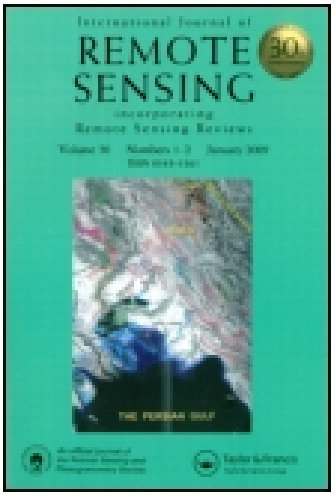

International Journal of Remote Sensing

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tres20

Statistical and neural pattern recognition methods for dust aerosol detection

P. Rivas-Perea[a], J. G. Rosiles[b] & J. Cota-Ruiz[c]

[a] Department of Computer Science, Baylor University, Waco, TX, USA

[b] Science Applications International Corporation, El Paso, TX, USA

[c] Department of Electrical and Computer Engineering, Autonomous University of Ciudad Juarez, Mexico

Published online: 27 Aug 2013.

To cite this article: P. Rivas-Perea, J. G. Rosiles & J. Cota-Ruiz (2013) Statistical and neural pattern recognition methods for dust aerosol detection, International Journal of Remote Sensing, 34:21, 7648-7670, DOI: 10.1080/01431161.2013.822660

To link to this article: http://dx.doi.org/10.1080/01431161.2013.822660

International Journal of Remote Sensing, 2013
Vol. 34, No. 21, 7648–7670, http://dx.doi.org/10.1080/01431161.2013.822660

Statistical and neural pattern recognition methods for dust aerosol detection

P. Rivas-Perea[a]*, J.G. Rosiles[b], and J. Cota-Ruiz[c]

[a]*Department of Computer Science, Baylor University, Waco, TX, USA;* [b]*Science Applications International Corporation, El Paso, TX, USA;* [c]*Department of Electrical and Computer Engineering, Autonomous University of Ciudad Juarez, Mexico*

(*Received 2 January 2013; accepted 21 June 2013*)

The dust storm phenomenon is a growing research area motivated by the potential harm that dust aerosols can cause to human beings, and by advances in remote-sensing technology. Nonetheless, formal studies on pattern recognition methods for dust aerosol detection are lacking and necessary. In this article, we present dust aerosol detection methods using statistical pattern recognition classifiers. Particularly, we investigate a support vector regression (SVR) approach and a large-scale approach to SVR. The feature set consists of multispectral thermal emissive bands from the Moderate Resolution Imaging Spectroradiometer (MODIS). We utilized four near-infrared bands: B20 (3.66–3.84 μm), B29 (8.40–8.70 μm), B31 (10.78–11.28 μm), and B32 (11.77–12.27 μm). Numerical performance evaluation shows that SVR outperforms other neural network-based classifiers in terms of a balanced error rate. Visually, both SVR paradigms accurately detect dust storms. The models demonstrated a strong ability to find non-trivial relationships within the spectral bands. The proposed detection methods are shown to be soil-independent and surface-invariant.

1. Introduction

Climate change exacerbates meteorological phenomena such as dust storms, which by themselves represent an increasing health threat to human beings (Rivera Rivera et al. 2009; Loomer 2006; Gail 2007). Therefore, an understanding of dust storm genesis, formation, propagation, and composition is important to reduce their impact and also to study their effect. Fortunately, advances in remote sensing such as in multispectral instrumentation permit the imaging of atmospheric and Earth materials based on their spectral signature over the optical range. In particular, dust aerosols propagated through the atmosphere in the form of dust storms can be detected through current remote-sensing instruments.

Some of the most relevant systems are based on data from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument and from the aerosol optical depth (AOD) product, which is also provided by National Aeronautics and Space Administration (NASA) (Levy et al. 2006; Sorek-Hamer et al. 2013). However, the production of the AOD product requires a considerable amount of processing time that may introduce an unnecessary delay before its information can be processed for aerosol analysis. Other approaches are based on the so-called 'band-math', where simple operations between bands are used

*Corresponding author. Email: Pablo_Rivas_Perea@Baylor.edu

to provide a visual (and subjective) display of the presence of dust storms (Chen, Lin, and Lee 2006; Ackerman 1997; Rivera Rivera et al. 2009). Other multispectral image-based research focuses on the analysis of dust storm transport by observing its trajectory (El-Askary et al. 2004; Mahler et al. 2006; Prospero et al. 2002; Rivera Rivera 2006; Lee et al. 2009; Janugani et al. 2009). However, these findings require more accurate dust analysis and detection methods (Khazenie and Lee 1992; Kanellopoulos, Austin, and Wilkinson 1997; Kast et al. 1977; Tso and Mather 2009; Hillger and Ellrod 2003; Agarwal et al. 2007; Miller 2003; San-Chao et al. 2006).

Furthermore, given the large amounts of data produced by the MODIS instrument, it is also desirable to have automated systems that assist scientists in finding or classifying different Earth phenomena. For example, Aksoy et al. developed a visual grammar scheme that integrates low-level features to provide a high-level spatial scene description of land cover and land usage (Aksoy et al. 2005). As far as the authors are aware, only a few similar automated schemes for dust aerosol detection based on statistical pattern recognition techniques have been reported (Rivas-Perea and Rosiles 2010; Rivas-Perea et al. 2010; Chacon-Murguía et al. 2011).

This research selects features based on spectral signatures already known to express dust aerosols (Ackerman 1997; Hao and Qu 2007; San-Chao et al. 2006). Such features allow high performance, accuracy, and real-time detection of dust aerosols. These features are analysed using a support vector regression (SVR) approach, and we include a comparison with the maximum-likelihood (ML) classifier, a probabilistic neural network (PNN), and a feed-forward neural network (FFNN). We will show that the SVR approach provides better detection and representation of dust storm events.

This article is organized as follows. Section 2 discusses the features selected from MODIS bands according to a literature review, as well as the event objects of study. Section 3 presents an overview of the classification algorithm. Section 4 addresses a numerical comparison of the results and discusses visual results over both land and ocean, followed by a discussion of segmentation. Finally, Section 5 presents the conclusions of this research.

2. Feature and event selection from MODIS data

The MODIS instrument is carried on board the NASA Terra and Aqua satellites. MODIS provides Earth information over 36 spectral bands, easing analysis of the atmosphere, ocean, land, and more. The data available in MODIS Level IB make possible dust storm analysis. Dust storm visual assessment can be achieved using MODIS bands B1, B3, and B4, since these match the human visual spectrum (Levy et al. 2006; Team et al. 1997). For visualization purposes, we can produce RGB-like images by mapping $R = \text{B1}$, $G = \text{B4}$, and $B = \text{B3}$. The colour images produced in this research were enhanced using a non-linear function (Fanning 2012; Rivas Perea 2011). As an example, Figure 1 shows an RGB (pseudo-coloured) image of the southwestern US area.

In 1997, Ackerman demonstrated that using bands B32 and B31 improves dust storm visualization (Ackerman 1997). Later, Hao et al. established that a linear combination of bands B20, B29, B31, and B32 can be utilized for dust storm visualization (Hao and Qu 2007). More recently, Alkhatib et al. demonstrated that B20 is particularly relevant in the study of dust aerosols (Alkhatib and Cabrera 2011). Based on these findings, we designed a classification scheme using the following MODIS thermal emissive bands: B20 (3.66–3.84 μm), B29 (8.40–8.70 μm), B31 (10.78–11.28 μm), and B32 (11.77–12.27 μm); our previous research using such bands can be found in Rivas-Perea et al. (2010), Rivas-Perea and

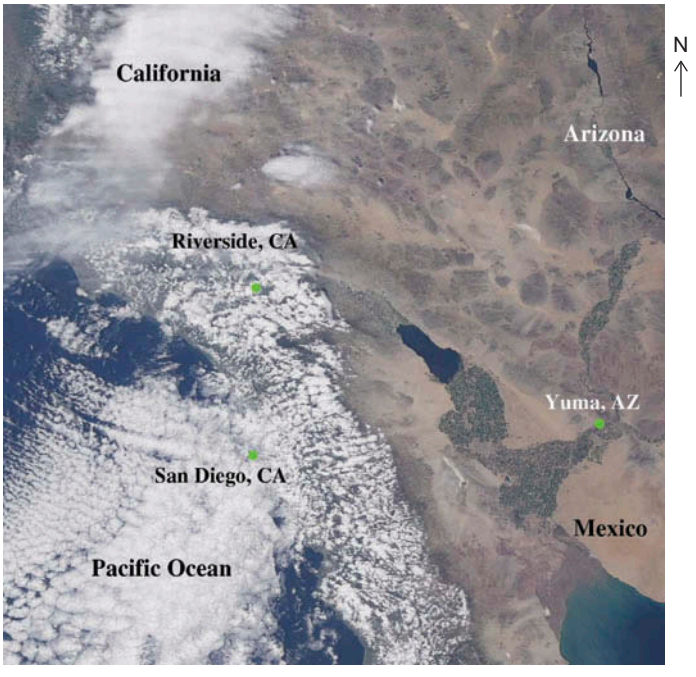

Figure 1. True colour RGB composite of the southwestern USA, 15 April 2001, 18:05 UTC. Satellite: Terra. Instrument: MODIS.

