

BAMM gives misleading rate estimates in simulated and empirical datasets

Andreas L. S. Meyer,^{1,2} Cristian Román-Palacios,² and John J. Wiens^{2,3} 

¹Graduate Program in Zoology, Universidade Federal do Paraná, Curitiba, PR 81531–990, Brazil

²Department of Ecology and Evolution, University of Arizona, Tucson, Arizona 85721

³E-mail: wiensj@email.arizona.edu

Received April 25, 2018

Accepted July 20, 2018

In a previous paper, we used simulations and empirical data to show that BAMM (Bayesian Analysis of Macroevolutionary Mixtures) can give misleading estimates of rates and rate shifts. In simulations, BAMM underestimated rate shifts across every tree analyzed, and assigned incorrect rates to most clades in most trees. In empirical analyses, BAMM behaved as expected from simulations, and assigned different rates to clades when clades were analyzed alone versus across the tree (i.e., with rate heterogeneity). Rabosky recently criticized our paper, focusing primarily on the idea that our comparison of BAMM to another approach (method-of-moments estimators of Magallón and Sanderson, or MS estimators) was unfair to BAMM. Here, we provide further evidence that BAMM gives misleading rate estimates in empirical studies. We then describe how Rabosky's own method comparisons were either acknowledged as being problematic or were described inaccurately (to favor BAMM). Finally, we show that the MS estimators can perform well when rates vary over time, despite untested assertions that they require constant rates to be accurate. Many other methods are available for analyzing diversification rates: we argue that BAMM should be avoided for estimating both diversification rates and rate shifts.

KEY WORDS: BAMM, diversification, extinction, macroevolution, simulations, speciation.

Analyzing diversification rates among clades can offer insights into the factors that underlie species richness patterns and the processes that cause speciation and extinction. Numerous methods have been developed to estimate these rates (review in Morlon 2014). However, the accuracy of most methods remains poorly understood.

Meyer and Wiens (2018; MW hereafter) addressed the accuracy of two methods for estimating diversification rates, focusing especially on BAMM (Rabosky 2014). BAMM has become widely used to estimate both rate shifts across trees and rates of individual clades. MW showed that BAMM performed poorly, by several criteria. Most conspicuously, there were relatively weak relationships between true and estimated rates among clades across simulated trees. This occurred because BAMM severely underestimated the number of rate shifts across each tree. Each tree had 10 clades, and each clade had a distinct speciation rate, extinction rate, and diversification rate (speciation–extinction). However, BAMM estimated (on average) only two changes in

diversification rates across each tree. Two rate shifts would be consistent with only three different diversification rates across each tree of 10 clades. Thus, BAMM assigned incorrect rates to most clades in most trees. The estimated rates were often quite different from the true rates. However, MW showed that BAMM improved when applied to single clades with no variation in diversification rates (MW: Fig. 2). MW also demonstrated that these problems applied to empirical data. Specifically, when applied across 15 snake families (MW: Fig. 6), BAMM sometimes estimated very different rates for the same families depending on whether BAMM was applied across the whole tree or to clades in isolation. These mismatches supported predictions from the simulations.

MW also compared BAMM to another method for estimating diversification rates (method-of-moments estimators; MS estimators hereafter: Magallón and Sanderson 2001). MW showed that the MS estimators performed better than BAMM when BAMM was used as in empirical studies (to estimate rates across trees

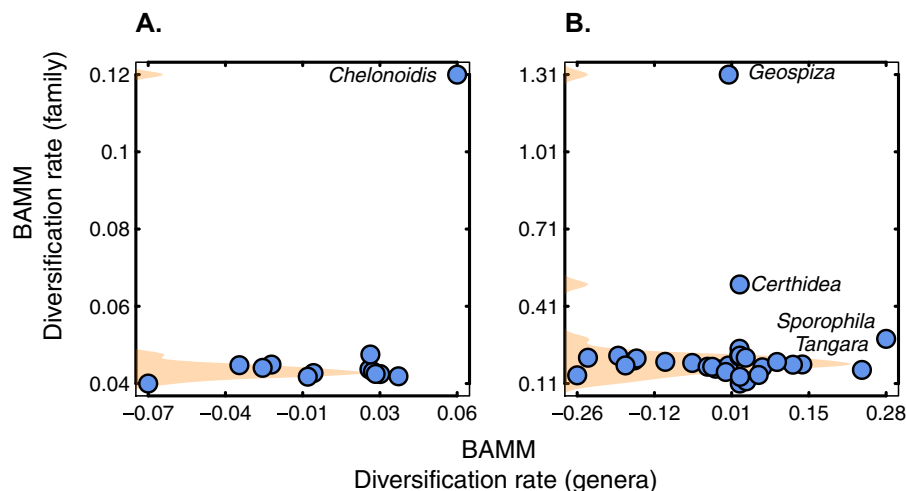


Figure 1. Comparison of two sets of diversification rate estimates from BAMM for (A) genera of tortoises (Testudinidae), and (B) genera of tanagers and relatives (Thraupidae). On the y-axis, diversification rates are estimated when BAMM is applied across the entire family-level phylogeny (and allowed to estimate rate shifts, as normally applied in empirical studies), and estimated diversification rates are then extracted for each genus. On the x-axis, diversification rates are estimated for each genus separately, treating that genus as the entire tree (this is not how BAMM is usually applied, but appears to be more accurate in simulations). The two approaches disagree strongly about what the actual diversification rates are for each genus. Yet, both estimates are from BAMM. Therefore, BAMM must be estimating incorrect diversification rates.

with heterogeneous diversification rates; MW, Table 1). MW also showed that the accuracy of the MS estimators was not contingent on rates being homogeneous or constant within clades (MW: Fig. 3).

