Monitoring effects of a controlled subsurface carbon dioxide release on vegetation using a hyperspectral imager

Charlie J. Keith a, Kevin S. Repasky a,*, Rick L. Lawrence b, Steven C. Jay c, John L. Carlsten d

a Electrical and Computer Engineering, Cobleigh Hall Room 610, Montana State University, Bozeman, MT 59717, United States
b Land Resources and Environmental Sciences Department, United States
 c Spatial Sciences Center, United States
d Physics Department, United States

ABSTRACT

A hyperspectral imaging system was used to monitor vegetation during a subsurface controlled release of carbon dioxide (CO₂). From August 3 to 10, 2007, 0.3 tons CO₂/day were released through a 70 m horizontal pipe located at a nominal depth of 1.8 m below the surface. Hyperspectral images of alfalfa plants were collected during the controlled release and used along with classification tree analysis to study changes in the reflectance spectra as a function of perpendicular distance from the horizontal pipe. Changes in the reflectance spectra near the red edge (650–750 nm) were observed over the course of the controlled release experiment for plants within a perpendicular distance of 1 m of the release pipe. These results indicate monitoring vegetation over a carbon sequestration site has the potential to allow monitoring of the integrity of the CO₂ storage.

1. Introduction

The atmospheric concentration of carbon dioxide (CO₂) in the past 250 years has increased from an average 280 parts per million (ppm) in the preindustrial era to a current global average concentration of 384 ppm (Barnola et al., 1987; Cuffey and Vimeux, 2001; Masarie and Tans, 1995; Monnin et al., 2001; Scripps, 2007). The rising level of atmospheric CO₂ during the industrial age is mainly attributed to burning fossil fuels such as coal, oil, and natural gas (Masarie and Tans, 1995; Scheffer et al., 2006; Tans, 2006a) with annual emissions resulting from burning fossil fuels increasing from 23.5 GtCO₂/year in the 1990s to 26.4 GtCO₂/year from 2000 to 2005 (Keeling et al., 2005). There is growing international concern that the increase in CO₂ released in the atmosphere is altering the global climate and environment (Alcamo and Kreileman, 1996; Hansen, 2004; IPCC, 2001b; Norby and Luo, 2006; Shackleton, 2000; Tans, 2006b; Vinnikov and Grody, 2003). International efforts are underway to curb the emission of CO₂ into the atmosphere as a result.

One effort aimed at decreasing the emission of CO₂ into the atmosphere involves the capture of CO₂ at its source, such as a power generating facility. This CO₂ is then stored in a geologic storage site such as depleted oil wells, underground coal seams, or in deep saline formations, effectively removing the CO₂ from the atmosphere (Herzog, 2001; IPCC, 2001a, 2005; LBNL, 2000; Xu, 2004). Carbon capture and storage in geologic sites have an estimated potential capacity of 1680 GtCO₂ (LBNL, 2000; Mingzhe et al., 2006; Xu, 2004).

Three industrial scale carbon capture and storage projects are currently operating. The Sleipner Saline Aquifer Storage Project removes 1 million tons of CO₂ and stores it in a deep sub-sea brine-filled sandstone formation (Korbol and Kaddour, 1995). The Salah Gas Project began in 2004 and is currently storing 1 million tons of CO₂ in a depleted gas field (Knott, 2004). A third project located in Saskatchewan, Canada is using CO₂ captured by the Dakota Gasification Plant located in North Dakota for both enhanced oil recovery and CO₂ storage in the Weyburn oil field (Whittaker, 2004; Whittaker et al., 2002). Several smaller pilot projects for CO₂ storage are also under investigation (LBNL, 2000).

An important issue to ensure the successful storage of carbon dioxide in geologic sequestration sites is the ability to monitor these sites for leakage. The three main causes of leakage include leaking injection wells, leakage from improperly sealed abandoned wells, and leakage through geologic faults and fractures (Benson et al., 2005; Hepple, 2002; Knauss et al., 2005; Wilson et al., 2003). Initial studies of geologic storage sites indicate that for carbon storage to effective, seepage rates must be less than 0.1–0.01%/year (Benson et al., 2005).

