

# HYPERSPECTRAL SULFUR DETECTION USING AN SVM WITH EXTREME MINORITY POSITIVE EXAMPLES ONBOARD EO-1

Lukas Mandrake<sup>(1)</sup>, Kiri L. Wagstaff<sup>(1)</sup>, Damhnait Gleeson<sup>(1)(2)</sup>, Umaa Rebbapragada<sup>(3)</sup>, Daniel Tran<sup>(1)</sup>,  
Rebecca Castaño<sup>(1)</sup>, Steven Chien<sup>(1)</sup>, Robert T. Pappalardo<sup>(1)</sup>

<sup>(1)</sup>*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive  
Pasadena, CA 91109, <first.lastname>@jpl.nasa.gov*

<sup>(2)</sup>*Laboratory for Atmospheric and Space Physics, University of Colorado at Boulder, Boulder, CO 80303*

<sup>(3)</sup>*Tufts University, Computer Science Department, 161 College Ave, Medford, MA 02155, urebbapr@cs.tufts.edu*

*Abstract*—Onboard classification of remote sensing data is of general interest given that it can be used as a trigger to initiate alarms, data download, additional higher-resolution scans, or more frequent scans of an area without ground interaction. In our case, we study the sulfur-rich Borup-Fiord glacial springs in Canada utilizing the Hyperion instrument aboard the EO-1 spacecraft. This system consists of naturally occurring sulfur-rich springs emerging from glacial ice, which are a known environment for microbial life. The biological activity of the spring is associated with sulfur compounds that can be detected remotely via spectral analysis. This system may offer an analog to far more exotic locales such as Europa where remote sensing of biogenic indicators is of considerable interest. Unfortunately, spacecraft processing power and memory is severely limited which places strong constraints on the algorithms available. Previous work has been performed in the generation and execution of an onboard SVM (support vector machine) classifier to autonomously identify the presence of sulfur compounds associated with the activity of microbial life. However, those results were limited in the number of positive examples available to be labeled. In this paper we extend the sample size from 1 to 7 example scenes between 2006 and 2008, corresponding to a change from 18 to 235 positive labels. Of key interest is our assessment of the classifier's behavior on non-sulfur-bearing imagery far from the training region. Selection of the most relevant spectral bands and parameters for the SVM are also explored.

## 1. MOTIVATION

The advent of satellite-based remote sensing has permitted vast swaths of otherwise difficult-to-access Earth terrain to come under close, regular scientific scrutiny. Of particular interest are deserts, ice sheets, mountain ranges, and other nearly or totally inaccessible locales [1][2][3]. These studies and others yield invaluable data on natural and human-induced climate change, seasonal variation, global-scale dynamics, and a

host of other relevant inquiries. In the realm of missions to other planets, remote sensing enables the otherwise impossible goal of analyzing large, inhospitable regions too distant for direct human observation and study. However, the same challenges remain no matter the application at home or among the stars: remote observation generates far more data than can ever be successfully downlinked. For non-terrestrial missions, the requirement of time from the Deep Space Network creates severe financial and political burdens that are often responsible for the decision to end a mission on an otherwise still functioning spacecraft.

Several approaches have been employed to address this limitation. Data compression was legendarily utilized to rescue the science return of the Galileo spacecraft (high gain antenna deployment failure) [4] and continues to be a mainstay of satellite communication. Cooperation between in-situ spacecraft to relay data over the horizon or to boost transmission bandwidth has now also become possible [5]. Yet even as we increase our data throughput capability we are also increasing the capacity for our sensors to generate truly enormous data sets. Spacecraft now carry hyperspectral cameras capable of sensing hundreds of wavelengths simultaneously in a streaming, real-time manner [6]. Despite all this data production and transmission capability, the majority of data collected by any streaming camera is redundant or otherwise uninteresting after a region has been initially scanned. Instead, key regions with unusual or dynamic features may be selected as high priority while static areas, once surveyed, may receive lower priority. Time-based events may be predicted and captured in similar manner. But often, events are either unpredictable or not known to exist prior to discovery. Thus an onboard detector capable of deciding when a region is interesting based on predefined criteria would yield a much higher science return. Many such detectors have been created for a variety of specialized applications such as floods [7], volcanic eruptions [8], and novel image detection aboard Martian rovers [9]. In our case, we are supporting NASA's mission to discover living systems and their

evidence by creating a remotely sensed detector for sulfur compounds.

Prior work on precisely this question has been hampered by a lack of available labeled hyperspectral imagery [10]. We will be extending these results by including two orders of magnitude more data for both test and training sets.

## 2. TARGET: BORUP FIORD, ELLESMERE ISLAND

Set in the far north of the Canadian Arctic Archipelago (81°N, 81°W), Ellesmere Island abuts the northern edge of Greenland between the Lincoln Sea and Baffin Bay. Like other ice-covered terrain in our solar system, it does not appear at first glance to be particularly hospitable to life. Yet from the air can be seen vivid yellow stains along a glacial edge. Discovered by Benoit Beauchamp in the mid 1990's who also collaborated on later field investigations with Stephen Grasby, Damhnait Gleeson, and Marie-Eve Caron in 2006, these discolorations were identified as Supraglacial deposits associated with sulfur-rich springs flowing through and across a 200-meter thick glacier [11]. The telltale yellow stain was due to the presence of elemental sulfur, which dominates the composition of the deposits. The source of these minerals most likely relates to marine evaporite deposits of anhydrite ( $\text{CaSO}_4$ ) in the region's bedrock. Unfortunately, the various deposition structures and spring locations greatly varied between field campaigns. As the Fiord is only accessible for three months of the year, the activation cycles of the springs during the remaining year are not known nor are the conditions that spur their flows.

The springs themselves are interesting scientifically based only on the above, but more fascinating still is the presence of elemental sulfur, gypsum, and hydrogen sulfide gas. These three compounds contain sulfur in three different oxidation states. The presence of a complex redox system, in addition to such an abundance of elemental sulfur out of equilibrium with its environment, provided an indication that microbial mediation of the local geochemistry could be occurring. Testing the water from all ten known springs detected a rich microfauna of known and as yet unidentified microbes [12]. This arrangement of fissured permanent ice, sulfur-bearing subsurface rock, and seeping water is potentially analogous to the surface of Jupiter's moon Europa. Frigid beyond any place on Earth, Europa still demonstrates surface sulfur-bearing mineral taint on otherwise solid ice along fractures [13] The reddish-tinged lenticulae (round-shaped potential upwelling) and long double-lined fissures speak of a subsurface ocean likely rich with dissolved minerals [14]. Studying the microbes surviving beneath the Borup Fiord glacier might therefore also shed light on potential habitats

beneath Europa's surface. As elemental sulfur occurs naturally generally from either volcanic or anaerobic bacterial processes, producing a remote elemental sulfur detector is a step towards a microbial life detector both for the cryosphere of Earth and potentially for Europa as well.