Rosiles (2010), Rivas Perea (2011), and Chacon-Murguía et al. (2011). MODIS level 1B data contain 16 bit scaled thermal emissive bands that need to be recovered to their original scale in (W m^{-2} μm^{-1} sr^{-1}) units (Kopp and Lean 2011). The recovery process is given by

$$L = \kappa\,(\iota - \eta), \tag{1}$$

where $L \in \mathbb{R}$ denotes the recovered radiance, k is the radiance scale factor, η is the radiance offset, and ι is the scaled (i.e. raw) data. We select our feature vector as

$$\boldsymbol{F} = \left[L^{B20}\ L^{B29}\ L^{B31}\ L^{B32}\right]^{T}, \tag{2}$$

which consists of the four recovered radiances corresponding to the bands described by the superscripts of L; T denotes vector transposition.

In our experiments, we used 31 different events (i.e. multispectral images) corresponding to the southwestern US and northwestern Mexico areas, as reported in Novlan, Hardiman, and Gill (2007). All data samples[1] were downloaded using NASA's online Warehouse Inventory Search Tool (WIST; Kobler and Berberi 1991). Each of the 31 events produced 85 million feature vectors that were manually selected to be part of two possible classes: *dust present* and *dust not present.* The feature set is unbalanced, in that the number of vectors corresponding to dust is much lower than the number of vectors corresponding to dust-free areas in an image. For classification purposes, the data were separated into three sets: *training*, used to build the model; *validation*, used to determine the generalization capabilities of the model; and *testing*, used to determine the numerical performance.

3. Overview of pattern classification schemes

3.1. ML classifier

We introduce first the ML classifier, which has been extensively used by the remote-sensing community. The ML classifier is a Bayesian classifier that relies on probabilistic arguments. First, we define two possible events or classes, *dust* and *background*, for our particular problem. We define a discrete random variable C with domain $S_c = \{c_1,\ c_2\}$, where the two values are used to represent dust and background, respectively. The probability mass function of C will be denoted as $f_C(c)$. Next, we let $\boldsymbol{X}$ be a random vector in $\mathbb{R}^n$, which in our case $n = 4$; $\boldsymbol{X}$ represents the observable evidence (i.e. feature vector) used for detecting the presence of dust in the corresponding pixel. The *class-conditional* probability density function (pdf) of $\boldsymbol{X}$ is defined as $f_{\boldsymbol{X}|C}(\boldsymbol{x}|c)$.

Bayes decision theory provides the following two class classification rule (Duda, Hart, and Stork 2001):

$$\boldsymbol{F} \in c_1 \quad \text{if} \quad \psi_{c_1}(\boldsymbol{F}) > \psi_{c_2}(\boldsymbol{F}) \quad \text{else} \quad \boldsymbol{F} \in c_2, \tag{3}$$

for a feature vector $\boldsymbol{F}$, where $\psi_{c_i}(\boldsymbol{F})$ are discriminant functions defined as

$$\psi_{c_i}(\boldsymbol{x}) = f_{\boldsymbol{X}|C}(\boldsymbol{x}|c_i) f_C(c_i). \tag{4}$$

Without further knowledge of the seasonal and/or meteorological conditions of the area sensed by the MODIS apparatus, it is safe to assume that the dust and background classes are equiprobable. Hence, the $f_C(c_i)$ terms can be dropped, simplifying the classifier design to providing only the class-conditional or *likelihood* functions.

A common approach is to assume that $\boldsymbol{X}$ follows a multivariate Gaussian distribution. Furthermore, the problem is simplified using the log-likelihood of the Gaussian which results in the well-known discriminant functions

$$\psi_{c_i}(x) = \left(x - \boldsymbol{\mu}_i\right)^T \boldsymbol{\Sigma}_i^{-1} \left(x - \boldsymbol{\mu}_i\right) - \det(\boldsymbol{\Sigma}_i), \tag{5}$$

where $\boldsymbol{\mu}_i$ is the mean feature vector for class c_i, $\boldsymbol{\Sigma}_i$ is the corresponding feature's covariance matrix, and det $(\cdot)$ is the determinant function. The mean vectors and covariance matrices are obtained from training data using ML estimators. Mathematically, the discriminant functions, $\psi_{c_i}(\cdot)$, are quadratic functions that measure the distance between a feature vector and a mean vector weighted by $\boldsymbol{\Sigma}_i$. Hence, a feature vector $\boldsymbol{F}$ will be classified to the closest of the two classes.

3.2. PNN-based detection

Specht's PNN is a neural network that learns (i.e. is trained) in a supervised manner (Specht 2002). It is widely used in pattern recognition applications. The PNN is inspired by Bayesian classification and does not require training (Chettri and Cromp 1993). The network estimates the pdfs of the feature vectors assuming Gaussian distributions. The PNN has a four-layered architecture, as shown in Figure 2.

The first layer is an input layer receiving the features, $\boldsymbol{F} \in \mathbb{R}^{n\times 4}$. The second layer contains exponential functions, $\varphi(\cdot)$, at each node, and the node's count corresponds to the

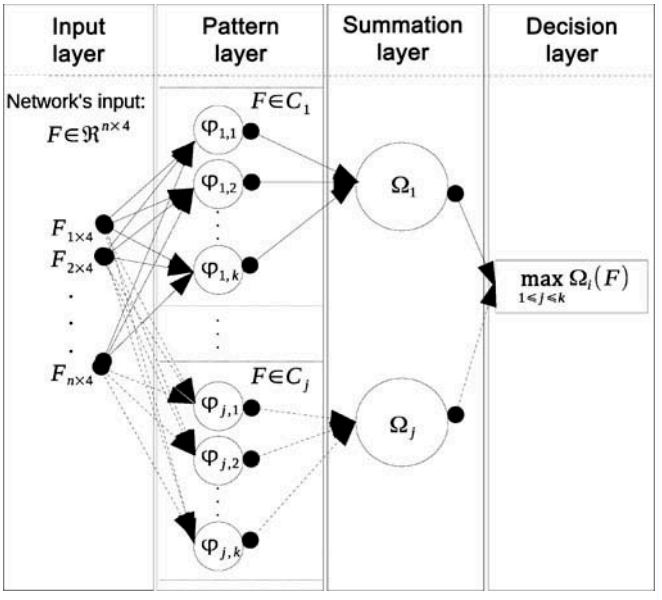

Figure 2. Hybrid architecture of the probabilistic neural network. Note the probabilistic nature embedded in a neural architecture.

number of κ samples belonging to the jth class. These nodes are called pattern units and are fully connected to the input nodes. The pattern layer's output is denoted by

$$\varphi_{jk}(\boldsymbol{F}) = \frac{1}{(2\pi)^{\frac{d}{2}}\sigma^d} e^{-\frac{1}{2\sigma^2}\left(\boldsymbol{F}-\upsilon_{jk}^{\boldsymbol{F}}\right)^T\left(\boldsymbol{F}-\upsilon_{jk}^{\boldsymbol{F}}\right)}. \tag{6}$$

The third layer contains summation units to complete the probability estimation. The number of summation units equals to the number of classes. The jth summation unit receives input only from those pattern units belonging to the jth class. This layer computes the likelihood of $\boldsymbol{F}$ being classified as C. This is done by averaging and summarizing a neuron's output belonging to the same class:

$$\Omega_j(\boldsymbol{F}) = \frac{1}{N_j}\sum_{k=1}^{N_j}\varphi_{jk}(\boldsymbol{F}), \tag{7}$$

where N_j is the number of examples in class j.

The last layer is called the 'decision layer'. It classifies $\boldsymbol{F}$'s pattern according to the Bayesian decision rule given by

$$F \in C_j \quad \text{if} \quad C_j\left(\Omega_j(\boldsymbol{F})\right) = \arg\max_{1\leq i\leq j}\ \Omega_i(\boldsymbol{F}). \tag{8}$$

Thus, the maximum of the summation nodes' output characterizes the PNN's general output. The function $\Omega_j(\cdot)$ gives the probability of the jth class. This allows us to generate probabilistic visualizations of dust storms.

The parameter σ is estimated by the method of Srinivasan et al. (Ramakrishnan and Selvan 2007). This method involves a pre-normalization phase consisting of two steps: first, subtract the mean μ_F from the training feature vector $\boldsymbol{F}$; second, divide $\boldsymbol{F}$ by its standard deviation σ_F. Then, σ would be the absolute difference between the two smallest normalized variances. This completes the PNN modelling since it needs no training phase.

3.3. *Feed-forward back-propagation neural network-based detection*

Multilayered FFNNs are of particular interest in pattern recognition and classification applications because they can approximate any square-integrable function to any desired degree of accuracy and can exactly implement any arbitrary finite training set (Hoogerheide, Kaashoek, and Dijk 2003; Gallant and White 1988). Therefore, we designed an FFNN to model a dust storm by approximating the true posterior pdf $f_{C|x}\,(C = j|\boldsymbol{X} = \boldsymbol{F})$. A simple FFNN contains an input layer and an output layer, separated by l layers (the set of l layers is known as the *hidden layer*) or neuron units. Given an input sample clamped to the input layer, the other units of the network compute their values according to the activity of the units they are connected to in the previous layers. In this model, we consider the particular topology where the input layer is fully connected to the first hidden layer, which is fully connected to the second layer and so on up to the output layer.

