Available online at www.sciencedirect.com
ScienceDirect
Remote Sensing of Environment 107 (2007) 138–148
Remote Sensing of Environment
www.elsevier.com/locate/rse

A data-mining approach to associating MISR smoke plume heights with MODIS fire measurements

Dominic Mazzoni [a,c,*], Jennifer A. Logan [b,c], David Diner [a,c], Ralph Kahn [a,c], Lingling Tong [b,c], Qinbin Li [a,c]

[a] Jet Propulsion Laboratory, California Institute of Technology, United States
[b] Division of Engineering and Applied Sciences, Harvard University, United States
[c] Department of Computer Sciences, University of Texas, Austin, United States

Received 4 March 2006; received in revised form 11 August 2006; accepted 15 August 2006

Abstract

Satellites provide unique perspectives on aerosol global and regional spatial and temporal distributions, and offer compelling evidence that visibility and air quality are affected by particulate matter transported over long distances. The heights at which emissions are injected into the atmosphere are major factors governing downwind dispersal. In order to better understand the environmental factors determining injection heights of smoke plumes from wildfires, we have developed a prototype system for automatically searching through several years of MISR and MODIS data to locate fires and the associated smoke plumes and to retrieve injection heights and other relevant measurements from them. We are refining this system and assembling a statistical database, aimed at understanding how injection height relates to the fire severity and local weather conditions. In this paper we focus on our working proof-of-concept system that demonstrates how machine-learning and data mining methods aid in processing of massive volumes of satellite data. Automated algorithms for distinguishing smoke from clouds and other aerosols, identifying plumes, and extracting height data are described. Preliminary results are presented from application to MISR and MODIS data collected over North America during the summer of 2004.

Keywords: MISR; MODIS; Smoke plumes; Boreal fires; Injection height; Aerosol transport; Data mining; Support vector machines; Neural networks

1. Introduction

The injection height of smoke plumes from forest fires is a large source of uncertainty in transport models used to predict the effect of emissions from fires on air quality and climate. It is well known that crown fires generate sufficient energy to loft smoke plumes above the boundary layer (Cofer et al., 1996; Lavoue et al., 2000), facilitating long-range transport of gases and particulate matter (e.g., Bertschi et al., 2004; Colarco et al., 2004; Kahn et al., in press). A large fraction of smoke aerosols remain in the near-surface boundary layer, and do not form discrete "plumes" that are the focus of this paper. Emissions that rapidly escape the boundary layer are more likely to contribute to long distance transport.

Case studies have shown that smoke from large boreal fires can be injected to the lower stratosphere by supercell convection (Fromm & Servranckx, 2003). The frequency of high-altitude (and thus long-lifetime) smoke injection has not been quantified systematically (Fromm et al., 2004). It is possible that boreal and mid-latitude fires may become more common in the future as a result of global warming (e.g., Brown et al., 2004; Flannigan et al., 2000). Understanding the impacts of fires on air quality and climate requires, in part, the use of transport models to relate particle and gas emissions to their downstream dispersal. Observations of aerosol injection are necessary to initialize and validate the models and to develop relationships between injection height and local surface and meteorological conditions. To facilitate progress in this area, we are using data

* Corresponding author. Google, Inc., 604 Arizona Ave., Santa Monica, CA, 90401.
E-mail address: dominic@minorninth.com (D. Mazzoni).

0034-4257/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2006.08.014

from the Terra satellite to obtain statistics on the geographic distribution, extent, orientation, and injection height of plumes (Averill et al., 2005). Terra data acquisition began in February 2000, and NASA recently approved an extension of the mission through 2009. This paper describes the specific process we used to automatically find and extract measurements from smoke plumes, using machine learning techniques and custom image analysis algorithms. We are in the process of refining these algorithms, but our initial results demonstrate the utility of the automated approach.

The development work discussed here uses data collected from June to September 2004 over North America. We selected this period for initial study because of record setting fires in Alaska and the adjacent Yukon Territory of Canada. More than 2.6×10^6 ha burned in Alaska and 1.7×10^6 ha in the Yukon Territory (CFS, 2004; NIFC, 2004). Records were also set for the number of days with reduced visibility caused by wildfire smoke, 42 days in Fairbanks, compared to the previous record of 19 days in 1977, as noted by Averill et al. (2005).

Smoke plumes from these fires were intercepted by aircraft on the INTEX-NA field campaign over the United States which took place from July 1 to August 15 (Singh et al., 2002, in press), and enhanced CO was observed by MOPITT (Measurements Of Pollution In The Troposphere) as a continental scale plume over North America (e.g., Pfister et al., 2005). A case of pyroconvection was documented in June over Alaska, with aerosol enhancements observed near the tropopause (Damoah et al., 2006). Knowledge of the injection heights of the emissions from severe fires such as those in 2004 is required for a quantitative assessment of their effects on atmospheric composition. The work described below is a first step towards providing this information, for plumes immediately downwind of fires.

The Terra Multi-angle Imaging SpectroRadiometer (MISR) instrument observes the Earth in reflected sunlight with a 10:30 AM local time equator crossing, and its typical data collection mode is to observe the Earth globally at nine different view zenith angles in four spectral bands (446, 558, 672, and 866 nm) (Diner et al., 1998). The fore–aft cameras are paired in a symmetrical arrangement and acquire images with nominal view angles, relative to the Earth's surface, at 0°, 26.1°, 45.6°, and 70.5°. In its global observing mode, the nadir camera data in all bands, and the red band data of all of the off-nadir cameras are downlinked at the full spatial resolution of the instrument, 275 m. All other channels are averaged on-board to 1.1-km resolution. The swath width observed in common among all nine cameras is about 380 km. Complete coverage between ±82° latitude is obtained every 9 days. Absolute geolocation uncertainty for the nadir camera is about ±45 m, and relative co-registration errors among the nine cameras are typically less than 275 m.

MISR data make possible unique smoke plume identification and characterization approaches. The use of oblique-angle imagery from MISR enhances plume sensitivity because of the longer optical path through the atmosphere, and the combination of multiangle and multispectral information assists in distinguishing smoke from clouds or other types of aerosols (Mazzoni et al., 2006-this issue). Furthermore, automated pattern matching algorithms (Moroney et al., 2002; Muller et al., 2002; Zong et al., 2002) take advantage of the stereoscopic nature of MISR data, and as part of MISR operational data processing determine the geometric parallax (horizontal displacement) that occurs for a given plume due to its altitude above the surface. Pattern matching is aided by the moderately high spatial resolution of MISR imagery and the 14-bit radiometric depth. Photogrammetric algorithms using accurate camera geometric models transform the derived parallaxes into cloud-top or plume-top heights. Using the nadir and near-nadir cameras, as is done for the standard MISR product, the quantized precision of the resulting height field is ±560 m. Height accuracies for low clouds have been validated to a few hundred meters (Naud et al., 2004); since the technique is purely geometric, comparable accuracy is expected for smoke plumes. Altitudes for clouds as well as smoke, dust, and volcanic plumes are routinely retrieved, and reported on a 1.1-km resolution geolocated grid. However, the MISR standard stereo product does not provide a scene classifier along with the heights which identifies whether the observed target is cloud or aerosol. In this paper, we describe the use of a Support Vector Machine approach for providing this classification in conjunction with height extraction from the MISR stereo product.

