April 1

**Desk**

Google Earth Engine video tutorials, began Google Earth Engine Notes.docx

Imported Mapbiomas to GEE by using asset id provided by mapbiomas

Looked at script for mapbiomas to connect the land use type to the classification number; the <http://mapbiomas.org/pages/methodology> link also seems to lead to useful info about how this dataset was generated

April 2

**Desk**

SCYM paper reading (crop model)

Listed initial GEE tasks to focus future GEE tutorials/learning

April 3

**Desk**

Prospectus outline v2

GEE tutorial and took notes

April 4

**Meeting**

To regionalize Brazil, drive categorization by agriculture instead of all the regression parameters that Dave’s using. Group Brazil by crop varieties; group by soy maturity classes. Soy maturity classes are a good approximator for climate.

For Mapbiomas, Avery’s group classifies different types of farmland. We might want to have a 100km buffer around urban areas and filter out large bodies of water. For non-soybean agriculture, count it as agriculture (i.e. only look at forest cover vs. non forest cover)

For crop model SCYM, if we use the model we wouldn’t be starting with the High Plains regression. Also, why aren’t we just using APSIM directly and ignore the LAI stuff in SCYM? Doing LAI stuff will allow us to bias correct yield calculations, but we can bias correct using actual high resolution yield dataset that Avery has. We need to think about what we want out of crop model. If we get LAI and NDVI out of crop model, why should it be used when we already have good yield data? For now, start with APSIM – can get it myself.

Wind direction: For now, ignore the effect of wind direction on how land use affects climate/yield. This is because wind in Brazil move in all directions so most areas of the neighborhood will be upwind at some point in time, so assume it’s isotropic. It would be interesting to define the neighborhood on different sides of the pixel, running the model, seeing how the neighborhood definition affects the result, and using that result to say something about wind in Brazil.

For prospectus and research questions:

Question 2. Avoid having to say stuff about farmer decision making and agent based models. Look at sensitivity to decision-relevant thresholds. What levels of deforestation have the potential to adversely impact agribusiness and how does that look across Brazil? Answer this by applying this model. Hypothesis: there’s spatial variation in how strong the sensitivity of agribusiness to land use change via climate will be across Brazil. Sensitivity might have different “pathways” – different sectors of agricultural economy might have different sensitivities.

Question 3. Land use change feedbacks to hydrological processes and biological processes. How land use change in our synthetic case also changes hydrological processes (ET, surface water. Precipitation will be the input)

April 8-9

**Desk**

Lit review of AgroServe related papers

April 10

**Desk**

Mapbiomas – the integrated maps downloaded as ee.Image('projects/mapbiomas-workspace/public/collection2\_3/mapbiomas\_collection23\_integration\_v1') seem to be the most useful, but need to figure out what land classes correspond with each “number”

New script in GEE: ExploreMapbiomas

Loaded Brazil outline, created histogram of pixel values in Brazil to look at the land class numbers

Mapbiomas site has Brazil biome shp files, which I downloaded and uploaded to GEE as an asset called BrazilBiomes

Made histogram of a 1000km buffer around practice point using ee.Reducer.frequencyHistogram(). Don’t use ee.Reducer.histogram() because it messes up the land use pixel values (i.e. if the lower pixel value is 4, it moves the histogram to start with 4 at the zero position)

April 11

**Desk**

Avery says: We have a 500 m soymap on bitbucket. I think we ought to overlay it on the the 30 mapbiomas cropland layer and use the intersection for our soy modeling. i’m going to spend a little time reviewing the repo and make sure its intelligible and then invite you to access it. He also gave the Mapbiomas legend

**Meeting**

Crop model:

Crop simulation modeling team in the Marcos Costa lab: meet with them to learn about the crop model.

We need to regionalize Brazil by maturity class (instead of biomes) because each maturity class will have its own set of parameters in the SCYM model

There is GEE code with dataset for regression model (rain parameters as fcn of climate, land use, etc). Dave has it – Dave will give me all data to run regression model.

Need to get a sense of data needed in order to work with SCYM.

We will use SCYM instead of the simpler APSIM because SCYM allows us to un-bias spatial estimates of yield. For example, the yield dataset may only have yields near roads or other places that are easy to access. We can use SCYM to calibrate the model based on yield data, and then use SCYM to estimate yields in other parts. Also, there are many farmer management decisions that we need to model.

