May 1

**Desk**

Updated the “workflow” document to reflect most recent Sally and Avery meetings

May 2

**Desk**

Began notes on INLAND from Skype with Gabriel; started a document in which to take notes on INLAND

Downloaded NCL

**Meeting**

For determining representativeness of yield data, use Landsat (make sure it’s the same spatial resolution as yield data). Then compute GCVI from Landsat. Try to see if we can back out planting date from phenology – ask Avery. Then, divide GCVI by phenology periods, like planting date, and see what GCVI is represented by each phenology section. Also get out of this a measure of variation in planting date. Do all this in GEE.

For sampling additional soy pixels for simulation, look at Morgan’s streamflow paper for parameter matching. Set some target number of points per class, and if there aren’t enough points in each class, redefine the classes OR play around with an uneven number of points in each class.

For management conditions: rotations will be determined by harvest date; we will also have planting date. Maturity class contains a bunch of varieties. So for each variety, we only have planting date and harvest date. Maybe make the presence of rotations a random variable.

Do we really need to knock out management scenarios before we put things into INLAND? Since after accounting for variety, there shouldn’t be too many planting/harvest dates to fiddle with? then we wouldn’t need to compare simulated phenology to phenology data, and simulated yield to yield data, to choose the most likely management practices. For calibration, instead of guessing management, can we first constrain management as well as we can? For example, come up with management groups (e.g. early planting, early harvest) – classify the yield data according to management groups, then calibrate for each management group. So just one calibration, but making sure the calibrated parameters are good across all management groups.

May 3

**Desk**

Map out different ways to handle management

Summarize questions to ask Gabriel and Avery in Brazil

Installed NCL, which requires tcsh to run (which is a shell).

May 4, 6

**Desk**

Map out management

May 7

**Meeting**

Got AgroServe Data list from Sally – look at it and begin actually doing work. Choosing between Landsat and MODIS: which one matches yield resolution the best? Make sure can use both Landsat and MODIS.

Focus on SCYM for the past right now, NOT SCYM for the future. Work on learning the model and SCYM for the past. Carve out a near-term to do list and get started on it, because we will think about future research questions later when we know what’s feasible.

May 11 to 15

**Meetings**

See May 22 meeting powerpoint for a summary, and also Meetings May 14 and 15 Gabriel for a description of how INLAND is calibrated

May 21

**Desk**

Finished summarizing Brazil meetings in May 22 2018 powerpoint

Begin planning for sending a LAI timeseries example set for soy pixels

May 22

**Desk**

Iterated with Gabriel about calibration summary powerpoint

Began code Landsat Phenology in GEE. Basic setup of, given a year and pixel, masking out cloudy days, calculating LAI (fake equation though) and plotting a timeseries.

May 23

**Desk**

In Landsat Phenology in GEE, added in Google Fusion Table for rally points – but still need to know how to read them in.

**Meeting**

For me

Dave doesn’t interpolate gridded weather. We shouldn’t interpolate either – 500m is smaller than correlation length scale of rainfall. For temperature, work at native resolution because it changes quickly across space. Work out the scale we’re working at before deciding if interpolation is necessary.

For calculating vegetation indices, and making sure clouds aren’t in it, look at actual images to see if simpleCloudScore is good enough. Also note cloud shadows.

Try MODIS because it might be better than Landsat due to temporal resolution – esp with clouds

Ask Avery or Gabriel:

Are there other sources of spatial variation within regions besides management? Like significant variation in weather?

For the single vs double cropping data: ask Avery if this data represents the farmers’ intention vs what they actually did. Ask if the single vs double crop maps

Is there any sort of ground truth planting date for validation?

Calibration

Planting date and harvest date should be set exogeneously. It’s weird to have a calibration depend on planting, cycle length, and then to make management package calibration bend over backwards to compensate for the planting, cycle length errors. We should set planting and cycle length exogenously, but if we don’t know planting and cycle length, we should do different scenarios.

