Masters Project Activities Record

Thur 31/3/22

Read some introductory papers on insect declines, and one on meta-analysis

Fri 1/4/22

Continued reading introductory papers on insect declines

Mon 4/4/22

Worked on project proposal – improved upon the very rough draft by fleshing it out, including references and a Gantt chart. Sent to Samraat.

Tue 12/4/22

Meeting with Joe:

* Difference between Synthetic analysis and Meta-analysis
  + Synthetic analysis – done by PREDICTS - pulled out abundance data from different papers – built mixed effects models using glmer package (for non-linear mixed effect models). Data was nested – therefore needed random effects. Use model to eg predict biodiversity based on land use
  + Meta-analysis – authors have studied effect of driver on biodiversity. Collect multiple of these papers. Pull out effect size and standard error. Average across studies.
  + Mine will be more on the side of mixed effects meta-analysis – working with effect size and standard error
* PREDICTS database:
  + Nested structure of predicts - SSBS – source, study, block, site
  + Info on coordinates (lat/long), land use intensity, taxonomic group
  + NHM previously have written functions such as one to calculate diversity index across all data – I wont be building a function like that. But will probably write analogous functions e.g. for mean, effect size, SE
  + NHM also written function for doing log total abundance mixed effect model – so predict log total abundance based on other data
* Joe will ask Andy tomorrow about getting me the database
* In the meantime, work on intro
* Check with other students about getting a pass for NHM
* Joe is in next week from Tue 19th – Fri 22nd
* To find Andy's office/labs:
  + Main entrance from Cromwell road
  + Go into Hintze hall (with whale)
  + Turn right, follow down corridor with dinosaurs on the left
  + Reach Darwin centre
  + Andy's office is on 6th floor
* Can join the lab meetings

Finished reading the Sánchez-Bayo and Wyckhuys, 2019 paper about the different insect orders and different drivers of decline.

Began collecting and reading papers about the PREDICTS database, and papers that use data in this database.

Wed 13/4/22

Read and made notes on papers about the PREDICTS database, and studies that use the PREDICTS database.

Read and made notes on papers about meta-analyses.

Started working through a YouTube tutorial on SQL - SQL Tutorial - Full Database Course for Beginners. Have got basic info on databases, and have installed MySQL and PopSQL. Not sure if using SQL for project but think it will be useful to learn anyway.

Thur 14/4/22

Joe emailed me to keep working on background reading and write 2 essays – one on insect declines, and the other a more technical piece on meta-analyses in the context of insect ecology.

Read and made notes on insect decline papers.

Continued working through the SQL YouTube tutorial

Tue 19/4/22

Travelled into the NHM for the 1st time. Arrived at 11am. Got ID pass, email login, did bit of tour etc.

Continued reading papers on insect declines. Left NHM at 5:15pm.

Wed 20/4/22

Went into NHM, arrived at 10:30, left at 5:45.

Went to Joe's seminar on his PhD work – looking at climate change and land use on pollinators and crop production.

Continued reading papers on insect declines.

Thur 21/4/22

Reading papers on insect declines.

Had meeting with Joe:

* Insect declines essay will be more general, whereas meta-analysis essay will be more focused, more refined, technical piece.
* When doing a meta-analysis, due to reproducibility debate in science, there is a systematic way to do them. Use Christina's files in GLITRS channel as a starting point.
* After essays, we can then think about protocol structure and code pipeline.
* Ideally, after my work, someone can go to Scopus/Web of Science, and search for terms relating to their question e.g. effect of pesticide on insect BD, and then will return literature on this. Extract data and put into consistently formatted spreadsheets.

Fri 22/4/22

Read through last few papers on insect declines

Highlighted all notes on insect declines, meta-analyses, predicts papers

Mon 25/4/22

Putting together structure for insect declines essay. Put all notes onto about 10 pages.

Also made table of primary papers and their key info.

Tue 26/4/22

Wrote insect declines essay – 1400 (including references) – 40 references

New PhDs talks:

* Iona Cunningham-Eurich – parasitoid wasps
  + Evolutionary biology
  + Parasitoid wasps lay eggs on prey
    - Generalists vs specialists – which came first? What genomic aspects are involved?
    - Often used as biological control so important effects on agriculture
    - Is there plasticity in host choice?
* Sonia Papadaki – microbes in Arctic endolithic habitats
  + Get clues from extreme environments on Earth in order to study astrobiology
  + Vast amount of microbes that live in Antarctic and Arctic
  + Endolithic – inside rock
  + What microbial life is present? How do the communities function? What biosignatures do they have?
* Justin Isip – How do anthropogenic threats reorganize insect communities?
  + See Eggleton, 2020 on numbers of insects etc.
  + Insects are a major part of biodiversity
  + Huge increase in insect declines in last 10 years
  + BUT how accurate are these reports?
    - Biases – e.g. taxonomically and geographically
  + Will use PREDICTS database – contains biodiversity data and how it responds to anthropogenic threat (mainly land-use)
  + Also use GLiTRS database – threat model incorporating multiple lines of evidence to describe trends in insect biodiversity
  + Which subsets of insect data show greatest response to land use change/intensity?
  + Structural equation modelling – go more towards causation (rather than correlation). How different things interact within a network.
* Joe Minus – Integrating fossils and genomes to unravel the pattern of mammalian diversification through time
  + Reconstruct mammal evolutionary tree using recently developed methods
  + Mammals now are much bigger than Mesozoic mammals
* Joel Heath – Role of tectonics and environment on the distribution of dinosaur diversity
  + Focus on Mesozoic as this was a period of big global change – e.g. fragmentation of Pangaea, extinction events

Wed 27/4/22

Travelled into NHM, arrived at 10:30, left at 5:45.

Read meta-analysis papers

Thur 28/4/22

Read 11 more meta-analysis papers, so now have 24. Should be enough to be able to start condensing and organizing the notes.

Fri 29/4/22

Went into NHM – arrived at 10. Left at 5:45.

Meeting with Joe and Andy:

* For meta-analysis essay:
  + Don't need to know all details/in depth everything about meta-analyses. More just be aware of it.
  + Have to write essay in context of insect ecology but not many review papers specifically on this, just on ecology in general. But apply stuff I'm reading to insects. E.g. use random effects model when high heterogeneity, which is the case for insects.
* Send Andy proposal again to sign
* Need to read Christina's info on meta-analysis.
* Next steps for project:
  + Someone else is working on something which will be upstream of my work. They are putting together a database with uploads
  + When this is complete, I can use the code I've written to query from these data entry sheets
  + Could use Christina's sheets as dummy data? From multiple duplicated data entry sheet, pull/put results in R
  + Once the GLiTRS specific datasheet is finished, can make my code applicable to this
  + If time at end of project, would be nice to put together a shiny app, which uses R code to make websites. Front-end vs back-end development. Should be relatively easy to make a basic one. See <https://joemillard.shinyapps.io/Real_time_SAI/>
    - Joe can do session on shiny apps at some point
  + Instead of Christina's sheets, could wrangle together different meta-analyses that people have already done. These should largely map onto the ideal data structure, but some will have some missing fields.
  + Could just start with one dataset and use metafor to try out meta-analysis and see what it can do and how it works. Joe can ask Christina for this?
  + Do one steep learning curve at a time. Wait until this flattens off to start the next thing.

Read last few MA papers, put together plan for essay.

Mon 2/5/22

Condensed plan for meta-analyses essay.

Wrote meta-analyses essay.

Tue 3/5/22

Edited both essays, made more concise, sorted out references. Submitted.

Wed 4/5/22

Went into NHM, arrived at 10:30.

Sorted out remaining issues with endnote, edited the imperial\_college\_harvard style so I don't have to make so many manual changes. Made record of things I learnt about endnote.

Recapped SQL stuff I've already done and did some new stuff.

Had GLITRS meeting:

* Purpose of meetings is to raise any problems we've had with our projects
* We are all working on things to do with the GLITRS project but all doing little bits of it, so the meetings are good to see the bigger picture.
* Sanchez-Bayo paper claims to have done MA but there is no details hardly at all on how they went about this. It's basically vote-counting, but this isn't stated (possibly because it's a discredited method)
* Read Eggleton paper and Miličić paper
* The living planet index? – but this is just based on vertebrates?
* Space for time analyses? – read paper on this?
* Overview of how masters projects link up:
  + Daero and I are doing the MA pipeline to overall assess how different insects are affected by different threats. Pipeline includes loading files, manipulating the data, and analysing the data.
  + Daero will take initial inputs – each input is an already completed MA. He will then put together a protocol and data entry sheet so each MA (which are currently quite different) will be in the same standardised format. These will be in R markdown. When people do MA in future, they can add to this. All these together will eventually make up a database.
  + Based on this, I will write code which combines all of these so effectively I'm doing a meta-analysis of meta-analyses.
  + Will make shiny app in which you can filter for different groups, threats, etc. and produce a graph which will change based on the input.
  + Yeahji is doing something a bit separate – she is filling in gaps in the PREDICTS database
  + Anna – doing a meta-analysis
  + All our projects are the same length.
* Joe will send over Christina's sheets to practise on while I wait for Daero's output

Next things:

* Metafor tutorial
* Side-project – learning SQL
* Keep up with reading any papers that come up
* Start looking at shiny apps?
* Start looking at how to make an R package?

Thur 5/5/22

SQL tutorial

Read a few more papers.

Fri 6/5/22

Went into NHM – arrived at 10, left at 4:15.

Email from Joe:

* Saved Christina's files that Joe sent over.
* Need to start taking a look at Christina's meta-analytic scripts here (<https://github.com/CristinaRaw/NHM_MMA/tree/main/Scripts>)
* Christina performed a nice meta-analysis on the effect of agriculture on arthropod biodiversity which is relevant for GLITRS, but more importantly the pipeline for each individual submission will likely look similar to what she's already written.
* I think the first thing to do is to reproduce the whole of Christina's pipeline (by pipeline I just mean R code). Eventually what we're aiming for is a set of scripts that will take a set of protocols and data entry sheets for individual meta-analyses, do all the cleaning and processing automatically, and then run a meta-meta-analysis that we can visualise/download from a Shiny app.

Read few papers that I needed to catch up on.

Continued SQL tutorial – ¾ done now. (3hr15m / 4hr20m)

Started metafor tutorial – did 40 mins of <https://www.youtube.com/watch?v=IkduL5iRdqo&t=5048s>

* Want to set working directory to be in my onedrive – can't do this normal way. Need to click the three dots …



Writing workshop with Andy:

* Read How to Write a Lot by Paul Silvia
* Useful to have journal I like in mind so you can refer to it for structure/style
* Would be good to deconstruct a paper I like to see why I like it
* Get structure right before polishing
* Start writing as soon as you think you know what you're doing
* Save words by saying things in the right section
* Intro
  + Why your research is worth doing?
  + Why is it a problem that we don't know?
  + Why no one else has answered it before?
    - New technique
    - Not enough info/evidence until now
    - Only just become a question – no one thought about it before now
  + Give flavour to next section
* Methods
  + Where info from and what you have done with it
  + In thesis (not so much in paper) – why that method? Justifications
  + Needs enough info for it to be reproducible
  + Have R markdown of your analysis
* Results
  + In parallel with your methods (use subheadings – can lose near end if necessary)
  + Not necessarily in chronological order / the order you did things in – should build
  + Don't repeat things you've shown in figures/tables, maybe highlight few key bits
  + Don't have figure and table of the same info – which shows info best?
  + Caption should let figure stand alone without having to refer to the text
* Discussion
  + What is the main thing that is now clear / remain unclear despite focused look at it?
  + Most important results comes first
  + What difference do the results make to what we think we now know?
  + Use literature to put into context
  + Caveats/limitations – could these affect your conclusions? How could you find this out? What future work would answer this?
  + Balanced appraisal of your work
* Paragraph = 1 clear point
  + Start with what the topic is in the 1st sentence
  + Then evaluate, explore, give examples, make links
  + Then end with a link to the next paragraph
* 1st question of viva is commonly – tell me about your work. Structure your answer like your abstract
* Write thesis twice if possible – don't get stuck trying to edit the first one
* See nature abstract structure hand-out

Sat 7/5/22

Finished SQL tutorial

Continued metafor tutorial

Sun 8/5/22

Finished metafor tutorial

Started working on CV for Kew Bioinformatician job

Mon 9/5/22

Weekly lab meeting 10am:

* Finished my 2 essays last week on insect declines and meta-analysis
* Metafor tutorial to get to know what kinds of things it can do
* Have now got Christina's data spreadsheets and scripts so I can start looking through those to see the kind of pipeline I will need to put together
* Sharepoint – andy will try and find MAs from other students that have already been completed so I can also look at these after Christina's.
* Interesting to see if I glean the same conclusions as Christina has done
* Lab social lunch – Tue 31st May

Started going though Christina's work

* Looked through and made notes on the 2 spreadsheets I was sent, along with the 4 screening R scripts and the writing r script

Started going through shiny apps tutorial <https://ourcodingclub.github.io/tutorials/shiny/index.html> - about half way through this.

Tue 10/5/22

Windows command prompt:

* cmd
* help for command commands
* The **dir** command, which is short for **directory**, will list the contents of the folder that you're currently in.

Had a think about how I'm going to use github – will probably have to do my code and pushes to github on my linux laptop? And will do latex on there too. Use this one for Word.

Continued going through Christina's scripts – from 00. to 12.

Finished shiny app tutorial – made app which is available via github or [https://r26dnk-grace-skinner.shinyapps.io/tutorial/](https://r26dnk-grace-skinner.shinyapps.io/tutorial/?_ga=2.197517999.1244292373.1652113646-714609040.1652113646)

Wed 11/5/22

Went into NHM – trains were delayed so got in at 11. Left at 5:45.

Finished going through and making notes on the rest of Christina's scripts

GLITRS meeting (Andy, Joe, Daero, Anna, Yeahji):

* Yeahji – found 2 reports – aiming to increase the number of species we have data on from Korea (currently 6).
* What biodiversity metric is best? Everyone does it differently and say that it means biodiversity, but they're not directly measuring biodiversity. Which is best depends on your question / situation – e.g. pollination rate mainly determined by number of pollinators, rather than number of species, apart from for a few species.
* Abundance – if stays same may miss changes such as rare species being replaced by common
* Daero will not focus just on one BD metric – I can filter later on
* Neither is there a single best sampling method
* Work package 2 within GLITRS – up to 1 hr long weekly Tuesday lunchtime teams meeting – good for each of us to present/run our ideas/dataset at one of these to get expert opinion. Do when would like broader input into what we're doing.

Started working through R package tutorial - <https://ourcodingclub.github.io/tutorials/writing-r-package/>. Have made a package, tomorrow will start the next section on uploading to git.

Had chat with Christina – Joe needs to send me the rest of the files that Christina shared with him. And not sure how much of the process I need to replicate e.g. do I need to do the screening process?

Thur 12/5/22

Went into NHM, got in at 10:30.

Finished making R package and uploaded it to Github.

Journal club meeting (12:30-1:45):

* Reproducible workflow
* Doing work e.g. commercial customers, or for policy decision that we are not getting independent reviewers to look at so need a clear protocol to follow
  + This is in contrast with papers that go through peer review
* Currently – have markdown document that people read through BUT not sure if people actually go through it line by line
* Things we could do:
  + Unit testing
  + Pair programming? – 1 person coding, another person checking what you're doing. Best if 2 people have similar abilities? Or different abilities?
  + Code review – checking that projects are reproducible
  + More defensive code writing – e.g. if function is given incorrect parameters, then code it so it returns a message saying why there is an error etc.
  + Data club – help align the way code is written by discussing how we would approach certain problems
* Even for things that go through peer review, would still be good to have some sort of system in place
* Increasingly, we have to publish code as well as paper
* Have lots of comments on our writing, but not so much on our code
* Set up folder in which people can put in useful resources
* Have starter pack for new students who come into the lab e.g. use lab github? How want code to look like? How to comment? Etc.
* Have coding/writing club session after weekly lab meeting?
* R weekly podcast?
* Check basic quality assurance list for your own code before sending to someone else
* Better for code to be checked as you go, rather than done all at the end
* Without a license on your github repo, other people technically can't use it without your permission.
* Style guides – tidyverse? Styler can be used to format your code in this way

Installed styler package – now have an add-in in R where you can 'style active file' and get it to follow the tidyverse style guide

Meeting with Joe:

* Don't need to do screening stage - just try and replicate analysis stage for now
* When people do Mas in the future, there will probably always be a manual stage in which people will have to extract the data from papers etc
* But once have protocol and data spreadsheet, I can use my functions/shiny app on it
* Would be good to make another git repo, renaming files etc.
* Then can have a think about abstracting it
* Christina was using multilevel models – metafor couldn't do the things she wanted to do
* Joe has sent over zipped file
* I can try and replicate Christina's graphics, and put it into shiny app
  + Could have drop-down for BD measure, or taxonomic order, country etc. in user interface (ui) bit
  + Data\_file %>% filter(order = input$tax) %>% ggplot()
* Good idea to split app.R into 2 files – 1 with ui, one with server
* What should I do about git?
  + Git through rstudio
    - Use projects
    - Create git repo on create new project
    - Have r, data folders in this
    - Data – data sheets e.g. christina's upstream data file (clearly named with numbers)
    - Then go to github – make new repo
    - Git remote add origin in shell
    - Developer tools – personal access token – url and pac token.google authenticate git with pat token
    - Git window in rstudio – have to of installed git on rstudio
    - And can commit straight from here
    - And then push from command line (tools – shell)

Fri 13/5/22

Read and made notes on few papers. Thought "Consistent signals of a warming climate in occupancy changes of three insect taxa over 40 years in central Europe" was particularly good.

