

# Airborne imaging spectroscopy to monitor urban mosquito microhabitats

David R. Thompson<sup>a</sup>, Manuel de la Torre Juárez<sup>a</sup>, Christopher M. Barker<sup>b</sup>, Jodi Holeman<sup>c</sup>, Sarah Lundein<sup>a</sup>, Steve Mulligan<sup>c</sup>, Thomas H. Painter<sup>a</sup>, Erika Podesta<sup>a</sup>, Felix C. Seidel<sup>a</sup>, Eugene Ustinov<sup>a</sup>

<sup>a</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena CA

<sup>b</sup>Center for Vectorborne Diseases, University of California, Davis, CA

<sup>c</sup>Consolidated Mosquito Abatement District, Selma, CA

## Abstract

West Nile (WNV) is now established in the continental United States, with new human cases occurring annually in most states. Mosquitoes in the genus *Culex* are the primary vectors and exploit urban stagnant water and swimming pools as larval habitats. Public health surveys to monitor unmaintained pools typically rely on visual inspections of aerial imagery. This work demonstrates automated analysis of airborne imaging spectroscopy to assist *Culex* monitoring campaigns. We analyze an overflight of Fresno County, CA by the Airborne Visible Infrared Imaging Spectrometer instrument (AVIRIS), and compare the spectral information with a concurrent ground survey of swimming pools. Matched filter detection strategies reliably detect pools against a cluttered urban background. We also evaluate remotely sensed spectral markers of ecosystem characteristics related to larval colonization. We find that commonly used chlorophyll signatures accurately predict the probability of pool colonization by *Culex* larvae. These results suggest that AVIRIS spectral data provide sufficient information to remotely identify pools at risk for *Culex* colonization.

**Keywords:** Imaging Spectroscopy, Disease Vector Control, West Nile Virus, Green Swimming Pools, Matched Filter Detection, Urban Environments

## 1. Introduction

Remote sensing has long contributed to infectious disease prediction and warning systems (Linthicum et al., 1987; Washino & Wood, 1994). Spatial epidemiological studies have used satellite data to map environmental conditions associated with disease vector habitats. They typically correlate environmental variables such as land use, vegetation indices, temperature, and elevation with relative vector abundance or pathogen transmission (Beck et al., 2000; Kalluri et al., 2007). Instruments used for this purpose include the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and Moderate Resolution Imaging Spectroradiometer (MODIS). Urban areas pose a special challenge: they are heterogeneous collections of residential and commercial areas, parks, and other land use types with potential vector habitats on much smaller spatial scales (Reisen, 2010). Characterizing these sparse microhabitats requires different remote sensing techniques.

A particular concern is West Nile Virus (WNV), which spreads in urban areas by transmission between birds and mosquitoes in the genus *Culex*. These mosquitoes often colonize stagnant water during their aquatic immature stages, and features such as open containers or unmaintained swimming pools provide key habitats (Caillouët et al., 2008; Reisen et al., 2008). Unmaintained residential swimming pools, or *green*



Figure 1: *Culex* mosquitoes in urban areas often use unmaintained “green pools” as larval habitat. Image: Santa Clara Vector Control District (2013); Franklin (2013)

pools, are especially problematic. These neglected pools become stagnant with accumulated organic matter, and often harbor *Culex pipiens* mosquito larvae (Figure 1).

Current WNV mitigation efforts generally rely on street-level monitoring and treatment campaigns. Previous work has used remote sensing products such as airborne images to monitor individual pools, enabling more effective *Culex* treatment programs by identifying risk areas and flagging specific households for direct intervention. For instance, Reisen et al. (2008) demonstrate an airborne survey of Bakersfield, CA with high resolution color imagery. These images clearly reveal neglected

Email addresses: david.r.thompson@jpl.nasa.gov (David R. Thompson), mtj@jpl.nasa.gov (Manuel de la Torre Juárez), cmbarker@ucdavis.edu (Christopher M. Barker)

green pools. However, manual inspection is necessary to catalog these habitats, so the approach is better suited for a single snapshot in time than sustained monitoring campaigns that track the evolution of the vector habitats. Since then, local providers have continued to refine the manual image inspection using GIS pool catalogs and higher spatial resolution (Franklin, 2013). There have also been efforts to automate the image analysis. Kim et al. (2011) propose a fully automated method to locate pools in GeoEye satellite images, detecting pools from a *Normalized Difference Water Index* (NDWI) score followed by a morphological classification. This is effective at finding pools, but assessing their condition still relies on manual inspection. To our knowledge no previous study has quantified a link between remotely sensed pool color and *Culex* colonization that would permit automated assessment of pool condition from airborne instruments.

This study contributes to the literature in two ways. First, we evaluate a new class of sensor — airborne imaging spectrometers — to assist *Culex* monitoring campaigns. These instruments, such as the Airborne Visible Infrared Imaging Spectrometer (AVIRIS), typically measure reflected light over large areas in wavelengths from 370 nm to 2500 nm. The wide bandwidth and high spectral resolution permit a suite of powerful approaches for pool detection and classification. In this work, we apply a matched filter approach to identify pool locations. Our second contribution is to quantify the relationship between remotely observed spectral attributes and the presence of *Culex* larvae. To this end, we combine AVIRIS imagery of an urban environment with reference surveys by vector control authorities. We construct models relating pool health to common spectral indicators of water quality. The results indicate a strong relationship between typical signatures of algal chlorophyll and *Culex* colonization.

## 2. Methodology

### 2.1. Data Acquisition

Our study analyzed a survey overflight of Fresno, California (USA) that took place on 30 Sept. 2011. WNV is endemic in the area; the 2012 year had 24 confirmed infections in Fresno County and 479 in the state at large, resulting in 19 fatalities (California Department of Public Health West Nile Website, 2012). The Airborne Visible Infrared Imaging Spectrometer (AVIRIS) (Green, 2008) overflew the city on a Twin Otter turboprop aircraft under clear atmospheric conditions, acquiring spectra in the 370 – 2500 nm range with 10 nm spectral resolution and 3.7 m ground sampling distance. It imaged 60 km<sup>2</sup> of urban terrain with residential communities and commercial districts with occasional parks, canals and open reservoirs. Figure 2 (Left) shows a typical orthorectified subframe. We applied the Atmospheric / Topographic Correction for Airborne Imagery (ATCOR) algorithm (Richter & Schläpfer, 2012) to compensate for scattering and absorption. At the same time the AVIRIS measured radiances were normalized with the solar irradiance to produce Hemispherical Directional Reflectance Factors (Schaeppman-Strub et al., 2006), hereafter “reflectance.”