Given an input $\boldsymbol{F} \in \mathbb{R}^{n\times 4}$, the value of the jth unit in the ith layer is denoted $h_j^i\,(\boldsymbol{F})$, with $i = 0$ referring to the input layer, $i = l + 1$ referring to the output layer. We refer to the size of a layer as $\left|\boldsymbol{h}^i\,(\boldsymbol{F})\right|$. The default activation level is determined by the internal bias b_j^i of that unit. The set of weights W_{jk}^i between unit $h_k^{i-1}\,(\boldsymbol{F})$ in layer $i - 1$ and unit $h_k^{i-1}\,(\boldsymbol{F})$ in layer i determines the activation of unit $h_j^i\,(\boldsymbol{F})$ as follows:

$$h_j^i\,(\boldsymbol{F}) = \Phi\Big(a_j^i\,(\boldsymbol{F})\Big), \tag{9}$$

where $a_j^i\,(\boldsymbol{F}) = \sum_k W_{jk}^i h_k^{i-1}\,(\boldsymbol{F}) + b_j^i$, $\forall i \in \{1, \ldots, l\}$, with $h^0\,(\boldsymbol{F}) = \boldsymbol{F}$, and $\Phi = \mathrm{sigm}\,(\cdot)$ is the sigmoid activation function defined as $\mathrm{sigm}\,(a) = \frac{1}{1+e^{-a}}$, which could be replaced by any desired activation function. Given the last hidden layer, the output layer is computed similarly by

$$\begin{aligned} o(\boldsymbol{F}) &= h^{l+1}(\boldsymbol{F}), \\ \boldsymbol{h}^{l+1}(\boldsymbol{F}) &= \boldsymbol{\Phi}\big(\boldsymbol{a}^{l+1}\,(\boldsymbol{F})\big), \end{aligned} \tag{10}$$

where $\boldsymbol{a}^{l+1}\,(\boldsymbol{F}) = \boldsymbol{W}^{l+1}\boldsymbol{h}^l\,(\boldsymbol{F}) + \boldsymbol{b}^{l+1}$, and the activation function $\boldsymbol{\Phi}$ is of the softmax type in order to define probabilities of classes (see texts such as (Bishop 1995; Canty 2007) for a detailed development). Thus, when an input sample F is presented to the network, the application of (9) at each layer will generate a pattern of activity in the different layers of the neural network and produce an output at the end.

The FFNN requires a training phase to build the model. In this training phase, we used the 'Levenberg–Marquardt' algorithm along with with a back-propagation strategy to update the weights $\boldsymbol{W}$ and biases $\boldsymbol{b}$. As a learning function, we used the well-established method of gradient descent with momentum weight and bias. The FFNN training phase ends when any of the following conditions holds:

- a number of 100 epochs (i.e. training iterations) is reached;
- the actual mean squared error (MSE) is 0; and
- the gradient step size is less than or equal to 1×10^{-10}.

A well-established technique for preventing over-fitting in the training was also implemented. This technique consists of partitioning the training set into two sets, training (80%) and validation (20%), such that when the MSE has not been decreased in the past five iterations using the internal validation set, then the training phase stops and rolls back to the weights associated with the previous minimum MSE.

3.4. Typical and large-scale linear programming support vector machines-based detection

In this research, we also used a linear programming support vector regression (LP-SVR), which was trained with a reduced training data set as described in the next section. We formulate an $l_1 - \text{SVR}$ {i.e. norm-1 SVR) problem as the following constrained optimization problem:

$$\begin{aligned} &\min_{\alpha,\xi} \ \|\alpha\|_1 + 2\eta \sum_{i=1}^{N} \xi_i \\ &\text{s.t.} \quad \begin{cases} j_k - \sum_{i=1}^{N} \alpha_i \boldsymbol{K}(\boldsymbol{F}_k, \boldsymbol{F}_i) - b \le \varepsilon + \xi_k \\ \sum_{i=1}^{N} \alpha_i \boldsymbol{K}(\boldsymbol{F}_k, F_i) + b - j_k \le \varepsilon + \xi_k \\ \xi \ge 0 \end{cases} \\ &\text{for} \quad k = 1, 2, \ldots, N, \end{aligned} \tag{11}$$

where α is the Lagrange multiplier associated with the support vectors (SVs). The summation in the cost function accounts for the ε-insensitive training error, which forms a tube where the solution is allowed to be defined without penalization. The constant $\eta > 0$ describes the trade-off between the training error and the penalizing term $\|\alpha\|_1$. The variable ξ_i is a non-negative slack variable describing the ε-insensitive loss function. Finally, the variable b is a bias and $\boldsymbol{K}(\cdot,\cdot)$ is a valid kernel function (see (Mercer 1909; Courant and Hilbert 1966)). The parameter vector α and the bias b are the unknowns and can take on any real value.

Since, by definition, the requirement of an LP-SVR is to have the unknowns greater than or equal to zero, we typically decompose such variables in their positive and negative parts. Therefore, we denote $\alpha = \alpha^+ - \alpha^-$, and $b = b^+ - b^-$. Then, in order to pose the problem as a linear program in its canonical form and in order to use an interior point method solver, Problem (11) must have no inequalities; thus, we need to add a slack variable u, which leads to the following linear programming problem:

$$\begin{aligned} &\min_{\alpha^{\pm},\, b^{\pm},\, \xi,\, \mathrm{u}} \ \sum_{i=1}^{N} \left(\alpha_i^+ + \alpha_i^- + 2\eta\xi_i\right) \\ &\text{s.t.} \quad \begin{cases} -\sum_{i=1}^{N} \left(\alpha_i^+ - \alpha_i^-\right) \boldsymbol{K}(\boldsymbol{F}_k, \boldsymbol{F}_i) - b^+ + b^- - \xi_\kappa + u_k = \varepsilon - j_k \\ \sum_{i=1}^{N} \left(\alpha_i^+ - \alpha_i^-\right) \boldsymbol{K}(\boldsymbol{F}_k, \boldsymbol{F}_i) + b^+ + b^- - \xi_\kappa + u_k = \varepsilon + j_k \\ \alpha_k^+, \ \alpha_k^-, \ b^+, \ b^-, \xi_k, u_k \ge 0 \end{cases} \\ &\text{for} \quad k = 1, 2, \ldots, N, \end{aligned} \tag{12}$$

which is one of the formulations we used along with an interior point solver, a radial-basis function (RBF) kernel with parameter $\sigma = 0.125$, a value of $\eta = 0.5$, and $\varepsilon = 0.1$; see (Rivas Perea 2011) for more details regarding this formulation.

4. Results

To summarize the description from previous sections, the proposed dust detection process consists of the following steps. First, multispectral data from MODIS Level 1B thermal emissive bands B20, B29, B31, and B32 are recovered using (1). Second, for each image location, a feature vector $\boldsymbol{F}$ is formed (2) and used as input to each classifier. Third, dust detection is performed by observing the result of (3) for ML, (8) for PNN, (10) for FFNN, and (12) for both the LP-SVR and its large-scale version LS LP-SVR. In this section, we address the results of comparing the proposed dust storm detection models.

4.1. Classifier training

As explained in Section 2, in our experiments, we account for approximately 85 million feature vectors from 31 events. Some pattern recognition approaches like the PNN are very limited by the number of samples that can be used to construct the model. As will be recalled, the PNN requires as many neurons as the number of samples provided times the number of classes. Therefore, we aim to use the least possible number of samples for training the classifiers.

In this article, the training set selection is based on Mather's criteria (Mather, Tso, and Koch 1997; Kanellopoulos, Austin, and Wilkinson 1997) and the Karhunen-Loeve Transformation (KLT). Mather's criteria establish that the number of training samples for each class must be at least 30 times the number of bands used as features. In our case, we selected 60×4 feature vectors per class (the feature vectors consist of four features). This decision was driven by the fact that we want neural models with low complexity and, hence, cheap hardware implementation capabilities. Furthermore, the number of 240 samples per class matches a confidence level of 99.9% according to a well-established sample number selection criterion described in Congalton and Green (2008). Thus, for each class (dust and non-dust), we have 240 vectors. The selection of the training set was performed with KLT (i.e. principal component analysis). In this case, the KLT was used to find the most representative feature vectors. For each class, we computed the covariance matrix and then used the KLT to identify those feature vectors associated with the largest initial 240 eigenvalues.

4.2. Comparison between classifiers

We selected four performance metrics to evaluate the three proposed classification methods. These metrics are processing time per feature vector, precision, accuracy, and the area under the receiver operating characteristics (ROC) curve (AUC). The AUC is a widely used metric because of its superiority in reflecting the true performance of a classification system. The precision and accuracy metrics can be computed as follows:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{13}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN}, \tag{14}$$

where *TP* is the total number of 'True Positives', *FP* is the number of 'False Positives', and *TN* ('True Negative') and *FN* ('False Negative') represent similar quantities.

Sometimes these metrics may become biased towards false positive counts, and considering we have many examples of the non-dust class, we use the balanced error rate (BER) defined as follows:

$$\mathrm{BER} = \frac{1}{2}\left(\frac{FP}{TN + FP} + \frac{FN}{FN + TP}\right). \tag{15}$$

Clearly, the BER meets the classical misclassification rate if there is an equal number of positive and negative examples, in which case $TN + FP = FN + TP$ (see Cawley 2006).

For this evaluation, we selected eight worldwide events from the NASA WIST database. Reduced training and validation feature sets were selected using the procedure described in the previous subsection. Each of these sets consisted of 240 feature vectors. All of the remaining feature vectors were used as test and validation sets divided evenly, i.e. 42.5 million each.