Data captured by the MODerate-resolution Imaging Spectroradiometer (MODIS) instrument (Barnes et al., 1998), co-located with MISR on the Terra spacecraft, provides invaluable information about fires, including fire occurrence maps and mean radiative power. Burned area products are being developed with the algorithms of Roy et al. (2005) for the entire MODIS data set. MODIS observes the Earth in 36 spectral bands from 0.4 to 14.4 μm. Its scan pattern sees a 2330-km swath, providing near global daily coverage. MODIS's thermal-infrared sensing capabilities give it the ability to detect active fires with high temporal resolution. By combining MISR and MODIS data over hundreds of fires, and supplementing the Terra data with meteorological information, we are developing a statistical database that will make it possible to empirically relate a fire's power and local atmospheric conditions to the resulting smoke plume injection height. While case studies have shown that this technique is feasible (Kahn et al., in press), performing this computation for hundreds of smoke plumes is daunting, not only due to the calculations required, but also due to the relative rarity of smoke plumes in terabytes of satellite images and the challenge of finding them.

2. Method

Our method was developed based on the assumption that it is unreasonably time-consuming to identify and extract data from every smoke plume manually. Using machine learning and data mining techniques, we have developed an approach for extracting smoke plume and fire data automatically from MISR and MODIS imagery and higher-level data products and retrieving several properties including the plume direction and injection height. Since this automated system cannot identify plumes with perfect accuracy, all individual plumes found are

spot-checked before being included in our database. This human validation does not take very much time and provides an appropriate balance between mostly automated processing and scientific integrity.

Our method is designed to find as many plumes as possible, while minimizing the number of false detections. So as not to create an unreasonable amount of manual labor, we sought to achieve an automatic false detection rate of no more than 3 or 4 times the number of true detections. Thus in our automated algorithm we describe below, we sometimes made design decisions to eliminate certain potential plumes, knowing that it would reject a small number of true plumes, because it would also reject a far greater number of false detections. Unfortunately, this does mean that some plumes are not detected by the current system. We discuss the implications of missed plumes in Section 4.

2.1. Scene pruning

To facilitate the fusion of MISR and MODIS data, it is necessary to define a common coordinate system for the data products from the two instruments. The MISR swath is narrower and a strict subset of the MODIS swath. We therefore found it advantageous to conduct the plume search in the Space Oblique Mercator coordinate system used for MISR's level 1 and 2 data products. In addition, we define a "scene" to consist of one MISR block, comprised of 512×128 pixels at 1.1-km resolution. Our data mining strategy starts with two tests designed to quickly focus on scenes that likely contain smoke plumes. This way the more expensive algorithms to trace the shape of plumes and extract their properties are run on only a fraction of the data. As discussed below, our initial tests reject the vast majority of scenes.

2.2. Identifying fire locations using MODIS

We start with the MODIS Thermal Anomalies product to look for evidence of fires (see Fig. 1). This is a daily Level 3 product at 1-km resolution, and it classifies each pixel as fire, cloud, water, or land, with three levels of fire confidence (high, nominal, or low). It uses the 4 μm brightness temperature and the difference between the 4 mm and 11 μm brightness temperatures in comparison to nearby pixels to detect fires and other thermal anomalies (Kaufman & Justice, 1998). We map the MODIS product onto the MISR grid using a fast nearest-neighbor mapping. More accurate registration is not necessary because we allow several kilometers of tolerance between our estimate of the fire location and the observed source of the smoke plume.

After the MODIS thermal anomalies have been projected onto the MISR grid, the initial pruning occurs at the scene level. If one scene has no high or nominal confidence fire detections at all, it is immediately rejected from further analysis. We found that this eliminates 82.5% of all scenes. (The performance statistics cited in this paper are derived from application of our methodology to the fire season over North America, and may not be appropriate for other regions, such as the tropics where fire occurrence is more frequent.) Any scene that has one or more fire pixels on that day is retained for further analysis.

2.3. Identifying smoke in MISR data using a support vector machine

The next step in the data mining is to identify which pixels contain smoke. To do this, we leverage earlier work (Garay et al., 2005; Mazzoni et al., 2006-this issue) in which we developed a scene classifier able to distinguish between cloud, smoke, dust, land, water, and ice/snow using a combination of spectral, angular, and texture features from MISR (see Fig. 2). Each MISR pixel at 1.1-km resolution is classified independently — this is the same resolution as the stereo-derived height product. The classifier was trained using Support Vector Machines (SVMs) (Cortes & Vapnik, 1995), a modern and powerful machine learning technique. The SVM takes only radiance data from MISR and geometric information as input. We trained the SVM using hundreds of hand-labeled scenes containing those six classes, and tested it by applying the resulting classifier to new scenes. On our independent validation scenes, we determined that the SVM was able to distinguish successfully between all of these classes, though with an accuracy of only 81% at the 1.1-km pixel level. This is sufficient to make it usable since most plumes are over 40 km long and are several kilometers wide; therefore, hundreds of pixels are part of the plume. As long as most of them are classified correctly, we will detect the presence of smoke. Furthermore, the classifier is biased in favor of smoke: it is approximately 2.7 times more likely to misclassify a non-smoke pixel as smoke than it is to misclassify a smoke pixel as another class. This bias works to our advantage, as we would prefer to find a few extra false positives than to miss an important smoke plume that we could have analyzed. Using different terminology, *recall* is more important to us than *precision*.

As a second stage in pruning, we eliminate scenes that do not have a significant amount of smoke. In our current implementation, we eliminate scenes that have less than 2% smoke, because such scenes generally did not contain smoke plumes sufficiently large enough to provide usable data. Eliminating this pruning step would possibly lead to more small smoke plumes being found, at a cost of 15% more scenes to be analyzed.

Although the majority of remaining scenes contain evidence of smoke and fires, only a small fraction of them contain distinct smoke plumes. In some instances, an entire scene can be blanketed with smoke, and one or more large plumes are clearly embedded within it. Our next step is to identify plumes by shape. Our initial attempts involved a template-matching approach, where we compared known smoke plume images to objects seen in the scenes. This successfully found only about 25% of the plumes. We abandoned this approach, realizing that there is too much variability in smoke plume shape. Our preferred approach involves several steps, as described in the next sections.

