



Bayesian selection of growth models with systematic error estimation

Ndour Alassane-Anand

Aims, objectives

Growth models are important as they touch multiple domains from fundamental biological cell growth to complex man-made measurement such as Gross Domestic Product (GDP). Therefore, the selection of the underlying growth process of a time series is just as prominent a problem. Naturally, it is crucial to select the best functional form as the real one is often unknown, and a wrong selection invalidates any following inference (Nguimkeu, 2014).

In this context, the aim of this project is to create an empirically driven Bayesian model selection process for growth functions in a noisy environment. The selection process would first be created and validated on generated data, tested on one or multiple suitable datasets and evaluated with corresponding literature. Throughout this project, an emphasis will be put on biological models, particularly cell counting issues as described by Harris et al. (2016). However, this does not exclude other similar problems to be studied in this framework.

A consequence of such a model selection would be to create a confidence interval through which one could evaluate the certainty of selecting a specific growth model over another. Through Bayesian modelling, the parameter distributions will also be obtained and can serve for interpretation.

The time plan and different milestones of the analysis are given in the Gantt chart figure 2.

Outcomes / deliverables and beneficiaries

The deliverable outcome of this project consists of a code that encompasses a model selection process as well as the parameter estimation of the function of interest. There may be multiple models used to select ‘the best one’, all of which will also be included along with corresponding robustness checks. Finally, the evaluation of the model on the generated data and on existing datasets will be provided. A commercial partner might be included, in which case the selection process shall be used on the provided dataset. One important metric that will be provided is the confidence in selecting the best functional form.

Research question(s) and scope

The key question answered here is: How can one make a model selection process of growth functions in a noisy environment while estimating systematic errors?

From there on many questions trickle that constitute the main research components of the analysis.

- I. How to deal with noisy data in a time series construction?
- II. How to evaluate growth while taking into account systematic errors (such as measurement errors)?
- III. What evaluation methods yield a good model selection process?
- IV. How to estimate the confidence of the selected model?

A discussion of the relevant literature is presented here to form an analysis path which is outlined in the methodology section.

I. Literature review

In cell biology, counting cells is one of the most widespread exercises and is used for instance to determine the effects of a drug on cell growth dynamics (Harris et al., 2016). Oftentimes, end point measures are performed

which access the count of cells a long time after the treatment period following which the functional form described by the data is eye-balled or fit with a logistic regression using Least Square or Maximum Likelihood Estimation (MLE) to identify coefficients (Millar 2011). However, as pointed out by Harris et al (2016), this approach can be flawed due to systematic errors. Recent works such as (Albert and Mafart 2005, Pouillot 2003) in population models can help address those flaws by introducing Bayesian methods. Furthermore, Harris et al. (2016) helps minimize the uncertainty gap by creating an algorithm to obtain full posterior probabilities. In a Bayesian manner, this method estimates the distribution of parameter values of a growth function by taking into account potential systematic errors.

However, if the researcher does not know which functional form to fit, a model selection process is necessary. A similar problem arises in other disciplines such as in economics when researchers are uncertain of the underlying growth model and its drivers. This is illustrated by Nguimkeu (2014) where the author presents a model selection process to choose between Gompertz and Logistic Growth Models. Often times, to determine the coefficients, researchers use model averaging as demonstrated by Claeskens and Lid Hjort (2010) which consists of computing a weighted average of the candidate models.

There are several manners to determine assigned weights and Claeskens and Lid Hjort (2010) emphasize that using a selected model as the best one without taking into account the other candidates could lead to over-optimistic tests and biased inference. It is important to keep this in mind once coefficient estimation is necessary. However, the question of determining the best model still remains. Dormann et al. (2018) highlight three main paradigms to select weights for model averaging. As model averaging is intimately linked to the quantification in selection problems, it can be seen as a good starting point:

- The first relies on purely Bayesian theory and although intuitive it becomes computationally difficult due to exact calculation of the marginal likelihoods (Lambert, 2018). Bayes factor would be categorized here; As it is a topic of interest in this project, we define it in further detail: consider two models M_a and M_b that we wish to compare.

