



1-5-2018

Estimating the Creation and Removal Date of Fracking Ponds Using Trend Analysis of Landsat Imagery

Rutherford V. Platt
Gettysburg College

David Manthos
SkyTruth

John Amos
SkyTruth

Follow this and additional works at: <https://cupola.gettysburg.edu/esfac>

 Part of the [Environmental Monitoring Commons](#), and the [Oil, Gas, and Energy Commons](#)

Share feedback about the accessibility of this item.

Platt, Rutherford V., Manthos, David, and Amos, John. "Estimating the Creation and Removal Date of Fracking Ponds Using Trend Analysis of Landsat Imagery." *Environmental Management* (5 January 2018).

This is the author's version of the work. This publication appears in Gettysburg College's institutional repository by permission of the copyright owner for personal use, not for redistribution. Cupola permanent link: <https://cupola.gettysburg.edu/esfac/91>

This open access article is brought to you by The Cupola: Scholarship at Gettysburg College. It has been accepted for inclusion by an authorized administrator of The Cupola. For more information, please contact cupola@gettysburg.edu.

Estimating the Creation and Removal Date of Fracking Ponds Using Trend Analysis of Landsat Imagery

Abstract

Hydraulic fracturing, or fracking, is a process of introducing liquid at high pressure to create fractures in shale rock formations, thus releasing natural gas. Flowback and produced water from fracking operations is typically stored in temporary open-air earthen impoundments, or frack ponds. Unfortunately, in the United States there is no public record of the location of impoundments, or the dates that impoundments are created or removed. In this study we use a dataset of drilling-related impoundments in Pennsylvania identified through the FrackFinder project led by SkyTruth, an environmental non-profit. For each impoundment location, we compiled all low cloud Landsat imagery from 2000 to 2016 and created a monthly time series for three bands: red, near-infrared (NIR), and the Normalized Difference Vegetation Index (NDVI). We identified the approximate date of creation and removal of impoundments from sudden breaks in the time series. To verify our method, we compared the results to date ranges derived from photointerpretation of all available historical imagery on Google Earth for a subset of impoundments. Based on our analysis, we found that the number of impoundments built annually increased rapidly from 2006 to 2010, and then slowed from 2010 to 2013. Since newer impoundments tend to be larger, however, the total impoundment area has continued to increase. The methods described in this study would be appropriate for finding the creation and removal date of a variety of industrial land use changes at known locations.

Keywords

Trend analysis, Landsat, BFAST, Hydraulic fracturing, Impoundments, Marcellus shale

Disciplines

Environmental Monitoring | Environmental Sciences | Oil, Gas, and Energy

Estimating the Creation and Removal Date of Fracking Ponds using Trend Analysis of Landsat Imagery

Rutherford V. Platt

Gettysburg College

300 N. Washington Street

Gettysburg, Pennsylvania 17325

David Manthos

John Amos

SkyTruth

213 W Washington St

Shepherdstown, WV 25443

Corresponding author: Rutherford V. Platt, rplatt@gettysburg.edu, 717-337-6078

Citation:

Platt, R.V., D. Manthos, J. Amos. 2018. Estimating the Creation and Removal Date of Fracking Ponds Using Trend Analysis of Landsat Imagery. Environmental Management. 10.1007/s00267-017-0983-4.

Acknowledgements

Thanks to SkyTruth for the FrackFinder PA Dataset. Thanks also to Gettysburg College research assistants Micaela Edelson '17, Sam Donnelly '17, Caitlin Rasche '17, and Cody Kiefer '17.

Abstract

Hydraulic fracturing, or fracking, is a process of introducing liquid at high pressure to create fractures in shale rock formations, thus releasing natural gas. Flowback and produced water from fracking operations is typically stored in temporary open-air earthen impoundments, or frack ponds. Unfortunately, in the United States there is no public record of the location of impoundments, or the dates that impoundments are created or removed. In this study we use a dataset of drilling-related impoundments in Pennsylvania identified through the FrackFinder project led by SkyTruth, an environmental non-profit. For each impoundment location, we compiled all low cloud Landsat imagery from 2000-2016 and created a monthly time series for 3 bands: red, near-infrared (NIR), and the Normalized Difference Vegetation Index (NDVI). We identified the approximate date of creation and removal of impoundments from sudden breaks in the time series. To verify our method, we compared the results to date ranges derived from photointerpretation of all available historical imagery on Google Earth for a subset of impoundments. Based on our analysis, we found that the number of impoundments built annually increased rapidly from 2006-2010, and then slowed from 2010-2013. Since newer impoundments tend to be larger, however, the total impoundment area has continued to increase. The methods described in this study would be appropriate for finding the creation and removal date of a variety of industrial land use changes at known locations.

Keywords

trend analysis, Landsat, BFAST, hydraulic fracturing, impoundments, wastewater, Marcellus shale

Introduction

Hydraulic fracturing, or fracking, is a process for extracting natural gas from layers of shale rock under extreme pressure. Gas production through hydraulic fracturing produces massive amounts of wastewater that represents a potential public health risk. Once a well has been drilled, millions of liters of water, chemicals, and sand are injected at high pressure. Flowback water then returns to the surface for several weeks after hydraulic fracturing, but before gas production begins (Vidic et al. 2013). Thereafter, produced water returns to the surface along with the gas produced by the well. Frack ponds are temporary open-air earthen impoundments that store flowback and produced water.

Many of the chemicals used in fracking are known carcinogens and have been shown to cause health effects related to the skin, respiratory and gastrointestinal systems, kidneys, and endocrine systems (Colborn et al. 2011). People can be exposed to frack chemicals via a number of pathways, including from contaminated groundwater, treated wastewater in rivers and streams, and water in frack pits and ponds. Groundwater may be contaminated when there is poor well integrity. Studies in Pennsylvania, West Virginia, and Ohio have found that there are elevated methane concentrations in drinking well near natural gas wells (Darrah 2014, Jackson et al. 2013, Osborn et al. 2011). Wastewater may be contaminated when treatment is incomplete; treated water in Pennsylvania has been found to have elevated levels of chloride, bromide, barium, and radium above background levels (Warner et al. 2013). Wastewater volume has increased substantially since 2008 and has increasingly been treated on-site, reused, or disposed through injection (Rahm et al. 2013). As of 2016 the U.S. Environmental Protection Agency (EPA) banned the disposal of wastewater at public sewage plants (EPA 2016). Untreated wastewater in frack ponds may also be highly contaminated. A study of evaporation pits in New Mexico revealed high concentrations of chemicals on EPA lists of reportable toxic chemicals (Colborn et al. 2011). Frack ponds are a potentially major source of hazardous air pollution since they contain easily evaporable chemicals such as formaldehyde, acrylamide, naphthalene, and others (Shonkoff et al. 2014). In Pennsylvania, unconventional natural gas activity metrics were found to be associated with increased

asthma exacerbations (Rasmussen et al. 2016). Despite the air pollution potential, no states require air monitoring of waste materials from fracking and other oil and gas facilities (EPA 2014).

To better evaluate the potential public health effects associated with fracking impoundments, it is important to identify where and when fracking activities and wastewater storage have taken place. Past studies have used both manual and automated image classification to identify the location of oil and gas infrastructure. Manual image interpretation has long been successfully used to distinguish oil and gas infrastructure from other types of disturbances (Pasher et al. 2013). A study focusing on the Northern Great Plains successfully used Rapid Land Cover Mapping (RLCM), a manual photo interpretation procedure, to determine the area affected by energy development (Preston and Kim 2016). Other studies have used ancillary data as a starting point for manual classification. For example, researchers examined the (often inaccurate) point locations of wells in the Colorado Oil and Gas Conservation Commission's dataset and then manually digitized actual well location and infrastructure (Baynard et al. 2017). While manual classification is usually accurate, it is difficult to scale up to large areas or time series.