Rabosky's (2018; R2018 hereafter) response to MW is problematic on several fronts. Most importantly, showing that BAMM can give similar accuracy to the MS estimators under the simplest conditions (no rate variation) does not solve the problem of BAMM being inaccurate under more realistic conditions. In the first section, we show empirical results that further prove that BAMM gives inaccurate rate estimates in the real world. In our second section, we address the comparison of BAMM and the MS estimators. For example, we will show that R2018 failed to perform a valid comparison of these methods, and that some results in R2018 were described inaccurately (to favor BAMM). Finally, we will further test the idea that the MS estimators require constant diversification rates to be accurate.

BAMM Gives Misleading Results in Empirical Studies

MW introduced an approach to evaluate the accuracy of BAMM in empirical studies. They compared diversification rates estimated for individual clades when estimated across an entire tree (with many clades) relative to rates estimated for those clades in isolation. In theory, rate estimates for clades in isolation should be more accurate, since they should be less prone to the problem

of underestimating rate shifts across a large and heterogeneous tree. Most importantly, if the two sets of rates differ, then BAMM must be giving inaccurate rate estimates (because two different rates for the same clade cannot both be correct). MW applied this approach to 15 snake families, and found a nonsignificant relationship between the two sets of rates. This weak relationship was driven primarily by two species-poor clades, which had the lowest diversification rates in isolation, but intermediate rates when BAMM estimated rates across the entire tree. The differences between estimates were striking (Pareatidae: whole-tree estimate = 0.10494 species/Myr: individual-clade = 0.02276; Sybonophiidae: whole-tree = 0.09816; individual-clade = 0.00371).

Here, we present similar results from two additional datasets: thraupid birds (tanagers) and testudinid turtles (tortoises). These analyses are based on well-sampled time-calibrated trees for each family, which were analyzed for a study on island diversification (Román-Palacios and Wiens 2018). We summarize here the supplementary analyses from that study, which used BAMM. For methods and results, see Appendix S1 here (trees in Supplementary Files S1, S2). We first estimated diversification rates for each genus within each family based on the whole family-level phylogeny (allowing BAMM to estimate the placement of rate shifts). We then analyzed each genus separately, treating each genus as the entire tree.

For tortoises, when BAMM is applied across all 13 genera simultaneously, BAMM estimates two rate regimes (Fig. 1A): a high diversification rate for *Chelonoidis* (0.125 species/Myr)

Table 1. Full results from MW comparing the performance of BAMM and the MS estimators, when BAMM is applied to single clades without rate shifts.

Method	Epsilon	Taxon sampling	Mean absolute error (%)	Proportional error (bias, %)	r^2 (true and estimated rates)
BAMM	estimated	100%	27.8	-10.1	0.70
MS-crown	0	100%	47.1	44.1	0.73
MS-crown	0.5	100%	59.2	57.9	0.77
MS-crown	0.9	100%	26.3	-15.8	0.87
MS-stem	0	100%	50.2	44.0	0.81
MS-stem	0.5	100%	30.1	18.9	0.85
MS-stem	0.9	100%	32.8	-32.8	0.87
BAMM	estimated	50%	26.2	-7.4	0.66
MS-crown	0	50%	59.2	57.9	0.73
MS-crown	0.5	50%	47.1	44.1	0.77
MS-crown	0.9	50%	26.3	-15.8	0.87
MS-stem	0	50%	50.2	44.0	0.81
MS-stem	0.5	50%	30.1	18.9	0.85
MS-stem	0.9	50%	32.8	-32.8	0.87
BAMM	estimated	25%	22.5	-0.3	0.78
MS-crown	0	25%	67.0	65.7	0.63
MS-crown	0.5	25%	54.1	51.1	0.68
MS-crown	0.9	25%	29.4	-12.3	0.83
MS-stem	0	25%	50.2	44.0	0.81
MS-stem	0.5	25%	30.1	18.9	0.85
MS-stem	0.9	25%	32.8	-32.8	0.87

Compare to Table 1 of R2018. Note that only results based on the same data (same % species sampled) are actually comparable. R2018 suggested that these results show that BAMM outperforms the MS estimators, but given complete taxon sampling (100%), BAMM does not perform best by the preferred criterion of R2018 (mean absolute error). Instead, the best-performing MS estimator (crown, epsilon=0.9) outperforms BAMM under these conditions. Based on one of the preferred criteria of MW (r^2 for relationship between true and estimated rates), BAMM is the lowest-ranked method under all conditions, except when taxon sampling is only 25%. All results are from MW: BAMM results are from Figure 2 of MW, but also including results of two additional performance metrics (not included in Fig. 2) and including comparisons to the MS estimators for these same clades.

and low rates for all other genera (~0.05 species/Myr). When BAMM is applied to these clades (genera) in isolation, the results are strikingly different (Fig. 1A): *Chelonoidis* is only 0.058, and the other clades have a broad range of lower rates (0.038 to -0.070). There is almost no overlap between rate estimates for the same clades from the two analyses. The relationship between the two sets of rates is weak ($r^2 = 0.2389$, $P = 0.0901$) and would be effectively zero except that both agree that *Chelonoidis* has the highest rates.