2.4. Creating a smoke mask image

In order to accurately identify the shape of each object that might be a plume, we need to first generate a mask image for

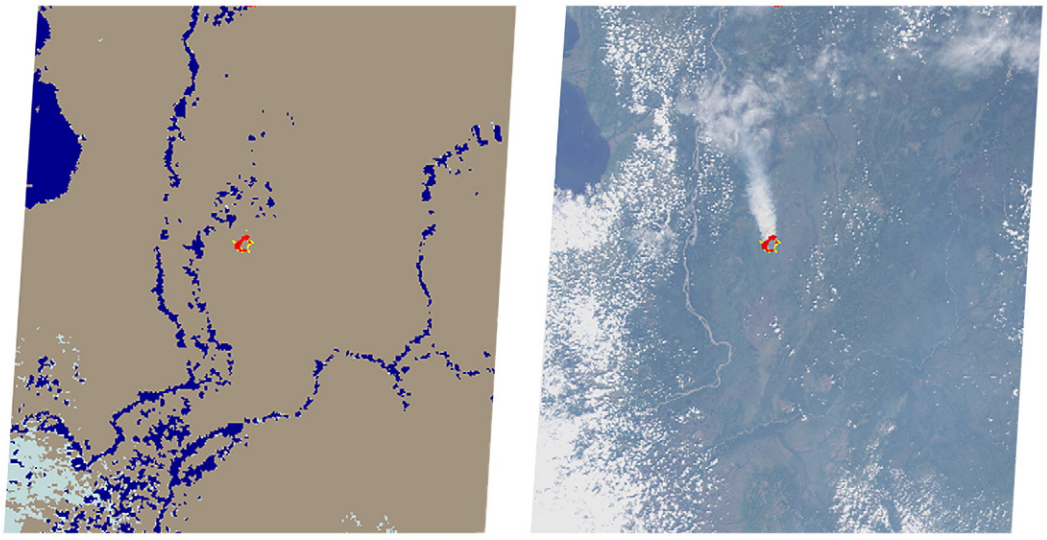

Fig. 1. Left, the MODIS thermal anomalies product. Fires are indicated in red (high confidence) or yellow (low confidence). Right, an overlay of the MODIS fire detections on the MISR image of a smoke plume. MISR orbit 24313, blocks 38–40, July 13, 2004. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

each scene identifying pixels that might belong to a smoke plume. The straightforward approach is to use the SVM to classify every pixel in the scene as smoke or not smoke. This is successful for scenes that contain a smoke plume surrounded by non-smoke pixels, but it fails when smoke pervades the entire scene, though one or more plumes are still visible. For the latter type of scene, a simple brightness threshold often does an excellent job of masking the plumes.

We apply the brightness threshold as follows: let p_{SVM} be the percentage of pixels classified as smoke by the SVM. Choose p_{Br}, the percentage of smoke pixels that will be masked out using a brightness threshold, to be:

$$p_{\mathrm{Br}} = \max(p_{\mathrm{SVM}} - 25\%, 0)$$

Thus if 100% of the pixels in the scene are classified as smoke by the SVM, we choose the brightness threshold so that 75% of the pixels are masked out. If 25% or fewer of the pixels are classified as smoke by the SVM, we do not mask out anything due to brightness, and in-between, some fraction of the pixels are masked out. This is illustrated in Fig. 3: there is pervasive smoke in the MISR image on the left, and in fact over 90% of the pixels are classified as smoke by the SVM, so the SVM mask reveals no interesting structure. However, when about 70% of those pixels are masked out using a brightness threshold, the result is the image seen on the right, where several smoke plumes are clearly visible. We found that 22% of scenes had more than 25% smoke and thus required dynamic thresholding.

2.5. Finding distinct smoke plumes

Given a smoke mask image such as the one in Fig. 3, our goal is now to identify specific plumes. We found template-matching approaches to be inadequate because of the wide variety of possible shapes for smoke plumes. Instead, we sought

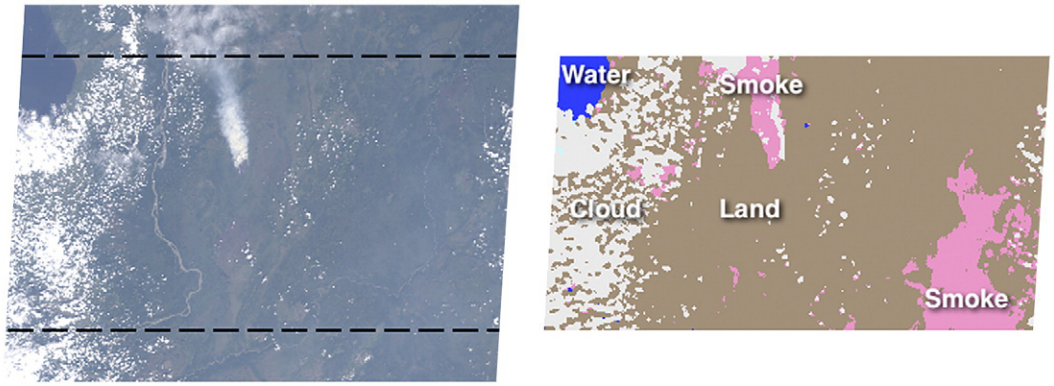

Fig. 2. Left, a true-color image from MISR's AN (Nadir) camera. On the right, the result of the support vector machine (SVM) classification of the image. The plume is clearly detected, as is additional smoke in the lower-right corner of the image, even though it is practically invisible in the nadir image. MISR orbit 24313, blocks 38–40, July 13, 2004.

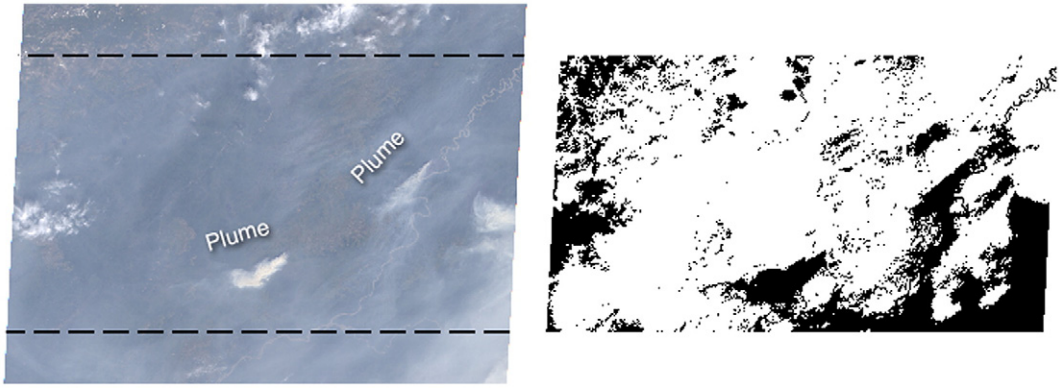

Fig. 3. The SVM classifies this scene as entirely smoke, so the SVM cannot be used to determine the shape of the plume. The image on the right shows the smoke mask computed using dynamic radiance thresholding, and two distinct plumes are visible. MISR orbit 24109, blocks 35–37, June 29, 2004.

to find shapes in the image consisting of a connected set of pixels, with a long thin shape, and a tip near a detected fire source. We think this definition is broad enough to catch most plumes, while specific enough to find a manageable number of false positives.

Here we describe the specific algorithm we implemented to find smoke plumes. This algorithm is somewhat arbitrary and other approaches could conceivably improve upon it, but it works in practice. Our approach in a nutshell is to start with pixels near a fire source that could possibly be the tip of a plume, and explore its connected region to determine if it is long and thin.