Automated image classification has also been used to identify oil and gas infrastructure, and is more practical over larger areas or time series. In one study, automated detection of oil and gas infrastructure was found to be highly accurate when the infrastructure is associated with large high-contrast forest clearings (Baker et al. 2013, He et al. 2010). However, another study found that when the areas surrounding the oil and gas infrastructure are spatially complex, such as when vegetation regrowth appears, automated detection is characterized by a high rate of commission (Salehi et al. 2014). In areas characterized by sparse vegetation and bare ground automated extraction works poorly (Garman and McBeth 2014). Recently, studies have leveraged temporally dense time series to improve classification accuracy. For example, one study in Alberta, Canada accurately quantified land disturbance from oil and gas activities using Normalized Difference Built-up Index (NDBI) calculated from annual Landsat best available pixel (BAP) composites 2005-2013 (Chowdhury et al. 2017). Another study, also in Alberta, used all available summer Landsat imagery 1985-2012 to accurately identify abrupt changes in the

Normalized Difference Wetness Index (NDWI) associated with oil and gas infrastructure (Pouliot and Latifovic 2016).

To identify both the location and date of fracking impoundments, we used a hybrid method that applies both manual and automated procedures. The location of fracking impoundments is determined using crowdsourced (manual) image interpretation while automated trend analysis of Landsat imagery is used to identify when the impoundment existed. The result is the first dataset of fracking impoundments that includes the estimated date of creation and removal of the impoundments. The dataset is used to calculate the change in number of impoundments and area of impoundments over time, and has potential applications to a variety of environmental and public health research.

Methods

Study Area and Data

The study area for this study is the Marcellus Shale region of Pennsylvania (Fig. 1). The Marcellus Shale is an extensive area of marine sedimentary rock that contains extensive natural gas reserves and underlies parts of New York, Ohio, West Virginia, and Pennsylvania.

Fig. 1: Study Area, Impoundment Locations, and Landsat Path/Row

Impoundment Locations

To identify the location of impoundment ponds in the Marcellus Shale region of Pennsylvania, we used the ‘FrackFinder PA’ dataset created by SkyTruth, a remote sensing nonprofit. The dataset was created through crowdsourcing of a manual image analysis process. The process for identifying fracking impoundments comprised four main steps (Wurster 2014):

1. SkyTruth experts identified the probable location of active wellpads by compiling and analyzing publically available data about unconventional wells reported to the Pennsylvania Department of Environmental Protection (PADEP 2016).
2. Crowdsourced volunteers visually inspected those locations to identify active wellpads using high-resolution aerial survey photography on 2-3 year intervals.
3. Crowdsourced volunteers visually identified all standing bodies of water within 0.5 km of wellpads.
4. Crowdsourced volunteers classified standing bodies of water as fracking related.
5. SkyTruth experts verified crowd response at each stage, and digitized fracking related ponds.

Through this process, SkyTruth identified impoundment ponds that existed in the 2005-2013 timeframe. We used the centroid of the impoundment polygons to represent the point location of the pond (Fig. 1).

Landsat Imagery

To estimate when the impoundments were created and removed, we used Landsat imagery. We downloaded 839 Landsat images from four path/rows (15/31, 16/31, 17/31, 17/32), covering most of the Marcellus shale region of Pennsylvania (Fig. 1). The images came from all available Landsat satellites (5, 7, and 8) January 2000-December 2016, and had less than 30% cloud cover. There were an average of 12 images available per year for a given impoundment point, and 80% of the images fell during the leaf-on period (April-October). To improve comparability between images, we used the Landsat USGS Surface reflectance products (USGS Landsat Surface Reflectance High Level Data Products 2016).

Data Processing

With the above data, we used trend analysis to estimate the date that the impoundments were created and removed. Our methods proceeded in two main steps: (1) creating a temporal composite of all available low-cloud cover Landsat imagery, and then (2) conducting a trend analysis of the temporal composite at each impoundment point (Fig. 2). Trend analysis has been used to identify forest disturbance events and regrowth using MODIS (Schmidt et al. 2015) and Landsat (Czerwinski et al. 2014, DeVries et al. 2015, Hermosilla et al. 2015, Hamunyela et al. 2016). In this case, it is used to identify the creation and removal of fracking impoundments.

Fig. 2: Overview of data processing

Image Compositing

Image compositing was conducted using *R* (R Core Team, 2015) the Raster package (Hijmans 2015), and the Zoo package for time series analysis (Zeileis and Grothendieck 2005). For all 839 images we masked out clouds, cloud shadows, snow, and Landsat 7 gaps using the CFMask layer that is packaged with the USGS surface reflectance product (ESPA Cloud Masking – Release Notes, 2016). We then created temporal composites for NDVI, red, and near infrared (NIR) for each path/row. Temporal composites are multiband rasters where each band represents a different image date (in this case 839 distinct bands). Then, at the location of each of the 1221 impoundment points, we extracted value from each band of the NDVI, NIR, and red temporal composites.

Trend analysis

To conduct the trend analysis, we created a monthly time series for each of the temporal composites (NDVI, NIR, and red) for each impoundment point in the 2000-2016 timespan (1221 impoundment points x 3 temporal composites = 3663 time series). The time series were developed in several steps. First, we

created irregular time series from all available imagery in each temporal composite. The number of observations in each time series depends on factors such as cloud cover and which Landsat satellites were active at a given time. Next, we interpolated missing values for each time series and then aggregated to monthly values (Example output: black line, Fig. 3).

The monthly values have a strong seasonal component (i.e. vegetation ‘greening’ in summer and ‘browning’ in winter). To identify and remove the seasonal variation, we applied the Seasonal Decomposition of Time Series by Loess (STL) procedure (Cleveland et al. 1990). The STL procedure uses the loess smoother and separates time series into seasonal, trend, and remainder components. The “de-trended” data shows the smoothed trend over time after removing seasonal and remainder components (Example output: red line, Fig. 3).

Fig. 3: NDVI for a single impoundment point, showing both monthly interpolation and de-trended (seasonal effects removed)

Next we identified breakpoints in the time series using the BFAST Package in *R* (Verbesselt et. al 2010). BFAST general model is of the following form:

$$Y_t = Tt + St + et \quad (t = 1, \dots, n),$$

where Y_t is the data at time t , Tt is the trend component, St is the seasonal component, et is the remainder component (noise), and n is the number of observed values. The trend component Tt is fitted as piecewise linear model and the seasonal component is fitted as a harmonic model. In this study, the seasonal component (St) had already been removed by STL so St was set to zero.

BFAST iteratively identifies abrupt breaks in the time series from the slope and intercept of the segments of the piecewise linear model shown in Tt . Fig. 4 shows an example of de-trended NDVI and the fitted piecewise linear model (Tt). The dotted lines show the breaks identified by BFAST, creating a temporal segmentation of each time series. For each linear segment in Tt , we can identify the starting digital

number (DN), ending DN, and change in DN. In the next step this information is used to estimate impoundment creation and removal.