To reduce computational cost and keep similar constraints to the estimation of classifier parameters, we decided to use the reduced training sets on the ML, PNN, FFNN, and LP-SVR. Note that the ML and FFNN can handle larger training sets compared with the PNN and LP-SVR, whose training and model complexity grow with the data; however, since the reduction of the feature vectors number is performed with a statistical method aimed to preserve the statistical properties of the data, there is not much gain in using the large training set for ML and FFNN; moreover, in using the same reduced training set for the four classifiers, we achieve fairness in comparing performance results. The performance metrics were computed using only the validation set that produced the results shown in Table 1. Note that since we have a large number of test samples, it is sufficient to use these overall metrics to compare classifiers, i.e. we do not have a problem with data insufficiency.

For comparison purposes, we decided also to test a large-scale method based on a LP-SVR (LS LP-SVR) scheme recently proposed by Rivas et al. (Rivas-Perea and Cota-Ruiz 2012; Rivas Perea 2011). LS LP-SVR allows us to perform learning with the complete training set (i.e. 42.5 million feature vectors) rather than the reduced set. We use this formulation to compare results to those classifiers trained with the reduced training set. The proposed scheme uses an interior point method (IPM) to solve a sequence of smaller LP optimization sub-problems according to a decomposition strategy. Then we perform chunking, preserving the support vectors of the sub-problem at each iteration. We take advantage of the quadratic rate of convergence of IPM on the proposed LP-SVR method to find the global solution to the regression/classification problem in fewer iterations. LS LP-SVR was

Table 1. Classifier performance. Rank of classifiers is shown in parenthesis.

	Precision	Accuracy	AUC	BER	Processing time (millisecond)
ML	0.5255 (5)	0.6779 (5)	0.4884 (5)	0.2259 (5)	**0.0141** (1)
PNN	**0.8080** (2)	**0.8816** (2)	0.7035 (3)	0.0536 (3)	0.2393 (5)
FFNN	0.7664 (4)	0.8412 (4)	0.6293 (4)	0.0729 (4)	**0.0459** (2)
LP-SVR	0.7907 (3)	0.8678 (3)	**0.7117** (**2**)	**0.0502** (**2**)	0.0809 (3)
LS LP-SVR	**0.8295** (1)	**0.9104** (1)	**0.7349** (1)	**0.0318** (1)	0.0974 (4)

Note: Bold type numbers indicate the best two cases per column.

thus trained using the complete training set, and once the SVs were identified, the remaining non-SVs were used as part of the testing set; also let us recall that this is only used for comparison purposes.

From Table 1, we observe that among the neural classifiers, the PNN always outperforms the FFNN. Among the statistical classifiers, LP-SVR clearly outperforms the ML classifier. In fact, the ML classifier should not be used according to the results because it misclassifies samples almost half the time. This could have different causes, e.g. the multivariate Gaussian model is not adequate, or perhaps the training data set was not large enough to provide good parameter estimates. From the table, we can also see that LP-SVR is the best performer from the four classifiers of interest; this is attributed to the fact that the training stage of LP-SVR solves an optimization problem that finds the best classification (based on regression) function from the training data available. The PNN also performs fairly well, and we could attribute this success to the fact that the PNN is very close to a mixture of Gaussian classification problem formulation which is also known to perform very well in binary classification problems (Yang, Zwolinski, and Chalk 1998; Yang and Chen 1998). Overall, the FFNN and LP-SVR show a good balance of speed, little error, and good accuracy and precision. Finally, we see that there is a benefit to using a large training set from the LS LP-SVR classifier; the metrics are better compared with all others. However, the resources to train the LS LP-SVR are substantially higher and the metric improvements are rather modest for our purposes. Since the training is done off-line, it could also be considered a good alternative when the resources are available.

The processing time is an important measure while modelling real-time processing systems. In the case of the MODIS instrument, a complete scan (i.e. 10×1053 pixels at 36 bands at 1 km resolution) is produced every 6.25 seconds. In accordance with the feature vector used in this research, MODIS would produce approximately one feature vector every 0.5 milliseconds. Thus, a real-time system must perform a classification in less than or equal to that time. Table 1 also shows the processing time per scan in milliseconds. Therefore, since for all classifiers the time required to classify a feature vector is less than 0.25 milliseconds, they can be considered suitable for real-time detection at 1 km resolution. In contrast, the typical MODIS AOD product would take much longer and, moreover, it is released at 10 km resolution.

In addition to the typical comparison of methods, we can measure the statistical significance of the results shown in Table 1, employing the Friedman test (see Demšar (2006) for more details), and rank the classifiers according to their performance. Table 2 shows the rankings and average rank for each classifier in the sixth row.

Clearly, the two proposed LP-SVR classifiers are better ranked than all of the other classifiers in the average case. Using the Friedman test, we conclude that the classifiers

Table 2. Classifier ranking according to Friedman test.

	ML	PNN	FFNN	LP-SVR	LS	LP-SVR
Precision	5	2	4	3	1	
Accuracy	5	2	4	3	1	
AUC	5	3	4	2	1	
BER	5	3	4	2	1	
Processing	Time	1	5	2	3	4
Average	4.2	3	3.6	2.6	1.6	
Critical difference	$CD = 1.960\sqrt{\frac{2 \times 3}{6 \times 5}} = 0.876$					

perform differently and the null hypothesis is rejected, showing a chi-squared Friedman statistic of $x^2 = 7.840$ with probability $p = 0.0976$. Focusing on the two better ranked classifiers (i.e. LP-SVR and LS LP-SVR), we computed the critical difference (CD) between any two different classifiers, which is 0.876. The actual difference between the two best classifiers is 2.6–1.6 = 1. Therefore, since the difference between the two best classifiers is beyond the critical difference, i.e. $CD = 0.876 < 1$, we conclude that the LS LP-SVR approach is significantly better than the second best classifier. The same conclusion that the LS LP-SVR performs significantly better than any other classifier applies to the remaining classifiers and there is no need to compute the remaining differences. The critical difference value was computed using a significance level of $\alpha = 0.05$. Note, however, that the difference between the rankings of the LP-SVR (without the large-scale approach) and the PNN is not greater than the critical difference, meaning that they are not significantly different and they perform very similarly, in a statistical sense. Nonetheless, the LP-SVR still performs significantly better than the FFNN and ML. For more details of the statistical analysis presented here, the reader is encouraged to consult Demšar (2006) and García and Herrera (2008).

In terms of the visual assessment of dust storm detection, we divided the results into two categories: dust over land, and dust over ocean. This division is intentionally made since the methods were modelled using only dust over land data. Thus, the validation using events with dust over ocean will demonstrate the generalization capabilities of the methods.

4.3. *Generalization of dust detection over land*

A major difficulty in detecting dust over land is the variability of the dust physical properties depending on the region. Thus soil variability can adversely impact the performance of the classifiers. The reader must recall that the models were constructed using data from southwestern US events. Therefore, here we demonstrate that the classifiers trained with these data have the ability to detect dust independently of soil type (i.e. the same classifier parametrization can be used for different regions around the world regardless of variation in soil composition).

We consider a case of dust storm detection from 10 April 2001, covering the North of Mexico, Texas, and New Mexico. A pseudo-coloured image is shown in Figure 3(*a*), with the region of interest (RoI) depicted in Figure 3(*j*). Figure 3(*b*) shows the result reported by the MODIS AOD level 2 product, where higher intensities are associated with more dense aerosols. Figures 3(*c*) and (*d*) show the result using the approaches of Ackerman and Hao, respectively (Ackerman 1997; Team et al. 1997; Hao and Qu 2007). With Hao's method, brighter intensities are associated with a higher thermal dust index (TDI). These methods are well-established visualization methods and not classification methods; in other words, they do not automate the task of finding a dust storm. Here we use them to ease the visual location of the dust. Figures 3(*e*)–(*i*) depict dust storm probability using ML, PNN, FFNN, LP-SVR, and LS LP-SVR, respectively. Clearly, PNN, FFNN, and SVR approaches perform well from the visual assessment perspective, while the ML method shows conservative results.

Figure 4(*a*), however, shows a dust storm that occurred in a very different region and soil type. This dust storm event is from 5 June 2009, which covered a very large area in the Middle East. As before, RoI is shown in Figure 4(*j*). Figure 4(*b*) shows the result reported by the MODIS AOD product. Figures 4(*c*) and (*d*) show results using the approaches introduced by Ackerman and Hao, respectively, where brighter intensities are associated with

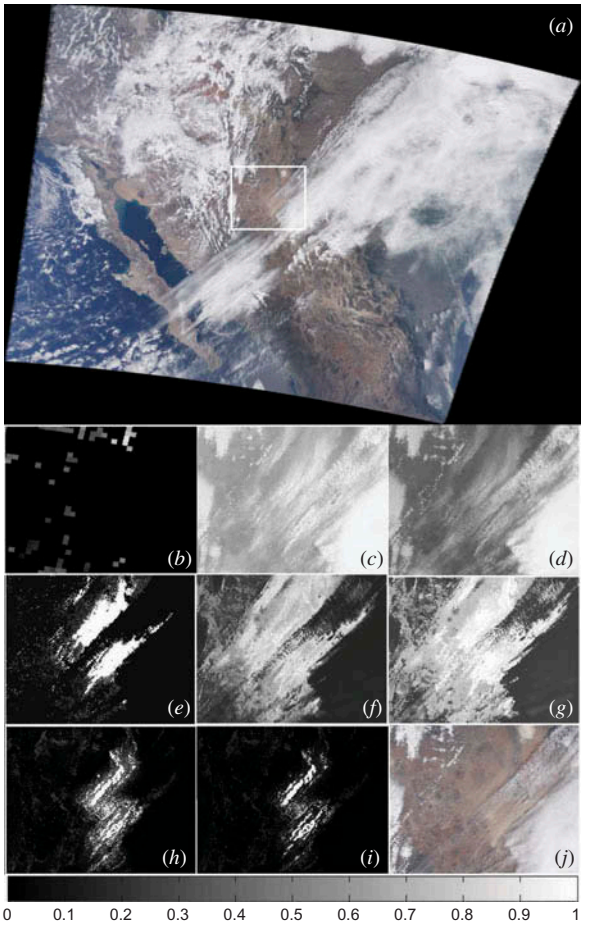

Figure 3. Dust storm over the southwestern USA, 10 April 2001, 18:05 UTC. Satellite: Terra. Instrument: MODIS. (*a*) True colour image. (*b*) MODIS AOD. (*c*) Result using Ackerman's method. (*d*) TDI according to Hao's method. (*e*)–(*i*) Dust storm probability using ML, PNN, FFNN, LP-SVR, and LS LP-SVR, respectively. (*j*) Region of interest.

a higher TDI. Dust storm probability using ML, PNN, FFNN, LP-SVR, and LS LP-SVR, respectively, is presented in Figures 4(*e*)–(*i*). Clearly, the classifier models have the capacity to detect different types of dust.