## 3. INSTRUMENT: HYPERION ON EO-1

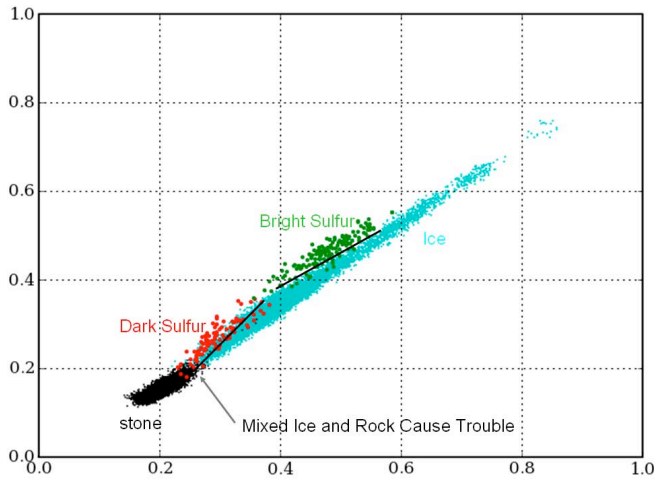
The Hyperion imager aboard the near-polar-orbiting Earth-Observing-1 spacecraft (EO-1) orbiting at 700 kilometers altitude resolves the Earth's surface to 30 meter resolution simultaneously in 220 distinct spectral bands ranging from 0.38 to 2.5  $\mu\text{m}$  (near ultraviolet to short-wave infrared) [6]. The first hyperspectral sensor to operate from space, it captures 7.5 kilometers by  $\sim 100$  kilometers of land area with each image. Designed to assist mining, geology, forestry, agriculture, and environmental management through the classification of surface type and features, it has successfully generated over 27 Tb of raw data available as various products from the USGS.

Although EO-1 has 220 bands of hyperspectral data available via the post-processed, downloaded products, onboard operation only has access to 12 bands at a time. Thus, any autonomous algorithm must select the 12 most relevant bands to the inquiry at hand. Selecting which of the 220 bands to examine is of paramount interest. One consideration is the existence of decreased signal to noise ratios at increased wavelength (band number) [15]. We expect therefore that our selection process will likely favor shorter wavelength bands.

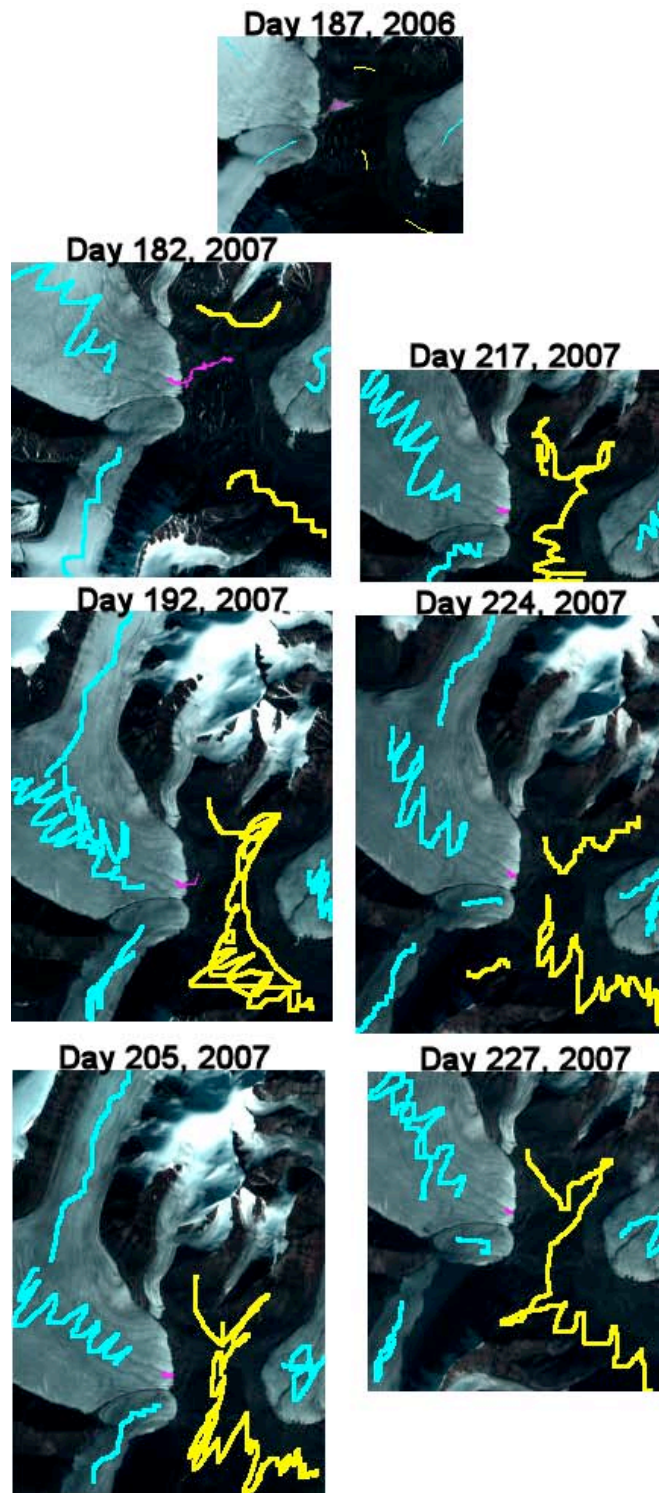
## 4. DATA PROPERTIES

We used seven flyover images of Borup Fiord as our test and training dataset (Table 1). Individual pixels were labeled by Damhnait Gleeson using the commercial ENVI toolkit, aerial photos, and the PixelLearn image labeling software developed by the Machine Learning Group at the Jet Propulsion Laboratory [16]. Each image is 256 pixels wide and of variable length (mean 3200 pixels) with 220 frequency bands available for each pixel. See Table 1 for the number and type of labels used in each image. We utilize the L1R Hyperion product [17] as it has been processed to reflectance (as opposed to radiance) to help mitigate such effects as sun angle. Further, we have taken seven similarly sized flyovers of regions far from the sulfur-bearing glacial springs that we label "Sulfur-Free" as a control for false positives, as shown in Table 2. Graphical versions of these provided labels are provided in Figure 1. Original results on this study were provided based entirely on training/testing utilizing the single 2006 image (Figure 1 top) [10].