Windows laptop:

* Word for making notes
* Everything on onedrive, and also often save a copy to my documents
* Push to NHM project git repo from onedrive? – Can't find anywhere which says this is possible
  + Could make onedrive folder the remote repository (rather than github online) and clone this into a local folder on computer. And then push changes to the onedrive repo BUT this wouldn't work very well with Linus laptop because hard to access onedrive on Linux (can only see online version)
  + <https://www.permikkelsen.dk/how-to-host-your-git-repository-on-onedrive.html> follow this if ever wanted onedrive to be remote repo

Linux laptop

* Can access documents from onedrive online
* Need for latex
* Don't think I can install onedrive on linux – onedrive only for windows and mac
* So back up would be through git
* Can set up R with github?
* But not sure if easier to do through terminal as I know how to, and not just pushing R stuff

Installing and setting up git on windows laptop:

* Downloaded from https://git-scm.com/download/win
* Installed here: C:\Program Files\Git
* Kept all default settings
* With the install, you get Git Bash
  + Git Bash is an application for Microsoft Windows environments which provides an emulation layer for a Git command line experience.
  + Bash is an acronym for Bourne Again Shell. A shell is a terminal application used to interface with an operating system through written commands.
  + Bash is a popular default shell on Linux and macOS. Git Bash is a package that installs Bash, some common bash utilities, and Git on a Windows operating system.
  + Git Bash has the same operations as a standard Bash experience.
* Changed name of folder with project files to CMEE\_NHM\_Project (in onedrive)
* Made CMEE\_NHM\_Project repo on github
* Made CMEE\_NHM\_Project folder in my documents, copied contents of CMEE\_NHM\_Project folder in onedrive into this
* Go into mydocuments CMEE\_NHM\_Project folder in git bash (right click – gitbash here)
* Git init
* Create README: echo "CMEE Master's Project with the Natural History Museum" > README.md
* Check configuration: git config –list
  + Git config user.name gls21
  + Git config user.email gls21@ic.ac.uk
* Check files have been created: ls -al
* Make git ignore:
  + echo -e "\*~ \n\*.tmp" > .gitignore
  + cat .gitignore
* git status
* git add -A
  + Gives warnings about LF and CRLF but think ok
* git status
* git commit -m "Full first commit of master's project"
* git status
* Get https url from github repo https://github.com/gls21/CMEE\_NHM\_Project.git
  + git remote add origin <REMOTE\_URL>
* git remote -v # Verifies the new remote URL
* git push origin master
* Had to do some authentication thing

Think about gitignore?

Large files? - We suggest that you include files larger than some size in your .gitignore. For example, you can use the following bash command:

find . -size +100M | cat >> .gitignore

Seemed fine: Can copy whole folder (including .git etc) into onedrive folder as extra back-up

Setting up git on linux laptop:

* git clone <https://github.com/gls21/CMEE_NHM_Project.git>
* Successfully got it onto Linux
* Created test file and tried to push but says: Support for password authentication was removed. Please use a personal access token instead
* Follow this on Linux laptop:
  + From your GitHub account, go to **Settings** => **Developer Settings** => **Personal Access Token** => **Generate New Token** => **Fillup the form – clicked repo box** => click **Generate token** => **Copy the generated Token**, it will be something like ghp\_sFhFsSHhTzMDreGRLjmks4Tzuzgthdvfsrta
  + Now try git push origin master again
  + Put in the token as the password (paste it in)
  + Think I've configured it so it doesn't ask for a token password every time. Did:
    - Git config –global credential.helper cache
    - Git pull
  + Was able to push another new file without having to put the token in
* Will then have to git pull origin master each time I switch laptops before I can push

**Need to do git pull origin master and git push origin master on both laptops**

Sun 15/5/22

Finished CV and wrote cover letter for job

Mon 16/5/22

Lab meeting 10am:

* Went through Christina's scripts on meta-analysis of meta-analyses to get overview of the process she went through. This week going to try and replicate some of it, starting with the analysis stages
* Did tutorial on shiny apps and made a basic one of these, in preparation for using this as a tools to present the findings of my project
* Did R packages tutorial so can now make a basic R package
* See <https://r-graph-gallery.com/> for e.g. R graphs

Finished cover letter for job.

Wrote relevant skills and abilities section for job application.

Tue 17/5/22

Finished relevant skills and abilities section for job application.

Submitted job application.

Started replicating Christina's analysis process – did scripts 20 and 21.

Wed 18/5/22

Went into the museum – got in at 10:30, left at 5:45.

Working on reproducing Christina's analysis – did scripts 22, 23, 25, 26, 28, 29

GLITRS meeting (Joe, Andy, Anna, Yeahji, Daero):

* Yeahji struggling to get replies from people for data
* Conditions on Andy giving reference for job:
  + Copy of cv you've applied with
  + Bullet points of things to mention
  + Nag as deadline approaches until Andy has said he's sent it
* Daero – how to do multilevel MA in R?
* Insect declines workshop at NHM next Tuesday afternoon – looking for people to help get people from reception to boardroom
* Joe and Andy writing some code where you can put in location co-ordinates and get out environmental raster data. These could be used as moderators in a MA.

Did most of an R markdown tutorial - <https://ourcodingclub.github.io/tutorials/rmarkdown/>

Thur 19/5/22

Finished R markdown tutorial.

Did dplyr tutorial to get reminder of how to use this package.

Meeting with Joe:

* What will report look like?
  + Ideally more like a methods paper (unless have strict guidelines) e.g. like methods in ecology and evolution – see ones that have made a package or shiny app
    - <https://besjournals.onlinelibrary.wiley.com/journal/2041210X?utm_source=google&utm_medium=paidsearch&utm_campaign=R3MR425&utm_content=LifeSciences&gclid=Cj0KCQjw1ZeUBhDyARIsAOzAqQIuoB1H49gmRTCDu92ZnzE2l7wiX3SeBsNSKcd896V1c2wO_E9ZEj4aAiAUEALw_wcB>
    - <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13897>
    - Methods will describe engineering of a process, rather than details of doing a meta-analysis
  + Look at past CMEE projects?
  + Check guidebook
* Christina hasn't really got any graphs of her results – has tables of extracted coefficients from models
  + Organise chat with Christina and Andy to discuss different decisions she made to see if we will also use the model Christina settled on
* Need to strip everything back to just the set of models that Christina uses. Will then add the graph script to this. Then can put this into a shiny app.
* After pipeline is finished, need a way of coping with each new addition of data - it needs to be the same format
  + Way to validate data as new one is added?
* Read Christina thesis for better idea of her thought processes

Went through tidyverse style guide documentation - <https://style.tidyverse.org/index.html>

Fri 20/5/22

Went into NHM, arrived at 10:30. Left at 5:15

Condense Christina's replicated files

* 22\_preparing\_dataset\_for\_modelling\_LRR
* 26\_analysis\_LRR?
* 28\_model\_diagnostic\_with\_influence.ME\_and\_robust\_analysis
* 29\_making\_summary\_tables

Meeting with Christina and Joe (Fri 20/5 5pm and Mon 23/5 11am):

* Is the LONG data extraction spreadsheet the final output – is this made with an R script? Because I'm not currently reproducing it
  + Think it's lots of outputs put together
  + Christina designed the data extraction spreadsheet originally
* Thought process to using final robust model
* Graphical outputs – I might need to think about these
* Where did formula for converting % to LRR come from? – From a source
* Can't work out changing -ve % bit. How does -30% become +70% - it doesn't? -30 becomes +130
  + Baseline for LRR is 100 – above 100 increase, below 100 decrease.
  + Already positive percentages – add 100.
  + Christina make a figure of this
* How does swapping the columns round affect things – doesn't conventional always have to be treatment? And conservation the control? i.e. the treatment isn't always the conventional one, and the control isn't always the conservation one?
  + Guess it doesn't because the LRR will reflect this i.e. will be lower when treatment is conventional, compared to conservation
  + Lowest LRR will ever be is 0 (no change) because when it was -ve, we swapped the control and treatments around.
* When you pivot\_longer, the control\_LRR 0 values column disappears – I know this is the point of pivot\_longer but what was the point in adding them?
  + Then in the next bit, you have a bit on with control 0s and without control 0s – but where actually are they?
  + Still has same amount of columns – I thought you were condensing columns 37 to 51?
    - 37 and 51
* Scoping – see what's out there – crops/livestock species – see what to focus on – crops per area / density per livestock. Is project doable? Pilot phase to refine what you're going to do / how to approach it
* Screening – 1 phase for each commodity. Download relevant info for each paper. Sometimes multiple rounds. Left with papers to assess at full text according to criteria in protocol
* Metafor – wide format better
* Wide to long when decided metafor wasn’t suitable
* Extraction – made spreadsheet with variables needed to assess question, plus papers to assess at full text. Then pilot the spreadsheet – take 10 papers and extract data from them. Drop/add columns as necessary.
* Cleaning – make naming of things more consistent. Decided LRR made more sense. Wide to long when decided can't use metafor. Can't use metafor without mean, sample size, and variance. MAs and reviews don't report these.
  + Adriana might have used this other approach too?
* Analysis – 26analysisLRR
  + When make it long format – all control become 0. Horizontal line in qq plot shows control 0s. Control zeros in LRR column.
  + Random ID and Crop to account for similarities among results
  + Compare all treatments to conventional (most intensive category)
  + In future, if have more data, it may be useful to include the interaction between biodiversity metric and treatment
  + Could drop 0s entirely – but doesn't make any difference to the predictions
  + Including crop as random effect doesn't improve model fit but keep in because makes sense according to question
  + Rethink which model is best as more data is added? Maybe influential data won't be influential for e.g.
  + Robust module accounts for influential data
* Analysis – 28modeldiagnosticwithinfluence.ME
  + Check out influential data
  + But these weren't outliers so left them in
  + Sigtest – does removing influential data affect size/significance of estimates?
  + So decided to try robust package
    - Have to tune the fit / increase efficiency – followed tutorial/example – Christina doesn't fully understand these bits either – see vignette
    - Robust model accounts for things but overall results are pretty much the same as the Bayesian model
* More like predicts analysis, than metafor MA. But also different again
* Put all curation in same script – remove from analysis bits
* Still include Bayesian models? For future
  + 2 separate scripts – one for Bayesian, one for robust
  + Work from robust for shiny app
* Think about how to put into shiny app
* Carry on condensing and organising
* Defensive programming for accounting for structure of new data
* Shiny app – try building maximal model. If can't, try next model instead
  + Not available for now – try including interaction – will show up in app once no rank deficiency issue

Empty spreadsheet? – Christina send over

How to deal with the literature without going mad workshop (Andy):

* Why cite the literature and what type should you cite?
  + Go look here for more – to direct the reader to a source with more info
    - Cite technical papers for where you've followed the methods of someone else
  + Give credit to people who said it / acknowledge the work of others / avoid plagiarism
    - Cite primary paper where this idea started.
    - Or could cite a recent review for the history of a subject.
    - Or indirect citation – e.g. Heckel 1869 in review 1989.
  + Give point some evidence / giving weight to argument – not just me saying it. Show there is a body of evidence
    - Cite a strong source of evidence (good paper).
    - Could cite MA / review.
    - Highly cited paper.
    - In field-leading journals.
    - Government / UN / World Economic reports good for intros / discussion. Good for what matters to society. Very under-used.
  + Contextualising what you've found – related to others. To show you're aware of literature and how your work links
    - Reports for broad focus
    - Narrow focus - What is citing the papers that you cite at the end of your intro – these will be likely be relevant to your work. See what other people are doing that are doing a similar thing to you.
    - Set up google scholar alert – do this regularly to keep up with the literature discussion
* Cite a paper for its main point – or can be seen as lazy
* Highlight how citation is relevant to the point you're making
* You do not have to have a reference for every statement you make – don't cite unnecessarily
* 120 references for master's thesis is excessive – aim for ~50 but very wide range of what is acceptable
* Put data you are citing such as PREDICTS as an appendix
* Stick to primary papers, reviews, reports (rather than articles)
* Author, date citations (not numbers)
* How should you read a paper?
  + Evaluate the authors interpretation of their results – is it actually showing what they say it is
  + Can generally trust that the authors have been competent technically in a peer-reviewed paper
  + When making notes, use 2 different colours – one for authors interpretation, one for your interpretations
  + Maybe print out most important papers and really go through in detail
  + Write a point(s) for which that source is a good point / take home message in own words (include citation in this) with sentence from paper which is the underlying evidence that the paper reports. This is the sentence that goes in the 1st draft of a manuscript. Can easily go back to these for future manuscripts so don't have to go back to primary papers – can just search for keywords in your 'good points' and re-use these
  + Fine to re-use points if they are your own words – self-plagiarism isn't a thing because it's your own words
  + If doesn't look relevant, stop reading straight away
  + Tag points/paragraph for certain topics
  + Don't need to read all papers completely top to bottom
  + Use software – think justin uses notion
  + Don't have time to keep going back to the same papers
* Decide how much time you are going to spend on reading each week – start with an aim of 4 hours a week – don't go over

Next week:

* LaTeX set-up
* Structure of thesis
* Make shiny app be in 2 files
* Notion? Or at least making key point in my own words for each paper?

Mon 23/5/22

Lab meeting 10am:

* Applied for job
* Replicated Christina's analysis – currently going through thought process with Christina and Joe. This week going to think about some ways of presenting the findings graphically and start trying to put it into a shiny app
* R markdown, dplyr, styler tutorial
* Literature workshop

Meeting with Christina and Joe 11am – finished going through rest of stuff from Friday's meeting – notes are with notes made on Friday.

Continued condensing Christina's scripts into what is relevant for me.

Started thinking about plotting results and made very basic initial plot:

Chart, bar chart

Description automatically generated

Eventually got the graph to this stage:

Chart, box and whisker chart

Description automatically generated

Had slight issue with git because I ended up with 2 versions of christina\_mma (one capitalised, one lowercase) in git hub online (but not in files)?

* I did git config core.ignorecase false to make it recognise that I've changed file names from capitalised to lower case but this didn't fix.
* In the end I had to move the folder out of the git repo, then push. Then put it back in, and then push again. This fixed it.

Had to put in another PAT when git pulling from Linux laptop?

Have done a back-up on a memory stick today too.

Tue 24/5/22

Went into NHM, arrived at 10:30am. Left at 5:45pm.

Now have:

* 00\_preparing\_dataset\_for\_modelling\_LRR
  + Input:
    - 04.Excel\_Magpie\_Crops\_WithControl\_Quantitative\_spreadsheet (1).xlsx
  + Output:
    - 06.Excel\_Dataset\_to\_model\_LRR\_WIDE.xlsx
    - 07.Excel\_Dataset\_to\_model\_LRR\_LONG.xlsx
* 01\_bayesian\_analysis\_LRR
  + Input:
    - 07.Excel\_Dataset\_to\_model\_LRR\_LONG.xlsx
    - 06.Excel\_Dataset\_to\_model\_LRR\_WIDE.xlsx
  + Output:
    - model.rds
* 02\_robust\_analysis\_LRR
  + Input:
    - 07.Excel\_Dataset\_to\_model\_LRR\_LONG.xlsx
  + Output:
    - blmer\_model.rds
    - robust\_model.rds
* 03\_making\_summary\_tables
  + Input:
    - blmer\_model.rds
    - robust\_model.rds
  + Output:
    - blmer\_coefficients.txt
    - robust\_coefficients.txt

Need to know the inputs and outputs so I can make sure I can move stuff around without issues – want to put all data curation/cleaning in 1 script. Did this today.