For thraupids, the results are even more striking (Fig. 1B). When BAMM is applied across the entire family-level tree, *Geospiza* has a high rate (1.30), *Certhidea* an intermediate rate (0.50), and all other genera have lower rates (0.10–0.30). When rates are estimated for each clade (genus) separately, the results are again very different. Rates are highly variable but much lower, ranging from 0.28 to -0.25. Instead of *Geospiza* and *Certhidea* having the highest rates, the genera *Sporophila* and *Tangara* have

the highest rates. *Geospiza* and *Certhidea* have intermediate rates, surprisingly close to zero (both genera are Darwin's finches, famous for their rapid diversification: e.g., Schluter 2000; Coyne and Orr 2004). There is no significant relationship between the two sets of rates ($r^2 = 0.002$, $P = 0.8127$).

For both tanagers and tortoises, the disparate results between analyses are not simply caused by clades with very few species (Román-Palacios and Wiens 2018). For example, both *Chelonoidis* (tortoises) and *Geospiza* (tanagers) have >5 species each but have very different rates in the two analyses. Thus, even though rates may not be accurately estimated for very small clades in isolation, the problem is clearly more general.

Some readers might be surprised by how few rate regimes are estimated across these phylogenies. But the numbers of rate shifts are similar to those in simulations (mean = 2.3 per tree; MW). These results are also concordant with previous BAMM

analyses of these groups. Rodrigues and Diniz-Filho (2016) inferred only two rate regimes across all turtles, with all testudinids most frequently sharing an identical rate. Burns et al. (2014) also inferred only three rate regimes across all thraupids (i.e., two rate shifts).

Overall, these results show that BMM must be giving rate estimates that are incorrect in the real world: the two different sets of rates are very different, yet both are from BMM. In our simulations, rates from separately analyzed clades are most accurate. This seems plausible for snakes, but less clear for testudinids and thraupids. We think that any future studies that use BMM should demonstrate that similar problems do not apply to their analyses.

BMM versus the MS Estimators

COMPARING THE METHODS

MW compared BMM to the MS estimators, and found that the MS estimators performed well by several criteria. MW cautioned that MS estimators are not necessarily the best approach for estimating diversification rates. However, they can be applied when only clade ages and species richness are known, without a detailed phylogeny within each clade (which is necessary for most other methods, including BMM). MW tested the MS estimators under various conditions. Importantly, they found that they do not require homogeneous diversification rates within clades to be accurate (MW, Fig. 3), despite claims that they require “constant” rates within clades (e.g., Rabosky 2009; Rabosky and Adams 2012; Rabosky et al. 2012; Rabosky 2018). That is, the MS estimators performed well even when MW did not “give them” the location of rate shifts.

R2018 stated that the comparison between BMM and the MS estimators was unfair. Given this, one might expect a valid, fair comparison of BMM and the MS estimators from R2018. Figure 2 of R2018 appears to represent this comparison. Yet, after describing these results, R2018 stated “The preceding exercise is not offered as a serious comparison of BMM and MS estimators.” In these analyses, R2018 compared the proportional error of BMM and MS estimates for all clades with 10 or more species in a tree simulated by MW. Here are some reasons why we think these analyses are problematic: (1) R2018 only analyzed one of the 20 trees simulated by MW. Yet, R2018 criticized MW for only analyzing only “2%” of the clades that they could have analyzed. (2) The rate estimates for each clade are not equivalent between methods in this figure. The BMM estimates incorporate information from the entire tree, whereas rate estimates for the MS estimators are for each clade in isolation. A more fair comparison would be to estimate the diversification rate for each clade in isolation for both methods. Using the data from

Figure 2 of MW, we did so. This yields plots for BMM and the MS stem estimator that are almost indistinguishable (Fig. S1; but this ignores BMM’s problems with rate heterogeneity: MW, Table 1). (3) BMM was given two interlinked advantages, related to clade size and nonindependence of BMM estimates for individual clades. Given a single clade in isolation, both methods often perform more poorly with smaller clades (Figs. S1, S2). In Figure 2 of R2018, the MS estimators are forced to separately estimate rates for hundreds of smaller clades, whereas BMM is not. Thus, if BMM accurately estimates the rate regime for a larger clade, this same estimate is counted repeatedly for each of the smaller clades within it (even though these clades and rate estimates are not independent). Furthermore, larger clades are given more weight in determining overall accuracy (because each large clade contains many smaller clades that are counted repeatedly as separate clades). In contrast, MW equalized the contributions of large and small clades when assessing the overall accuracy of each method (and explicitly warned about this problem), and only compared rate estimates for separate, nonoverlapping clades. The goal should be to accurately estimate rates for both large and small clades. Overall, Figure 2 of R2018 seems designed to favor BMM.