We can show without any processing that distinction between sulfur and non-sulfur (ice and rock) pixels will be challenging simply by selecting two arbitrary wavelength bands and plotting the labeled data to investigate the degree of separation as shown in Figure 2. The sulfur pixels are well mixed with the ice pixels with some overlap with the rock pixels as well. Further, two subpopulations are evident in the pixels labeled as sulfur; we distinguish these as “Bright Sulfur” and “Dark Sulfur” and will discuss this distinction further in Section 6. Note that the “Dark Sulfur” points are especially easy to confuse with rock and similarly dark ice.



**Figure 2.** All labeled data (pixels) shown as a scatter plot with two arbitrary wavelengths as the axis. The black separation lines are drawn by hand for illustration only.



**Figure 1.** Expert-provided labels for source images of Borup Fiord. Springs are at the edge of the glacier. (yellow = rock, cyan = ice, violet = sulfur)

**Table 1.** Test / Train Data Source Detail from Borup Fiord

Scene ID	X size (zonal)	Y Size (meridional)	Sulfur Labels	Ice Labels	Rock Labels
EO1H0570012006187110PY	256	3399	18	75	58
EO1H0570012007182110KF	256	3519	110	1322	579
EO1H0570012007192110KF	256	3176	35	3016	2357
EO1H0570012007205110KF	256	3455	25	2146	2067
EO1H0570012007217110KF	256	3460	22	1643	1186
EO1H0570012007224110KF	256	3183	19	2027	1712
EO1H0570012007227110PF	256	3247	16	1800	1526

**Table 2.** “Sulfur-Free” Data Source Detail far from any sulfur source

Scene ID	X size (zonal)	Y Size (meridional)	Location
EO1H0201122008017110KF	256	3242	Getz Ice Shelf, Antarctic
EO1H0270062008086110KF	256	3176	Tyr 76, Arctic
EO1H0300052008101110KF	256	3187	Thule AFB, Arctic
EO1H0451162006327110PF	256	3242	Ross Ice Shelf 1, Antarctic
EO1H0630112006093110KF	256	3399	Tuktoyaktuk, Arctic
EO1H0692442004121110KZ	256	6587	War Hunt Ice Shelf, Arctic
EO1H2161062007085110KF	256	3410	Larsen Ice Shelf, Antarctic

## 5. METHODS

*Pixel Classification: Support Vector Machines (SVMs)*

Key to all our methodology here will be the use of the SVM [18]. This method was originally selected by Castano et al [10] based on empirical performance on Hyperion hyperspectral data and the substantial expertise available for consultation as well as their excellent performance for linear kernel implementations. These classifiers attempt to construct a hyperplane (in our case, within a 12 dimensional space of reflectances) that separates two classes of previously labeled data. A linear kernel SVM is defined by  $n+1$  parameters: a weight ( $w_i$ ) for every training example ( $x_i$ ) and a scalar bias term ( $b$ ). These weights and bias are determined during the training step by solving a quadratic programming problem with one degree of freedom: the regularization parameter  $C$  which determines how strongly to penalize training errors. Classification of a new example  $x$  is achieved by:

$$F(x) = \text{sign}\left(\sum_{i=1}^n w_i (x_i \cdot x - b)\right)$$

where the linear dot product can be replaced by a kernel function  $K(x_i, x)$  that implicitly maps each point into a new (possibly infinite dimensional) feature space in which linear separability is achievable. For this study, we examined both a linear kernel (simple dot product) and a Gaussian kernel parameterized by  $\gamma$ , the Gaussian width:

$$K(x_i, x) = e^{\left(\frac{-\|x_i - x\|^2}{\gamma}\right)}$$

which generally provides higher accuracy, due to increased representation power, and is an example of what we might be able to achieve given more spacecraft resources than generally available on EO-1.

*Feature Selection*

To accommodate the restrictions imposed in onboard computation, we must select 12 wavelength bands from the potential 220 available for Hyperion data. This is a classic feature selection problem, and we will compare the following methods of band selection:

1) Greedy Forward Feature Selection [18]: Starting with the empty set, train a classifier based on a currently accepted set of bands and test performance as baseline. Then exhaustively add each band one at a time evaluating freshly trained classifiers for each new set. Select the one with the highest performance. Continue until 12 are selected. This procedure is known to be weak against the case of pairs or larger sets of data that must be admitted together to add information to the classifier.

2) Recursive Feature Elimination (RFE) [18]: Starting with the entire 220 available bands, train a classifier on this accepted set and calculate performance. Then exhaustively remove each remaining band one at a time evaluating freshly trained classifiers for each new set. Select the one with the highest remaining performance.

Continue until only 12 bands remain. This procedure incorporates groups of bands which together are meaningful but alone are less relevant, though it still does not explore the full space of possibilities.

3) Domain expert manual selection: Request a knowledgeable party (Damhnait Gleeson) to guess, calculate, or intuit which 12 bands “should” be important for classifier generalization based on science, intimate instrument knowledge, and past experience.

4) RFE based only on the 2006 data [10]: Same as (2) above, but using only the single source image from 2006. This is all that was previously available as labeled data for this region and is included for comparison.

For each of these four sets of potential bands of interest, we trained linear kernel 2-class (sulfur / non-sulfur), 3-class (sulfur, ice, rock), and 4-class (bright/dark sulfur, ice, rock) SVMs [19]. The linear kernel was utilized for computational speed and simplicity due to the constrained speed and integer arithmetic of onboard operations. Statistics on success were measured via per-image cross validation, where one of the seven source images in Figure 1 / Table 1 was used as a held-out test set and the others were used to train the SVM.

We evaluated the impact of a large range of different values for  $C$ .  $C$  trades off penalties for errors on the training data against model complexity; small values permit better generalization while large values more accurately fit the training data. The “best” result will be some set of bands (one of the four sets selected above),  $C$  value, and number of permitted classes. This configuration can then be coded and uploaded into EO-1 for autonomous sulfur detection.

#### *Labeled Data Filtration*

In our search for a highly accurate classifier, we allowed the possibility that our labeled examples were less than pure. Labeling these pixels manually is a challenging task even for experts. Marginal, or even incorrect labels may be present especially given the coarse spatial resolution of the labels (Figure 1). Umaa Rebbapragada has constructed a label confidence evaluation method based on averaging across multiple EM clustering’s of the labeled data named Pair-Wise Expectation Maximization (PWEM) [20][21]. In its general formulation, PWEM suggests that improved results may be obtained by permitting a reweighting of input labels based on these results; however, in our case we perform the more coarse procedure of filtration of any labeled example less than 75% consistent with others of its same class.

