Des Moines Lobe Milkweed Augmentation Simulations

Working folder is C:\Users\tgrant\Documents\Monarch Butterflies\Projects\Des Moines Lobe MW Augmentation.

**12/8/20 – 12/11/20**

First need to find GIS files for the state.

Shapefiles appear to be in Box under ArcGIS and Techs root folder. Altogether the Final Shapefiles folder is 5.5 Gb. Wonder how it will run bringing in shapefiles from Box or if I will have to move them to harddrive. Hard-drive has 35 Gb space.

There’s also a folder under Monarch Butterflies on the hard drive called Techs and ArcMap. Not much there though.

The MonarchsTech.mxd file in Box has a few layers but not the big regional shapefiles.

I’m going to start a new .mxd to start fresh.

Started blank and added county.shp from MonarchsGIS/shapefiles. Which put dataframe coordinate system into NAD1983 UTM Zone 15.

Found a shapefile called ecoregions.shp that has Des Moines Lobe. Need to figure out what we define Des Moines Lobe as. The Griffiths et al. 1994 paper looks to be the basis for the ecoregions.shp shapefile, which I downloaded from the DNR website before it went obsolete. Can’t find ecoregions.shp on new website, geodata.iowa.gov, which is frustrating because may need to cite it. In the grant proposal we cited USDA 2006, a nation-wide document about major land regions. I downloaded it and saved it in the lit folder. This was turned into MLRA (Major Land Resource Area) shapefiles I have seen somewhere. In the pdf, Des Moines Lobe is number 103 on p 322.

There’s a shapefile I added called mlra\_a\_ia.shap that seems to just be cut from the national MLRA. It’s generally the same as ecoregions.shp, but with some minor differences.

Emailed Steve and he agreed to just do Des Moines Lobe. I’ll use ecoregions.shp because it has smooth edges. Should be able to cite DNR or just Griffith et al. 1994.

Added tech shapefiles which had to be transformed from GCS\_WGS\_1984.

Five tech shapefiles occur in DM Lobe. First clip each one to DM Lobe so have smaller shapefiles to work with.

There was overlap in some of the tech shapefiles. Well, just Cory’s. So clipped those to the counties so they won’t overlap. After some angst about how I would remove overlapping polygons.

After merging all the shapefiles into one shapefile called **DMLModel.shp**, the shapefile something like 800 Mb. It has 871,502 polygons. Recalc’d area in m^2 in column Aream2. Need to run some elimination. Looking at everything selected with < 1000 m2, most is along edges. I ran eliminate on everything under 1000 m2 first. Then did 1300 m2. Then recalc’d the area. I had the techs eliminate everything up to 2000m2. So eliminated everything below 2000m2 and called it good. The subsequent files are called DMLModel2, DMLModel3, etc.

I think it’s ready to prepare to run on the high performance computers. Well no, I need to change to lat/long projection and get all the probEggs right.

Apparently the most recent notes on running the model on ISU HPC machines is in Spatial Toxicological Impact Analysis - Working Notes.docx. I should make a separate document. Some other things I need to do based on reading those notes on p33.

*Thurs – 12/10/20*

Sent email to steve about the milkweed scenarios and foliar insecticide scenarios.

Cleaning up the shapefile. Deleting extra fields. Deleted Aream2 and calc’d area in Shape\_Area, which is needed for RS. So DMLModel4.shp has no extraneous fields and all the fields needed for RS. But ProbEggs and probMove need to be populated. Need to find .cal files. The .cal files for the toxsims paper are called ToxSimsScenXB.cal, where X = 1, 2, 3, or 4 and are in the Monarch GIS folder. Just using the ToxSimsScen1B.cal code gives a bunch of 999’s so need to change the code. Changed code and saved the cal file as DML1.cal in Monarch GIS folder.

Phone call with Steve – we can’t think of any reason to change the milkweed augmentation scenarios from the last paper, so we’re going to run those. Steve asked about CRP, but decided better not to try to add random CRP. Steve found a typo in the submitted files for the toxsims paper where probEggs was 0.65 instead of 0.065. I changed it in both versions of the supp files on my hard drive and in Cybox. For foliar insecticide effects, we agreed not to add the spray drift zone for now. His idea was to potentially do one county in the north area to get an idea of how it is up there. But basically just talking about back of the envelope calcs on potential effects in the review paper. At this point I think we are both leaning toward folding this analysis into the review paper. He also wanted me to look at Kelsey’s final presentation to see her results on how to reconcile it with modeling.

Deleted DMLModel2.shp and DMLModel3.shp to save disk space.

*Fri – 12/11/20*

Yesterday found some overlapping polygons. Not sure how they got there. Found online how can use union tool on a single shapefile to find overlapping areas. <https://gis.stackexchange.com/questions/75418/erasing-overlapping-polygons-from-same-layer-using-modelbuilder>

Didn’t seem to do what I wanted. Deleted the shapefile it made. Trying intersect tool with DMLModel4.shp as the only input. That worked and produced a shapefile just of overlapping areas. There are 6047 overlapping polygons, which range from 2400 m2 to about 0 m2. It’s less than 1% of polygons so I’m going to try first to run the model without fixing it. Because I’m not sure how to fix it. So moving forward, ran repair geometry on DMLModel4.shp. Took 6 hours to run repair geometry.