Phenology: is plant growth stages.

Talk to Jake before trying to use SCYM.

Work plan:

Talk to Jake and Gabriel about SCYM.

Coordinate Eric and David about inputs and outputs needed for SCYM.

Ask Avery about yield data, soybean maps, maturity class maps.

Prepare for meeting with Eric and David by: reading about SCYM; listing datasets I’ll need and datasets I’ll produce; think about amending SCYM by adding a model calibration using actual yield data (we have this, but original SCYM paper didn’t).

Another question is, we’ll get from Dave the rain parameters; however, will different rain realizations change the LAI vs. yield line in the crop model? Perhaps do this offline.

Categorizing neighborhoods by % forest, % non-forest:

Eric has already characterized all pixel for Brazil.

All pasture is non-forest. Consider leaf cover (perennial? Not perennial?) and rooting depth in order to see whether a particular piece of vegetation is

The “pixel” is literally a pixel of 30m. Aggregation might need to happen for climate (MODIS), but not necessarily for crop model.

Get rid of edge neighborhoods.

Project general background:

The research question about how land use change in synthetic case also changes ET, surface water: to determine this, use a simple hydrologic model. Hypothetical land use change -> precip changes -> hydrological model -> ET and surface water. The link between land use change and climate has never been done before, so it’s interesting and new.

“synthetic case”/”toy model” means: some of the processes we’re modeling is represented in a stylized way. What we stylize is the way we let land use change. We make it something that “turns as a knob” or something that responds to something. Don’t need full agent model. Can use the toy model across a map.

April 12-15

**Desk**

Read crop model/AgroServe related papers

Prep for Dave/Eric meeting

Wait for Dave/Eric and Jake/Gabriel meetings

April 16

**Desk**

Got 500m soymap (lc\_morgen) to overlay on Mapbiomas. Imported them on GEE in filed ExploreLCMorgen.

Sally’s emails:

Regarding calibrating SCYM using yield data:

* I think I would do this (but you should think about this too - because you are the driver of this I think - I don't have an "answer" necessarily) ---
* Crop Sim --> gives a timeseries of phenology and a final yield.
* Bias correct yields based on observations (I think this could be as simple as regressing observed yields on modeled yields)
* Do the regressions with the pseudo RS observations using the bias corrected yields (so that rather than recalibrate the crop model, you basically will throw the bias correction into the regression step)
* Map with RS data

Regarding what I should do now:

* It might be good to start looking at this from the point of view of how you'll use this information for the implementation of SCYM.  So presumably that map will also need to be used to get the RS data for the regression and mapping, to select points at which to do the model runs and compare to observations and so on.
* Before doing any of this in GEE, I'd encourage you to draw up a set of algorithmic steps that get you from start to finish.  I think you'll find that the actual implementation and making of maps and so on is a trivial part of that process, and the best place to put effort right now is in working out what the process consists of!

Regarding crop model:

IBIS is the crop model that is in INLAND.

My understanding right now is that we'll start by finding all the parameters in the regression equation:

1) use RS datasets of actual weather as input to the crop simulation. Run the crop simulation.

2) bias-correct the simulated yield using some regression equation like this: (yield data) = A\*(simulated yield)

3) find corrected yield using: (corrected yield) = A\*(simulated yield)

4) use the corrected yields and their corresponding pseudo RS observations to find the beta values in the SCYM regression equation below, where W is the monthly gridded weather and RM\_d are pseudo RS observations.

On the first piece -- my thinking is that at the moment, we'll use SCYM firstly for mapping yields.

We can do this through to your point (4) and I agree on what you're suggesting to that point.

My suggestion would be that as far as developing a workflow and so on, you start with this - thru step (4).

The outcome here is clearly a timeseries of mapped soy yields across the ag frontier in Brazil for the history of our datasets, and looking just at that and what it tells us about how yields are varying as ag expands and these land use feedbacks to climate kick in should be very interesting - I think for you it should be Chapter 1.

As to the inputs - we have precip & temp.  Check out [http://clima.org.br/en/archives/460](http://clima.org.br/en/archives/460" \t "_blank) for the others.

Can you please do a little reading into this and be prepared to chat with Eric on Thursday about getting the CLIMA data into our Earth Engine / Database situation?