Then Bayes theorem states that:

$$P(M_i | data) = \frac{P(data|M_i) \times P(M_i)}{P(data)} \text{ where } i \in \{a, b\} \text{ (eq.1)}$$


Following this, the Bayes factor B_{ij} introduced by Jeffreys (1935), is defined as:

$$B_{ij} = \frac{P(M_i|data)}{P(M_j|data)} \times \frac{P(M_j)}{P(M_i)} \text{ where } i, j \in \{a, b\} \text{ and } i \neq j \text{ (eq.2)}$$

Note here that if we assume uninformative priors, Bayes factor boils down to the ratio of the posterior.

- The second describes empirically driven approaches that are more common in machine learning tasks (Cross Validation or jackknife model averaging) - these methods are defended by Lambert (2018) and their success in machine learning testifies to their effectiveness;
- The third is methods that rely on information criteria such as the Akaike information Criterion (AIC) or the Bayesian Information Criterion (BIC). They are well established, and their low computational cost make them suitable in many fields of study such as social science disciplines Raftery (1995) or ecological population model selection (Dormann et al., 2018).

The purely Bayesian methods are computationally difficult and as pointed out by Kass and Raftery (1995) the BIC (closely linked to the AIC) gives an approximation of Bayes factor - which he notes is well suited for scientific communication (Linares, 2015). In a case such as Harris et al. (2016) where the marginal likelihoods are explicitly calculated, although expensive, a Bayes factor selection process makes sense. This is further backed by Gelman and Rubin (1995) who advocate against BIC and suggest using Bayes factor for selection problems. Also, the computational cost of Bayesian estimation has been reduced by methods such as Monte Carlo Markov Chains (MCMC) as pointed out by Steel (2015).



It is sometimes argued that for model selections, Bayes factors are not only expensive but also subjective (Turner, 2012; Steel, 2015). This is because to compare models M_a and M_b with Bayes factor, one would need to compare the ratio of their likelihoods (eq 2) - assuming uninformative priors. In such a case if likelihood ratio = 1, then the models are equivalent and if the ratio is vastly greater than 1, M_a is preferred. However, if the ratio is slightly larger than how would one determine the actual cutoff to certify which model is actually a better selection. Here, some common rules similar to p-value conventions are used Kass and Raftery (1995) although they remain subjective guidelines.

Furthermore, to compute the Bayes factor the researcher needs to provide the priors of each model. If she has a preference between one of the models, the amount by which the priors would be set to reflect this is difficult to quantify.

From this discussion, although it appears that data driven selection methods seem less flawed for model selection, some studies from the literature point out that it is not the case (Shao 1993). This is shown by Gronau and Wagenmakers (2018) who present evidence through experiments on three datasets that Cross validation in a Bayesian setting does not strongly support a simple but performant model as Okharm's razor would suggest - and as it is advised in the principles of model selection Mackay (2003). Furthermore, they demonstrate that the priors play an asymptotic role in the results of cross validation. These findings should be kept in mind for the problem at hand. It remains nonetheless true that cross validation performed with distributions is still a well-established methodology (Vehtari 2018) that is recommended in M-open cases (i.e. where the true model is not within the compared models which is often the case as it is unknown).

Furthermore, there have been interesting developments in combining Bayesian methods and cross validation as they are not mutually exclusive methods and can contribute to robust estimates. Such works include Bürkner et al. (2019) where the authors aim to improve upon leave-future-out cross-validation (LFO-CV) - an adaptation of leave-one-out cross-validation (LOO-CV) to timeseries - to reduce computation time. As with LOO-CV, LFO-CV is computationally costly since for each fold the posterior must be reevaluated. Vehtari et al. (2016) introduce a computationally faster method of LOO-CV using Pareto smoothed importance sampling. Bürkner et al. (2019) build on this model to construct a time-series equivalent algorithm. Furthermore, an interesting Bayesian model selection process with cross validation was produced in a population assessment in an ecological context (Link and Sauer, 2015).

Importantly, during this discussion, we have not made explicit mention of sample size as the datasets worked on here are large enough to disregard small sample properties.