Shiny apps:

* global.R - <https://mraess.rbind.io/2018/07/the-awesomeness-that-is-the-global-r-file-or-how-to-clean-up-your-shiny-app/>
  + Think of the global.R file as a file that is being run once before your app starts. That means you can use it for all sorts of data processing, running models, and, of course, to load in your data.
  + Any R objects that are created in the global.R file become available to the app.R file, or the ui.R and server.R files respectively! This makes building a dashboard a lot cleaner. Now, you can handle a lot of your pre-processing in the global.R file and then feed the results into the server.R file simply by using the object. No need to use source() to load anything either.
* Then also have separate ui.R and server.R files
* Also have data directory to store the data

Meeting with Joe:

* 2 files - is it just ui.R and server.R, don't need one that brings them together? i.e. global.R, which I think you can use to write functions and run models, on top of loading packages and data
  + global.R could be used to set environment options
  + Don't bother with this for now – just stick with ui.R and server.R
* Not sure how to filter – filter before model? Or after? i.e. filter what goes into model OR filter what results are displayed after model is run? E.g. Have choices over which treatments to display results for?
  + For now at least, filter after
  + It will slow down app a lot if has to run model every time
  + shiny is slow for ggplot anyway – want to minimise the amount of work it has to do on the fly
* Reactivefilereader – so can take into account updates to the spreadsheet as new data is added?
  + Think this will have to go in server.R
  + Have seen some people loading data and packages in global.R BUT will this work with reactivefilereader? Not sure it will – think it has to go in server.R?
  + Eventually, this will be something to consider, but leave for now

Overall:

* Shiny:
  + Don't bother with global.R or reactivefilereader, at least for now
  + I need to run the model as Christina has done, and save this to disk in the shiny app directory. Then, in the server file, I can read in the model, and produce ggplot from here based on what the user wants to see
* Read this blog post about R projects <https://www.tidyverse.org/blog/2017/12/workflow-vs-script/>
  + Have 1 project for normal scripts, 1 for shiny
  + R projects makes relative paths easier as working directory is set as the one your project is in
  + Also removes the need for rm(list=ls()) as environment is cleared automatically when R is restarted
* Talk to andy about approach – compare each individual treatment to its control
* For other scripts (not shiny):
  + Make it so that all inputs are at top of script, and all outputs at end (don't have write to disk in middle of script)
  + Only have one output per script
  + Completely separate Bayesian and robust – WIDE and LONG formats needed for Bayesian, only LONG format needed for robust
  + Use ggsave() rather than pdf() to save images of plots
    - 1st argument is file name
    - dpi for resolution
    - scale

<https://www.tidyverse.org/blog/2017/12/workflow-vs-script/>

* Don't have setwd() or rm(list=ls()) in script
* I suggest organizing each data analysis into a project: a folder on your computer that holds all the files relevant to that particular piece of work. I’m not assuming this is an RStudio Project, though this is a nice implementation
* The problem is that rm(list = ls()) does NOT, in fact, create a fresh R process. All it does is delete user-created objects from the global workspace. Many other changes to the R landscape persist invisibly and can have profound effects on subsequent development. Any packages that have been loaded are still available. Any options that have been set to non-default values remain that way. Working directory is not affected.
* Daily work habit: Restart R very often and re-run your under-development script from the top
* Lastly, rm(list = ls()) is hostile to anyone that you ask to help you with your R problems. If they take a short break from their own work to help debug your code, their generosity is rewarded by losing all of their previous work

Continued working on organising files – split robust and Bayesian analyses into 2 separate R projects. Currently in the process of organising these.

Wed 25/5/22

ggsave()

* defaults to saving the last plot that you displayed
* automatically detects the file format from the file name
* You can change the output size with height and width options. You change the unit with the units option. By default, in (inches) is used.
* Chart

  Description automatically generatedYou can control the resolution of the output image by specifying DPI (dots per inch) using the dpi option. The default DPI value is 300, but you can specify any value suitable for the output image, including “retina” (320) or “screen” (72). 600 or higher is recommended when a high resolution output is required.

Finished organising files – now have 1 R project for Bayesian analysis and 1 for robust analysis.

Started making shiny app:

output$table <- renderTable({

d()

})

* Why the curly brackets as well as the ()?
  + You could do renderTable( d() ) if you wanted to, but the braces make it feel more like a function.
  + Whenever you have a function argument that's more than just a single expression, you need to use curly braces
    - By more than a single expression meaning because the code block is made up of three distinct expressions e.g. assigning to x, assigning to bins, and calling hist
* Chart, box and whisker chart

  Description automatically generatedChart, box and whisker chart

  Description automatically generatedGot lots of errors the first few times I tried to preview the shiny app but eventually managed to produce something than ran without errors.

GLITRS meeting (Andy, Anna, Yeahji, Daero):

* Andy will send another dataset over to me following this meeting.
* Yeahji needs to send official form to researchers to get data.
* Daero – putting together an R markdown which people conducting MAs can use. Ideally will put together something which develops a user's search string for them.
* Anna – learning METAGEAR and metafor. Has done the 1st screen of studies.
* How do we cite other people's code that we've re-used? Until now: "Based on template code provided by xyz, I implemented the analyses in R". BUT now heading towards code that is more widely shared so if available, can just cite it like you'd cite a paper.
* Suggestions for my shiny app:
  + Have sample sizes on right side of plot.
  + Shiny app – have option to plot LRR or % change.
  + Socialising – ask Christina, Victoria, Joe for suggestions for design of shiny app.
  + Longitude and latitude – more will get filled in as we go. Be aware that this might be good to include later on (especially with the code Andy and Joe are currently writing that returns location info based on the inputted co-ordinates).

Thur 26/5/22

Went into NHM, arrived at 10:30, left at 5:45.

Shiny app:

* Did a lot of playing around with this today.
* Had an error "warning longer object length is not a multiple of shorter object length". Fixed due to:
  + == does element-wise checking for equality. I suspect what you want to do is find which of the elements of dih\_y2$MemberID are also in the vector memb. To do this you would want to use the %in% operator. So had to swap == input$treatment for %in% input$treatment
* Added in significance stars – took quite a long time to get it to do what I wanted it to do. Not sure if there is a better way?
* Need to add in sample size
* Need to have option to plot LRR or % change
* Need figure legend and intro to app
* Sort out error when no inputs are selected
* Having to load data in both ui and server files?
  + Reactive in the server file, but can't make it reactive in the ui file
* Not sure if I'm making the app in the most efficient way, but whenever I ty to condense the code, I end up with errors.
* What happens if you change the dataset required by a deployed shiny app?
* Deployed app for 1st time <https://r26dnk-grace-skinner.shinyapps.io/meta_meta_analysis/>

Found some papers that report their shiny apps in methods in ecology and evolution – need to read these to get ideas for how to write my report.

Downloaded Notion desktop app - need to ask Justin how he uses it.

Fri 27/5/22

Shiny app:

* Added in the option to plot percentage change or adjusted\_LRR
  + Had to go back to original plotting robust model script and add in columns for percentage change confidence intervals to these could be used to plot percentage change
  + Involved using varSelectInput in ui script, and then specify the variable with !!input$metric in the server script
  + In the end had to use an if statement, to get it to plot one graph if the user selects one option, and the other graph if they select the other. Couldn't just use input$metric throughout because couldn't change the error bars etc using this
* Add an intro to the app
* Added a figure legend which updates based on what the user chooses to plot
* Fixed error that occurs when no agricultural systems are selected in the app by using req(input$treatment) in renderPlot{()}
* Added in sample sizes table and table legend under plot
  + This might have to be rearranged in future
  + Had to make sample sizes table in 1\_robust\_modelling and save in data and models folder (Mon 30/5/22 update: no longer need this table as I cbind-ed the sample sizes to the robust-coefficients table)
* Re-deployed app

Started reading papers that report their shiny apps.

Sat 28/5/22

Finished reading papers I'd found that report on shiny apps / packages.

Mon 30/5/22

Lab meeting 10am:

* Finished condensing and re-organising relevant scripts from Christina
* Made shiny app
  + Get ideas/opinions
* Read some papers by researchers who have designed and reported their shiny apps

Ran git config --global credential.helper store on Linux laptop to try and get it to stop asking for PAT every time. Not sure this has worked.

Set up LaTeX document on Linux laptop.

Shiny app:

* Completely changed layout – from sidebarLayout to using FluidRows
* Adjusted sample size table to the order matches the order of agricultural systems shown in the graph, and includes conventional as the 1st row
* The agricultural systems choices are now across 2 columns (used custom CSS which I found online)
* Included link to github code, along with github icon
* Re-deployed

Tue 31/5/22

Went into NHM, arrived at 11 (tubes were delayed), left at 5:45pm.

Clear-up of lab and social lunch, dippy exhibition, sciences library.

Chat with Joe:

* Discussed shiny app – no major changes needed at this point
* Take at a look at <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3901538.2/> about taxize package – might be more useful than the papers in Methods in Ecology and Evolution?
* Start writing intro and methods
  + Why useful? Motivation for doing?
  + Case study/how to use?
  + Where data from?
* Ask around about shiny app
* Take a look at <https://golemverse.org/> - to turn shiny app into R package

Wed 1/6/22

Worked out travel expenses for project so far.

Read paper that I like that I’m going to dissect why I like it.

GLITRS meeting (Andy, Joe, Anna, Yeahji, Daero):

* Anna – screening – probably being too generous with selecting studies at the moment. May need to reduce the number of studies further by having more strict criteria for acceptance.
* Daero – writing protocol format. Trying to work out what variables to use e.g. taxa, geography, metric etc.
* Yeahji – trying to understand how PREDICTs spreadsheet works and how to add data to this.
* Me
  + Need to try and include option over whether to rescale the x axis or not as users select and unselect agricultural systems. Good if they could choose whether to have a fixed x axis (based on max and min values we have in the dataset at that point) or have one that rescales depending on the agricultural system selected. Probably use if statement. If can’t have option, then just do fixed axis. Looked into radio buttons for this?
  + Need to think about how we are going to adapt for new datasets being added?
* There is a workshop on Monday afternoon that would be worth joining (at least for the 1st bit) – Joe will send link.

Thur 2/6/22

Added in option over whether to have a fixed x axis or an x axis which re-scales to the shiny app – only issue is that the axis labels are different lengths which slightly alters the width of the box as different agricultural systems are selected.

Finished going though Engelhardt et al., 2022 paper and why I like it.

Fri 3/6/22

Intro and methods – trying to put a structure together.

Mon 6/6/22

Lab meeting 10am:

* Shiny app – completely changed the layout. Added in option over whether to rescale the x axis or not
* Reading papers that report their shiny apps / R packages
* Started thinking about structure of intro and methods
* This week:
  + Go through papers I’ve read already and write key point for each
  + Intro
  + Student conference
  + Expert elicitation workshop later today

Applied for Syngenta Bioscience Data Scientist job.

Expert elicitation workshop:

* Global Insect Threat-Response Synthesis
* Collate global evidence on threats to insects and how insects are responding
* Part of the project involves expert elicitation – asking experts to estimate the responses of insects in data-deficient taxa and regions
* This workshop is about informing the structuring of the expert elicitation
* Aim to generate ranking of importance of different threats (according to IUCN threat categories) to different orders
* How important is each threat in driving population changes in each order?
* Agriculture and aquaculture – to do with land-use. Pesticides etc. are covered under pollution category.
* How do we assess importance of a threat?
  + Exposed, Sensitive, Low adaptive capacity
    - If has all three of these – at greatest risk
    - Think holistically around this
  + Discuss threat category cards – assign to high, medium, low, and not important
  + Once have broad ranking, will the refine this further
  + Consider abundance, richness, and diversity
  + Don’t worry about intensity of threat at the moment – this will be refined later
  + Consider effect averaged globally
  + Can add stickers to each card to add additional information – data deficient, positive/mixed impact, synergistic effects, facilitatory effects
* Lepidoptera break-out room - Gavin Broad and Ralph Hobbs
  + Sorting the different threat categories into not important, low, medium, or high importance
  + Then have to rank cards within each threat category
* After break-out rooms – everyone else can have input on the other categories too
* Feedback
  + Is synergistic sticker too much to think about on top of everything else?
  + How to cope with functional differences between groups?
  + How to measure global extent of issues raised?

Started going through literature review and adding in key point for each paper that I can cite each paper for in my thesis – did 9 papers.

Tue 7/6/22

Continued going through lit review and adding key points for each paper that I could include in my report – done 44 in total now (35 today).

Meeting with Joe:

Questions/thoughts before meeting:

* Show app – now includes option to re-scale x axis
* Guidance on writing – what will go in each section?
* Need some more long-term goals – sense of lack of direction at the moment
* Andy sent me over data from Hanyu but not sure what I’m meant to be doing with that at this point?
* Andy essay comments?

Talked about:

* In terms of the shiny app adjusting the box size based on the longest agricultural system name, it’s not too much of a problem at the moment because the names are relatively similar in length. BUT may want to encourage future researcher to keep names of agricultural systems relatively short.
* Long-term goal = Platform onto which we can upload lots of MA and then run models and display results based on what interested in
  + Done via Daero creating standardised protocol that all future MA will follow
    - Protocol = Set of parameters in which the MA will be carried out based on – methods you write before doing MA
    - Ideally re-purpose things that Christina did. At the moment my app is based on what Christina did but eventually it will be used to analyse data produced using Daero’s standardised protocol. So don’t want it to be too different or will need a lot of alterations later on.
  + Then my bit can rbind all of these together and use the shiny app to analyse and display results
  + Think eventually the model will have to be run based on parameters the researcher is interested in e.g. taxonomic group
    - Issue is that the robust model takes a long time to run so can’t really have the shiny app doing this directly
    - Joe’s app https://joemillard.shinyapps.io/Real\_time\_SAI/ - trends tab – reads in files from aws and the actual processing is not done in the shiny app
    - So eventually we might want something like this where the model is run elsewhere and the shiny app just pulls the info from it and produces the ggplot
* Need to have meeting with Andy and Daero too
* For now continue getting skeleton of report together – can send to Joe (and Andy) for comments at any point
* Good stage to chat to some of the other GLITRS researchers to get their thoughts on next steps. Put together few slides and present shiny app at next meeting Tue 14th June 1pm.

Wed 8/6/22

Went into NHM – arrived at 10:30am. Left at 5:45pm.

NHM Student Conference – went to couple of the morning talks. Then spoke to Royal Entomological Society and looked at student posters. Went to last talk by Eileen Cox, director of post grad studies.

Finished going through lit review and adding key points for each paper that I could include in my report – done 53 in total now (9 today).

* Didn’t add shiny app papers to this because these are for getting ideas of how to structure report, I won’t necessarily be citing them
* Didn’t add MA technical papers because not sure how relevant they are for now

GLITRS meeting (Andy, Joe, Yeahji, Anna, Daero):

* Yeahji – learning how to put data on PREDICTS database and continuing emailing people for data
* Anna – refining search criteria
* Daero – put together backbone of protocol – how search literature, extract data, etc.
* Me - What to cover in slides
  + Difference between predicts synthetic models and MA models
  + Need standardised formats of data entry and protocol
  + Need way of filtering different things depending on what interested in
  + Overview of what I’ve done and what based on

Thur 9/6/22

Went into NHM, arrived at 10:30.

Read Hallmann et al. (2017) paper in more detail on train.

Went through and highlighted all the key points from the papers I’ve gone through.

Made very big structure for intro.

Went to the afternoon talks at the student conference.

Fri 10/6/22

Wrote 2000 word 1st draft of intro – just need to finish off last couple of paragraphs.

Mon 13/6/22

Went into NHM after spending weekend in London. Arrived at 10am. Left at 3:30pm.

Weekly lab meeting:

* Last week:
  + Expert elicitation workshop
  + Student conference
  + Wrote key point for each paper
  + Wrote 1st draft of intro
* Sorted out meeting with Andy and Joe at 11:45am Tue
* Andy in museum Tuesday and Thursday
* Friday R markdown session with Adriana – 2pm
* Don’t get off train at Ascot due to races
* At end of project, write some ‘lessons learned’ – what went well, what didn’t, what you’d do differently.

Finished intro

Presentation preparation:

* Re-deployed app with updated changes so can just use link in presentation
  + Use library(rsconnect)
  + rsconnect::deployApp()
* Put slides together outlining:
  + Differences between synthetic analyses and meta-analyses
  + Premise of my project
  + Overview of the process so far
  + Shiny app demonstration
* To do presentation:
  + Settings – display – multiple display – extend these displays
  + Have powerpoint ready to go on laptop
  + Join teams call on laptop
  + Share monitor screen, click laptop screen, then press f5 to start presenting
  + Once at last slide – open up web app
  + Then stop sharing
  + Then re-share window with app in

Tue 14/6/22

Practised presentation.

Meeting with Joe and Andy 11:45am

* Go over slides for GLITRS – all good
* Essay feedback – still in the works
* Questions:
  + Feeling I have a lack of direction at the moment
  + At the moment – the shiny app works by the model already being run and it is the extracted coefficients from the model that are fed into the shiny app
  + But surely we want a system where the model can be run in the app based on user choice for what they are interested in?
  + BUT very slow to run robust model in shiny app so not possible to do this the current way we’re doing it
  + I think right at the start of the project, we said about writing R functions that pull out the info the researcher is interested in based on their questions? Is this still an option or not possible until we have multiple MAs all with the same standardised data entry sheet?
  + Incorporating the other variables based on long/lat?
* Answers / what was discussed in meeting:
  + Re-run model only when upload new dataset
  + Then user can chose what they want to plot – agricultural system or other variables
  + Plot one variable and have corresponding checkboxes
  + Different tab for each threat class
  + Different model for each threat class
  + Adding in new data:
    - See how Hanyu’s data maps onto Christina’s
    - Could subset Christina’s’ dataset and show can add this in
    - Have upload new meta-analyses button in app
    - Will have to have checks to check that what is uploaded is what we want
      * Is it a csv
      * Right set of columns?
      * Authentication?
      * Don’t share link
      * This isn’t in required format – check this document – Daero’s standardised sheet

GLITRS meeting 1pm – present what I’ve done so far

* Lynn Dicks
  + See shiny app from conservation evidence which does similar thing. Paper called ‘Dynamic meta-analysis’. Shiny app called Meta dataset
  + Ask Christina for the definitions of the agricultural systems
  + Would be great if I could include a map of where each of the studies comes from? To how globally representative the results are
* Rob Cooke
  + Look into interactions between drivers?
* Nick Isaac
  + Consider the audience for the app
    - Just use internally for NHM? Or published as separate piece?
  + Good to have a tab in the app which gives intro
    - Include table of number of observations BUT break down further i.e. number of studies, number of observations etc. (show hierarchy)
    - Give definition of log response ratio – what is it a log response ratio of?
    - What does the reference log response ratio of 0 mean?
    - Explain other terms e.g. Bt
* Andrew Bladon
  + Also liked the idea of having more columns in the table to show where the data comes from
  + And agrees about the comment on understanding the definitions of the agricultural system and what they are based on? Where they came from?
* James (Stellenbosch University) and Charlotte Outhwaite also in meeting, along with Andy and Joe

Wed 15/6/22

Read “Dynamic meta-analysis: a method of using global evidence for local decision making” paper – describes a shiny app which is very relevant to mine.