There may well be other datasets we can look at for humidity, I believe MODIS can be used for incoming solar (check), and we can potentially use wind reanalysis data for wind speed, although it'll be a pain.

BUT, as far as the next piece goes - we have *not* made a sensible decision about how to do the crop modeling with the simulated weather.

What we are stuck on is whether it is more sensible (in terms of results and computer time) to go down the line of somehow having a separate phenology model that gives us your "fake LANDSAT" data (Avery's suggestion), or whether it is more sensible to calibrate the crop model to the yield data (basically taking your "A\*simulated yield" bit and trying to get that A into the mode parameters so that the model - simulated yield match is maximized) and then do lots of the long runs you discuss.

Ultimately, I will be tasking you with making a decision about how to do this.  I think you'll want to get a sense of how labor intensive model calibration across all cases is going to be, and secondly whether there are any sensible ways to have a simple weather--> phenological dates model.  Minimally we could prescribe a constant relative phenological progression that doesn't interact with the weather.  Some sensitivity analyses and testing about this is needed.  What you should be thinking about is how we could go about making a decision here - and what our options are.  We don't have a finalized answer yet.

As far as getting modeled radiation, wind and humidity ... Dave does \*not\* have all those on his radar.  I think radiation could come from a standard annual curve that is "reduced" by some factor when its cloudy (i.e. when it rains).

We need to work out the humidity piece.

And the winds we may just use a seasonally averaged wind speed - I don't think they're actually a major issue.

I would take the AWE-Gen paper and read how it does the humidity and temperature and wind and radiation pieces - we may be able to steal a lot of things from it.

April 17

**Desk**

Read IBIS/INLAND papers and documentation. Downloaded INLAND code.

Looked at CLIMA data documentation, downloaded an example file and sample MATLAB code to visualize it.

Advanced stochastic weather generator – read about it

Continue drafting workflow in AgroServe folder

April 18

**Meeting**

Skype call with Gabriel, Avery and Jake: notes are in the AgroServe document “Crop Model Meeting April 18 2018”. Main points:

* We want to develop a crop yield model that’s rooted in a crop simulation model and observations of meteorology and crop phenology. Then we’ll have a robust platform to predict crop yield under simulated climates.
* We want to run INLAND at a few points and compare INLAND simulations to vegetated indices. Basically we’re predicting both yield and phenology with SCYM (so use two predictive models).
* Use LAI to have sense of representativeness of the observational yield dataset.
* As for using either pseudo RM vs actual RM: use pseudo RM for model selection and actual RM to predict yields in regression equation.
* Send Avery an updated diagram of the workflow, which incorporates the gridded dataset.
* We are NOT simulating all soybean pixels. Hopefully we will get enough inference information from the crop model without modeling all pixels. Phenology will help us understand if we’ve sampled the points for the simulation well.
* Remote sensing: decide on resolution, MODIS or Landsat.

April 19

**Meetings with Avery, Dave, Sally**

General

* Think about spatial correlation of environmental metrics
* Ask Dave for his paper repository in Paper

About crop model

* For “actual” INLAND run, we don’t know management so we’ll run for a bunch of management. Hope to rule out some management types based on yield (identify the credible management combinations) – retain the credible management practices for our counterfactual simulations. We might need to vary soil parameters; we might not have layered soil data. May need to go and measure them. Maybe vary different soil categories.
* As to what parameters we’re calibrating vs fixing, that’s a question for Gabriel. Need to ask them what their parameters are, which represent management and therefore solve exogenously.
* If yield data isn’t representative, may need to vary the importance of those calibrations. Maybe use county level info for calibration.
* By using yield data, we are presuming that different management techniques that affect NDVI in the same way will also affect yield in the same way. But we should test this by subsetting by management.
* Work flow is as follows: we will use INLAND (with real weather) to create pseudo phenology timeseries and pseudo yields. We have real meteorology. We will regress doing yield\_sim = fcn(SIMULATED phenology, REAL meteorology) for all potential observation date pairs. For all of these regression equations, we will pick certain regression equations that have high R2 (Lobell’s stuff) AND pick out which regression equations work the best for predicting ACTUAL YIELD.
* We won’t run INLAND with simulated weather. We need simulated phenology too. We need to do another SCYM process where the outcome variable is phenology (phenology = fcn of meteorology and other controls) maybe estimate the inflection pts, curve characteristics of phenology. People have done maize yields in US, and actually this phenology step isn’t a main control on the yields we get.
* We don’t want to run INLAND for simulated weather because of computation time and flexibility. This is why we need regression. To do regression, we need a phenology model. Not running INLAND everywhere means we don’t need to make assumptions about every single place.
* Yield data: Worries about near the road part. But also because they don’t have articulate sampling strategy for where they go. They have 30 major agricultural regions (of soy functional types) – they try to hit each of those regions.

