

A Classic and Neural Probabilistic Approach to Remote Sensing: The Dust Storm Detection Problem

Pablo Rivas-Perea¹, Jose G. Rosiles¹,
Mario I. Chacon Murguia² and James J. Tilton³

¹ The University of Texas El Paso,
Department of Electrical and Computer Engineering, El Paso TX 79968, USA

² Chihuahua Institute of Technology,
Graduate Studies Department, Chihuahua Chih, Mexico

³ NASA Goddard Space Flight Center,
Computational and Information Sciences and Technology Office,
Greenbelt MD 20771, USA

Abstract. Dust storms are a natural severe weather condition. A recent study in 2009 found correlation between lung diseases and dust storm events. Since then, more research has been done for dust air-borne suspended particle (aerosol) analysis. However, there is paucity of formal methods in machine learning. Particularly, we study a classic and hybrid neural probabilistic approach to alleviate the lack of specialized classification methods. The Maximum Likelihood Estimator (MLE) and the hybrid Probabilistic Neural Network (PNN) approaches are discussed. The features utilized are Moderate Resolution Imaging Spectroradiometer (MODIS) thermal emissive spectral bands. We utilized four near infrared bands: B20 (3.660 – 3.840 μm), B29 (8.400 – 8.700 μm), B31 (10.780 – 11.280 μm), and B32 (11.770 – 12.270 μm). Numerical performance evaluation show that the hybrid approach (PNN) performed better than the classic (MLE). Visually, both accurately detect dust storms. The classifiers demonstrated a strong ability to find non-trivial relationships within the spectral bands. Both methods demonstrated to be soil-independent and surface-invariant detection methods. The proposed methods can be effectively utilized in understanding dust storm phenomena.

1 Introduction

During the past 20 years, society has been experiencing an increase in severe weather conditions: strong winds, storms, hurricanes, etc. These conditions force human beings to adapt their way of life. Research has been done to understand human adaptation implications. In desert regions, dust storms occur as a natural part of a climate system. In other regions, dust storms are a rapidly increasing phenomenon, suggesting a modification to known climate models. Recent findings in dust storm analysis show a correlation between lung diseases and

dust storms [1]. Ackerman demonstrated the feasibility of dust storm analysis using satellite data [4]. Later, geophysicists [2] and geoscientists [3] introduced formal methodologies for dust storm visualization and enhancement. Indeed, NASA Terra and Aqua satellite data can be processed producing the Aerosol Optical Density (AOD) product utilized for air-borne suspended particle (aerosol) identification. Although much work has been done to date, most of the work is limited in spatial resolution. Moreover, there is paucity of formal machine learning detection methods. The need for computational intelligence specialized approaches is evident. The purpose of this work is to increase specialized machine learning methods, and to provide high resolution models. We modeled and compared classic and hybrid neural-based probabilistic detection methods: the Maximum Likelihood Estimation (MLE) [5] and the Probabilistic Neural Network [10]. The output of both provide a 1 km spatial resolution detection. It is hoped that the information presented here may be useful in identifying opportunity areas to those who are responsible for rapid response teams or aerosol identification research. Section 2 of the paper introduces the analysis for multi-spectral band selection. The MLE and PNN models are explained in Section 3 and 4, respectively. Section 5 presents experimental results followed by brief results and findings discussion. Finally, conclusions are drawn in Section 6.

2 Spectral Bands Selection

The MODIS instrument is built in NASA’s Terra and Aqua satellites. MODIS provides the Earth’s information in 36 spectral bands. It eases the analysis of different phenomena such as atmospheric and sea temperature, surface reflectivity, etc. All these data make possible the analysis of dust storms. These 36 bands are available in MODIS Level 1B file organization. Dust storm visual assessment can be achieved using MODIS bands $B1$, $B3$, and $B4$ which correspond to human visual perception. An RGB composite true color image can be produced by mapping: $R = B1$, $G = B4$, and $B = B3$. Figure 1 shows an RGB composite true color image of a dust storm off West Africa. Besides bands $B1$, $B3$ and $B4$, Hao *et al.*[3] demonstrated that bands $B20$, $B29$, $B31$, and $B32$ can also be utilized for dust storm visualization. Ackerman *et al.*[4] demonstrated that the usage of bands $B32$ and $B31$ improve dust storm visualization contrast. Based on these findings, we designed a classification scheme using the following thermal emissive bands: $B20$ ($3.660 - 3.840\mu m$), $B29$ ($8.400 - 8.700\mu m$), $B31$ ($10.780 - 11.280\mu m$), and $B32$ ($11.770 - 12.270\mu m$).

2.1 Radiance Recovery and Feature Vector

A ”recovered” radiance is a 16 bit MODIS thermal emissive band recovered to its original scale ($W/m^2/\mu m/sr$). The recovery process is given by

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

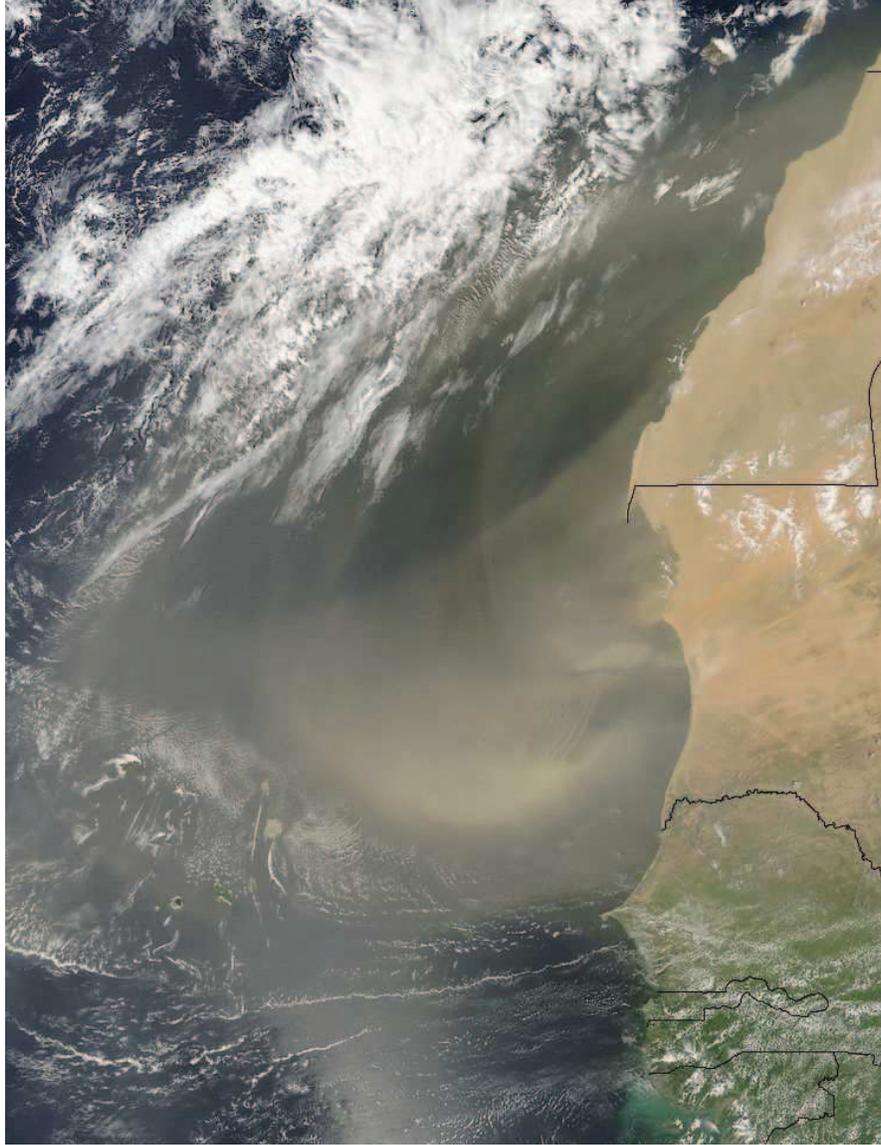


Fig. 1. True color RGB composite of dust storm off West Africa. Date 09/21/2009. Time 14:50 UTC. Satellite: Aqua. Courtesy of NASA MODIS Rapid Response Team.

where L denotes the recovered radiances, κ are the radiance scales, η are the radiance offsets, and ι are the scaled intensities (raw data). Our feature vector $F_j \in \mathbb{R}^{n \times 4}$ consists of the following recovered radiances

$$F = [L^{B20}, L^{B29}, L^{B31}, L^{B32}], \quad (2)$$

where F_j is a set of features associated to the j -th class, and the superscript of L refers to the spectral band.

2.2 Data Set Collection

In our experiments, we selected 31 different events corresponding to the south-western US and north-western Mexico area [9]. Only 8 were used for validation. The data samples were downloaded using NASA’s WIST online tool. The selection of modeling (training) and validation set was performed randomly. The complete data set provides approximately 75 million feature vectors.