Did some planning because now nearly half-way through project.

|  |  |
| --- | --- |
| Week/Date | Tasks to complete |
| 10 – 13/6 | * Wed afternoon/Thursday – Methods * Sat – Understanding how tabs work in Shiny apps |
| 11 – 20/6 | * Make intro tab, include updated/expanded table (on number of observations for each agricultural system) in this * Run model in the app (not just use extracted co-efficients) * Include button for uploading new data in intro tab |
| 12 – 27/6 | * Getting new data added successfully with checks in place * Start on results section |
| 13 – 4/7 | * Play with models – adding in models for different threats/variables – have different tab for each? * Finish 1st draft of results |
| 14 – 11/7 | * Start discussion |
| 15 – 18/7 | * Finish 1st draft of discussion |
| 16 – 25/7 | * Have initial report complete by end of this week – including abstract and acknowledgements |
| 17 – 1/8 | * Additional tweaks e.g. map of where studies are from * Re-drafting of report * Turn into R package as well? * Transfer report to LaTeX * Sort out github |
| 18 – 8/8 | * Same as week 17 |
| 19 – 15/8 | * Same as week 17 |
| 20 – 22/8 | * Same as week 17 * Hand-in |
| After hand-in | * Presentation/viva |

Started writing methods.

Updated R to version 4.2.0 (2022-04-22 ucrt) -- "Vigorous Calisthenics" – because of this, will have to re-install packages as I need them.

Glitrs meeting (Andy, Victoria, Daero, Anna, Yeahji):

* Updates from me since last week:
  + Wrote intro, which I’ve sent to Andy and Joe to look at
  + Attended Glitrs meeting to present what I’ve done so far and get some feedback and suggestions
  + Read paper that was suggested to me ‘Dynamic meta-analysis’ which is very relevant to what I’m doing – not sure why I didn’t find it before
  + Started on methods
  + Next steps for app
    - Include intro tab to give more background to app
    - Include upload button so new MA can be added – the model will have to be re-run when this occurs
    - Then can start thinking about having other tabs within the app for other variables/threats apart from agricultural system
* Me – because it is the robust models that are slow, use different model but make it clear that it is not a robust model. Good for data exploration without the need for R. Ideally need shiny app that does analysis on the fly. As more MAs added, may make sense to move away from robust model anyway. Just make limitations clear. Perhaps could include option for both?
  + If can do analysis on the fly: option to filter by location e.g. show me just Europe
* Anna – starting with data extraction. Has blank spreadsheet from Christina. Few issues with metagear package for extracting data from figures.
* Daero – finalised topic of MA - effect of pesticides and insecticides on Odonata species abundance – found around 350 original papers. Protocol also needs more work – could send out protocol protocol for now with bits unsure on in square brackets.
* Yeahji – entering data into PREDICTS – currently using Excel, but need to switch to R.
* Give critical but balanced appraisal of your work/decisions

Thur 16/6/22

Went into NHM, arrived at 10:45 (delays on district line).

Read Sanchez-Bayo paper in detail.

Attended seminar online: Climate change impacts on the physical environment of lakes by Iestyn Woolway

Continued writing methods.

Looked into tabs in shiny apps:

* Think I’m going to want to use navbar pages (multiple distinct sub-components each with their own sidebar, tabsets, or other layout constructs).
  + Because tabs are displayed horizontally, there’s a fundamental limit to how many tabs you can use, particularly if they have long titles. navbarPage()and navbarMenu() provide two alternative layouts that let you use more tabs with longer titles.
  + navlistPanel() is similar to tabsetPanel() but instead of running the tab titles horizontally, it shows them vertically in a sidebar. It also allows you to add headings
* Graphical user interface, text, application

  Description automatically generatedOr tabsetPanel

Manage to produced working app that has tabs:

* Have intro tab including the table of frequency of observations of each agricultural system
  + Table isn’t reactive to the agricultural systems that the user selects anymore because it didn’t really make sense for it to change based on changes made in a different tab
* Have agricultural systems tab which has the graph and user input options
* Have third tab that is not used for anything so far

Also changed the theme of the app. Used flatly bootswatch theme with slight customisation.

Re-deployed app.

Started looking at file uploads in shiny apps:

* Use fileInput()
* Will need to use req() to make sure file is uploaded before running subsequent code
* The accept argument allows you to limit the possible inputs. BUT only gives the browser a hint as to the type of file the server expects – so probs need more checks than this.

Started looking into progress bars for when models are running:

* Use withProgress()
  + <https://www.quantargo.com/help/r/latest/packages/shiny/1.6.0/withProgress>
  + Not sure I’m going to be able to use this because need to use it with [incProgress()](https://rdrr.io/pkg/shiny/man/withProgress.html) which requires being able to divide the big task into a known number of small pieces that each take roughly the same amount of time.
* May have to use transient notification
  + See section 8.2 Notifications <https://mastering-shiny.org/action-feedback.html#transient-notification>
  + Can hopefully code it so the notification disappears upon completion
* Or try the waiter package to give buffering circle

Fri 17/6/22

Bug network prep

Mon 20/6/22

Bug network work. Fieldwork – collected 5 individuals of 10 species, that we had to identify. Took back to lab and rehydrated.

Tue 21/6/22

Bug network work. Labwork – picked 5 leaves off each individual plant and weighed fresh mass. Scanned the leaves from each plant. Worked out way to get imagej to work.

Edited and submitted seminar diary.

Read Crossley et al., 2020 paper in detail.

Wed 22/6/22

Bug network work. Measured all leaf surface areas with imageJ.

Applied to Met Office Scientist in Data Assimilation job.

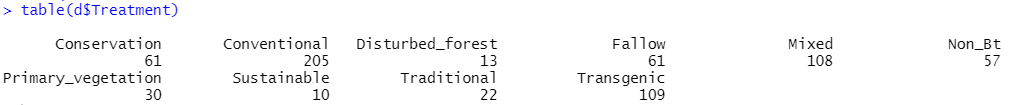
GLITRS meeting (Victoria, Anna, Daero):

* Updates from me since last week:
  + Made a good start on methods
  + Made tabs in shiny app
  + Last 2 days – bug network work. Global research project to understand the impact of insects and fungi on plant communities. Field and lab work. So far identified and collected plant specimens from field, and weighed fresh mass, and calculated leaf surface area. Got to measure dry mass on Friday once leaves have dried out.
* Anna – working on data extraction using metaDigitise
* Daero – working on scoping and screening for his meta-analysis

Things that will needed to be included in the app so I can run the model in the app and extract coefficients from it:

* Final version of spreadsheet in data folder - 07.Excel\_Dataset\_to\_model\_LRR\_LONG.xlsx
  + This will need to be saved in the data\_and\_models folder of the app
  + For now, in the server file, just read in this single datasheet
  + Eventually will have to read in all datasheets and then combine them
* Then run model on it
* Extract the coefficients from the model and calculate other columns needed
* Also make sample size table based on 07.Excel\_Dataset\_to\_model\_LRR\_LONG.xlsx
* Merge sample size and extracted coefficient model
* Then have the rest of the server file as it is? So basically just adding in previous steps to what is currently in the shiny app.

Potential future issue? – would be difficult to add in data that looked at treatments/agricultural systems outside of the 10 I have now.

* Issue comes mainly from when you extract the coefficients from the model. Rather than renaming the treatments manually by putting the treatments in a list, could you just change them by using unique$treatments from an older dataframe? Think they are in alphabetical order at this point so should be possible.
* Now the main issue is that we want Conventional to be 1st?

Then start thinking about making a complex sample size table – use paper numbers (there are 23 papers in total).

Thur 23/6/22

Fixed potential future issue outlined above:

* In 1. Removed renaming of agricultural systems when making sample size table. So ones with 2 words have underscores
* Wasn’t too bad to fix – in 2\_plotting\_and\_tables.R script, I just used the row names of the sample size table to name the robust coefficients treatment column.
* So now shouldn’t be an issue for new MAs to include new agricultural system types

Ran model from within the app:

* From 0\_preparing\_LRR\_dataset.R script, save 07.Excel\_Dataset\_to\_model\_LRR\_LONG.xlsx into the shiny app data\_and\_models folder – saved as christina\_mma
* Issue: was originally going to load this into the server file and run the model in here BUT the ui.R file currently uses the robust\_coefficients table which contains the extracted coefficients from the model.
* So thinking of making a global.R file which runs the default model using all the data.
  + This will be run once at the opening of the app.
  + The coefficients will be extracted from this.
  + And then this object will be available to the ui.R and server.R files.
  + The ui.R can use this as it does currently
  + And the server.R can either use this default model to plot, or run different models based on user input e.g. just look at Europe. (Not sure what type of model this will be yet, or how this will be implemented)
* Think for now, the 1st thing to do is implement the global.R file and obtain the extracted\_coefficients table that both ui.R and server.R can use
  + They both currently use robust\_coefficients.csv straight from data\_and\_models so they will still both end up reading in the same table, but that table will be created in global.R rather than already being present in data\_and\_models
  + Have done this and the app works.
* Have re-deployed the app – it is quite slow to start up now

Making a more complex sample size table:

* Not doing this in the shiny app to start with
* A picture containing graphical user interface

  Description automatically generatedHave produced the following table:
* Not sure it quite gets across what I want
* E.g. want to get across that there are 22 studies in total. Some look at multiple agricultural systems, and within each agricultural system looked at, some include multiple instances of each
* Is there a way to show a table which has each paper, and then total agri systems looked at, and total instances of all agri systems. AND then have detailed table for each paper, which shows details of which agri systems looked at and the number of instances of each
* In Shiny app: Default: show overview table as detailed in bullet point above. And then have user input to show more detailed info (as detailed above) on specific paper chosen by user?
* So have made this table:

Table

Description automatically generated

* And then made this as an example of a detailed table of a specific paper (which the shiny app will do automatically for any selected paper)

Table

Description automatically generated

Fri 24/6/22

* Put updated versions of tables into the app, with the user choice of which paper to show full details for
  + Default – don’t show any full details of any paper to start with
  + In table legend, put “see references tab for full paper details”
* Made references tab with more details on each paper included in the app
  + Had to alter the initial spreadsheet a little because there was a couple of mistakes:
    - Yaap paper has 2 slightly different versions of the author, so made them the same
    - In finding unique combinations of paper\_number and paper\_ID, Delaquis.E appears twice. Most of them have the paper ID as 36, but a couple have it as 35. So I changed the ones that have 35 to 36 for consistency.
* In the agricultural systems tab, I included a button that the user can click to display a description of each agricultural system. There is also a hide button to hide this
  + Spent about 3 hours trying to do this. The best I could get for a long time was being able to show the table upon clicking the show button and then hiding it again by pressing the hide button. But then couldn’t re-show the table.
  + Eventually got it to work using the reactiveValues() function with actionButtons()
* Re-arranged the server.R file so the order follows the order of the tabs in the app

Sat 25/6/22

* Re-ordered agricultural systems in the paper details table in intro tab to be alphabetical
* Started looking it being able to upload data to the app. Not sure if this is actually possible. I have included an upload option in the app in a new ‘upload data’ tab but doesn’t do anything for now.
  + Used fileInput()

Mon 27/6/22

Read van Klink et al., 2020 paper in detail.

Lab meeting 10am:

* Had presentations from Victoria, Christina, and Justin in preparation for their actual presentations later in the week

Read Gillespie et al., 2022. Landscape-scale drivers of pollinator communities may depend on land-use configuration.

Just thinking about next steps – want user to be able to run own model. So for now, going to give the user the option to filter the data by the metric used e.g. biomass, diversity. And then run the model based on this filtered dataset. Then can adapt later to whatever I want e.g. location. The framework will be in place. Still robust model?

* Started on the ui bit of this
* Need to do the server bit

Tue 28/6/22

Went into NHM, arrived at 10:45am. Left at 5:45pm.

Corrected mistake in spreadsheet – biodiversity\_metric\_category was spelt wrong so changed this in the 0\_preparing\_LRR\_dataset.

Working on the server bit of running a custom model.

* Have managed to get this to work. So users can filter the data by their choice of studies to include based on which biodiversity metric category was studied. Then they can press a button to run the model. Loading symbols appear until the model is run and the table of coefficients and the figure appear.
* Stopped some errors occurring when it was trying to plot before the model had run. Used req() again
* I’ve altered the table so it has the correct column and row names
  + BUT still need to add in extra coefficients e.g. adjusted\_LRR
* Graph it produces at the moment is just so I could prove it works, it’s not the right graph yet

Wed 29/6/22

* Split agricultural systems tab into 2 tabs – one for default model, one for custom model
* Have expanded the table for the custom model. Now has adjusted\_LRR as well as things like % change and sample sizes (frequency of studies included in the model by agricultural system).
* Included the agri systems definitions in the custom tab too.
* Added table of coefficients for default model.
* If no biodiversity metric categories are selected, it shows the message “Please select at least one biodiversity metric category.”
  + But currently shows it 4 times, I want to just get it to show once?
* Trying to include correct graph for the custom model – don’t think I will bother with user inputs on choosing agricultural systems and re-scaling x-axis as too complicated because not all agricultural systems are included once you filter the data based on studies which have certain biodiversity metric categories. Could still have option to plot adjusted\_LRR or percentage\_change.
  + Custom model graph is pretty much working after from some weird error bars?
  + Also plotting conventional – want conventional in table but not graph

GLITRS meeting (Andy, Joe, Daero, Anna, Yeahji):

* Daero – started extracting data from 11 papers. Comparing areas with and without pesticide application on Odonata abundance
  + What if a future student wants to do a time-series analysis rather than a meta-analysis? Probably wouldn’t use the MA protocol in those cases.
  + Protocol needs to be flexible in some way for future students to use
    - Better format would be long-format data e.g. if comparing 2 areas, this needs to be over 2 rows (rather then the 1 is over now)
    - If only given 1 number e.g. for LRR, the other number is 1.
* Me – showed shiny app updates
  + Aws server
  + Define class for spreadsheet upload
* Yeahji – thinking about analysis process for data. Uploaded datasets to predicts.
* Anna – extracting data. Some papers don’t make their sample size clear.

Thur 30/6/22

* Fixed the issue with the custom model graph from yesterday. Now have correct graphs and tables for both default and custom models.
* Re-organised both ui and server code.
* Re-deployed app and tested – seems to work fine. Does take a little while to load the app due to it running the model.
* Sorted out table and figure numbering.
* Have added in the option for the user to download the table of coefficients produced by their custom model.
* Also need to change things around slightly so I can give the user the option to download the R model summary too. Currently don’t have this saved as an object so can’t access it.
  + Sorted this so it is possible
  + Coded it so button is not enabled until a model is run
  + Did the same for the default model so users can also download these results
  + Coefficients download is a csv which opens in excel
  + R model summary is a txt file
* Added in some extra dividing lines to improve layout
* If you run a custom model with just “NA”, it doesn’t work. This is because the NAs are actually NAs, and not “NA” as a character string. So I went back to the 0\_preparing\_LRR\_dataset script and changed the instances of NA to “unclassified”.
* Redeployed app – all working

Project update meeting with Joe:

* Uploading data
  + Probably need to try and use something like onedrive or dropbox
  + Users can then write a file which is stored in a directory in onedrive/dropbox
  + And when the model is run – read in all the files from that directory
  + Think about api keys? Put it in a key file but make sure to gitignore this file so it is not pushed to git
    - And then source the key in the server file and subsequently refer to it as an object
  + Because the datasets I have currently are quite different, and sorting out the uploads is not likely to be easy, just make some dummy data by duplicating Christina’s data for now and get it working with this
* Think about a name for the app
* Don’t worry that Christina’s data is not just based on insects – ok because shiny app is a proof of concept
  + Could think about filtering Christina’s data to just include insects?

Updated plan:

|  |  |
| --- | --- |
| Week/Date | Tasks to complete |
| End of 12 – 27/6 | * Writing – structure of results? |
| 13 – 4/7 | * Uploading data successfully using dummy data * Writing – improve intro, continue methods, start results |
| 14 – 11/7 | * Uploading new datasets (not dummy data) * Writing – finish results, start discussion |
| 15 – 18/7 | * Play with models – adding in models for different threats/variables – have different tab for each? * Finish discussion |
| 16 – 25/7 | * Have initial report complete by end of this week – including abstract and acknowledgements |
| 17 – 1/8 | * Additional tweaks e.g. map of where studies are from * Re-drafting of report * Turn into R package as well? * Transfer report to LaTeX * Sort out github |
| 18 – 8/8 | * Same as week 17 |
| 19 – 15/8 | * Same as week 17 |
| 20 – 22/8 | * Same as week 17 * Hand-in |
| After hand-in | * Presentation/viva |

Fri 1/7/22

Worked on methods section of report based on updates made to the app.

Mon 4/7/22

Lab meeting 10am:

* Updates from me since last Monday:
  + Introduction: Masters student making shiny app (we based application written in R) so users can interactively analyse and visualise meta-analysis data on insect biodiversity trends
  + Tested positive for covid over the weekend so won’t be able to come in this week – will miss statistics session and staff summer party
  + Last week:
    - Users can now run a custom model
    - User can download their results
  + This week
    - Writing and uploading data – start with dummy data

Read Shackleford et al., 2021. Dynamic meta-analysis - a method of using global evidence for local decision making in detail.