We defined a simple metric to identify tips in the mask image by looking at an 11×11 square of pixels at a time. We require that the center pixel of the square is in the mask (i.e., it is classified as smoke), that there is a connected line of smoke pixels all the way to the edge of the square, and that the percentage of smoke pixels in the square is strictly between 20% and 40%. This successfully finds the tip of most plumes, but also finds several times as many additional false tips. We reject tips that are not within a 15-pixel radius (about 16.5 km) of a MODIS fire detection. This large radius is used to account for the fact that some plumes are very narrow near the source, or the fire is not hot enough to be detected by MODIS near the tip of the plume. While there are occasional examples of real plumes that did not have MODIS fire detections nearby, they are relatively rare; the alternative of not rejecting tips too far from a fire results in a large number of false positives.

Next we explore the region of pixels connected to that tip to determine if it is long and thin. The specific algorithm is described below and illustrated in Fig. 4. The threshold values cited in the following algorithm were chosen by trying many possible values systematically and choosing the values that gave the best accuracy on a test set of plume images.

1. Let p_0 be the coordinates of the pixel at the tip of the plume. Compute the set S of all pixels that are classified as smoke and are connected to p_0 by a continuous chain of adjacent smoke pixels. If this set is large, we only consider pixels within a certain radius around p_0 (we chose a radius of about 60 pixels, or 66 km).
2. If $|S|$ (the number of pixels in the set S) is too small, we reject the shape as not being large enough. Shapes that are too small are not only more likely to be false positives, but even

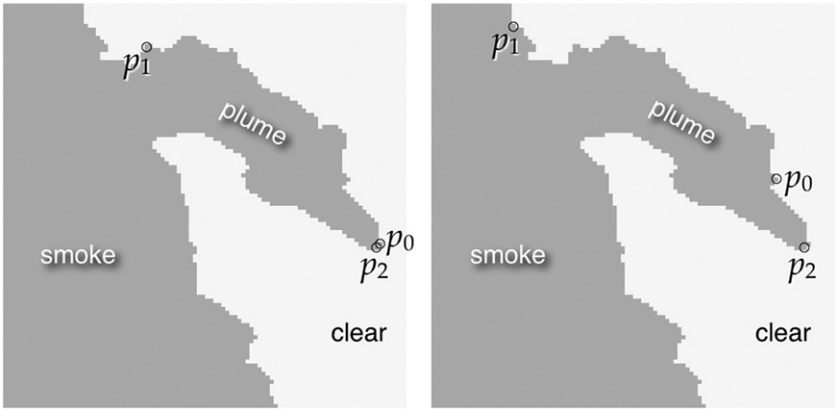

Fig. 4. Illustration of the algorithm used to identify smoke plumes, given an image that has already been separated into smoke and non-smoke pixels. Given a starting point p_0, the algorithm finds the set of all connected smoke pixels within some maximum radius. If p_1 is the farthest point from p_0 within the set and p_2 is the farthest point from p_1 within the set, then we have found the start of a valid smoke plume if p_0 is close to p_2, as in the example on the left. When p_0 is far from p_2, as in the example on the right, the point we found is not the tip of a smoke plume (though it may be a different point on a smoke plume).

if they are real plumes they are less likely to be large enough for us to determine the smoke plume height with the automated stereo height algorithm. We require that $|S| \geq 250$.

3. Let p_1 be the point in S farthest from p_0. We require that $\|p_1 - p_0\| \geq 22$, to ensure that the shape is long enough to be a possible plume.
4. Let p_2 be the point in S farthest from p_1. If p_0 really was the tip of the plume, then p_2 and p_0 should be the same, but if p_0 was a false tip and was in the middle of the set of pixels, then p_2 will be very different than p_0. We require that $\|p_2 - p_0\| \leq 12$; if so, we replace our previous tip with p_2 — experiments showed that p_2 is actually a more accurate guess as to the actual source.
5. Let p_{ctr} be the center of mass of the pixels in S, which is easily computed as the average of the x coordinates and average of the y coordinates of all pixels in S. The vector $v = (p_{ctr} - p_0)/\|p_{ctr} - p_0\|$ is our approximation of the direction of the plume.
6. Finally, we eliminate from S all points which are not within a 50-pixel (about 55 km) radius of p_2, as these are less likely to be useful for measuring properties or computing the injection height.

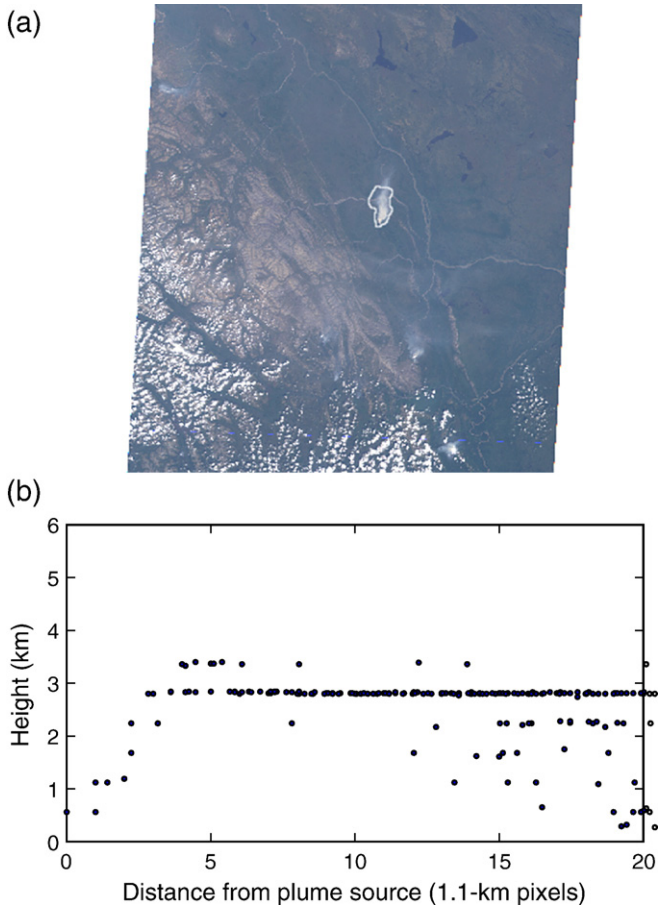

Fig. 5. Above, an image from MISR's AN (Nadir) camera with the smoke plume mask outlined. Below, a plot of the extracted stereo height (above sea level) of each pixel in the plume as a function of its distance from the source. The rise of the plume over the first 3–4 km can be seen. MISR orbit 24166, blocks 38–40, July 3, 2004.