Fig. 4: NDVI for a single impoundment point, showing both de-trended and piecewise linear model. Vertical dotted lines show the breaks identified by BFAST (Breaks For Additive Season and Trend) algorithm

Estimating impoundment dates

Impoundments have a fairly predictable lifecycle. When an impoundment is created, vegetation is cleared and then the impoundment is lined with plastic and filled with water. The impoundment may then be filled and emptied of wastewater many times. When an impoundment is removed, first the liquid is emptied and liner removed. Finally, the surface is bulldozed and re-seeded. Using an expert knowledge approach similar to Pouliot and Latifovic (2016), a ruleset was developed to classify four main stages in impoundment lifecycle (Table 1). Each of these four stages is associated with rapid spectral changes that BFAST identifies as breakpoints. The rulesets use the magnitude and duration of change in NDVI and NIR to distinguish the stages. NDVI easily distinguishes vegetation from water, and thus is used for identifying the early stages of an impoundment (Table 1). NIR best distinguishes water from bare ground and is thus used for identifying the later stages of an impoundment (Table 1).

Table 1: Ruleset for identifying stages in impoundment lifecycle

Stage	Description	Ruleset
1	Vegetation cleared	Date of break that starts the largest decline in NDVI in the time series. NDVI must decline below 0.45.
2	Impoundment lined/filled with liquid	Date of break that ends the decline in NDVI described in stage 1. Once impoundments are created, they are often repeatedly filled and emptied of wastewater.
3	Impoundment emptied/liner removed	Date of break that starts the largest increase in NIR in the time series. The increase in NIR must take place after stage 2 and must rise above .25.
4	Impoundment area bulldozed and reseeded	Date of trend break that ends the increase in NIR described in stage 3.

There is no precise definition for the date of the creation and removal of impoundments. Therefore we defined the ‘creation’ of an impoundment as the time right before the impoundment is lined/filled with liquid (4/5 the distance between stages 1 and 2). Similarly we defined the ‘removal’ of the impoundment as right after it is emptied for the last time (1/5 the distance between stages 3 and 4). These distances optimize the accuracy of the classification, as described in the next section. Using these classification rules, the model estimated a creation date. The model also determined whether the impoundment had been removed and if so estimated a removal date. Fig. 5 shows an impoundment point with both a creation and removal date.

Fig. 5: Modeled impoundment creation and removal. Solid vertical lines represent the dates at which an impoundment is estimated to be created and removed. Dashed vertical lines represent the

dates for which imagery is available on Google Earth (images shown at the top of the figure). The blue letter represents the class as determined by an image analyst (H=herbaceous, T=trees, L=impoundment with liquid, E=empty impoundment, B=Bare)

Data Analysis

Using the impoundment creation dates, we evaluated the number of new impoundments built each year disaggregated by the size of the impoundment. We defined “large” impoundments as impoundments that are 0.5 hectares or greater in size, and “small” impoundments as ones smaller than 0.5 hectares.

From the creation and removal dates, we created a frequency distribution of the number of impoundments over time. This was compared to the hectares of impoundments and the number of active wellpads in the Pennsylvania Unconventional Natural Gas Wells dataset (Whitacre and Slyder 2015). We defined an active wellpad as the centroid of a cluster of unconventional wells within 50m of each other with at least one active well in a particular year.

In addition, we calculated descriptive statistics for the following:

- % removed: the percentage of impoundments that have been removed as of December 2016.
- Duration removed: the average duration (in years) of removed impoundments.
- Duration persistent: the average duration (in years) of persistent impoundments as of December 2016.

We evaluated whether the size of the impoundment is correlated with the year built, % removed, and duration removed. We also evaluated the correlation between the year built and duration removed.

Accuracy assessment

To assess the validity of the analysis, we compared the trend analysis-derived dates to date intervals

derived from photointerpretation for all 314 impoundment points in Landsat Path/Row 16/31. For each point there were between 5-8 historical images available in Google Earth, most commonly the years 2005, 2008, 2009, 2010, 2012, 2013, 2014, 2015 and 2016. The image analyst classified the points for each available image in one of 5 land classes (H=herbaceous, T=trees, L=impoundment with liquid, E=empty impoundment, B=Bare). Fig. 5 gives an example for a single impoundment point.

The process was repeated by a second image analyst. If the two image analysts disagreed, a third analyst made the final decision about the class. Land cover classes were then consolidated into two classes:

- 'impoundment absent' -- classified as herbaceous (H), trees (T), or bare (B).
- 'impoundment present' -- classified as impoundment with liquid (L) or empty impoundment (E).

For each impoundment, two intervals were defined: 'creation range' and 'removal range'. The creation interval range for a point is the span of time between the first 'impoundment present' image and the previous 'impoundment absent' image. The removal range for a point is the span of time between the last 'impoundment present' image and the first 'impoundment absent' image. The trend analysis derived dates of creation/removal were compared to the creation and removal ranges. If the modeled dates fell within the creation and removal ranges, the date estimate was considered valid.

Using this strategy for all 314 impoundment dates, we calculated the percentage of creation and removal dates that were valid for both large and small impoundments. In addition, we conducted a chi-square test to evaluate whether impoundment removal was significantly different between the trend analysis derived dates and photointerpretation derived dates. Finally, we conducted a paired t-test to evaluate whether duration removed and duration persistent were significantly different between the trend analysis derived dates and photointerpretation derived dates.

Results

Accuracy Assessment

We found that overall 82% of creation dates and 71% of removal dates were valid (i.e. they fell within the creation and removal range, Table 2). However, the accuracy varied substantially by the size of the impoundment. Large impoundments had a higher proportion of valid estimates, especially for removal date (93% valid). In a few cases, errors were caused when impoundments were removed but not yet reseeded by the end of the time series, making them spectrally similar to empty impoundments that are still in use. In contrast, removal dates for small impoundments were only 53% valid. This may be due to the mixed pixel problem; within a single Landsat pixel, small pits or impoundments can be mixed with surrounding land covers. When pixels are mixed, the temporal signal of the impoundment may become indistinct. It is also common that small on-site pits and impoundments share a cleared area with wells; even when the impoundment has been removed the wellpad may remain.

The accuracy of the estimated creation date varied much less by size of impoundment (79% valid for small impoundments vs. 85% for large). This may be because the clearing of vegetated land to create impoundments is typically a clear and spectrally distinctive event that covers a large area. One source of error is that, in a small number of cases, impoundments were built in locations that had been previously cleared for other purposes, so no ‘break’ was recorded in the time series.

Table 2 – Accuracy assessment, validation subset

	All Impoundments		Large Impoundments (≥ 0.5 ha)		Small Impoundment (< 0.5 ha)	
	Creation	Removal	Creation	Removal	Creation	Removal
# Valid	247	215	115	126	132	89
# not valid	56	88	21	10	35	78
% Valid	82%	71%	85%	93%	79%	53%

The chi-square test indicated that there is a significant difference between the number of impoundments predicted to be removed by the trend analysis vs. photointerpretation ($\chi = 145, p = .000$). Most of the time (84%) the model correctly predicts whether or not an impoundment has been removed (Table 3). This varies somewhat by size; 91% of “large” impoundments are correctly predicted, while 78% of “small” impoundments. The most common error is when the model indicates removal when photointerpretation indicates the impoundment is persistent (Table 3).

It is notable that for small impoundments the model is relatively accurate in predicting if the impoundment is removed (78% accurate) but poor at predicting a valid date (53% accurate). It may be that the impoundments and the surrounding area are spectrally heterogeneous late in the lifecycle of the impoundment (i.e. there are many spikes in NIR, and it is not always clear which is indicative of the removal of the impoundment).