Put together structure of results section and wrote this – have about 400 words. (Doesn’t seem like a lot?)

Included t value in model output tables.

Moved upload data tab to tab before references as makes more sense to have it here.

Looking into uploading data. https://shiny.rstudio.com/articles/persistent-data-storage.html:

* Functions like write.csv() and saveRDS() save data locally, but this won’t work for a shiny app. App is distributed across different servers, which means that if a file is saved during one session on some server, then loading the app again later will probably direct you to a different server where the previously saved file doesn’t exist.
* Local storage is generally faster than remote storage, but it should only be used if you always have access to the machine that saves the files.
* Remote storage means saving data on another server, usually a reliable hosted server such as Dropbox, Amazon, or a hosted database.
* If using OneDrive: use R package Microsoft365R: Enables access from R to data stored in 'Teams', 'SharePoint Online' and 'OneDrive', including the ability to list drive folder contents, upload and download files, send messages, and retrieve data lists.
* If using dropbox: you can use the [rdrop2](https://github.com/karthik/rdrop2) package to interact with Dropbox from R.
* Currently looks like most help available for and most viable option would be dropbox
* Dropbox:
  + rdrop2 package
  + Note that rdrop2 can only move existing files onto Dropbox, so we still need to create a local file before storing it on Dropbox.
  + Each response is its own file, different responses can have different fields, so it's easy to change the form to have different fields (though doing that will make aggregating all the responses more tricky).
  + You will also need to add authentication to rdrop2 with any approach [suggested in the package README](https://github.com/karthik/rdrop2#accessing-dropbox-on-shiny-and-remote-servers). The authentication approach I chose was to authenticate manually once and to copy the resulting .httr-oauth file that gets created into the Shiny app’s folder.
    - Do this – but remember to git ignore this file so it is not pushed to git
    - Think save token to rds file – see below
    - Graphical user interface, text

      Description automatically generatedUse drop\_auth function from r2drop – use in server function – read in RDS file where it is stored
    - drop\_download Download a file from Dropbox to disk.
    - drop\_get Downloads a file from Dropbox
      * use dtoken - The Dropbox token generated by drop\_auth
    - drop\_upload Uploads a file to Dropbox
      * Again use dtoken
      * Path = dropbox folder name where data is stored
    - drop\_dir List folder contents and associated metadata.
* When saving multiple files, it is important to save each file with a different file name to avoid overwriting files. There are many ways to do this. For example, you can simply use the current timestamp and an *md5 hash* of the data being saved as the file name to ensure that no two form submissions have the same file name. Use sys.time()
  + Get user to input their names and somehow include this in the file name?
* Because have to save file locally first: save to temporary directory using tempdir(). When you start a new R session a temporary directory is created. Location of the directory depends on your system and configuration. Function tempdir() returns path to the current temporary directory.

Process followed:

* Installed rdrop2 into R and loaded.
* Ran drop\_acc() which opened web browser and accepted to let rdrop2 access my dropbox. Then closed and returned to R, which showed my account info.
* Ran token <- drop\_auth(), then saveRDS(token, "dropbox\_authentication\_token.RDS") to save into MMA folder. Made .gitignore and included this file in it
* Firstly, just see if I can get a file uploaded from the shiny app to dropbox
* Then can think about next steps –
  + If uploaded new data, will have to re-run analysis on all available data. Will have to be done in the server app – make a button which the user can click to update the analysis.
  + This way, could leave global.R file alone? BUT doesn’t fix the potential drawback that the app is slow to start-up.
* Made some dummy data, which I will try to upload to dropbox from shiny app
* In global.R file, added a line which reads in the dropbox authentication token
* Included rdrop2 package in both global and server
* Have managed to get it so I can upload a file from the shiny app to dropbox. Can repeat for a 2nd file etc.
* Lots of challenges with it not uploading to start with, but also not giving me any errors
* Problem with it just saving the local data path rather than the actual data but fixed this
* Have written function in global.R which is then used in server.R to upload the file

Tue 5/7/22

Moved all packages that need to be loaded into the global.R file.

Thinking about changing the app so it doesn’t take so long to start up, and can react immediately when new data are added.

* Think I want it so that the default model is not run in the global.R file because when the user starts up the app, it takes a long time to load.
* Instead, think I want the app to open straight away and then the user can press a button to run the default model within the app
* The issue is that the ui.R file uses the object produced after running the model in the global.R file for certain things e.g. the selections for which agricultural systems to plot
* Could get the names for this kind of thing by finding unique agricultural systems from Christina’s dataset (without having to run the model)
* BUT if new data are added and these have additional agricultural systems, these won’t be visible to the ui.R file and won’t be available as an option for the user to select.
* SO the global.R file might end up a bit empty, but that is fine. It can include:
  + Loading packages
  + Loading authentication token for dropbox
  + Function to save new data to dropbox
  + Loading agricultural systems definitions
* No point loading the data in global.R because we want it to be a reactive object – want it to update every time new data is added.
* There is a thing called uiOutput() and renderUI(), which I think will get over my problem of the ui not being reactive to changes in the data (see below)

From <https://mastering-shiny.org/action-dynamic.html>:

* The user interface is defined statically when the app is launched so it can’t respond to anything that happens in the app. In this chapter, you’ll learn how to create dynamic user interfaces, changing the UI using code run in the server function.
* uiOutput() and renderUI() allow you to create and modify the user interface while the app is running
* uiOutput() inserts a placeholder in your ui. This leaves a “hole” that your server code can later fill in.
* renderUI() is called within server() to fill in the placeholder with dynamically generated UI.
* Might take a fraction of a second to appear after the app loads. That’s because it’s reactive: the app must load, trigger a reactive event, which calls the server function, yielding HTML to insert into the page. This is one of the downsides of renderUI(); relying on it too much can create a laggy UI.

I have made reactive ui outputs for all outputs that needed to be reactive. SO, for e.g., if new data with new agricultural systems are added, then this will now be included as an option for the user to select.

Think I have successfully removed all reliance of the ui.R file on the global.R file running a model at the start-up of the app. So now should be able to run the default model in the server.R file once the user clicks the button for it to run this model.

Change app so data is loaded within server file (and is reactive). And then have added a button for the user to click to run default model. Had to add () into the reactive ui output bits i.e. data(). Remove the running from global.R. App is also now much faster to start up.

Tried to check whether uploading data was still possible.

* Unfortunately, it was not because apparently the access tokens expire after 4 hours!!!!
* For now, I made a new one using token <- drop\_auth(new\_user = TRUE) which flushes out the expired token. Then did saveRDS(token, "dropbox\_authentication\_token.RDS") to overwrite token file. I could successfully upload data again. BUT this is not a long-term solution!
* Read a lot and I can’t find a way to fix this currently
* Think I may have to switch to googlesheets – see: https://debruine.github.io/shinyintro/data.html

Pushed changes to github, but don’t want to re-deploy app until fixed issue with uploading data.

Wed 6/7/22

Set up uploading data with googlesheets, rather than dropbox. https://debruine.github.io/shinyintro/data.html:

* googlesheets4 package provided by tidyverse
* Installed and loaded googlesheets4 into R
* Authorisation for Apps
  + First, you need to get a token and store it in a cache folder in your app directory.
  + We're going to call that directory .secrets.
  + Run the following code in your console (NOT in an Rmd file). This will open up a web browser window and prompt you to choose your Google account and authorise "Tidyverse API Packages".
  + With shiny app folder as working directory, run (in console): gs4\_auth(email = "my\_email", cache = ".secrets")
  + This took me to googlesheets online to confirm the link
* Added secrets folder to gitignore
* Now, you can include the following code at the top of your app.R script to authorise the app to read from and write to your files: gs4\_auth(cache = ".secrets", email = "my\_email").
  + I put this in global.R file
  + Am I not giving out my email by doing this?
  + Could I put email in another folder and not push this to git either, and then app could read it in?
  + Do end up doing this eventually – see later on
* Try different way – create google service account? – Too complicated and couldn’t work out how to do it https://robocorp.com/docs/development-guide/google-sheets/interacting-with-google-sheets
  + Go to <https://console.cloud.google.com/> and make new project (called Shiny app MMA Grace Skinner)
  + Click on ENABLE APIS AND SERVICES.
  + Find and enable the Google Sheet API
  + Create new credentials to the Google Sheets API. Select Other UI from the dropdown and select Application Data. Then click on the What credentials do I need? button.
  + Created API key
  + Created service account and downloaded .json key – saved this
  + Got lost and gave up on this method

Go back to original way:

* Put email in email.txt, in email folder
* Can read this into R using as.character(read.table("email/email.txt")) – this goes in the global.R file
* Put email in gitignore
* Folders are not very good in gogglesheets. Guess I will have to start the name of every file upload with the same thing so when I load in all the files, I can use same\_name\_\*.csv
* OR use one sheet with multiple tabs in it: If you don't specify the tab name(s), the sheet will be created with one tab called "Sheet1". I recommend making only one sheet per app and saving each table in a separate tab.
  + # Set up initial sheet - put christina's data here? This will be tab1
  + id <- gs4\_create("test\_sheet")
  + # Add second tab
  + sheet\_add(ss = id, sheet = "tab2")
* Have created the initial spreadsheet with Christina’s data in the 1st tab using this script: save\_christina\_data\_to\_googlesheet.R

Have managed to code the shiny app so you can now successfully upload a file to googlesheets. You have to put your first and second name in first, and then the browse button is enabled so you can choose a file to upload. The new file is named with the first and second name inputted and a time stamp. It gets added as a new tab within the shiny\_app\_data gogglesheet.

Am hopeful this time that I won’t have an issue with the token expiring – sometimes get the message “Auto-refreshing stale OAuth token.” when start-up app, which seems a good sign.

Have put the sheet ID in a .txt file, and also now put the email.txt file in a details\_for\_googlesheets folder. These will be in the gitignore so won’t get pushed to git.

Next need to load data from gogglesheets, rather than data folder.

* Practise from testing.R first – managed to load all sheets and rbind them together
* Moved this code into shiny app code, and seems to be working pretty well
* Have tried it with just Christina’s data and the models run fine
* Then also tried in with Christina’s initial data and 4 lots of dummy data – few issues but overall not bad:
  + In references, some papers are repeated? – Due to slight differences in names by some having carriage return \r and others having newline \n. Not sure where each is coming from? Filter by paper\_ID so just shows 1 of each?
    - Graphical user interface, text

      Description automatically generatedFixed this by altering code that makes this table slightly
  + In the default and custom models, the conventional reference level is not adjusting to 0 correctly.
    - Because it is taking away 0.02 no matter what, but this won’t always be the case when data is added
    - So fixed it by changing it to - robust\_coefficients\_all$Original\_LRR[1] rather than -0.02
    - Ran app and seems to of fixed it

At the moment, the app does not include the new data in the tables or figures even if the models are run again. You have to load up the app again to do so. Not sure why this is because the data is meant to be reactive?

* Maybe will just have to be that you have to restart the app to be able to include the new data in the analysis.
* Look into this tomorrow

Made a few changes before git pushing:

* Tidied up the ui, server, and global files because they had lots of versions of the code at different stages. Did make a back-up before I did this.
* Removed the dropbox authentication things from the app folder.
* Updated the gitignore

GLITRS meeting (Joe, Anna, Daero, Yeahji):

* Update from me since last week:
  + Written some of results section
  + Working on uploading data - First attempted with dropbox, after spending whole day working on it on Monday, it seemed to be working absolutely fine
  + Tuesday – I didn’t like the fact that the app took ages to start up when opened, so changed whole app around so it doesn’t have to run a model upon start-up so it starts-up immediately. Once open, the user can press a button to run the default model if they wish.
    - Had to use some reactive UI outputs
  + Then tried to check if uploading data was still working and it was not. Because the dropbox tokens only last 4 hours.
  + Couldn’t find a way around this so today have tried to switch to using googlesheets instead, which I have worked on today and currently seems to be working
  + Give demonstration of this
  + Next I need to load in the data from the googlesheet, rather than the data folder of the shiny app
* Anna:
  + Got data for 1st intervention, though 2nd one appears slightly more difficult
* Daero and Yeahji – been unwell so not much to update on

Summary for Wednesday – linked shiny app to googlesheets, can upload data from the shiny app to the googlesheet, and the app loads in the data from the googlesheet too. Hopefully still works tomorrow.

Thur 7/7/22

Can I get the app to automatically update/refresh when new data is added? Or will the user have to refresh the app to see the changes?

* Can’t seem to find a solution to this
* The data is coded to be reactive but is not actually updating when new data is added
* The user has to refresh the page to see outputs which include their uploaded data
* Not sure this is too big of an issue?
* Have added a message to the upload data tab which says please refresh the app once you have uploaded data

Using the “maps” package, I included a map with points of where data comes from in the intro tab. Unfortunately only 58 of 676 datapoints have coordinates (say this in app). But good proof-of-concept.

Meeting with Joe 2pm:

* Discuss why app doesn’t automatically update when new data added to googlesheets – not sure why it’s not working. Think it’s fine for now based on how many people will be using the app and how many will be uploading data etc.
* For the models that currently are not working due to lack of data – use tryCatch()?
* Most important thing for now is to get the checks on the file uploads so can re-deploy app without fear of misuse.
  + Make sure file is a csv
  + Make sure it has certain columns
  + Make sure name that people input only contain letters (no %,$,& etc.)
  + Have a look at php/code injections – stop code going into input
* For methods:
  + Include sub-sections/headings e.g. philosophy behind app, then more specific things
  + Start more generally – e.e. why shiny? What it is? Easy way to combine analysis and visualisation. Easy to deploy.

Messaged Christina about how she would describe adjusted log response ratio.

* So on it’s own the conventional result of 0.02 (or whatever is it) doesn’t really mean anything. It has to be compared to the other agricultural systems to make any sense. E.g. if conservation has LRR of 0.04, then you can conclude this has higher biodiversity than conventional.
* Christina then adjusted the conventional to 0, because it makes more sense for interpretation and visualisation. The difference between 0.04 and 0.02 is 0.02, so conventional might as well be 0, and conservation 0.02. It is the difference that is important.
* Conventional agriculture is highly impactful on BD, and usually has low levels of BD, so Christina was trying to see which systems are less impactful than this

Some models actually just don’t work because not enough data, so would be good to add an error message if the model fails. Tells the user they need to add more data in, though this model may work in future if more data is available.

Text

Description automatically generated

* Efficiency does not run at all (think all 0 – not quite)
* Enzymatic activity – no conventional – model runs but output is not correct
* Reproduction – starts but then doesn’t work (not enough data?)
* Survival - starts but then doesn’t work (not enough data?)
* Make sure conventional is in the dataframe (fix enzymatic activity)
  + Sorted this out using a validate(need()) catch before attempting to
* Make sure not all 0 (fix efficiency)
  + Unfortunately they are not quite all 0 so this won’t work
  + So this will have to be caught through the model not running
* Give user message if model returns an error (fix reproduction and survival [+ now efficiency])

|  |  |
| --- | --- |
| Week/Date | Tasks to complete |
| 14 – 11/7 | * Checks on uploading new datasets * Other tweaks/things to add in * Writing – start discussion, improve intro |
| 15 – 18/7 | * Play with models – adding in models for different threats/variables – have different tab for each? * Finish discussion, improve/finish all other sections |
| 16 – 25/7 | * Have initial report complete by end of this week – including abstract and acknowledgements |
| 17 – 1/8 | * Additional tweaks * Re-drafting of report * Turn into R package as well? * Transfer report to LaTeX * Sort out github – move app to own public repo, add README |
| 18 – 8/8 | * Same as week 17 |
| 19 – 15/8 | * Same as week 17 |
| 20 – 22/8 | * Same as week 17 * Hand-in |
| After hand-in | * Presentation/viva |

Next:

Need to get checks on data uploads in place before I redeploy the app. Then want to test with someone else that they can upload data.

Mon 11/7/22

Lab meeting:

* Staff sale tomorrow and Wednesday in the museum
* Scite.ai – see connected papers – see which papers cite the paper you are looking at
* Update on progress from last week:
  + Introduction: Masters student making shiny app (web based application written in R) so users can interactively analyse and visualise meta-analysis data on insect biodiversity trends
  + Started on results section
  + Changes to shiny app
    - App starts up straight away as soon as user opens it. They can click a button to run a default model if they wish.
    - Making it possible for the user to upload their own data. Gets stored remotely in googlesheets.
  + This week need to put some checks in place on the data that can be uploaded to make sure it’s the right kind of data

Working on showing the user error messages when the model doesn’t run:

* Used tryCatch to try and run the model. If doesn’t work, it stops and produces an error message
* I have got these error messages to be grey (rather than red) by using tags$head(tags$style(".shiny-output-error{color: grey;}")) in the custom model section of the ui script
* Displays the error message twice – one for graph and one for table, which is a bit annoying but taking too long to sort so will leave for now.

Tue 12/7/22

Started to read Hallmann, 2020 paper in detail

Checks on data uploads:

* Maybe try something along these lines:

Graphical user interface, text, application, email

Description automatically generated

* Make sure file is a csv

Graphical user interface, text

Description automatically generated

* Make sure it has certain columns
* Make sure name that people input only contain letters (no %,$,& etc.)
  + input$first\_name
* Check not already uploaded – don’t want duplicates
* Have a look at php/code injections
  + PHP Object Injection is an application level vulnerability that could allow an attacker to perform different kinds of malicious attacks, such as Code Injection
  + Code injection vulnerabilities exist when an attacker can submit executable input to a program and trick the software into running that input. This gives the attacker a channel by which they can circumvent any security restrictions put in place by the program’s author.