**12/14/20 – 12/18/20**

*Mon*

Steve postponed phone call until tomorrow.

Noticed find identical and delete identical tools. I remember I tried using these before. I should look back in notes and see if they worked. I found in the old Shapefile Protocol v6 that I used with the techs, I used Find Identical to find overlapping polygons, but then they were just deleted by hand. Couldn’t find any other use of this term. So again, just gonna try model as it is. I don’t think the overlaps should be a significant issue.

Wanted to run integrate tool next. Not sure it does what I think it does. Entered one dataset, DMLModel4.shp, and left XY tolerance blank (default). Tried to find XY tolerance of DMLModel4.shp, but couldn’t find it. Should be in environment settings, wherever those are. Doesn’t show it in Geoprocessing Results window either.

Integrate took about 2 hrs to run. Then recalc’d Shape\_Area. Then moved files to C:\Users\tgrant\eclipse-workspace\Monarchs\data and changed to lat/long in Analysis Shapefiles.mxd. Then tried to load Eclipse and discovered Java is not installed on this computer. So got IT to install it.

*Tues*

Eclipse works. Next problem is that I need code to get random initial start in Des Moines Lobe, which is irregular shaped. I started to try some code. I saved a backup of ContextBuild.java in C:\Users\tgrant\Documents\Monarch Butterflies\Projects\Des Moines Lobe MW Augmentation.

Gonna load up TeresaTestData2.shp to work in and test code because it’s really small. It runs, but code needs some thought.

*Wed*

After mulling the code, I think I was making it way too complicated. Problem is I’m not sure if JTS geometry function does what I want.

Called Dave (Grant, my brother) and after about an hour sharing screen we got it to work.

*Thurs*

Added counter to see how many random locs are discarded.

RS throwing errors about accessing stuff on the internet. Could update to 2.8, but I don’t think it will be an issue when not using the front end GUI.

Need to change bounce back code to make sure it lands in shapefile. Should be able to use same code block.

Ready to test DML shapefile.

Lat/longs of corners:

NW: -95.8171, 43.5047

SW: -95.7295, 41.5258

SE: -93.0510, 41.5582

NE: -93.0527, 43.5394

Runs out of memory in RS, so next ready to try it on HPC machines. When making .jar, for the record, I used the parameters from the spatial toxicology paper, which are in the spreadsheet I copied to this folder. Max probEggs varies though. For first run, used number of agents already in RS, which seems to be 200x50, or 200x?.

Just ran with everything at defaults that were already there. See how it goes.

Job 5917. It ran with no errors. And 10,000 println’s for the number of discarded points. I copied the discarded points from Notepad++ to Excel in Model Runs and Parameters.xlsx. 63% of points are discarded.

*Fri*

Using normal approx. to binomial distribution (could use binom package in R but sample size is very large) found that the discarded points correspond to a different area in RS. It’s only off by 3%, but the reason must be because I used the max outer lat’s and long’s to make square in RS. In ArcMap, when drawing a square, lats and longs are not constant. Anyway, it’s a minor thing, but I was curious. The model code appears to be working perfectly.

Also calculated how many monarchs I need. At the rate of 10,000 per Story county, it comes out to 212,961 per the Des Moines Lobe. I’ll use 200,000.

Trying to run model with 100 instances and 500 monarchs per. Changed SBATCH to request 100 tasks and 3 hrs.

Can’t run 100 tasks though per 1 node. Tried requesting 2 nodes and at least it submitted job 6062.

Forgot to delete discarded points prntln. Now commented out.

Also getting some out of memory errors. Not clear if they are resolved or failed. So far it seems 5 lines failed, but the others are running? Not sure. 1 hr in. Probably have to go to 400x100 and run it 5x. Seemed to run fine on most instances, except 5. Not sure if 5 instances of 50 or 5 lines of 100. Used outputcombiner.sh. Gonna run 400x100 then quit for the week. Some of the instances failed again though.

Left it running, so need to run output combiner on Mon.

*Mon need to figure out how to get all instances to run, why getting the error, or if I can just combine what instances do run. Also free up disk space.*

Deleted a bunch of GIS imagery used to develop GIS model for Iowa, so I hope I don’t need it again. Deleted NAIP and hillshade and CDL imagery. Goodbye. Freed something like 20 Gb.

**12/21/20 – 12/??**

*Mon*

Trying more memory to get model to run without errors. Well, can’t add much more memory that 500G. So I ran 500x100 again and it ran without errors this time. Weird. I was going to show the errors to the researchIT guys and ask them what to do.

I also downloaded the new SLURM files from the Repast-L listserv. I’m not using them, but can if I have to.

If 500x100 continues to work well, I just need to run that 4x.

*Tues*

Got 4 runs of 50,000 each, now need to pull them into R. Where did I pull R code from for DMLsims.R? Somewhere in Tox Sims project I guess, but didn’t make notes about where I got it.

**12/29/20-12/31/20**

*Tuesday*

Stopped R running today. It ran for 6.66 days (!) and only finished half of the dataframe. It got to i=457,006, so it finished to row 457,005 out of 868,618, so 52.6% complete.

