**Student Final Feedback**

**1. Miary**

Dear Miary,

Many thanks for your amazing contributions to E2M2 this year! It was truly a pleasure to have you as a student. Please see below for my feedback on your final presentation, and please don’t hesitate to reach out with any further questions or clarifications.

**Overall**

I love your project! You’ve picked a tractable system, collected excellent data, and it lends itself well to both statistical and mechanistic modeling. Congratulations. I think you are well on your way to completing the work you need to publish a paper this year, and I would be very happy to help you along this path if you have questions.

**Statistical Model**

Looks really great! One thing to note is that because this is a ‘Gaussian’ distribution, which is the default in R, you don’t actually need to make it a generalized model! You can just make it a linear model incorporating random effects, using the function lmer() in the lme4 package in R. This is good because this function rtypically uns into fewer convergence errors than you get with glmer().

Additionally, you can test two different hypotheses based on the way that you represent ‘year’ in this model. If you write the code as you currently have it:

lmer(SVL~mean\_temp + mean\_precipitation + (1|year), family=’gaussian’, data=my.data)

this addresses the question, *how does SVL vary with temp and precipitation?* But if you represent year as a fixed effect interacting with your temp and precip, like this:

lm (SVL~mean\_temp\*year + mean\_precipitation\*year, family=’gaussian’, data=my.data)

then you are testing the question, *how does SVL vary temp and precipitation as they change throughout the years?* I don’t know how many years of data you have, but if you had a long times series (5+ years), this would allow you to show that temp and precip increased or decreased with time (i.e. due to climate change) and that that affected the body size of the chameleon.

**Mechanistic Model**

Nice work here! In this case, I would not make growth a box but instead a state that impacts size. You can have ~3 boxes that represent different size classes (i.e. small, medium, and large) and model the impacts of temperature on their transitions between these boxes. If they transition faster, it means they have a higher growth rate. There is a whole class of models called ‘integral projection models’ that is used for this kind of work, and in reality, you have A LOT of boxes (100s-1000s!), each with a constant transition rate that lets you track the growth.

**Next Steps**

These are all great! Your statistical question and your paper are well within reach. Your plan for a hatching time model is well in line with what we typically use integral projection models for (see above), and it is extremely interesting to me! How many years of data do you have? Please get back in touch later this year after you’ve advanced the statistical work, and we can help you to build this hatching time model.

Great work, again!

Best,

Cara

**2. Ella**

Dear Ella,

Many thanks for your amazing contributions to E2M2 this year! It was truly a pleasure to have you as a student. Please see below for my feedback on your final presentation, and please don’t hesitate to reach out with any further questions or clarifications.

**Overall**

Fantastic work in the course overall. It was a true pleasure to have you and watch you develop and learn throughout the week. Your final models are in great shape, and I am excited to see where you take this work next!

**Statistical Model**

Great start here with the statistical model! You will likely find that some of these predictor variables are autocorrelated, meaning that they vary predictably together. So, for instance, habitat type will likely be predictive of both rainfall and temperature, which will also be correlated with each other. To account for these correlative effects, you can represent the predictor variables as interacting terms, written as temperature\*rainfall\*habitat\_type rather than as independent variables.

What does your study design and data look like? I am wondering if you need any random effects to account for your study design.

**Mechanistic Model**

Also looks great here! I love where this model ended up. Like I pointed out in class, you many want to include an influence arrow from the lemur to the forest natural regeneration rate (r) to show how the forest growth increases with more lemur seed dispersal. Your model could actually test the hypothesis of whether this is an important variable to show the contributions of the lemur population to the forest health. In this case, in your equations, the first term in the F equation would be written as rFL instead of rL.

Also, I know you mentioned this in your presentation, but it looks like you represented alpha as lambda in the diagram!

**Next Steps**

Nice plan! I would caution against ‘adding more variables’ because, remember, we want to construct the simplest model possible and only add variables where needed. In order to actually fit your mechanistic model, you will need a time series of lemur density and forest density in the same region, ideally with as many time points as possible, I’ll bet that exists in the literature for Ranomafana. If you found those data, I’d be happy to show you how to construct the model to fit it.

Great work, and please keep in touch!

Best,

Cara

**3. Cathucia**

Dear Cathucia,

Many thanks for your amazing contributions to E2M2 this year! It was truly a pleasure to have you as a student. Please see below for my feedback on your final presentation, and please don’t hesitate to reach out with any further questions or clarifications.