In the weeks before and after the overflight, the Consolidated Mosquito Abatement District (MAD) conducted field surveys of suspect swimming pools in the area. Inspectors recorded pools as *breeding* or *nonbreeding* depending on whether mosquito larvae were found. They also categorized pool conditions as one of: *Dry*, if the pool was empty; *Light green*, if the pool had low algal density indicative of a recent lack of maintenance; *Dark green*, if the pool had high algal density indicative of a long-term lack of maintenance; and *Blue*, if the pool was in normal condition. The breeding pools were always associated with the dark or light green conditions, while nonbreeding pools were always blue. Inspectors recorded the GPS location of the household and the date of the visit.

We began the analysis by matching the MAD survey entries to specific pixels in the AVIRIS image. This was complicated by the fact that the GPS records did not exactly correspond to pool locations; instead, they typically fell on the households’ front driveways. We determined the precise pixel position of each pool by inspecting high-resolution commercial satellite imagery (Google, 2012; Nokia/DigitalGlobe, 2012). Almost all pools were visually apparent both in the high-resolution satellite images and the AVIRIS data. Occasionally a small, dark, or shadowed pool was not obvious in the lower-resolution AVIRIS image. In these cases we located the appropriate AVIRIS pixel by referencing nearby landmark features. Our data set consists of the first 25 pools from both breeding and nonbreeding groups based on the incidental ordering of the Consolidated MAD database.

It was necessary to skip some database entries to preserve the quality of this sample. We ignored entries for pools that could not be seen in the high-resolution images (6 pools). This may have been caused by modifications between the time of the survey and the unknown date of the satellite image. It is also possible that some database entries referred to spas which were under patio covers and therefore invisible. We also skipped pools whenever neighboring landmarks were not clear enough to identify the corresponding AVIRIS pixel (7 pools). Finally, some pools visited very early had obviously dried out by the time of the AVIRIS overflight so that only the bright bottom was visible (4 pools). We continued to label valid pixels until reaching the total number of 25 from each category. Most pools were inspected within two weeks of the overflight and it is reasonable to expect they did not change significantly during this period.

### 2.2. Matched Filters for Pool Detection

We sought an automated procedure to transform an AVIRIS reflectance cube into a map of colonized and clean pools. This involves several distinct challenges. Finding pools requires suppressing false positives from urban spectral clutter. Then, assessing colonization potential requires an interpretable statistical prediction. Consequently we formulated the problem as two sequential steps of *pool detection* and *pool characterization*. We used different methods for each objective and then validated the two stages independently. The two steps appear in the center and left panels of Figure 2.

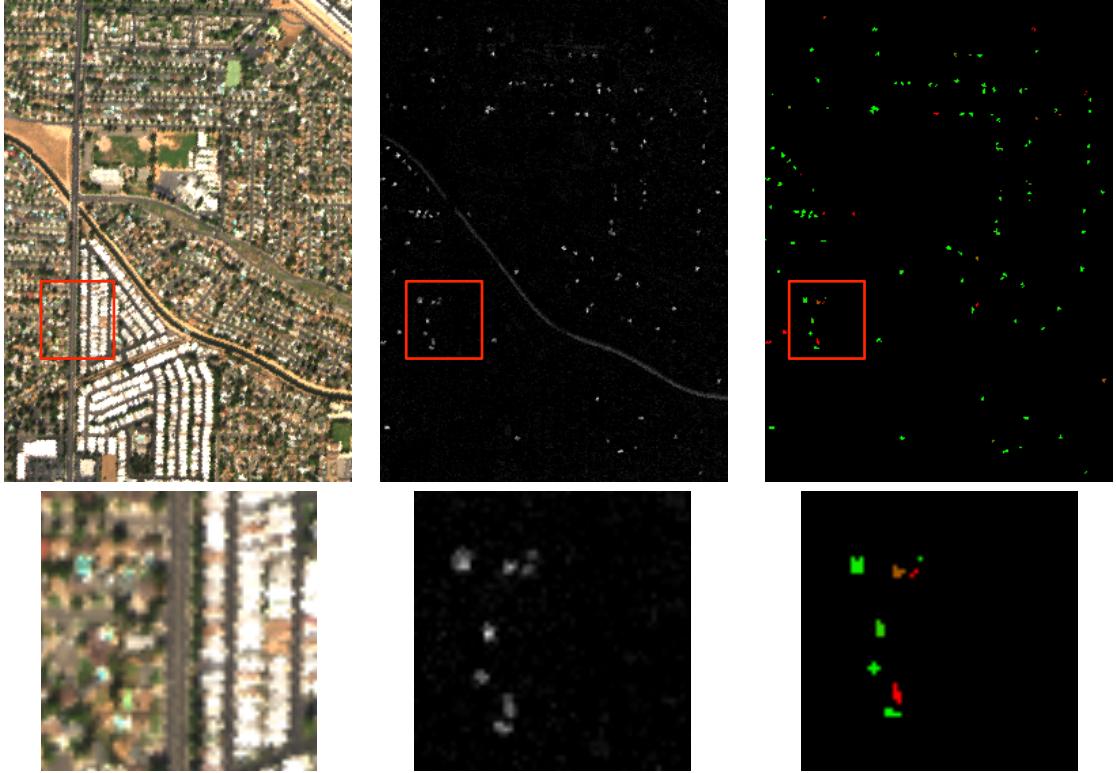


Figure 2: Left: A representative subframe of the AVIRIS image acquired in an overflight of Fresno, CA. The frame below shows a zoomed-in view of the red rectangular area, with several pools in different conditions. Center: Detection result showing pool locations. All detections are accurate. Pixel intensity correlates with the strength of the MF score. Right: We compute each pools' probability of colonization using its *chlorophyll-a* signal. Red values correspond to a high colonization probability. We removed spurious detections on the canal, since these would be easy to exclude using GIS data.

