June 1

**Desk**

Landsat Phenology and Landsat Phenology v2 in GEE – have to change the rally da safra csv file that Avery sent (“blah” for year and “lat and long” for x and y)

Masked clouds using Landsat surface reflectance pixel quality band, moved from looking at a single point to many points of rally data locations

Read Environmental Biophysics textbook

June 4

**Desk**

Landsat Phenology v2 in GEE – export LAI timeseries data as table. Includes making separate date and lat/long columns, getting rid of .geo column, reducing by soy points, adding pointID to soy points.

Landsat Phenology v3 in GEE – work on piecewise LAI function with the image.where() function, but need to make sure not all bands are replaced by LAI band

June 5

**Desk**

Re-read phenology finding papers to see if it’s possible to get planting, harvest date in GEE. Make initial plan for determining phenology from these papers.

June 6

**Desk**

Worked on Landsat Phenology v3. Tried more LAI equations – none of them look better than the gLAI equation using SR because they all give negative LAIs.

Begin Modis Phenology in GEE.

**Meeting**

For phenology from satellite: first think about how we will separate different peaks due to crops, and how to separate the weed peaks from the crop peaks. Think about an algorithm to divide the timeseries into chunks in which we expect a planting and a harvest date. Think about all the ways things could go wrong and try to make an algorithm that can account for those situations. Look for peak detection methods (i.e. filter out the weed peaks by looking at thresholds in peak characteristics). Harvest, at high resolution, should look like a period of steepest descent. So inflection may be questionable.

For LAI equations: try two things. (1) relate gLAI from SR to LAI; (2) ask Gabriel for a max LAI and adjust KrossSR LAI equation so that it generates LAI values that are between zero and his max LAI. Compare the two and see if they match. Ask Gabriel for Portuguese papers about calculating LAI for soy in Brazil.

June 7

**Desk**

Read MODIS papers to choose which MODIS product to use for phenology

**Meeting**

Clean up Avery’s Mapitoba planting date dataset – 100 farms and 100 questions per farm – and use it for validation of our planting date estimate

Two ways to use INLAND: either as a huge curve fitting exercise; or as representation of real world situations. In which case we need to think where parameters need to be conserved among locations; parameters should mean something quasi physical. There should be an ensemble of parameters that are applicable at more than one place and one time. The different end members approach might be more appropriate. It’s hard to use curve fitting in a predictive way.

We run multiple combos of parameters at each location for calibration. When we do yield regression, we control for things we can observe like planting date.

If we have a large dataset for Level 2 calibration, we won’t be overfitting to “bend over backwards” to account for a bad planting date. Is there robustness to planting date assumption within a certain (low, med, high) yield? Do you always get “low yield” for different planting dates?

Level 2 is just for identifying ranges of getting low vs high yield – see if the management parameters are independent of planting vs harvest date.

Rally dataset WDRVI: used MODIS and smoothed with TIMESAT. MOD09A1 – 8 day composite – has specific day of year of each pixel. MOD means Aqua – overpasses in the morning and has longer time domain, mornings tend to be more cloud free. MYD means Terra.

For Gabriel, negative LAI values during fallow is ok. He’s more worried about magnitudes. We care more about LAI = 1 to the max LAI.

Veg indices are a problem at low latitudes because of atmospheric composition there. Avery wants to look at SIF and radar backscatter for bias correcting.

MODIS LAI is more for natural vegetation. Be careful because they’re classified by land cover classes.

June 8

**Desk**

In GEE, Modis Phenology and Modis Phenology v2: work on getting actual image acquisition date into the output table and in the chart. Did two reducers, joined them by system:index, and then flattened into a new FeatureCollection, then finally put it in a chart.

Began Modis Phenology 3, where the chart generation happens in a separate script called PhenologyFunctions.js

June 11

**Desk**

Worked on getting LAI points from Modis.

Gabriel says to try KrossSR VI-LAI equation, and also the Nguy-Robertson GCVI-gLAI equation. The “gLAI” in that paper is measured the same way as he’s measured LAI, so gLAI in that paper might be okay.

Play around with ways to mask clouds in Modis MYD09A1 product – using two bitmask methods. Added a mask to filter out sensor zenith angles greater than a threshold. Don’t use NBAR correction because it might add a systematic error.

June 12

**Desk**

Clean pesquisa MATOPIBA planting date survey

Summarize planting date finding methods into a file for Avery; sent him via Slack

June 13

**Desk**

Looked up bitmasks in general and in GEE for MODIS – see June 13 PPT for summary

Began GEE file called Matopiba Survey to plot planting/harvest data alongside Modis derived LAI and then export the information

**Meeting – Sally**

For the MODIS derived phenology curves, perhaps set a max allowed LAI change to filter out spurious fluctuations.

Try to set partial clouds as OK, but then correct the resulting LAI values (check if this is possible)

Maybe throw out pixel-years where we suspect the peak LAI is being lost to clouds – but check how much data we’d be losing if we go this route.