3 Probabilistic Modeling

Let X be a random associated to F in (2). Let $f_C(C = j)$ be the probability mass function (PMF) associated to the j -th class. Let $f_{X|C}(X = x|C = j)$ be the conditional probability density function of the spectral band feature vector X having a value x , given the probability that the j -th class occur. This might be referred as the “data likelihood” function. The Maximum Likelihood Classifier (Maximum Likelihood Estimator, MLE) assumes unknown posterior probabilities. The MLE can be derived from Bayes theorem. It is an accepted method in remotely sensed data classification and analysis [5]. Therefore, we modeled the MLE classifier for the dust storm classification. Our models consider four classes: $C0$ = land/sea background, SM = smoke, $BLDU$ = blowing dust, and DS = dust storm. The DS class is of primary interest. From now on, PMF’s will be expressed as probability density functions (PDF’s). To obtain the maximum likelihood between the observed data and the prior probability for all classes, we can state the following decision rule:

$$\begin{aligned} x \in C = j \quad & \text{if,} \dots \\ & f_{X|C}(X|C = j)f_C(C = j) > \dots \\ & f_{X|C}(X|C = i)f_C(C = i), \end{aligned} \quad (3)$$

for all $j \neq i$, and appearing already simplified by removing the common factor $f_X(X = x)$. Then, assuming that the prior probability is normally distributed, we can rewrite the above terms by defining a function

$$\psi_k(x) = f_{X|C}(X = x|C = k)f_C(C = k), \quad (4)$$

that allows restating the decision rule as

$$x \in C = j \quad \text{if} \quad \psi_j(x) > \psi_i(x) \quad \forall j \neq i, \quad (5)$$

which is more convenient to handle. In MLE, the functions $\psi_k(x)$ are commonly referred to as *discriminant functions*. The function $f_C(C = k)$ is removed from (4) due to the high uncertainty about the true $f_C(C = k)$ PMF. Then, assuming that the discriminant functions follow a Gaussian distribution, $\psi_k(x)$ can be reduced removing the factor $-\frac{d}{2} \ln(2\pi)$, which adds no discriminant information to classification [6]. Also the common factor, $\frac{1}{2}$, can be removed. This leads to a commonly used simpler discriminant function:

$$\begin{aligned} \psi_k(x) = & -\det(\Sigma_{X|C}) - \dots \\ & (x - \mu_{X|C})^T \Sigma_{X|C}^{-1} (x - \mu_{X|C}) \end{aligned} \quad (6)$$

where $\det(\cdot)$ is the determinant function, $\Sigma_{X|C}$ the likelihood PDF's covariance matrix, $\mu_{X|C}$ denotes the mean vector, $(\cdot)^T$ denotes the transpose operation, and $\Sigma_{X|C}^{-1}$ the inverse covariance matrix.

3.1 Conditional PDF $f_{X|C}$ Parameter Estimation

Considering Gaussianity, the goal is to estimate the following parameters: the expected values vector $\hat{\mu}_{F|C}$ and the covariance matrix $\hat{\Sigma}_{F|C}$. We segmented the images based on already published remote sensing papers. The segmentation groups are dust storm, blowing dust, smoke, and background. These groups are associated with a particular class C . Then, we extracted and stored the subset of pixels associated to each class. Finally, we computed the sample mean $\hat{\mu}_{F|C}$ and the covariance matrix $\hat{\Sigma}_{F|C}$. The covariance matrix $\hat{\Sigma}_{F|C}$ is not ill-posed since we have many data samples available.

4 Hybrid Neuro-Probabilistic Modeling

Specht's Probabilistic Neural Network (PNN) is a semi-supervised neural network [10]. It is widely used in pattern recognition applications. The PNN is inspired by Bayesian classification and does not require training. It estimates feature vector PDF's 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 $F \in \mathfrak{R}^n$. The second layer contains exponential functions $\varphi(\cdot)$ in each node, and the node's count corresponds to the k samples count for the j -th class. These nodes are called pattern units ν_{jk}^F and are fully connected to the input nodes. The pattern layer's output is denoted by

$$\varphi_{jk}(F) = \frac{1}{(2\pi)^{\frac{d}{2}} \sigma^d} e^{-\frac{1}{2\sigma^2} (F - \nu_{jk}^F)^T (F - \nu_{jk}^F)}. \quad (7)$$

The third layer contains summation units to complete the probability estimation. There are as many summation units as classes. The j -th summation unit denoted as ϖ_j , receives input only from those pattern units belonging to the j -th class. This layer computes the likelihood of F being classified as C , averaging and summarizing neurons output belonging to the same class:

$$\begin{aligned} \Omega_j(\varphi_{jk}(F)) &= \frac{1}{(2\pi)^{\frac{d}{2}} \sigma^d} \frac{1}{N_j} \times \dots \\ &\sum_{i=1}^{N_j} e^{-\frac{1}{2\sigma^2} (\varphi_{ik}(F) - \varpi_i)^T (\varphi_{ik}(F) - \varpi_i)}. \end{aligned} \quad (8)$$

The last layer is the decision layer. It classifies F 's pattern according to the Bayesian decision rule given by

$$F \in C_j \text{ if, } C_j(\Omega_j(\varphi_{jk}(F))) = \max_{1 \leq i \leq j} \Omega_i(\varphi_{ik}(F)). \quad (9)$$

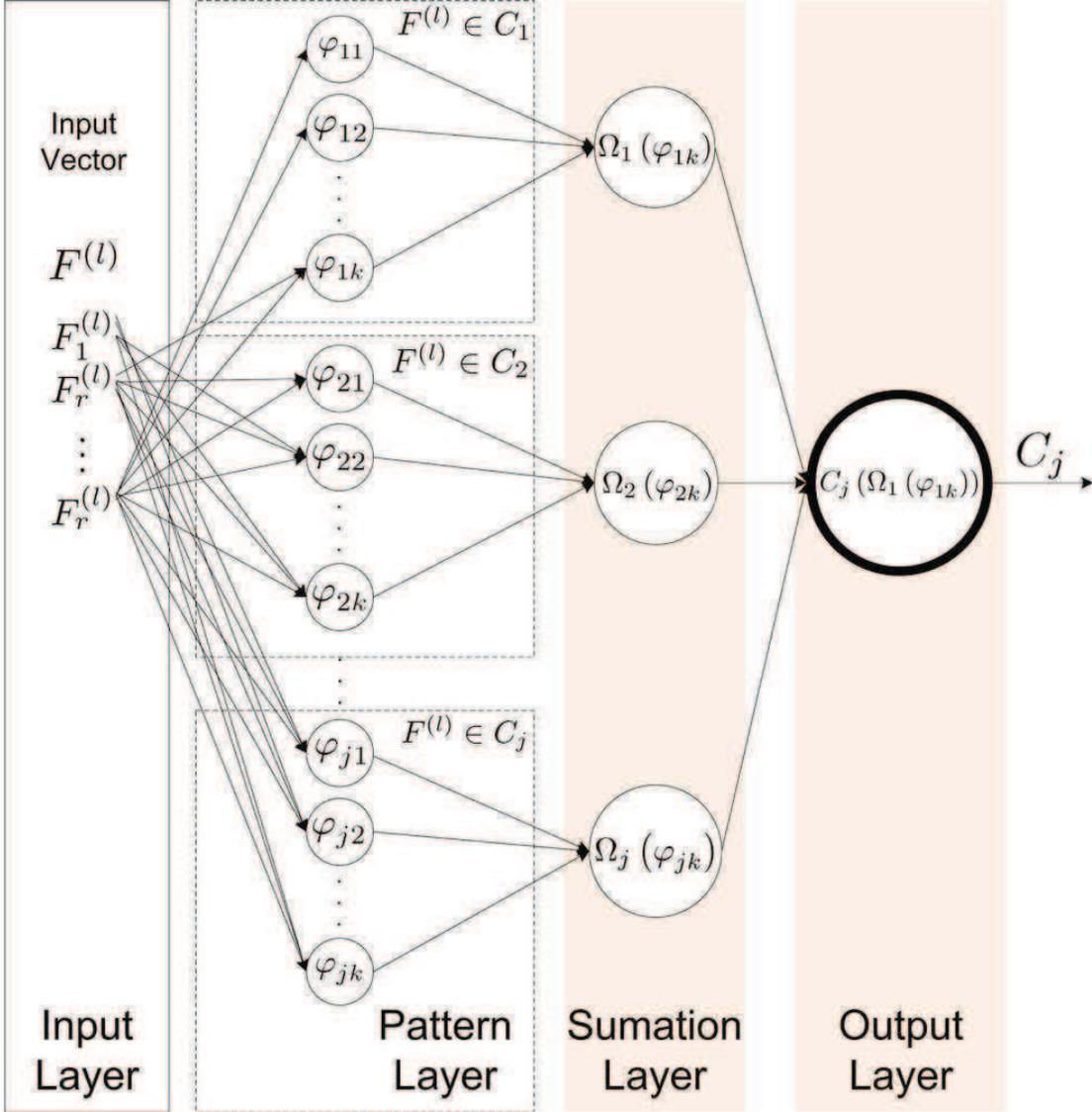


Fig. 2. The hybrid architecture of the Probabilistic Neural Network. Note the probabilistic nature embedded in a neural architecture.