A major conclusion of R2018 was that BMM performs better when it analyzes clades with no variation in diversification rates. The problem is that under more realistic conditions, with variable rates among clades, BMM performs poorly (MW; Moore et al. 2016). Nevertheless, R2018 presented his Table 1 as if it contradicted MW. But it shows exactly what one should infer from comparing Figures 1 and 2 of MW. Moreover, the description of the results in Table 1 of R2018 is inaccurate. First, the table legend states “Models are ranked from best-performing to worst-performing.” But this ranking conflates different methods with different datasets (i.e., datasets having different levels of taxon sampling). Only method results from the same dataset are directly comparable. Furthermore, the table excludes some data and methods, without explanation. Most importantly, there are no comparisons between the generally most accurate MS estimator (crown, $\epsilon = 0.9$) and BMM with 100% taxon sampling. When comparing these two methods on the same data (100% sampling), this MS estimator ranks higher than BMM (using R2018’s preferred ranking criterion). Moreover, given 50% sampling, this MS estimator is behind BMM by only 0.1%. Finally, Table 1 of R2018 shows that BMM performs best when 75% of the species in a clade are excluded. This seems problematic. Our Table 1 here shows all methods, grouped to only compare their accuracy on the same data. This table also shows relationships between true and estimated rates. BMM performs relatively poorly by this standard criterion, even under the simple (but unrealistic) conditions of constant rates, the conditions where BMM performs best.

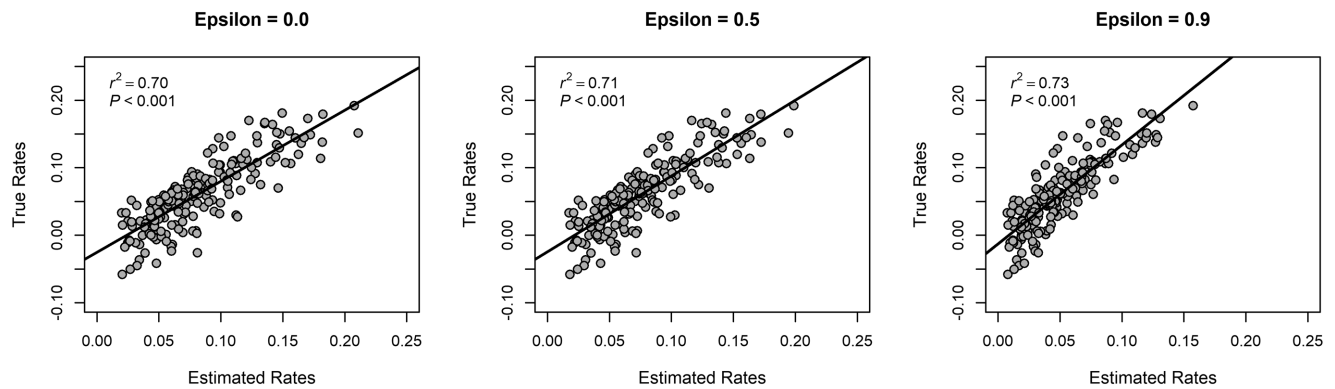


Figure 2. Strong relationships between true and estimated diversification rates for the crown-group MS estimators when rates are variable over time within clades. The results are based on five sets of simulations (40 replicates each, one clade per replicate) with: (A) linear change in speciation rates over time and constant extinction rates, (B) exponential change in speciation and constant extinction, (C) constant speciation and linear change in extinction rates over time, (D) constant speciation and exponential change in extinction, and (E) linear change in both speciation and extinction rates over time. Figure S5 shows the same relationships after excluding 150 replicates with the lowest variation in diversification rates over time.

10 POINTS ABOUT THE RESPONSE TO MEYER AND WIENS (2018)

In the section below, we briefly address numerous points made by R2018.

- (1) A major theme of R2018 is that MW gave the MS estimators “more information” than BAMM. In fact, the only information given to the MS estimators was each clade’s age and species richness. In contrast, BAMM was given the full species-level phylogeny within each clade (topology and branch lengths), and the correct taxon-sampling fraction. If MW gave BAMM the same limited information given to the MS estimators, BAMM would not work. Nevertheless, the MS estimators often yielded more accurate estimates of diversification rates than BAMM (MW: Fig. 1; Table 1).
- (2) R2018 stated that it was unfair that MW expected BAMM to detect rate shifts across each tree. But we think that everyone who uses BAMM expects that it will be able to automatically detect rate shifts (as stated in the title of the paper that introduced BAMM: Rabosky 2014). The results here and in MW suggest that BAMM does not detect rate shifts accurately.
- (3) Despite the impression given by R2018, MW did not simply perform one set of simulations in which they “gave” the location of the rate shifts to the MS estimators but not to BAMM. Instead, MW also performed analyses in which they explored the consequences of applying BAMM to clades with constant rates (MW: Fig. 2), and the MS estimators to clades with heterogeneous rates (MW: Fig. 3). These simulations showed that BAMM performed well with constant

rates, but that the MS estimators also performed well with heterogeneous rates.

