Preguntas para Luz:

1. Al parecer, debo usar link= inverse para distributiones gamma (como Geometridae)… Se pueden comparar modelos en donde algunos son inverse y otros son log (pq con inverse no converge?, ejemplo gf1).

<https://stat.ethz.ch/R-manual/R-devel/library/stats/html/family.html>

<https://www.sagepub.com/sites/default/files/upm-binaries/21121_Chapter_15.pdf>

Luz no sabe si se pueden comparar cuando algunos son inverse y otros son log. Ella cree que no estaría bien, porque son modelos diferentes.

Opcion: Eliminar el gf1 y decir en los metodos que no converge. O igual correrlo asi aunque no converja y probablemente el AIC va a salir bien bajito. En ambas opciones el gf3 y el gf5 salen como los mejores dos modelos.

Luz recomienda preguntarle a Gotelli – que si cuando comparo modelos con AIC puedo usar diferentes link.

Tambien debería chekiar lo que diga Ben Bolker. Puedo preguntarle por stack overflow (Ben Bolker glmm – frequently asked questions). Tiene un índice y puedo buscar por temas

<http://bbolker.github.io/mixedmodels-misc/glmmFAQ.html>

1. Cómo comparar dos modelos con deltaic < 2… Estoy usando model averaging, pero tengo problemas con el método (ejemplo af1 que me da un WARNING por REML, y no estoy segura si lo estoy interpretando bien).

<https://danstich.github.io/stich/classes/BIOL678/06_modelSelection.html>

Luz no usa model averaging. Ella interpreta todos los modelos con mejor AIC. Otra opcion es escoger el que tenga menos variables, que sería el más parsimonioso.

Igual ella dice que tal vez que un interval de confianza de 0-0.4 esta bien, lo que no debe es cruzar por cero… tengo que verificar esto.

1. Tengo que revisar los residuales de cada modelo?

* Técnicamente si… Tal vez lo hago solo con los modelos que salgan más importantes.

1. What should I trust most, the glmulti results or the new individual models?

Sjplot o sjtools – function plot\_models

* Permite graficar forest plots, los cuales sirven para ver los valores fijos o los valores random. Puedo ver que tan importante es cada uno de los factores. Podria servir para evaluar importancia de Moonlight y decidir si eliminarlo

Buscar info sobre REML en la function de aictab… Debo entender mejor que significa esto. Buscar que significa el Warning que me esta dando.

<https://stackoverflow.com/questions/54980399/when-should-i-specify-reml-false-in-lmer#:~:text=1%20Answer&text=To%20cite%20the%20author%3A%20%22It's,theoretic%20criteria%20such%20as%20AIC.%22>

if you are using aic to compare 2 models, you CANNOT use the REML parameter until after you selected your best fit model. Then use REML.1.) Fit models WITHOUT REML (REML = FALSE)2.) Compare using AIC3.) select the best model and fit with REML.

Tosty sugiere correr todo en PERMANOVA (permutational multivariate analysis of variance). Es un modelo no lineal. Tal vez asi pueda correr todos los modelos sin problemas.

* Multivariate Analysis of Ecological Communities in R: vegan tutorial – ver seccion de permutation tests
* funcion Adonis (funcion de Permanova en vegan)

20 octubre 2020

Progreso con modelos de moth paper #2 de maestría

* Removí Moonlight como efecto aleatorio de TODOS los modelos, luego de probar con cor.test que no afecta significativamente la diversidad de mariposas de ambos grupos.
* Limité el NMDS de plantas a solo 2 dimensiones, ya que lo tenía en 3. Así me aseguro de que toda la varianza está explicada por esos dos ejes.
* Geometridae florístico corre completo – mejor modelo incluye solo VegDiversity, y le sigue el modelo nulo con un delta AIC<2. Tengo que revisar el ‘estimate’, porque en tab\_model sale positivo pero en el modelo como tal sale negativo. ---- FALSO!!! Al final cuando volví a correr los modelos florísticos me comenzaron a salir errores de isSingular…. TENGO QUE VOLVER A BREGAR CON ESTO… ME LLEVA PUTA!!!!
* Arctiinae florístico tiene problemas – tres de los modelos tienen el error de isSingular, probablemente porque los sitios de palma se están separando mucho los datos. Tengo que pensar en cómo resolver este problema.
* Geometridae estructural corre completo – mejor modelo es el nulo, lo cual indica que ninguna de mis variables estructurales explica la diversidad de geométridas.
* Arctiinae estructural tiene problemas – uno de los modelos tiene el error de isSingular.
* Intenté transformar geometridae y correr los modelos con lmer, pero me sale el error de isSingular en algunos de los modelos. Creo que este error sale más a menudo cuando uso lmer en lugar de glmer.
* El error de isSingular me está saliendo cuando la varianza de Habitat es 0. Una opción puede ser remover Habitat del modelo por completo, ya que si su varianza es 0, esto no debe estar afectando mis resultados. Pero, al hacer esto, ¿puedo seguir comparando con AIC con los otros modelos en donde sí dejé Habitat como efecto aleatorio?