Table 3: Confusion matrix of impoundment state in validation subset

		Photointerpretation		User's Accuracy
		Persistent	Removed	
Trend Analysis	Persistent	131	13	91%
	Removed	35	124	78%
Producer's Accuracy		79%	91%	84%

We compared duration removed and duration persistent for the 84% of impoundments for which there is agreement about whether the impoundment has been removed. The paired T-test revealed no significant difference in duration removed ($p=.082$) and duration persistent ($p=.269$) between the trend analysis and photointerpretation derived dates (Table 4). We also compared duration for all impoundments. The paired T-test revealed a significant difference ($p=.006$), with the trend analysis predicting a slightly longer duration than photointerpretation. This indicates that errors in duration are likely associated with whether the impoundment is predicted to be removed.

Table 4: Paired T-test of duration of impoundments, validation subset

		Duration – – All	Duration – Removed impoundments	Duration - Persistent impoundments as of Dec 2016
Trend Analysis	Mean	4.54	3.1	5.6
	SD	2.1	1.8	1.4
Photointerpretation	Mean	4.2	2.8	5.8
	SD	1.9	1.1	1.5
Paired T-test	T	2.796	1.752	-1.111
	DF	295	130	121
	Sig	.006	0.082	.269

Findings

Based on the analysis of all 1221 impoundments, we found that 76% of impoundments were removed during the 2005-2016 timeframe. The duration of persistent impoundments was 5.2 years as of the end of 2016, while the duration of removed impoundments was 2.4 years (Table 5). We found that there is wide variability in the duration of impoundments, with some impoundments estimated to exist for less than a year and others over 9 years.

Table 5: Summary Statistics, duration of impoundments

	Duration – Removed impoundments	Duration- Persistent impoundments as of Dec 2016
Mean	2.4	5.2
SD	1.6	1.3
Max	8.2	9.2
Min	0.4	2.1

We found that there is a significant but small correlation between the size of impoundments (natural log) and the duration of removed impoundments ($R=.133$, $Sig=0.00$, $N=916$). There is also a significant but small correlation between year built and duration of removed impoundments ($R=-.118$, $Sig=0.00$, $N=916$). In contrast, there is a significant and larger correlation between the size of impoundment and the year built ($R=.469$, $Sig=0.00$, $N=1191$). The more recently built impoundments tend to be larger (Fig. 6). Furthermore, 90% of small impoundments have been removed, but only 40% of large impoundments have been removed. These finding reflects the fact that though large impoundments tend to have a similar duration to small impoundments, they also tend to be more recently built.

Fig. 6: Number of new impoundments built by year

Finally, we found that the total number of impoundments increased rapidly from 2006-2010 and then plateaued from 2010-2013 (Fig. 7). In contrast, the total area of impoundments increased all the way through 2013, reflecting the fact that recent impoundments are getting larger in size. In 2010, where lines cross in Fig. 7, the average area of impoundments reached one acre. For reference, active wellpads increased rapidly to 2011 and growth slowed through 2013 (Fig. 7).

Fig. 7: Count of active wellpads (derived from PADEP data), persistent impoundments (predicted by trend analysis), and impoundment area over time

Discussion

There are several illuminating trends in the number and size of impoundments over time. From 2006-2010 the total number of built impoundments increased rapidly. From 2010-2013 the number of built impoundments slowed to the point that it equaled removed impoundments, leading to a fairly constant number of impoundments. Since new built impoundments tend to be larger than the older removed impoundments, the total impoundment area continued to rise through 2013. The increase in the area of impoundments roughly mirrored the total number of active wells through 2013. From 2013-2016, the number of active wells declined slightly. Since we do not yet know the locations of impoundments created after 2013, we do not know if the total number of impoundment or impoundment area has also declined post-2013.

The duration of individual impoundments is short, averaging 2.4 years for removed impoundments from (2005-2016). Duration is also highly variable, with some impoundments existing for less than a year, and others lasting 8 or more years. Duration appears not to be changing over time, as it is minimally correlated to impoundment size or the year built. Knowing the locations and dates of impoundments can be useful for epidemiological studies. However, because impoundments are spatially and temporarily correlated with wells, may be difficult to distinguish the health effects of impoundments from other fracking-related activities.

We found that the trend analysis yielded a high level of accuracy for estimating impoundment creation and removal. In addition, the trend analysis estimates of duration are similar to photointerpretation estimates. That said, there were two major sources of error. The first major source of error relates to impoundment size. We found that impoundments with valid removal dates were 218% larger on average than impoundments with invalid removal dates (0.8 vs. 0.25 hectares), suggesting that it is easier for the

model to identify removal dates of large impoundments. In Pennsylvania the centralized storage pits must be double-lined with plastic, while smaller on-site ponds do have no such requirements. This difference in requirements may be one reason why there is a more obvious spectral shift when larger impoundments are removed.

The second major cause of error relates to impoundments that do not follow the usual temporal sequence. We observed several (uncommon) examples of non-standard sequences in the validation subset, including:

- The impoundment was created on land that was previously cleared of vegetation for other purposes.
- The construction of the impoundment was discontinuous -- weeds or vegetation grew in the middle of the sequence.
- Impoundment was renovated in the middle of the time series (e.g. emptied, liner removed, liner replaced, and then refilled).
- Impoundments were removed but not yet re-seeded by the end of the time series.
- Algae or other photosynthetic vegetation grew on the impoundment before it was removed.

With new regulations expected to take effect, there is reason to believe that the accuracy of the trend analyses will improve. The regulations will ban small on-site waste storage pits for unconventional well sites (PADEP 2016). Centralized impoundments would still be permitted, but drillers will need to apply for a residual waste permit. As we have seen, the removal dates of the larger, centralized impoundments are more likely to be valid. Also, the new regulations will introduce standards for building and removing impoundments, wastewater processing, site restoration, and remediating spills (PADEP 2016). The building and removal standards may reduce the number of impoundments that follow an atypical temporal sequence.

The trend analysis described in this study has good accuracy and yields important insights, but there is room for further methodological refinement. One promising strategy is the Continuous Change Detection and Classification (CCDC) algorithm, which looks at all available Landsat imagery and only flags a pixel as “changed” if it is different from the predicted pixel value for 3 consecutive images (Zhu and Woodcock 2014). The CCDC algorithm is adaptable to a wide range of land covers, and could potentially be tuned to specific features like impoundments. Another promising strategy is to combine trend analysis with geographic object based image analysis (GEOBIA), which segments images into objects, and then classifies objects based on spectral, textural, geometric, and contextual information. A study of industrial disturbances in Alberta Canada used GEOBIA to calculate vegetation condition in objects adjacent to disturbances, which was then used to distinguish oil and gas disturbances from other spectrally similar disturbances (Powers et al. 2015). Combining GEOBIA and trend analysis would allow researchers to classify spatiotemporal segments based on both temporal context (i.e. the approach of this study, as well as Pouliot and Latifovic 2016) and spatial context (i.e. the approach of Powers et al. 2016). Due to computational limitations, few studies combine GEOBIA and trend analysis, but it remains an important avenue of future research (Platt et al. 2016). Finally, trend analysis could be used for continuous monitoring of impoundments (Verbesselt et al. 2012), where image analysts can be sent images for photointerpretation only after a break in the time series near an impoundment location has been identified. These methods could be applied to a variety of ephemeral land use changes associated with oil and gas infrastructure or other industrial development

