CMEE Masters: Miniproject Assessment

February 14, 2022

Assignment Objectives: To address on a model-fitting problem using computational methods, and produce a written report, all in a coherent, reproducible, modular workflow under version control.

Student's Name: Lizzie Bru

Overall Miniproject Mark: 71%

Overall Project Organization

All directories in place and uncluttered.

You have included a `readme` file which briefly describes the project, lists the programming languages used (with version numbers) and the dependencies/packages needed, gives instructions on how to run the project and a brief summary of the key scripts.

You could have put the writeup LaTeX source files and pdf in a separate directory – this is what you should aim to do for your final dissertation.

Overall a clean and well organised project with good documentation. Nice job.

The Code

Your choice of coding tools is appropriate, nice to see a combination of R and Python being used for different tasks. For future reference, you could probably have saved yourself some headache in the R script by using str_split() rather than manually convert species labels to genus labels. Your use of packages is not excessive, which is good for your programming development and for reproducibility.

Your code is well commented, though you could remove large blocks of code that have been commented out since they are not ultimately part of the final workflow (as in miniproj_script.py L49-60). Your code is somewhat reasonably partitioned into two scripts (wrangling and analysis) but consider naming these more descriptively, and consider also separating out more individual tasks into their own scripts/files (e.g. model fitting, plotting, model comparison).

We encountered some errors in running your workflow. In miniproj_script.R we had to rename "gompertz_AICs2" to "gompertz_AICs", and in the same script we had to add "()" to one of your calls to the function "coord_flip" near the end of the script. Aside from this the workflow ran smoothly. You successfully fit 4 models (quadratic, cubic, logistic and Gompertz) and compare them using AIC. However, we note that you chose to log all the population sizes for the polynomial models. This is an unusual choice, and technically means you have chosen to investigate whether the log of the population follows a polynomial relationship w.r.t time, rather than the population itself. A better option might have been to fit the polynomial models to non-logged data, and the Gompertz models to logged data, and to manually calculate non-logged residuals for these so that you can still perform model comparison using AIC/BIC.

Running your project generated warnings regarding the number of iterations in nls.lm, lmdif. Ideally these should be addressed so that no warnings are given at all, but this is not always possible given time constraints, and is OK so long as the warnings aren't a sign of something fundamentally flawed.

Recall that you should write into your workflow commands that will delete all existing output files every time the workflow is run (they should be re-generated afresh). Also, put in checks so that the computational workflow aborts if any step in the analysis gives an error. Reporting that error to the user is a good idea too.

Your workflow incorporated progress updates displayed on the terminal, which is good. However these were a little hard to see among the other terminal outputs. In the future consider delineating them more – for example adding special characters makes messages stand out a lot better.

Your model ran in reasonable time (90s), with the model fitting and parameter sampling taking the bulk of the time.

Overall you have been commendably ambitious with your additional analyses but were held back by errors that we needed to fix before your workflow would run, and by a somewhat lackluster partitioning of the workflow into more specific scripts for individual subtasks.

Marks for the project and computational workflow: 68%

The Report

Very nice work. You completed an ambitious extension of the basic project brief, and analysed and interpreted results capably.

Title: Descriptive but slightly long.

Abstract: Good. Background/motivation sentence is a little generic. Study objectives, methods and results clearly conveyed. Take-home message a little vague but ok. (65%)

Intro: Also good. Motivation expanded upon alongside the background of model comparison and mechanistic vs phenomenological modelling with reference to the literature, though some specific factual statements are not explicitly referenced. Study objectives clearly stated, though could have specified which model comparison approach would be applied to the fitted models. (68%)

Methods: All essential elements present. Extra credit for fitting 4 models, for optimising initial parameter values with random sampling and for carrying out the thermal and other more advanced analyses. Computing tools section is present. (76%)

Results: First sentence regarding filtered data belongs in Methods. Credit for such a large proportion of subsets successfully fit to the Gompertz data. Fig 1 is a little random – would have been nice to see a few comparison plots with good/bad fits for different models. Otherwise, results succinctly and relatively clearly described and supported with figs/tables. (72%)

Discussion: Key findings summarised, reasonable discussion of what might account for differences in overall model performance, and thermally/medium/species-driven differences. Could nonetheless have tied things more closely to the existing literature. Honest appraisal of limita-

tions and potential future improvements/extensions. Slightly generic final paragraph. (72%)

(Some specific feedback is in the attached pdf, and we can also discuss more aspects of your write-up in our 1:1 feedback meeting)

Marks for the Report: 74%

Signed: Samraat Pawar & Alexander Kier Christensen

February 14, 2022

Notes on Assessment :

- This written feedback will be discussed in a 1:1 session scheduled after this assessment has been been given to you.
- The coursework marking criteria (included in this feedback at bottom) were used for both the computing and report components of the Miniproject Assessment. *In contrast*, Your final dissertation project marks are going to be based pretty much exclusively on the written report and viva (not code). Expect your final dissertation report to be marked more stringently, using the dissertation marking criteria (also included in this report).
- In the written feedback, the markers may have contrasted what you have done with what you should do in your actual dissertation. *This does not mean that you were penalized* — one of the main goals of the miniproject is to provide feedback useful for your main dissertation. However, there may be cases where what you have done is just really bad practise (for example missing line numbers or abstract), irrespective of whether it is a mini- or main- project report – you will be penalized in that case.
- The markers for this assessment are playing the role of somebody trying to understand and use your project organization and workflow from scratch. So it will seem like the feedback is particularly pedantic in places — please take it in the right spirit!