Thus, the maximum of the summation nodes output characterize the PNN's general output. The function $\Omega_j(\cdot)$ gives the probability of the j -th class. This allows us to generate probabilistic visualizations of our interest class: DS .

4.1 Spread Parameter Estimation

The parameter σ is estimated with Srinivasan's *et al.* method [7]. It requires a pre-normalization phase consisting on two steps: subtract the mean μ_F from the training feature vector F , and divide F by its standard deviation σ_F . Then, σ_F is the absolute difference between the two smallest normalized variances. This completes the PNN modeling since there is no need for a training phase.

4.2 PNN’s Large Sample Size Problem

To avoid the overwhelming processing of millions of training samples, we limited the sample number. We based our reduction method on Kanellopoulos criteria [8], and the Karhunen Loeve Transformation (KLT). Kanellopoulos criteria establishes that the training samples number must be at least three times the feature bands number. Therefore, in our PNN design we use at least three times the F feature vector size. However, to avoid shortcomings and given the data availability, we used the Kanellopoulos criteria twice. Thus, for each class we have a feature vector $F_j \in \mathfrak{R}^{24 \times 4}$, for all j . The feature vectors’ reduction was performed with KLT, also known as Principal Component Analysis (PCA). The KLT is typically used to reduce problem dimensionality. It projects the data to an uncorrelated data space. It is typically used to reduce the number of features elements. However, we used it to reduce the sample number, not the feature elements. We did not project the data to a different subspace. Instead, we identified the sample elements associated with the 24 largest covariance matrix eigenvalues. Then, we used those feature vectors to model the PNN, following the previously described model.

5 Experimental Results and Discussion

We selected three performance metrics to evaluate the two proposed classification methods. These metrics are

$$\text{Precision} = \frac{\sum TP}{\sum TP + FP}, \quad (10)$$

$$\text{Accuracy} = \frac{\sum TP + TN}{\sum TP + FN + FP + TN}, \quad (11)$$

as well as 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. In 10 and 11, TP stands for “True Positive,” FP “False Positive,” TN “True Negative,” and FN “False Negative.”

In our experiments we used approximately 56 million training feature vectors (23 cases). The remaining 19 million (8 cases) were used for validation. These feature vector are recovered radiances using 1 to produce F with 2. Due to the large sample size problem, we used the KLT approach to keep a vector $F_j \in \mathfrak{R}^{24 \times 4}$. Since we have four classes, the total size is $F \in \mathfrak{R}^{96 \times 4}$. Then, the MLE and PNN parameters were estimated and modeled from the resulting F using the training set. The performance was calculated using only the validation set. The numerical results were concentrated and averaged to produce Table 1, showing that the hybrid neuro-probabilistic approach is better than the MLE. Table 1 also include the processing time per feature vector in milliseconds.

Numerically, the hybrid approach performed better than the classic approach. In terms of the visual assessment of dust storm detection, visually, the methods

Table 1. Classifiers Performance. The hybrid method performs better than the traditional approach. The processing time shown here is in milli-seconds.

	Precision	Accuracy	AUC	P. Time
MLE	0.5255	0.6779	0.4884	0.0141
PNN	0.7664	0.8412	0.6293	0.2393

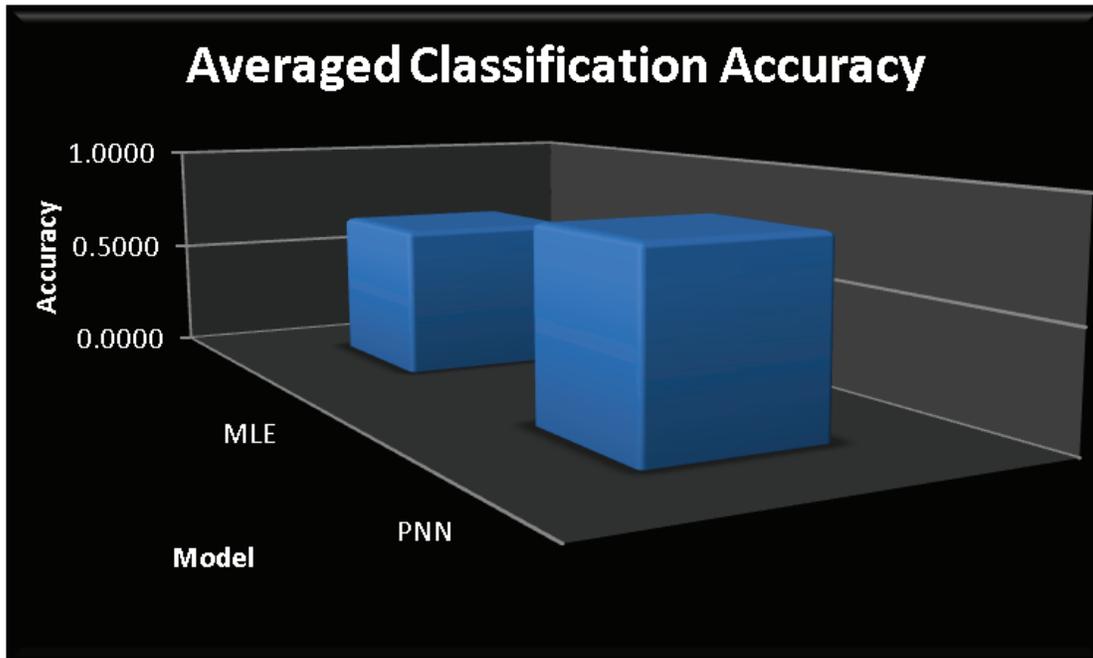


Fig. 3. Accuracy bar chart for MLE and PNN.

must show a higher probability in the region where the dust aerosol is present. Considering, for example, the dust storm of April 6, 2001, covering Chihuahua, Texas, and New Mexico, classification results are illustrated in Figures 5-9. Figure 5 shows the true color image. Figures 6-7 show the dust storm probability. Figures 8-9 show segmented images, following the mapping: *DS* to red, *BLDU* to green, *SM* to blue, and *C0* to black. Clearly, both methods perform well from the visual assessment perspective. Other examples in the the southwestern USA and northwestern Mexico area are shown in Figures 10-15.

Both methods performed well in training and validation cases. But, what about other events outside the southwestern US? Can the methods detect dust storms over the ocean, even if they were not trained for that? The unfortunate dust storm in West Africa shown in Figure 1 at the paper's beginning originated over land, and extended thousands of miles over the ocean. The complete satellite scan true color image is shown in Figure 16. The dust storm's probability for MLE and PNN is shown in Figures 17 and 18 respectively. Similarly, we present the detection of a dust storm in Australia on October 12, 2009, and also in southern Asia on February 8, 2010, in Figures 19-21 and Figures 22-24

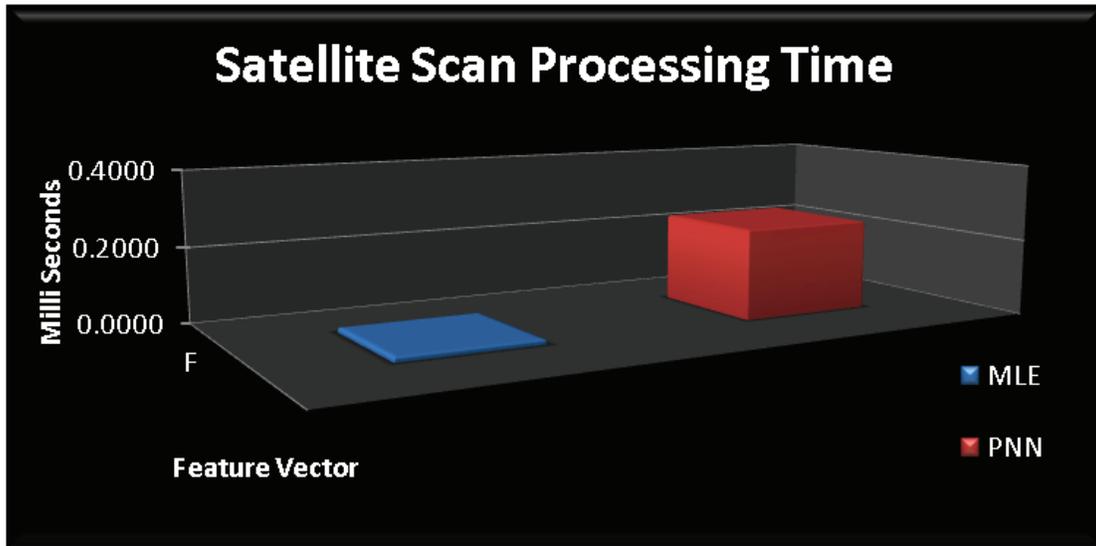


Fig. 4. Processing time bar chart for MLE and PNN. Results are in milli-seconds.

respectively. Both methods detected the dust storm over land and ocean. It is evident that both methods provide a strong ability to infer non trivial multi-spectral data relationships. PNN, due to its hybrid neural-probabilistic nature, performed better than MLE. On the other hand, MLE is quite conservative. For some applications this still is a good thing.