4.4. Generalization of dust detection over ocean

Dust storms can travel over hundreds of miles and in many instances propagate across oceans. It is possible for a satellite to pass over a region once the dust plume has moved from land to water. Since we are using thermal emissive information from multispectral data, there is an inherent change in dust reflectivity when measured in the ocean. Changes in reflectivity can have a direct impact on the statistical properties of the feature set and hence require training of a different set of classifiers to detect dust over ocean. This situation increases system complexity since now two sets of classifiers would be needed and the system will need an additional step to determine whether the background is land or ocean.

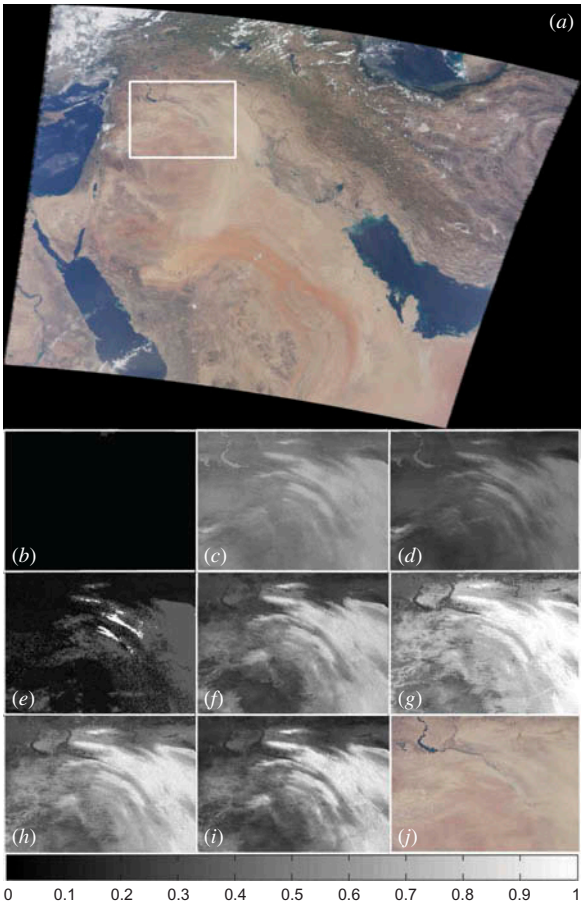

Figure 4. Dust storm over the Middle East, 5 June 2009, 07:50 UTC. Satellite: Terra. Instrument: MODIS. (*a*) True colour image. (*b*) MODIS AOD. (*c*) Result using Ackerman's method. (*d*) TDI according to Hao's method. (*e*)–(*i*) Dust storm probability using ML, PNN, FFNN, LP-SVR, and LS LP-SVR, respectively. (*j*) Region of interest.

Obviously, we would like to evaluate the generality of the classifier models in relation to this change in background class. We present two events of dust over ocean in Figures 5 and 6. We evaluate the classifiers trained with dust and land background (see previous subsection) directly with test data obtained from these events. The visualizations show that dust aerosols can be detected with the classifiers.

We note that despite their different geographical locations (Argentina and Australia), the two dust events were detected despite significant differences in dust composition. Furthermore, the methods can accurately discriminate similar signatures like smoke and clouds. It is evident that the methods provide a strong ability to infer non-trivial multispectral data relationships. In particular, PNN, due to its hybrid neuro-statistical nature, performed better than ML.

4.5. Testing a dust propagation period

The methods presented in this research can also be tested using MODIS data for consecutive days belonging to the same dust event. The aim of this analysis is to determine the

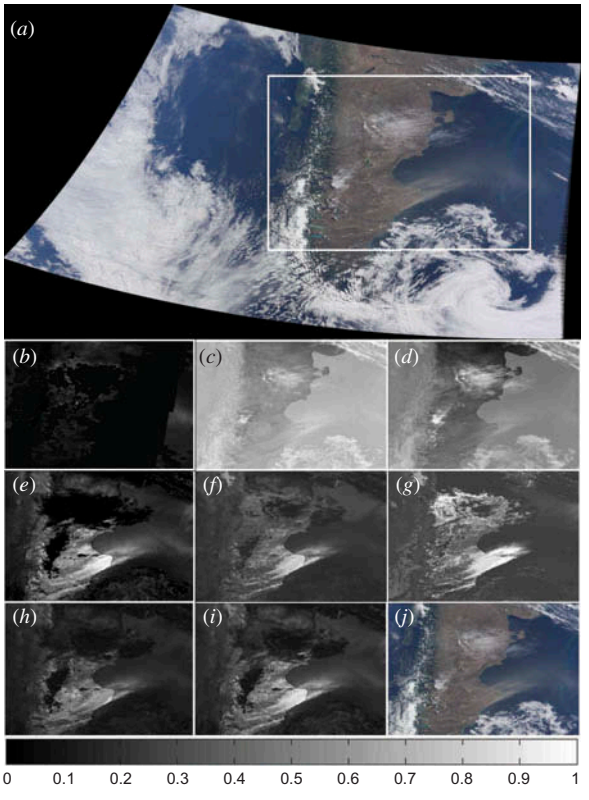

Figure 5. Dust storm over Argentina, 24 January 2010, 14:40 UTC. Satellite: Terra. Instrument: MODIS. (*a*) True colour image. (*b*) MODIS AOD. (*c*) Result using Ackerman's method. (*d*) TDI according to Hao's method. (*e*)–(*i*) Dust storm probability using ML, PNN, FFNN, LP-SVR, and LS LP-SVR, respectively. (*j*) Region of interest.

ability of classifiers to follow dust aerosols as they are transported across different geographical regions. The case study is a dust event within the USA. This event initiated on 18 October 2012 and was first captured by MODIS at 16:55 UTC near Nebraska, Colorado, and Kansas, as shown in Figures 7(*a*) and (*j*). The figure depicts the response with the same eight methods shown in previous figures. The pictures show the dust storm as it was observed in two different overpasses (i.e. 16:55–17:00 and 18:35 UTC). These occurred after the satellite had scanned the Earth with a complete turn. Since the dust storm was captured first by MODIS detectors near the western side of the instrument with respect to Earth, and then after a complete turn, it was captured with the detectors near the eastern side, this may cause a few issues. For instance, when aerosols, clouds, or other elements have moved, or reflectivity properties have changed with respect to the previously acquired data for that geographical region, this issue causes one group of pixels to look different to others (Figures 7(*a*) and (*j*)). One solution could be to preserve only the most recent data and ignore the previously acquired data that overlap the geographical region previously covered by MODIS sensors. However, we decided to preserve all the more recent data only when MODIS detector data overlap on a particular geolocation; otherwise, the data from a previous overpass are preserved. The intention of this is to test the method's ability to classify in the case of potentially changing thermal emissive data. Clearly, in the visual range,

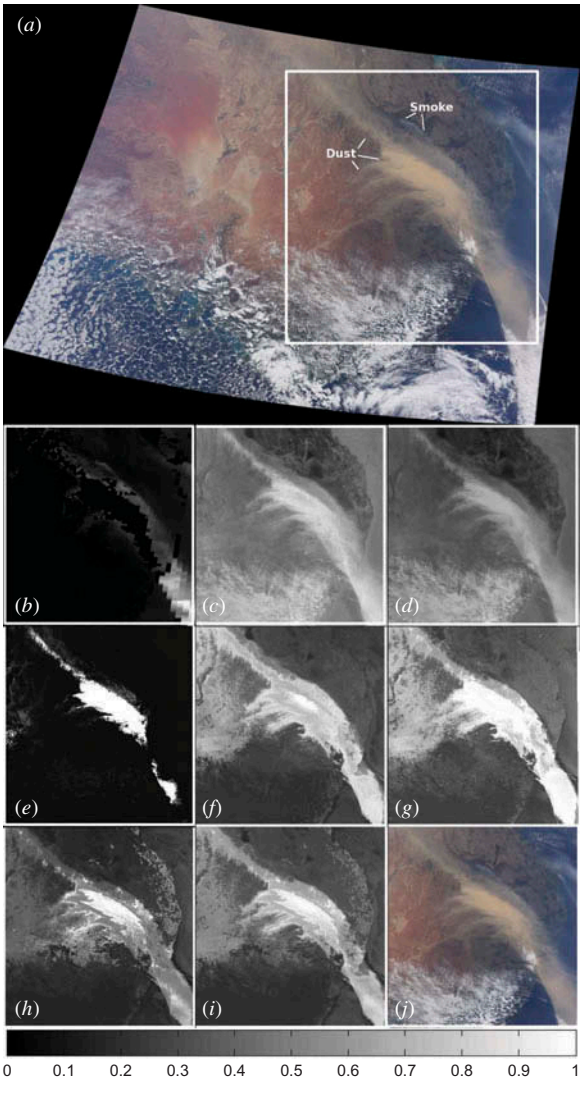

Figure 6. Dust storm over Australia, 26 September 2009, 00:35 UTC. Satellite: Aqua. Instrument: MODIS. (*a*) True colour image. (*b*) MODIS AOD. (*c*) Result using Ackerman's method. (*d*) TDI according to Hao's method. (*e*)–(*i*) Dust storm probability using ML, PNN, FFNN, LP-SVR, and LS LP-SVR, respectively. (*j*) Region of interest.