Sources of uncertainty in the model: planting date and harvest date (constrain with phonological information, physical bounds); soil and weather and model structural error (won’t do much about that); fertilization, variety, other farmer decisions and physiology parameters that aren’t affected by farmer decisions (these are our biggest unknowns – but instead of treating them separately, calibrate them together. They are all equally unknown. Therefore this would do away with needing two levels of calibration).

She suggests:

First, take a block of years (and probably a region?)

Second, for each year, use phenology to assign approx. planting and harvest date to each pixel-year

Third, classify dates into 9 timing groups: early, middle, late planting; short, med, long cycle length. Each pixel-year will be classified into one of the 9 timing groups, and run INLAND with the assigned “average” timing.

Fourth, set a common physiological parameter combo set for the year group and region. maybe n = 10000 parameter sets. “Common” across all 9 timings.

Fifth, run INLAND for one of the 9 timing groups (i.e. planting date and harvest date are exogeneous) and all n physiological parameters. For each parameter combo, get performance statistics like RMSE for each param combo and timing group.

Sixth, average over all timing groups per param combo and identify the top 100 parameter combinations. (average the RMSE for each timing group or average the actual parameters and run INLAND again?

Seventh, run the ensemble of 100 parameter combinations to get an average yield and confidence interval. This will give the historical yield map (so not SCYM?)

Issues:

* Runtime for INLAND? Can we replace running n = 10000 physiological parameter sets with INLAND calibration set – i.e. the members of the genetic algorithm, and retain the best set? Can we really assign approx. planting and cycle length from phenology? Mixed pixel effect? Can we replace cycle length with [harvest date – planting date], since the different parameter sets would technically capture the cycle length, varieties, etc.
* We need to average across timing groups because we don’t want the physiological parameter values to be contingent upon the timing. However, why do we need to average parameters across timing? If we get phenology map, then it’s okay to calibrate a separate set of params for each timing group and not combine them again. Also, doesn’t it make sense that different timing groups will correspond to different varieties with different physiological parameters? Also it sounds like we’re using INLAND to generate the historical yield map, instead of SCYM – but didn’t we agree to use SCYM due to computation time?

May 25

**Meeting with Gabriel**

We should try to estimate planting date like Sally says but there will still be some uncertainty. We should definitely do a smaller range than onset to greenup.

We’re going to account for different planting dates in SCYM by running diff planting dates and the same crop parameters.

When we set different planting dates for a pixel, we’re just propagating uncertainty; we’re not assuming a pixel has a specific planting date. It’s still weird to assume a very large range of planting dates.

We’re all trying to account for uncertainty in timing. Gabriel likes the idea of better constraining planting date, but thinks we should still have uncertain input there. Sally wants to have all uncertainty in timing to come out in the calibrated parameters.

Even using phenology will have very high RMSE in planting dates. So may have Monte Carlo treatment of timing.

Spotpy – ways to deal with uncertainty.

Gabriel thinks that having a common set of parameters means that we’re going to get the average management, not the extremes in management. It would only get an estimate of uncertainty. Sally’s going at our Level 1, to get an averaged management. We can talk about getting extremes in management later.

Gabriel wants one set of ensembles for each planting date assumption and each cycle length assumptions.

Look at WDRVI in rally points for an idea of phenology.

**Desk**

Need to change rally points csv from “x and y” label to “lat and long” label for it to work in GEE

May 30

**Desk**

Read papers about how to get LAI for soybeans from satellite data, with focus on something to apply to getting LAI in GEE

**Meeting**

Last week’s meeting with Sally was focused on calibrating a model in general. If we’re only calibrating for SCYM, then it’s more of a curve fitting exercise and therefore what we’ve proposed with Avery, Gabriel in Brazil is okay – just try and see if we can constrain the planting and harvest dates more by looking at phenology.

By next week, do a lit review about finding planting date and harvest from phenology, look at how feasible it is. Try it in GEE.

Put in GEE a way to get a sense of the spread in planting, harvest dates and how different pixels look from each other.

May 31

**Desk**

Read papers about how to get planting and harvest date out of satellite data