The final part of this study reports on the use of a Gaussian kernel, an advance over the linear kernel. A

second hyper-parameter,  $\gamma$ , must be swept across for generalization testing in addition to the standard SVM  $C$ . We do not anticipate being able to upload an algorithm to EO-1 based on a Gaussian, but it is interesting to examine just how much better such a classifier may perform. Future missions are likely to have increased capability and to be able to make use of Gaussian SVMs.

## 6. RESULTS

### *Band Selection*

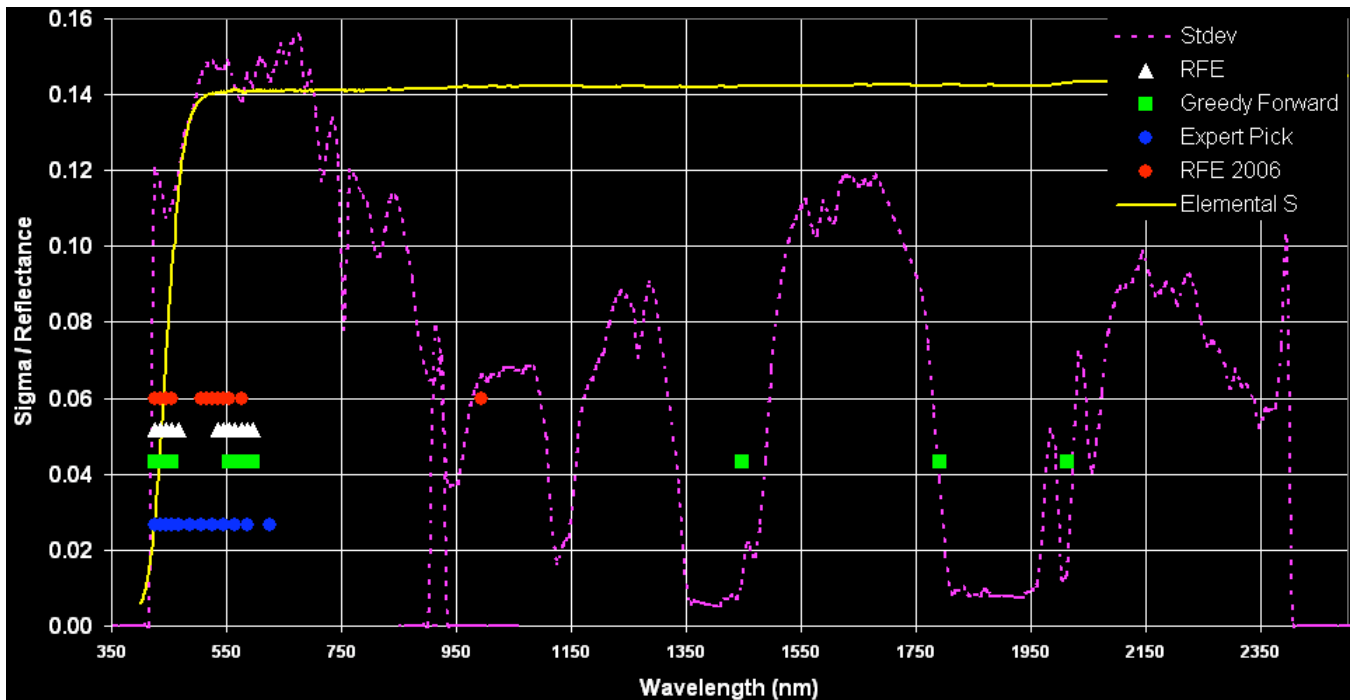
The bands of interest generated by the three feature selection methods are shown in Figure 3 and Table 3. The standard deviation was calculated across all labeled examples for a given wavelength (2585 measurements for each of the 200 bands). The overall standard deviation pattern indicates the sensitivity of the Hyperion instrument [15] and is largely influenced by atmospheric absorption properties.

Spectrum reflectance values should theoretically vary only between 0 (no reflectance) and 1 (total reflectance). However, as the wavelength increases so does the noise, eventually yielding obviously erroneous inclusions well outside this range. Unfortunately, the greedy forward selection method preferred some higher wavelengths that proved detrimental to its generalization later in the process. Due to computational limitations on runtime, the forward method performance was not pursued beyond this point.

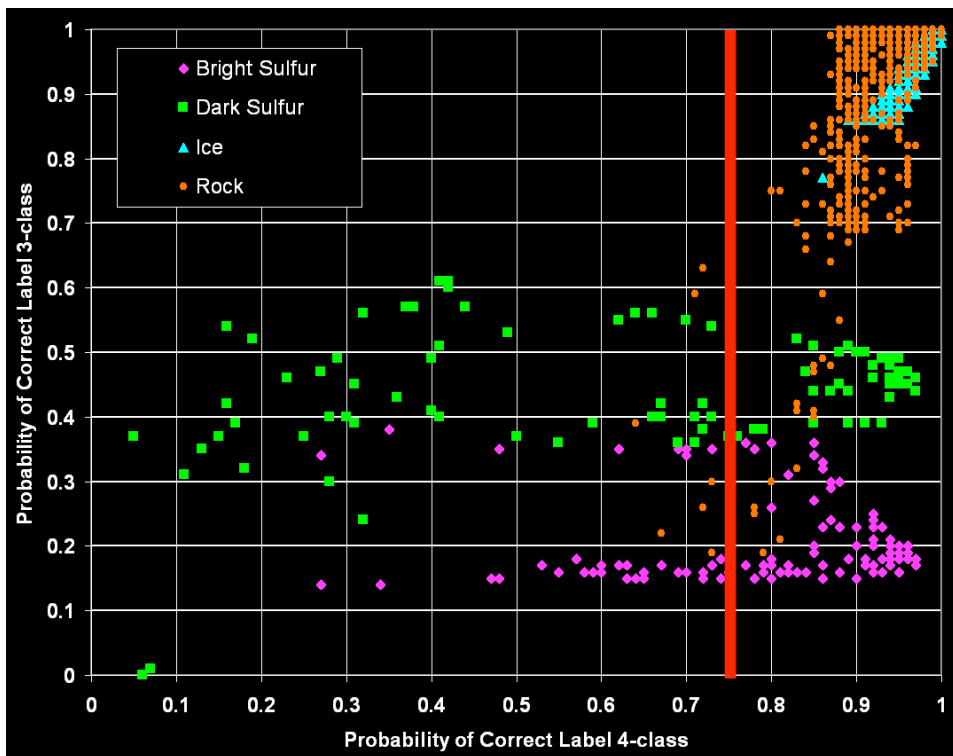
**Table 3. Band Selection Results ( $\lambda$  nm)**

<b>RFE 2006 only</b>	<b>RFE</b>	<b>Greedy Forward</b>	<b>Expert Pick</b>
426	426	426	426
436	436	436	436
446	446	446	446
456	456	456	456
506	466	556	466
516	536	566	486
526	546	576	506
536	556	586	526
546	566	596	546
556	576	1448	566
576	586	1791	586
994	596	2013	626

We should also consider the emitted spectrum of our target material. Elemental sulfur [22] has a rather unusual reflectance spectrum, ranging from 7% below 430 nm to ~90% above 500 nm. Other than this transition between low and high, there is no structure to its spectra. Thus, we expect detection to be very challenging. Figure 3 superimposes this reflectance over the selected bands normalized to 100% reflectance at top of graph.