the picture will look strange because the most recent overpass (upper left diagonal side) will reflect sunlight differently (Figure 7(*j*)). Also, note that the AOD and the methods of Ackerman and Hao also suffer because of possible changes in thermal emissive properties of aerosols and land (see Figures 7(*b*)–(*d*), respectively). The classification methods we explore in this research seem to be robust to these changes, except for the ML method (see Figures 7(*e*)–(*i*)). However, there seems to be an inverted diagonal crossing the picture that presents several false positives. These false positives are associated with MODIS detectors that are scanning on far edges, at large angles, or inclinations. Nonetheless, FFNN,

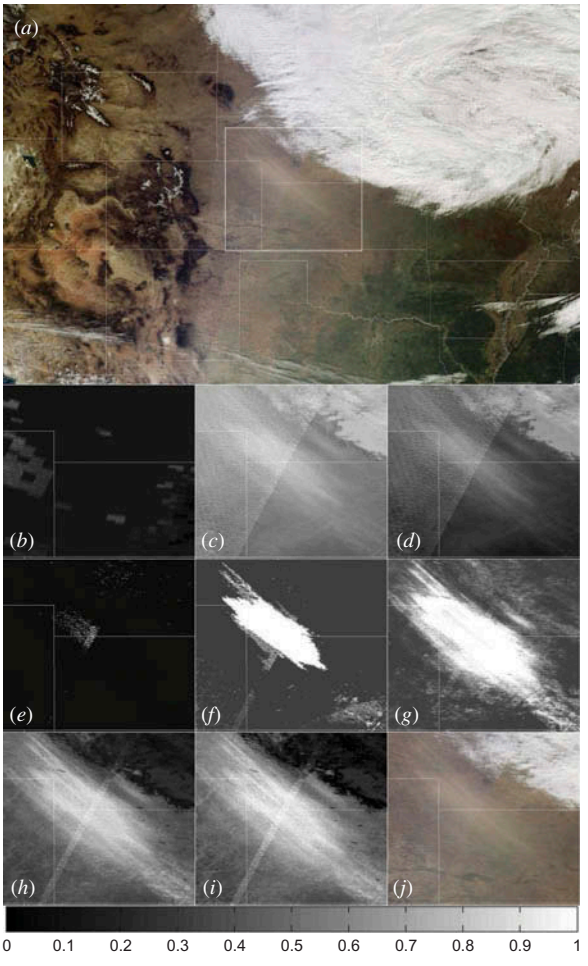

Figure 7. Dust event in central USA, 18 October 2012, 16:55, 17:00, and 18:35 UTC. Satellite: Terra. Instrument: MODIS. (*a*) True colour image. (*b*) MODIS AOD. (*c*) Result using Ackerman's method. (*d*) TDI according to Hao's method. (*e*)–(*i*) Dust storm probability using ML, PNN, FFNN, LP-SVR, and LS LP-SVR, respectively. (*j*) Region of interest.

LP-SVR, and LS LP-SVR seem to perform well, even displaying dust aerosols not trivially visible in the true colour picture and elucidating the dust plume's origin.

The dust event that started on 18 October 2012 produced large amounts of dust aerosols that were transported from the centre to the south and then to the eastern region of the USA, on the following day. Strong winds facilitated the transport of dust aerosols and even changed the trajectory that was originally thought to be southeast to northeast. This can be observed in Figures 8(*a*) and (*j*), which depict the same dust event but one day later. From these figures, we can conclude that dust aerosols were transported across Texas, Oklahoma, Arkansas, and Louisiana. At the time of MODIS overpass, on 19 October, the dust seems to have reached the states of Mississippi, Alabama, Georgia, Tennessee, and North Carolina. MODIS data correspond to 16:00–16:05, and 17:40 UTC. Figure 8(*b*) indicates that the AOD product provides significant data indicating high concentrations of aerosols. The methods of Ackerman and Hao, as well as the FFNN, provide a poor visualization of the

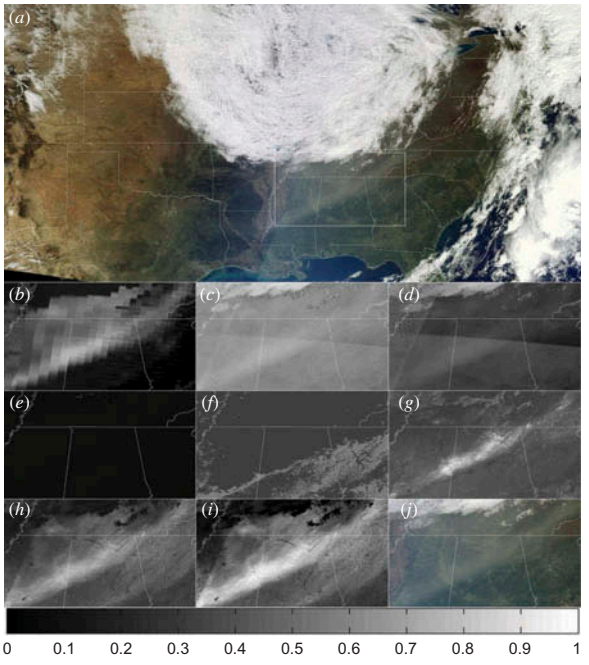

Figure 8. Dust event near the east coast of the USA, 19 October 2012, 16:00, 16:05, and 17:40 UTC. Satellite: Terra. Instrument: MODIS. (*a*) True colour image. (*b*) MODIS AOD. (*c*) Result using Ackerman's method. (*d*) TDI according to Hao's method. (*e*)–(*i*) Dust storm probability using ML, PNN, FFNN, LP-SVR, and LS LP-SVR, respectively. (*j*) Region of interest.

dust (see Figures 8(*c*), (*d*), and (*g*), respectively). The LP-SVR approach in either of its versions (classic or large scale) is able to provide a good detection of dust aerosols, as shown in Figures 8(*h*) and (*i*). In contrast, the PNN and the ML approaches fail to provide helpful information about the presence/absence of dust aerosols, especially the ML approach (see Figures 8(*e*) and (*f*)).

The same dust event observed on different, yet sequential, dates is typical of large dust storms such as the one presented here. Other typical big dust storms occur frequently in northwest Africa and in the Middle East. For such events, the LP-SVR approach seems to be a good alternative over other methodologies, as demonstrated here.

4.6. Segmentation

There is a one-to-one relationship between a feature vector and a pixel spatial location. This automatically converts the detection algorithms into an image segmentation method, and can be seen in the examples of the previous subsections. Besides the binary classification/segmentation, it is possible to provide more useful segmentation results for the analyst by taking advantage of the classifier characteristics. As we have seen in the previous subsection, the classifiers are capable of separating dust from background. Once each decision is mapped to its spatial location, a segmented image is automatically obtained, which can be used as a thematic map for further analysis.

Consider the ML, PNN, and FFNN classifiers. It is well known that these schemes aim to approximate the posterior class pdfs $f_{C|X}\,(cj|^{x})$ for $j = \{1, 2\}$. The two-class Bayes

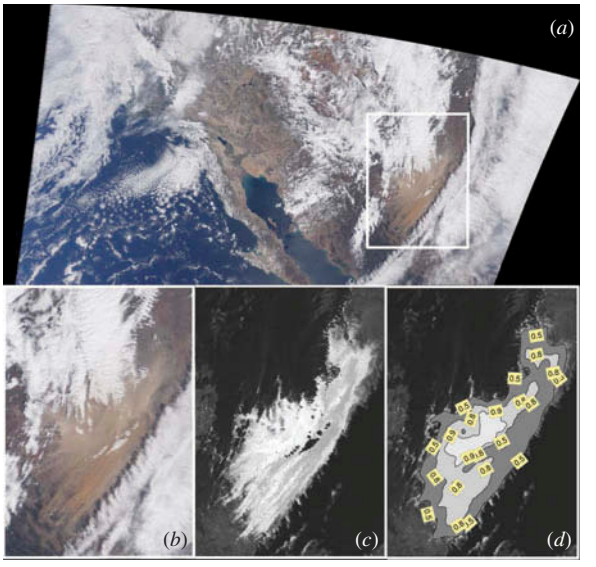

Figure 9. Dust storm over southwestern USA and northwestern Mexico, 6 June 2001, 18:30 UTC. Satellite: Terra. Instrument: MODIS. (*a*) True colour image. (*b*) Region of the dust storm. (*c*) Dust storm probability using PNN. (*d*) Classification/segmentation regions found using $t = 0.5$, $t = 0.8$, and $t = 0.9$.

classification rule can be also interpreted under the hypothesis testing framework, where ratio tests of class-conditional pdfs (i.e. likelihood ratios) have been used extensively. For instance, when there is no access to prior probability information, the Neyman–Pearson decision rule can be applied using a threshold γ value, which is selected to achieve a certain false alarm rate (Duda, Hart, and Stork 2001).