**Overall**

Wow, I am so impressed with your project and plans! You are clearly very advanced and have a lot of exciting data and research ahead of you. I hope we were able to help you advance your work—please feel free to reach out throughout the year ahead if you have more specific questions about your own dataset where we might be able to offer guidance!

**Statistical Model**

This looks fabulous! I don’t really have any constructive feedback because I think you know it will work (since it worked in the paper you referenced), and now you just need to implement. One thing I might ask is whether this MMRR function can incorporate either (a) interacting effects of predictor variables or (b) a correlation matrix. I would expect that climatic and geographic distance will be autocorrelated, and it would be good for your model to be able to account for this.

**Mechanistic Model**

Nice work here too—I love the pictorial diagrams explaining the different mechanistic processes for how we get at different genetic outcomes.

Your data are a bit difficult to model mechanistically, as I am sure you have realized. This model structure has aspects that will work, but you probably won’t be able to construct anything quite exactly like this. However, you can do something similar. There are a number of published nucleotide substitution rates, based on assumptions of genetic drift, and in using these to construct a phylogenetic tree, you are essentially testing mechanistic processes to find the model that most parsimoniously explains your data. You can test whether your phylogenetic tree can be explained exclusively by drift or whether the number of synonymous substitutions / the number of non-synonymous substitutions at a given site is higher than 1, suggesting evidence of positive selection in the genome.

**Next Steps**

This is a great research plan! I really like that you are starting out with such an advanced modeling framework and study design already before you even have the data. This is the best way to do science! Construct a model to formulate a hypothesis, collect the data to test the model, then test the model.

Great work!

Best,

Cara

**4. Hoby**

Dear Hoby,

Many thanks for your amazing contributions to E2M2 this year! It was truly a pleasure to have you as a student. Please see below for my feedback on your final presentation, and please don’t hesitate to reach out with any further questions or clarifications.

**Overall**

Fantastic work in the course as a whole, and great job developing these models! I think your question is fascinating and quite tractable—meaning that you should actually be able to make it work. I really like that you are formulating these research questions before you actually collect the data. Having a model-guided study design is one of the most important parts of science!

**Statistical Model**

Nice job here! I sort of doubt you will be able to actually collect data on the number of eggs lost, so why not construct your response variable as the inverse of this: the density of eggs in a given fecal sample. You can then have the same predictor variables but your data will not rely on needing two timepoints to show the loss of worms. You can incorporate a random effect to track individuals inthose where you do have multiple time points, and this should track the changing trajectory of the egg count for you. It would look like this:

glmer(egg density in fecal sample~age+sex+weight+dose+location+admin\_time +(1|lemur\_ID), family=”poisson”, data=mydata)

That way, if one individual has been sampled many times with different doses of phytomedicine, the model will account for this.

Also, note that age, sex, and weight will likely all be autocorrelated. You could model these as interacting predictors (using a \* instead of a +) to account for this.

**Mechanistic Model**

This looks great! Since you are interested in the dose of the phytomedicine, you could also model this as a two population model and have phytomedicine in the lemur gut as a separate box. As the phytomedicine grows, it could then have an influence arrow on the rate of worm loss. If you could collect detailed time series data for a few individual lemurs on (a) their longitudinal consumption of this plant and (b) their egg counts in feces, then you’d have a chance of being able to actually fit this model.

**Next Steps**

Great plan! I did not realize you were doing this in captivity until I read this slide! I think it is a very tractable plan, and I’m excited that you will get to control the experiment so well. Please feel free to reach out if you have any questions as you design your study.

Best,

Cara

**5. Fifi**

Dear Fifi

Many thanks for your amazing contributions to E2M2 this year! It was truly a pleasure to have you as a student. Please see below for my feedback on your final presentation, and please don’t hesitate to reach out with any further questions or clarifications.

**Overall**

Great job with this project! You’ve clearly been reading and learning a lot! I am very impressed with your plans for the year ahead and excited for where you will take this project next. Try not to be so hard on yourself about the presentation—I feel like you get so nervous and then so angry with yourself. I once gave a research presentation in front of a big audience and just completely froze on stage. I had to sit down and collect my thoughts before I could continue with the slides. Happens to us all. Big hug!