- (4) R2018 argued against comparing methods based on statistical relationships between true and estimated rates. Yet, Rabosky (2014) suggested that BAMM was superior to MEDUSA (Alfaro et al. 2009), based (in large part) on BAMM’s stronger relationships between true and estimated rates. We think relationships between true and estimated rates are crucial. Specifically, we want to know if methods will estimate low rates when the true rates are low, and high rates when the true rates are high. MW showed that, across each tree, BAMM broadly underestimates the true variation in rates among clades, leading to weak relationships between true and estimated rates. Similar patterns appear to occur in empirical datasets (Fig. 1).
- (5) R2018 argued that analyzing higher taxa (e.g., families) is problematic, that comparing their diversification rates is misleading, and that there is no reason to simulate each higher taxon as having a single diversification rate. However, many previous papers by this author simulated diversification rates of higher taxa, each with a single rate (e.g., vertebrate orders: Rabosky 2009; ant families: Rabosky 2010; families of plants, animals, and fungi: Rabosky et al. 2012). These simulations followed the same design that R2018 criticized MW for using. Moreover, we have found that comparing diversification rates of higher taxa can offer insights into the traits that drive large-scale patterns of diversity, often revealing strong relationships between traits and diversification rates (e.g., Wiens 2015; Wiens et al. 2015; Bars-Closel et al. 2017; Jezkova and Wiens 2017; Moen and Wiens 2017). Furthermore,

diversification rates of higher taxa are generally strongly related to those of the clades within them (Scholl and Wiens 2016).

- (6) R2018 acknowledged that BAMM underestimates the number of rate shifts, but defended BAMM as being “conservative.” That is, it only assigns clades different rates if there is strong support for those differences. MW showed the downside of this: BAMM frequently assigns incorrect rates to clades. In the main results of MW, most clades in most trees had the wrong rates assigned to them (i.e., an average of three rate regimes were assigned among 10 clades, where each clade had a different rate regime). Most importantly, incorrect rates were assigned regardless of whether the estimated and true rates were similar (MW; Fig. S5). Furthermore, in 15% of 200 simulated clades, the true rates were outside the 95% posterior intervals of the estimated rates, even though in many replicates these intervals were extremely broad (MW; Fig. S3). Indeed, in some cases these intervals simultaneously ranged from strongly negative rates (−0.10 species/Myr or more) to very high positive ones (0.20 or more, even >0.40). Thus, “conservative” for BAMM includes giving misleading results with strong statistical support.
- (7) Another consequence of BAMM’s “conservativeness” (i.e., its failure to detect real variation in diversification rates among clades) is that analyses using BAMM may fail to find real relationships between diversification and other variables (e.g., traits, habitats, regions). We have observed results suggesting this pattern in our own analyses (Hutter et al. 2017). Given this, studies that have found negative results using BAMM (e.g., nonsignificant relationships) should be viewed with particular caution.
- (8) The simulation results of MW were actually consistent with Rabosky’s (2014) simulations testing the performance of BAMM. Specifically, his results showed that the accuracy of BAMM for estimating speciation rates, extinction rates, and rate shifts declined dramatically as the number of rate regimes simulated across each tree increased from 1 to 5. The simulation results of MW merely show what happens when the number of regimes increases to ~10.
- (9) As one of many tests, MW compared the frequency with which BAMM and the MS estimators correctly estimated which of two sister clades had the higher diversification rate (MW; Fig. 5). They showed that the MS estimators correctly determined which clade had the higher rate far more often than BAMM. Figure 4 of R2018 is intended to address these results. However, it cannot overturn them, since it does not compare the MS estimators and BAMM. Moreover, the point of Figure 4B is particularly unclear: it shows that the MS estimates are not always identical to the

correct diversification rate, but without any evidence that BAMM performs any better under the same conditions (see instead MW, Fig. 5).

- (10) The main conclusion of MW was not that the MS estimators are the best approach for estimating diversification rates, but rather that BAMM gives problematic results. Focusing on making a perfectly equivalent comparison seems like a distraction from the fundamental problems of BAMM. Moreover, the most equivalent comparison (Table 1) involves applying BAMM to clades with uniform, constant rates. Even under these unrealistic conditions, the relative performance of BAMM is entirely contingent on the performance criterion used (and excluding many species in each clade).

DO THE MS ESTIMATORS REQUIRE CONSTANT RATES?

A common theme in R2018 and numerous previous papers (e.g., Rabosky 2009, 2010; Rabosky and Adams 2012; Rabosky et al. 2012) is that the MS estimators require constant diversification rates within clades. Nevertheless, we are unaware of papers (besides MW) that tested whether variable rates within clades actually decrease the accuracy of these estimators. MW found that treating clades with different diversification rates as a single clade had little impact on the accuracy of the MS estimators, based on the deviance between true and estimated rates.

Here, we expand on those simulations in two ways. First, we clarify the initial results of MW by examining relationships between true and estimated rates. Second, we address (possibly for the first time) how variation in diversification rates within a clade over time impacts the accuracy of the MS estimators.