So need to figure out how to run on the HPC machines or it’ll take forever.

Got email from researchIT. Need to run R script as a batch run, instead of interactively. I don’t know why I didn’t think of that. Did I do it that way before? I don’t think so, I think I ran it interactively. So today figured out how to write a sbatch script file for R and it seems to be working. I’m also running it on speedy2 to get it done as fast as possible. I estimated 8 days, but who knows how long it will take. I think I can check the output file to see progress because I am doing the print(i) thing.

But then couldn’t get large files to copy to pronto, so emailed IT.

IT had no problem copying a large file to my folder, so it must be a problem on my end. I restarted computer and VPN and it worked fine. But then if I come back later it won’t (because computer went to sleep?? Who knows) work, but worked when restarted computer. Grr.

Anyway, all 4 files are running now.

Deleted some test output files that are not part of the final four to free space. Left the other files, just deleted the largest output file.

**1/5/2021**

R batches have all finished running. I moved .Rdata files to hard drive. I may want to move some other files in the future, like the scripts and .out files. Next also need to run Scen 2 and 4. But now need to work on reviewer comments for tox paper.

**2/18/2021**

I ran the 4 runs on the ISU HPC machines. Now I have 5 large dataframes to combine (2 from map 1, 1 each from the other 3).

I moved a bunch of output files from my LSS network drive to the ToxSims folder.

I noticed the output files in my research drive are called repast.out for R runs, because in the R sbatch file I forgot to change the filename.

Finally did get them all combined into one dataframe for Scenario 1.

**2/24/2021-2/27/2021**

Back from working on accepted ms. Steve was interested in seeing differences between counties, so trying to add county back in to the dataframe.

So after some looking and lots of ideas, I found that even the original technician shapefiles don’t have the county in them. Which is why I never added it in, in the first place. So now looking for some what to spatially query each polygon and add the county it is in to a field.

For scenario 1 dataframe that I already made, the other problem is matching the counties to the right rows in the dataframe. The PolygonID is not exactly the same as the FID from the analysis shapefile (DMLModel4LL.shp). I added notes to the workflow for running RS to strongly consider adding your own polygonID, because I have been confused before about the PolygonID field that comes out of RS. I checked if polygon\_area is unique and it usually is but not always. The order may be the same or close.

Looks like spatial join is what I need to try.

Spatial join doesn’t work. It does some calculations about area of the join, but doesn’t have a field for the actual county.

Going to try the Attribute Asst. Never mind, seems to need some sort of account, so would probably have to go through IT.

But looking closer, discovered Spatial Join under Analysis | Overlay in ArcToolbox does what I need. Not sure why spatial join in the right click menu for layers is called the same thing. Using this tool, I tried “one to one” and the “WITHIN” parameter, so should join features than are within counties. For polygons that overlapped multiple counties (even just ever so slightly), the county field was left empty, so tried just the default “INTERSECT” option. Doesn’t take long to run, unlike the other spatial join, which took hours. Got wrong county for some polys. Found the “HAVE\_THEIR\_CENTER\_IN” option, tried that. So the final one is Spatial\_Join\_DML4\_Counties3.shp.

I imported that attr table into R and it seems to match the Denresults.1.B dataframe fine. I subtracted the areas in both to check if they were different and the result is all 0 except for 201 entries that are very small (10^-13) so probably numerical. So I guess the assumption that they are in the same order is true.

Next just need to calculate summary stats for landcover types and counties, etc.

Seem to have the eggs laid calculated by landcover type and county.

Next need to calculate area of each landcover type in each county.

So I have learned that Brooke apparently didn’t quite do her counties right. They are missing MWROW and nonGMO corn and beans. Can probably gloss over the problem because it was only Story, Polk, Jasper, and Marshall. Jasper and Marshall are very small in the DML and Polk wouldn’t have much of those things anyway.

Wetlands and water are the same in the .csv, so probably a mistake in the dataframe. Fix HabArea dataframe on Monday.

**3/1/2020-3/5/2020**

Talked to Steve about the errors in the shapefile. We decided to fix errors in shapefiles before running Scen 2 and Scen 3, then run Scen 1 again while writing the paper.

Turns out MWROW0 is there, it’s just labeled MWROW\_0 and I failed to capture that in my R code. But it does appear that the Non-GMO corn and beans are not there. So I used the old protocol (Shapefile Protocol v6.docx from C:\Users\tgrant\Documents\Monarch Butterflies\Techs and ArcMap) to add the non gmo corn and beans to the current shapefile. Made a copy (Spatial\_Join\_DML4\_Counties3Copy.shp) in case something goes wrong. Subset out corn and beans and renamed the CLASS\_NAME. It recalc’d area, which is wrong because it was in lat/long proj. Erased the nonGMO fields out of the main shapefile to make Spatial\_Join\_DML4\_Counties4.shp. Merged the file with erased fields with the nongmo fields to make Spatial\_Join\_DML4\_Counties5.shp. Used DML1.cal to set probEggs and probMove. Sorting indicates they are accurate. To recalc area, need to project again and calculate shape\_area. So, did that. This time made a field called GISPolyID with unique ID. The question is how to get it into RS. To calculate, when you open the Field Calculator, there is a link at the bottom called About Calculating Fields. Near the bottom of that help file, is a section called “Accumulative and sequential calculations” that gives some Python code for sequential numbers. Didn’t seem to work though. Downloaded the attr table and put into R can found it didn’t go above 99999, I think because precision was 5. So deleted field and made new one with precision 8. So it did work this time, I checked it in R.

