 2001, *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(11), 1222.
Cadavid, A.C., Lawrence, J.K., Ruzmaikin, A.: 2008, *Solar Phys.* **248**, 247. doi:10.1007/s11207-007-9026-2.
Crouch, A., Barnes, G., Leka, K.: 2009, *Solar Phys.* **260**, 271.
Curto, J.J., Blanca, M., Martínez, E.: 2008, *Solar Phys.* **250**, 411. doi:10.1007/s11207-008-9224-6.
de Toma, G., White, O.R., Chapman, G.A., Walton, S.R., Preminger, D.G., Cookson, A.M., Harvey, K.L.: 2001, *Astrophys. J. Lett.* **549**. L131. doi:10.1086/319127.
de Toma, G., White, O.R., Chapman, G.A., Walton, S.R., Preminger, D.G., Cookson, A.M.: 2004, *Astrophys. J.* **609**, 1140. doi:10.1086/421104.
Ermolli, I., Berrilli, F., Florio, A.: 2000, In: Wilson, A. (ed.) *The Solar Cycle and Terrestrial Climate, Solar and Space Weather* **SP-463**, ESA, Noordwijk, 313.
Ermolli, I., Criscuoli, S., Centrone, M., Giorgi, F., Penza, V.: 2007, *Astron. Astrophys.* **465**, 305. doi:10.1051/0004-6361:20065995.
Foukal, P., Lean, J.: 1988, *Astrophys. J.* **328**, 347. doi:10.1086/166297.
Gelman, A., Jakulin, A., Grazia Pittau, M., Su, Y.S.: 2008, *Ann. Appl. Stat.* **2**(4), 1360. doi:10.1214/08-AOAS191.
Geman, S., Geman, D.: 1984, *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721.
Geman, S., Bienenstock, E., Doursat, R.: 1992, *Neural Comput.* **4**(1), 1.
Gorsuch, R.L.: 1983, *Factor Analysis*, 2nd edn. Lawrence Erlbaum, Hillsdale.
Gyori, L.: 1998, *Solar Phys.* **180**, 109.
Gyori, L., Baranyi, T., Turmon, M., Pap, J.: 2004, *Adv. Space Res.* **34**, 269.
Harvey, K.L., White, O.R.: 1999, *Astrophys. J.* **515**, 812. doi:10.1086/307035. Cited in text as HW1999.
Hastie, T., Tibshirani, R.: 1996, *J. Roy. Stat. Soc. (B)* **58**(1), 155.
Jones, H.P., Ceja, J.A.: 2001, In: Sigwarth, M. (ed.) *Advanced Solar Polarimetry – Theory, Observation, and Instrumentation* **CS-236**, Astron. Soc. Pac., San Francisco, 87.
Jones, H.P., Duvall, T.L. Jr., Harvey, J.W., Mahaffey, C.T., Schwitters, J.D., Simmons, J.E.: 1992, *Solar Phys.* **139**, 211.
Jones, H.P., Branston, D.D., Jones, P.B., Wills-Davey, M.J.: 2000, *Astrophys. J.* **529**, 1070.
Jones, H.P., Branston, D.D., Jones, P.B., Popescu, M.D.: 2003, *Astrophys. J.* **589**, 658.
Jones, H.P., Chapman, G.A., Harvey, K.L., Pap, J.M., Preminger, D.G., Turmon, M.J., Walton, S.R.: 2008, *Solar Phys.* **248**, 323.
Krivova, N.A., Solanki, S.K., Fligge, M., Unruh, Y.C.: 2003, *Astron. Astrophys.* **399**, 1. doi:10.1051/0004-6361:20030029.
Li, S.Z.: 2009, *Markov Random Field Modeling in Image Analysis*, 3rd edn. Springer, New York.
Lugosi, G., Nobel, A.: 1996, *Ann. Stat.* **24**(2), 687.
McAteer, R.T.J., Gallagher, P.T., Ireland, J., Young, C.A.: 2005, *Solar Phys.* **228**, 55. doi:10.1007/s11207-005-4075-x.

McLachlan, G., Peel, D.: 2000, *Finite Mixture Models*, Wiley, New York.
Pap, J.M., Turmon, M., Floyd, L., Fröhlich, C., Wehrli, C.: 2002, *Adv. Space Res.* **29**, 1923. doi:10.1016/S0273-1177(02)00237-5.
Pham, D.L., Xu, C., Prince, J.L.: 2000, *Annu. Rev. Biomed. Eng.* **2**, 315. doi:10.1146/annurev.bioeng.2.1.315.
Preminger, D.G., Walton, S.R., Chapman, G.A.: 2001, *Solar Phys.* **202**, 53.
Preminger, D.G., Walton, S.R., Chapman, G.A.: 2002, *J. Geophys. Res.* **107**, 6. doi:10.1029/2001JA009169.
Ripley, B.D.: 1988, *Statistical Inference for Spatial Processes*, Cambridge Univ. Press, New York.
Schrijver, K., Hurlburt, N., Mark, C., Freeland, S., Green, S., Jaffey, A., Kobashi, A., Schiff, D., Seguin, R., Slater, G., Somani, A., Timmons, R.: 2008, AGU Fall Meeting Abstracts, 1619.
Scott, D.W.: 1992, *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley, New York.
Simonoff, J.S.: 1998, *Smoothing Methods in Statistics*, corrected edn. Springer, New York.
Turmon, M.: 2004, In: Antoch, J. (ed.) *Compstat 2004 – Proceedings in Computational Statistics*, Physica, Heidelberg, 1909.
Turmon, M.J., Pap, J.M.: 1997, In: Babu, G.J., Feigelson, E.D. (eds.) *Proc. Second Conf. on Statistical Challenges in Modern Astronomy*, Springer, New York, 408.
Turmon, M., Mukhtar, S., Pap, J.: 1997, In: Heckerman, D., Mannila, H., Pregibon, D., Uthurusamy, R. (eds.) *Proc. Third Conf. on Knowledge Discovery and Data Mining*, MIT Press, Cambridge, 267.
Turmon, M., Pap, J., Mukhtar, S.: 2002, *Astrophys. J.* **568**(1), 396. doi:10.1086/338681. Cited in text as TPM2002.
Wenzler, T., Solanki, S.K., Krivova, N.A., Fröhlich, C.: 2006, *Astron. Astrophys.* **460**, 583. doi:10.1051/0004-6361:20065752.
Worden, J.R., White, O.R., Woods, T.N.: 1998, *Astrophys. J.* **496**, 998. doi:10.1086/305392.
Zhang, Y., Brady, M., Smith, S.: 2001, *IEEE Trans. Med. Imaging* **20**(1), 45.
Zharkova, V.V., Aboudarham, J., Zharkov, S., Ipson, S.S., Benkhalil, A.K., Fuller, N.: 2005, *Solar Phys.* **228**, 361. doi:10.1007/s11207-005-5623-0.