6 Conclusion

Two formal methods for classification were studied in the dust storm detection problem: Maximum Likelihood Estimator (MLE) and the Probabilistic Neural Network (PNN). MLE is a classic, but PNN is a hybrid probabilistic inference mechanism embedded in a neural architecture. We analyze MODIS thermal emissive spectral bands to select appropriate feature vectors. Sample feature vectors were reduced using a known criteria in remotely sensed data. From these reduced sample feature vectors, the models and parameters were estimated. After numerical performance evaluation, the hybrid PNN approach performed better than MLE. Visually, both methods perform accurate detection and classification. Both PNN and MLE were modeled using known cases from the south-western US and north-western Mexico over land observations. However, when tested over land in different countries with different types of soils and, thus, different type of dust, both methods provided accurate detection. Furthermore, both were able to classify over-ocean dust storms. Findings suggest that both MLE and PNN are soil-independent and surface-independent classifiers. Moreover, important information about dust transport can be obtained by tracking low probability dust storm detections. Both methods perform 1 km spatial resolution classification, which improve traditional Aerosol Optical Density (AOD)-based methods at 10 km spatial resolution. The presented

methods can be utilized in the analysis of stratospheric dust, thereby, helping researchers in the understanding of dust aerosol activity and transport.

Acknowledgments

The author P.R.P performed the work while at NASA Goddard Space Flight Center under the Graduate Student Summer Program (GSSP) and mentored by Dr. James C. Tilton. This work was partially supported by the National Council for Science and Technology (CONACyT), Mexico, under grant 193324/303732, as well as by the University of Texas El Paso Graduate School Cotton Memorial Funding. The author M.I.C.M. wants to thank DGEST for the support in the dust storm detection project.

References

1. Rivera Rivera, N.I., Gebhart, K.A., Gill, T.E., Hand, J.L., Novlan, D.J., and Fitzgerald, R.M., 2009. Analysis of air transport patterns bringing dust storms to El Paso, Texas. Symposium on Urban High Impact Weather, AMS Annual Meeting, January 2009, Phoenix., JP2.6.
2. Khazenie, N.; Lee, T.F., "Identification Of Aerosol Features Such As Smoke And Dust, In NOAA-AVHRR Data Using Spatial Textures," Geosc. and Rem. Sens. Symp., pp.726-730, 26-29 May 1992.
3. Xianjun Hao, John J. Qu, 2007, Saharan dust storm detection using MODIS thermal infrared bands. Journal of Applied Remote Sensing, Vol. 1, pp. 013510.
4. Ackerman, S.A., Strabala, K.I., Menzel, W.P., Frey, R., Moeller, C., Gumley, L., Baum, B.A., Seeman, S.W., and Zhang, H.. Discriminating clear-sky from cloud with MODIS algorithm theoretical basis document (MOD35). ATBD-MOD-06, version 4.0, NASA Goddard Space Flight Center.
5. Aksoy, S.; Koperski, K.; Tusk, C.; Marchisio, G.; Tilton, J.C., "Learning bayesian classifiers for scene classification with a visual grammar," Geosc. and Rem. Sens., IEEE Trans. on, vol.43, no.3, pp. 581-589, 2005.
6. John A. Richards and Xiuping Jia, Remote Sensing Digital Image Analysis, Springer, 2006.
7. Ramakrishnan, S. and Selvan, S. Image texture classification using wavelet based curve fitting and probabilistic neural network. Int. J. Imaging Syst. Technol. 17, 4 pp. 266-275. 2007.
8. Kanellopoulos, G. G. Wilkinson, J. Austin, Neurocomputation in Remote Sensing Data. Analysis, Springer-Verlag, 1997.
9. D. J. Novlan, M. Hardiman, and T. E. Gill, "A synoptic climatology of blowing dust events in El Paso, Texas," 16th Conference on Applied Climatology, AMS, J3.12, pp. 13
10. D. F. Specht, "Probabilistic neural networks for classification, mapping, or associative memory," IEEE international conference on neural networks, vol. 1, num. 7, pp. 525-532, 1988.

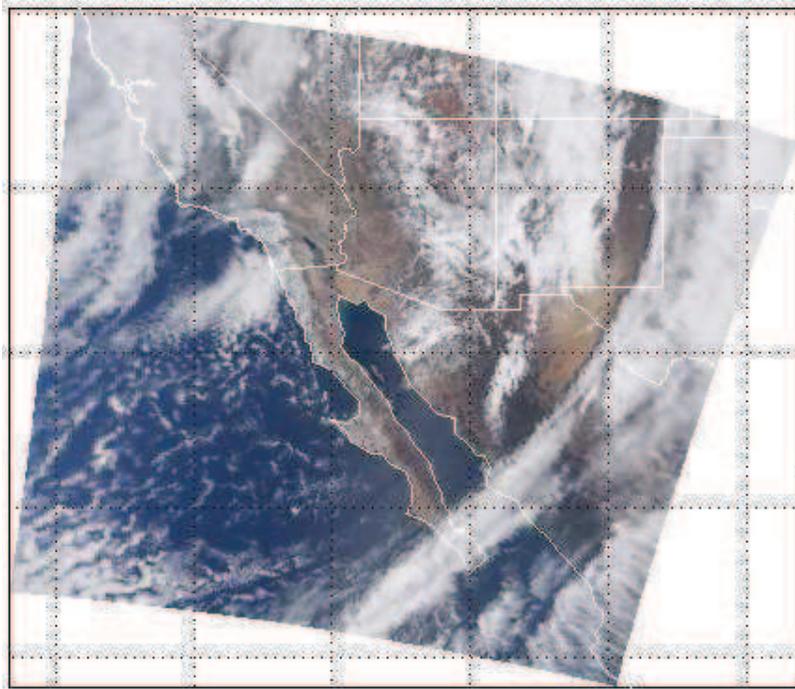


Fig. 5. Dust storm event on April 6th 2001. True color image R=B1, G=B4, and B=B3.

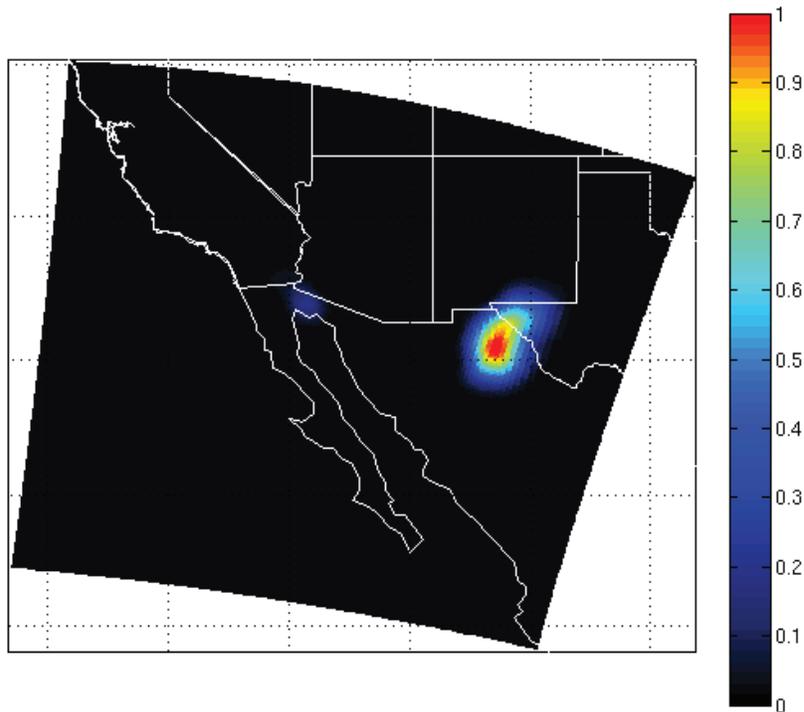


Fig. 6. Dust storm event on April 6th 2001. Dust probability MLE.