Graphical user interface, text, application

Description automatically generated

Have put checks in place on the names inputted. The browse button to upload a file is disabled unless first and second name is not empty and only contains letters. The user is given feedback if they input an unacceptable name.

If name is acceptable, the upload button is enabled and the user can chose a file to upload. Upon meeting some checks (currently whether the file is a csv), the user is shown a preview of their data.

The upload to googlesheets button is now enabled and the user can click this. If uploaded successfully, a completion message appears. (This message disappears if the user changes the first or second name input or file input).

Have added in a check on the data to make sure it includes LRR and Treatment columns – can change this later if necessary.

Started trying to add in a check on the data to make sure it is not a duplicate of one that already exists in the googlesheets.

* Trying with identical() and all.equal() functions

Wed 13/7/22

Went into NHM, arrived at 10:30am.

Working on putting the final check in place to make sure the data is not a duplicate of data that already exists.

* Think I have mostly sorted out this check
* Fixed a lot of little things which ultimately meant that the wrong data or wrong inputted name could still be submitted.
* Think all checks work pretty much fine now, a bit slow in cases.

Used shinydisconnect package to display a message if the whole app fails / disconnects from the server. Added this as the header argument in navbarPage so it applies to all tabs. Seems to work – if I run the shiny app, and then stop R running from RStudio, an error message appears in the online app.

Fixed an issue with the downloading results button in the custom models tab. I’ve coded it so that you can’t click the download buttons unless a successful model has been run. So now if the model fails, the user can’t click these buttons. Before, they still could and it end up failing to download when they did.

Continued reading Hallmann 2020 paper in detail.

Thur 14/7/22

Careers appointment 10am.

GLITRS meeting (Joe, Daero, Anna, Yeahji):

* Updates from me:
  + Have added in custom error message for when a model doesn’t run – uses tryCatch()
  + Have added in the checks on data that can be uploaded
    - Name only contains letters
    - File is a csv, contain the columns Treatment and LRR, is not a duplicate of one that already exists
    - Had to do lots of little fixes to stop the wrong thing being uploaded even with these checks – e.g. could put an acceptable name in and upload the file to the app, but then could change the name and upload it to googlesheets
  + Fixed a few other little things
    - e.g. not enabling the download results button unless the model ran successfully
    - e.g. a message appears if the whole app fails / disconnects from the server
  + Next – couple of other things to sort, and then need to redeploy the app and test it still works. And then need to put some effort into writing because been focusing on the coding recently.
* Anna – editing drafts of these. Continued with data extraction for crop rotation – not sure if there is enough data though. There is enough data for the other things - inter-cropping and margins.
* Daero – working on analysis and writing results.

Mon 18/7/22

Lab meeting:

* Update from me
  + Putting checks in place on the data that users can upload
  + Few other little bits that have been annoying me
  + Undergrad graduation
  + This week – focus more on the writing, job interview
* Tips for interview
  + Don’t worry if mess up a question up – don’t let it affect subsequent questions
  + Have 90 second summary of project ready – what did, why matters
  + Take your time – say ‘let me think about that’
  + Do ‘Power posture’ beforehand
  + Don’t be afraid to ask for clarification on questions – are you asking me … ?
  + If get question you don’t know the answer to, say something like ‘I don’t really know but I would go about it like …’
  + If get question on specific thing you don’t know about/have experience with, use another experience that relates to it – shows you can see how things relate
  + Afterwards, write down what went well what didn’t, lessons learned etc. so can know for next time

Re-deployed app and tested.

* Not giving correct error message when model doesn’t run?
  + Trying to use safeError() to see if that will work

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

* Upload to googlesheets button was still greyed out once data has been uploaded to shiny app the 1st time I tried after re-deploying – not sure why. But tried again and been fine since

Disabled download default model buttons until model has run – so can’t get an error if user clicks it too early.

Tried to re-deploy app but wouldn’t let me – think it did it too recently. Waited an hour and tried again and it worked. And the safeError fixed the issue of the wrong error message displaying.

Finished reading Hallmann 2020 paper in detail.

Read Outhwaite, 2020 paper in detail.

Tue 18/7/22

Job interview prep

Wed 19/7/22

Job interview prep

Thur 20/7/22

Job interview prep

Job interview – technical and personal for Bioscience Data Scientist position at Syngenta

Fri 21/7/22

Caught-up on emails etc., refreshed what I’ve been doing and what I need to do next.

Mon 25/7/22

Lab meeting 10am

* Update from me since last week:
  + Mainly preparing for and doing job interview

Meeting with Joe and Christina 1:30pm

* App currently based on studies that are MAs themselves
* New MA data will be based on studies that are primary studies
* How are they comparable?
* Christina has almost raw data that is fed into the robust model
* New MA data will have effect sizes and variances and then these would need to be used in a metafor model. BUT Christina’s data couldn’t be used like this as don’t have the mean and variance values needed to calculate effect size
* BUT could the new studies be analysed in the same way as Christina has done?
  + E.g. Hanyu – think data is in wide as has control and treatment in separate columns?
* What if multiple MAs use the same study? Don’t want to over-inflate influence of this
  + Discard duplicates
* How easy to filter out insects from Christina’s data?
  + Column for class so could do this quite easily
* For new MAs
  + Upload 1 row – overall LRR for the whole MA
  + Or upload whole lot – LRR for each study, then have extra code at start to do some calculations

1. Metafor needs wide format, app needs long format.
2. Send hanyu’s to joe

Started writing discussion – did about 400 words.

Tue 26/7/22

Bugs, birds, and beasts day – 10:30 – 5:00

* Volunteered on the ‘bug hunt’ activity where we helped children to use sweep nets to catch insects and spiders, and then tipped them onto a white sheet so the children could hold them and try to identify them
* Helped on the seed ball stall where we got children to make their own seedballs using seeds, soil, clay, and water

Wrote another 500 words of discussion – have 900 in total now.

Wed 27/7/22

Editing introduction, condensing section on literature review of insect population trend studies.

GLITRS meeting (Andy, Joe, Daero, Yeahji):

* Daero – protocol format, next going to start thinking about discussion
* Yeahji – introduction and methods, getting code ready for analysis on data already have. And then if can add more before the deadline, can add these in if time.
* Me:
  + If something doesn’t fit in intro, might go into discussion – if don’t follow down potential line, could be future work
    - Consider moving rather than cutting
  + Look at marking criteria?
  + Think about the fact that the app isn’t just relevant to insect declines – other fields also need this tool to synthesise literature
  + Structure: start with the general issue of needing a tool to synthesise literature, and use insect trends as an example for how the tool could be used. OR say insect declines need a tool, but the tool developed could be applied more generally.

Thur 28/7/22

Expanding last section of intro on the shiny app side. Sent this new version to Joe and Andy.

Working on the methods section, including adding sub-headings.

Fri 29/7/22

Went into NHM, arrived at 10:30am, left at 6:15pm.

Continued working on methods – sent to Joe and Andy.

Worked on results – is now 800 words.

Crystal Palace Dinosaurs Sci-Fri talk 4:30-5:30

Sat 30/7/22

Continued working on results – sent to Joe and Andy.

Continued working on discussion – thought of quite a few bits to add and started writing some of these bits. Other sections still need writing and I think the organisation/structure needs work.

Mon 1/8/22

Lab meeting 10am:

* Chat with Joe and Christina about adding in new data to the Shiny app
* Volunteered on the bugs, birds, and beasts day at Silwood – did bug hunt
* Intro, methods, results – got good draft for each of those.
* Will finish off discussion in the next couple of days – then will send all over together
* Crystal palace dinosaurs Sci-Fry talk
* Things I can tweak to the app in terms of how it looks, README, latex
  + Sweave – useful for writing anything where lots of numbers will change e.g. a database methods paper
    - Sweave is a function in the statistical programming language R that enables integration of R code into LaTeX documents. The purpose is "to create dynamic reports, which can be updated automatically if data or analysis change"
* Probs need meeting with Joe at some point this week
* Do mini-presentation in coding club on Friday 2-4pm
  + Tell narrative of thesis
  + Makes few slides with main figures
  + Couple of mins on main points of narrative

Reading for discussion:

Robustlmm handbook:

* robustlmm provides functions for estimating linear mixed effects models in a robust way. The main workhorse is the function rlmer; it is implemented as direct robust analogue of the popular lmer function of the lme4 package.

<https://cran.r-project.org/web/packages/robustlmm/vignettes/rlmer.pdf>:

* As any real-life data, data modelled by linear mixed-effects models often contain outliers or other contamination. Even little contamination can drive the classic estimates far away from what they would be without the contamination. At the same time, datasets that require mixed-effects modelling are often complex and large. This makes it difficult to spot contamination. Robust estimation methods aim to solve both problems: to provide estimates where contamination has only little influence and to detect and flag contamination.
* Fixed terms are added in the usual R formula notation, and random terms are specified in parentheses. Random effects are defined in conjunction with a grouping factor. The grouping factor is separated from the random effect by a pipe symbol “|”

Linear mixed effects models:

* Mixed effect models include fixed and random effects. So we could look at the effect of fixed effect agricultural system on response variable LRR while controlling for the random effect location.
* Fixed effects = explanatory variables.
* Random effects = usually grouping factors for which we are trying to control – categorical.

Meta-analyses fixed, random, mixed effects models:

* Fixed effect = assuming all studies share a common true effect and only one source of variation is present: the within-study (sampling) variance. Get an overall effect size.
* Random effect = also accounts for between-study variance, termed heterogeneity. Better for ecological studies. Get an overall effect size but the effect size of a particular study could lie anywhere between the confidence intervals.
* To identify causes of variation, it is common to subsequently run mixed-effects models (or meta-regression) involving moderators
* The moderators will ‘mop-up’ some of the unexplained variance – i.e. the reason variance exists is not random, but due to these moderators. How do the moderators influence the effect size?

Sub-group analysis / meta-regression - <https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/subgroup.html>

* My app is a sort of like sub-group analysis but full sub-group analysis is run by running a meta-analysis which does an individual meta-analysis on each subset. So in my Shiny app, we would say the subsets are defined by biodiversity metric, and then you would get a result for each of these.
  + BUT mine could still be argued as sub-group – as dynamic meta-analysis does
* Mixed effects model / meta-regression model accounts for the fact that observed studies deviate from the true overall effect due to sampling error and between-study heterogeneity. More importantly, however, it also uses one or more variables xx to predict differences in the true effect sizes
* Multiple meta-regression, for example, allows us to include not only one, but several predictor variables
* rma function in {metafor}. This function runs a random-effects meta-analysis, which is extended to mixed-effects meta-regression models when moderators are added.
* On top of all this, can then have multi-variate or multi-level models – you have to define the (nested) random effects – there are different levels of hierarchical clustering and you have to define the structure of these

Signed up to Metadataset to trial it. Was going to replicate the example outlined in the paper but can’t actually do it. App I used is different – likely a prototype with less functionality so not actually very helpful.

Go through metadataset video with Joe?

* Seems like when user selects a filter e.g. US as location, the meta-regression included location as a moderator, and the model output includes estimates for all locations. And then I would assume they just plot the US estimate? But the numbers don’t match? Think this is because location is not the only moderator. And I can’t see all the model output.
* For sub-group analysis, number do match

Include crop and paper ID as random effects in model but don’t ever mention these again? And don’t seem to have much effect either. BUT needed to account for non-independence of the data?

A screenshot of a computer

Description automatically generated with medium confidence

Tue 2/8/22

Went into museum, arrived at 10:30am, left at 5:45pm.

Worked on discussion.

Response Ratio = Ratio of means

Useful book - https://bookdown.org/robcrystalornelas/meta-analysis\_of\_ecological\_data/#motivation-for-this-book

Define meta-meta-analysis – combining the data collected for multiple individual meta-analyses and analysing it altogether.

The metafor model is based on studies which all investigate the same thing in the 1st place. So then meta-regression investigates how other variables influence this intervention.

We cannot combine overall effect sizes from individual meta-analyses that investigated different interventions

Why use relational database over just using Excel? Database vs spreadsheet:

* A relational database run by an RDBMS usually stores data much more than regular spreadsheets such as Excel can process at ease. At its full capacity, Excel can store about 1 million rows and 16,000 columns of data. Those numbers are relatively small compared to what a database system can store.
* We can indeed store up to about 1 million rows of data in an Excel sheet. We can still scale up by using multiple sheets. However, doing so will cause Excel’s data processing speed to drop, and looking up for any single rerecord of data will become a time-consuming task. On the other hand, an RDBMS can process millions of records of data without losing speed.
* Multiple users can access and work on a single database simultaneously over a network or on the Internet. Although spreadsheets are originally designed for a single user, an updated version of a spreadsheet such as Google Sheets can also provide multi-user functionality. For a database, multiple users can update or manipulate the data at the same time, but the RDBMS will carry out only a single process at any point in time. For a spreadsheet like Google Sheets, however, a multi-user session can be confusing. It can also cause data integrity loss because data can be deleted, overwritten, or changed simultaneously in unintended ways due to human errors. This error type will not happen in an RDBMS because the data input type is predefined during the database design process. Human error is eliminated because, in a database, each record in a table will not accept data of a type it is not assigned to keep. Therefore, a database can preserve data integrity better than any usual spreadsheet software.
* A user should aim for an RDBMS if he needs to store a large amount of data, preserve its integrity, and process it quickly.
* For the easy retrieval and updating of data, efficiency, data consistency, data integrity, speed, and security, relational databases are definitely the structure to opt for.

What are the disadvantages of using Google Sheets as a database?

* Limited functionality. Spreadsheet apps aren’t designed to be used as databases, so compared to more sophisticated database management systems, Google Sheets’ functionalities are limited. For example, you can’t store videos or images, and there’s no built-in interface for querying or updating the information.
* Limited storage. Google Sheets’ limit on data volume makes the platform unsuitable for large teams and companies looking to scale. The platform has a limit of 10 million cells.
* Not Scalable. Because of its limited storage capacity, Google Sheets isn’t scalable. Additionally, if many team members are working on the spreadsheet at the same time, the software slows down and you might get more errors.
* Security concerns. As opposed to a dedicated system that requires user authentication and designated permissions, files on Google Sheets could be shared with anyone via email. And while Google does have measures in place to protect your data, if your business email is hacked, your data can easily be leaked or tampered with.
* Performance. Since Google Sheets wasn’t designed to store large amounts of data, it’s not really possible to design a complex database. The platform gets slower and slower the bigger the dataset, which is annoying for teams of all sizes.
* Prone to chaos. Speaking of large amounts of data, when your data volume grows, you might find yourself scrolling through hundreds, or thousands, of rows and columns. It’s not that intuitive to make such huge spreadsheets readable or digestible.

Why Shiny app is slow when try to open 2 connections at once:

* R is a single threaded application which means that a Shiny application cannot serve two different users at precisely the same time.
* As your applications get more complex, requiring more time to service a single request, and as more users interact with the application simultaneously, you may find that the user experience for your applications does not meet your expectations.
* A running application will have at least one Application Instance. You can add additional instances if the application is hosted on a paid tier. SO I can only have 1.
* With an application instances, there are workers, which can be connected to by multiple users
* To overcome issue of it being slow when multiple users are connected to 1 worker, consider lowering the maximum number of connections per worker, and possibly increasing the maximum number of workers.

Sent complete draft with intro, methods, results, and discussion to Joe and Andy.

Wed 3/8/22

Things to discuss with Joe on Friday:

1. Go through metadataset video
   1. Seems like when user selects a filter e.g. US as location, the meta-regression included location as a moderator, and the model output includes estimates for all locations. And then I would assume they just plot the US estimate? But the numbers don’t match? Think this is because location is not the only moderator. And I can’t see all the model output.
   2. For sub-group analysis, number do match
2. Include crop and paper ID as random effects in model but don’t ever mention these again? And don’t seem to have much effect either. BUT needed to account for non-independence of the data?
3. Shiny app is slow when 2 connections are open. – have a look at shinyproxy
4. Been thinking a lot about the future of the app
   1. Basically the input into metafor is effect sizes, but these effect sizes must all be comparable i.e. they must be comparing the same intervention.
   2. We cannot combine overall effect sizes from individual meta-analyses that investigated different interventions using metafor.
   3. Dynamic meta-analysis is able to use metafor because it has loads of studies, and then these are filtered to ones that just study a particular intervention and a meta-analysis can be conducted on these. We don’t have enough data to do this.
   4. Trying to work out how different MAs are comparable using robust models
      1. If one investigates the effect of agricultural systems (all not invaded with invasive species):
         1. Difference in LRR is due to the agricultural system

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Treatment | Treatment type | LRR | Control | Control type |
| Control | Conventional | 0.00 | Not Invaded | Invasive species |
| Treatment | Conservation | 0.43 | Not Invaded | Invasive species |

* + 1. And another investigates the effect of invasive species (all in conventional):
       1. Difference in LRR is due to the invasive species

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Treatment | Treatment type | LRR | Control | Control type |
| Control | Not invaded | 0.00 | Conventional | Agricultural system |
| Treatment | Invaded | 0.30 | Conventional | Agricultural system |
| Control | Not invaded | 0.00 | North America | Location |
| Treatment | Invaded | 0.30 | North America | Location |

* + 1. BUT then Christina’s papers do not all look at agricultural systems so maybe fine?
  1. MA compared to MMA
     1. MA = Input effect sizes from individual studies investigating similar intervention into metafor to get overall effect size for that intervention
     2. MMA 1: Do multiple MAs to get lots of overall effect sizes and do MA of these effect sizes from MAs. But would have to study similar things, and there would likely be lots of overlap in papers included
     3. MMA 2: Just take all the unique effect sizes from individual studies from all the MAs and do big MA on all these
  2. Robust models are quite a different way of doing it – allow you to include data from papers that investigate different interventions

1. Thesis feedback
   1. Example in results for the custom model is ok, and if not, do they have any suggestions?
   2. Is what I’ve written in the methods about the data not just being insect data ok?
2. Do I need to reference all the papers used in the app?