<https://github.com/lme4/lme4/issues/518>

<http://bbolker.github.io/mixedmodels-misc/glmmFAQ.html#singular-models-random-effect-variances-estimated-as-zero-or-correlations-estimated-as---1>

* Otra opción es usar glmer para arctiinae también, especificando en el modelo que tengo una distribución Gaussian con link = identity. Aquí me da un warning de que debería usar lmer directamente, pero puedo ignorar este warning?
* Cosas por revolver
  + Error isSingular en varios modelos. Que hago?
  + Cambiar lmer for glmer Gaussian para arctiinae, se puede? Puedo ignorar el warning?
  + Creo que forzar el nmds a 2 ejes me causó que comenzaran a salir errores en los modelos de geometridae… Lo dejo en 2 ejes o lo devuelvo a 3? Usar Past para calcular importancias de cada eje.
* Tengo que buscar ayuda para resolver el problema de isSingular… Ya llevo demasiado tiempo trancada en esto y no vale la pena seguir rompiéndome la cabeza intentando resolverlo sola. Opciones: Scott Merrill, Gotelli, Easton White, Ollin.

Opciones luego de plática con Nati – 22 oct 2020

* Cambiar mi variable respuesta… Por ejemplo, utilizar nmds 1 de las mariposas en lugar de Fisher. Ahí tal vez puedo aumentar mi n, porque podría tener 120 sitios en lugar de 20. Otra opción podría ser que mi variable respuesta sea diversidad funcional, utilizando los datos del paper con Dominik.
* Si aumento mi matriz de datos a 120 líneas (1 línea por sitio por fecha), puedo meter tiempo en mi modelo. Nati dice que hay una manera de decirle al modelo que hay repetición temporal, por ejemplo utilizando variables aleatorias compuestas como tiempo/hábitat.
* Si utilizo otras variables (e.g. riqueza o abundancia) como variable respuesta, evaluar distribución Poisson, que se utiliza para datos con muchos ceros:
  + A Poisson or negative binomial regression model, with robust standard errors to account for overdispersion, would be my first approach to the data. If excess zeros are still a problem, you can use a zero-inflated negative-binomial model.
  + Why do you want the data normally distributed? That kind of thinking seems very outdated, given all the advances in statistical theory and computer-assisted statistical analysis techniques. Fit various distributions to the data and examine which one most closely approaches the distribution of your sample. Compare the goodness of fit statistics of different ways of modeling the data as they are (without any transformation). My first guess, without looking at your data, is that one of these will provide the best fit: (1) zero-inflated negative binomial regression, (2) zero-inflated poisson regression, (3) zero-inflated beta regression
    - [http://onlinelibrary.wiley.com/doi/10.1002/bimj.200390024/abstract](https://www.researchgate.net/deref/http%3A%2F%2Fonlinelibrary.wiley.com%2Fdoi%2F10.1002%2Fbimj.200390024%2Fabstract)
    - [http://math.usu.edu/jrstevens/biostat/projects2013/rep\_ZIP.pdf](https://www.researchgate.net/deref/http%3A%2F%2Fmath.usu.edu%2Fjrstevens%2Fbiostat%2Fprojects2013%2Frep_ZIP.pdf)
    - [http://www.sciencedirect.com/science/article/pii/S0167947311003628](https://www.researchgate.net/deref/http%3A%2F%2Fwww.sciencedirect.com%2Fscience%2Farticle%2Fpii%2FS0167947311003628)