We can use a similar system where the posterior probability estimated for dust, $f_{C|X}(c_1|x)$ output by a classifier can be threshold to a preselected probability value. Threshold γ can be perceived as the assumption that a given pixel corresponds to dust. Then, the decision rule is set as

$$\begin{array}{lll} X & \text{is dust} & \text{if } f_{c_1|x}(C = j|X) > \gamma \\ X & \text{is background} & \text{otherwise,} \end{array} \tag{16}$$

where γ denotes the threshold in the interval $0 \leq \gamma \leq 1$, establishing the desired confidence level. As an example of segmentation let us consider the case of the dust storm of 6 April 2001, whose pseudo-colour image and region of interest are shown in Figures 9(*a*) and (*b*), respectively. The *posterior* probability of dust as approximated by the PNN approach is shown in Figure 9(*c*). The (superimposed) segmentations, using three thresholds, $\gamma = [0.5,\ 0.8,\ 0.9]$, are shown in Figure 9(*d*).

There are several implications in the use of γ for segmentation. First, this parameter can be easily controlled by an image analyst to gain understanding of the extent and distribution of the dust storm. From Figure 9(*d*), we can conclude that the posterior pdf maps into a spatial distribution of dust which shows a strong and compact concentration of dust over the region given by $\gamma = 0.9$, and decreases as the distance to the dust storm 'centroid' increases. The way the map is displayed is also very informative; for instance, the

concentric regions could be established according to some pre-defined intervals of γ values in order to establish the impact of a dust storm on asthma cases as it spreads over an urban area. Finally, we could envisage using this segmented output in an automated image analysis system, which could extract size, shape, direction, and other parameters of interest according to the remote-sensing application.

4.7. Impacts on human health

As recent studies have shown, there is a direct correlation between exposure to high levels of aerosols and increase in mortality rate from cardiovascular and respiratory illnesses and lung cancer (Rivera Rivera et al. 2009). Particularly, dust storms are an inadvertent cause of severe environmental and economic hazards. Therefore, our findings will help researchers in the further study of the dust storm phenomenon. The LP-SVR method, proposed here, will help in the understanding of dust storm genesis, formation, and propagation. This research, if implemented in near-real-time (NRT) systems, could be used to alert communities to the severity of dust events, and also has the potential of working under NASA's Land Atmosphere Near Real-time Capability for EOS (LANCE). Thus, if this research is implemented within the LANCE framework, it could be used to warn communities in a timely manner, especially if such communities lack high-resolution radar. Further research will include our findings in the area of implementing the proposed method in the LANCE framework. A prototype version is under test at the moment of writing this article, and further information about this can be found in the blog www.ismmalab.com, which details many other dust events and related information.

5. Conclusion

We compared four methods for dust storm detection: the ML classifier, the PNN, the FFNN, and the LP-SVR. ML and FFNN are classic mechanisms, while PNN is a hybrid probabilistic inference mechanism embedded in a neural architecture that in this case demonstrated good performance; we also explored an SV learning approach based on the solution to a linear program for regression (LP-SVR). Comparison among classifiers was made using different performance metrics, after the classifiers were modelled using a reduced training data set. This comparison also included a large-scale approach to LP-SVR (LS LP-SVR), which allowed us to use the complete (non-reduced) training data set. Numerical results showed that the reduced training data set contains significant data for modelling and performing dust detection based on very few feature vectors. More explicitly, we extracted features from MODIS thermal emissive spectral bands B20, B29, B31, and B32. Sample feature vectors were reduced following the concept introduced by Mather. The reduction involved the usage of KLT to preserve the statistical distribution of the features. The model's parameters were estimated from these reduced sample feature vectors, aiming to approximate true posterior class membership probabilities.

Among the neural network-based classifiers, the hybrid approach, PNN, performed better than the FFNN after a numerical evaluation of different performance metrics; however, neither was significantly better in the statistical sense. The LP-SVR approach was demonstrated to be better than the ML among statistical classifiers. Visually, all four methods performed accurate detection.

ML, PNN, FFNN, and LP-SVR were modelled using known cases from the southwestern USA and northwestern Mexico over land observations. Nevertheless, the

methods provided accurate detection when tested over land at different geographical regions with different soil types. Furthermore, all methods were able to classify dust storms over the ocean. Numerical and visual results suggest that the ML, PNN, FFNN, and LP-SVR classifiers are soil-independent, which in turn suggests that dust aerosol physical properties over land and ocean are very similar when analysed within the near-infrared spectral range.

Moreover, tracking low-probability dust aerosols may provide seminal dust transport information, i.e. dust storm impact and geographical extent can be addressed by means of low-probability dust detection analysis.

Furthermore, the four methods currently output 1 km spatial resolution results, which improves traditional AOD-based methods at 10 km spatial resolution.

In general, the four methods can be effectively utilized in the analysis of stratospheric dust, thereby helping researchers in the understanding of dust aerosol activity and transport. Among the four methods, however, an LP-SVR-based approach would be the best choice for the researcher.

Acknowledgements

PRP performed part of this work while at NASA Goddard Space Flight Center under the supervision of Dr James C. Tilton. This work was supported in part by the National Council for Science and Technology (CONACyT), Mexico, under grant 193324/303732, and by the University of Texas at El Paso. The partial support of the Secretaría de Educación Pública - Dirección General de Relaciones Internacionales (SEP-DGRI) is also acknowledged.

Finally, the authors acknowledge the support of the Large-Scale Multispectral Multidimensional Analysis (LSMMA) Laboratory (www.lsmmalab.com), and also credit and acknowledge that all MODIS data were acquired via the Earth Observing System Data and Information System (EOSDIS), 2009. Earth Observing System Clearing House (ECHO)/Warehouse Inventory Search Tool (WIST) Version 10.X [online application]. Greenbelt, MD: EOSDIS, Goddard Space Flight Center (GSFC) National Aeronautics and Space Administration (NASA). URL: https://wist.echo.nasa.gov/api/

Note

1. The complete list of data granules referred to in Section 2 can be found in Rivas-Perea (2009). These events correspond to NASA Terra MODIS instrument. The reference will show level 1B file names.

References

Ackerman, S. A. 1997. "Remote Sensing Aerosols Using Satellite Infrared Observations." *Journal of Geophysical Research* 102 (D14): 17069–17080.

Agarwal, A., H. M. El-Askary, T. El-Ghazawi, M. Kafatos, and J. Le-Moigne. 2007. "Hierarchical PCA Techniques for Fusing Spatial and Spectral Observations with Application to MISR and Monitoring Dust Storms." *IEEE Geoscience and Remote Sensing Letters* 4: 678–682.

Aksoy, S., K. Koperski, C. Tusk, G. Marchisio, and J. C. Tilton. 2005. "Learning Bayesian Classifiers for Scene Classification with a Visual Grammar." *IEEE Transactions on Geoscience and Remote Sensing* 43: 581–589.

Alkhatib, M. Q., and S. D. Cabrera. 2011. "Multiresolution and Directional Filtering Techniques for Detecting Dust Storm Direction in Satellite Imagery." In *Proceedings of SPIE* 804880480V. Bellingham, WA: SPIE.

Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*. New York: Oxford University Press.

Canty, M. J. 2007. *Image Analysis, Classification and Change Detection in Remote Sensing: With Algorithms for ENVI/IDL*. Boca Raton, FL: CRC Press.

Cawley, G. C. 2006. "Leave-One-Out Cross-Validation Based Model Selection Criteria for Weighted LS-SVMs." In *IEEE International Joint Conference on Neural Networks, 2006. IJCNN'06*, 1661–1668. Vancouver: IEEE.

Chacon-Murguía, M., Y. Quezada-Holguín, P. Rivas-Perea, and S. Cabrera. 2011. "Dust Storm Detection Using a Neural Network with Uncertainty and Ambiguity Output Analysis." *Pattern Recognition* 6718: 305–313.

Chen, J. P., M. L. Lin, and H. H. Lee. 2006. "Mineral Dust Emission Estimated from Satellite Data." In *Proceedings of SPIE, the International Society for Optical Engineering*, 629901. Orlando, FL: Society of Photo-Optical Instrumentation Engineers.

Chettri, S. R., and R. F. Cromp. 1993. "Probabilistic Neural Network Architecture for High-Speed Classification of Remotely Sensed Imagery." *Telematics and Informatics* 10: 187–198.

Congalton, R. G., and K. Green. 2008. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton, FL: CRC Press.

Courant, R., and D. Hilbert. 1966. *Methods of Mathematical Physics. Vol. II: Partial Differential Equations*. New York: Interscience.

Demšar, J. 2006. "Statistical Comparisons of Classifiers over Multiple Data Sets." *The Journal of Machine Learning Research* 7: 1–30.

Duda, R. O., P. E. Hart, and D. G. Stork. 2001. *Pattern Classification. Pattern Classification and Scene Analysis*. New York: Wiley.