The detection step identifies pixels containing pools. Previous work in pool detection by Kim et al. (2011) uses high-resolution imagery containing color and morphological information. They use the Normalized Difference Water Index (NDWI), a ratio of green and near-infrared channels, to flag water pixels. In contrast, we used a full-spectrum detection algorithm to analyze a wide spectral range. We excluded atmospheric absorption bands, analyzing all channels within the intervals from 400–1206 nm, 1433–1732 nm, and 1957–2500 nm. We used a matched filter, a classical strategy for subpixel target detection in spectral data (Manolakis et al., 2001). A linear mixing model treats the observed spectral reflectance in  $d$  channels,  $\mathbf{x} \in \mathbb{R}^d$ , as convex linear combination of a background distribution with the target spectrum  $\mathbf{t} \in \mathbb{R}^d$ . It models the background as a multivariate Gaussian distribution with mean  $\mu$  and covariance matrix  $\Psi$ , ignoring independent additive measurement noise (Stocker et al., 1990). It is most common to estimate background means and covariances from the data. For the set of observed reflectances  $X = \{\mathbf{x}_i\}_{i=1}^n$ , we have:

$$\mu = \frac{1}{N} \sum_{\mathbf{x}_i \in X} \mathbf{x}_i \quad \Psi = \frac{1}{N} \sum_{\mathbf{x}_i \in X} (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T \quad (1)$$

For a target  $\mathbf{t}$  combined with the background at a mixing fraction  $\phi \in [0, 1]$ , the measured reflectance is a perturbed normal distribution (DiPietro et al., 2010):

$$\mathbf{x} = (1 - \phi)N(\mu, \Psi) + \phi\mathbf{t} \quad (2)$$

The optimal matched filter is a projection that best separates the distributions in which the target is present and absent - these differ only by a constant factor, having equivalent covariance statistics. The matched filter is defined as:

$$\alpha_i = \frac{(\mathbf{x}_i - \mu)^T \Psi^{-1} (\mathbf{t} - \mu)}{(\mathbf{t} - \mu)^T \Psi^{-1} (\mathbf{t} - \mu)} \quad (3)$$

The matched filter score  $\alpha_i$  approximates the target mixing fraction  $\phi_i$ .

In practice the background is not perfectly Gaussian and numerically extreme outliers may have large but purely incidental projections  $\alpha$ . This is particularly true for urban environments containing a diverse mixture of synthetic and highly reflective materials. We adopted a false-alarm mitigation strategy first proposed by DiPietro et al. (2010) to reject these outlier pixels. Here a second score,  $\beta_i$ , estimates the likelihood of the observed reflectance given the appropriate fractions of Gaussian background and target. This indicates whether, after subtracting the target fraction, the remainder of its mixed spectrum is representative of the background. Under weak assumptions (DiPietro et al., 2010), this equates to the Mahalanobis distance with respect to the background covariance,  $\beta_i$ , measuring the distance between the pixel under test and its expected value at the predicted mixing fraction.

$$\mu(\alpha_i) = \alpha_i \mathbf{t} + (1 - \alpha_i) \mu \quad (4)$$

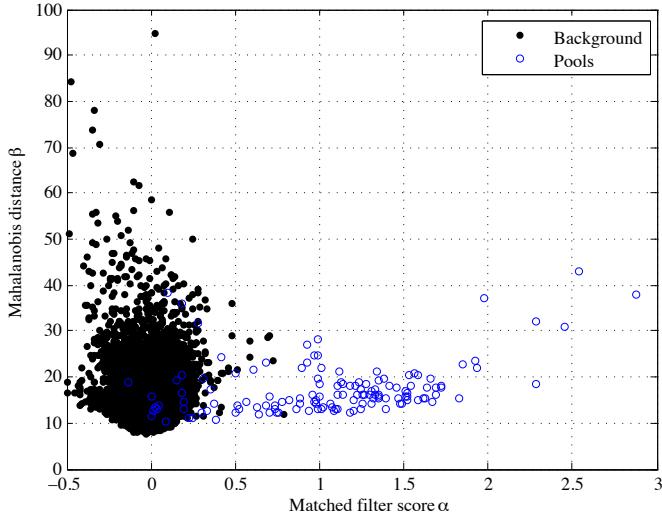


Figure 3: Matched filter and outlier rejection scores  $\alpha$  and  $\beta$ , for each pixel in the test subframe.

$$\beta_i = (\mathbf{x} - \mu(\alpha_i))^T \Psi^{-1} (\mathbf{x} - \mu(\alpha_i))^T \quad (5)$$

Large  $\beta$  values indicate outlier pixels. Figure 3 shows an example of  $\alpha$  and  $\beta$  values for pixels in the test subframe. Most pool pixels were easily separable from the background distribution.

In order to make the detection decision one must combine the  $\alpha$  and  $\beta$  scores into a scalar score. This combined score is typically a linear combination of  $\alpha_i$  and  $\beta_i$ , but the optimal weighting depends on characteristics of the data. Here we set the coefficients by injecting synthetic spectra of pools into an urban scene at mixing ratios between 0 and 1, producing  $\alpha$  and  $\beta$  scores for both target and background pixels. We then fit a linear discriminant which best separated the two populations, using a logistic regression mapping onto class labels 1 and 0 for the pool and non-pool classes, respectively (Bishop, 2006). Other linear discriminants are possible - for example, Foy et al. (2009) use a Support Vector Machine. In practice, we found performance was insensitive to the precise form of the decision boundary and any reasonable combination improved performance *vis a vis* matched filters that did not include the  $\beta$  score.