So now shapefile is reconstituted, so back to repair geometry and integrate, just to be safe. Repair geometry taking a lot of time to run. Looked back in notes and took 6 hours last time. It is finding things to fix though. Finished after 5 hours. Started Integrate, which should take ~2 hrs. Ran again in the morning, because seemed to get stuck when VPN connection lost overnight. Took 5 hours to run, when I finally got it to work. Restarted comp, ArcMap, and VPN. Recalc’d area – now it’s ready for Scen 1 run again.

Fixed HabArea in R, with new numbers for MWROW0. Will have to wait until Scen 2 runs to get nonGMO numbers. Copied new numbers into EggsLaid and HabArea.xlsx, but need to check numbers. Wetlands in HabArea.csv was also wrong, so fixed that. EggsLaid and HabArea.xlsx should be correct now, until I run Scen 1 again.

The probEggs file for Scenario 2 from the last paper was *ToxSimsScen2B.cal*. So using that for the next shapefile.

Maybe found error in Developed Open Space and Developed Low Intensity – was 0.005 in DML1.cal sometimes, but in Table S15 it was 0.003. But I don’t think any fields actually get assigned 0.005. But there is 0.005 in the RS output or in the shapefile, because the categories in the .cal file weren’t in the shapefile.

Got Scen 2 shapefile into RS. Trying to get RS to load GISPolyID and put it in export file. I think I got GISPolyID into the RS output file, maybe.

Ran on HPC machines and got an error about long and double. I assume related to GISPolyID.

**3/8/2021-3/12/2021**

Got Dave to help me with code, which seems to be working now. It’s running on HPC machines. Only 14 of 100 ran, the rest ran out of memory. Trying 750 G instead of 500 G in the SBTACH file (--mem=750G). Trying 900G, 750 didn’t work. But later discovered mem is hard coded at bottom of the sbatch file.

Tues going to try 750 G hard-coded and write new sbatch file while it’s running. Couldn’t figure out new sbatch file so gave up and emailed IT. They reminded me there is a memory setting in slurm\_repastwrapper.sh for each Java instance. So I set those to 64g and the overall job mem to 700G and it seems to be working. Edited the workflow file to make this clearer, or at least as clear as mud.

It seems to have run, but .err file shows some errors. Can’t find anything wrong with output files though. After doing outputcombiner, renamed combined data to Scen2 – I’ll put all Scen2 data in there for awhile.

Need to figure out how to use the LSS to store my output files. They keep saying /work is not backed up. Have to request space though. I don’t think we have requested space. Decided to delete cumeggsperzone.txt files for Scen 1, since I’ll be running them again. They were too big to recycle, so they are gone from my hard drive. They are still on /work for now. Copied Scen 2 files from first run to hard drive. Can’t remember if I need them on hard drive before I combine them in R. Really need to look at hard drive again and find some stuff to delete. But really, these files are easy to reproduce in a few hours.

Wed morning started 3rd run of Scen 2. Seemed to be some errors. Output file only half normal size, so gonna run again. Ran again, but again some didn’t run with out of mem error. Not sure why. According to output combiner, instances 17, 36, and 41 had nothing in them. But monarch output file shows really only a few instances fully ran. I think I just have to take what I can get and combine until I get the 200,000 monarchs I need.

Thurs morning – finished a bunch of maps. Trying 1T memory in sbatch. Then need to add up how many monarchs I have so far. 1T memory actually seemed to work, no errors. Also ran in only 120 mins, so everything is gud. It ran so smoothly, I ran it again to get all instances in one file. But last time didn’t run that good either. Maybe I underestimated the overall mem, because 64g\*100 is 6,400 not 640.

To start counting how many instances and monarch agents I have, I copied monarch.txt files to hard drive, but not the big cumeggsperzone.txt files. I checked how many unique run numbers. I have over 400 instances run, so just need to combine them properly now. See notes in spreadsheet and DMLSims.R.

So I’m getting code ready to combine instances and found 2 things that happened in GIS with the GISPolyID. After I processed the files, I lost a single shapefile somewhere, so now I have 868,617. The other thing is that I made GISPolyID before I lost the single shapefile, so GISPolyID matches PolygonID sometimes, but sometimes its 1 off. I’m sorting/filtering off PolygonID, so actually it doesn’t cause any problems in the results. So at least the GISPolyID is another check and I’m figuring out whats going on.

Only 96, instead of 97, instances seem to be showing up in Scen2.1.

Added more code to look at monarchs.txt. Added some histograms to see if there were as many steps in each run. It does show only 96 finished in Run 1 and 97 in Run 2. It looks like that aborted instance in Run 1 was the only one like that. For the rest, monarchs did the full number of steps in each instance.

Now have to figure out how to leave out one instance in Run 7 to get an even 400 instances.