Try Terra in addition to Aqua

Try cloud masking the MODIS LAI product

Try to fuse Landsat and MODIS?

Think about taking remote sensing class with Iryna

Read into the physical meaning of VI-LAI equations, maybe that will help us tweak it.

June 14

**Desk**

Clean pesquisaMatopiba data by adding errors to the verbal dates and read into GEE’s Matopiba Survey script. The pesquisaMatopiba\_cleaned\_toGEE.xlsx file doesn’t have all the survey questions; only the ones with planting/harvest date and yield. The pesquisaMatopiba\_cleaned\_full.xlsx file has all the data.

**Meeting, Avery and Gabriel**

PAM data is county level data that the government produces about soy and corn yield and acreages

For LAI points:

Overlay the Matopiba/rally points with Land cover class in Mapbiomas and CAR field polygon data so we know that the LAI info was taken for an actual soy point rather than a road, etc. wait for Jake to do this.

To get more LAI points, look at Landsat-MODIS fusion (STAIRS) from Kaiyu Guan group – ask Avery for code.

Try Sentinel data and Terra in addition to Aqua

Data

Double crop info is on Github and also on Tufts cluster.

Avery will share stuff on Github

I need to put cleaned Matopiba planting/harvest date info as my first repository in Github

Learn Github

Getting planting/harvest from phenology

Use data about variety to get cycle length information, which can then inform the frequency of the wavelet used to smooth EVI series

Gabriel will send information on the relationship between planting date and observed emergence

Test a weighted average of different ways to estimate planting/harvest – try an ensemble of methods

For the retreat:

Propose a method for filtering and combining sensor data

Propose a method for estimating (and taking ensemble of) planting and harvest dates

Propose a workflow – i.e. where to keep data, where to do smoothing analysis, etc

Create map of the amount of data that we have

Propose a way to do Gabriel’s sensitivity scenarios for management

June 15

**Desk**

GitHub tutorials; put MatopibaSurvey on as a repository

Imported Brazil regions into GEE; it’s just called “regions” and it’s under my assets

Started Matopiba Survey v2 and v2, which looks at more than one point over a given year – and incorporates the errors into planting/harvest date range

**Meeting Gabriel**

He prefers Matopiba LAI and planting/harvest dates on the same file. Put it on GitHub.

June 17-18

**Desk**

Prep AgroServe retreat

Read remote sensing basic textbook

June 19

**Desk**

Work on Matopiba Survey v3 in GEE; add on CAR field polygons and Mapbiomas to see if survey points overlay with them. Also, before this, had only Modis Aqua; added Modis Terra.

June 20-21

AgroServe Retreat

General project info

Goal of the project is: line item in cost-benefit analysis for agribusiness. More broadly, we are looking for feedback loops, thresholds, safe operating space and climate-agri-econ nexus

Cerrado is grassland and forest. It’s a political boundary for environmental policy. There’s a lot of new agriculture in the region, and a lot more still available. WEF (a group of agribusinesses like Walmart, land investment, etc all along the supply chain) made an announcement to decrease deforestation in Cerrado. The CWG (Cerrado Working Group – NGOS + agribusinesses) propose a pledge to stop deforestation, regardless of legality. They therefore need information about ecosystem services to determine details on stringency, etc. it’s tricky because there are many types of land, such as public undesignated areas, protected areas, reserve areas on farm, etc.

Our time period is 2030, so we can ignore “chemical” climate change (GHGs)

Cattle farmers are smaller, less modern, less educated than soy planters; less likely to comply with laws

SEALS: a coarse land use map is used to generate a high-res land use map based on land suitability

(Rittel and Weber, 1973) – decision making under deep uncertainty

Outcomes of AgroServe:

Maybe put our data on AgroIdeal – a WebGIS with maps used by NGOs

Produce a talking points memo on model results, i.e. ecosystem services in economic terms, robustness of estimates, how we calculate it, etc

Stakeholders care a lot about likelihood of recurring bad years and spatial correlation of risk (e.g. soy mill that sources in a region – how likely will there be a change in typical climate?

To do

Jake will give me CAR polys – area explorada (which has managed regions only) – overlay this with Matopiba data points to see if the ones in area explorada have LAI values that make more sense

Match up Matopiba and CARpolys by farmer name

See if Dave or Jake can upload GOME data

Atmospheric stuff and off nadir stuff for MODIS makes it unacceptable to calculate LAI from MODIS alone; SIF helps with atmospheric stuff and Landsat helps with off-nadir stuff

Overlay MODIS and Landsat for a one-week plant/harvest attempt. Use harmonic smoothing in GEE to start smoothing LAI curve.

Fit planting date to onset map for running hypothetical pixels. Ask for onset map from Gabriel.

Think about ways to do a basic statistical crop model instead of INLAND (for the summer)

Think about paper on plant/harvest date. Will it be a methods paper (RS? Hydrology? Agri?) or application paper?

