Agricultural Classification of Multi-Temporal MODIS Imagery in Northwest Argentina Using Kansas Crop Phenologies

September 17, 2014 background.pdf

Jarrett Keifer
Department of Geography

./logo.eps

#### RESEARCH QUESTIONS

#### Can I...

- develop a phenological classification toolset?
- extract crop signatures from Kansas data?
- classify an Argentina study area with the Kansas signatures?

#### RESEARCH QUESTIONS

#### Can I...

- develop a phenological classification toolset?
- extract crop signatures from Kansas data?
- classify an Argentina study area with the Kansas signatures?

#### RESEARCH QUESTIONS

#### Can I...

- develop a phenological classification toolset?
- extract crop signatures from Kansas data?
- classify an Argentina study area with the Kansas signatures?

# OUTLINE

- 1. Background
- 2. Study Areas
- 3. Data and Methods
- 4. Results and Discussion
- 5. Conclusion



- ▶ 1998 to 2002: 940,000 ha deforested
- ► Ley de Bosques passed in November 2007
  - Classified red, yellow, and green areas through the Ordenamiento Territorial de los Bosques Nativos (Land Management Order for Native Forests, OTBN) passed in 2009

- ▶ 1998 to 2002: 940,000 ha deforested
- ► Ley de Bosques passed in November 2007
  - Classified red, yellow, and green areas through the Ordenamiento Territorial de los Bosques Nativos (Land Management Order for Native Forests, OTBN) passed in 2009

- ▶ 1998 to 2002: 940,000 ha deforested
- ► Ley de Bosques passed in November 2007
  - Classified red, yellow, and green areas through the Ordenamiento Territorial de los Bosques Nativos (Land Management Order for Native Forests, OTBN) passed in 2009

## Deforestation in Argentina, 2006 to 2011

Time Period	<b>Hectares Deforested</b>
2006 to Ley de Bosques (2007) Ley de Bosques to OTBN (2009) OTBN to 2011	573,296 473,001 459,108
Total	1,505,405

#### Deforestation in Argentina, 2006 to 2011

Time Period	<b>Hectares Deforested</b>
2006 to Ley de Bosques (2007) Ley de Bosques to OTBN (2009) OTBN to 2011	573,296 473,001 459,108
Total	1,505,405

- ► Deforestation has remained extremely high
- ► Questions the effectiveness of the *Ley de Bosques*

- ► Argentina's soybean cultivation has continually increased
  - ▶ 5 million ha in 1993 to 19 million ha in 2011

- Soy production highly mechanized
- ► Over 99 percent of Argentine soy is genetically modified
  - ► Resistance to glyphosate = heavy pesticide use
- ► Capital requirements cut out small producers

- ► Soy production highly mechanized
- ► Over 99 percent of Argentine soy is genetically modified
  - ► Resistance to glyphosate = heavy pesticide use
- ► Capital requirements cut out small producers

- ► Soy production highly mechanized
- ► Over 99 percent of Argentine soy is genetically modified
  - ► Resistance to glyphosate = heavy pesticide use
- ► Capital requirements cut out small producers

- ► Soy production highly mechanized
- ► Over 99 percent of Argentine soy is genetically modified
  - ► Resistance to glyphosate = heavy pesticide use
- ► Capital requirements cut out small producers

- ► Prevailing perception that soy drives deforestation
- Deforestation research has neglected to analyze specific crop cover

- ► Prevailing perception that soy drives deforestation
- Deforestation research has neglected to analyze specific crop cover

# Goal

Develop a crop mapping toolset which is efficient and economical

# Goal

Develop a crop mapping toolset which is efficient and economical

# Why is this important?

- ► Better understanding of the dynamics of deforestation
- ► More effective land management policies

# Goal

Develop a crop mapping toolset which is efficient and economical

## Why is this important?

- Better understanding of the dynamics of deforestation
- ► More effective land management policies

# Goal

Develop a crop mapping toolset which is efficient and economical

## Why is this important?

- ► Better understanding of the dynamics of deforestation
- ► More effective land management policies

## Problem

#### Problem

- ► A Vegetation Index (VI) can help with crop identification
  - ► Normalized Difference Vegetation Index (NDVI)