References

- Baker BA, Warner TA, Conley JF, McNeil BE (2013) Does spatial resolution matter? A multi-scale comparison of object-based and pixel-based methods for detecting change associated with gas well drilling operations. *International Journal of Remote Sensing* 34(5): 1633-1651
- Chowdhury S, Chao DK, Shipman TC, Wulder MA (2017) Utilization of Landsat data to quantify land-use and land-cover changes related to oil and gas activities in West-Central Alberta from 2005 to 2013. *GIScience & Remote Sensing* 54(5):700-720
- Cleveland RB, Cleveland WS, McRae JE, Terpenning I (1990) STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics* 6, 3-73
- Colborn T, Kwiatkowski C, Schultz K, Bachran M (2011) Natural gas operations from a public health perspective. *Human and Ecological Risk Assessment* 17(5):1039-1056
- Czerwinski CJ, King DJ, Mitchell SW (2014) Mapping forest growth and decline in a temperate mixed forest using temporal trend analysis of Landsat imagery, 1987-2010. *Remote Sensing of Environment* 141:188-200
- Darrah TH, Vengosh A, Jackson RB, Warner NR, Poreda RJ (2014) Noble gases identify the mechanism of fugitive gas contamination in drinking-water wells overlying the Marcellus and Barnett Shales. *Proceedings of the National Academy of Sciences* 111(39):14076-14081
- DeVries B, Verbesselt J, Kooistra L, Herold M (2015) Robust monitoring of small-scale forest disturbances in a tropical montane forest using Landsat time series. *Remote Sensing of Environment* 161:107-121
- EPA (2014) Review of State Oil and Natural Gas Exploration, Development, and Production (E&P) Solid Waste Management Regulations. <https://www.epa.gov/sites/production/files/2016->

[04/documents/state_summaries_040114.pdf](#)

EPA (2016) Pretreatment Standards for the Oil and Gas Extraction Point Source Category.

https://www.epa.gov/sites/production/files/2016-06/documents/uog-final-rule_fact-sheet_06-14-2016.pdf

ESPA Cloud Masking – Release Notes (2016) <https://github.com/usgs-eros/espa-cloud-masking>

Garman, SL, McBeth, JL (2014) Digital representation of oil and natural gas well pad scars in southwest Wyoming: U.S. Geological Survey Data Series 800, 7 p., <http://dx.doi.org/10.3133/ds800>.

Hamunyela E, Verbesselt J, Herold M (2016) Using spatial context to improve early detection of deforestation from Landsat time series. *Remote Sensing of Environment* 172:126-138

He, Y, Franklin SE, Guo X, Stenhouse GB (2011) Object-oriented classification of multi-resolution images for the extraction of narrow linear forest disturbance. *Remote Sensing Letters* 2:147-155

Hermosilla T, Wulder MA, White JC, Coops NC, Hobart GW (2015) Regional detection, characterization, and attribution of annual forest change from 1984 to 2012 using Landsat-derived time-series metrics. *Remote Sensing of Environment* 170:121-132

Hijmans RJ (2015) raster: Geographic Data Analysis and Modeling. R package version 2.4-20. Available at: <http://CRAN.R-project.org/package=raster>

Jackson RB, Vengosh A, Darrah TH, Warner NR, Down A, Poreda RJ, Osborn SG, Zhao K, Karr JD (2013) Increased stray gas abundance in a subset of drinking water wells near Marcellus shale gas extraction. *Proceedings of the National Academy of Sciences* 110(28): 11250-11255

Osborn SG, Vengosh A, Warner NR, Jackson RB (2011) Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. *Proceedings of the National Academy of Sciences* 108(20):8172-8176

- Pasher J, Seed E, Duffe J (2013) Development of boreal ecosystem anthropogenic disturbance layers for Canada based on 2008 to 2010 Landsat imagery. *Canadian Journal of Remote Sensing* 39(1):42-58
- Pennsylvania Department of Environmental Protection (PADEP) (2016) Final Regulations for Oil and Gas Surface Activities (Amendments to 25 Pa. Code Chapters 78 and 78a, Subchapter C).
<http://files.dep.state.pa.us/PublicParticipation/Public%20Participation%20Center/PubPartCenterPortalFiles/Environmental%20Quality%20Board/2016/February%203/Fact%20Sheet%20for%20Final%20Ch%2078%20Regulation.pdf>
- Platt RV, Ogra MV, Badola R, Hussain SA (2016). Conservation-induced resettlement as a driver of land cover change in India: An object-based trend analysis. *Applied Geography* 69: 75-86.
- Pouliot D, Latifovic R (2016) Land change attribution based on Landsat time series and integration of ancillary disturbance data in the Athabasca oil sands region of Canada. *GIScience & Remote Sensing* 53(3):382-401
- Powers RP, Hermosilla T, Coops NC, Chen G (2015) Remote sensing and object-based techniques for mapping fine-scale industrial disturbances. *International Journal of Applied Earth Observation and Geoinformation* 34:51-57
- Preston TM, Kim K (2016) Land cover changes associated with recent energy development in the Williston Basin; Northern Great Plains, USA. *Science of the Total Environment* 566:1511-1518
- R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL. <http://www.R-project.org/>
- Rahm BG, Bates JT, Bertoia LR, Galford AE, Yoxtheimer DA, Riha SJ (2013) Wastewater management and Marcellus Shale gas development: trends, drivers, and planning implications. *Journal of Environmental Management* 120:105-113

- Rasmussen SG, Ogburn EL, McCormack M, Casey JA, Bandeen-Roche K, Mercer DG and Schwartz BS (2016) Association between unconventional natural gas development in the Marcellus Shale and asthma exacerbations. *JAMA Internal Medicine* 176(9):1334-1343
- Salehi B, Chen Z, Jefferies W, Adlakha P, Bobby P, Power D (2014) Well site extraction from Landsat-5 TM imagery using an object- and pixel-based image analysis method. *International Journal of Remote Sensing* 35(23):7941-7958
- Schmidt M., Lucas E, Bunting P, Verbesselt J, Armston J (2015) Multi-resolution time series imagery for forest disturbance and regrowth monitoring in Queensland, Australia. *Remote Sensing of Environment* 158:156-168
- Shonkoff S, Hays J, Finkel M. (2014) Environmental Public Health Dimensions of Shale and Tight Gas Development. *Environmental Health Perspectives*. DOI:10.1289/ehp.1307866
- USGS Landsat Surface Reflectance High Level Data Products (2016)
http://landsat.usgs.gov/CDR_LSR.php
- Verbesselt J, Hyndman R, Newnham G, Culvenor D (2010) Detecting Trend and Seasonal Changes in Satellite Image Time Series. *Remote Sensing of Environment*, 114(1), 106-115.
doi:10.1016/j.rse.2009.08.014
- Verbesselt J, Zeileis A, Herold M (2012) Near real-time disturbance detection using satellite image time series. *Remote Sensing of Environment*, 123: 98-108
- Vidic RD, Brantley SL, Vandenbossche JM, Yoxtheimer D, Abad JD (2013) Impact of Shale Gas Development on Regional Water Quality. *Science*. Vol 340, Issue 6134
- Warner NR, Christie CA, Jackson RB, Vengosh A (2013) Impacts of shale gas wastewater disposal on wastewater disposal on water quality in western Pennsylvania. *Environmental Science and Technology* 47(20): 11849-11857

Whitacre JV, Slyder, JB (2015) Carnegie Museum of Natural History Pennsylvania Unconventional Natural Gas Wells. Pittsburgh, PA: Carnegie Museum of Natural History.

