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: Congjia Chen

Overall Miniproject Mark: 70%

Overall Project Organization

All directories in place and uncluttered.

You have included a `readme` file that describes the project, languaged used, dependencies/packages, repo structure and a list of scripts with arguments explained. For future projects, pick a naming convention and stick with it! CamelCase is great, and so is snake_case, a mix of both is not ideal (naming comes with practice, so don't worry too much, but do bear it in mind).

Overall a relatively clean and well organised project. Nicely done.

The Code

Your choice of coding tools is generally sensible. Preferring to write all the major code in R is ok, but do remain open to Python or C in future for more computationally expensive projects. You are relatively sparing in your use of packages, which is good both for your own development as a programmer, and for the reproducibility of your project.

Your code is reasonably well commented, and you have split the coding up quite rationally into a series of different scripts. That said, the internal logic of some scripts is somewhat hard to follow in places, and recall that it is generally considered best practise to define functions at the start of a script rather than throughout the main body of the code. In future you may also want to consider writing separate scripts to hold all your functions and the main body of your code.

Your workflow ran without error, which is great. You successfully fit 6 models (linear, quadratic, cubic, logistic, Gompertz and Baryani) and compare them using AIC, AICc and BIC. However, we note that you chose to log all the population sizes even for the polynomial (and logistic) 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 and logistic models to non-logged data, and the Gompertz and Baryani models to logged data, and to manually calculate non-logged residuals for these so that you can still perform model comparison using AIC/BIC.

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.

You incorporated progress updates, printed to the terminal. This is great for the user to know what is going on at a finer level, and helps troubleshoot if needed. Your use of special characters helped these progress updates stand out from other terminal output. Nice job.

Your project ran in reasonable time (110s), with most of the time spent on model fitting.

Overall, a solid project with ambitious additional analyses, and an error-free workflow. Your code is a little hard to follow in places, consider keeping variable names more informative or streamlining the logic of your code a little in future.

Marks for the project and computational workflow: 72%

The Report

You appear to understand the relevant concepts and the tools at their disposal, and have made a good stab at a rather ambitious set of analyses over 6 different models. However you are being rather let down by a distinctly unfocused writing style that is unfortunately confusing in places and ultimately produces a report that struggles to truly let the quality of the underlying work shine through.

Title: Concise but pretty vague.

Abstract: Meets the brief but a little unfocused. The abstract reads like a sequence of not entirely related sentences rather than as one cohesive whole (I appreciate this might be partly a language complication). Study objectives not clearly conveyed before mentioning findings either. (59%)

Introduction: More cohesive than the abstract, which is good. I like the effort to empirically demonstrate the increased interest in modelling over time, but Fig 1 seems to show some kind of periodic boom-and-bust behaviour rather than steadily increasing interest. Background of the mechanistic models and their relationship to growth stages is well described, though you do not discuss the phenomenological models that you use later. The research question however is clearly stated, and good that you specified the model comparison techniques as well. (67%)

Methods: All essential components appear to be present, though it doesn't seem that data were filtered according to number of samples. Models are clearly defined, as is AIC/BIC. Initial parameter estimates not discussed. Extra credit for fitting 6 models, attempting model comparison beyond just AIC, and attempting analysis of temperature, measurement style and medium. Computing tools section is present. (71%)

Results: Linear model comparison is well presented, I like the multipart figure in particular. I like the idea of comparing AIC, AICc and BIC to determine which is most useful...but it's not clear exactly how you're doing this from the text. Fig 4 is a rather confusing way of showing the strength of each model type in my view – stick to the summary as in Fig 5 or plot a genuine distribution of AIC values for each model. The main result is ultimately communicated in all this, but somewhat messily. Fig 6 on the other hand is a nice breakdown of different trends in the input data and how this affects different model fits. Might be better suited to Discussion? Careful with the text size in your figures (in particular Fig 5 text is too small, and the x-axis of Fig5a appears to be mislabelled as counts rather than percentages). (65%)

Discussion: The discussion of starting value selection strategies is more suited to the Methods,

except where it is mentioned as a limitation of the study and a possible avenue for future improvement. You have included a treatment of weaknesses/fortes of each model, though it is a little scattered and hard to distill an ultimate conclusion from. You should clearly define "segmented model strategy" somewhere as well. You also interpret the results of the additional (temp/medium/etc. . .), but here your writing seems somewhat unfocused. I think the important fundamentals are there, and will shine through if you can work on the clarity of your writing! (63%)

(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: 68%

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!