#### Problem

- ► A Vegetation Index (VI) can help with crop identification
  - ► Normalized Difference Vegetation Index (NDVI)

$$NDVI = rac{
ho_{NIR} - 
ho_{red}}{
ho_{NIR} + 
ho_{red}}$$

$$NDVI = rac{
ho_{NIR} - 
ho_{red}}{
ho_{NIR} + 
ho_{red}}$$

- ► is a ratioing index
- minimizes multiplicative noise
- ► has issues with non-linearity and additive noise

$$NDVI = rac{
ho_{NIR} - 
ho_{red}}{
ho_{NIR} + 
ho_{red}}$$

- ► is a ratioing index
- minimizes multiplicative noise
- ► has issues with non-linearity and additive noise

$$NDVI = rac{
ho_{\ NIR} - 
ho_{\ red}}{
ho_{\ NIR} + 
ho_{\ red}}$$

- ► is a ratioing index
- minimizes multiplicative noise
- ► has issues with non-linearity and additive noise

# Problem

#### Problem

Must be able to classify crops by type

## Questions

- ► What if two crops have similar VI values on a single date?
- ► How does one determine a crop's VI values?

## Question

What if two crops have similar VI values on a single date?

## Question

What if two crops have similar VI values on a single date?

#### **Answer**

Use imagery from multiple dates.

## TIME SERIES IMAGES

# NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) Sensor

- ► Terra and Aqua satellites
- ► Each images the Earth once per day
- Composite 16-day NDVI imagery at 250-meter resolution

#### TIME SERIES IMAGES

# NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) Sensor

- ► Terra and Aqua satellites
- ► Each images the Earth once per day
- Composite 16-day NDVI imagery at 250-meter resolution

# NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) Sensor

- ► Terra and Aqua satellites
- ► Each images the Earth once per day
- ► Composite 16-day NDVI imagery at 250-meter resolution

# Time Series Image (TSI)

- ► Each band is a 16-day VI composite
- Bands are sequential composites
- Contains enough bands to cover an entire growing season

# Time Series Image (TSI)

- ► Each band is a 16-day VI composite
- Bands are sequential composites
- ► Contains enough bands to cover an entire growing season

# Time Series Image (TSI)

- ► Each band is a 16-day VI composite
- Bands are sequential composites
- ► Contains enough bands to cover an entire growing season

# **Key Points**

- A TSI pixel shows VI values over time
- Each crop's phenology exhibits a unique temporal signature

# CROP TEMPORAL SIGNATURES

Graphics/wardlowCropSignatures.png

(From Wardlow and Egbert 2005)

# Question

How does one determine a crop's VI values?

# Question

How does one determine a crop's VI values?

### **Answer**

Existing approaches require training sites.

# Problem

What if you don't have training sites?

# **Key Points**

- A TSI pixel shows VI values over time
- Each crop's phenology exhibits a unique temporal signature

# **Key Points**

- ► A TSI pixel shows VI values over time
- ► Each crop's phenology exhibits a unique temporal signature

Sounds a lot like hyperspectral remote sensing...

# Idea

Could we use a hyperspectal-like method to fit known crop signatures to unknown pixels?

 ${\tt Graphics/transformations.pdf}$ 

Two-Step Filter (TSF) method from Sakamoto et al. (2010)

- ► Two steps: (1) wavelet smoothing and (2) curve fitting
- ► Curve fitting can fit reference signature to unknown pixels

# TSF Equation 1

$$RMSE = \left[\frac{1}{365/s} \sum_{x=j(0), j(1)...}^{n} (f(x) - g(x))^{2}\right]^{\frac{1}{2}}$$

### where

- ▶ *n* is the number of dates in the TSI
- $\blacktriangleright$  f(x) is the temporal signature for a given pixel in a dataset
- x is the DOY, as defined by j(y)

# TSF Equation 2

$$g(x) = yscale \times h(xscale \times (x + tshift))$$

### where

- ▶ *yscale* and *xscale* are coefficients controlling the vertical and horizontal scaling of a reference signature h(x)
- ▶ tshift is a constant representing the horizontal shift, in days, of h(x)
- $\triangleright$  x is the DOY

 ${\tt Graphics/transformations.pdf}$ 

# TSF Equation 1

$$RMSE = \left[\frac{1}{365/s} \sum_{x=j(0), j(1)...}^{n} (f(x) - g(x))^{2}\right]^{\frac{1}{2}}$$

Minimizing Equation 1 with appropriate constraints on *yscale*, *xscale*, and *tshift* will find the fit of a a reference signature to a pixel.

The signature with the lowest RMSE provides the most probable identification.

# TSF Equation 1

$$RMSE = \left[ \frac{1}{365/s} \sum_{x=j(0), j(1)...}^{n} (f(x) - g(x))^{2} \right]^{\frac{1}{2}}$$

Minimizing Equation 1 with appropriate constraints on *yscale*, *xscale*, and *tshift* will find the fit of a a reference signature to a pixel.