<http://maps.carnegiemnh.org/index.php/projects/unconventional-wells/>

Wurster K (2014) SkyTruth FrackFinder PA 2005-2013. Methodology.

http://skytruth.org/wordpress/wp-content/uploads/2014/10/SkyTruth_FrackFinder_PA_2005-2013_Methodology.pdf

Zeileis A, Grothendieck G (2005) zoo: S3 Infrastructure for Regular and Irregular Time Series. Journal of Statistical Software, 14(6), 1-27. <http://www.jstatsoft.org/v14/i06/>

Zhu Z, Woodcock CE (2014) Continuous change detection and classification of land cover using all available Landsat data. Remote Sensing of Environment, 144: 152-171

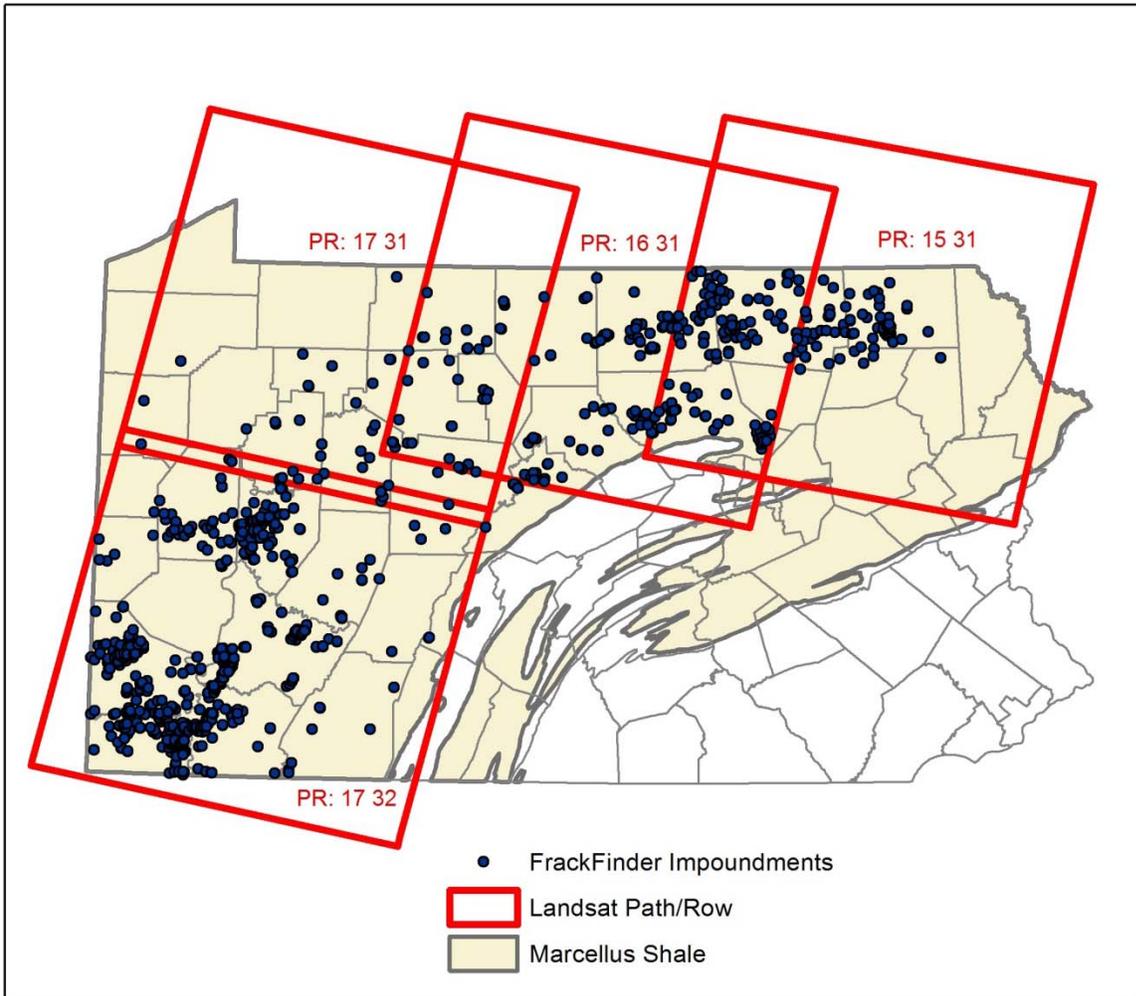


Fig. 1: Study Area, Impoundment Locations, and Landsat Path/Row

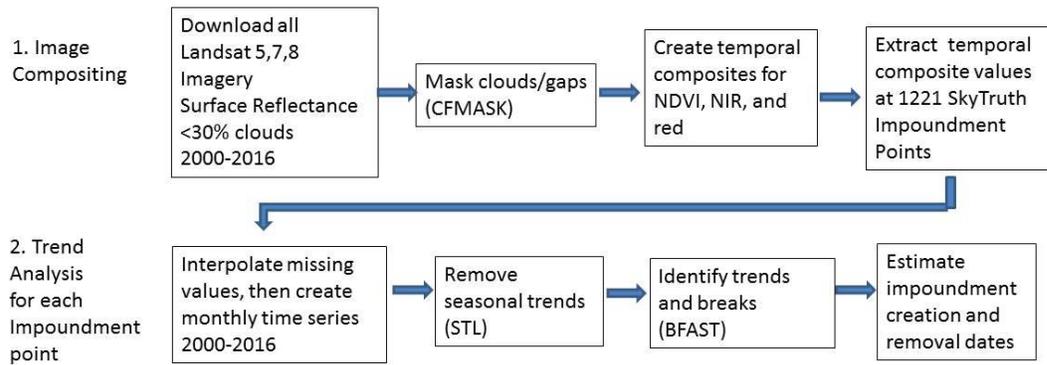


Fig. 2: Overview of data processing. Acronyms used:

- CFMASK: C Function of Mask algorithm, identify fill, cloud, cloud confidence, cloud shadow, and snow/ice in *Landsat imagery*.
- NDVI: Normalized Difference Vegetation Index
- NIR: Near infrared
- STL: Seasonal Decomposition of Time Series by Loess
- BFAST: Breaks For Additive Season and Trend