*Figure 3. Band analysis including standard deviation across all labeled examples and acceptance by a selection method. Selection method picks have been placed with an arbitrary Y value on the graph for illustration only. The "Elemental S" line shows the % reflectance of elemental sulfur with 100% normalized to the top of the graph.*



*Figure 4. PWEM estimate of label correctness. Everything left of the red line was filtered as being too uncertain.*

### Expanding from 3 to 4 Classes

An early attempt at this analysis showed abundant false positives (FP) using the entirety of the sulfur training set and a 3-class SVM ( $\sim 10^5$  FP events per downlinked image). We suspected that there were multiple populations (including potentially false labels) within our sulfur training set. We applied a K-means clustering method [23] to the sulfur labels that immediately yielded two distinct populations: one with high mean reflectance and one low. We divided our sulfur labels according to the clustering results, yielding four classes, and discovered that most of the false positives were from the “dark” sulfur class. Therefore, although this division was not obtained via manual labels, it still provided a very useful refinement of the labeled data and, as we will show, led to improved performance. Physically, sulfur that occurs within rock rich pixels is extremely difficult to distinguish from sulfur-free rock-rich pixels. Table 4 shows the breakdown per image of bright and dark sulfur labeled examples.

**Table 4.** Test / Train Data Source Detail from Borup Fiord

Scene ID	Bright Sulfur Labels	Dark Sulfur Labels
EO1H0570012006187110PY	17	1
EO1H0570012007182110KF	26	74
EO1H0570012007192110KF	22	13
EO1H0570012007205110KF	22	3
EO1H0570012007217110KF	22	0
EO1H0570012007224110KF	19	0
EO1H0570012007227110PF	16	0

### Filtering Poorly Labeled Examples

Armed with our three band sets of interest (RFE 2006, RFE, and Expert Pick), we now investigate the mislabeling comparison previously discussed. PWEM [20] identifies for each labeled example the probability that it was correctly labeled. Figures 4 & 5 show the distribution of these percentages for both the 3-class and 4-class labelings. Ice labels proved to be the most reliable, which is not surprising since selecting pure ice examples in the image is quite easy. Rock was similarly well labeled with a few remote outliers. However, sulfur was highly contentious. Examining the Y-axis of Figure 4 (3-class labeling), we see not a single sulfur label exceeds 0.75 probability of being correct. Separating into bright and dark (X-axis) we immediately rise to 104 / 144 bright sulfur labels as above 0.75 confident, and 39 / 91 dark sulfur labels above 0.75 confident. At this point, we discarded any label not above 0.75 confident as too uncertain for use. We will report these results as “4class filtered.” This necessary step also has the unfortunate effect of removing 40% of our already very limited

positive examples. However, again, we will see that this leads to an improvement in performance. The low-confidence pixels tended to occur near ice-rock boundary areas. The 0.75 filter limit was determined simply by empirical observation that the “clear” cases of ice and rock are rarely below this value while both sulfur classes span the entire range of probability. Any more stringent requirement would have seriously jeopardized our already limited positive example labeled data.

### SVM Performance Metrics

Let us now define our metrics. The X-axis of all comparison graphs is the F-measure (F) [24] as calculated by the harmonic mean of recall (R) and precision (P):

$$S_{Correct} = |L_B \cap (D_B \cup D_D)|$$

$$S_{False} = |D_B \cap (L_I \cup L_R)|$$

$$S_{Missed} = |L_B \cap (D_I \cup D_R)|$$

$$P = \frac{S_{Correct}}{S_{Correct} + S_{False}}$$

$$R = \frac{S_{Correct}}{S_{Correct} + S_{Missed}}$$

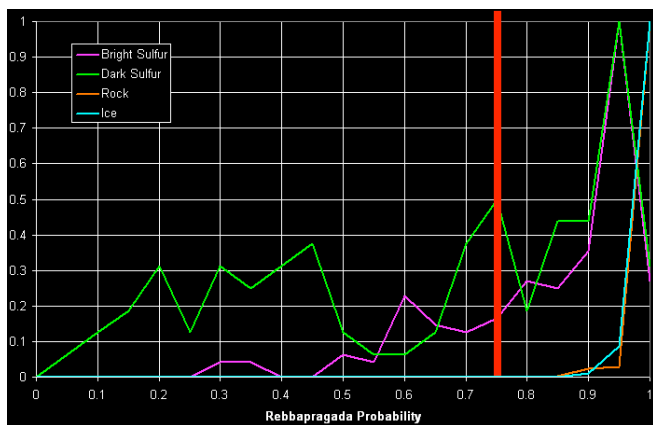
$$F = \frac{2PR}{P + R}$$

where  $L_B$  are pixels labeled Bright Sulfur,  $L_D$  are labeled Dark Sulfur,  $L_I$  are labeled Ice, and  $L_R$  are labeled Rock. Similar nomenclature is used for D to refer to detected (SVM classified) pixels.  $S_{Correct}$  is the total number of correct detections of bright sulfur (or dark sulfur as bright sulfur, to give the benefit of the doubt),  $S_{Missed}$  is the total number of failed detections of labeled bright sulfur (identified as anything but bright sulfur), and  $S_{False}$  is the total number of incorrect bright sulfur identifications (ice or rock identified as bright sulfur). We also compute two other statistics, both addressing false positives identified outside the original labeled pixels. If sulfur is detected in any unlabeled pixels within the source images, these are considered “likely false positives” and referred to as Likely FPOS on the graphs. The name derives from the fact that we cannot be certain there are not other sulfur sources elsewhere in this image, but they certainly should be small in number compared to the entire image. The second statistic of this type we call “Sulfur-Free FPOS” as it represents the number of pixels in the seven new scenes taken very far from the known sulfur source

(images of clouds, opposite hemisphere ice sheets, frozen ocean, etc.). Any positives found in the Sulfur-Free images are simply mistakes.