The signature with the lowest RMSE provides the most probable identification.

# Problem

What if you don't have training sites?

# Problem

What if you don't have training sites?

# **Answer**

The TSF equations allow the classification of unknown pixels using a library of crop signatures.



# KANSAS STUDY AREA

- ► 2012 Kansas top crops:
  - ▶ Winter wheat
  - ► Corn
  - ► Soy
- Ground truth: USDA Cropland Data Layer

Graphics/KSstudysite.pdf

# KANSAS STUDY AREA

# Kansas Study Site Planting Dates (adapted from Shroyer et al. 1996)

Crop	Planting Date Range
Wheat	25 September to 20 October
Corn	1 April to 10 May
Sorghum	15 May to 20 June
Soybeans	5 May to 10 June

# DEPARTMENT OF PELLEGRINI Graphics/argentinaOverview landscape.pdf

Graphics/pellegrini75to14\_landscape.pdf

# Deforestation in Pellegrini, 2001 to 2011

Time Period	Hectares	Percent of	Hectares
	Cleared	Land Area	per Year
2001 to 2005	5,968	0.9	1,492
2006 to 2011	75,249	10.9	15,050

# Deforestation in Pellegrini, 2001 to 2011

Time Period	Hectares	Percent of	Hectares
	Cleared	Land Area	per Year
2001 to 2005	5,968	0.9	1,492
2006 to 2011	75,249	10.9	15,050

Annual rate of clearing increased over 1000%!

Top crops in Pellegrini, 2001 to 2005

- ► Soy
- ► Corn
- ▶ Winter Wheat

(From Volante et al. 2005)



- ► 250-meter MODIS 16-day composite VI imagery
- ▶ 30-meter 2012 USDA Cropland Data Layer
- ▶ 30-meter Landsat 8 OLI satellite imagery
- ► Pellegrini boundary shapefile

- ► 250-meter MODIS 16-day composite VI imagery
- ▶ 30-meter 2012 USDA Cropland Data Layer
- ▶ 30-meter Landsat 8 OLI satellite imagery
- ► Pellegrini boundary shapefile

- ► 250-meter MODIS 16-day composite VI imagery
- ▶ 30-meter 2012 USDA Cropland Data Layer
- ► 30-meter Landsat 8 OLI satellite imagery
- ► Pellegrini boundary shapefile

- ► 250-meter MODIS 16-day composite VI imagery
- ▶ 30-meter 2012 USDA Cropland Data Layer
- ► 30-meter Landsat 8 OLI satellite imagery
- ► Pellegrini boundary shapefile

► 2014 Pellegrini Land Cover vector dataset

### PELLEGRINI DATA COLLECTION

- ▶ Data collection in Pellegrini 12 March to 3 April
  - ► 400 random sample points
    - ▶ Direct observation
    - ▶ Interviews with farmers
    - ► Satellite image interpretation
  - ► Agricultural practices and planting/harvesting dates

#### PELLEGRINI DATA COLLECTION

- ▶ Data collection in Pellegrini 12 March to 3 April
  - ► 400 random sample points
    - ▶ Direct observation
    - ► Interviews with farmers
    - ► Satellite image interpretation
  - ► Agricultural practices and planting/harvesting dates

#### PELLEGRINI DATA COLLECTION

- ▶ Data collection in Pellegrini 12 March to 3 April
  - ► 400 random sample points
    - ► Direct observation
    - ► Interviews with farmers
    - ► Satellite image interpretation
  - ► Agricultural practices and planting/harvesting dates

#### PELLEGRINI DATA COLLECTION

- ▶ Data collection in Pellegrini 12 March to 3 April
  - ► 400 random sample points
    - ► Direct observation
    - ► Interviews with farmers
    - ► Satellite image interpretation
  - ► Agricultural practices and planting/harvesting dates

# PROCESSING WORKFLOW

- 1. Reproject the MODIS composite VIs
- 2. Assemble composite VIs into TSIs
- 3. Extract crop signatures from the Kansas TSI
  - 3.1 Identify pure pixels (e.g. non-mixels)
  - 3.2 Use the CDL to isolate each crop
  - 3.3 Identify phenological groups using k-means clustering
  - 3.4 Extract pixel values for each group and average
- 4. Fit the Kansas signatures to the Kansas TSI using the TSF method
- 5. Classify the Kansas RMSE rasters and assess accuracy
- 6. Fit the Kansas signatures to the Argentina TSI
- 7. Classify the Argentina RMSE rasters and assess accuracy