CO₂ leaking from geologic storage sites might affect the vegetation above the storage site in an observable way. The leaking CO₂ initially might cause fertilization and stimulate plant growth. As the concentration of CO₂ builds up in the soil and displaces soil oxygen, however, plant stress begins to occur. Anomalous plant growth and plant stress might be monitored by utilizing different parts of the reflectance spectra generated by a
hyperspectral imaging spectrometer (Dawson and Curran, 1998; Smith et al., 2004a,b; Wilson et al., 2004), providing an indirect method of monitoring carbon sequestration sites for CO2 leakage. The ability to monitor carbon sequestration sites over long periods of time for signatures of the effects of leaking CO2 on vegetation might provide a cost effective monitoring technique for large areas associate with the sequestration sites.

Hyperspectral imaging is used to study scenes at several spectral bands typically in the visible to the short-wave infrared region, 400–2500 nm. Hyperspectral imaging spectrometers work by simultaneously capturing an image in many narrow adjacent spectral bands (Fig. 1). Electromagnetic energy is recorded for each spectral band for each pixel within the image. A radiance spectrum can then be generated for each pixel from the image by plotting the relative brightness for each spectral band as a function of wavelength. This process is repeated for each of the pixels.

The spectral response from an alfalfa plant (Fig. 2) demonstrates several of the spectral features associated with the spectra of vegetation resulting from the chlorophyll and the structure of the plant cells. In a healthy plant (Fig. 2a), chlorophyll absorbs light in the visible part of the spectrum, 400–650 nm, with stronger absorption in the blue and the red as compared to the green part of the visible spectrum, thus producing a small reflectance peak within the green wavelength range, near 550 nm for this plant. The red edge results from the presence of the chlorophyll, which has strong absorption in the visible but not in the infrared part of the spectrum, and spongy leaf mesophyll, which results in high reflectance in the near infrared portion of the spectrum. The location of the red edge shifts toward lower wavelengths by 5–10 nm as plants become stressed due to a decrease in the chlorophyll (Smith et al., 2004a,b). Also, absorption of light in the red part of the spectrum decreases and the red edge tends to flatten out as the plant becomes stressed (Fig. 2b). The high reflectance in the near infrared part of the spectra results from interactions with the internal cellular structure of the leaves. The spectral reflectance drops after about 1300 nm with two water absorption bands that further decrease the reflectance spectra at 1400 and 1900 nm. An oxygen absorption line accounts for the sudden perceived change in the reflectance spectrum at about 760 nm because the signal-to-noise ratio of the spectrum becomes low here. Monitoring changes such as these in the spectrum can give a sign of elevated CO2 levels in the soil.

The method employed for detecting differences in spectra used Random Forest classifiers constructed from training reference data (Breiman, 2001). The training data is a set of plant spectra with known characteristics, such as being nearer to the CO2 release pipe for a longer period of time, leading to greater exposure to elevated CO2 levels. The classifier finds which characteristics of the spectrum give the best separation among the training set classes.

Fig. 1. Schematic of the data structure of hyperspectral images.

Fig. 2. Example reflectance spectrum from an alfalfa plant taken from above. The majority of spectra were collected from the plant leaves. Spectra from (a) a healthy plant and from (b) an unhealthy plant are shown.
This can then be used to predict the classification of plant spectra not from the training set.

From these classifiers a measure related to the plants’ exposure to elevated CO₂ levels can be derived. Regression analysis was used to verify the relationship between this measure and the distance from the plant to the CO₂ release pipe. The results of these analyses are described below.

This paper is organized as follows. Section 2 describes the controlled release facility. In Section 3, the data collection at the controlled release facility is described. An explanation of Random Forest classifiers is provided in Section 4. The data analysis technique is described in Section 5. A discussion of the results from the controlled release experiment is presented in Section 6. Finally, some brief concluding remarks are presented in Section 7.

2. Controlled release facility

The Zero Emission Research Technology (ZERT, 2005) field site is located on a relatively flat 12-ha agricultural plot at the western edge of the MSU-Bozeman campus in Bozeman, Montana at an elevation of 1500 m above sea level. Typical of the geology of the Bozeman area, thick sandy gravel deposits are overlain by several meters of silts and clays with a topsoil blanket. The water table at the site is quite close to the ground surface. The depth to the water table from the ground surface was approximately 1.6–1.7 m during the release of CO₂ occurring August 3–10, 2007. From measurements made the previous year, the direction of the ground water gradient was estimated to be 17° west of north. Although highly variable, measurement of wind direction at the site indicated winds predominantly out of the southeast in the summer months.