**Statistical Model**

Great start here! Typically, to measure species richness, you’ll want to calculate one or more measures of diversity (alpha vs. beta diversity or some composite measure like a Shannon’s diversity index, etc.). In this case, the response variable is going to be that measure of diversity, which will be a ‘Gaussian’ distribution, not a ‘Poisson.’ If you want to make the response variable counts of a specific type of microbe, that would be poisson, but in this case, I think that using the diversity index as a gaussian makes the most sense.

How are you planning to measure food variability? If you want to actually correlate diet breadth with microbe richness, then we might want to start DNA barcoding the fecal samples. This makes a lot of sense to me, but I’ll need to look into a few places to where we can send the samples to do this affordably. The other option would be to have this be predicted based on diet from hair isotope samples (what Anecia is working on), or we could do both.

Note that food\_variability is going to be autocorrelated with season, so you will probably want to show these as interacting effects: season\*food\_variability instead of season+food\_variability. Also, if you use a gaussian distribution, you can model this in a simple linear model using lm() with no need for glm(). You can incorporate species as a random effect like this:

+ (1|species)

In lmer() to model all three fruit bat populations at once.

**Mechanistic Model**

This is a great start! Instead of one class of m from the food, you will actually need all three which you can represent as m1f,m2f,m3f and then your colonization rates will show the difference in rates by which they establish in the gut: m1g,m2g,m3g. Other than that, it looks great, and I really love the interacting effects of competition in here. We could truly build this model—and show how it gets disrupted from infection, but we’re going to need to barcode the feces for foods as well, I think. Nice job!

**Next Steps**

Good plan! We’ll get you started on the lab work at some point this year! I know it is hard to find time for this when you spend so much time in the field. Great work, Fifi! I really appreciate all you’ve done with us this year.

Best,  
Cara

**6. Sylviane Miharisoa**

Dear Sylviane,

Many thanks for your amazing contributions to E2M2 this year! It was truly a pleasure to have you as a student. Please see below for my feedback on your final presentation, and please don’t hesitate to reach out with any further questions or clarifications.

**Overall**

Really fantastic job with your research plan! I was so impressed with your presentation yesterday. It looks like you’ve got some amazing data and some great plans for how to model it. I really think you will go far with this project.

**Statistical Model**

This looks great! I expect that you will find that your three predictor variables (altitude, temperature, and precipitation) to be highly autocorrelated. To account for this, you can model them as interacting predictors, using a \* instead of a +. So your R code would look like this:

glm(presence/absence~altitude\*temp\*precip, family=”binomial”, data=mydata)

If you have multiple sites in your study design that have been sampled each at different altitudes (I don’t know what the data look like), then you could add a random effect of site using the glmer() function.

**Mechanistic Model**

This also looks great—a highly relevant and important model! I asked you in the presentation if you thought that the different processes would respond to climatic change in different ways. To show these differences in your model diagram, you can incorporate a subscript nezt to each of your seasonal theta terms (something like this: ) to show that the seasonality is specific to the interaction in question.

Will you really be able to collect data on all these different life stages of mosquito across so many altitudes? Impressive!

**Next Steps**

This looks good. Would you need to modify your mechanistic model if it turns out that the two Aedes species are in competition? Would this be relevant across all altitudes? How might your demonstrate these changes?

Nice work! Please don’t hesitate to reach out with any further questions!

Best,  
Cara

**7. Fandresana**

Dear Fandresana,

Many thanks for your amazing contributions to E2M2 this year! It was truly a pleasure to have you as a student. Please see below for my feedback on your final presentation, and please don’t hesitate to reach out with any further questions or clarifications.

**Overall**

Amazing job in E2M2! You have such clean and beautiful data—it is a real treat to see a system that can be modeled so easily and appropriately. I have no doubt that you will go far with this project.

**Statistical Model**

Great start on this statistical model. How are you planning to measure species richness? Typically, this is measured using some sort of diversity index—alpha or beta diversity or a Shannon diversity index, etc. If that is the case, then your response variable will most likely be Gaussian, not Poisson. If, by contrast, you would rather your response variable be the count of number of species in a given fragment, then you are correct that this is a Poisson model. Do you think you would make different hypotheses about the expected results for these different model structures?

Also, typically island biogeography typically predicts that species number will result both from fragment size AND from the distance of each fragment to a mainland source population. Do you think that secondary predictor has a role in explaining species distributions in your own system?