Things I need to do

* I need to come up with a list of what scientific questions I’ll answer – give Avery, Gabriel a draft of my prospectus.
* Must learn GEE stuff from Jake. He will be in Boston until August, maybe wait until he comes to bay area in June.
* Get through gridded, validated historical yields workflow by this summer. (go through SCYM workflow with INLAND results). To do validation, hold out some of the Rally dataset. Do cross-validation.

April 23-24, 2018

**Desk**

Update workflow based on conversations with Avery on April 19

Draft Research Questions for Avery

Literature review

April 25, 2018

**Meeting with Gabriel**

* Management isn’t an explicit input to the model – it has explicit planting date, cycle length. Management will be implicit in the parameters we calibrate in the data. The model doesn’t have fertilization, only nitrogen. If we calibrate with yield data with a bunch of management types, we will essentially be having an average of the management types.
* We can calibrate at least four regions with representative management. Soybean farmers are usually high tech, usually huge farms because it’s expensive to get into the soybean business. Management won’t be a huge issue if we account for regional differences.
* In the same region, there’s some variation, but it’s not as big of a difference as among different regions
* We’re also calibrating actual plant parameters, like temperature sensitivity function. We’re going to set ranges of parameters to actual values. Assumes no till management.
* They have maps of soil parameters up to more than 2m depth. Soil is an input. Soybean roots rarely go over half a meter. Soil database contains clay, silt %.
* Vicosa for a couple days.
* The model works on Macs, try it on laptop – may need to download packages. The hard part is getting the libraries. The installation is the same. It will just be ./configure, make. Just install netcdf library.

**Meeting with Sally**

* For running INLAND with actual data: take subset of additional pixels that aren’t representative of yield data and only run INLAND for those pixels? [ask Avery - What pixels are we actually running INLAND for? A subset?]
* For management, do a hybrid of what Gabriel said about having average management in a region and what Avery is saying about modeling a bunch of management conditions. First use phenology and knowledge of a region to say something about the most feasible management conditions, then model them and look at how phenology responds to each management condition. We don’t want to completely have an average management condition because we want to say something about the uncertainty introduced by the different kinds of management.
* For management: rule down a feasible set of management scenarios for each region (like Gabriel mentioned), and since we don’t know management from LAI, run INLAND for that subset of feasible management.
* The goal of having multiple management is for uncertainty. Ideal: calibrate model for known management at a subset of locations.
* Maybe we can tell soy, soy-corn stuff from RS, then guess planting dates. Assume maturity class is fixed for the region.
* To get gridded, historical yield: don’t need to know actual management; have a weighting of potential management scenarios
* For regression of yield (SCYM), either throw in the yield from weighting management options; OR or do regression separately for each management (does phenology really take care of management for us?)
* Historical gridded yield map will be produced from historical LAI information!
* Phenology SCYM stuff: we need to generate phenology info that’s pertinent to yield through weather. Not sure exactly what we’re going to do. Will be able to decide when set the phenology.
* Won’t go onto simulating weather and phenology SCYM stuff until after we’ve done the gridded, historical yield stuff.
* In terms of picking pixels to model in INLAND, avoid cities b/c there might be pollution, economics. Don’t buffer out major roads. It’s ok to include near coasts.

April 26

**Desk**

Iterated with Gabriel about how to install netcdf, gfortran, netcdf-fortran, and INLAND on Mac

April 27

**Desk**

Gabriel says: for calibrating INLAND, We use a tool we developed here called OPTIS, which employs a generic algorithm for calibration. He shared a book pdf about crop models.

Online NCL course: <http://www.ncl.ucar.edu/get_started.shtml>

Drafted new version and diagram of research questions – send to Sally and then Avery

April 28-29

**Desk**

Read Gabriel’s crop model book and other lit review. Wrote draft of research questions.

April 30

**Desk**

Write 2 page summary of SCYM workflow for Avery