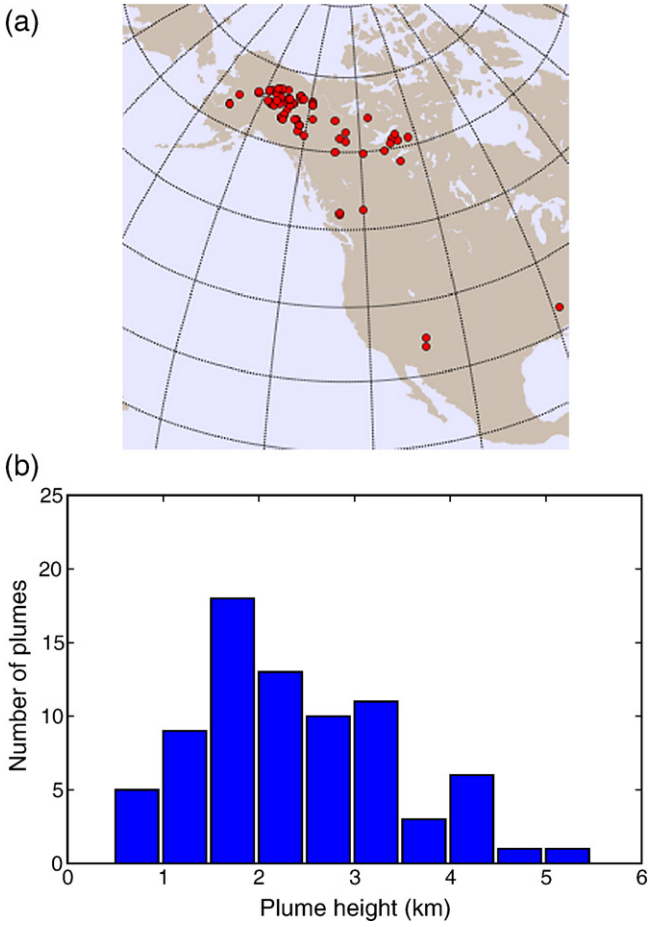

Fig. 6. Above, the geographic distribution of smoke plumes that were found using our automated approach. Below, the distribution of smoke plume heights (above sea level) we computed. There were 77 plumes found, during the summer of 2004.

Out of all plumes found, it is not uncommon to end up with several detections that are actually of the same plume, but with different tips that are off by only a couple of pixels. Therefore we perform an arbitration step, where we randomly choose one plume out of all overlapping detections, and eliminate the others. After this step is done, for each plume we save the tip p_2 and the set of points S for further analysis.

2.6. Limitations of this approach

There are several examples of smoke plumes that we would like to be able to detect but our algorithm is currently unable to identify. It was not a design goal to find plumes that originate outside of the MISR swath, because we would have less information about the fire, so our approach deliberately rejects such examples. Smoke plumes that are very near the edge of the MISR swath can confound the algorithm. A more common example of a failure was when two independent smoke plumes that originate very near each other merge to form a single shape. Most of the time our algorithm rejects such examples. We expected cases of mixed clouds and smoke to cause problems, but we found several examples of smoke plumes with

Table 1
Table of data collected from 77 smoke plumes observed over North America in the summer of 2004

Date (2004)	UTC time	Terra orbit no.	Lat.	Lon.	Orient.	Plume height (km)	Fire area (km^2)	Mean radiative power (MW)
6/19	21:16:58	23963	66.483	−138.593	259°	1.9	20	56.5
6/19	21:16:58	23963	67.161	−146.330	10°	1.9	148	143.6
6/19	21:17:18	23963	67.187	−146.469	27°	2.0	148	143.6
6/19	21:17:18	23963	66.937	−145.367	32°	2.0	95	171.0
6/19	21:17:18	23963	65.581	−141.263	279°	1.5	26	43.8
6/19	21:17:18	23963	65.648	−142.905	260°	2.0	0	10.8
6/19	21:17:39	23963	64.917	−141.803	295°	1.9	51	39.6
6/21	19:30:10	23991	52.105	−125.120	290°	3.6	8	70.3
6/22	21:47:49	24007	67.140	−146.396	79°	5.2	148	143.6
6/23	20:53:12	24021	63.823	−138.368	272°	2.7	9	51.4
6/23	20:53:12	24021	63.904	−142.363	35°	4.8	64	109.6
6/23	20:53:33	24021	63.428	−142.653	8°	4.1	61	116.7
6/25	20:53:33	24050	63.114	−137.436	285°	3.0	12	49.9
6/25	20:53:33	24050	62.254	−137.297	330°	4.4	48	75.0
6/29	21:54:20	24109	65.708	−152.227	258°	1.8	85	69.0
6/30	20:58:22	24123	66.370	−134.425	243°	1.7	50	64.2
6/30	20:58:22	24123	66.876	−138.697	319°	1.9	84	74.3
6/30	20:58:43	24123	66.876	−138.697	317°	2.0	84	74.3
6/30	20:58:43	24123	66.097	−134.707	226°	1.6	102	92.9
6/30	20:58:43	24123	66.218	−142.239	256°	1.9	44	75.6
6/30	20:58:43	24123	65.837	−140.593	265°	2.0	63	62.6
6/30	20:59:03	24123	65.816	−140.759	271°	2.2	63	62.6
6/30	20:59:03	24123	65.503	−141.311	303°	2.8	40	79.1
6/30	20:59:24	24123	63.644	−139.038	257°	2.6	120	195.2
6/30	20:59:24	24123	63.476	−142.850	306°	4.5	46	92.1
7/02	20:46:21	24152	66.357	−134.434	256°	2.2	50	64.2
7/02	20:46:21	24152	66.680	−137.735	345°	2.2	57	57.5
7/02	20:46:21	24152	66.104	−134.711	223°	1.7	102	92.9
7/03	18:20:29	24165	35.069	−111.326	48°	2.2	60	244.0
7/03	18:20:50	24165	34.043	−111.459	354°	3.4	140	83.9
7/03	19:51:26	24166	64.141	−127.219	50°	2.5	11	26.1
7/03	19:51:46	24166	62.660	−124.059	22°	3.5	21	126.0
7/03	19:52:07	24166	61.879	−125.596	44°	3.1	14	40.5
7/03	19:52:07	24166	61.509	−124.155	11°	3.9	5	152.7
7/04	20:35:02	24181	63.090	−137.282	304°	3.1	10	69.9
7/05	19:38:44	24195	64.478	−117.431	19°	3.6	1	23.3
7/10	21:36:10	24269	65.782	−152.377	120°	2.1	60	44.3
7/11	19:02:41	24282	61.053	−110.263	24°	4.1	3	50.2
7/11	20:40:53	24283	64.058	−133.825	102°	3.1	26	66.1
7/11	22:19:26	24284	64.516	−157.644	61°	1.1	26	44.7
7/11	22:20:07	24284	62.807	−158.756	63°	1.1	58	71.3
7/12	21:23:27	24298	65.806	−145.586	73°	4.4	31	62.0
7/12	21:23:48	24298	65.089	−147.220	68°	2.9	101	72.4
7/13	18:50:20	24311	61.413	−111.646	193°	1.5	12	34.9
7/13	22:07:45	24313	62.819	−158.778	10°	2.6	58	71.3
7/17	21:42:00	24371	65.195	−147.339	354°	3.2	101	72.4
7/17	21:42:00	24371	65.088	−147.134	32°	3.3	101	72.4
7/17	21:42:21	24371	65.088	−147.134	37°	3.1	101	72.4
7/18	20:47:25	24385	61.724	−135.456	337°	4.1	26	41.0
7/22	18:44:08	24442	61.821	−110.724	152°	2.8	17	121.2
7/22	18:44:08	24442	61.248	−107.278	79°	2.9	11	73.1
7/24	18:32:48	24471	58.386	−110.330	82°	1.7	51	79.9
8/01	19:23:54	24588	52.266	−125.260	252°	3.3	40	81.8
8/03	19:08:46	24617	60.849	−112.028	332°	2.1	37	64.4
8/03	19:09:06	24617	59.995	−113.971	331°	1.6	5	50.3
8/10	20:53:03	24720	63.744	−142.477	288°	2.4	25	35.6
8/10	20:53:03	24720	63.692	−143.124	47°	1.8	262	78.4
8/11	19:58:27	24734	60.174	−126.723	42°	2.5	12	37.6
8/17	19:21:38	24821	59.826	−119.532	177°	1.9	83	68.6
8/17	19:23:22	24821	52.578	−120.129	193°	2.8	1	18.5
8/20	21:28:58	24866	67.448	−143.365	253°	1.4	41	84.9
8/20	21:28:58	24866	67.273	−145.424	248°	1.3	60	67.8
8/20	21:28:58	24866	66.179	−142.163	277°	1.5	65	76.9
8/20	21:29:18	24866	66.785	−148.653	258°	0.9	108	76.9