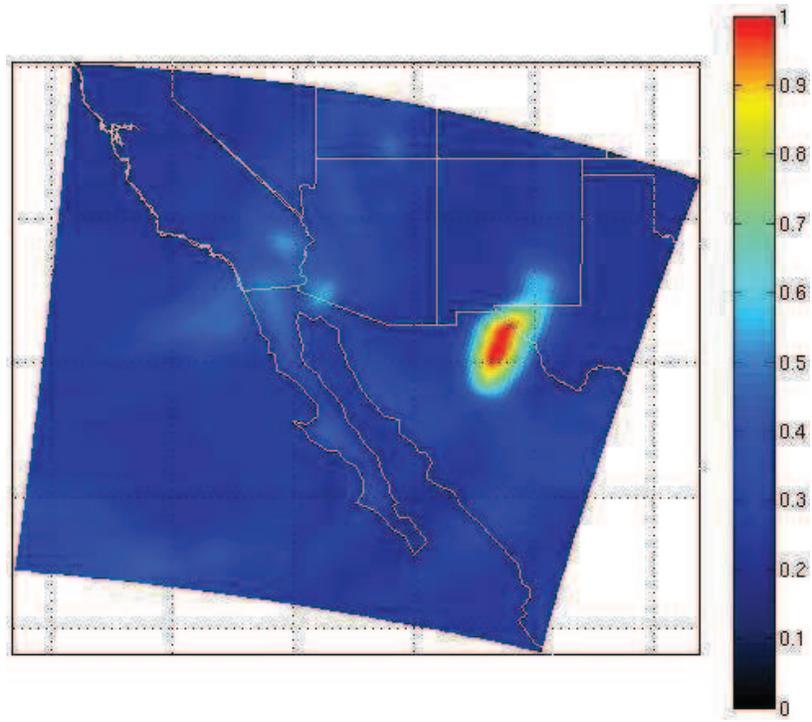


Fig. 7. Dust storm event on April 6th 2001. Dust probability PNN.

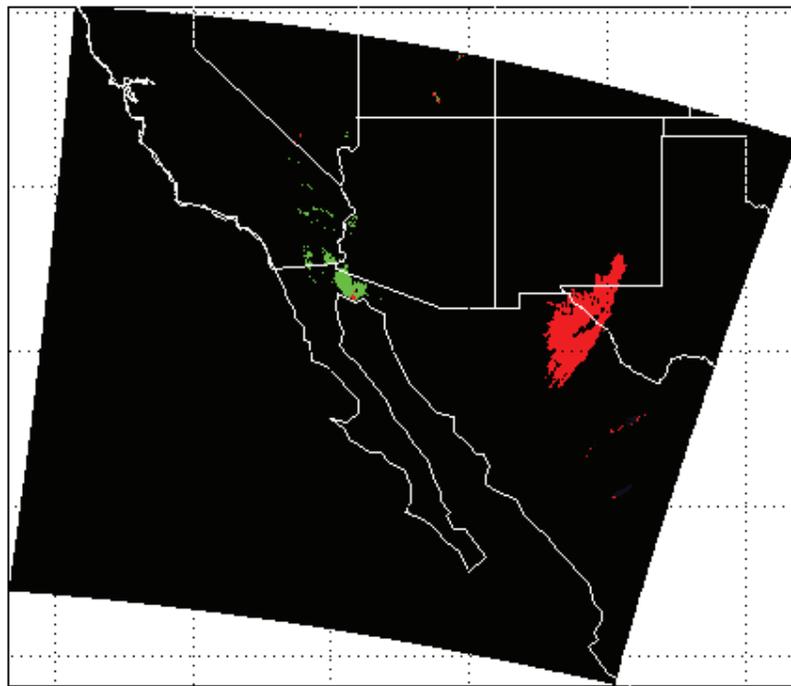


Fig. 8. Dust storm event on April 6th 2001. Segmentation MLE.

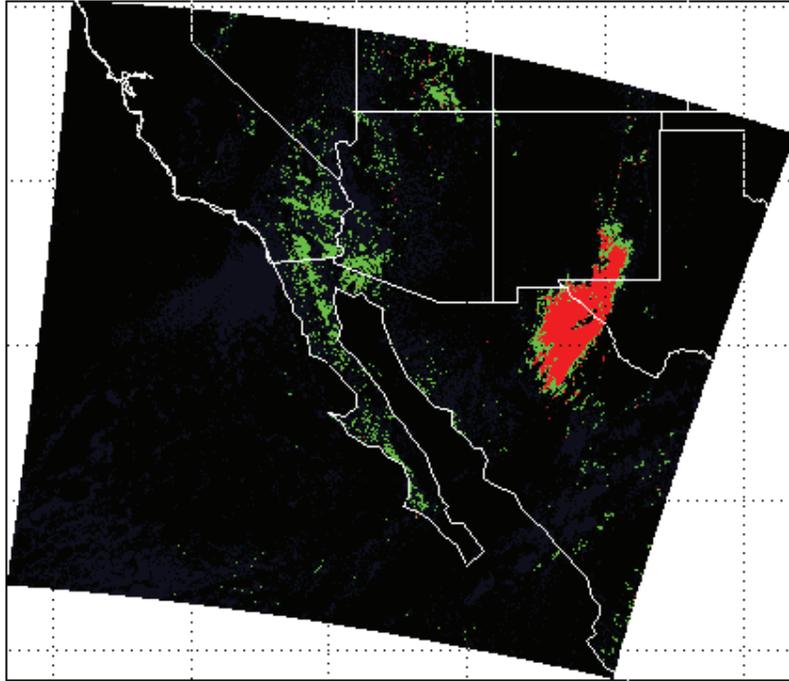


Fig. 9. Dust storm event on April 6th 2001. Segmentation PNN.

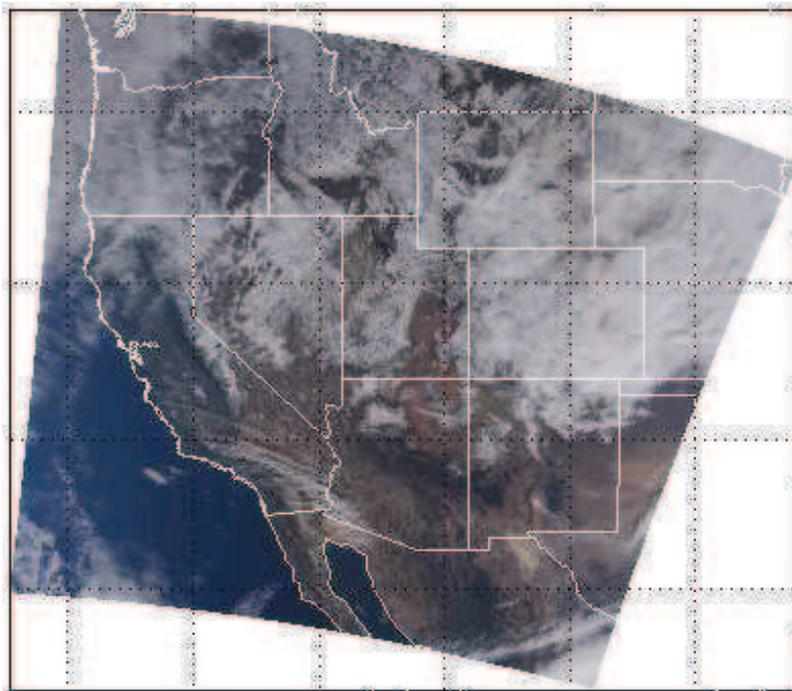


Fig. 10. Dust storm event on December 15, 2003. True color image R=B1, G=B4, and B=B3.

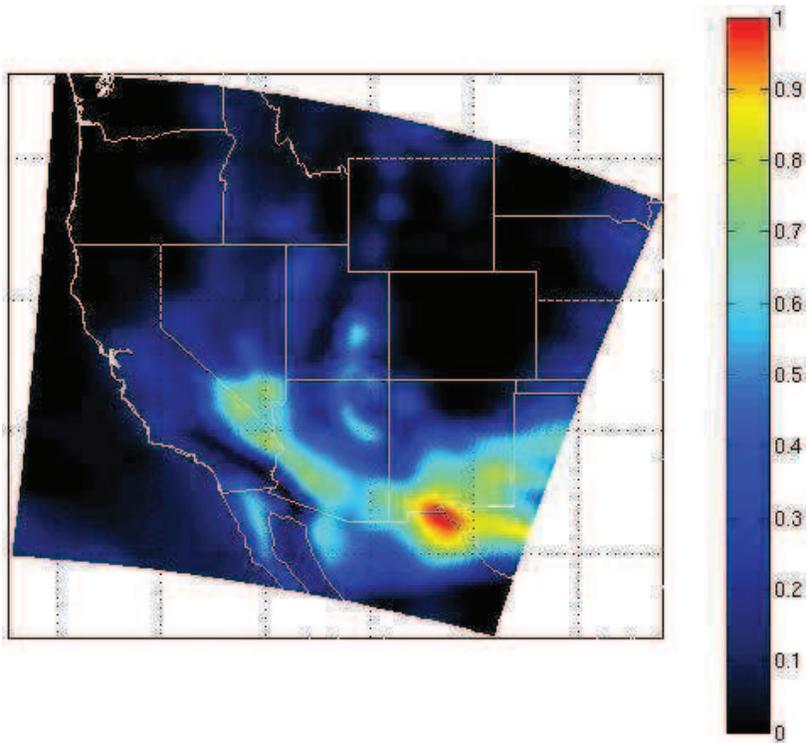


Fig. 11. Dust storm event on December 15, 2003. Dust probability MLE.