The matched filter algorithm required a representative spectral library. We gathered these target spectra directly from pools that were visible in urban AVIRIS scenes. For these tests we formed a target vector  $\mathbf{t}$  based on the mean spectrum of the pools identified from the MAD survey database. The middle panel of Figure 2 shows the result, with pixel brightness indicating the strength of the match.

We then evaluated true and false detection rates by exhaustively labeling all visible pools in a subframe of the AVIRIS overflight. The MAD database was not suited for this purpose because it did not include all pools in its geographic region. Therefore it could not identify a false positive; spurious detections within its boundaries could simply have been unrecorded pools. Fortunately even small pools were quite easy to see in the high-resolution satellite remote sensing data (Google, 2012; Nokia/DigitalGlobe, 2012), so we used this as a standard for

evaluation. We examined all houses appearing in a fixed geographic area of approximately 2.0 square kilometers outside the ground survey area. This subframe was selected to contain range of different commercial and residential areas (Figure 2). We used high resolution satellite images to build a comprehensive list of all pools in the test scene (137 in total). Next, we marked one pixel of each pool and treat detection scores within 10 m (3 pixels) of the mark as successful. This radius was large enough to capture the largest pools, and small enough that localization errors would not matter to a ground survey team. We masked out a canal that transected the subframe, since large water features would be easy to exclude with GIS data in a fully automated system.

### 2.3. Pool Characterization

After the initial *detection* step we used reflectance spectra to characterize the condition of each pool. Mosquito colonization of unmaintained pools often coincides with accumulation of spectrally distinctive algae and other organic material. Studies by Reisen et al. (2008) state that larval colonization is associated with lower concentrations of pool chemicals, which also lead to algal blooms. We hypothesized that commonly used spectral water quality indicators, such as measures of algae, chlorophyll and suspended solids, might predict mosquito colonization. Most previous water quality analyses involve controlled laboratory environments (Han, 1997) or remote sensing of large natural bodies of water (e.g. Kallio et al., 2001; Yacobi et al., 1995; Dekker et al., 2002, and references therein). Residential pools introduce additional complications, such as an optical path that includes reflection from the pool bottom and the presence of non-water materials in the AVIRIS pixel. However, pools are physically isolated from each other and permit many independent trials in a small geographic area.

We first considered spectral signatures related to algae. Algae-laden water contains several diagnostic spectral features independent of the presence of other suspended solid material. These features appear in studies based on laboratory measurements (Han, 1997), *in situ* field measurements (Yacobi et al., 1995), and remote sensing data (Dekker et al., 2002). Algae exhibit low reflectance in the 400 – 500nm region. There are multiple reflectance peaks caused by local minima in pigment absorption - a strong peak between 550 and 580 nm, and another at 650 nm (Dekker et al., 2002). There is a reflectance minimum near 670 due to absorption by chlorophyll. Reflectances then increases to a 700 nm *NIR peak* caused by an interaction of cell scattering, pigment, and absorption. The *chlorophyll-a* concentration is a common primary inversion parameter for remote sensing of phytoplankton pigments. It typically involves a ratio of reflectances in the region from 670 – 720 nm (Kallio et al., 2001). Other ratios are possible; Hoogenboom et al. (1998) estimate these concentrations with the ratio of bands at 715 nm and 667 nm, capturing chlorophyll and pheophytin pigment features. We also considered spectral measures of Turbidity, used here as a proxy for Total Suspended Solid matter (Kallio et al., 2001). In general, suspended sediment has been shown to cause an increase in reflectance across the 600-710nm range (Han, 1997). In algae-laden water this feature remains strongest

at approximately 700 nm, and Kallio et al. (2001) measure turbidity with the reflectance at 710 nm.

In this study a logistic regression model (Bishop, 2006) represented the relationship between one or more spectral parameters and the presence of mosquito larvae. It estimated the probability  $P_\theta$  that a pool harbored *Culex* larvae based on one or more variables of interest. For independent variables  $\mathbf{v} = \{v_1, \dots, v_n\}$  and free parameters  $\Theta = \{\theta_0, \dots, \theta_n\}$  the model was:

$$P_\theta = \frac{1}{1 + \exp[-(\theta_0 + \theta_1 v_1 + \dots + \theta_n v_n)]}. \quad (6)$$

We report performance from models formed using each of the aforementioned spectral attributes. We also fit a multivariate model with a wide spectral range, using all channels between 400 nm and 2500 nm except the atmospheric H<sub>2</sub>O absorption bands as before. We fit all model coefficients using standard maximum likelihood methods detailed by Bishop (2006), without regularization terms.

We evaluated performance by comparing *breeding* and *nonbreeding* classifications against the MAD reference. We estimated classification accuracy rates using Leave One Out Cross Validation (LOOCV) (Dreiseitl & Ohno-Machado, 2002). Specifically, we fit coefficients using all but one data point and then applied the resulting model to classify the held out datum. We repeated the procedure over the entire dataset and computed the expected accuracy rate. This gave an unbiased estimate of how a similar classifier trained on the entire dataset would perform on new instances from the population (Kearns & Ron, 1999).

### 3. Results

#### 3.1. Pool Detection Results

We first evaluated matched filter detection by comparing performance to the Normalized Difference Water Index (NDWI) from McFeeters (1996) and Kim et al. (2011). Their work uses green and near infrared bands of satellite data having higher spatial resolution but lower spectral resolution. For an accurate comparison we spectrally coarsened AVIRIS radiance values to match the GeoEye bands (Red: 510-550nm, NIR: 800-910nm). We also considered an NDWI score computed from the high spectral resolution, atmospherically corrected reflectances (Table 1).