June 22

**Meeting – Avery, Jon, Gabriel**

We care about random fluctuations of the weather, and about variation across space in yield – one approach is yield = fcn(mean climate in country, year-to-year variation in climate in the country) – looked at precipitation and temperature

Extreme heat, vapor pressure deficit, rainy season properties, onset duration, dry spells in rainy season, flooding rain at harvest: the biggest things that matter. The effects of these depend on when in the phenology it occurs. Divide the growing season into vegetative and fruiting.

Variation in soy yield isn’t enough; need to look at propensity to double crop, yield of second crop, ability to grow crops at all

Include soy yield, double crop or not, and corn yield

Need to think about correlations among climate variables.

Assume farmers are doing best they can – how much are they constricted by planting date, etc based on the changing climate?

Land scenarios are for generating climate scenarios – but land use scenarios have nothing to do with what pixels we actually model. Model for representative sample of actual soy pixels.

Idea: bisect by before and after peak LAI, calculate the climate variables before and after it, then train the crop model. We’re talking about a local fixed effects (“de-meaning”). Approximate peak LAI using change in onset. Then the climate variables (before and after peak) will depend on the time of the peak date.

With Gabriel - Come up with LAI peak estimation strategy for rally points and for Mapbiomas cropland area in the Cerrado municipalities. Write code that finds both first and second peak. Right now we only care about timing. Do something like Timesat – come up with greenup and browndown dates, cycle length proxy. Variables: first peak, second peak, greenup and browndown; for each: 25th, 50th and 75th percentile for day of year. DOY should be zero on Aug 1. Implement this in Earth Engine.

Write up dataset: area of soy in each municipality, area of double cropped maize in each municipality, temp, precip, greenup/browndown dates, Xavier dataset, local prices of soy and maize to get county level changes in revenue. Everything we need to do crop model regression. I’ll own assembly of the dataset, coordinate with Jake and Gabriel.

Rally data is pointwise so it’s associated with a particular management strategy, but doesn’t have info about harvest loss. Also do municipal data for yield and do panel regression.

Gentine paper about SIF – read in detail

June 23

**Desk**

Timeseries analysis tutorial in GEE

Worked on Timeseries Analysis GEE script; overlaid Landsat and MODIS LAI to compare… Landsat doesn’t look great, is usually lower than MODIS. Perhaps begin harmonic fitting without Landsat

Worked on a function to turn DOY into an actual date

June 24

**Desk**

Worked on Timeseries Analysis v2 GEE script; did the most basic harmonic series on Modis for only single crop example and no peak detection, used fitted harmonic terms to calculate max, min and inflection points. Assumed a period of one year; but we could instead take a Taylor series expansion of the cosine function and get frequency as a “linear” term and then calculate frequency instead of having to guess one. This has the danger of “overfitting” – we have only amplitude, frequency, and phase to calculate from six fitted parameters. Doesn’t make physical sense.

June 25

**Desk**

Worked on Timeseries Analysis v3 in GEE script; try to apply smoothing, then take derivatives of the smoothed function in order to get estimate of peak times and inflection point locations. The first derivative itself was smoothed again. This is as a first-step peak detection for future harmonic fitting, since the missing data points and the use of system:time\_start instead of actual pixel acquisition dates will mess up the analysis. Decided to ignore the second derivative, since it just gets really messy. The max of LAI will determine the peak location and the max of derivative will determine the inflection point location. The difference between location of max LAI and max LAI derivative will help determine estimate of

Started Timeseries Analysis v4 in GEE – implement finding max LAI, max first derivative. There is an issue that the max occurs in the second peak which probably isn’t soy; try to work around it by finding the max before March 1 (or some value determined by the region?)

June 26

**Desk**

Tested Timeseries Analysis v4 in GEE for all Matopiba points – decide will need single/double cropping map to more accurately find the peaks.

Worked on a way to get second peak timing (by basically repeating the method to find first peak), did this for a single point in Timeseries Analysis v5.

**Meeting with Gabriel**

If we can separate single vs double cropping, then use April 1 as the second peak cutoff date instead of March 1.

Be careful of the “sanitary period” in Matopiba.

He will look for price data, and has posted municipal yield data. He can extract municipal climate data if needed.

June 27

**Desk**

Work on Timeseries Analysis v6, which converts the previous point-based peak estimation dates into fully image analysis

**Meeting with Avery**

Maybe use onset as start of window?

Do the peak finding stuff only over cropland; may need to resample mapbiomas to same resolution as Modis.

Switch from analyzing LAI to analyzing EVI

Augment repo for Matopiba dataset on Github

June 28

**Desk**

Work on Timeseries Analysis v7, with user interface functionality and overlays of mapbiomas and lc\_morgen, and ability to export as GEE asset; need to wait for newer maps of these from Jake

June 29

**Desk**

Updated Matopiba survey on Github

Read crop model (statistical) papers