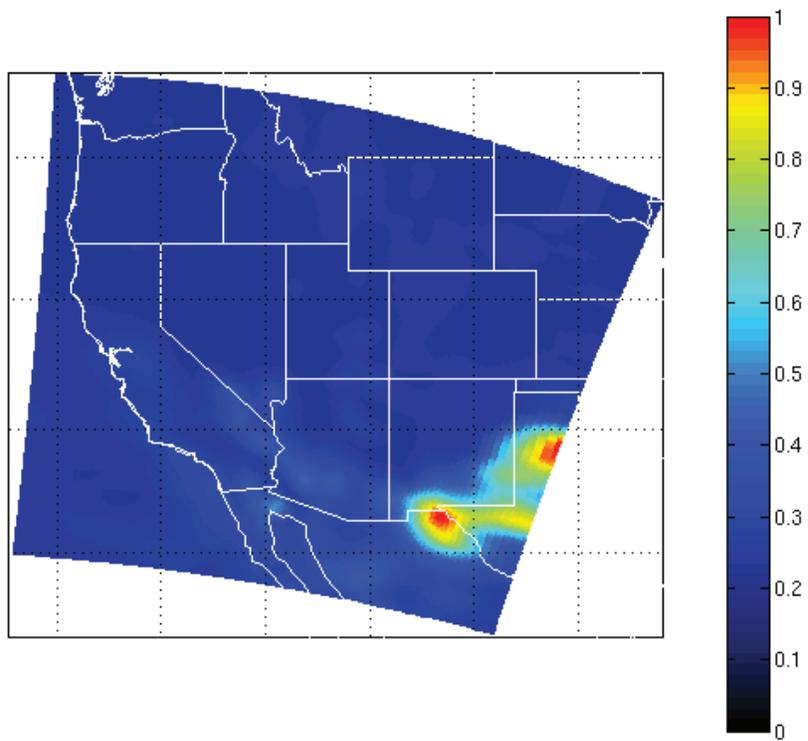


Fig. 12. Dust storm event on December 15, 2003. Dust probability PNN.

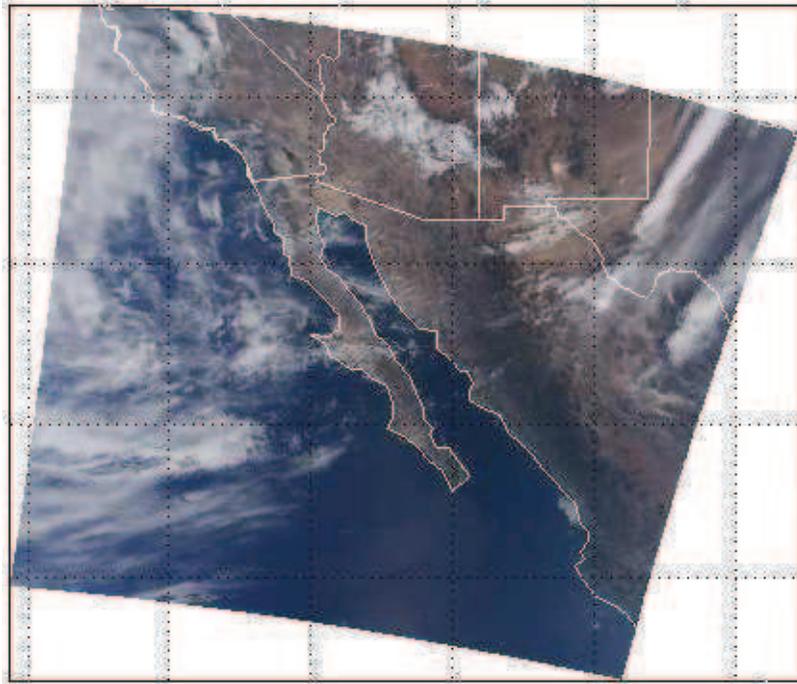


Fig. 13. Dust storm event on February 19, 2004. True color image R=B1, G=B4, and B=B3.

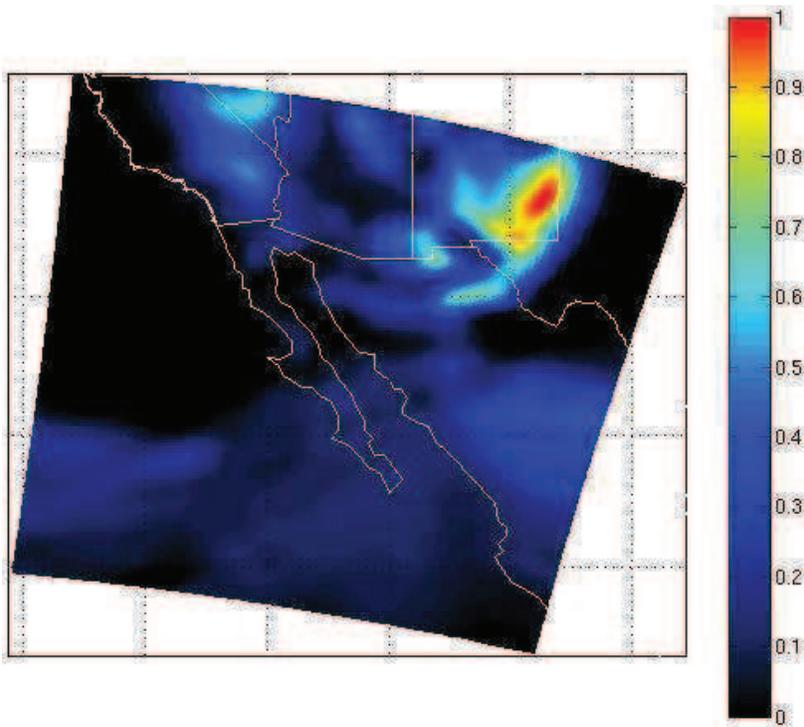


Fig. 14. Dust storm event on February 19, 2004. Dust probability MLE.

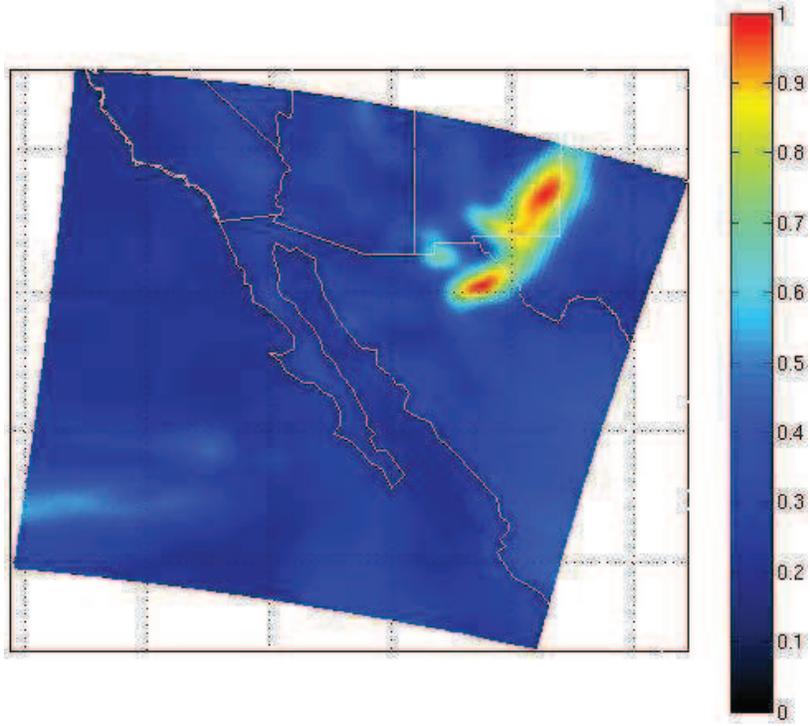


Fig. 15. Dust storm event on February 19, 2004. Dust probability PNN.