The CO₂ was injected through a 100 m long, 10.16 cm diameter stainless steel pipe installed in the shallow subsurface at a depth of 1.8 m with the center 70 m of the pipe being slotted. The horizontal pipe runs southwest to northeast approximately 45° off true north beneath the field. This orientation was chosen to allow resolution of vector components perpendicular to the well of potential CO₂ transport both in ground water and by wind.

The well is partitioned by a system of packers into six isolated and independent zones to improve the chances of getting an even distribution of CO₂ release along the length of the pipe and to increase the flexibility of the system. Because each zone is plumbed separately and has a dedicated mass flow controller, CO₂ flow in each zone can be controlled independently. A uniform flow rate was delivered to the six zones resulting in a total release of 0.3 tons CO₂/day during the August 3–10, 2007 release. The CO₂ injection lasted for an 8-day period. The CO₂ flow rate was chosen in the following way. For a 500 MW fossil fuel burning power plant, approximately 4 Mtons of CO₂ per year could be captured and sequestered. Over a 50-year period this would result in a total storage of 200 Mtons of CO₂. The area of the pipe is assumed to be approximately 0.33% of the area of a typical geologic fault. The flow rate was chosen such that an annual seepage rate through the fault of approximately 0.02% of the total amount stored would be mimicked.

3. Data collection

Hyperspectral imagery was collected using a commercial imager (Resonon, 2008) mounted on a tripod, placing the hyperspectral imager at a distance of about 1 m from the plants being imaged. The hyperspectral imager collected 640 × 640 pixel images covering a ground area of approximately 20 cm × 20 cm. The hyperspectral imager has 160 spectral channels in the 400–900 nm spectral range with a spectral band resolution of 3.21 nm. This spectral range is the primary area in which effects due to plant stress may be observed. In particular, this range includes the red edge, which is the transition region between high absorption in the visible region due to chlorophyll and high reflectance in the near infrared region due to spongy leaf mesophyll. The red edge experiences the greatest degree of spectral change as plant health deteriorates.

Much of the plant matter at the site had senesced at the time of the experiment. An exception was the alfalfa plants. A number of alfalfa plants were thus chosen for observation at various positions about the release line (Fig. 3). Care was taken in choosing the plants such that they were of approximately the same health and flowering stage.

Data was collected by repeated measurements of the selected plants over the course of the CO₂ release. It was acquired at nearly the same time each day, about solar noon. The hyperspectral imager was positioned opposite to the sun such that the angle between the hyperspectral imager’s field of view and vertical was about the same as between the sun and vertical.

Data collection began 2 days after the start of the release and finished 1 day after the end of the release. All of the plants selected were imaged each day except when inclement weather conditions prevented the completion of the data collection. This occurred for part of the data collection on August 5 and for the entire day of August 9.

4. Explanation of random forest classifiers

The Random Forest classifiers (Breiman, 2001) use an ensemble (or forest) of tree-structured classifiers (Breiman et al., 1984), each of which uses a random subset of the training set. The training set contains data samples which have been labeled as belonging to definite classes before the analysis has begun. The tree-structured classifiers are constructed by separating the training set by recursive binary splitting, where the splits are chosen based on the explanatory variables of the training set samples. For each split, a random subset of the bands is considered for selecting the best split. Each tree is grown by continuing to split until each of the lowest nodes contains only a single class. When presented with new data, each tree in the forest returns its own classification of the data. The Random Forest then decides on the classification by taking the plurality vote of the classes returned by each of the trees.
It has been shown (Breiman, 2001) that the Random Forest does not overfit the learning set. The Random Forest misclassification rate approaches the Bayes misclassification rate as the number of trees grows large for training sets that well represent the population.

5. Analysis of the spectral data

The purpose of the analysis of the data collected using the hyperspectral imager was to detect effects in the spectra of the vegetation resulting from the elevated subsurface CO₂ concentrations associated with the CO₂ release. We assumed that distance from the CO₂ release pipe was a reasonable surrogate for CO₂ concentration. The analysis approach developed was based on Random Forest classifiers, described above. The training set included as classes plants that had experienced long exposure to elevated CO₂ levels and ones that have not.