For each configuration, we train on six of the seven labeled images and use the seventh as test, then rotate through each of the images as test for cross-validation. The precision, recall, F-measure, “Likely FPOS” and “Sulfur-Free FPOS” are then computed from the appended list of all tested pixels, i.e. precision, recall, and F-measure are calculated at the end of the cross validation, not for each cross-validated image.

For all linear kernels, we evaluate the SVM hyper-parameter  $C \in \{10^{-1}, \dots, 10^5\}$  over 36 values ranging logarithmically. For the Gaussian kernel example, we use these same  $C$ 's but also vary the Gaussian width parameter  $\gamma \in \{10^{-2}, \dots, 10^3\}$  by factors of ten. These ranges spanned the empirically observed dynamic region between numeric instability and the SVM's total failure to generalize. The number of intermediate values ( $36 \times 6$ ) was chosen based on the machine time available using JPL's Nebula supercomputer.



**Figure 5.** Histogram of label correctness. Everything left of the red line was filtered. Note that the histograms have been normalized to 1 for comparison despite each class being substantially different in number of labeled examples

### SVM Performance

Figure 6 shows the performance of the SVMs for various class numbers and kernel type, and mislabeled filtration strategies. On the right we see the Sulfur-Free FPOS results along the Y axis, while on the left we find the Likely FPOS results. Note that there are always more Likely FPOS than there are Sulfur-Free FPOS. This makes sense, as the Likely FPOS is a harder problem (separate sulfur from nearby sulfur-like terrain) than separating entirely non-sulfurous terrain from sulfur-bearing surroundings. On the X axis in both plots is the F-measure, with each point representing a given SVM configuration of  $C$ , class number, set of 12 wavelength

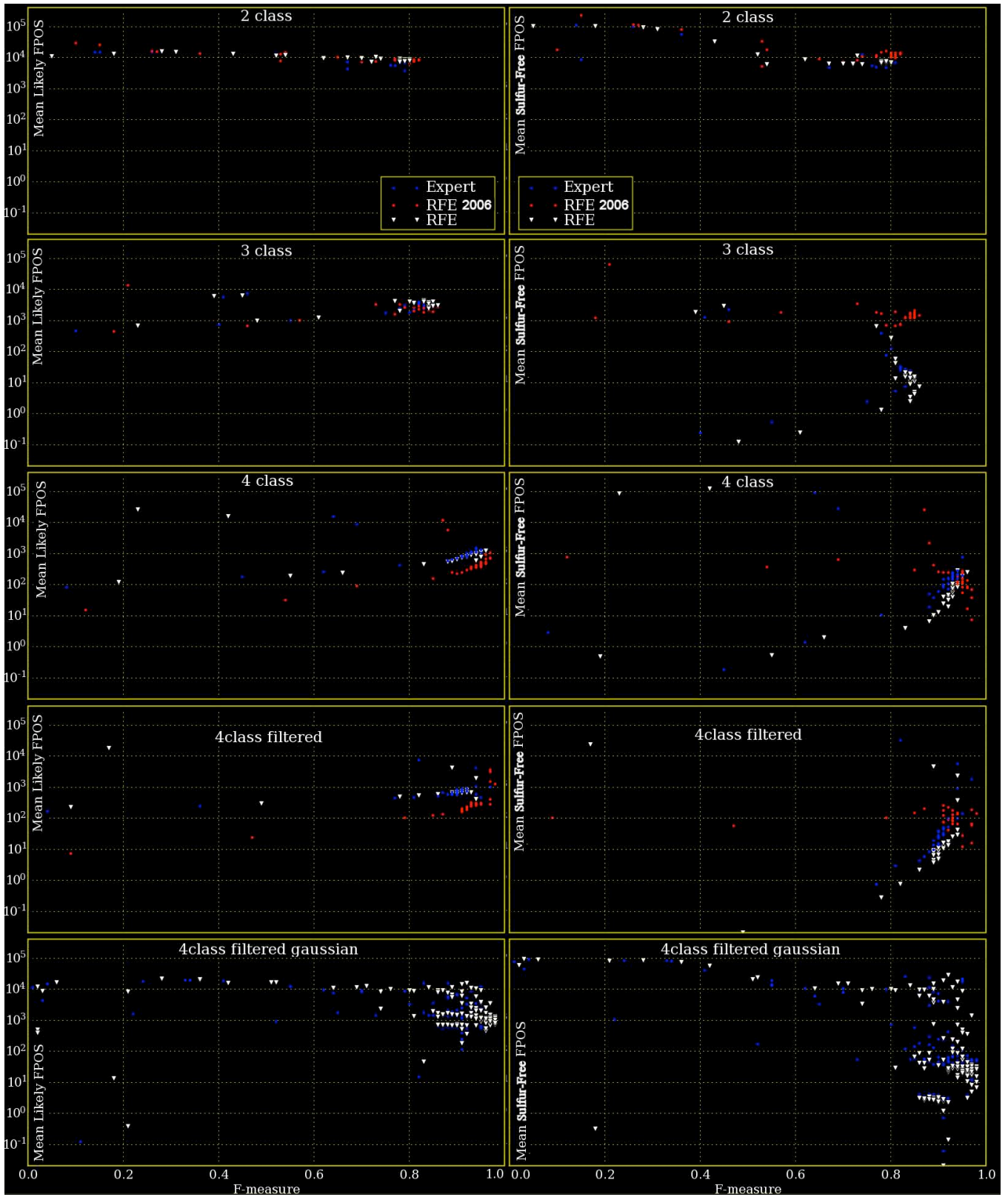
bands, and in the case of the Gaussian kernel the width hyper-parameter  $\gamma$ . We have separated the sets of 12 bands (Expert, original RFE result using only 2006 image, and current RFE result) into blue, red, and white points respectively. The optimal results would be points lying at the bottom-right of each graph with zero Likely / Sulfur-Free FPOS (Y) and F-measure of one (X). Please note the y-axis is a log scale in number of false positives.

The linear kernel 2-class (sulfur / non-sulfur) SVM shows poor performance. While several configurations achieve an impressive F-measure of up to 0.80 - 0.82 for all three band selections, there are more than  $10^4$  Sulfur-Free FPOS. Similarly poor generalization occurs in the Likely FPOS of the unlabeled examples from the source image. Thus, though the labeled data was fit, it failed to generalize to more remote examples.