Both the Matched Filter and NDWI produced scores for every pixel. To detect discrete pools one must threshold this score at a level which balances the number of true and false detections. Figure 4 represents this tradeoff with a Receiver Operating Characteristic (ROC) curve (Fawcett, 2004). Desirable curves (near the upper left) have many true detections and few false positives. We compare NDWI detection strategies with the matched filter approach, cautioning that our NDWI statistic does not include the additional morphological analysis of Kim et al. (2011) due to the lower spatial resolution. The matched filter performed better than the NDWI approach using GeoEye-equivalent and AVIRIS spectral resolutions. It detected 122 of 137 pools without false alarms. A false detection

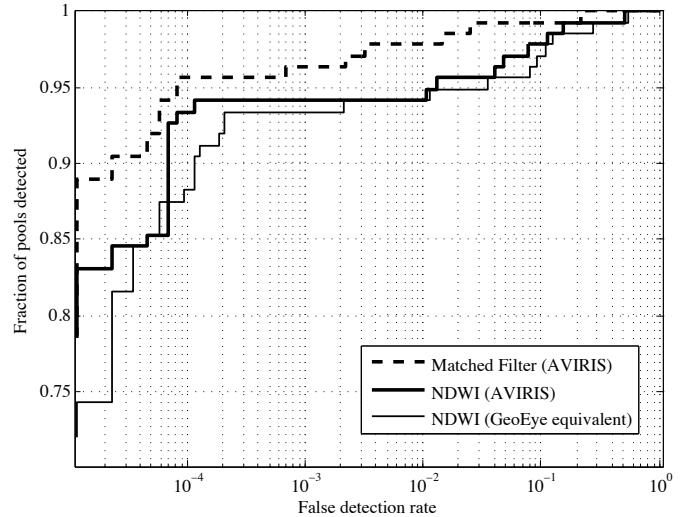


Figure 4: Performance comparison for spectral detection methods based on matched filtering and the Normalized Difference Water Index (NDWI).

rate of  $10^{-4}$  revealed 131 real pools, providing 95% recall with approximately 7 false alarms per square kilometer. This was comparable to the accuracy reported by Kim et al. (2011) who use a combination of color and morphological analysis. Pools showed distinctive signatures combining high reflectivity in visible wavelengths with a 850nm water absorption feature. This made them relatively simple to detect, and there was no single predominant cause of errors. Rare synthetics like plastic roofing materials occasionally caused false positives. There was occasionally confusion inside dark shadows, a common challenge for urban remote sensing (Lachérade et al., 2005; Dell'Acqua et al., 2005). Shadows could also prevent detection when overhanging trees or buildings obscured a real pool.

#### 3.2. Pool Characterization Results

The spectra of *breeding* and *nonbreeding* groups separated visually into distinct populations. Figure 5a shows the median spectrum from each group, with confidence bars indicating the middle quartiles. Figure 5b shows the first derivative of the spectrum. The *breeding* pool data was consistent with all algae chlorophyll signatures noted by Han (1997). A lower slope in the 400-500nm range may indicate algal absorption. A 550nm green peak was present, though it also appeared (somewhat blue-shifted) in clean pools. A concavity in the *breeding* pool spectra was consistent with a 670nm chlorophyll absorption feature. A relative peak at 700nm appeared in *breeding* pools and may also be related to chlorophyll (Gitelson, 1992). Finally, the *breeding* pools also exhibited a shallower slope in the 500-600nm range.

We evaluated each of these spectral attributes for predicting the *breeding* / *nonbreeding* distinction. Table 1 shows leave-one-out model prediction accuracies and Kappa statistics (Fielding & Bell, 1997) at the best-performing thresholds. Each  $R_{a,b}$  represents the mean reflectance value between wavelengths  $a$  and  $b$  (Lunetta et al., 2009). Turbidity performed only slightly better than random. The NDWI index correlated somewhat with

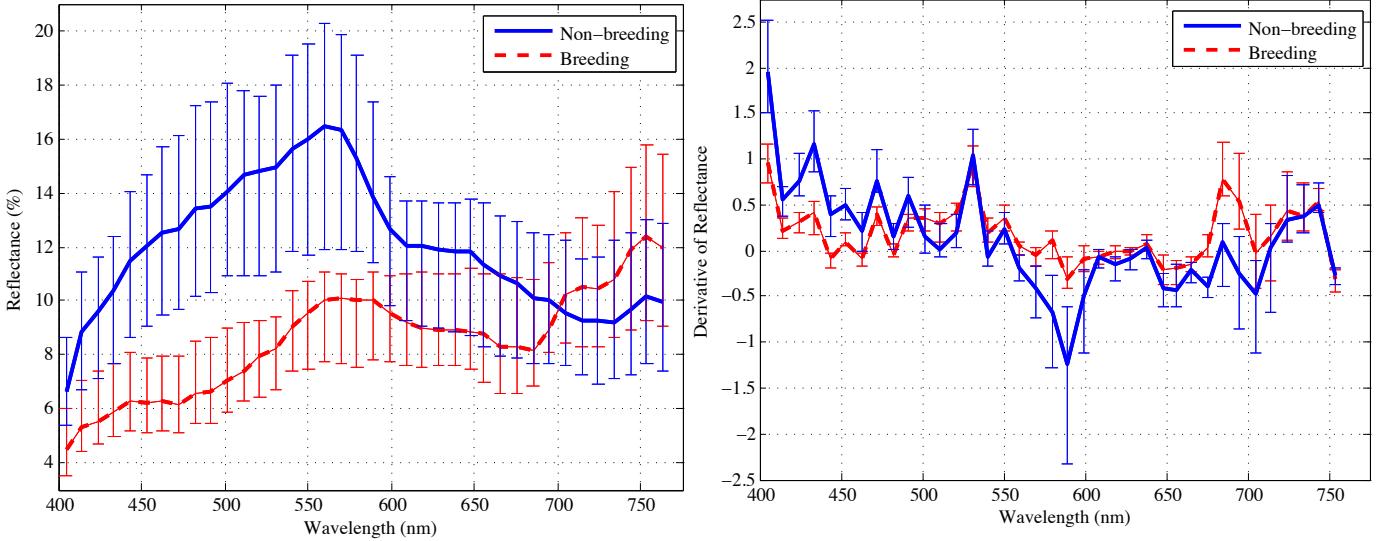


Figure 5: Spectra from pools surveyed by the Consolidated Mosquito Abatement District. 25 Pools from each class are represented. Error bars indicate the median and middle quartile radiance values for the reflectance spectra (top panel), their first derivative with respect to wavelength (bottom panel).