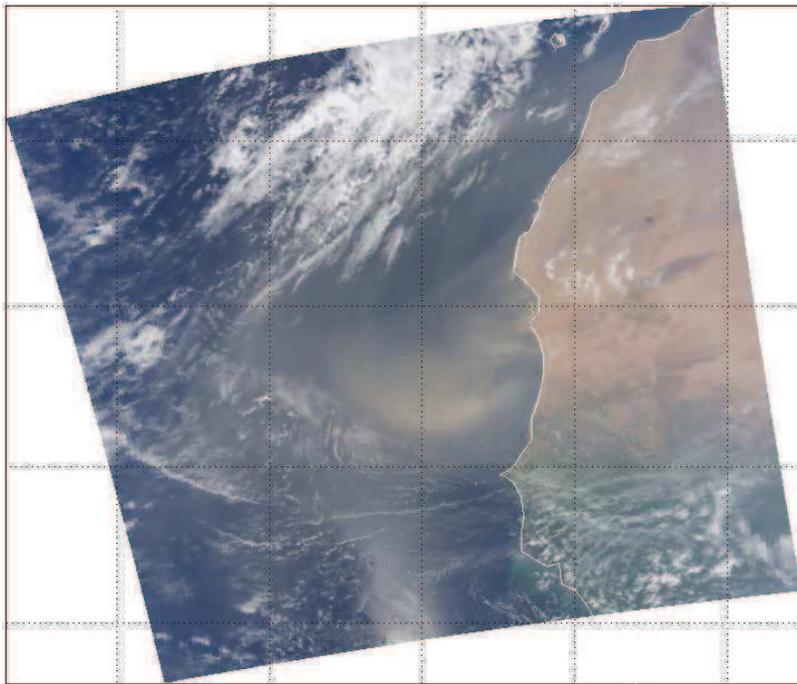


Fig. 16. Dust storm event on September 21, 2009. True color image R=B1, G=B4, and B=B3.

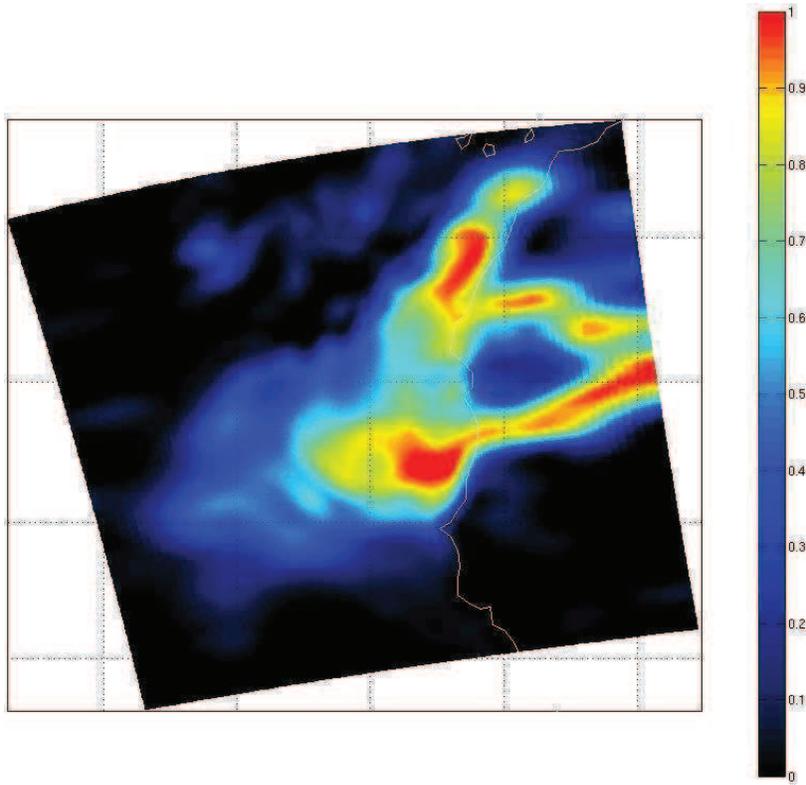


Fig. 17. Dust storm event on September 21, 2009. Dust probability MLE.

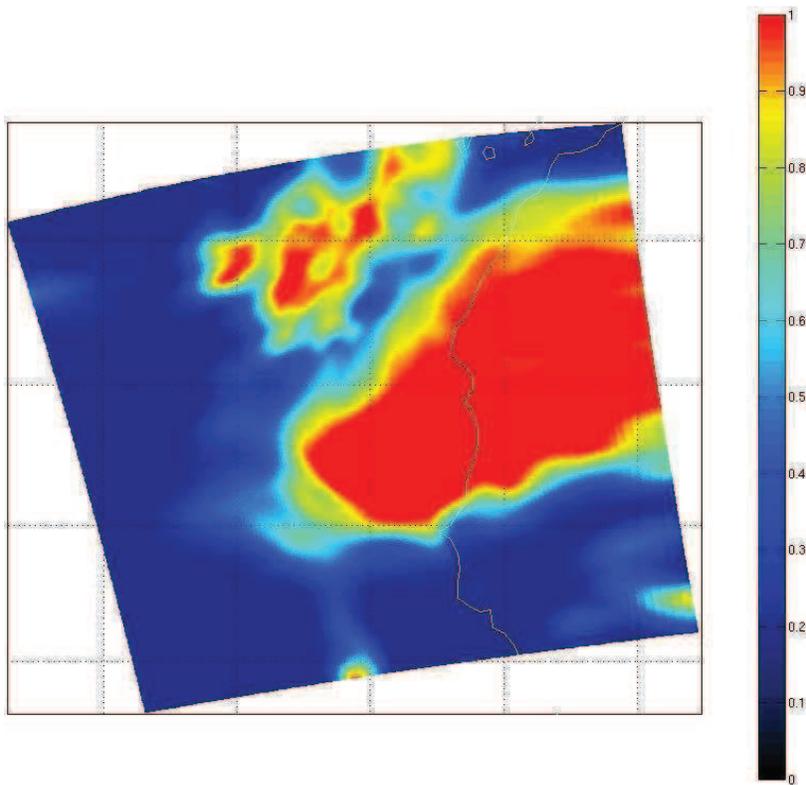


Fig. 18. Dust storm event on September 21, 2009. Dust probability PNN.

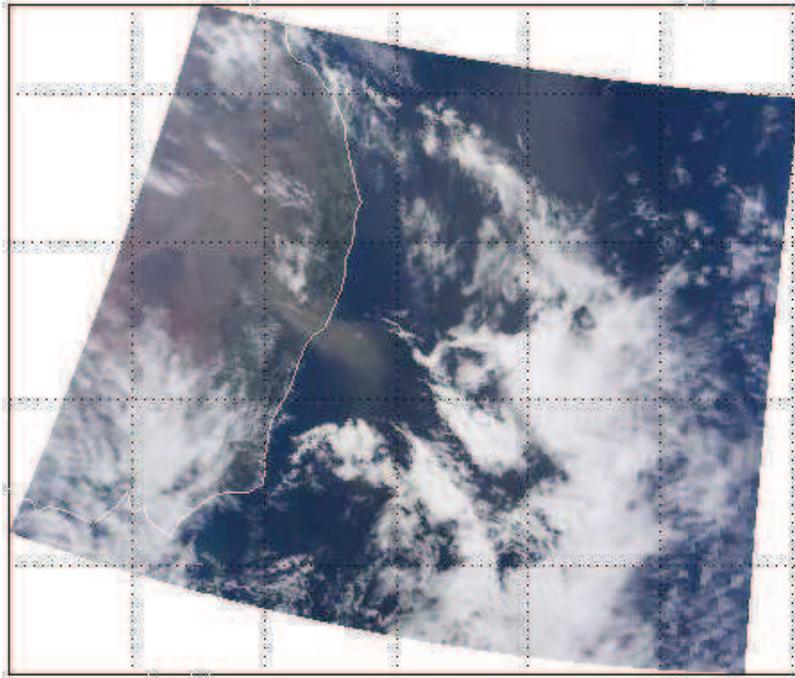


Fig. 19. Dust storm event on October 12, 2009. True color image R=B1, G=B4, and B=B3.