Friday morning – subset Run 7 to leave out one instance, running now. So I have jobs running on the HPC machines, longest should take about 3.5 days, so done Monday or so. Then can move on to Scen 3.

Removed all large dataframes from the RStudio workspace.

**3/15/21 – 3/19/21**

Worked on survival analysis for NFWF and MJV first.

In phone call with Steve, he liked Figs 1, 2, and 4. So deleted 3 and now 4 is 3. I’ll add scale bars and another panel for 3.

Running Scen 3 in RS. Gonna try 16 Gb per instance and 1.8 T memory. 1.8T didn’t work, so tried 8g and 1T. 16g and 1T worked. Can’t seem to request more than 1T. Seems to be going good.

Got all of Scen 2 compiled and exported to Excel. Need to figure out what figures and tables to make for paper.

Next time I ran model on HPC machines, some instances didn’t run using 16g and 1T. Tried 8g and 1T, which was pending for over a day, but only about a quarter of instances ran. By Saturday, I have all of Scenario 3 agent-based model runs done. Next need to combine with R jobs.

Survival estimation model – after meeting with NFWF and MJV on Thurs 3/18/2021, decided to try a model with longer development time based on A. syriaca development. Looking at Niranjana’s data, I decided to use 15 days for 70F/21C, and calculated degree days based on that. The new model is in C:\Users\tgrant\Documents\Monarch Butterflies\Projects\Immature Survival Estimation\NFWF and MJV project and the bug file is called MonarchModelSyriaca.bug. I compared it to the site 73 analysis I did for MJV and the survival estimate was lower (~9% vs 13% - it’s a bad dataset). So gonna run it and compare to some of the results from the paper. Maybe then upload it to GitHub.

**3/22/2021 – 3/26/2021**

Monday talked to Steve and did interview with Ann Robinson for landscape toxicology paper. Got agent-based model results compiling in R jobs – should take 4-5 days if they don’t get paused. Talking to Niranjana Tuesday about changing survival model. I ran 6 data sets but want to firm up syriaca model before finishing.

Worked most of the week on survival estimation. Made a separate working notes file in the NFWF and MJV folder.

On Friday, R jobs compiling finished on the HPC machines.

**3/29/2021 – 4/2/2021**

Monday finished running survival estimation for each site. Next need to do meta-analysis and look at stage estimates. Drafted email to NFWF and MJV.

Tuesday switched back to DML. Got code to combine 4 R jobs, will run at end of day because I think it will take most of the day. So starting to run Scenario 1 again.

Got data for Scen 3 compiled and put into spreadsheet with Scen 2. Haven’t made any graphs, etc., yet, waiting on Scen 1 to finish. Submitted job, but it’s held up by priority.

Couldn’t get any to run this week, ran smaller and smaller jobs and couldn’t get enough mem.

**4/5/2021 – 4/9/2021**

Rescheduled meeting with Steve to Wed morning.

Just trying to get any job through the HPC machines.

I have 8g for each instance, which should be 1G for each instance and 20G for 20 instances. But trying 170G for 20 instances in case something wrong. Can’t get even 20G jobs to run. So emailed researchIT.

Until hear back from them, read Kelsey’s stuff I guess and work on stage survival.

Email from researchIT reminded me I forgot to specify memory in the second place in the sbatch file. So I added a sentence in caps in workflow document to remind me to do that.

Got a job to run with 25,000 monarchs. But it got killed because it ran over time limit, which never happened before, so maybe they changed that. No CumEggsPerZone files got written to, so have to delete entire run.

Tried 1T RAM, got stuck in pending based on priority. Dropped to 500G RAM. Still pending priority.

Got it to run consistently with 500G RAM. Got 6 of 8 runs done now. Been getting the code to calculate mean lifetime eggs laid per monarch. Could be interesting.

Still need to make graphs/figures/writeup for stage survival.

Got all repast runs done for Scenario 1, so I’m all done with RS unless we think of another scenario to run.

Calculated average eggs laid for each scenario. It’s a little interesting, the point estimates and histograms.

On Monday, need to get R jobs running to combine Scenario 1 simulations. Have to run 8 jobs instead of 4. Need to work on reporting stage survival. Right now, writeup is in separate file, so need to add to DML/AFRI review paper writeup. Possibly graph density of the three eggs laid histograms, to see how they compare. Maybe start working on graphs etc for Scen2 and Scen3.

Also send email to NFWF and MJV about monarch survival model.

**4/12/2021 – 4/16/2021**

Submitted 4 R jobs to HPC machines. All are pending. Queue seems more full than ever this morning. Might want to think of other ways to get the jobs done. Or if the code can run faster somehow if using different commands instead of dplyr.

Also thinking of sending new survival model to coauthors Flockart et al. Also thinking of going through AFRI grant after meeting and looking into fragstats or anything else I have missed.

It’s now apparent that the total eggs laid by monarchs and total eggs laid in the polygons differs by a few tenths of a percent. Still thinking about why. For the paper, I’m going to use the numbers from the eggs per zone.

All 8 R jobs are now running. I emailed IT to ask why I can’t use sstat to see how much RAM my model is using. It looks like they are running at a rate that they will be done in 3.5 days.

So now turning my attention to the methods and results writeup, where I will add some from the DML sims and then work on the survival analysis.