Table 2: Performance of models formed by multiple spectral attributes, with Kappa coefficients and predictive accuracy for the *breeding / nonbreeding* distinction. The model based on chlorophyll outperformed every combination except for the full spectrum data.

CCombination	Kappa	Predictive accuracy
Turbidity only	0.11	59.0%
NDWI only	0.53	75.6%
NDWI, Turbidity	0.69	84.6%
Chlorophyll, NDWI	0.87	93.6%
Chlorophyll, Turbidity	0.87	93.6%
Chlorophyll, NDWI, Turbidity	0.87	93.6%
<b>Chlorophyll only</b>	<b>0.89</b>	<b>94.8%</b>
<b>Full spectrum</b>	<b>0.92</b>	<b>96.2%</b>

*Culex* infestation, but the chlorophyll indices achieved better performance. The ratio of 685 – 691 nm and 670 – 677 nm bands noted in Kallio et al. (2001) had the best cross-validation error.

Table 2 shows prediction accuracy for models formed from combinations of multiple variables. The combinations used the best-performing band ratios from Table 1. Notably, the turbidity feature bolstered the predictive performance of the NDWI, suggesting that these indicators may be complementary. However, chlorophyll signals alone achieved a far better cross-validation error rate, outperforming any combination of multiple variables with the sole exception of the full-spectrum data that performed only slightly better. Figure 6 illustrates logistic regression models made from two different pairs of variables. The left and right panels show models formed by pairing NDWI with chlorophyll and turbidity attributes. Each plot indicates probability isocontours at the 5%, 50%, and 95% levels. The populations separated most cleanly along the *chlorophyll-a* feature.

The simplicity and physical interpretability of the *chlorophyll-a* relationship recommends it for pool char-

acterization. With spectra expressed as fractional surface reflectance, the top-performing logistic regression model using chlorophyll was simply:

$$P_{\Theta} = \frac{1}{1 + \exp(-240 + 243.3c)} \quad \text{for } c = \frac{R_{685-691}}{R_{670-677}} \quad (7)$$

This assumed an even balance of *breeding* and *nonbreeding* pools, so widely unbalanced distributions would require appropriate weighting to preserve accurate probability estimates.

#### 4. Discussion

This study demonstrates that airborne imaging spectroscopy can assist WNV vector control. For the task of pool detection, AVIRIS reflectance data provides performance comparable to the GeoEye studies of Kim et al. (2011), without the high spatial resolution of that instrument. The full spectrum detection is valuable to find green pools having subtle water absorption signatures. After detection, the spectral data is uniquely helpful for characterizing pool condition. We find that remotely sensed *chlorophyll-a* signatures predict mosquito colonization. A band ratio related to *chlorophyll-a* approaches the predictive power of a full-spectrum model, classifying colonized and clean pools with over 94% accuracy for this dataset. Separating detection and characterization allows both steps to be validated and refined independently. For example, the detection could add morphological or GIS cues, while characterization could incorporate seasonal priors on colonization probabilities.

In reality the observed spectrum is affected by the bottom reflectance as well as optical properties of the water (Kim et al., 2010). Residential pools have a wide range of synthetic colors and materials. Many pool bottoms appear green to the human eye and can be mistaken for algae in visible images. Spectroscopic data overcomes this problem by revealing the unique spectral signatures of algal pigmentation and scattering. Our

Table 1: Definition and predictive accuracy of spectral water quality variables.

Variable	Value	Kappa	Predictive accuracy	
NDWI	$(R_{555-565} - R_{800-900})/(R_{555-565} + R_{800-900})$	0.53	75.6%	Kim et al. (2011)
Chlorophyll- <i>a</i>	$R_{704-713}/R_{670-680}$	0.74	87.2%	Hoogenboom et al. (1998)
	$R_{699-705}/R_{670-677}$	0.82	91.0%	Kallio et al. (2001)
	$R_{685-691}/R_{670-677}$	0.89	94.8%	Kallio et al. (2001)
	$R_{699-714}/R_{661-667}$	0.77	88.5%	Kallio et al. (2001)
	$R_{674}/R_{693}$	0.87	93.6%	Lunetta et al. (2009)
	$R_{704-713}/R_{675-685}$	0.82	91.0%	Richardson (1996)
Turbidity	$R_{705-714}$	< 0.0	55.1%	Kallio et al. (2001)
	$R_{699-705}$	< 0.0	55.1%	Kallio et al. (2001)
	$R_{705-714} - R_{747-755}$	0.11	59.0%	Kallio et al. (2001)
	$R_{705-714}/R_{747-755}$	0.09	56.4%	
Full spectrum	$[R_{400} \dots R_{1206}, R_{1433} \dots R_{1732}, R_{1957} \dots R_{2500}]$	0.92	96.2%	