The analysis was begun by choosing a training set. The spectra for the training set were taken from two plants near the CO₂ release pipe to maximize the effects of the CO₂ on the plants. The spectra from these two plants taken on August 6th were classified as healthy for the training set. The spectra from these same two plants taken August 11th were classified as unhealthy for the training set. These dates correspond to early and late in the experiment, respectively. Spectra from these plants were not included in any subsequent analyses. Bias due to individual characteristics of the plants was accounted for by using the same plants on each date, thus allowing analysis of change rather than relative plant vigor.

From each plant image, 102,400 pixels were selected for analysis. This represents one fourth of the total image area and was taken out of the center of the image. Using the image center ensures that only the alfalfa plant being observed is included, since many of the images also included senesced grasses and soil about their edges. This helps to ensure that these extraneous spectra are not included. Every 100 pixels were averaged to reduce the effects of noise, resulting in 1024 spectra from each plant.

Several statistics may be analyzed after construction of the classifier. The out-of-bag misclassification rate, which is a good estimate of the true misclassification rate, was 1%. An estimation of the importance of each band can be examined as well. For this statistic, a variable becomes more important if it is used more often to produce good splits. A plot of the relative importance of the bands revealed that for distinguishing between healthy and unhealthy spectra, most of the important bands are near the red edge of the spectra (Fig. 4). This is expected as explained above.

6. Experimental results

The classifier can be used to analyze spectra from new imagery. The proportion of pixels classified as healthy was taken as a measure of the health for the plant in the image. The proportion of pixels classified as healthy versus distance from the CO₂ release pipe was plotted for each day that data was taken during the controlled CO₂ release (Fig. 5). One point at 17.2 m appeared to be a possible outlier (Fig. 5(f)), so subsequent analysis will consider this date both with and without that point.

The health measure remained fairly high even near the CO₂ release pipe for the days earlier in the experiment. For the two last days in the experiment, August 10 and 11, the health measure dropped to nearly zero for the plants near the carbon dioxide release. August 9 data was not collected due to clouds, so the transition from mostly healthy plants at the beginning to very unhealthy plants near the CO₂ release pipe at the end might be more or less gradual than observed here. No rise in plant health appeared for moderate distances from the CO₂ release pipe or at a time partway into the experiment before the effect became negative. Any fertilization effects due to the CO₂ were either not present or masked by other factors such as generalized plant senescence.

The surface CO₂ flux was non-uniform along the length of the CO₂ release pipe (Lewicki et al., 2007). It would be assumed that non-uniform surface flux would correspond to non-uniform underground concentrations. Thus, plants along the release pipe would receive a variable dosage of CO₂ even though their distances perpendicular to the release would be the same. This effect may reduce the statistical relationship between observed changes in plant health and distance from the CO₂ release pipe.

A linear regression between the distance to the release and the health measure was evaluated for the data from each day. Residual plots indicated that a log transformation of distance would improve regression fits, so this was also analyzed (distances of zero were replaced with 0.01 to avoid undefined log values).

The Adjusted $R^2$ values can be compared between the regressions as the response variable (distance or log(distance)) was the same for each. The Adjusted $R^2$ values indicate the portion of sample variance that the model accounts for and are thus a measure of the goodness of the regression. The $p$-values indicate the significance of the regressions and may be compared as well (Table 1) (Ramsey and Schafer, 2002). It would be expected that a regression between plant health and distance from the CO₂ release would not be significant at the start of the experiment because there was no factor causing vegetation to change as a function of distance from the pipe. Accordingly, the Adjusted $R^2$ values should be small as the model accounts for very little of the variance. As effects caused by the CO₂ becomes more pronounced it would be expected that a regression would become more significant over time. $p$-Values for August 10 and thereafter were highly significant and substantially lower than before that date. The Adjusted $R^2$ values were large at the same time. $p$-Values were not significant.