In Table 2 (see also Fig. S3), we summarize relationships between true and estimated rates when rates are homogeneous (constant) within clades and heterogeneous (two clades treated as one; as in Fig. 3 of MW). True rates for heterogeneous clades are defined based on the mean of the true rates for the two subclades, as in MW. The results for the crown-group MS estimators are as expected from Figure 3 of MW: heterogeneity has little impact on relationships between true and estimated rates. However, for the stem-group MS estimators, there is a dramatic decrease when rates are heterogeneous (from $r^2 = 0.70$ – 0.74 to $r^2 = 0.22$ – 0.23). This decrease is most likely explained by the extremely long stems associated with some sister clades (Fig. S1 of MW). Removing the 32 replicates with long stems (>20 Myr) increases accuracy considerably (e.g., $r^2 = 0.47$ – 0.59 ; $n = 34$; Fig. S4). Importantly, our index for the “true” rates here does not incorporate these long, unbranched stems, even though these stems are included in rate estimation. Thus, much of the mismatch between “true” and estimated rates here may not be error at all. This should be explored further. However, simple heterogeneity in rates between

Table 2. Impact of rate heterogeneity on MS estimators, based on relationships between true and estimated rates.

Method	Epsilon	r^2 (constant)	r^2 (heterogeneous)
MS-crown	0	0.61	0.67
MS-crown	0.5	0.64	0.67
MS-crown	0.9	0.73	0.63
MS-stem	0	0.71	0.23
MS-stem	0.5	0.74	0.23
MS-stem	0.9	0.70	0.22

These results are from the same analyses as in Figure 3 of MW, which used estimates of deviance to assess accuracy (instead of relationships between true and estimated rates). Sixty-six pairs of sister clades were used. For constant rates, the diversification rate is estimated for each of the 132 clades separately, and the true diversification rate is the true speciation rate minus the true extinction rate. For heterogeneous rates, each of the 66 sister clades was treated as single a clade (despite having different diversification rates), and the true rate is based on the average of the simulated diversification rates of the two sister clades. These relationships are shown graphically in Figure S3. Note that the weaker relationships for the stem-group estimator seem to be strongly influenced by the long, unbranched stems linking pairs of clades: these stems are not included in the calculation of “true” rates, which may greatly increase the apparent error. Indeed, removing sister pairs with long stems (>20 Myr) greatly improves relationships for the stem estimator (Fig. S4).

subclades does not appear to be the problem (given the crown-group results).

We also performed new simulations in which diversification rates changed over time within clades. Data are given in Table S1, and trees in Supplementary File S3. We performed five sets of simulations (40 replicates each) with speciation and/or extinction rates changing linearly or exponentially over time: (1) constant extinction rates and linearly variable speciation rates, (2) constant extinction, exponentially variable speciation, (3) constant speciation, linearly variable extinction, (4) constant speciation, exponentially variable extinction, and (5) linearly variable speciation and extinction rates. The functions used are shown in Table 3. Trees were simulated using the *rbdtree* function in the R package APE version 4.1 (Paradis et al. 2004).

Speciation and extinction rates were chosen to be similar to those simulated by MW, but ensuring that diversification rates were positive for at least part of each simulation replicate. MW simulated speciation rates that ranged from 0.01 to 0.30 speciation events/Myr (million years). For the simulations with time-variable speciation rates and constant extinction rates, we selected random starting and ending speciation rates from 0.01 to 0.30 events/Myr (uniform distribution). We then selected a random but constant extinction rate, but limiting the rate so that it varied between 0.001% and 90% of the highest speciation rate during that

replicate (either starting or ending). For simulations with constant speciation and time-variable extinction rates, we selected random starting and ending extinction rates from 0.001 to 0.27 events/Myr. We then selected a random but constant speciation rate that was at least 10% higher than the lowest extinction rate (either starting or ending), but without exceeding 0.30 events/Myr. For time-variable speciation and extinction rates, we selected random starting and ending speciation rates that varied from 0.01 to 0.30 events/Myr. We then selected random starting and ending extinction rates, between 0.001 and 0.27 events/Myr. However, we again constrained the rates so that either starting or ending extinction rates were smaller than the highest speciation rate. The rate at which speciation and extinction rates changed over time was determined by the parameter x , which was calculated using simulated starting and ending rates and the equations in Table 3. For all groups, the age of each tree was randomly selected from a uniform distribution from 15 to 100 million years, similar to the ages used by MW for the constant-rate clades. Analyses were carried out using R version 3.4.4 (R Core Team 2018).

We estimated net diversification rates for each tree using crown-group MS estimators. Given that diversification rates were not constant over time, we did not delete a basal clade to ensure that there was a stem-group age for each tree, as in MW. Following standard practice, we used three relative extinction fractions (epsilon): low (0), intermediate (0.5), and high (0.9). Diversification rates were estimated using the R package GEIGER version 2.0.6 (Harmon et al. 2008; Pennell et al. 2014).

Given the variability in rates over time, assessing the correct rate is less straightforward than if rates are constant. Here, we separated each tree into 100 evenly-spaced time segments, starting from the crown age of each tree. Since each segment has a specific value of speciation and extinction, we calculated the diversification rate for each segment as the difference between the speciation and extinction rates for that segment. To estimate the true net diversification rate for each tree, we calculated the mean of all 100 segments. This should reflect the overall diversification rate for the tree. The mean over time should be preferable to estimating the true rate based on a mean across all branches, which would bias the true rate to favor those time segments that generated the highest richness (i.e., those times with higher diversification rates).