The linear kernel 3-class (ice, rock, and sulfur) SVM shows immediate improvement in all three band selections with remarkable improvement in the Expert picked bands and the new RFE results. Sulfur-Free FPOS were reduced by two orders of magnitude to a mean of  $\sim 40$  FPOS per Sulfur-Free image. Likely FPOS remained high, however, with a mean of  $\sim 3400$  per image. The mean and maximum F-measure also increased with a new top achievement of 0.86.

The linear kernel 4-class (ice, rock, “bright” sulfur-on-ice, and “dark” sulfur-on-rock) SVM demonstrated two immediate consequences relative to the 3-class performance. The first was that the Sulfur-Free FPOS actually increases to a mean of  $\sim 70$  pixels per image. Likely FPOS was dramatically reduced yielding a mean of  $\sim 800$  false positives per image. Most of the later improvement was due to the decision to define all sulfur on rock “dark sulfur” labels as unreliable and no longer count them towards our accuracy requirement. The mean F-measure increased further with a mean of  $\sim 0.9$  and a maximum of 0.96. As we increased our performance by filtering out a portion of our positive labeled instances, a more rigorous method to remove contaminated training examples became attractive.





**Figure 6.** SVM performance for 2-4 classes, with and without mislabel filtration, using linear and Gaussian kernels. X axis are F-measure while y-axis are FPOS from the same training image (Likely) or from distant non-sulfur-bearing images (Sulfur-Free). Note the Y axis is a log scale.



**Figure 7.** Location of "Likely False Positives." Note the geographic separation of bright/dark sulfur in the middle frame (violet/green) shows the origin of the bimodal population among sulfur. Red is a false positive.

Employing PWEM's confidence estimation method on the 4-class data and rejecting anything less than 0.75 confident, the mean F-measure of 0.90 was maintained while reducing the mean Likely FPOS to ~600 per image. Sulfur-Free FPOS reduced further to a mean of 24 pixels per image. We take these results to indicate that mislabeled training examples were indeed corrupting our results and generating false positives. While further filtration of example labels might in principle yield further improvements, we have already discarded 40% of our very limited positive examples for the gains shown here and further filtration is unlikely to improve performance. For EO-1 Hyperion effort, we would select one of the blue or white indicators on the graph representing a particular combination of band (Expert Pick or RFE) and C hyper-parameter (3-20 is roughly comparable in performance). This linear kernel SVM would then be uploaded for operation. Unfortunately, we still expect ~500 false detections on images near Borup Fiord.

Extending these results to Gaussian kernels (using only Expert and RFE bands) yields the last set of graphs. Note that both C and the Gaussian width ( $\alpha$ ) are hyper-parameters here resulting in 252 configurations per band selection set instead of 36 for a linear kernel. The Gaussian SVMs achieved a mean F-measure of 0.93 and maximum F-measure of 0.98 using the 4-class filtered dataset, exceeding the best performance of the linear kernel. Mean sulfur-Free FPOS remains at ~ 50 pixels per image, while mean Likely FPOS remains high at again mean ~1000 detections per image. However, certain parameter pairings do show significant promise.

#### *False Positives in the Training Images "Likely FPOS"*

Our inability to extinguish unwanted false positives on the training images is troubling. A best result image using the Gaussian kernel results with C = 0.25, Gaussian width = 1000, and RFE band selection showing the location of these false detections is provided in Figure 7. The accuracy for this case was an F-measure of 0.83, 0 sulfur-free FPOS, and ~50 Likely FPOS per image. The top of the figure shows the original image (visible wavelengths only). The middle panel shows the location of the original labeled data where cyan is ice, yellow is rock, violet is sulfur on ice, and green is sulfur on rock. The last panel shows white for all labeled locations after filtration based on PWEM's confidence estimation and red for "bright sulfur" false detections that were reported as Likely FPOS. We immediately observe a correlation between small icy valleys, glacier edges, and false detections. Thus, the filter has indeed learned to recognize terrain similar to where sulfur occurs but still has trouble detecting the (very faint) yellow tint to the sulfur-bearing region.

This points out a weakness in this multiclass SVM application, namely that if a user provides examples of clean, unambiguous areas (such as our ice and rock labels) but then specifies a minority class within ambiguous background detail (such as sulfur only existing in ice-rock mixture pixels), the filter may accurately (according to the training data) begin to predict that mixes of ice and rock are good determinants of sulfur presence. Thus, we have constructed both a sulfur-detector and a rock-ice mixture detector at the same time.

## 7. CONCLUSIONS

Overcoming the practical challenges of autonomous detection oftentimes exceeds in difficulty the mathematical treatment or data interpretation itself. In our case, we faced erroneously labeled examples, a bimodal population within our labeled sulfur examples, a single geographic sulfur region for training, a limitation of only 12 out of 220 bands to examine, and restriction to a linear kernel. Even extending to the Gaussian kernel, we found our ability to remove false detections in highly similar (nearby) terrain to that which housed the sulfur example was lacking, resulting in at best ~50 false detections throughout a single mean image measuring ~ 256 x 3200 corresponding to 0.006% false alarm rate. This is significantly superior to the figure cited by Castano of 0.2% based only on 2006 data. While these may seem like small numbers, we are attempting to detect small events isolated in frequency and region size. Thus false positives are very costly should such a detector be utilized to schedule additional satellite observation runs or downlink data.

However, in the development and evaluation of this classifier we did determine that while three classes of terrain was a good fit to the data (ice, rock, and sulfur), to eliminate the majority of false positives especially on data far from the source images we were forced to carefully prune our input labels using filtration based on the PWEM estimation method. We also recognized that of the bimodal sulfur population, the “dark sulfur” class (sulfur occurring on rock) was the majority of the false positive source. By discarding the dark class as unreliable, we obtained significant improvement in our false positive rate overall but especially in distant imagery where less than one event in a mean 256 x 3200 image is expected. All of this was done while preserving a representative F-measure of ~ 0.9 with regard to the filtered labels.

Of feature selection methods, we demonstrated that Recursive Feature Elimination (RFE) remains an excellent tool to pare down uninformative, corrupted, or noisy satellite wavelengths when constructing such a detector, empirically better than a greedy forward method. We also showed that the addition of

substantially more data than past work yielded a much more successful classifier (using bands selected by RFE vs. RFE 2006 only). Still, lack of positive labeled examples remains the most significant source of error. In particular, labeled sulfur sources that do not all geographically collocate would greatly enhance our generality and help to reduce false positives.