Wrote out basic structure of app and how it is put together to help explain it to someone else. (structure\_of\_shiny\_app)

Made README for whole project and one specifically for the Shiny app

* They are both currently in my master’s project repository (private)
* Eventually think I want to make 2 new repos (both public)
  + One with most of the stuff done for my project (but not data files and authentication info that can’t be public)
  + One for the Shiny app alone

GLITRS meeting 3pm (Joe, Daero, Yeahji):

* Update from me since last week:
  + Mainly been working on write-up which I have sent over
  + Been thinking about adding new data into app and how it would all work – but can discuss with Joe on Friday
  + Would be good to restructure methods so it’s less just describing what the app does and how it’s structured. Joe will link paper in comments that will be helpful for this.

Presentation for Friday – made good start on putting this together

* Tell narrative of thesis in couple of minutes
* Make couple of slides on background/intro, couple of slides on results, and a slide on interpretations/discussion
* Think I should include screenshot of the upload data tab in thesis?

Thur 4/8/22

Finish making presentation and practised it.

Got report feedback from Joe.

Started working on improving the design/visuals of the app. And adding more info to the intro.

Fri 5/8/22

Went into NHM. Arrived at 11 due to tube delays. Left at 6.

Meeting with Joe 11am:

* Discussed re-structuring of thesis, especially methods and results, to focus it more around the intentions of the app, rather than being descriptive. Include bits on effectiveness of the app e.g. how long it takes to run a model, how long it takes to upload file. Does it change if multiple users are using the app?
* Discussed next steps for the app and how to upload new data to it

Presentation on thesis so far and listened to presentations from others (2-4pm):

* To do presentation:
  + Settings – display – multiple display – extend these displays
  + Have powerpoint ready to go on laptop
  + Join teams call on laptop
  + Share monitor screen, click laptop screen, then press f5 to start presenting
* Daero, Sanisha, me, and Jonathon presented
* Asked about why robust models were used
* Asked about checks put on new data uploads
* Asked about why I think relational database would be better than googlesheets spreadsheet
* Good for methods and results to include comparison to other tools that do similar kind of thing. Find additional ones to dynamic MA (even if in other fields such as climate). Compare the features
* Could this kind of tool be useful for other synthesis projects e.g. Conservation Evidence which works around expert opinion, rather than data?
* Asked about the audience for the app. Who will use it?
* Asked about if there is an equivalent way to assess publication bias with this approach, as metafor does will funnel plots
* Could get some people to test it out and score it on different criteria – ask Pippa Gittings about this?
* At end of presentation, have a few slides prepared on things that you think people will ask
* Include videos of the app in use (rather than screenshots). Probs don’t do live demo as reliant on technical things going right, and will take up too much time of the presentation.

Continue working on design of shiny app.

Sat 6/8/22

Sorted out git as not working yesterday

* Got error when tried to push saying I needed to pull before I could push because changes has been made elsewhere
* Worked out this was because I made changes to the README online
* Did back-ups of everything
* Did git pull and it seems to be fine – doesn’t look like I’ve lost anything.
* Then was able to git add, commit, push
* Made few further changes and was able to push these too so think all fine
* Found something called git stash which may be worth looking into if this happens again

Finished working on the design of the app.

* Made everything consistent between and within tabs
* Made buttons bigger
* Made tables bigger
* Made names of agricultural systems in the figure bigger
* Change custom model figure and table legends

Changed naming of some things

* Changed t to t-value
* Changed biodiversity metric category to biodiversity metric
* Where possible, changed LRR to log response variable – wouldn’t be so easy to do when the words are taken from the data

Re-deployed app

Sun 7/8/22:

Think about moving stuff into new git repositories and how it would work

* Make 2 new github repos – both public
  + One for Shiny app
    - Remember to gitignore authentication bits
    - Won’t need Christina’s data in data folder
    - Change ‘link to code’ in shiny app code
    - Take README from whole project README
  + One for whole project
    - Gitignore certain things
    - Don’t put datasets in there
  + Rename other one ‘private’?

Read guidebook about thesis:

* Cover Page: Include a cover page bearing the project title, your name and the month and year of submission. In addition, the following text must appear at the bottom of the cover page: A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science/Research at Imperial College London Submitted for the MRes/MSc in XXXX
* • Declaration: The first page inside the cover must provide a brief declaration of the contributions made by you and by others to your project. Key points to address are:
  + Was the data provided to you or did you collect or assemble it?
  + Were you responsible for data processing or cleaning, if required?
  + Were any mathematical models developed by you or by your supervisor?
  + What role, if any, did your supervisor play in developing the analyses presented?
* Data and Code Availability statement: At the end of your Main text, before the References section, you must provide a statement titled “Data and Code Availability”, where you name a data (e.g., Dropbox, FigShare, Zenodo, etc) and a code (e.g., Dropbox, GitHub, etc.) archive from where the data and code can be obtained that will allow replication of your results. The code may be in the form of a single script file.
* All reports should have an Abstract, Introduction, Methods, Results, Discussion and possibly, a separate Conclusions section.
* The main text should not exceed 6,000 words in length (excluding figures, tables, references, and appendices/supplementary information).
* You should use a ‘normal’ font at 11 point or 12 point size. We recommend Helvetica, Arial or Times New Roman – similar fonts are also fine. Do not use highly stylized or bitmap fonts. You may have any number of references, but note that excessive referencing will be as frowned upon as inadequate referencing!
* Use a in text citation style that has author names (e.g., Smith & Thomas, 2015).
* Computer Programs. If the program has been published, cite the reference, include it in the reference list and provide a brief outline of the methods it uses.
* All figures must be numbered and have a caption that is sufficiently detailed to explain the main features of the content by itself. All figures must be referred to in the main text of the thesis. Put the figures in appropriate points in the text, close to the text that refers to them.
* The main body of the text should use 1.5 line spacing and page numbering should be used. The thesis margins should be at least 2 cm and the main text font size should not be smaller than 11 point.
* CMEE Masters project proposals should be written in LATEX
* You must submit an electronic version of your project report on Blackboard and your class git repository (CMEE Masters).
* We suggest that you bring a printed copy of your Thesis to your External and Internal Vivas.

Read guidebook about presentation and viva:

* The structure and content of the presentation should match that of the final dissertation.
* Each student will undertake a 30-min Viva with two Examiners

Read through thesis feedback.

Mon 8/8/22

* Lab meeting 10am
  + Update since last week:
    - Made README
    - Did adjustments to design of app etc.
    - Presentation in coding club
    - Got thesis feedback from Joe which I’m going to work on this week

Went through intro and improved (using Joe’s comments).

Tue 9/8/22

Went into NHM, arrived at 10:30am. Left at 18:15pm.

Sent follow-up email to Ecosulis.

Working on methods section of report

* Read through Joe’s comments
* Read through Joe’s paper he is writing about an app used for near real-time monitoring of conservation culturomics using Species Awareness Index
* Reading about ShinyProxy (see below)
* Put together new structure for methods, move parts of writing around as to where they may be included in the new structure
* Started writing first couple of paragraphs

Reading about ShinyProxy:

* ShinyProxy is a software that can serve containerized web applications – including Shiny apps – without limits on the number of concurrent users. It comes with free enterprise features, such as authentication and authorization
* When deploying a Shiny application with ShinyProxy, the application is simply bundled as an R package and installed into a Docker image. Every time a user runs an application, a container spins up and serves the application.
* Shinyapps.io
  + Shinyapps.io is an online service for hosting Shiny apps in the cloud.
  + Many customers use shinyapps.io to prove out some concepts, build out a prototype, or just run it for a short period of time for their own purposes, while others are using it as a core component of their analytical offerings within a larger online property.
  + The service runs in the cloud on shared servers that are operated by RStudio. Each application is self-contained and operates on either data that is uploaded with the application, or data that the code pulls from third-party data stores, such as databases or web services.
  + You don’t need to own a server or know how to configure a firewall to deploy and manage your applications in the cloud.
  + It has both free and paid plans.
  + The fastest and simplest way to deploy shiny apps is to use the hosting service provided by RStudio at [www.shinyapps.io](http://www.shinyapps.io/) (see [official documentation](http://docs.rstudio.com/shinyapps.io/)).
  + The alternative to shinyapps.io is to host shiny apps yourself by running a Shiny server on your own physical infrastructure or cloud. Although this option requires more initial set-up and internal resources, it offers the following advantages:
    - The data does not leave your organisation
    - You retain control of infrastructure and performance
    - More authentication options
  + There are currently two products in the market:
    - RStudio’s Shiny Server. There is a free version, as well as a “Pro” version with built-in authentication and improved performance. It runs on Linux.
    - Open Analytic’s ShinyProxy. It is free, open source and offers equivalent features to Shiny Server Pro in terms of authentication and performance management. It is based on Java and Docker and can be run directly on a Windows OS
* In this article we will cover three options available to host a shiny app. These options are **a)** **Shinyapps.io b) Shiny Server and c) Shiny Proxy.**We will compare these options based on efforts to use, stability of shiny app after hosting, pros and cons
  + Shinyapps.io
    - For hosting Shiny web apps (applications).
    - Once you have build your shiny app you can directly connect to shinyapp.io from RStudio using rconnect and upload the app to shinyapps.io server. Once its uploaded you can directly visit the link
    - Pros:
      * No installation required.
      * Hosting is easy as the server maintenance is taken care by shinyapps.io.
      * shinyapp.io provides a dashboard to monitor the hosted app metrics/logs etc.
    - Cons:
      * You can host only 5 Apps.
      * You will get only 25 active Hours on the server.
  + Shiny Server
    - Builds a web server specifically designed to host Shiny apps.
    - Shiny Server will host each app at its own web address and automatically start the app when a user visits the address.
    - When the user leaves, Shiny Server will automatically stop the app.
    - Shiny Server comes in two version. The Open Source version with limited feature and paid version with annual subscriptions.
    - Pros:
      * You can host as many apps you want.
      * You will get unlimited uptime hours.
      * You can customise the server configurations and authentications.
    - Cons:
      * One can have 20 sessions at a time, which is a headache when you have a larger team (more that 20).
      * Restrictions in Server configurations.
      * Required basic understanding of server configurations due to which Set up and installations need efforts.
  + Shiny Proxy
    - In order to get the feel of enterprise features, shinyProxy is a very good option. "ShinyProxy is your favourite way to deploy Shiny apps in an enterprise context. It has built-in functionality for LDAP authentication and authorisation, makes securing Shiny traffic (over TLS) a breeze and has no limits on concurrent usage of a Shiny app."
    - For every user shiny proxy starts a docker-container and maintain a separate session for every user.
    - Pros :
      * Unlimited app hosting
      * Unlimited concurrent users
      * Better user based Authentications
      * Highly stable
      * Once hosted minimal down time
    - Cons :
      * Since for every user shiny proxy initialise a container, it might consumes the RAM of the system/ server.
      * User has to wait till the container is initialised.
      * Requires effort to set up the platform
* Shiny-server can share memory between multiple instances of the same app while each instance of a ShinyProxy app is completely independent.
* Multiple instances of the same app will share the same R instance and processor thread in Shiny-server. So many concurrent users may experience a slowdown of a compute-heavy app when hosted using Shiny-server. In ShinyProxy, each user will get their own Docker container/app which could be distributed between multiple processors, supporting better scaling of compute-heavy apps.
* Shiny Server is limited to one R process per app, which potentially can serve multiple user sessions (connections to the app). This is totally fine for apps that don’t have many users, but it doesn’t work well for apps that will have large amounts of users. When you have many users, they all will be served by one process and will inevitably block each other.
* In contrast, on Shiny Server Pro, you can have multiple R processes per app. This means that many concurrent users can be distributed between separate processes and are served more efficiently.
* ShinyProxy from OpenAnalytics is an open source alternative for serving Shiny apps. It provides many features required by enterprise applications. Its architecture is based on docker containers, which isolate the app’s environment and can serve not only Shiny apps but other apps, too (like Dash). The key difference is that ShinyProxy starts a new app instance for each new user.
* In my opinion, shinyapps.io is more suitable for small personal projects or data science team who just want to focus on getting the model right and do not need massive computing power in each instance.
* Shiny Server or Shiny Server Pro on own premise or cloud
  + There are two versions of Shiny Server (both of them are developed by RStudio), one is the Shiny Server Open Source, and the other is the commercially licensed Shiny Server Pro. The former is free while the latter will set you back $9,995 / year with its base plan (20 concurrent users).
  + Shiny Server Pro natively supports a variety of password authentication schemes and a centralized management console for all your apps. But crucially, with Shiny Server Pro, you can launch multiple R processes per app, which can then serve multiple users concurrently. While the open-source version only supports one R process per app. Since R is single-threaded, sharing an R process would mean that users cannot do anything (not even load the UI) if R is occupied by other users.
* ShinyProxy on own infrastructure or cloud
  + With Docker, each user session has its own isolated workspace and R process, circumventing the limitation of the single thread described above.
  + ShinyProxy is a powerful tool which provides enterprise features while being open-sourced. However, setting it up might not be straightforward for people who are still finding their way in Docker and relevant tools.
* Shiny app served by ShinyProxy and hosted on AWS
* Shiny Server is a companion program to Shiny that builds a web server designed to host Shiny apps.
* Shiny server
  + Builds a web server specifically designed to host Shiny apps.
  + With Shiny Server you can host your apps in a controlled environment, like inside your organization, so your Shiny app (and whatever data it needs) will never leave your control.
  + You can also use Shiny Server to make your apps available across the Internet when you choose. Shiny Server will host each app at its own web address and automatically start the app when a user visits the address. When the user leaves, Shiny Server will automatically stop the app.
  + Shiny Server runs on a variety on Linux distributions.

Wed 10/8/22

Worked on methods section – wrote 2nd draft of methods based around the structure I put together yesterday.

Joe messaged to ask about the random intercepts of the model.

* Brought up the fact that the control variable differs among treatments. What that means is that each treatment LRR is measured relative to a different baseline, meaning even if you set the reference factor to 'Conventional', the interpretation of the LRR is not really relative to 'Conventional'.
* I’ve been thinking about this for a while but didn’t know how to phrase it, or have a solution
* Joe has suggested including Control as a random intercept in the model.
* This will require some adjustment to the ‘Control’ column to make sure the categories match those in the ‘Treatment’ column
* Alter the app to include these changes

Why do we log transform response ratios? - <https://stats.stackexchange.com/questions/433529/why-do-we-log-transform-response-ratios>

* Let's imagine you were meta-analyzing two studies. The first found that some treatment doubled the rate of the outcome i.e. 200% or 2/1 = 2. The second study, however, found that the same intervention cut the rate of the outcome in half i.e. 50% or 0.5/1 = 0.5. In this case, I would expect the meta-analyzed RR to be consistent with the "null effect" - an RR of 1. But averaging an RR of 2 and 0.5 results in a combined RR of 1.25. Taking the natural log first ensures that the average RR is 1. Or 0?
  + 2 + 0.5 / 2 = 1.25
  + ln(2) + ln(0.5) / 2 = 0
* Make positive and negative change symmetrical

Why is the log response ratio better than the standardised mean difference?