Although many established methods for Bayesian model selection exist and systematic error evaluation algorithms have been established, there does not seem to be an empirically driven case of linking these two areas of study in a time series setting. This project aims to bridge this gap by a series of experiments to create a Bayesian model selection process for growth that can take into account systematic errors.

II. Methods and Limitations

Technology and Environment

The model selection tool will be built in Python 3.7 using scientific libraries (i.e. numpy, scikit-learn, pandas, matplotlib) including PyMC, a package used for Bayesian estimation in Python. An environment using Anaconda will be created for this purpose and specifications of package versions will be provided for the model selection to

run adequately. The source code from Harris et al. (2016) was obtained. This was written in Python 2.7 and will be used in this project and updated to Python 3.6 accordingly. Git will be used as a version control and Google Drive will be used as a backup measure. As some packages of specific statistical techniques described above have been written in R (e.g. LFO-CV - Bürkner, 2019), it will likely be used in analysis process.

Methodology

The aim of the project is to combine systematic parameter estimation in a growth setting with model selection process. This will take place through different experiments for which several indicators will be monitored and compared. A typical experiment takes the form given in figure 1. The remainder of this section follows the structure of figure 1 by expanding on each of its components and finishes by relating the framework to a concrete example.

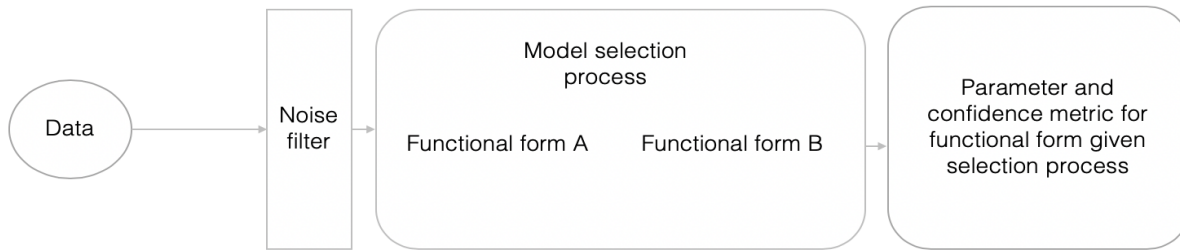


Figure 1. Process of experiment

Data

To operate in a controlled environment, the data will be simulated. This offers several advantages as pointed out by Kéry and Royle (2016), such as, first and foremost ‘know[ing] the truth’ behind the data. The importance of noise is critical to building a robust model so a noisy dataset will be generated. To do so, a synthetic dataset following one of our functional forms will be generated. We then add a term ε which designates the error. ε can be divided into two components: the random error and the systematic one. To simulate the data packages such as SimPy will be used.

The systematic error is given by Harris et al. (2016) and corresponds to the experimental error. This component, constant through time, ‘is an objective metric of the avoidable error’. Indeed, it represents the systematic error rate that can occur in an experiment driven by automated task - which would have a fairly constant distribution over time. Some examples are the error rate of the cell counting machine or the pipetting measurement error. Mostly this will be the form considered in this project which is rooted in a biological counting experiment - the same error type would not appear if we were measuring GDP growth for instance. There are other forms of systematic error such as the one proposed by Ghosh and Raychaudhuri (2007) which are more generalizable and can be considered.

The random error corresponds to the noise in the data and can take different forms; a classic white noise error term can be used here. Note that the generated noise process will be stochastic whereas the noise filter deterministic implying that both have a role to play in the overall construction. Of course, the aim is to build a robust model that detects the functional form under heavy noise.

Following the simulated data and experiments, real world data application is considered. The following datasets can be used in this context:

The dataset used in Harris et al. (2016) was generated using an open-source Python program for cell analysis called CEGANA. At this time, it is uncertain if this program is still available. If it is, this shall be used to generate data similar to Harris et al. (2016). This would be an ideal application as the Bayesian parameter estimation process used here is an application of Harris et al. (2016) 's analysis. Working with the same data should validate a logistic function compared to any linear functional form.

Another dataset to consider comes from Link and Sauer (2015). Here the researchers use a dataset of North American Bird population to assess the validity of their Bayesian cross validation method. This, if obtained, can lead to an interesting discussion contrasting the different methods.