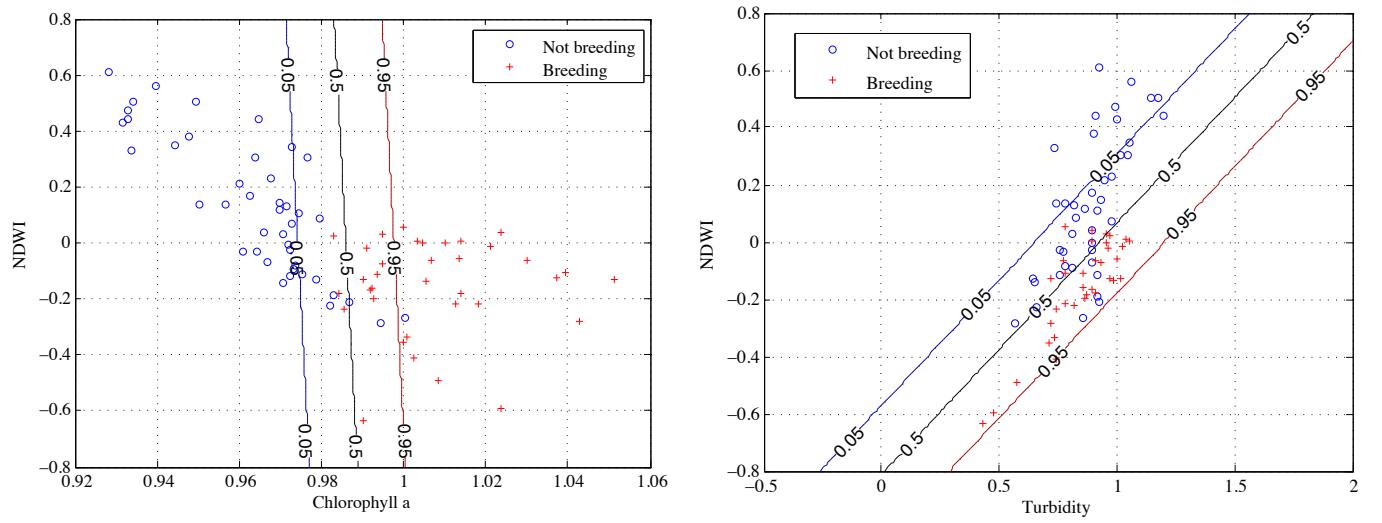


Figure 6: Logistic regression models related spectral attributes to the presence of *Culex* larvae. Left: A model combining NDWI and chlorophyll attributes. Chlorophyll showed the strongest predictive relationship of all the spectral ratios. Lines indicate probability isocontours at 5%, 50%, and 95% levels. Right: a similar model using NDWI and Turbidity attributes.

model uses this to directly estimate the probability of colonization; it averages over pool bottom, surface and volume effects to produce an expedient result that is readily interpretable by end users. However, it is possible that the spectral data could also estimate other physical attributes. For example, the semianalytical treatment of Lee et al. (1999); Kim et al. (2010) models optically deep water by expressing subsurface reflectance as a function of optical properties such as absorption and diffuse attenuation. One could add bottom reflectance and depth terms to model shallow water such as pools. Such models could be augmented with algal backscattering and absorption terms to estimate physical chlorophyll concentrations. This would tie *Culex* colonization to a physical quantity rather than a spectral proxy, with the advantage that chlorophyll concentrations could be measured *in situ* for direct validation. Predictions grounded in these physical attributes could also be more consistent across space and time. Unfortunately this approach requires pure homogeneous water pixels, which are not usually available for small pools in our dataset. Future airborne instruments will have far higher spatial resolution, resulting in one or more pure water pixels that could be used to fit these physical models.

For now, we have presented a preliminary study showing statistical relationships based on spectral attributes. These are stable for this dataset and provide sufficient information to assess and predict *Culex* colonization. Inevitably the limited temporal and spatial scope leaves some possibility for systematic biases. Seasonal factors like temperature, geography, and even regional demography might influence the suitability of unmaintained pools as habitat. This makes further study at different times and locales important. Future studies could also control for biases due to factors such as pool size. In this work we observed that unmaintained breeding pools were typically smaller than the *nonbreeding* pools. They more rarely filled entire AVIRIS pixels, which might increase the level of sub-pixel mixing with nearby materials. It would be surprising if such effects alone explained the strong chlorophyll correlation observed here. Nevertheless, it would be valuable to account explicitly for pool size in future studies.

Overall this work provides an initial proof of concept that airborne imaging spectroscopy could be useful for mosquito control and public health agencies charged with controlling WNV. Naturally, identifying green pools is only one aspect of a comprehensive vector control program. Other vector sources include small containers, drains and non-pool water features. On larger scales, it is also possible that remote reflectance spectroscopy could help to identify and characterize other important immature mosquito habitats, such as open wetlands, stormwater management structures, groundwater retention basins, or flooded cropland.

## Acknowledgments

We thank Joao Teixeira, Robert Staehle, and Michael Gunson whose suggestions and support contributed significantly to this research. This work was supported by a Jet Propulsion Laboratory Research and Technology Development Grant. AVIRIS

data is available courtesy NASA and the Jet Propulsion Laboratory. The Consolidated Mosquito Abatement District provided spatial data on swimming pool conditions. The aerial image in this article was provided courtesy Russ Parman and the Santa Clara County Vector Control District, and Robert Franklin of Aerial Services, a division of TeamBuilders, Inc. C. M. Barker is supported by grant U01 EH000418 from CDC, Climate Change: Environmental Impact on Human Health and a grant from NASA, Earth-Sun Science Applied Sciences Program, and by the Research and Policy for Infectious Disease Dynamics (RAPIDD) program of the Science and Technology Directorate, Department of Homeland Security and Fogarty International Center, National Institutes of Health. Copyright 2013 California Institute of Technology. All Rights Reserved. U.S. Government support acknowledged.