Stage survival estimates are a problem. Looking at CI’s for estimates in the accepted paper, from C:\Users\tgrant\Documents\Monarch Butterflies\Projects\Projection+Tox Model\Post-Poster Analysis\Natural and Insecticide Survival Analysis\StageSurvival.RData. CI’s are wide, often hitting 1.00 for later stages, but lower CI is not so bad.

Graphed the stage-specific estimates in Excel and tried to make sense of them. Wrote it up a bit in the DML writeup. Not sure what is going on, but appears to be identifiability problem especially with egg, 1st, and 2nd instars. From this point on, needs new simulations, also could look back at old simulations, where I never looked at stage-specific estimates and compared to the numbers they were simulated under.

Going to drop this for now, maybe until get to talk to Steve, and look at the fragstats issue.

Thurs – thinking about fragstats stuff today. Found old folder on Spatial Analysis with some papers and R package that does everything from Fragstats. Thoughts from today

* Most of these programs use rasters. If I convert my shapefiles back to rasters, they have been highly processed so how would that go?
* The landscape is so homogeneous the most interesting thing I can think of is looking at grassland and pasture landscape issues. Is there any patterns? Also, the thing about smaller patches have higher egg density I think is an edge effect. Monarchs get hung up on edges cuz they don’t want to leave. Directional flight inside rapidly leads to edges and so they spend more time on edges. Smaller plots have greater edge to area ratio, so more eggs per ha. This is the low hanging fruit I could look at first. Aggregation and patterns of grassland could be interesting, but few aggregations in DML.
* Drew a figure of egg density by percent cover grassland, with different lines/series for different window/radius sizes.

Friday – In DMLSims – Spatial Analysis.mxd I took the shapefile for Scenario 3 (it doesn’t matter which scenario because they are the same except for probEggs) and made a new field CLASS2 which combines the other CLASS\_NAME categories into 17 landcover categories. Then I converted the shapefile (…UTMS3.shp) to raster, with 30 cell size. Seemed to work fine, took less than 30 mins to make. Overlap with shapefile seems good. No roadside polygons got left out, as far as I can tell.

**4/19/21 – 4/23/21**

Combine R jobs

Get egg density into shapefiles

Get QGIS and LeCos

Compare to no pesticide scenarios in Grant 2021

Discovered that one of the R codes for the HPC machines was wrong. Code for both wrote output as “DMLDenresults\_1B.4.RData”. So need to run one again. Both used the run 4 RS output file, so no matter which one finished first, the file now called “DMLDenresults\_1B.4.RData” is actually the output for run 4. So just need to do run 3.

Couldn’t get raster to load in QGIS. Seems .ovr file is not really a raster. Went back to ArcMap and exported raster as DML.tif.

On the issue of “attribute tables” for rasters in QGIS: <https://gis.stackexchange.com/questions/259389/how-to-read-a-raster-value-table-vat-in-qgis?noredirect=1&lq=1>

Basically that functionality is not there yet. So I will need to match up the 17 kinds of landcover pixels in QGIS (which are simply numbered) with their text labels.

Need to work at a smaller scale, maybe smaller than county scale to make sure it’s doing what I want.

Run 1.3 finished, combining 8 runs for Scenario 1. So all done with HPCs for now.

Total eggs laid in the individual monarch output files doesn’t exactly match the total eggs in the eggs per zone output file. So that’s weird. I’m using the eggs per zone for reporting. Difference is well under 1%. I assume it has something to do with the RS code.

Sent Steve DML sim results. Emailed Tyler Flockhart prelim results on new DD model.

**4/26/2021 – 4/30/2021**

Raster approaches don’t seem right for what I want. Going back to polygon approach, trying to calculate some of these indices for polygons in ArcMap.

For spatial analysis of Scenario 1, I joined Denresults.1.csv to Spatial\_Join\_DML4\_Counties5\_UTM.shp, then exported only Grass/Pasture (there are 2 CLASS\_NAMES for this) to a new shapefile called Grass-P.Scen1.shp. EggDensity transferred over. Cleaned it up by deleted some fields.

Trying some analysis tools in ArcToolbox, in the Spatial Statistics Tools toolbox. Average nearest neighbor output report: <file:///C:/Users/tgrant/Documents/ArcGIS/NearestNeighbor_Result_6732_11196_.html>. Shows significant clustering because nearest neighbors closer than expected. High-Low Clustering Report shows significant “Low-Clusters”, whatever that is: <file:///C:/Users/tgrant/Documents/ArcGIS/GeneralG_Result_6732_11196_.html>. Ripley’s K function shows always a higher observed than expected K, not sure what that means. Spatial Auto-correlation (Moran’s I) shows highly clustered again: <file:///C:/Users/tgrant/Documents/ArcGIS/MoransI_Result_6732_11196_.html>. Didn’t do incremental spatial auto-correlation, cuz not sure what to use for inputs. Exploratory Regression was interesting but I can do better in R.

In working on a regression between egg density and area, perimeter, and fractal index (mostly abandoned shape index), looked at overplotting issues, linear regression, polynomial regression (no use), LOESS smoothing (works well, but is an exploratory not confirmatory analysis because no underlying model per se), and GAM regression which worked very well.