Dataset from a commercial partner subject to availability.

Noise filter

One characteristic of the data for which we need to select a model is its inclusion of noise. Although Harris et al. (2016) provide a method which estimates systematic error, it will likely be important to pre-process any time-series to dampen the effect of random noise in the signal. Here there are many alternatives and for simplicity, the first filters considered are a moving average and an exponential filter (Brockwell and Davis, 2016).

Although not the primary focus of this study, using a range of filters will be helpful in building a robust model. As prescribed by Brockwell and Davis (2016) 'The choice of smoothing filter requires a certain amount of subjective judgment, and it is recommended that a variety of filters be tried in order to get a good idea of the underlying trend'. Therefore, after constructing models with basic filters, more advanced ones shall be used.

Model selection process

The model selection process constitutes the heart of the project. To begin with basic Bayesian models (without the systematic error) are estimated given a time series and compared. Initially, the comparison between linear functionals form or sigmoid models are considered. The first estimation of these is done in a straightforward fashion (i.e. evaluate the posterior given the integral) for small number of parameters. As the number of parameters increases, MCMC will become necessary. To compare the models and apply a selection criterion the initial methods considered are Bayes factor (for which the whole posterior estimation is necessary) and LFO-CV. Their respective metrics (B_{ij} for Bayes factor and the accuracy for LFO-CV) can serve as our confidence in the selection. For B_{ij} we can use guidelines provided by Jeffrey's scale (Wasserman, 2000) to create a monotonic relationship between compared models. Both estimation methods will be compared in terms of how well they classify a specific dataset and the required computational cost necessary to do so.

As a second step, the analysis by Harris et al. (2016) will be adapted: the systematic error will be added to the functional forms after which the selection process outlined above is repeated.

Note that during these two steps the priors will be held constant and remain non informative. They can be modified later (with caution as outlined by Young and Pettit, 1996) to observe its effects on the posterior.

Within each functional form, there could be different models (many independent variables in a linear function for instance). In this case, a more robust model can be built using model averaging (Claeskens and Lid Hjort, 2010). By averaging on the k best models, we would not obtain 'over-optimistic' parameter estimations. Through this strategy we should have a robust model and then compare the functional forms as described above.



Application example

Similarly, to Harris et al. (2016), an example where this framework could be applied is provided. Consider a biologist performing cell count growth. The task is done on two separate sets, one which is reacting to a drug and the other being a control group. The biologist has some prior knowledge of the growth functional forms of the cells in both cases. Now if she had to estimate the growth at a specific period for an unknown sample, she would use a model selection to determine what the best model is (a classification issue) and then estimate using the techniques described above. The confidence of the estimation would in fine be determined by the estimation and the confidence in the selection process. Therefore, an important aspect of the project will be to create a level of confidence based on these two components.

Limitations

It is noteworthy to mention that model selection is criticized as it implies relative model performance which does not signal a model's intrinsic value. If one were to consider a different functional form, it might surpass the model chosen by a model selection technique.

To assess the strength of the selection process and robustness of the result, multiple techniques are available. As outlined by Tsangarides (2004) model averaging described above acts as a robust method for estimation. Furthermore, checking if the cross-validation accuracy and Bayes factor coincide in terms of selected model further sheds light on result robustness. Finally, changing the data generation process and observing the modelling consequences is also a tool to verify the results of a given selection process.

The computation of Bayesian posteriors constitutes a risk as well as a limitation in this project. Here, other than code optimizations that can help ease the computation and using MCMC when necessary, there is no easy alternative to alleviate this risk. Furthermore, using Information Criteria as indicators before computing the posteriors might alleviate the risk of time wasted on unnecessary computation.

Finally, it is noteworthy to mention that the systematic error outlined by Harris et al. (2016) has an assumption that usually holds in well executed experiments: the constant distribution of error through time. This for example invalidates this method if we used different measures or instruments to evaluate one growth process.

Ethics review

There are no strong ethical concerns identified in this project. If the data of a commercial partner is used, a Non-disclosure agreement will be signed. For more information on ethical concerns please refer to the Ethics review form attached at the end.