Table 1 (*continued*)

Date (2004)	UTC time	Terra orbit no.	Lat.	Lon.	Orient.	Plume height (km)	Fire area (km^2)	Mean radiative power (MW)
8/20	21:29:18	24866	66.315	−149.208	266°	0.8	85	43.6
8/20	21:29:39	24866	65.113	−148.552	202°	1.6	27	36.3
8/22	21:16:55	24895	66.204	−142.114	268°	1.3	65	76.9
8/22	21:16:55	24895	66.364	−146.364	220°	0.7	25	37.7
8/22	21:16:55	24895	66.141	−145.023	197°	0.8	32	53.0
8/22	21:17:15	24895	65.361	−144.663	265°	1.0	0	0.0
8/22	21:17:15	24895	64.927	−146.104	275°	2.1	85	54.7
8/22	21:17:15	24895	64.094	−142.445	323°	2.9	16	39.3
8/22	21:17:36	24895	63.334	−142.208	305°	3.2	5	29.7
8/27	21:35:04	24968	67.196	−147.094	237°	1.2	35	67.5
9/07	21:17:14	25128	65.530	−145.580	47°	2.4	8	48.8
9/12	21:35:25	25201	66.485	−149.520	277°	2.0	11	37.4
9/28	16:48:00	25431	33.386	−88.894	175°	1.2	13	108.9

An electronic version can be found at this URL: http://www-misr.jpl.nasa.gov/mission/data/plume_data/PlumeDB-2006-02-06.csv.

pyrocumulus clouds that were successfully detected by our system.

2.7. Computing the injection height and other properties

For each plume found, and for which associated stereo heights have been retrieved, we determine several properties associated with the plume from the MISR and MODIS data. The most important of the properties we derive from the MISR data is the maximum observed height of the plume, as one of our primary goals is to empirically relate the plume's injection height to the power of the fire and local weather conditions.

MISR has two different stereo-derived height products: with and without wind correction. Because there is approximately a 1-min delay between the time when each of MISR's cameras images each ground pixel, disparities in the location of features that appear in multiple cameras could be due to either to parallax resulting from the feature's height or to true displacement resulting from advection, or as is most likely, a combination of both. Because the density of stereo height retrievals that do not have the wind correction applied is larger, our initial derivation of plume height makes use of the non-wind-corrected values. Correction for wind is then applied as a final step.

Fig. 5 shows a plot of non-wind-corrected heights above sea level of the pixels in a particular plume, as a function of their distance from the tip of the plume. The plot clearly shows the rise of the plume and its stabilization at around 3 km. Note that the vertical resolution of the height calculation has a precision of approximately 560 m, which is why the heights appear to be in discrete bins. In theory, the injection height we are computing would be the maximum height obtained by the plume given a plot such as the one in Fig. 5. However, frequently the height retrieval gives erroneous results for as many as 10% of the pixels in a plume, sometimes due to stereo matching errors, or due to detection of higher clouds that were not screened out. As a result, we found it necessary to add some heuristics to eliminate outliers and estimate the true injection height. Specifically, given the set of height retrievals H in km, we first define a threshold $\max(\text{median}(H)*1.5, \text{median}(H)+1.5)$, and eliminate all heights above this threshold. This eliminates higher clouds. Then, instead of taking the maximum of H, we take the average value of the top 10% of H — this appears to give a better estimate of the true maximum height in the presence of a small amount of noise in the retrieval. Note that these heuristics were developed while examining a few dozen initial plumes. Analysis of more data could lead to more formal ways of eliminating outliers.

As a final step in the determination of injection height, a wind correction equation is applied to the results. The MISR stereo algorithm is designed to separate the apparent motion due to height from the true motion due to wind using camera triplets, but this calculation does not work on all pixels and sometimes results in isolated blunders. To deal with potentially problematic MISR wind retrievals, we use the following approach: first, our system retrieves all heights within the plume without any wind correction. Then, all stereo height retrievals within a 3-block area that have both a "non-wind-corrected" and "wind-corrected" value are regressed and a linear fit is obtained. This provides a height-dependent correction equation for this local region, essentially allowing the estimation of a mean wind correction to the heights while ignoring small errors in the wind retrieval or blunders in individual pixels.

3. Results and analysis

We have analyzed four months of data from the summer of 2004 (June through September) over North America. Our initial study coincides with the time period studied by the Intercontinental Chemical Transport Experiment North America (INTEX-NA) field campaign (Singh et al., 2002, in press), motivated by the relatively large number of fires in Alaska and the Yukon territory that summer. As noted in Averill et al. (2005), more than 2.6×10^6 ha burned in Alaska, and poor visibility was recorded in Fairbanks on 42 of the 92 days of summer.

We processed a total of 39,105 individual scenes (MISR data blocks) for June to September of 2004 over North America. As previously noted, 82.5% of these were immediately pruned due to lack of MODIS fire detections, and another 15% of the remaining were pruned due to lack of smoke detections by the SVM classifier. Of the remaining scenes, our analysis software found distinct plumes in only 196 of them. A total of 325