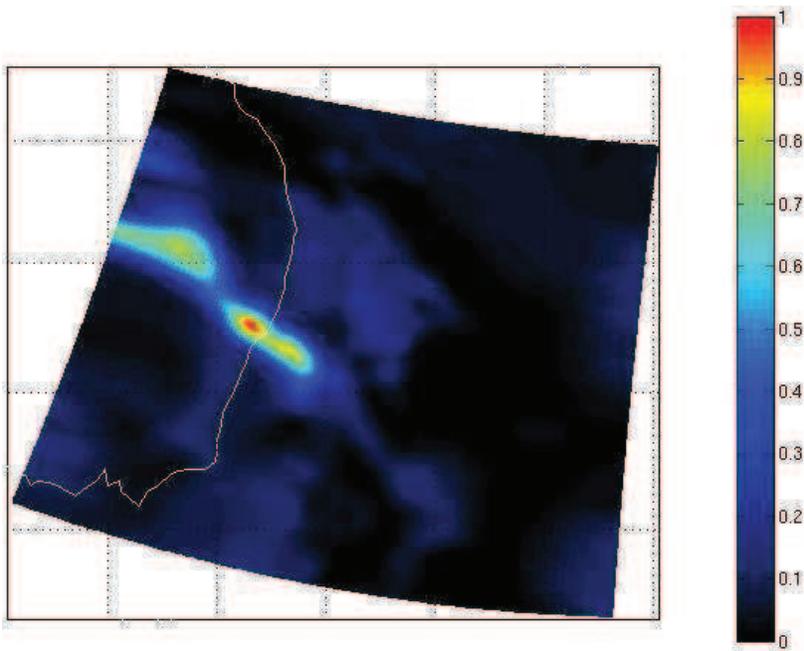


Fig. 20. Dust storm event on October 12, 2009. Dust probability MLE.

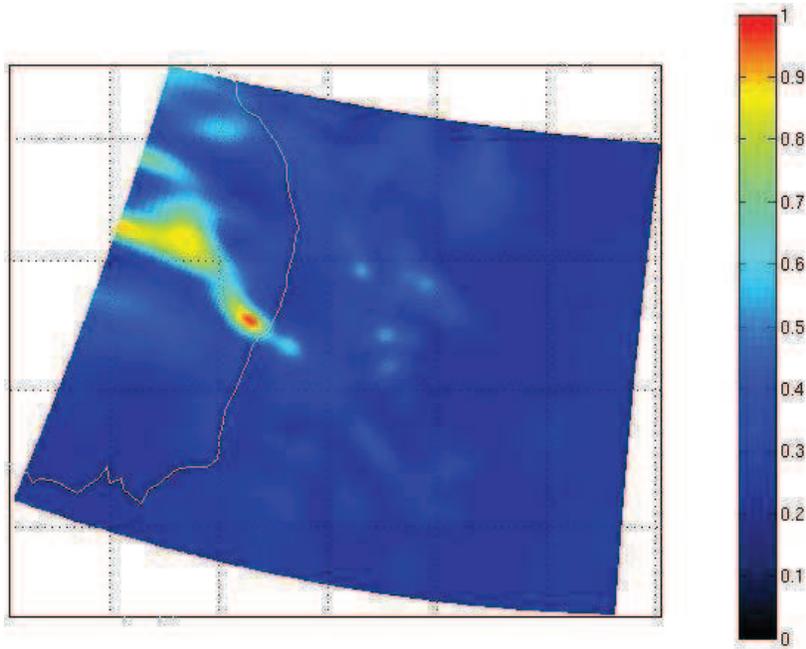


Fig. 21. Dust storm event on October 12, 2009. Dust probability PNN.

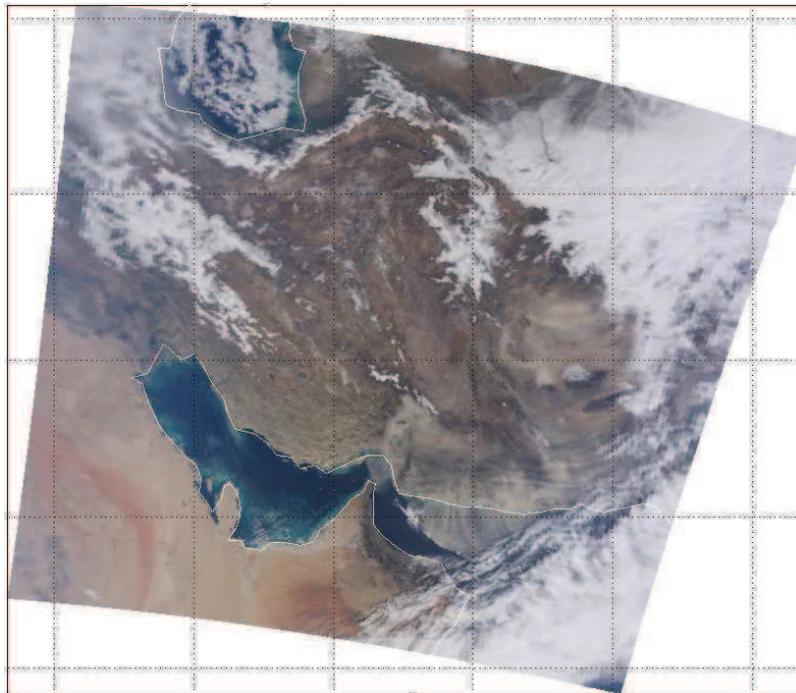


Fig. 22. Dust storm event on February 8, 2010. True color image R=B1, G=B4, and B=B3.

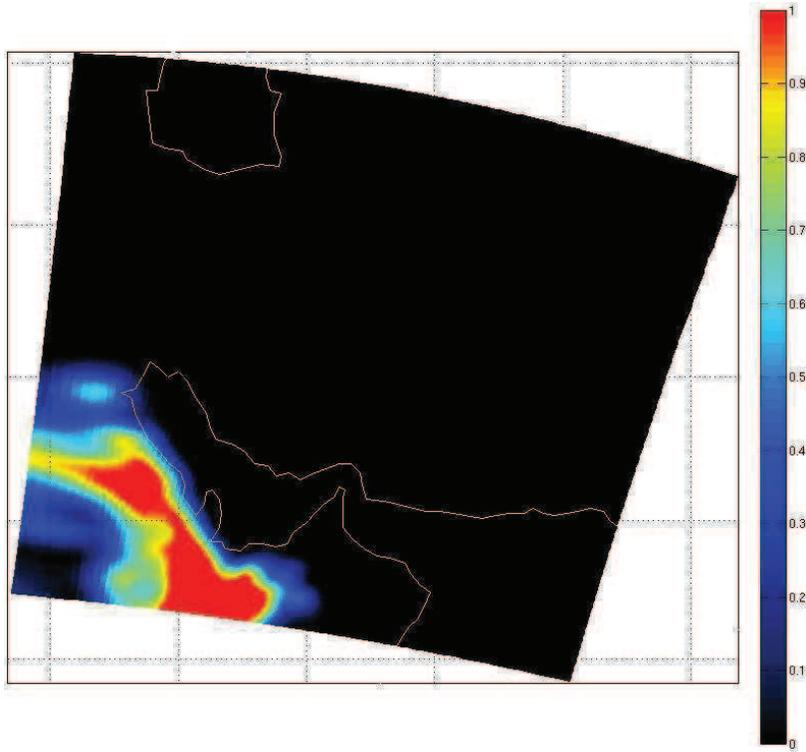


Fig. 23. Dust storm event on February 8, 2010. Dust probability MLE.

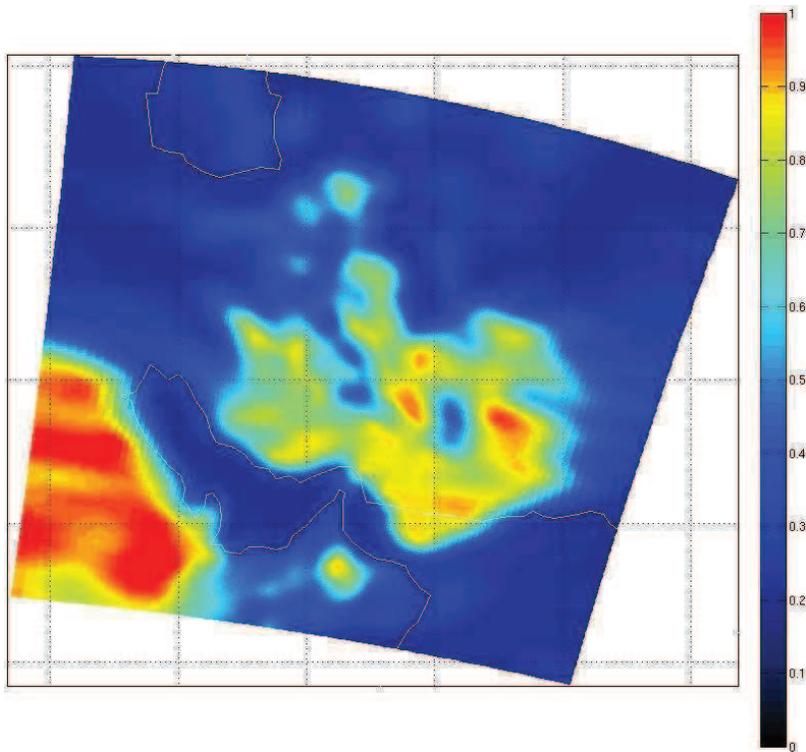


Fig. 24. Dust storm event on February 8, 2010. Dust probability PNN.