<table>
<thead>
<tr>
<th>Date</th>
<th>Health to distance regression</th>
<th>Health to log(distance) regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adj. $R^2$</td>
<td>p-Value</td>
</tr>
<tr>
<td>5 August</td>
<td>0.128</td>
<td>0.067</td>
</tr>
<tr>
<td>6 August</td>
<td>0.024</td>
<td>0.649</td>
</tr>
<tr>
<td>7 August</td>
<td>0.023</td>
<td>0.178</td>
</tr>
<tr>
<td>8 August</td>
<td>0.009</td>
<td>0.261</td>
</tr>
<tr>
<td>10 August</td>
<td>0.216</td>
<td>0.003</td>
</tr>
<tr>
<td>11 August</td>
<td>0.085</td>
<td>0.044</td>
</tr>
<tr>
<td>11 August (b)</td>
<td>0.220</td>
<td>0.002</td>
</tr>
</tbody>
</table>
Fig. 5. Plots of health measure versus distance. Health measure for (a) August 5th, (b) August 6th, (c) August 7th, (d) August 8th, (e) August 10th, and (f) August 11th. Plants included in the training set are not shown.
and Adjusted $R^2$ values were small before August 10 except for August 5, which, while significant, had a much higher $p$-value and larger Adjusted $R^2$ value than the later dates. The higher Adjusted $R^2$ value of The August 5 data may be due to the fact that fewer data points were collected this day than on others, resulting in fewer degrees of freedom, 18, for the regression compared to about 35 degrees of freedom for the other days. This was due to fewer data points being taken on August 5 as a result of inclement weather. The significance of the August 5 regression could potentially be related to underlying site properties (such as soil moisture), but we lacked the data for such analysis. These results are largely consistent with a threshold-type response to the CO2, where the threshold was reached between August 8th and 10th, causing the health to decrease rapidly.

Threshold-type responses are a type of nonlinear response common in biological systems. Some examples are those involving positive-feedback processes such as proteolysis and blood coagulation (Jest et al., 1993). Threshold-type functions, the response variable does not change (or changes slowly) with respect to the input variable until the input reaches some value called the threshold. After this value, the response variable responds much more greatly. In the case of this experiment, the threshold in the degree of treatment (CO2 exposure) occurred at some time between August 8 and 10, after which the response (plant health) responded greatly.

The results of the CO2 release experiment can be summarized by looking at the fraction of pixels classified as healthy, versus position for plants that have not yet been exposed to large quantities of CO2 shows little variation in the health measure moving perpendicularly across the CO2 release pipe (Fig. 6). The health measure versus position after the CO2 release had been active for some time shows a very pronounced depression in the health of the plants near the zero position, or directly over the CO2 release pipe (Fig. 7). Away from the CO2 release pipe, the health remains at a higher value. The information in these plots indicates that there was an effect, likely caused by elevated subsurface CO2 levels, that can be discerned from the spectral data collected by the analysis techniques presented above.

7. Conclusions

The results of the analysis indicate a threshold response of plant health to the injected CO2. The logarithm transform of the distance values gave more significant regressions for all of the data. The regressions from the health measure to distance have lower significance values for the days of August 5th to 8th. On August 10th and 11th, the significance of the regressions is far greater than for the earlier days. In each group of days before and after the apparent threshold has been reached there is no clear trend to the significance of the regressions, indicating a threshold response of the plant health to the elevated subsurface CO2 resulting from the injection.

This method of analysis assesses the health of plant spectra. The important bands used in the classifier are near the red edge of the spectra associated with the chlorophyll in the plant. These spectral bands near the red edge change as a plant’s health changes. Applying the classifier to hyperspectral images in the manner described gives a measure of health for plants in the images that vary with distance from the CO2 release in an expected manner. For the 0.3 tons CO2/day injection rate, the plant health deteriorated the most within a perpendicular distance of 0.5 m on either side of the release pipe.

### Symbols

<table>
<thead>
<tr>
<th>Definition</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon dioxide</td>
<td>CO$_2$</td>
</tr>
<tr>
<td>Gigatons carbon dioxide</td>
<td>GtCO$_2$</td>
</tr>
</tbody>
</table>

### Units

<table>
<thead>
<tr>
<th>Definition</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meter</td>
<td>m</td>
</tr>
<tr>
<td>Centimeter (10^{-2} m)</td>
<td>cm</td>
</tr>
<tr>
<td>Nanometer (10^{-9} m)</td>
<td>nm</td>
</tr>
<tr>
<td>Megawatt (10^6 W)</td>
<td>MW</td>
</tr>
<tr>
<td>Megaton (10^6 tons)</td>
<td>Mton</td>
</tr>
<tr>
<td>Parts per million</td>
<td>ppm</td>
</tr>
</tbody>
</table>
Acknowledgements

This work was supported by the Montana Board of Commercialization Technology grant number 08-44, National Aeronautics and Space Administration grant number NNX06AD11G, and the Department of Energy (DOE) grant number DE-FC26-04NT42262. However, any opinions, findings, conclusions, or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the DOE.

References