**Mechanistic Model**

Great start here! To model competition in different patch sizes, you’ll actually need to have 4 boxes: species1-fragmentA, species2-fragmentA, species1-fragmentB, species2-fragmentB and then you are correct that you’ll have different rates of dispersal between these patches and different birth and death rates that are influenced by the patch size. If one species does better in a smaller patch, you might multiply its death rate by patch size to show that it dies more quickly in bigger patches, while if another species does better in larger patches, you could multiply its birth rate by patch size. In order to fit a dynamical model to these data, you’ll need a time series of population changes over time, or you’ll need age of the chameleons to aprroximate time.

**Next Steps**

Next steps look great! If you are planning future field work, I’d be happy to help you make a mechanistic model to design an appropriate field study to get you the answers that you need.

**8. Aina**

Dear Aina,

Many thanks for your amazing contributions to E2M2 this year! It was truly a pleasure to have you as a student. Please see below for my feedback on your final presentation, and please don’t hesitate to reach out with any further questions or clarifications.

**Overall**

Great work in class this year! It was fun to get to know you a little better after all those years seeing each other in the office at IPM. You’re working with a really impressive dataset, and you have a lot of great results that have already and will in the future emerge from it.

**Statistical Model**

Nice, clear, straightforward question! I would suggest that you give the model a nested random effects structure to account for the fact that you have multiple unique sampling points. Since you have both urban and rural sampling sites for each locality and your hypothesis is particularly linked to the question of urban vs. rural (or distance from the forest), I would suggest the following structure for the model:

glmer(status~age+sex+urban/rural+(1|locality), family=”binomial”, data=serohantamad)

**Mechanistic Model**

This also looks great! As Jean-Michel suggested, It will be fascinating to test whether a model with or without waning immunity will fit these data better, and it will also be useful to fit the model across a range of statistically estimated cutoffs for the seropositive threshold.

Since you assume that S and R will vary by age, you should represent them as S(a) and R(a) in the boxes as well.

**Next Steps**

These sound interesting, but I think you have a lot to do with the data you already have! My understanding is that your statistical model is already accepted at EID—is that correct? If you’d like some help publishing the age-seroprevalence mechanistic model, I would be delighted to be involved. I talked to Jean-Michel yesterday, and he said it would be no problem, so if you are interested and would like some help, please let me know!

Great work and see you in a couple months!

Best,

Cara

**9. Sedera**

Dear Sedera,

Many thanks for your amazing contributions to E2M2 this year! It was truly a pleasure to have you as a student. Please see below for my feedback on your final presentation, and please don’t hesitate to reach out with any further questions or clarifications.

**Overall**

Great job in the course this year! It was really fun to have someone with such a deep public health perspective. Your work is so important and your contributions already very profound. I hope we were able to help you achieve your research goals.

**Statistical Model**

This looks really good, and it appears that you already have very significant results. Since age is so central to your research questions, instead of representing age as a binomial variable (above and below 45), you could instead incorporate it in the model as a numeric, and then you’ll get a correlation between the age as a number and the probability of death during treatment. That will leverage a lot more power in your dataset!

**Mechanistic Model**

This mechanistic model also looks great! It is completely valid as is, but for simplicity, I would just suggest making the death rate during treatment a function of age. The new case rate and the rate of recovery may very well vary with age as well, but you’ll want to start by fitting the simplest model first and then add in these more complex constructions if they are needed to best recapitulate your data if the simplest model fails. It is fairly likely that just with one age-dependent parameter, you’ll get the result you expect without needing to add in the others.

**Next Steps**

These are fascinating and important next steps! There is a wealth of modeling literature on the interactions of HIV and TB in driving disease outcomes in many populations. If you end up finding appropriate data on this subject and need help with the analysis, please don’t hesitate to reach out for help! It’s a topic of great interest to me.

Congratulations, and best,

Cara

**10. Cédrique**

Dear Cédrique,

Many thanks for your amazing contributions to E2M2 this year! It was truly a pleasure to have you as a student, and I was so impressed with everything you brought to the workshop. Please see below for my feedback on your final presentation, and please don’t hesitate to reach out with any further questions or clarifications.

**Overall**

Amazing work! You clearly have a firm grasp of all modeling concepts and are well on your way to doing amazing things in science. I hope we were able to provide some useful new perspectives that were not too basic for your level!