We then assessed the accuracy of the MS estimators. First, we tested the relationship between true and estimated diversification rates across all 200 trees for each epsilon value, using linear regression. We also calculated the mean percentage deviance between the true and estimated rates across all 200 trees, repeating this for each relative extinction fraction. Finally, we calculated the mean absolute proportional error between the true and estimated rates.

Table 3. Processes and functions used to simulate the 200 trees used in our analyses of time-variable diversification rates.

Process		Function	
Speciation	Extinction	Speciation	Extinction
Linear	Constant	$\lambda(t_i) = \lambda_0 + xt_i$	$\mu(t_i) = \mu_0$
Exponential	Constant	$\lambda(t_i) = \lambda_0 \exp(xt_i)$	$\mu(t_i) = \mu_0$
Constant	Linear	$\lambda(t_i) = \lambda_0$	$\mu(t_i) = \mu_0 + xt_i$
Constant	Exponential	$\lambda(t_i) = \lambda_0$	$\mu(t_i) = \mu_0 \exp(xt_i)$
Linear	Linear	$\lambda(t_i) = \lambda_0 + xt_i$	$\mu(t_i) = \mu_0 + xt_i$

$\lambda(t_i)$ is the speciation rate at time t_i , $\mu(t_i)$ is the extinction rate at time t_i , λ_0 is the initial speciation rate, μ_0 is the initial extinction rate, and x is the rate change parameter.

These analyses (Fig. 2) showed strong relationships between true and estimated rates among clades ($r^2 = 0.70$ – 0.73), even though diversification rates were never constant over time within clades. Indeed, constant rates yield similar relationships ($r^2 = 0.66$ – 0.79 , with full taxon sampling; MW: Table 1). The results remained similar when we included only the 50 clades with the most variable diversification rates over time ($r^2 = 0.64$ – 0.67 ; Fig. S5).

The mean deviance and absolute error could be high (Fig. S6). The deviance was associated with a bias toward overestimating diversification rates. Indeed, true rates could be negative, but the MS estimators only estimate positive rates. However, the overall deviance appeared to reflect some outliers, and removing results outside the 95% quantile showed more limited bias and error (Fig. S7). Most importantly, the MS estimators estimate low rates for clades with true low rates, and high rates for clades with true high rates (Fig. 2).

Overall, these results show that the MS estimators can be robust to variation in diversification rates within clades, both among subclades and over time. Thus, the MS estimators do not require constant rates to be accurate.

Finally, many readers may ask: why not analyze these clades with BAMM? Our goal here was only to evaluate whether MS estimators require constant rates. Furthermore, we doubt that BAMM would greatly outperform the MS estimators under these conditions. For example, the relationship between true and estimated diversification rates for MS estimators with changing rates is similar to that of BAMM with constant rates ($r^2 = 0.70$; Table 1). Previous simulations suggest that BAMM will perform worse with time-variable rates than with constant rates (e.g., Rabosky 2014).

SUMMARY

Here, we address the accuracy of BAMM and the MS estimators, responding to a paper by Rabosky (2018) that addressed our earlier study (Meyer and Wiens 2018). We provide additional evidence that BAMM generates misleading diversification-rate estimates

in the real world (i.e., different BAMM estimates for the same clade give strongly inconsistent results). We show that the main analyses of R2018 in which BAMM and the MS estimators were compared do not support the conclusion that BAMM is more accurate, and some conclusions were based on inaccurate reporting of results. We show that the MS estimators do not require constant diversification rates to be accurate. Overall, the MS estimators remain useful for estimating diversification rates (especially given limited phylogenetic information within clades), but we assume better methods will replace them soon. Regardless, we strongly caution against using BAMM in empirical studies.

AUTHOR CONTRIBUTIONS

ALSM, CRP, and JJW designed the study, performed analyses, and wrote the paper.

ACKNOWLEDGMENTS

J.J.W. was supported by U.S. National Science Foundation Grant DEB 1655690. We thank C. Ané and M. Noor for helpful comments that greatly improved the article.

DATA ARCHIVING

All data are included in the Supplementary Materials

LITERATURE CITED

- Alfaro, M. E., F. Santini, C. D. Brock, H. Alamillo, A. Dornburg, G. Carnevale, D. L. Rabosky, and L. J. Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc. Natl. Acad. Sci. USA* 106:13410–13414.
- Bars-Closel, M., T. Kohlsdorf, D. S. Moen, and J. J. Wiens. 2017. Diversification rates are more strongly related to microhabitat than climate in squamate reptiles (lizards and snakes). *Evolution* 71:2243–2261.
- Burns, K. J., A. J. Shultz, P. O. Title, N. A. Mason, F. K. Barker, M. L. Scott, and I. J. Lovette. 2014. Phylogenetics and diversification of tanagers (Passeriformes: Thraupidae), the largest radiation of Neotropical songbirds. *Mol. Phylogenet. Evol.* 75:41–77.
- Coyne, J. A., and H. A. Orr. 2004. *Speciation*. Sinauer Associates, Sunderland, MA.
- Harmon, L. J., J. Weir, C. C. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24:961–964.