The use of a Gaussian kernel did demonstrate a higher F-Measure than was possible with the linear kernel with strong reduction of the FPOS. This indicates that when it becomes possible to operate a Gaussian kernel aboard a spacecraft such as EO-1, it should be pursued. However, the CPU time cost for such algorithms must always be weighed against their improvement in performance. We are currently investigating an implementation for the EO-1 spacecraft.

A variety of options for further work on this front exist in the absence of new positive labels. One possibility would be to provide yet another class for “rock/ice mixture” separate from the sulfur positive examples to assist the classifier in avoiding nearby false positives. This would require further example labels from the domain expert that may or may not always be available. We treated our problem as though this was not an option.

A second line of attack would include nonlinearity in into the model by computing functions of the input channels (differences, averages, ratios, etc. of various wavelength bands). This may enhance the contrast of the sulfur signature or eliminate confounding input such as ice/rock mixture information.

Third, we may examine the SVM raw output directly for each classification. If the output for the “best fit” class is itself still negative, we might conclude that there is no sufficiently explanatory class and disregard the SVM’s classification. This could potentially reduce false positives dramatically.

## 8. ACKNOWLEDGEMENTS

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Copyright 2009. All rights reserved. Special thanks to the JPL Supercomputing and Visualization Facility for their support.

Expert label generation and analysis was carried out by Damhnait Gleeson under the NASA Planetary Geology and Geophysics Program grant NNX07AR28G to RTP.

## REFERENCES

- [1] R. M. Goldstein et al, "Satellite Radar Interferometry for Monitoring Ice Sheet Motion: Application to an Antarctic Ice Stream", *Science*, Vol. 262, No. 5139, pp. 1525-1530, doi: 10.1126/science.262.5139.1525, December 1993.
- [2] C. J. Tucker et al, "Expansion and Contraction of the Sahara Desert from 1980 to 1990", *Science*, Vol. 253, No. 5017, pp. 299-300, doi: 10.1126/science.253.5017.299, July 1991.
- [3] A. Rango et al, "Average Areal Water Equivalent of Snow in a Mountain Basin Using Microwave and Visible Satellite Data", *IEEE Geoscience and Remote Sensing*, Vol. 27, Issue 6, pp. 740-745, doi: 10.1109/36.35962, Nov 1989.
- [4] Kar-Ming Cheung et al, "Changing the Coding System on a Spacecraft in Flight", *Information Theory Proceedings*, ISBN: 0-7803-0878-6, Jan 1993.
- [5] R.C. Hastrup et al, "Mars Network for Enabling Low-Cost Missions," *Acta Astronautica*, vol 52 (2-6), pp 227-235, doi:10.1016/S0094-5765(02)00161-3, 2003.
- [6] S. G. Ungar et al, "Overview of the Earth Observing One (EO-1) Mission", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, No. 6, doi: 10.1109/TGRS.2003.815999, June 2003.
- [7] S. Chien et al, "Autonomous Science on the EO-1 Mission," *International Symposium on Artificial Intelligence Robotics and Automation in Space (i-SAIRAS)*, Nara, Japan, uri: <http://hdl.handle.net/2014/7155>, May 2003.
- [8] A. G. Davies et al, "Autonomous Volcanic Activity Detection with ASE on EO-1 Hyperion: Applications for Planetary Missions," *Bulletin of the American Astronomical Society*, vol. 35, May 2003.
- [9] R. Castaño, "Current Results from a Rover Science Data Analysis System," *IEEE Aerospace Applications Conference Proceedings*, doi: 10.1109/AERO.2005.1559328, Dec 27<sup>th</sup>, 2004.
- [10] R. Castaño, "Onboard Detection of Active Canadian Sulfur Springs: A Europa Analogue," *International Symposium on Artificial Intelligence Robotics and Automation in Space (i-SAIRAS)*, Los Angeles, Feb. 25-29<sup>th</sup>, 2008.
- [11] S. E. Grasby et al, "Supraglacial Sulfur Springs and Associated Biological Activity in the Canadian High Arctic – Signs of Life Beneath the Ice," *Astrobiology*, 3(3): 583-596, doi: 10.1016/S0375-6742(03)00026-8, Sep 1<sup>st</sup> 2003.
- [12] S. E. Grasby et al, "Biogeochemical Sulphur Cycle in an Extreme Environment – Life Beneath a High Arctic Glacier," *Journal of Geochemical Exploration*, vol. 78-79, pp. 71-74, doi: 10.1016/S0375-6742(03)00026-8, May 2003.
- [13] R. W. Carlson et al, "Sulfuric Acid on Europa and the Radiolytic Sulfur Cycle", *Science*, Vol. 1, No. 5437, pp. 97-99, doi: 10.1126/science.286.5437.97, Oct 1999.
- [14] C. Sotin et al, "Europa: Tidal Heating of Upwelling Thermal Plumes and the Origin of Lenticulae and Chaos Melting," *Geophysical Research Letters*, vol. 29, No. 8, doi: 10.1029/2001GL013844, 2002.
- [15] J. S. Pearlman et al, "Hyperion, a Space-Based Imaging Spectrometer," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, No. 6, doi: 10.1109/TGRS.2003.815018, Jun 2003.
- [16] D. Mazzoni, "PixelLearn," *NASA Tech Briefs / Software*, NPO-42082, Oct 31, 2006.
- [17] Hyperion Data Products site  
[http://eo1.usgs.gov/userGuide/index.php?page=hyp\\_pr od](http://eo1.usgs.gov/userGuide/index.php?page=hyp_pr od)
- [18] Vapnik, *Statistical Learning Theory*, Wiley-Interscience, ISBN-10: 0471030031, Sept 16<sup>th</sup> 1998.
- [19] Chih-Wei Hsu, Chih-Jen Lin, "A Comparison of Methods for Multiclass Support Vector Machines," *IEEE Transactions on Neural Networks*, vol. 13, Issue 2, pp. 415-425, doi: 10.1109/72.991427, Mar 2002.
- [20] U. Rebbapragada et al, "Improving Onboard Analysis of Hyperion Images by Filtering Mislabeled Training Data Examples," *2008 IEEE Aerospace Applications*, February 5-12<sup>th</sup>, 2009.
- [21] U. Rebbapragada and C. E. Brodley, "Class Noise Mitigation Through Instance Weighting," *Lecture Notes in Computer Science, Machine Learning: ECML 2007*, Vol. 4701/2007, doi: 10.1007/978-3-540-74958-5\_71, Sept. 8<sup>th</sup>, 2007.
- [22] ASTER Spectral Library, JPL/CalTech, Pasadena, California. Copyright© 1999, California Institute of Technology. ALL RIGHTS RESERVED.