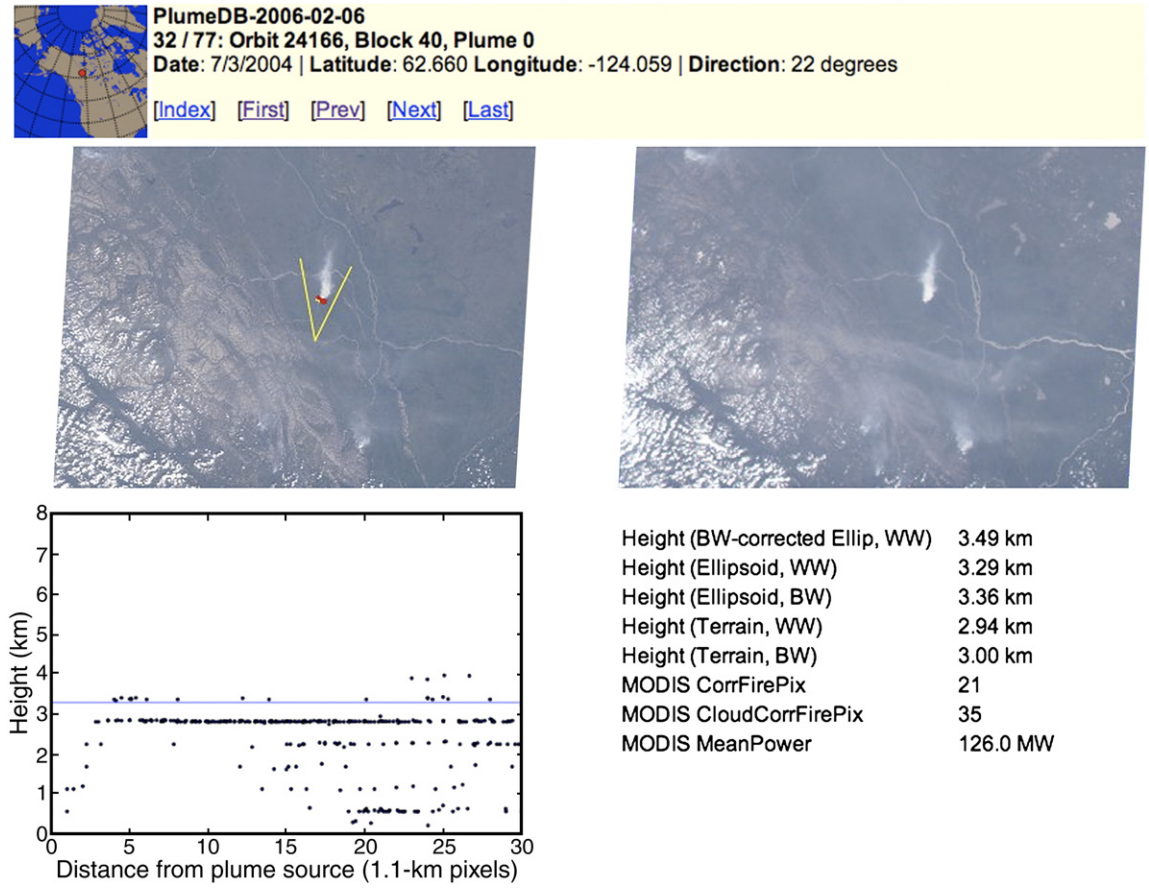

Fig. 7. A page from our website showing detailed information and images about all of the plumes we have analyzed. Available from the following URL: http://www-misr.jpl.nasa.gov/mission/data/plume.html.

candidate plumes were identified (some scenes contained more than one plume).

We manually examined each of the 325 potential plumes to determine if each was a proper detection. It turned out that 138 of them were true detections and 187 were false positives. The false detections included long, thin branches of a real smoke plume, smoke from a distant source that was trapped in a long, thin valley (giving the appearance of a plume), and even thin smoke over top of a long, thin river (because the smoke is easier to detect over the water than the surrounding land). More than half of the false positives clearly contained smoke, but not a plume originating from a source on the ground. Despite the large number of false positives from the automated algorithm, this result is considered quite acceptable because the alternative would be human examination of a data volume two orders of magnitude larger. Of the properly detected plumes, 61 had inconclusive stereo height data from MISR's operational product, either because the plume was too small or thin, or because the plume was surrounded by other smoke and clouds that made it too difficult to separate the plume from its surroundings.

Fig. 6 shows a histogram of the 77 plume heights (above sea level) obtained using the automated algorithm. The range is from 0.72 km to 5.18 km, with a mean of 2.42 km and a median of 2.16 km. None of our plumes came close to reaching the lower stratosphere, possibly indicating that plumes reaching those heights are rare. MISR has observed such plumes before, such as the plume from the Chisholm forest fire near Edmonton, Alberta, Canada, which reached heights of 12–13 km above sea level, as observed by MISR on May 29, 2001 (Diner et al., 2004).

A summary of all of the data we collected on all 77 plumes is found in Table 1. For each plume, we indicate the date and UTC time, the orbit number of the Terra spacecraft, the latitude and longitude of the plume source, the orientation (clockwise with 0° for North), injection height in km (as measured by MISR's operational stereo height product with the corrections described in the text), the fire area as retrieved by MODIS, and the mean radiative power in MW as retrieved by MODIS. We have posted an electronic version of this table at the following URL: http://www-misr.jpl.nasa.gov/mission/data/plume_data/PlumeDB-2006-02-06.csv and subsequent updates will be available here: http://www-misr.jpl.nasa.gov/mission/data/plume.html.

On this site you will also find more detailed information about each plume, including a plot of the geographic coordinates on a globe for context, a plot of the pixel heights (like in Fig. 5), and an animation of images from the different MISR camera angles, making the three-dimensional structure of the plume more apparent. An example page from this site is seen in Fig. 7. We intend to update this site from time to time as we improve our algorithms and collect more data.

Not counting the time to develop and debug our system or acquire the raw data, processing the four months of data for this study required approximately three days of computation time on an ordinary Linux workstation. The raw data consumed 720 GB on disk. Examining the resulting plumes and filtering out false positives required about two hours of human time. Extrapolating these numbers, in order to achieve our goal of analyzing five years of data, we estimate that we will need approximately 11 TB of disk space, a month and a half of CPU time (though this is trivially parallelizable across multiple machines), and a week of manual labor to eliminate false positives.

4. Discussion and future work

We have shown how data mining methods applied to imagery and higher level data products from MISR and MODIS on Terra are capable of generating partially automated retrievals of smoke plume injection heights over large spatial areas. The methodology enables pinpointing smoke sources and injection heights, which are needed for accurate modeling of 3-D transport. The ability to deal with large volumes of data makes it possible to assemble a statistical database of information that can be used in realistic simulations of the effects of fires on air quality. The database we are developing will make possible tests for fire models that predict plume heights for specific fires, based on the energy release. We plan to investigate the associated meteorology and likely fuel consumption for the fire/plume events we have identified in Alaska and the Yukon.

More detailed information on plume profiles is potentially available from the CALIPSO lidar (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation), which launched in April 2006. Lidar measurements are also needed to validate the plume heights retrieved by MISR. However, the locations at which the CALIPSO ground track crosses a given plume will be an even greater matter of serendipity than is the case with MISR. The spatial and temporal sampling characteristics of MISR enable observation of any given mid-latitude location once every 4–5 days on average. Even so, many plumes that are visible within the broad MODIS swath are missed, and failure of any stage of our algorithms could also reduce the potential size of the database. Nonetheless, the methodology described here makes possible an empirical investigation of the relationship between source conditions and plume heights that could not be done prior to the launch of Terra. We will partially make up for plume undersampling through sheer volume, and are currently in the process of scaling this approach to many years of MISR and MODIS data over North America.