But I can’t explain very well what is going on. Except patch area and shape don’t predict egg density very well.

Exported the top 100 and bottom 100 egg density grassland/pasture polygons from Scen 1 shapefile (which I joined the egg density .csv to) to new shapefiles so I could look at them more carefully.

**5/3/2021 – 5/7/2021**

Talked with Steve, he wants to prioritize polishing what I’ve done already. Also wants me to meet with Peter Wolter. Later can look at top 100 and bottom 100 and I would like to run model again and get coordinates then look at point density to see if there are any patterns to movement on the landscape.

So today worked on polishing the text and figures I have.

Talked to Peter Wolter, he didn’t have really any different ideas. He suggested talking to Dr. Moen: <https://scse.d.umn.edu/biology-department/faculty-staff/dr-ronald-moen>

TODO:

Update GitHub with most recent Repast Simphony code and clean up HPC drive.

**Notes on Papers**

*Fisher, K.E., Adelman, J.S. and Bradbury, S.P., 2020. Employing very high frequency (VHF) radio telemetry to recreate monarch butterfly flight paths. Environmental entomology, 49(2), pp.312-323.*

This paper seems to be based just on the radio telemetry data, which was kind of a pilot study. Radio-telemeters were 220 g from Holohil. Sham-tags were 300 g watch batteries. Both attached with superglue. Sample size for the butterflies for which flight paths were constructed was 13. Sample size for sham-tagged vs free monarchs was much higher. Sham-tagged and free monarchs spent about the same amount of time in the air, but sham-tagged monarchs spent more time resting. There’s one figure of step length for monarchs and generally they observed that monarchs made a bunch of short steps then a long one, similar to Levy flight. Turn angles were most common around 0 and 180, but distributed throughout. They did that complicated analysis of location, but 98.65% of locations were estimated to be in the prairie, because any time they left the prairie they captured them. The 4-bearing method provides some aid in estimating location of butterflies that are moving, but IMO the error is too large for the scale of the field. At that field, could visually locate monarchs almost all the time. Could have used flags. But at larger scales could be considered, or situations where can’t see them often. Though later they said couldn’t always relocate butterflies without the transmitters. They conclude that the movement patterns they observed were qualitatively consistent with Grant et al. 2018. Paper conclusion says they are developing automated system.

The obvious question from this paper is should the agent-based model employ step lengths that vary.

*Fisher, K.E., Dixon, P.M., Han, G., Adelman, J.S. and Bradbury, S.P., 2020. Locating large insects using automated VHF radio telemetry with a multi‐antennae array. Methods in Ecology and Evolution.*

Describes the automated ratio telemetry system (ARTS) and the statistical model to estimate the location from the recorded signals. The ARTS was 4 towers in a square. The statistical estimation model takes advantage of changes in signal strength due to both angle and distance, which is cool. Antenna physics implies these effects are additive on the logarithmic scale. Parameters in the equation have to be calibrated. Calibration data had to be collected for each field and new location of antennae arrays. The rest of the paper is tests of the system with known locations. It works ok, but never quite as good as you might hope, but it’s the best thing available.

*Fisher, K.E., Hellmich, R.L. and Bradbury, S.P., 2020. Estimates of common milkweed (Asclepias syriaca) utilization by monarch larvae (Danaus plexippus) and the significance of larval movement. Journal of Insect Conservation, 24(2), pp.297-307.*

Intro is pretty good at listing issues with studying monarch larval behavior. They also state this a lab study, so predation, parasitism, and intra- and inter-specific competition is not present. In 2016 pilot study, 33 larvae were placed in cages with a single milkweed ramet/stem. All left the ramet 2-3 x and had to be placed on new plants. After that, they placed 2-4 stems per cage. Complete block design with equal numbers 2,3,4 stems per cage, if I understood correctly, and 108+144 cages between 2017 and 2018. 2017 focused on larval ramet abandonment behavior. 2018 focused on biomass utilization. 67% of larvae across both years survived to pupation, varied from 52% to 83% depending on trial. Development from neonate to pupation took 1.6 days (SD=1.6). They were in a greenhouse, so it was warm but temp was not constant and not reported in detail. Larvae always abandoned natal ramet, usually in 4th instar. Larvae were usually on the top portion of the plant and on the underside of leaves ~60% of the time (eyeballing graph, exact number not given). Larvae abandoned ramets on average 3 x before pupation. When provided with 4 ramets, they used on average 3.22 ramets, so IMO they only need 4 good milkweed stems. However, they say that search behavior for new ramets appears to be random, so the more ramets in the area the better. Larvae eat about the mass of one ramet, but get that mass from 3-4 plants. They note that eggs laid on isolated stems almost certainly won’t survive because they will wander away and die. Nor is there enough biomass for more than one larva. Some discussion on possible external cues causing abandonment, like increasing cardenolides, but could be internal IMO – its consistent enough. When larvae abandoned the plant, 25-50% of the biomass was eaten – is that what has been observed in the wild? They also suggest larval monitoring overestimates mortality because they assume larvae that are gone are dead. I don’t know, I would have to study that closer. May depend on the assumption that immigration=emigration. Looks like that is the main issue to me.