Timeline and work plan

For timeline and deliverables please refer to the Gantt chart (Figure 2)

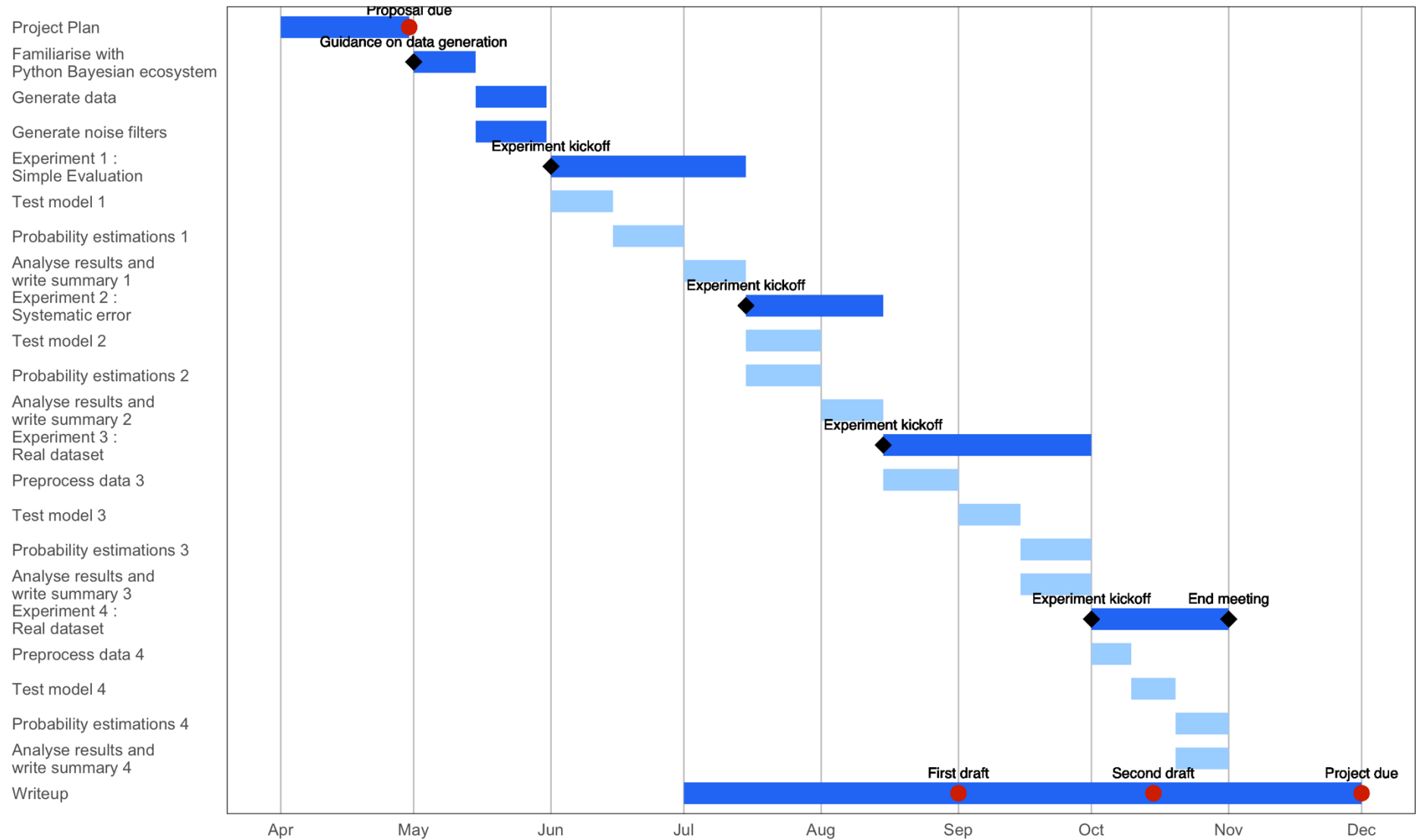
Risks

An assessment of potential risks and ways to mitigate them is given in the table below:

Description	Likelihood (1 – 3)	Consequence (1 – 5)	Impact (L x C)	Mitigation strategy
Computation cost : As the number of parameter increases probability density function estimations become infeasible	3	5	15	Estimation strategies such as MCMC, LFO-CV are meant to decrease computation time
Scripting Bayesian analysis can be challenging	3	5	15	Scientific libraries such as pymc3 will be used
Data loss / code loss / results loss (e.g. failure)	2	5	10	Use Git for code version control, Google Drive for backup storage of data and results
Data sensitivity: if confidential data is used	2	5	10	Do not share results or data with anyone other than supervisor and unless authorized avoid storage in third party (e.g. Cloud)
Some of the source code to be used is not available or usable	2	4	8	Will rewrite if necessary, with guidance of supervisor
Dataset obtention: Some of the data required is either from other studies or from a commercial partner. A part of the study is subject to its availability	2	3	6	Expand on simulated data or try and find an open source data set


Figure 2. Gantt chart (deliverables marked as red circles and meetings as black diamonds)

Here Writeup /summary writes include all steps including code cleaning



References

- Aki Vehtari (2018). *Comments on Limitations of Bayesian Leave-One-Out Cross-Validation for Model Selection* « *Statistical Modeling, Causal Inference, and Social Science*. [online] Columbia.edu. Available at: <https://statmodeling.stat.columbia.edu/2018/06/05/comments-limitations-bayesian-leave-one-cross-validation-model-selection/> [Accessed 27 Apr. 2019].
- ALBERT, I. and MAFART, P. (2005). A modified Weibull model for bacterial inactivation. *International Journal of Food Microbiology*, 100(1–3), pp.197–211.
- Brockwell, P.J. and Davis, R.A. (2016). *Introduction to time series and forecasting*. Switzerland: Springer.
- Bürkner (2019). *paul-buerkner/LFO-CV-paper*. [online] GitHub. Available at: <https://github.com/paul-buerkner/LFO-CV-paper/blob/master/LFO-CV.Rmd> [Accessed 27 Apr. 2019].
- Bürkner, P.-C., Gabry, J. and Vehtari, A. (2019). *Approximate leave-future-out cross-validation for Bayesian time series models*. [online] arXiv.org. Available at: <https://arxiv.org/abs/1902.06281> [Accessed 27 Apr. 2019].
- Claeskens, G. and Nils Lid Hjort (2010). *Model selection and model averaging*. Cambridge Cambridge University Press.
- Dormann, C.F., Calabrese, J.M., Guillera-Aroita, G., Matechou, E., Bahn, V., Bartoń, K., Beale, C.M., Ciuti, S., Elith, J., Gerstner, K., Guelat, J., Keil, P., Lahoz-Monfort, J.J., Pollock, L.J., Reineking, B., Roberts, D.R., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Wood, S.N., Wüest, R.O. and Hartig, F. (2018). Model averaging in ecology: a review of Bayesian, information-theoretic, and tactical approaches for predictive inference. *Ecological Monographs*, [online] 88(4), pp.485–504. Available at: <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecm.1309> [Accessed 27 Apr. 2019].
- Gelman, A. and Rubin, D.B. (1995). Avoiding Model Selection in Bayesian Social Research. *Sociological Methodology*, 25, p.165.
- Ghosh, K. and Raychaudhuri, P. (2007). *Tracing of Error in a Time Series Data*. [online] arXiv.org. Available at: <https://arxiv.org/abs/astro-ph/0701863> [Accessed 27 Apr. 2019].
- Gronau, Q.F. and Wagenmakers, E.-J. (2018). Limitations of Bayesian Leave-One-Out Cross-Validation for Model Selection. *Computational Brain & Behavior*.
- Harris, E.A., Koh, E.J., Moffat, J. and McMillen, D.R. (2016). Automated inference procedure for the determination of cell growth parameters. *Physical Review E*, [online] 93(1). Available at: <http://adsabs.harvard.edu/abs/2016PhRvE..93a2402H> [Accessed 27 Apr. 2019].
- Jeffreys, H. (1935). Some Tests of Significance, Treated by the Theory of Probability. *Mathematical Proceedings of the Cambridge Philosophical Society*, 31(02), p.203.
- Kass, R.E. and Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430), p.773.
- Kéry, M. and Royle, J.A. (2016). Introduction to Data Simulation. *Applied Hierarchical Modeling in Ecology*, pp.123–143.
- Lambert, B. (2018). *A student's guide to Bayesian statistics*. Los Angeles, Calif. Sage.
- Linares, A. (2015). *Introduction Overview Derivation BIC and Bayes Factors BIC vs. AIC Use of BIC*. [online] Academia.edu. Available at: https://www.academia.edu/11308132/Introduction_Overview_Derivation_BIC_and_Bayes_Factors_BIC_vs._AIC_Use_of_BIC [Accessed 27 Apr. 2019].
- Link, W.A. and Sauer, J.R. (2015). Bayesian Cross-Validation for Model Evaluation and Selection, with Application to the North American Breeding Survey. *Ecology*.
- Mackay, D.J.C. (2003). *Information theory, inference, and learning algorithms*. Cambridge: Cambridge University Press, p.,chap 28.
- Millar, R.B. (2011). *Maximum Likelihood Estimation and Inference : with examples in R, SAS and ADMB*. Chichester: Wiley.
- Nguimkeu, P. (2014). A simple selection test between the Gompertz and Logistic growth models. *Technological Forecasting and Social Change*, 88, pp.98–105.