## References

- Beck, L. R., Lobitz, B. M., & Wood, B. L. (2000). Remote sensing and human health: New sensors and new opportunities. *Emerg Infect Dis*, 6, 217–227.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer. ISBN:978-0-387-31073-2.
- Caillouët, K. A., Carlson, J. C., Wesson, D., & Jordan, F. (2008). Colonization of abandoned swimming pools by larval mosquitoes and their predators following hurricane katrina. *Journal of Vector Ecology*, 33, 166–172.
- California Department of Public Health West Nile Website (2012). <http://westnile.ca.gov/reports.php>.
- Dekker, A. G., Vos, R. J., & Peters, S. W. M. (2002). Analytical algorithms for lake water tsm estimation for retrospective analyses of tm and spot sensor data. *International Journal of Remote Sensing*, 23, 15–35.
- Dell'Acqua, F., Gamba, P., & Trianni, G. (2005). A preliminary study on separability of paving materials in shadowed hyperspectral pixels from a central urban area. In *Proceedings of ISPRS Joint Conferences: 3rd International Symposium Remote Sensing and Data Fusion Over Urban Areas (URBAN 2005) and 5th International Symposium Remote Sensing of Urban Areas (URS 2005)*.
- DiPietro, R., Manolakis, D., Lockwood, R., Cooley, T., & Jacobson, J. (2010). Performance evaluation of hyperspectral detection algorithms for sub-pixel objects. *Proc. SPIE*, 7695.
- Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics*, 35, 352–359.
- Fawcett, T. (2004). ROC graphs: notes and practical considerations for researchers. *Pattern Recognition Letters*, 27, 882–891.
- Fielding, A., & Bell, J. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental conservation*, 24, 38–49.
- Foy, B., Theiler, J., & Fraser, A. (2009). Decision boundaries in two dimensions for target detection in hyperspectral imagery. *Optics Express*, 17, 17391–17411.
- Franklin, R. (2013). Personal communication. *Aerial Services, a division of TeamBuilders, Inc.* <http://www.aerialpoolsurvey.com>.
- Gitelson, A. (1992). The peak near 700 nm on radiance spectra of algae and water: relationships of its magnitude and position with chlorophyll concentration. *International Journal of Remote Sensing*, 13, 3367–3373.
- Google (2012). Google maps. <http://maps.google.com/> accessed Aug. 2012, .
- Green, R. (2008). AVIRIS and related 21st century imaging spectrometers for Earth and space science. *High Performance Computing in Remote Sensing*, (p. 335).
- Han, L. (1997). Spectral reflectance with varying suspended sediment concentrations in clear and algae-laden waters. *Photogrammetric Engineering and Remote Sensing*, 63, 701–705.
- Hoogenboom, H. J., Dekker, A. G., & Althuis, I. J. A. (1998). Simulation of aviris sensitivity for detecting chlorophyll over coastal and inland waters. *Remote Sensing of Environment*, 65, 333–340.
- Kallio, K., Kutser, T., Hannonen, T., Koponen, S., Pulliainen, J., Vepsilinen, J., & Pyhlathi, T. (2001). Retrieval of water quality from airborne imaging

- spectrometry of various lake types in different seasons. *The Science of the Total Environment*, 268, 59–77.
- Kalluri, S., Gilruth, P., Rogers, D., & Szczur, M. (2007). Surveillance of arthropod vector-borne infectious diseases using remote sensing techniques: A review. *PLoS Pathog*, 3(10), e116.
- Kearns, M., & Ron, D. (1999). Algorithmic stability and sanity-check bounds for leave-one-out cross-validation. *Neural Computation*, 11, 1427–1453.
- Kim, M., Holt, J. B., Eisen, R., Padgett, K., Reisen, W. K., & Croft, J. (2011). Detection of swimming pools by geographic object-based image analysis to support west nile virus control efforts. *Photogrammetric engineering and remote sensing*, 77, 1169–1179.
- Kim, M., Park, J., & Tuell, G. (2010). A constrained optimization technique for estimating environmental parameters from CZMIL Hyperspectral and Lidar Data. In *SPIE Defense, Security, and Sensing* (pp. 769513–769513). International Society for Optics and Photonics.
- Lachéradé, S., Miesch, C., Briottet, X., & Le Men, H. (2005). Spectral variability and bidirectional reflectance behaviour of urban materials at a 20 cm spatial resolution in the visible and near-infrared wavelengths. a case study over toulouse (france). *International journal of remote sensing*, 26, 3859–3866.
- Lee, Z., Carder, K., Mobley, C., Steward, R., & Patch, J. (1999). Hyperspectral remote sensing for shallow waters. 2. deriving bottom depths and water properties by optimization. *Applied Optics*, 38, 3831–3843.
- Linthicum, K., Bailey, C., Davies, F., & Tucker, C. (1987). Application of remote sensing to arthropod vector surveillance and control. *Science*, 235, 1565–1569.
- Lunetta, R. S., Knight, J. F., Paerl, H. W., Streicher, J. J., Peierls, B. L., Gallo, T., Lyon, J. G., Mace, T. H., & Buzzelli, C. P. (2009). Measurement of water colour using aviris imagery to assess the potential for an operational monitoring capability in the pamlico sound estuary, usa. *International Journal of Remote Sensing*, 30, 3291–3314.
- Manolakis, D. G., Siracusa, C., Marden, D., & Shaw, G. A. (2001). Hyper-spectral adaptive matched-filter detectors: practical performance comparison. *Proceedings of SPIE*, 4381.
- McFeeters, S. K. (1996). The use of the normalized difference water index (ndwi) in the delineation of open water features. *International Journal of Remote Sensing*, 17, 1425–1432.
- Nokia/DigitalGlobe (2012). Nokia maps. <http://maps.nokia.com/> accessed on Aug. 2012, .
- Reisen, W. K. (2010). Landscape epidemiology of vector-borne diseases. *Annual Review of Entomology*, 55, 461–483.
- Reisen, W. K., Takashi, R. M., Carroll, B. D., & Quiring, R. (2008). Delinquent mortgages, neglected swimming pools, and west nile virus, california. *Emerging Infectious Diseases*, 14, 1747.
- Richardson, L. (1996). Remote sensing of algal bloom dynamics. *BioScience*, 46, 492–501.
- Richter, R., & Schläpfer, D. (2012). *Atmospheric / topographic correction for airborne imagery: ATCOR-4 User Guide*. Technical Report DLR-IB 565-02/12, Weissling, Germany.
- Santa Clara Vector Control District (2013). <http://www.sccgov.org/sites/vector/Pages/Vector-Control-District-Site-Home-Page.aspx>.
- Schaepman-Strub, G., Schaepman, M., Painter, T., Dangel, S., & Martonchik, J. (2006). Reflectance quantities in optical remote sensingdefinitions and case studies. *Remote Sensing of Environment*, 103, 27–42.
- Stocker, A. D., Reed, I. S., & Yu, X. (1990). Multidimensional signal processing for electro-optical target detection. *SPIE Signal And Data Processing of Small Targets*, 1305.
- Washino, R., & Wood, B. (1994). Application of remote sensing to arthropod vector surveillance and control. *Am. J. Trop. Med. Hyg.*, 50, 134–144.
- Yacobi, Y. Z., Gitelson, A., & Mayo, M. (1995). Remote sensing of chlorophyll in lake kinneret using highspectral-resolution radiometer and landsat tm: spectral features of reflectance and algorithm development. *Journal of Plankton Research*, 17, 2155–2173.