# 1. REPROJECTION

► Land Processes Distributed Active Archive Center's (LPDAAC) MODIS Reprojection Tool

- ► Python command line tool (PCLT) to stack composites
- ► Kansas TSI covered 2012 DOY 97 to 2012 DOY 273
- ► Argentina TSI covered 2014 DOY -13 to 2014 DOY 161
  - ► Aqua DOY 105 composite used in place of DOY 113 composite
  - DOY 129 interpolated from DOY 105 and DOY 145 composites

- ▶ Python command line tool (PCLT) to stack composites
- Kansas TSI covered 2012 DOY 97 to 2012 DOY 273
- ► Argentina TSI covered 2014 DOY -13 to 2014 DOY 161
  - Aqua DOY 105 composite used in place of DOY 113 composite
  - DOY 129 interpolated from DOY 105 and DOY 145 composites

- ▶ Python command line tool (PCLT) to stack composites
- Kansas TSI covered 2012 DOY 97 to 2012 DOY 273
- ► Argentina TSI covered 2014 DOY -13 to 2014 DOY 161
  - ► Aqua DOY 105 composite used in place of DOY 113 composite
  - DOY 129 interpolated from DOY 105 and DOY 145 composites

- ▶ Python command line tool (PCLT) to stack composites
- Kansas TSI covered 2012 DOY 97 to 2012 DOY 273
- ► Argentina TSI covered 2014 DOY -13 to 2014 DOY 161
  - ► Aqua DOY 105 composite used in place of DOY 113 composite
  - DOY 129 interpolated from DOY 105 and DOY 145 composites

- 3.1 Identify pure pixels (e.g. non-mixels)
  - ► Intersected MODIS pixel grid with vectorized CDL
  - ► Selected all features >= 53,000 m² in area

- 3.1 Identify pure pixels (e.g. non-mixels)
  - ► Intersected MODIS pixel grid with vectorized CDL
  - ► Selected all features >= 53,000 m² in area

- 3.2 Use the CDL to isolate each crop
  - ► Resampled CDL to MODIS grid by majority
  - ► Isolated the TSI pixels for each crop
    - ► Corn
    - ► Soy
    - ► Sorghum

- 3.2 Use the CDL to isolate each crop
  - ► Resampled CDL to MODIS grid by majority
  - ► Isolated the TSI pixels for each crop
    - ▶ Corn
    - ► Sov
    - ► Sorghum

- 3.2 Use the CDL to isolate each crop
  - ► Resampled CDL to MODIS grid by majority
  - ► Isolated the TSI pixels for each crop
    - ► Corn
    - ► Soy
    - ► Sorghum

- 3.3 Identify phenological groups using k-means clustering
  - ► ENVI k-means tool
  - ► Three clusters, 1.0% change threshold, 100 iterations

- 3.3 Identify phenological groups using k-means clustering
  - ► ENVI k-means tool
  - ► Three clusters, 1.0% change threshold, 100 iterations

- 3.3 Extract pixel values for each group and average
  - ► Pixels from each k-means cluster converted to points
  - ► PCLT to find mean values from points
    - ► Produces .ref signature files

- 3.3 Extract pixel values for each group and average
  - ▶ Pixels from each k-means cluster converted to points
  - ► PCLT to find mean values from points
    - ► Produces .ref signature files

- 3.3 Extract pixel values for each group and average
  - ► Pixels from each k-means cluster converted to points
  - PCLT to find mean values from points
    - ► Produces .ref signature files

Now we have crop signatures!

# 4, 6. Fit Kansas signatures to <insert study area here> TSI using TSF method

- Not using the TSF's wavelet filtering
- PCLT to fit reference signatures to a TSI using TSF equations
  - Creates a RMSE raster for each signature
  - ► xscale and yscale bounds: 0.6 to 1.4
  - lacktriangledown tshift bounds:  $\pm 10$  days Kansas, 120 to 140 days Argentina

- 4, 6. Fit Kansas signatures to <insert study area here>
  TSI using TSF method
  - ► Not using the TSF's wavelet filtering
  - PCLT to fit reference signatures to a TSI using TSF equations
    - Creates a RMSE raster for each signature
    - ▶ xscale and yscale bounds: 0.6 to 1.4
    - $\blacktriangleright$  tshift bounds:  $\pm 10$  days Kansas, 120 to 140 days Argentina

- 4, 6. Fit Kansas signatures to <insert study area here>
  TSI using TSF method
  - ► Not using the TSF's wavelet filtering
  - PCLT to fit reference signatures to a TSI using TSF equations
    - ► Creates a RMSE raster for each signature
    - ► xscale and yscale bounds: 0.6 to 1.4
    - ► *tshift* bounds: ±10 days Kansas, 120 to 140 days Argentina