* Log response ratio less susceptible to non-independence common in ecological MA
* The standardised mean difference is calculated by dividing the mean difference by the pooled standard deviation
* The log response ratio is the mean of one treatment divided by the mean of the other treatment, logged. i.e. not dependent on standard deviation.
* Non-independence results in standard deviation being lower than expected, which affects the standardised mean difference.
* From Noble, 2017:

Text

Description automatically generated

GLITRS meeting 3pm:

* Joe away the week beginning 22nd August (project hand-in week)
* So my plan is to send new draft over by early next week – send over complete draft
* Then implement the comments and send onto Andy for final comments

Joe meeting 4pm:

* Comments on thesis that need clarification:
  + Page 6 – geographic variation at what scale?
  + Page 10 - What qualities does this tool need for it to be amenable to a living review?
  + Page 10 - Where to put bit about R version used etc.?
  + Page 14
    - Derive some stats/graphs that evidence how well you've satisfied those intentions.
    - What does a typical instance of running a meta-analysis look like, perhaps visualised as a flow diagram?
    - Across a sample of interactive meta-analytic sessions how long does it typically take to perform? Which is the part of the process that takes the longest?
    - How long does it typically take to upload a file?
    - How do both of these factors relate to the user? (if you've got time you could ask a few people to test your app)
    - How much do the results change among selected variables? This is a crucial point that relates to the benefit in having an interactive platform
* Who is the audience for the app? – put in intro aims?
  + Short term while in prototype phase – people in GLITRS
  + Longer term - Insect biodiversity practitioners.
* Implement random intercepts – (1|Control) + (1|Paper ID)
  + Condense control agricultural system categories to match treatment categories – remove rows where the category doesn’t match one of the unique treatment categories
  + Make new column for whether it’s treatment or control
  + Remove control rows
  + Then include Paper ID and Control as random variables in model
  + Currently, model had ID and Crop as random variables. BUT think this is data point ID rather than paper ID, future data sets won’t necessarily have crop, and ID doesn’t explain any variance, and crop very little.
  + Joe is going to talk to Adrianna about the whole process
  + Joe is going to send me spreadsheet for doing this
* How to word why we log the response ratio?
  + Include that it makes positive and negative change symmetrical and give e.g. like above
* Check understanding of why LRR better than standardised mean difference.
  + Yes, makes sense
* Tuesday presentation – talk through app, I don’t need to prepare anything else?
  + Correct
* Changed design of app
  + All good
  + At some point, change name input bit
    - Change message about name being acceptable – if possible, just have empty string
    - Allow hyphens
* Can we test whether you can upload a file?
  + This worked
* Asked people to test – shall I just set up a call with them and do it with them rather than them doing it on their own?
  + Best to do call and do it with them
  + Need structure for each
    - Start by introducing app, and what set out to do
    - Tell them if they have any thoughts as they go, let me know
    - Let them explore
      * Get them to run a custom model
    - Looking to assess user experience
      * Qualitative feedback, word cloud
      * Google market research?
* Check formula for % change to LRR and vice versa?
  + Text

    Description automatically generated with medium confidenceIn preparing data for modelling script, LRR is obtained like this but the formula laid out doesn’t match what is run in R?
  + In the app, to convert percentage change to LRR:
  + Text

    Description automatically generatedOnline: <https://www.jepusto.com/files/ABAI-2019-Log-response-ratios.pdf> taken from this paper: Using response ratios for meta-analyzing single-case designs with behavioral outcomes
  + Seems like Christina justed accidently missed out the +1 in the R code. Probably didn’t spot it because the percentage change column isn’t used again anyway. When convert LRR to %change for the app, the correct formula is used
  + Probably best to alter this anyway though

Asked whether people would be willing to test the app on Teams.

Thur 11/8

Went into NHM, arrived at 10:30am.

Joe meeting (continued from yesterday) 11am. (See above)

Pu together structure for user testing:

1. Send them email with link to app, dummy data file, and after-testing link to questionnaire
2. Get them to open app
3. Intro to what going to do in this session. Can ask questions as going along.
4. Give brief introduction to app and what I set out to achieve
5. Give them a chance to have a look at the introduction tab
6. Then ask qualitative user experience research questions 1:
   1. What is your first impression of this platform?
   2. What do you expect to gain from using this platform? / When do you think someone would use this platform?
7. Go to agricultural systems tab, get them to run a custom model
8. Go to uploading data tab, get them to upload file
9. Then ask qualitative user experience research questions 2:
   1. Was anything surprising or did not perform as expected?
   2. Was the platform easy to understand and use?
   3. What did you like and dislike about the design of the platform?
   4. Do the different sections complement each other and fit together so the app works well as a whole?
   5. What words would you use to describe the platform?
10. Ask them to complete SurveyMonkey anonymous survey
11. Delete file they uploaded from Google Sheets

Other points:

Put together quantitative Likert scale questionnaire for after testing app with me – use survey money link: <https://www.surveymonkey.co.uk/r/PFY6Z2D>

* The introduction tab provided good background information in terms of the app’s purpose and where the data comes from.
* Running a custom model in the agricultural systems tab was straightforward.
* Uploading data in the uploading data tab was straightforward.
* The design of the app was appealing, clearly laid out, and made it easy to navigate.
* The app is a useful contribution to the field of insect biodiversity change.

Added few bits to methods after clarifying with Joe.

Made plan of what to do each day until deadline.

Setting up quantitative analysis of how long it take to run models, and upload data.

* Use system.time() – wrap this around the code you want to time
  + This won’t work in my Shiny code because need to be able to access it afterwards
* Using this thread: <https://stackoverflow.com/questions/50335756/how-to-time-reactive-function-in-shiny-app-in-r>
  + Suggest creating reactiveValues object for time
  + Looked at both Sys.time and proc.time but think proc.time is better because Sys.time does weird things with units. Think proc.time just keeps it as seconds.

Fri 12/8/22

Carrying on with quantitative analysis that I started yesterday (timing model fitting)

* Got reactiveValues() to work with proc.time() so could time the time taken from the model starting to be run to the graph and table being produced.
* Did practise of this – ran 5 models (of varying data sizes) 10 times and recorded
* Put into R to calculate mean and standard deviation
* Made graph
* Not sure whether to have sd or sem as the error bars?
  + Asked Joe – do SE or CI
* Will probs have to repeat this once the code has been updated to run the new models

Repeated a similar process for quantitative analysis for uploading data

* If possible, get time for uploading to shiny app, and time to upload to googlesheets
* Then can have stacked bar graph? – Not really because file sizes aren’t categories
* Make dummy data x10 (100,200,300,…,1000 rows of data) – repeat each 10 times
* Record file size as well as number of rows
* Ended up plotting time taken to upload to shiny, and time taken to upload to Google Sheets against file size.

Updated methods with this quantitative analysis.

User experience research/testing:

* Daero – 11:30am
* Justin – 12pm

Mon 15/8/22

Lab meeting 10am:

* Update from me since last week:
  + Worked on intro and methods for thesis
  + Efficiency testing - quantitative analysis – timing how long it takes to run models, and upload data
  + User experience testing
  + This week – carry on with user testing, change model slightly by changing the random variables, GLITRS meeting tomorrow to show them the app, finish 2nd draft of thesis

User experience testing:

* Victoria – 11am
* Adriana – 2pm
* Alexa – 3pm

Updates to Shiny app:

* Fixed some grammar errors
* Changed title to just be a bug icon, rather than words – as had feedback that this seemed like a tab you could click on
* Changed some instances of ‘Google Sheets’ to ‘database’ as user doesn’t really need to know its googlesheets and might think they have to login into it or something
* Made it clear in the intro tab what each tab is for

Coding club 2pm:

* Yeahji presentation on project
  + Focus on urbanisation as a threat to insects
  + Nice slides – good images, not too many words
  + What do you mean by 2-sided linear model? – means mixed-effect or 2-factor model
  + Some labels on graph quite small, colours not that easy to distinguish
  + Maybe switch the order in the boxplots so it reads A-Z rather than Z-A, or do it from lowest to highest value?
  + On the boxplot - the ones with negative LRR, are they the ones that decrease BD when going from Nature to Urban, or vice versa? Negative LRR = decrease BD when going from nature to urban
* Did app user testing with Adriana and Alexa

Tue 16/8/22

Went into NHM, arrived at 10:30am.

Changes to code/Shiny app:

* Change inputted name bit of uploading data
  + Allow spaces and hyphens
    - Spaces – use [:blank:] to match space and tab, but not new line
    - Hyphens –
    - To make sure name only contains letters, spaces, or hyphens: "[^A-Za-z-\\s]"
  + Change message from ‘name is acceptable’ – just changed it to a blank message if the name is fine
* Change % change formula in preparing dataset for modelling script
  + Changed to this: perc <- mutate(perc, LRR = log(Percentage\_change/100) + 1) (originally didn’t have the + 1)
  + Actually, not sure I need to change it. I don’t think Christina means percentage\_change as the column name suggests. I think 100% percentage\_change actually means one treatment is 100% of the other i.e. no change. So the log response ratio is actually 0
    - So if percentage\_change value is 100, the LRR comes out as 0
    - BUT surely no percentage change should give LRR of 1?
  + Just left as it was for now
* Changing random variables of model
  + Made new version of data
    - Put in robust\_analysis/changing\_data\_for\_new\_model\_structure.R
    - Condense control categories to match treatment categories
    - Deleted control rows
    - Save new version of data set to output\_files/christina\_MMA\_LRR\_long.xlsx and meta\_meta\_analysis/data\_and\_models/christina\_MMA\_LRR\_long.xlsx
  + Add to save\_christina\_data\_to\_googlesheet.R
    - Upload new version of data to Google Sheets
    - Delete old version of data from Google Sheets
  + Change model formula in server.R to include control and Paper ID as random effects

Testing app now done this big change:

* Going to have to change the intro tables slightly – as only including Treatment agricultural systems, but Control ones are missing. So a lot of them say they only studied 1 agricultural system, when they actually studied 2 (it’s just that the control one isn’t counting currently)
* Will probably also have to alter frequency column of model output tables
  + Actually this might be fine – the default model adds up to 334
* Not sure where fallow has gone from the output? Have 9 agricultural systems now, not 10?
  + It’s because it’s only a control agricultural system, not a treatment agricultural system
  + Remove fallow definition from table?
* Model runs faster now as “less” data
* More models with filtered data don’t run anymore as obviously “less” data
  + Only ‘diversity’ and ‘unclassified’ run if only 1 BD metric is selected
  + ‘biomass’ and ‘development’ don’t run on their own anymore
* Will have to change example of how results change when running different models as results now don’t change much from default model to model without biomass. Could just run ‘diversity’ on it’s own, because transgenic becomes significant
* Will have to make new dummy data with this new format and test uploading this
* Change error message when model doesn’t run to select additional or alternative BD metrics

GLITRS meeting 1pm (Andy, Joe, Justin, Rob Cooke, Tim Newbold, Nick Issac, Charlie Outhwaite

* Presented updated version of the platform – Joe put together some slides on the background, and I did a demo of the app
* Explaining why control is included as a random effect – because LRR will in part be affected by what the control is
* Improvements/suggestions/additions
  + Will lmer always be the best model?
    - At least provide model diagnostic plots in the app
    - Check whether random effects are normally distributed. Check random variables have enough levels
  + More options for downloads:
    - Download figure
    - Download RDS model object with attributes e.g. for date, environment, version, variables such as the BD metric(s) selected etc. So could repeat same analysis later on if wanted to
  + Include doi in paper references
* Could end up as 2 papers – one more conceptual, one more technical

Wed 17/8/22

Make new dummy data and test – seems to be working fine

Updated methods.

Wrote 2nd draft of results.

Wrote 2nd draft of discussion.

Sent report to Joe and Andy.

GLITRS meeting 3pm:

* Update from me since last week:
  + User testing – done with 5 people
  + Quantitative analysis – timing models and uploading data
  + GLITRS meeting yesterday to do app demonstration
  + Changes to app
    - Fixed grammar/spelling issues
    - Added bit in intro about what each tab is used for based on the user testing
    - Changed checks and messages on the inputting name bit of the app – allows spaces and hyphens too, doesn’t say ‘name is acceptable’
  + Can’t quite work out the % change and LRR conversion bit and why Christina doesn’t include the +1, but I think there’s probably a reason for it.
    - If you get that 1 treatment is 100% of the other i.e. 0% change, then to get LRR, you do log(100/100) + 1 = 1
  + Changed the data format and model used in app
    - Some changes to results
    - No fallow treatment anymore
    - Less custom models run
    - Intro tables messed up
    - Joe and Andy discussed this and decided its best to put data back to original format but still keep model in its new formula (with paper ID and Control as random variables – control not because it’s the best model based on AIC, but because the data structure means it needs to be there)
  + Finish 2nd draft by end of today
* Ask about data and code availability
  + Email Samraat
    - Ask mention about not feeling 100%
  + Andy: “The data analysed here are available, only for the purpose of reproducing the analyses shown here, at XXXX. The data will be made publicly available at the time of publication of papers currently being written. (And then also include a README in the same folder that also says it's only for that purpose.)”
* Switch GLITRS meeting next week from Wednesday to Monday from 2:30pm (Daero will do presentation at 2pm)
* Andy will do session on vivas (and interviews) at some point after we hand-in our reports
* In viva: If they criticise something, if it’s something you think is good, then argue why it is good. If not, say ‘as I said in my discussion…’ and give potential ways around it. Could also ask them what they would suggest.
* Make a list of things that you’re not really sure on yet – then work on these until the viva.
* Include something balanced (strengths and weakness) in the discussion

Thur 18/8/22

Changes to code/app:

* Put data back to original format, kept model changes
* Changed error message to say alternative or additional, rather than just additional biodiversity metrics when custom model doesn’t run
* Removed conventional row from output tables as doesn’t make sense to give a t-value for this one. Included the number of conventional data points in the legend
* Changed 2 to 1.96 for t-values for statistical significance
* Change names of download buttons

Re-deployed the app.

Repeated model efficiency tests – did 10 different models this time, 10 repeats of each. Updated results section with these results. Also updated timing uploading data figure (changed axis slightly).

Re-did the reactivity tests – compared biomass and diversity custom models and updated results.

Fri 19/8/22

Went through feedback from Joe – intro, methods, results, and discussion.

Put contents page into report, and wrote acknowledgements, data and code availability, and declarations sections.

N.B. Random intercept vs random effects: In the R model formula: When there is a 1 before the line, you are accounting for random intercepts (varying baseline levels) in your variable.

Sat 20/8/22

Read through report, cut down some words.

Wrote abstract.

Sent report to Owen and Grace.

Read Grace’s report.

Started making flow diagram for results.

Sun 21/8/22

Finished making flow diagram.

Went through Andy, Owen and Grace’s comments.

Sent report to Mahika.

t-value = difference between your sample and the null hypothesis. I.e. if t-value is 0, no difference between sampled value and null hypothesis. Can assume 1.96 is fine given large sample size, if not will have to use student’s t table.

Mon 22/8/22

Lab meeting 10am

* Update from me since last week:
  + Mainly been working on thesis – looking at comments etc.
  + Re-did some of the quantitative analysis stuff now the models are finalised
  + Presentation with Joe for GLiTRS
  + Few small tweaks to the app

GLITRS meeting 2/2:30pm

* Why models don’t necessarily take longer as more data included – should I put in discussion?
  + Decisiveness data set – hill climbing algorithm gets there faster. Needs more iterations. Around the info content of the data. If data add noise rather than signal, will make it slower
* Screenshot of table – is this a figure or a table? It’s a figure of a table. Should legend go above or below it?
  + Stick with figure
* Check data and code availability and declaration statement, and abstract
* Check comparison of 2 custom models – 1 only has ‘1’ data point.
  + t = 4.88; but note that this is based on single input meta-analysis
* Citing multiple sources at once – alphabetical or chronological?
  + Chronological makes most sense
* Ask Christina about metric definitions
  + Taxon richness

Went through Mahika’s comments

Started transferring report to LaTeX

* Made bib file with references
* Put all images/figures in a folder
* Put all over to LaTeX – just need to do formatting things:
  + Bigger title
  + Why does ‘A thesis submitted for..’ go wider than other lines
  + Equations/formulas
  + Links to app, code, repos etc. - included hyperref package
  + Made figure resolution/quality better
  + Figure legends – put figure x in bold, and with full stop
  + et al in italics – would have to edit style file, which I have no idea where it is
  + Sizing/organisation of figures
  + Figure 4 – can see weird lines?
  + Some references are weird – e.g. Analytics, Open.
  + Make references spacing bigger

Tue 23/8/22

Finished transfer to LaTeX – completed the things in the above list

Edited bib file so references would be how I wanted them. Had to manually add the doi’s. Used natbib with apelike style.

In .tex file, in preamble:

\usepackage{natbib}

\bibliographystyle{apalike}

In .tex file, at end:

\addcontentsline{toc}{section}{References} % to include references in contents

\bibliography{project\_report\_bibliography\_edited.bib}

Wed 24/8/22

Final tweaks to app:

* Check grammar, spelling
* Update time it says it will take to run a model
* Change bit in upload data where it says about seeing the standardised data entry spreadsheet
* Add in what name will be used for in upload data tab
* Re-deploy
* Re-take screenshot of uploading data

Sort out data archive

* Create imperial dropbox, put file in, then share with dropbox. And then anyone with the link can see the file.

Sort out code and git repos

* One for whole project, one for shiny app.
* Transfer appropriate files, and make new git repos for both, and push.
* Sort readmes and gitignore files too.

Update thesis with correct links.

Change link to code in Shiny app and re-deploy.

Final checks of thesis and submit.

* Update thesis word count.

Other notes:

Presentation

* Build around mini-presentation I did
* Make videos for presentation

Useful for further down the line?:

* When data no longer has same columns because the spreadsheets will be heterogenous: Use the bind\_rows() method from dplyr to combine data frames with different columns. The column names are number may be different in the input data frames. Missing columns of the corresponding data frames are filled with NA. The output data frame contains a column only if it is present in any of the data frame.

Other little things to do/could do:

* Upload description of study as well as the data? – Save to separate googlesheet?
* Download all data?
* Have halt button if user decides they want to stop running their model?
* Add year and doi to paper references
* Add definitions of BD metrics
* Download figure or RDS model object
* Work out why ‘survival’ takes so long to run a model on
* Did you add attributes to the model that can be downloaded? If so, you could mention the importance of being able to transparently link outputs to inputs and knowing what was done in the middle

Things I have changed on Christina’s spreadsheet:

* Biodiversity metric category – 6 entries in long format spreadsheet are NA. I will change these to be unclassified (like some others already are)
* Name of the biodiversity metric column is spelt slightly wrong – I’ve changed this
* Yaap paper has 2 slightly different versions of the author, so going to make them the same

Original plan:

Make functions/package that someone could use to search for terms relating to their question e.g. effect of pesticide on insect BD, and then will return literature on this. Extract data and put into consistently formatted spreadsheets.

Shiny app over R functions/package because more usable to everyone – anyone can use shiny app, not just people who can code in R. More visually appealing.