

Oh, what a tangled web we weave, when first we practise to misspecify our evolutionary models

Stephen Crotty

School of Mathematical Sciences, University of Adelaide

November, 2014



What is Model Misspecification?



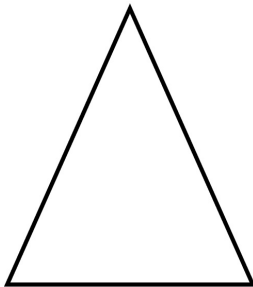
What is Model Misspecification?



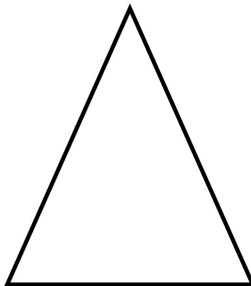
What is Model Misspecification?



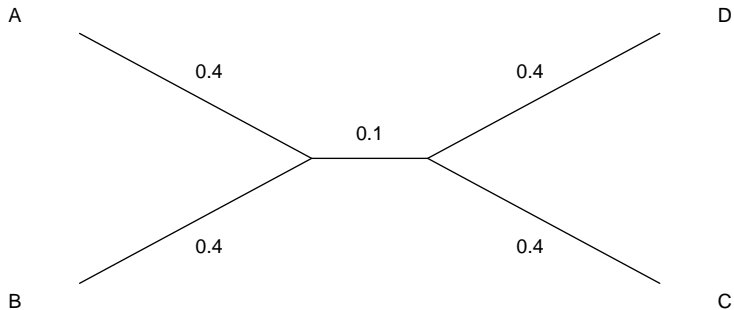
What is Model Misspecification?



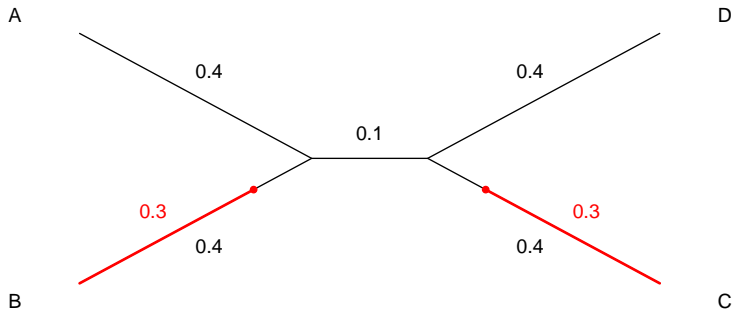
What is Model Misspecification?



What is Heterotachy?

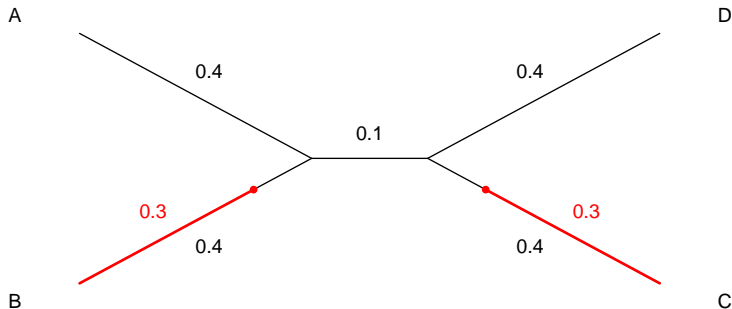


What is Heterotachy?