El-Askary, H., M. Kafatos, X. Liu, and T. El-Ghazawi. 2004. "Introducing New Approaches for Dust Storms Detection Using Remote Sensing Technology." In *Proceedings of 2003 IEEE International Geoscience and Remote Sensing Symposium, 2003. IGARSS'03*, Vol. 4,2439–2441. Toulouse: IEEE.

Fanning, D. W. 2012. "Bright MODIS Images." Accessed June 2009. http://www.idlcoyote.com/ip_tips/brightmodis.html

Gail, W. B. 2007. "Remote Sensing in the Coming Decade: The Vision and the Reality." *Journal of Applied Remote Sensing* 1: 12505.

Gallant, A. R., and H. White. 1988. "There Exists a Neural Network that Does Not Make Avoidable Mistakes." In *IEEE International Conference on Neural Networks, 1988*, 657–664. San Diego, CA: IEEE.

García, S., and F. Herrera. 2008. "An Extension on Statistical Comparisons of Classifiers over Multiple Data Sets for All Pairwise Comparisons." *Journal of Machine Learning Research* 9: 2677–2694.

Hao, X., and J. J. Qu. 2007. "Saharan Dust Storm Detection Using Moderate Resolution Imaging Spectroradiometer Thermal Infrared Bands." *Journal of Applied Remote Sensing* 1: 013510.

Hillger, D. W., and G. P. Ellrod. 2003. "Detection of Important Atmospheric and Surface Features by Employing Principal Component Image Transformation of GOES Imagery." *Journal of Applied Meteorology* 42: 611–629.

Hoogerheide, L. F., J. F. Kaashoek, and H. K. Dijk. 2003. "Neural Network Approximations to Posterior Densities: An Analytical Approach." *Econometric Institute Report* EI 2003–38.

Janugani, S., V. Jayaram, S. D. Cabrera, J. G. Rosiles, T. E. Gill, and N. R. Rivera. 2009. "Directional Analysis and Filtering for Dust Storm Detection in NOAA-AVHRR Imagery." In *SPIE Defense, Security, and Sensing*, 73341G. San Diego, CA: International Society for Optics and Photonics.

Kanellopoulos, I., J. Austin, and G. G. Wilkinson. 1997. *Neurocomputation in Remote Sensing Data Analysis*. Berlin: Springer-Verlag.

Kast, J. L., P. H. Swain, B. J. Davis, and P. W. Spencer. 1977. *ECHO User's Guide*. LARS Publication, 083077, West Lafayette, IN: Purdue University.

Khazenie, N., and T. F. Lee. 1992. "Identification of Aerosol Features Such as Smoke and Dust in NOAA-AVHRR Data using Spatial Textures." In *Geoscience and Remote Sensing. IGARSS'92. International Symposium on*, 726–730. Helsinki: IEEE.

Kobler, B., and J. Berberi. 1991. "NASA Earth Observing System Data Information System (EOSDIS)." In *Eleventh IEEE Symposium on Mass Storage Systems, 1991. Digest of Papers*, 18–19. Monterey, CA: IEEE.

Kopp, G., and J. L. Lean. 2011. "A New, Lower Value of Total Solar Irradiance: Evidence and Climate Significance." *Geophysical Research Letters* 38: L01706.

Lee, J. A., T. E. Gill, K. R. Mulligan, M. Domínguez Acosta, and A. E. Perez. 2009. "Land Use/Land Cover and Point Sources of the 15 December 2003 Dust Storm in Southwestern North America." *Geomorphology* 105: 18–27.

Levy, R. C., L. A. Remer, Y. J. Kaufman, D. Tanré, S. Mattoo, E. Vermote, and O. Dubovik. 2006. *Revised Algorithm Theoretical Basis Document: MODIS Aerosol Products MOD/MYD04*. Greenbelt, MD: NASA Goddard Space Flight Center.

Loomer, S. A. 2006. "Academic Research Opportunities at the National Geospatial-Intelligence Agency (NGA)." In *Proceedings of SPIE* 6233. Orlando, FL: SPIE.

Mahler, A. B., K. Thome, D. Yin, and W. A. Sprigg. 2006. "Dust Transport Model Validation Using Satellite-and Ground-Based Methods in the Southwestern United States." *Remote Sensing of Aerosol and Chemical Gases, Model Simulation/Assimilation, and Applications to Air Quality, Proc. SPIE* 6299: 1–11.

Mather, P. M., B. Tso, and M. Koch. 1997. "Geological Mapping Using Multi-Sensor Data: A Comparison of Methods." In *Neurocomputation in Remote Sensing Data Analysis: Proceedings of Concerted Action COMPARES* (*Connectionist for Pre-processing and Analysis of Remote Sensing Data*), 38–46. Berlin: Springer-Verlag.

Mercer, J. 1909. "Functions of Positive and Negative Type, and Their Connection with the Theory of Integral Equations." *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 209: 415–446.

Miller, S. D. 2003. "A Consolidated Technique for Enhancing Desert Dust Storms with MODIS." *Geophysical Research Letters* 30: 2071.

Novlan, D. J., M. Hardiman, and T. E. Gill. 2007. "A Synoptic Climatology of Blowing Dust Events in El Paso, Texas from 1932–2005." In *Preprints, 16th Conference on Applied Climatology, American Meteorological Society*, Vol. 3. San Antonio, TX: AMS.

Prospero, J. M., P. Ginoux, O. Torres, S. E. Nicholson, and T. E. Gill. 2002. "Environmental Characterization of Global Sources of Atmospheric Soil Dust Identified with the Nimbus 7 Total Ozone Mapping Spectrometer (TOMS) Absorbing Aerosol Product." *Reviews of Geophysics* 40: 1002.

Ramakrishnan, S., and S. Selvan. 2007. "Image Texture Classification Using Wavelet Based Curve Fitting and Probabilistic Neural Network." *International Journal of Imaging Systems and Technology* 17: 266–275.

Rivas-Perea, P. 2009. "Southwestern U.S. and Northwestern Mexico Dust Storm Modeling Trough Moderate Resolution Imaging Spectroradiometer Data: A Machine Learning Perspective." *Technical Report: NASA/UMBC/GEST Graduate Student Summer Program 2009*. http://gest.umbc.edu/student_opp/2009_gssp_reports.html

Rivas Perea, P. 2011. "Algorithms for Training Large-Scale Linear Programming Support Vector Regression and Classification." PhD thesis, The University of Texas at El Paso.

Rivas-Perea, P., and J. Cota-Ruiz. 2012. "An Algorithm for Training a Large Scale Support Vector Machine for Regression Based on Linear Programming and Decomposition Methods." *Pattern Recognition Letters* 34: 439–451. doi:10.1016/j.patrec.2012.10.026

Rivas-Perea, P., and J. G. Rosiles. 2010. "A Probabilistic Model for Stratospheric Soil-Independent Dust Aerosol Detection." In *Digital Image Processing and Analysis*. Tucson, AZ: Optical Society of America.

Rivas-Perea, P., J. Rosiles, M. Murguia, and J. Tilton. 2010. "Automatic Dust Storm Detection Based on Supervised Classification of Multispectral Data." *Soft Computing for Recognition Based on Biometrics* 312: 443–454.

Rivera Rivera, N. I. 2006. "Detection and Characterization of Dust Source Areas in the Chi-Huahuan Desert, Southwestern North America." Master's thesis, The University of Texas at El Paso.

Rivera Rivera, N. I., T. E. Gill, K. A. Gebhart, J. L. Hand, R. M. Fitzgerald, and D. J. Novlan. 2009. "Analysis of Air Transport Patterns Bringing Dust Storms to El Paso, Texas." In *Fourth Symposium on Policy and Socio-Economic Research, AMS Annual Meeting*, Phoenix, No. JP2.6.

San-Chao, L., L. Qinhuo, G. Maofang, and C. Liangfu. 2006. "Detection of Dust Storms by Using Daytime and Nighttime Multi-spectral MODIS Images." In *IEEE International Conference on Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006*, 294–296. Denver, CO: IEEE.

Sorek-Hamer, M., A. Cohen, R. C. Levy, B. Ziv, and D. M. Broday. 2013. "Classification of Dust Days by Satellite Remotely Sensed Aerosol Products." *International Journal of Remote Sensing* 34: 2672–2688.

Specht, D. F. 2002. "Probabilistic Neural Networks for Classification, Mapping, or Associative Memory." In *IEEE International Conference on Neural Networks, 1988*, 525–532. San Diego, CA: IEEE.

Team, M. C. M., S. Ackerman, K. Strabala, P. Menzel, R. Frey, C. Moeller, L. Gumley, B. Baum, C. Schaaf, and G. Riggs. 1997. "Discriminating Clear-Sky from Cloud with MODIS Algorithm Theoretical Basis Document: *MOD35*." In *ATBD Ref. ATBD-MOD-06, version 4.0*, 115p. Greenbelt, MD, NASA Goddard Space Flight Center.

Tso, B., and P. M. Mather. 2009. *Classification Methods for Remotely Sensed Data*. Boca Raton, FL: CRC Press.

Yang, Z. R., and S. Chen. 1998. "Robust Maximum Likelihood Training of Heteroscedastic Probabilistic Neural Networks." *Neural Networks* 11: 739–747.

Yang, Z. R., M. Zwolinski, and C. D. Chalk. 1998. "Fault Detection and Classification in Analogue Integrated Circuits Using Robust Heteroscedastic Probabilistic Neural Networks." In *Mixed Signal Testing Workshop, The 4th IEEE International*, The Hague: IEEE.