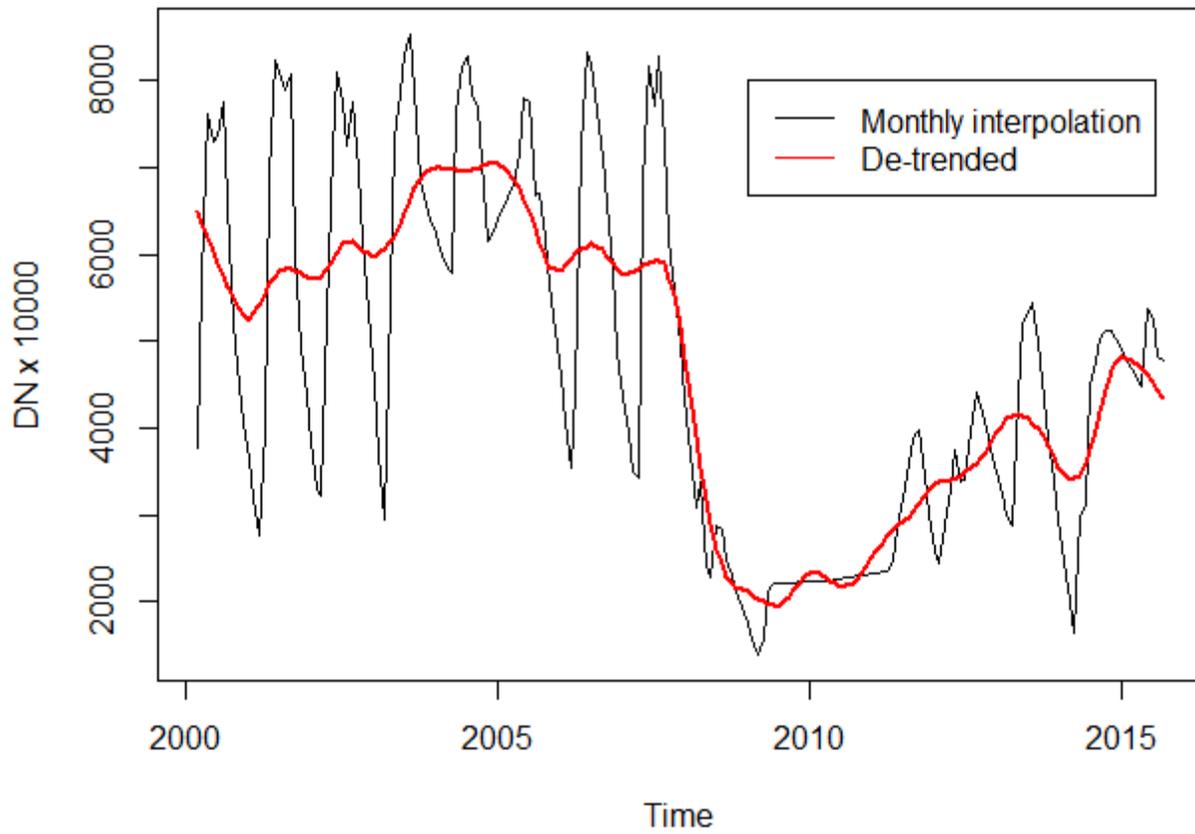


Fig. 3: NDVI for a single impoundment point, showing both monthly interpolation and detrended (seasonal effects removed)

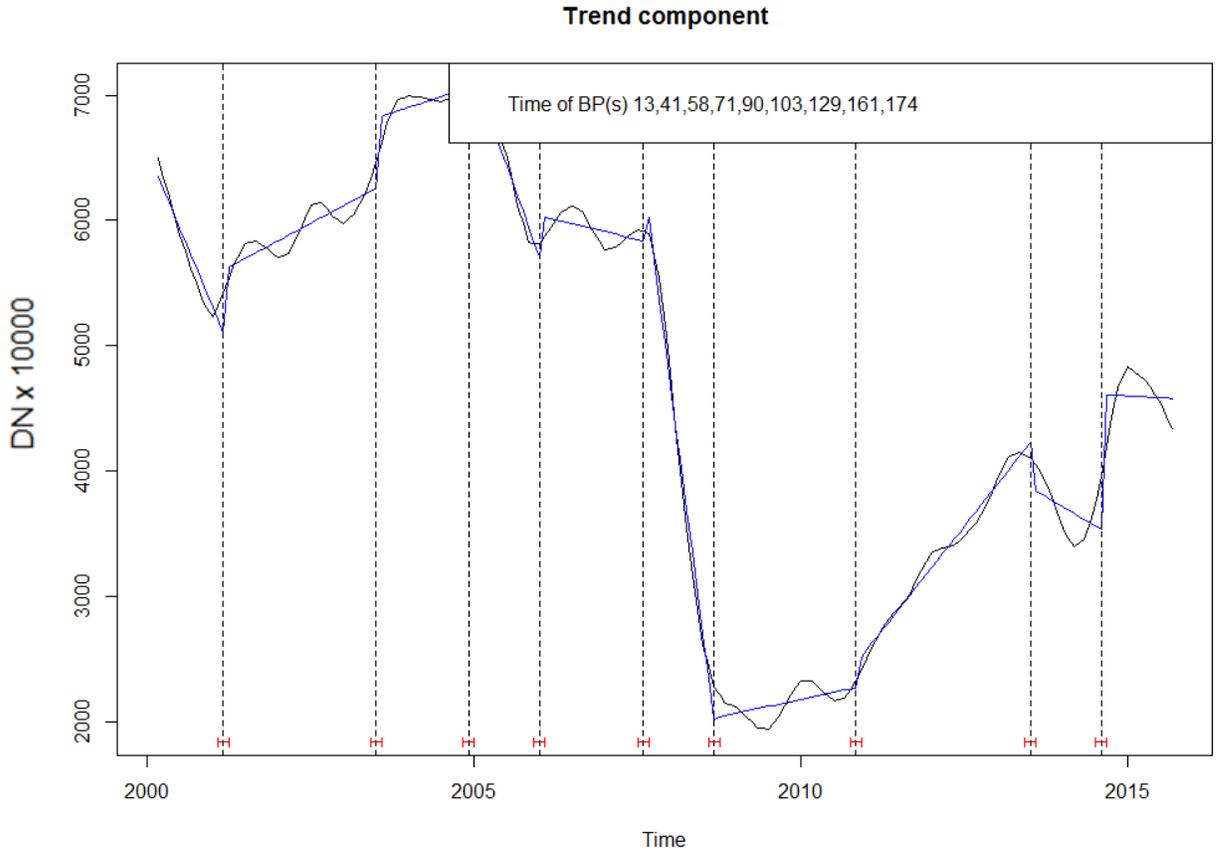
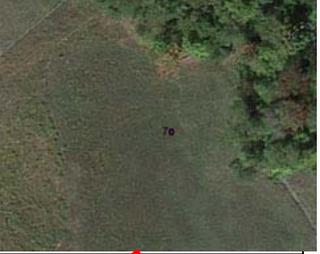


Fig. 4: NDVI for a single impoundment point, showing both detrended and piecewise linear model. Vertical dotted lines show the breaks identified by BFAST (Breaks For Additive Season and Trend) algorithm

5/27/2008 Herbaceous (H)	5/9/2010 Impoundment with liquid (L)	5/5/2013 Bare (B)	9/24/2016 Herbaceous (H)
			