Pouillot, R. (2003). Estimation of uncertainty and variability in bacterial growth using Bayesian inference. Application to *Listeria monocytogenes*. *International Journal of Food Microbiology*, [online] 81(2), pp.87–104. Available at: http://smas.chemeng.ntua.gr/miram/files/publ_330_10_6_2005.pdf [Accessed 27 Apr. 2019].

Raftery, A.E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, 25, p.111.

Shao, J. (1993). Linear Model Selection by CrossValidation. *Journal of the American Statistical Association*, [online] 88(422), pp.486–494. Available at: <https://www.jstor.org/stable/2290328> [Accessed 27 Apr. 2019].

Steel, M. (2015). *Bayesian time series analysis*. [online] The New Palgrave Dictionary of Economics. Available at: https://www.academia.edu/25066189/Bayesian_time_series_analysis [Accessed 27 Apr. 2019].

Tsangarides, C.G. (2004). *A Bayesian approach to model uncertainty*. Washington, Dc Imf.

Turner, R. (2012). *Why Gelman “hates” Bayesian model comparison*. [online] Available at: <http://www.gatsby.ucl.ac.uk/~turner/TeaTalks/bayes-model-comp/bayes-model-comp.pdf> [Accessed 27 Apr. 2019].

Vehtari, A., Gelman, A. and Gabry, J. (2016). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5), pp.1413–1432.

Wasserman, L. (2000). Bayesian Model Selection and Model Averaging. *Journal of Mathematical Psychology*, 44(1), pp.92–107.

Young, K.D.S. and Pettit, L.I. (1996). On priors and Bayes factors. *Journal of Econometrics*, 75(1), pp.113–119.

Research Ethics Review Form: BSc, MSc and MA Projects

Computer Science Research Ethics Committee (CSREC)

<http://www.city.ac.uk/department-computer-science/research-ethics>

Undergraduate and postgraduate students undertaking their final project in the Department of Computer Science are required to consider the ethics of their project work and to ensure that it complies with research ethics guidelines. In some cases, a project will need approval from an ethics committee before it can proceed. Usually, but not always, this will be because the student is involving other people (“participants”) in the project.

In order to ensure that appropriate consideration is given to ethical issues, all students must complete this form and attach it to their project proposal document. There are two parts:

PART A: Ethics Checklist. All students must complete this part. The checklist identifies whether the project requires ethical approval and, if so, where to apply for approval.

PART B: Ethics Proportionate Review Form. Students who have answered “no” to questions 1 – 18 and “yes” to question 19 in the ethics checklist must complete this part. The project supervisor has delegated authority to provide approval in such cases that are considered to involve MINIMAL risk. The approval may be provisional: the student may need to seek additional approval from the supervisor as the project progresses and details are established.