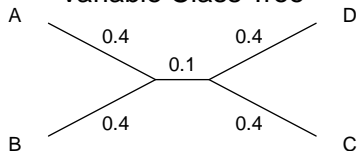


Event Tree vs Class Trees

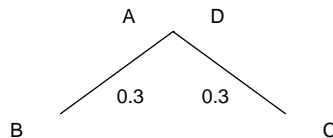
Event Tree



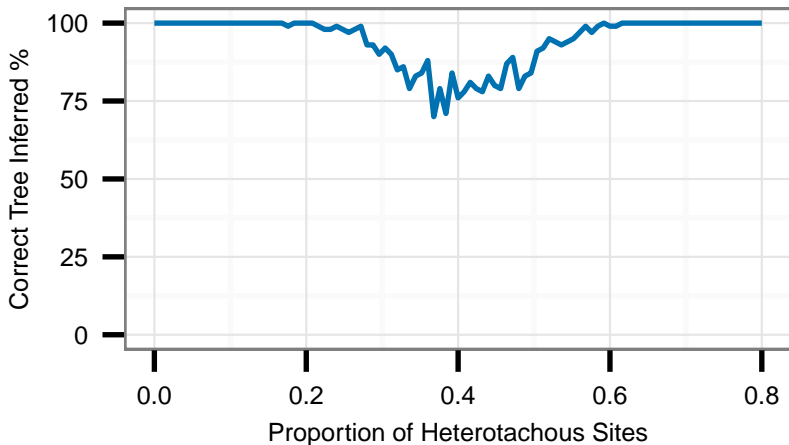
Variable Class Tree



Heterotachous Class Tree



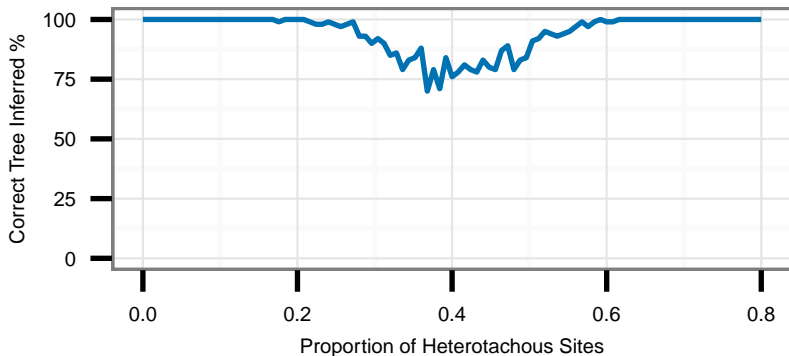
Early simulation results - Maximum Likelihood



Why the apparent recovery?

- The simulated data contained 3 classes of sites - variable, heterotachous and invariable.
- The proportion of variable sites was held constant.
- If the proportion of heterotachous sites is increased then the proportion of invariable sites is decreased, and vice versa.
- The inference was carried out assuming a JC model of evolution.

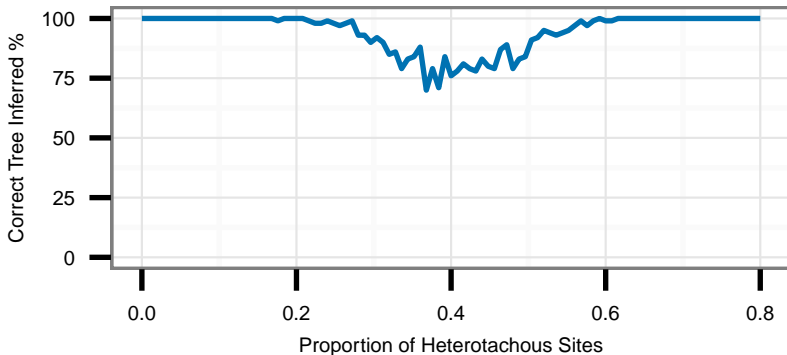
Competing misspecifications



Model Misspecification



Competing misspecifications



Misspecification due to Heterotachy



Misspecification due to Invariable Sites



Untangling the web

- ① Construct an asymptotic dataset, free of any stochastic variation.
- ② Construct models of evolution, each specifically designed to elicit a particular type of misspecification.
- ③ Carry out ML inference under each model and analyse results.

Constructing the asymptotic dataset

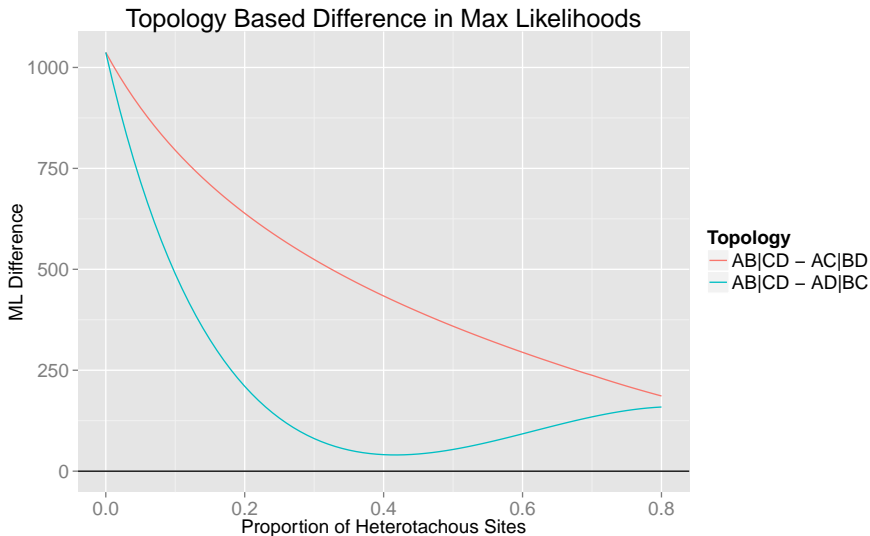
- 1 For each class tree calculate S , the asymptotic site pattern frequency vector for a given tree and model of evolution.
- 2 To find S for the event tree, take the weighted sum of the S vectors of each class tree.