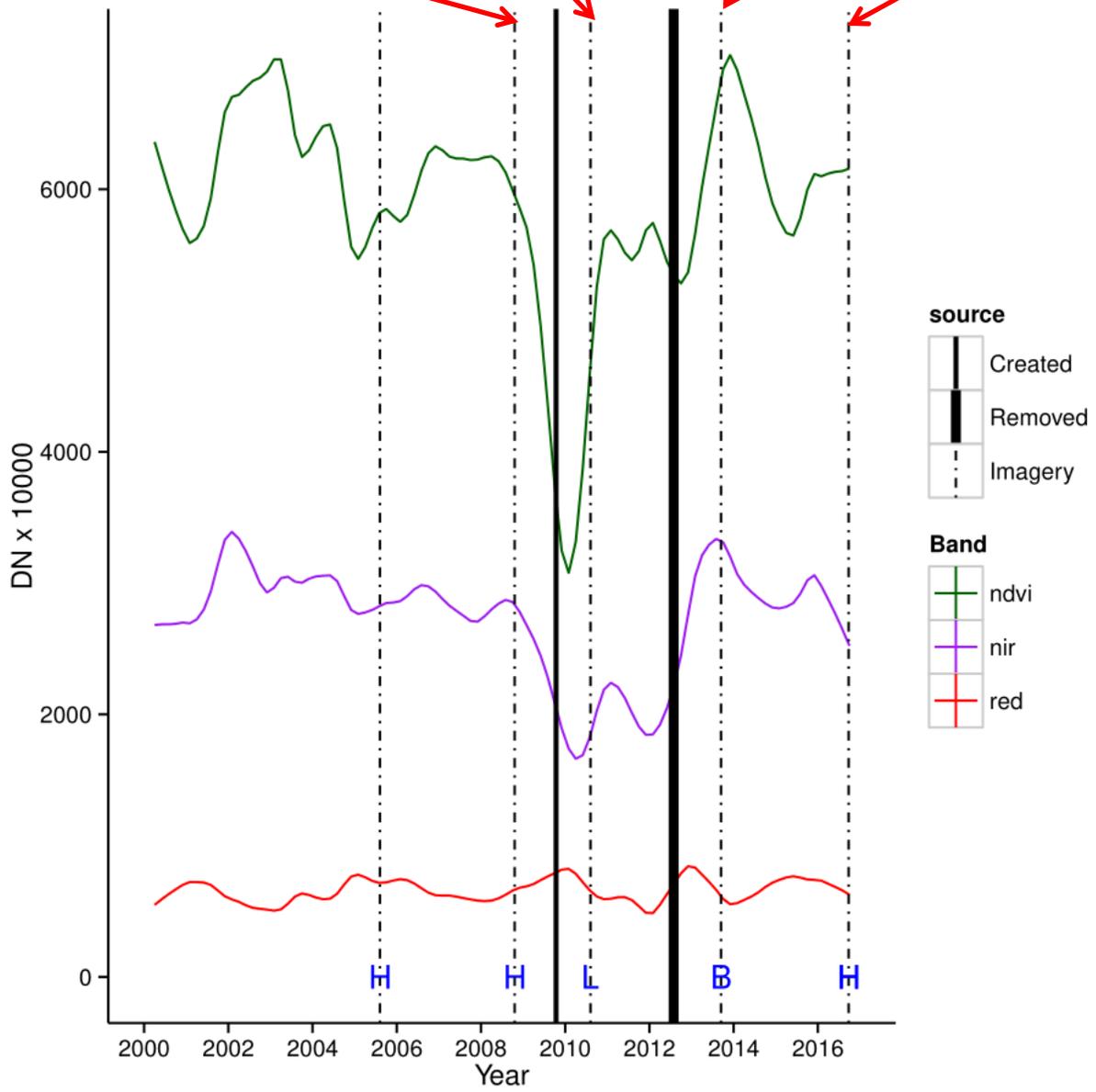


Fig. 5: Modeled impoundment creation and removal. Solid vertical lines represent the dates at which an impoundment is estimated to be created and removed. Dashed vertical lines represent the dates for which imagery is available on Google Earth (images shown at the top of the figure). The blue letter represents the class as determined by an image analyst (H=herbaceous, T=trees, L=impoundment with liquid, E=empty impoundment, B=Bare)

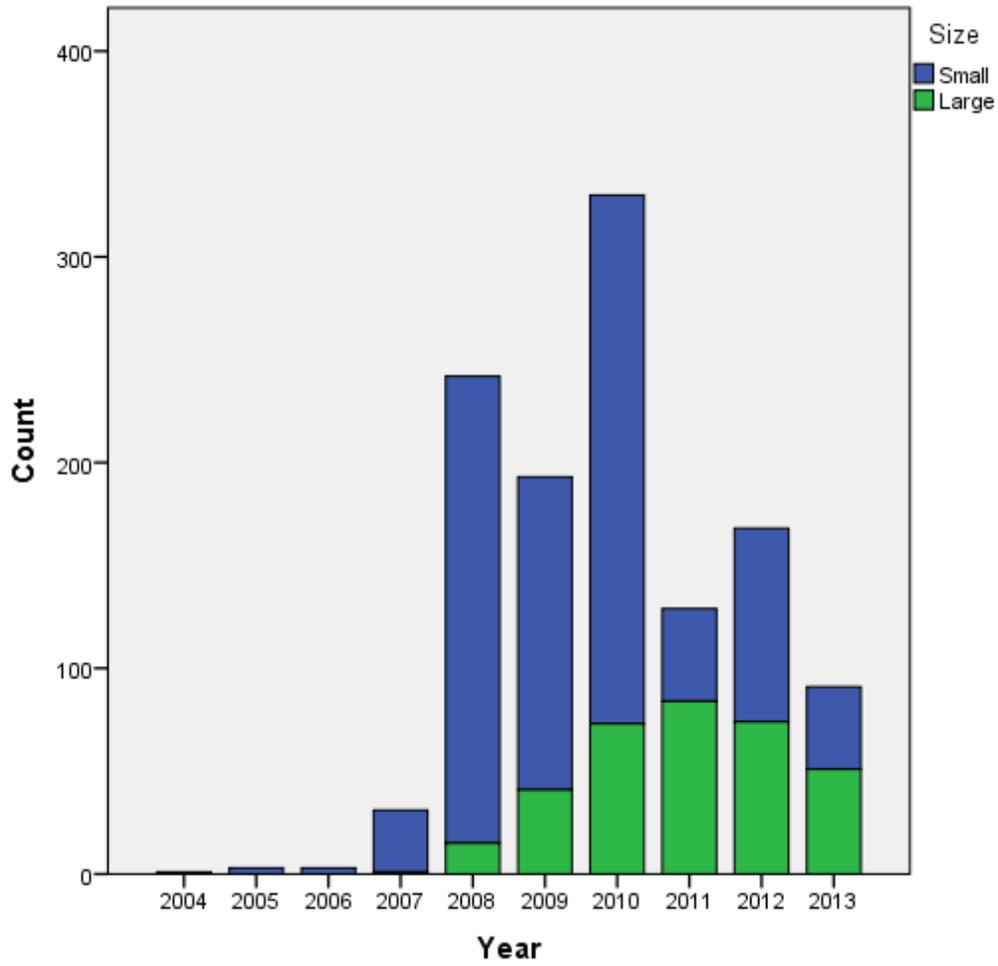


Fig. 6: Number of new impoundments built by year

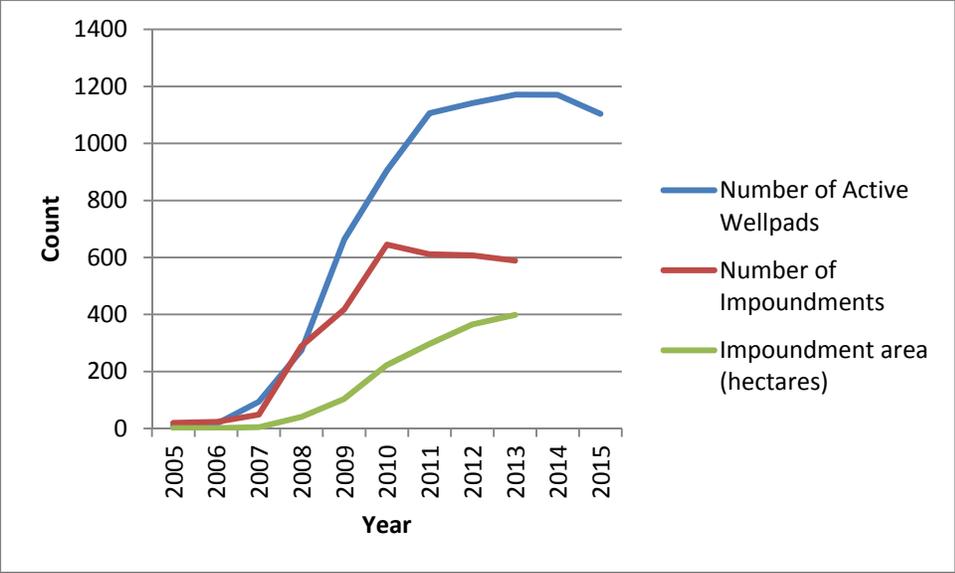


Fig. 7: Count of active wellpads (derived from PADEP data), persistent impoundments (predicted by trend analysis), and impoundment area over time