A.1 If you answer YES to any of the questions in this block, you must apply to an appropriate external ethics committee for approval and log this approval as an External Application through Research Ethics Online - https://ethics.city.ac.uk/		<i>Delete as appropriate</i>
1.1	Does your research require approval from the National Research Ethics Service (NRES)? <i>e.g. because you are recruiting current NHS patients or staff?</i> <i>If you are unsure try - https://www.hra.nhs.uk/approvals-amendments/what-approvals-do-i-need/</i>	NO
1.2	Will you recruit participants who fall under the auspices of the Mental Capacity Act? <i>Such research needs to be approved by an external ethics committee such as NRES or the Social Care Research Ethics Committee - http://www.scie.org.uk/research/ethics-committee/</i>	NO
1.3	Will you recruit any participants who are currently under the auspices of the Criminal Justice System, for example, but not limited to, people on remand, prisoners and those on probation? <i>Such research needs to be authorised by the ethics approval system of the National Offender Management Service.</i>	NO
A.2 If you answer YES to any of the questions in this block, then unless you are applying to an external ethics committee, you must apply for approval from the Senate Research Ethics Committee (SREC) through Research Ethics Online - https://ethics.city.ac.uk/		<i>Delete as appropriate</i>
2.1	Does your research involve participants who are unable to give informed consent? <i>For example, but not limited to, people who may have a degree of learning disability or mental health problem, that means they are unable to make an informed decision on their own behalf.</i>	NO
2.2	Is there a risk that your research might lead to disclosures from participants concerning their involvement in illegal activities?	NO
2.3	Is there a risk that obscene and or illegal material may need to be accessed for your research study (including online content and other material)?	NO
2.4	Does your project involve participants disclosing information about special category or sensitive subjects?	NO

	<i>For example, but not limited to: racial or ethnic origin; political opinions; religious beliefs; trade union membership; physical or mental health; sexual life; criminal offences and proceedings</i>	
2.5	Does your research involve you travelling to another country outside of the UK, where the Foreign & Commonwealth Office has issued a travel warning that affects the area in which you will study? <i>Please check the latest guidance from the FCO - http://www.fco.gov.uk/en/</i>	NO
2.6	Does your research involve invasive or intrusive procedures? <i>These may include, but are not limited to, electrical stimulation, heat, cold or bruising.</i>	NO
2.7	Does your research involve animals?	NO
2.8	Does your research involve the administration of drugs, placebos or other substances to study participants?	NO
A.3 If you answer YES to any of the questions in this block, then unless you are applying to an external ethics committee or the SREC, you must apply for approval from the Computer Science Research Ethics Committee (CSREC) through Research Ethics Online - https://ethics.city.ac.uk/ Depending on the level of risk associated with your application, it may be referred to the Senate Research Ethics Committee.		<i>Delete as appropriate</i>
3.1	Does your research involve participants who are under the age of 18?	NO
3.2	Does your research involve adults who are vulnerable because of their social, psychological or medical circumstances (vulnerable adults)? <i>This includes adults with cognitive and / or learning disabilities, adults with physical disabilities and older people.</i>	NO
3.3	Are participants recruited because they are staff or students of City, University of London? <i>For example, students studying on a particular course or module. If yes, then approval is also required from the Head of Department or Programme Director.</i>	NO
3.4	Does your research involve intentional deception of participants?	NO
3.5	Does your research involve participants taking part without their informed consent?	NO
3.5	Is the risk posed to participants greater than that in normal working life?	NO
3.7	Is the risk posed to you, the researcher(s), greater than that in normal working life?	NO
A.4 If you answer YES to the following question and your answers to all other questions in sections A1, A2 and A3 are NO, then your project is deemed to be of MINIMAL RISK. If this is the case, then you can apply for approval through your supervisor under PROPORTIONATE REVIEW. You do so by completing PART B of this form. If you have answered NO to all questions on this form, then your project does not require ethical approval. You should submit and retain this form as evidence of this.		<i>Delete as appropriate</i>
4	Does your project involve human participants or their identifiable personal data? <i>For example, as interviewees, respondents to a survey or participants in testing.</i>	NO