- 4, 6. Fit Kansas signatures to <insert study area here>
  TSI using TSF method
  - ► Not using the TSF's wavelet filtering
  - PCLT to fit reference signatures to a TSI using TSF equations
    - ► Creates a RMSE raster for each signature
    - ► xscale and yscale bounds: 0.6 to 1.4
    - ▶ tshift bounds:  $\pm 10$  days Kansas, 120 to 140 days Argentina

- 4, 6. Fit Kansas signatures to <insert study area here> TSI using TSF method
  - ► Not using the TSF's wavelet filtering
  - PCLT to fit reference signatures to a TSI using TSF equations
    - ► Creates a RMSE raster for each signature
    - ► xscale and yscale bounds: 0.6 to 1.4
    - ▶ tshift bounds:  $\pm 10$  days Kansas, 120 to 140 days Argentina

- 5, 7. Classify <insert study area here> RMSE rasters and assess accuracy
  - ► Signature with lowest RMSE value is best identification
  - ▶ Not all RMSE values should be considered
    - ► If all signatures have high RMSE values, none are a good match

- 5, 7. Classify <insert study area here> RMSE rasters and assess accuracy
  - ► Signature with lowest RMSE value is best identification
  - ▶ Not all RMSE values should be considered
    - ► If all signatures have high RMSE values, none are a good match

- 5, 7. Classify <insert study area here> RMSE rasters and assess accuracy
  - ► Signature with lowest RMSE value is best identification
  - ▶ Not all RMSE values should be considered
    - ► If all signatures have high RMSE values, none are a good match

- 5, 7. Classify <insert study area here> RMSE rasters and assess accuracy
  - ► Signature with lowest RMSE value is best identification
  - ► Not all RMSE values should be considered
    - ► If all signatures have high RMSE values, none are a good match

- 5, 7. Classify <insert study area here> RMSE rasters and assess accuracy
  - ► Signature with lowest RMSE value is best identification
  - ▶ Not all RMSE values should be considered
    - If all signatures have high RMSE values, none are a good match

#### Problem

How do we determine which RMSE values to consider?

# Problem

How do we determine which RMSE values to consider?

#### Problem

How do we determine which RMSE values to consider?

## Solution

Brute force through different threshold combinations, classifying and assessing the accuracy to find the best result.

#### Problem

How do we determine which RMSE values to consider?

# Solution, kind of

Brute force through different threshold combinations, classifying and assessing the accuracy to find the best result.

- 5, 7. Classify <insert study area here> RMSE rasters and assess accuracy
  - ▶ PCLT to iterate through user-defined range of thresholds
  - ► Classification and accuracy rasters with best combination

- 5, 7. Classify <insert study area here> RMSE rasters and assess accuracy
  - ► PCLT to iterate through user-defined range of thresholds
  - ► Classification and accuracy rasters with best combination

- 5, 7. Classify <insert study area here> RMSE rasters and assess accuracy
  - ► PCLT to iterate through user-defined range of thresholds
  - ► Classification and accuracy rasters with best combination



# **RESULTS**

- ► Summer 2014 Pellegrini ground truth
- ► Pellegrini agricultural practices
- Kansas crop signatures
- ► Kansas classification
- ► Pellegrini classification

#### PELLEGRINI GROUND TRUTH

- ➤ 378 of 400 sample points were identified
- Many additional fields were collected

Graphics/collecteddata.pdf

# PELLEGRINI GROUND TRUTH

# **Summer 2014 Pellegrini Land Cover Classes**

Cover Type	Hectares	Sample Points
Forested	389,541	247
Other	42,229	22
Corn	41,488	36
Pasture	35,057	37
Soy	27,498	24
Poroto	9,539	7
Nothing	3,057	3
Sorghum	1,646	2
Unknown	92,248	17
Omitted	52,052	5
Total	694,346	400

# PELLEGRINI AGRICULTURE

# **Key Dates for Pellegrini Summer Crops**

Crop	<b>Ideal Planting Range</b>	<b>Harvesting Begins</b>
Soy	15 December to 15 January	1 May
Corn	15 January to 15 February	1 June
Sorghum	15 January to 15 February	1 June
Poroto	15 January to 20 February	10 May

# KANSAS CROP SIGNATURES

Graphics/KSclustered.pdf

#### KANSAS CROP SIGNATURES

plots/allsigsKS.pgf Crop Signatures Extracted from the Kansas TSI Crop Clusters