The model experiment that needs to be run is trying a landscape with more patches vs the cleaned up matrix – expand on Zalucki’s cleaned up matrix idea with real landscapes, and probably more to explore. Because it’s not just how many stems are available, it’s how they are spread on the landscape and realized fecundity probably drops a lot when they fly through empty ag fields a lot.

*Fisher, K.E. and Bradbury, S.P. 2020a. Estimating perceptual range of female monarch butterflies (Danaus plexippus) to potted vegetative common milkweed (Asclepias syriaca) and nectar resources. Environ. Ent (Submitted).*

Steve provided a rough draft version under review. I think comments have been responded to, so it should be mostly final except for polishing. 188 monarchs released on sod farm downwind of potted plants – milkweed and blooming forbs. 49 incidental monarchs also observed. Conservative estimate of 125 perceptual range in that habitat. Conclude that placing habitat 50m apart would create functional connectivity. Argue that fitness costs increase as resources are spaced farther than perceptual range. Wild-caught monarchs were used, not lab monarchs. There is a histogram of turn angles in Fig 4, which could be useful for modeling. Also some figures of step lengths, which are time-based step lengths. Not 100% sure by what is meant by unimodal movement. In results, monarchs mostly flew with the wind, not toward resources. Did the incidental monarchs behave differently than the experimental ones? Should wind be included in the model? Someone needs to look at eggs laid per individual as a measure of fitness in different landscapes. Next step is estimating effective perceptual range.

*Fisher, K.E. and Bradbury, S.P. 2020b. Influence of resource abundance and habitat configuration on female monarch butterfly (Danaus plexippus) movement and habitat utilization patterns at a landscape-scale. Animal Ecology (In preparation).*

In this paper, 195 monarchs were released in more realistic environments – roadsides or mosaics of habitat. 77.9% of monarchs left their original habitat patch/class within 60 mins, no matter what type of habitat it was. They did cross fewer edges if started in a good habitat patch and crossed more edges if started in a crop field.

My thoughts on potential changes to movement model:

Things to consider when modeling:

* Time or space-based step length and associated distribution of step lengths
* Changing perceptual range in different habitats
* Turning angle distribution – which corr rand walk parameters correspond to observations
* Calibration with data from more habitat types could be just as useful
* Directional flight for long distances with no choices/steps
* Can test model with amount of time spent in each habitat

Another way to look at it – what’s wrong with the model now? What needs to be fixed? Is it giving wrong answers? Do we know?

**Themes for Review Paper**

* Modeling and empirical data
  + Insecticide laboratory data and spatial landscape modeling
  + Radio-telemetry and movement algorithm
* Multiple disciplines
  + Simulation modeling, statistics, toxicology, entomology, ecology, weed science, etc.
* Spatial configuration of habitat patches
  + So many things come down to spatial configuration – finding milkweed, how much milkweed is laid per patch, which patches get hit by insecticide drift and herbicide drift, how we can improve monarch production
  + Supremely important with monarchs because of vagility
* Scale – from single stems to continent wide
  + Population dynamics that scale from patches to continent wide – difficult to model, in addition to the fact that insect populations are difficult to model
  + Survival, a critical parameter – we now have better estimates

*The Big Questions*

How will our research help save monarchs?

What policy can be enacted to save monarchs?

Is current policy and management working?

*What are our take-home messages?*

Spatial configuration

State and local government-centered conservation? What is current thinking literature on this? Highlight massive efforts at state and local levels.

*Future work*

Movement algorithms

Spatial configuration simulations – theoretical

Predicting population growth rate and partitioning causes of increase or decrease – examples with other species?

Climate change – we have no research on this, but a big question for the future

**FILES**

**DMLModel4.shp** – this is the shapefile for Scenario 1 that I put into RS, after changing it to lat/long projection. Actual file in RS was **DMLModel4LL.shp**.

**Spatial\_Join\_DML4\_Counties3.shp** – this shapefile is just to get the counties for each polygon. I did a spatial join of **DMLModel4.shp** on **TechCounties.shp**.

After fixing the errors in one technicians shapefile, the new shapefile for Scenario 1 is **Spatial\_Join\_DML4\_Counties5\_UTM.shp**. I will run this again after I run Scen 2 and 3. Keep in mind that I created Story County separately for Grant et al. 2018 and Grant et al. 2021. In this shapefile, Story County was created by the technicians. The only difference should be that the nonGMO fields will be in different locations since they were randomly assigned. Still needs to be changed to lat/long proj to run. DML1.cal is probEggs. **Spatial\_Join\_DML4\_Counties5\_UTMLL.shp** is the lat/long file used in RS.

**Spatial\_Join\_DML4\_Counties5\_UTMS2.shp** is shapefile for Scen 2. *DML2.cal* is probEggs. **Spatial\_Join\_DML4\_Counties5\_UTMS2LL.shp** is lat/long file used in RS. Max ProbEggs and probMove is 0.07.

**Spatial\_Join\_DML4\_Counties5\_UTMS3.ship** is shapefile for Scen 3. *DML3.cal* is probEggs. **Spatial\_Join\_DML4\_Counties5\_UTMS3LL.shp** is the lat/long file in RS. Max ProbEggs and probMove is 0.07875.