**Statistical Model**

Statistical model structure looks totally perfect—nice work! One of the really cool things about statistical modeling is that you can use different model strucutres to test different questions with the same dataset. As it is, your model asks your exact question—what is the effect of shade on grass root biomass? This will give you a general trend that is controlled across species using your random effects structure. If, by contrast, you wanted to more explicitly test the hypothesis that different species might respond differently to this shade treatment, you could include species as a fixed interacting effect with shade (shade\*species), and the model will test the significance of each shade treatment for each grass species, with a random effect of replicate.

**Mechanistic Model**

This is also great—I love the simplicity. I know I mentioned this to you in our one-on-one meeting, but there is a whole class of models called integral projection models that is focused on modeling these growth processes—and they were actually originally devised for plants! In this case, the models do not actuall have one box—they have N1-Nx boxes where x can be sometimes thousands of age classes. And then you fit the growth rate that allows them to pass through from one size to the next. These are often the n incorporated into demographic models to show how size class influences things like probably of survival or reproduction.

**Next Steps**

Great plan! I love the idea of following up on the grass/forb interaction like we talked about last week. In an ideal world, what would you want to do next (post PhD) in science? I know you said that you were not that excited to continue as a postdoc with your current supervisors—would you rather have independent funding and your own lab at the university to explore other questions, or are you tired of grassland biomes in general? I’m just trying to get a sense of where you’d like to go next to see if we could manage to help in some way.

Thanks again for all your insights!

Best,

Cara

**Mentor Final Feedback**

**Liantsoa**

Dear Liantsoa,

Many, many thanks for all of your fantastic contributions to E2M2 this year. As a teacher, you were enthusiastic, dedicated, organized, and outspoken—a real leader on the mentoring team. As a researcher, it was equally rewarding to watch the way you absorbed new material and learned the science that underpins it and the many ways that it can be applied. Your final presentation was truly impressive—very well organized and elucidated, and it was delightful to listen to you present in French. As I mentioned throughout the week, I would be very excited to work with you and Durrell to help make conservation recommendations for the rewilding of these tortoises, and I’ll be in touch in the next month (likely with a grant proposal) to start taking steps in this direction.

You asked how to get to the point where you can do these things yourself—like Christian and Tanjona are always saying, the only way to learn R is to practice R. So you should take a figure—from my paper for instance—or even one of the figures that I produced for you already—and see if you can recreate it of your own accord from scratch. Using an example to then guide your own code construction is the only way you will learn.

Great work this week, and I am excited for the months and years ahead!

Best,

Cara

**Angelo**

Dear Angelo,

Congratulations on a fantastic week as a mentor at E2M2! I was very proud to see how much you have grown and learned in the past year—and especially how much you have taken ownership and initiative to learn R of your own accord. I really do think you are within reach of publishing your first first-authored paper this year.

As a teacher, you did a fantastic job engaging with and inspiring the students. My one piece of advice would be to work on trusting yourself more to be a figure of authority and lead the class as a whole. I think you are sometimes a little timid and over-nervous, but with time and practice, I expect you will get only more and more comfortable explaining the topics at hand.

As a researcher, you’ve done an amazing job putting together your data, analyzing it, and developing a story for a first paper. Sarah told me that she detected a large significant effect of the person collecting the samples in the ecto data, so it will be fascinating to see if that will disappear when you are modeling your data from the lab instead of simply from the raw database. I know I told you this in person, but it’s not really valid to incorporate k=4 as the smoother in your dataset. Christian may have told you that’s okay, but it’s really not because it forces the smoother to have 4 inflection points, meaning that it will show seasonality even if there is none. K=7 is the non-biased smoother presented in Simon Wood’s gam book. The other option would be to allow the data itself to choose how many smoothing knots it needs. To do this, you would write the smoother as:

s(﻿predictor, fx = FALSE, k=-1, bs = "cc")

Try doing that and comparing AIC with a model with 7 knots and also yours with 4. Let me know how it goes and I’ll advise you on next steps.

Given that the male and female bat fly seasonality does not look that different, I would also try fitting a model with one seasonal smoother that does not vary by sex and comparing AIC to see if that simpler model does an adequate job representing the data.

You can also extract the plot information from the gam object to make your own plot where you can play with the aesthetics (color, linetype, etc). See script for this here:

plot(gam1h, residuals = TRUE, pages =1)[[number\_of\_smoother\_you\_wish\_to\_extract]]

Great work! See you again soon!

Best,

Cara