Acknowledgements

Portions of the research described in this paper was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Jennifer Logan was funded by EPA grant RD-83227501-0 to Harvard University. MISR data were obtained from the Atmospheric Sciences Data Center located at the NASA Langley Data Active Archive Center (DAAC). The MODIS fire products were obtained from Earth Resources Observation and Science (EROS) Data Center.

References

Averill, C., Mazzoni, D., Logan, J., Tong, L., Diner, D., & Li, Q. (2005). Combining MISR and MODIS data to automatically catalog smoke plumes in north America. *The Earth Observer*, *17*, 11–12.

Barnes, W., Pagano, T., & Salomonson, V. (1998). Prelaunch characteristics of the Moderate Resolution Imaging Spectroradiometer (MODIS) on EOS AM-1. *IEEE Transactions on Geoscience and Remote Sensing*, *36*(4), 1088–1100.

Bertschi, I., Jaffe, D., Jaeglè, L., Price, H., & Dennison, J. (2004). PHOBEA/ITCT 2002 airborne observations of trans-pacific transport of ozone, CO, vocs, and aerosols to the northeast pacific: Impacts of Asian anthropogenic and Siberian boreal fire emissions. *Journal of Geophysical Research*, *109* (D23). doi:10.1029/2003JD004200

Brown, T., Hall, B., & Westerling, A. (2004). The impact of twenty-first century climate change on wildland fire danger in the western United States: An applications perspective. *Climatic Change*, *62*, 365–388.

Canadian Forest Service (2004). *National forest fire situation reports.* http://www.nrcan-rncan.gc.ca/cfs-scf/science/prodserv/firereport/archives_e.php

Cofer, W. R., Winstead, E. L., Stocks, B. J., Overbay, L. W., Goldammer, J. G., Cahoon, D., et al. (1996). *Emissions from boreal forest fires: Are the atmospheric impacts underestimated?* (pp. 834–839). Cambridge, MA: MIT Press.

Colarco, P., Schoeberl, M., Doddridge, B., Marufu, L., Torres, O., & Welton, E. (2004). Transport of smoke from Canadian forest fires to the surface near Washington, D.C.: Injection height, entrainment, and optical properties. *Journal of Geophysical Research*, *109*, D06203.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297.

Damoah, R., Spichtinger, N., Servranckx, R., Fromm, M., Eloranta, E. W., Razenkov, I. A., et al. (2006). A case study of pyro-convection using transport model and remote sensing data. *Atmospheric Chemistry and Physics*, *6*, 173–185.

Diner, D. J., Beckert, J. C., Reilly, T. H., Bruegge, C. J., Conel, J. E., Kahn, R., et al. (1998). Multiangle imaging spectroradiometer (MISR) instrument description and experiment overview. *IEEE Transactions on Geoscience and Remote Sensing*, *36*, 1072–1087.

Diner, D. J., Fromm, M. D., Torres, O., Logan, J. A., Martonchik, J. V., Kahn, R. A., et al. (2004). New satellite observations of upper tropospheric/lower stratospheric aerosols: Case studies over the U.S. and Canada. *Eos, Transactions of the American Geophysical Union*, *85*(47) (Fall Meet. Suppl., Abstract A21B-0744).

Flannigan, M., Stocks, B., & Wotton, B. (2000). Climate change and forest fires. *Science of the Total Environment*, *262*, 221–229.

Fromm, M. D., Bevilacqua, R., Stocks, B., & Servranckx, R. (2004). New directions: Eruptive transport to the stratosphere: Add fire-convection to volcanoes. *Atmospheric Environment*, *38*, 163–165.

Fromm, M. D., & Servranckx, R. (2003). Transport of fire smoke above the tropopause by supercell convection. *Geophysical Research Letters*, 30.

Garay, M. J., Mazzoni, D. M., Davies, R., & Diner, D. (2005). The application of support vector machines to the analysis of global datasets from MISR. *Proceedings of the Fourth Conference on Artificial Intelligence Applications to Environmental Science, San Diego, CA*.

Kahn, R. A., Li, W. -H., Moroney, C., Diner, D. J., Martonchik, J. V., & Fishbein, E., in press. Aerosol source plume physical characteristics from space-based multi-angle imaging. *Journal of Geophysical Research*.

Kaufman, Y., & Justice, C. (1998). *MODIS fire products algorithm technical background document (ATBD), version 2.2.* http://modis.gsfc.nasa.gov/data/atbd/

Lavoue, D., Liousse, C., Cachier, H., Stocks, B. J., & Goldammer, J. G. (2000). Modeling of carbonaceous particles emitted by boreal and temperate wildfires at northern latitudes. *Journal of Geophysical Research*, *105*, 26871–26890.

Mazzoni, D., Garay, M. J., Davies, R., & Nelson, D. (2006). An operational MISR pixel classifier using support vector machines. *Remote Sensing of Environment*, *107*, 149−158. doi:10.1016/j.rse.2006.06.021 (this issue MISR special issue).

Moroney, C., Davies, R., & Muller, J. -P. (2002). Operational retrieval of cloud-top heights using MISR data. *IEEE Transactions on Geoscience and Remote Sensing*, *40*, 1532−1540.

Muller, P., Madanayake, A., Davies, R., Diner, D., & Paradise, S. (2002). MISR stereoscopic image matchers: Techniques and results. *IEEE Transactions on Geoscience and Remote Sensing*, *40*, 1547−1559.

National Interagency Fire Center (NIFC). 2004 statistics and summary. http://www.nifc.gov/nicc/predictive/intelligence/2004_statssumm/2004Stats&Summ.html

Naud, C., Muller, J., Haeffelin, M., Morille, Y., & Delaval, A. (2004). Assessment of MISR and MODIS cloud top heights through inter-comparison with a back-scattering lidar at SIRTA. *Geophysical Research Letters*, *31*, L04114.

Pfister, G., Hess, P. G., Emmons, L. K., Lamarque, J. -F., Wiedinmyer, C., Edwards, D. P., et al. (2005). Quantifying CO emissions from the 2004 Alaskan wildfires using MOPITT CO data. *Geophysical Research Letters*, *32*, L11809.

Roy, D. P., Jin, Y., Lewis, P. E., & Justice, C. O. (2005). Prototyping a global algorithm for systematic fire-affected area mapping using MODIS time series data. *Remote Sensing of Environment*, *97*, 137−162.

Singh, H., Brune, W., Crawford, J. & Jacob, D. (in press). Overview of the summer 2004 Intercontinental Chemical Transport Experiment—North America (INTEX-A). http://www.espo.nasa.gov/docs/intex-na/Singh_INTEX-A%20Overview%20Paper.pdf

Singh, H., Jacob, D., Pfister, L., & Crawford, J. (2002). *INTEX-NA: Intercontinental Chemical Transport Experiment—North America.* http://cloud1.arc.nasa.gov/intex-na/

Zong, J., Davies, R., Muller, J., & Diner, D. S. (2002). Photogrammetric retrieval of cloud advection and top height from the multi-angle imaging spectroradiometer (MISR). *Photogrammetric Engineering and Remote Sensing*, *68*(8), 821−830.