$$S = p_{var} S_{var} + p_{het} S_{het} + p_{inv} S_{inv}$$

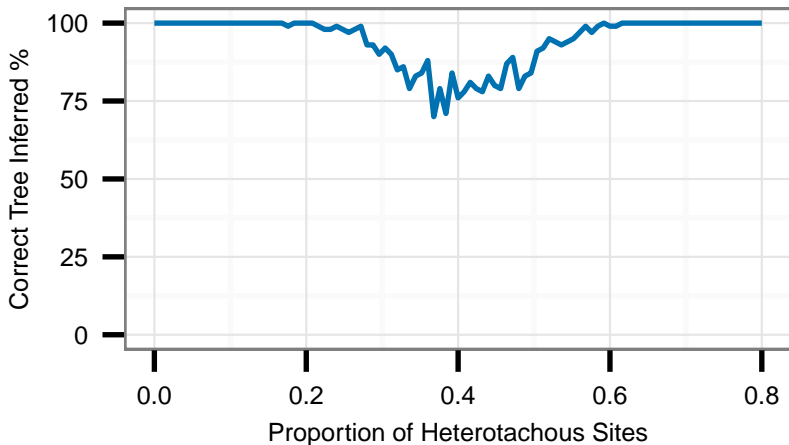
Models of Evolution

Model	Misspecification due to:
JC+I+H	No misspecification present
JC+H	Invariable sites only
JC+I	Heterotachous sites only
JC	Heterotachous and invariable sites

Results - JC Model



Early simulation results - Maximum Likelihood



Results - Inferred Tree under JC

Results - Inferred Tree under JC+I+H

Results - Inferred Tree under JC+H

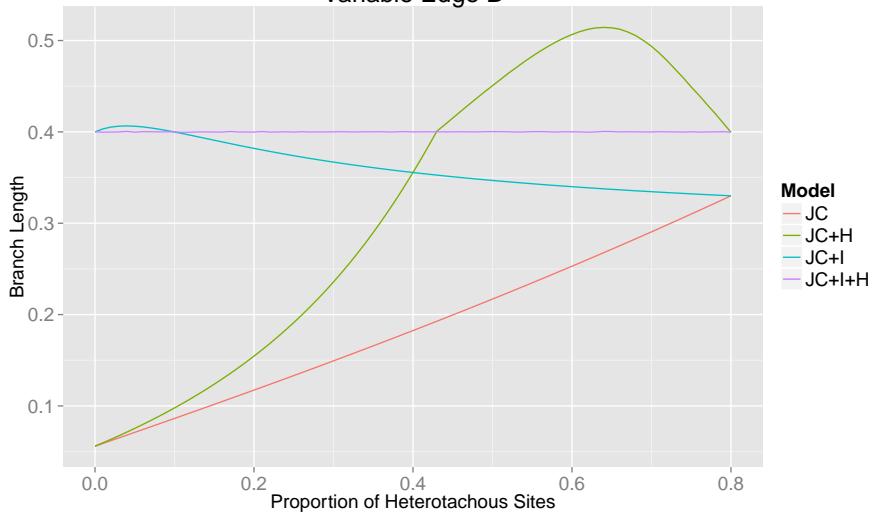
Results

Heterotachous Edge B

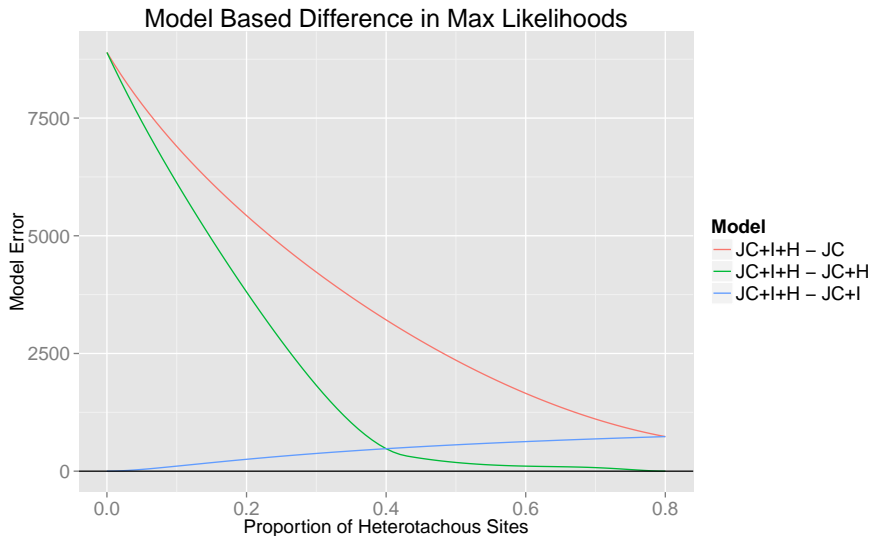


Results

Variable Edge B



Results



Future Work

- Develop a software package to map an event tree to a set of class trees and vice versa.
- Extend the optimisation algorithm to infer not only branch lengths, but also proportions of sites in each class.
- Apply the JC+I+H model to a real dataset.

Acknowledgements

I would like to thank my supervisory team for their input and guidance:

- Prof. Nigel Bean - University of Adelaide
- Dr Lars Jermiin - CSIRO
- Dr Barbara Holland - University of Tasmania
- Dr Jono Tuke - University of Adelaide



Shining a light on the situation

- There is a package available in R called Shiny
- The package allows the user to build their own web based applications.
- I have created an interactive Shiny App to facilitate easy interrogation of my results.
- You are most welcome to access it and have a play, I would appreciate any feedback.
- The web address is
<https://stephencrotty.shinyapps.io/Phylomania2014/>