- Hutter, C. R., S. M. Lambert, and J. J. Wiens. 2017. Rapid diversification and time explain amphibian species richness at different scales in the Tropical Andes, Earth's most biodiverse hotspot. *Am. Nat.* 190:828–843.
- Jezkova, T., and J. J. Wiens. 2017. What explains patterns of diversification and richness among animal phyla? *Am. Nat.* 189:201–212.
- Magallón, S., and M. J. Sanderson. 2001. Absolute diversification rates in angiosperm clades. *Evolution* 55:1762–1780.
- Meyer, A. L. S., and J. J. Wiens. 2018. Estimating diversification rates for higher taxa: BAMM can give problematic estimates of rates and rate shifts. *Evolution* 72:39–53.
- Moën, D. S., and J. J. Wiens. 2017. Microhabitat and climatic-niche change explain patterns of diversification among frog families. *Am. Nat.* 190:29–44.
- Moore, B. R., S. Höhna, M. R. May, B. Rannala, and J. P. Huelsenbeck. 2016. Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. *Proc. Natl. Acad. Sci. USA* 113:9569–9574.
- Morlon, H. 2014. Phylogenetic approaches for studying diversification. *Ecol. Lett.* 17:508–525.
- Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, M. E. Alfaro, and L. J. Harmon. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 15:2216–2218.
- R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rabosky, D. L. 2009. Ecological limits on clade diversification in higher taxa. *Am. Nat.* 173:662–674.
- . 2010. Primary controls on species richness in higher taxa. *Syst. Biol.* 59:634–645.
- . 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS One* 9:e89543.
- . 2018. BAMM at the court of false equivalency: a response to Meyer and Wiens. *Evolution* (in press).
- Rabosky, D. L., and D. C. Adams. 2012. Rates of morphological evolution are correlated with species richness in salamanders. *Evolution* 66:1807–1818.
- Rabosky, D. L., G. J. Slater, and M. E. Alfaro. 2012. Clade age and species richness are decoupled across the eukaryotic tree of life. *PLoS Biol.* 10:e1001381.
- Rodrigues, J. F. M., and J. A. F. Diniz-Filho. 2016. Ecological opportunities, habitat, and past climatic fluctuations influenced the diversification of modern turtles. *Mol. Phylogenet. Evol.* 101:352–358.
- Román-Palacios, C., and J. J. Wiens. 2018. The Tortoise and the Finch: testing for island effects on diversification using two iconic Galápagos radiations. *J. Biogeogr.* 45:1701–1712.
- Schluter, D. 2000. *The ecology of adaptive radiation*. Oxford Univ. Press, New York, NY.
- Scholl, J. P., and J. J. Wiens. 2016. Diversification rates and species richness across the Tree of Life. *Proc. Biol. Sci.* 283:20161335.
- Wiens, J. J. 2015. Explaining large-scale patterns of vertebrate diversity. *Biol. Lett.* 11:20150506.
- Wiens, J. J., R. T. Lapoint, and N. K. Whiteman. 2015. Herbivory increases diversification across insect clades. *Nat. Commun.* 6:8370.

Associate Editor: C. Ané

Handling Editor: Mohamed A. F. Noor

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Detailed methods for analyses of tortoises and tanagers using BAMM.

Appendix S1, Table 1. BAMM-based estimates of speciation, extinction, and diversification for each of the analyzed tortoise genera. Rates were calculated using either the whole phylogeny (family-level phylogeny) or each genus in isolation (genus-level phylogeny). Diversification rates for each genus were estimated as the mean speciation rate minus mean extinction rate.

Appendix S1, Table 2. BAMM-based estimates of speciation, extinction, and diversification for each of the analyzed tanager genera. Rates were calculated using either the whole phylogeny (family-level phylogeny) or each genus in isolation (genus-level phylogeny). Diversification rates (Div.) for each genus were estimated as the mean speciation rate minus mean extinction rate.

Table S1. Data for analyses of time-variable diversification rates.

Figure S1. Plots of absolute error in diversification rate estimation (absolute values of proportional error) relative to clade size (number of species per clade) and clade age. These results are based on Table S5 of MW (the first three trees from MW, 30 clades in total). Compare to Fig. 2 of R2018. Note that for all methods, each clade is independent and is analyzed in isolation. In contrast, in Fig. 2 of R2018, estimates for BAMM are based on the entire tree, and rate estimates for each clade are not independent (whereas the MS estimators are forced to estimate rates for each clade in isolation). The epsilon is 0.5 for both MS estimators, following Fig. 2 of R2018 (note that the crown-estimator with epsilon of 0.5 performs poorly relative to 0.9; MW). Relationships between variables are not strongly linear, and log-transformed relationships are shown in Fig. S2.

Figure S2. Same as Fig. 1, except that variables are natural-log transformed to yield stronger relationships between absolute error in diversification rate estimation (absolute values of proportional error) and clade size (number of species per clade) and clade age.

Figure S3. Relationships between true and estimated diversification rates for the MS estimators, comparing results when sister clades are treated independently (constant rates within clades) and when sister clades are treated as a single clade (heterogeneous rates within clades). A total of 66 clades were used. The relationships are summarized in Table 2, and are based on data from MW.

Figure S4. Relationships between true and estimated diversification rates for the MS estimators, comparing results when sister clades are treated independently (constant rates within clades) and when sister clades are treated as a single clade (heterogeneous rates within clades), but removing 32 clades with long stem branches (>20 million years). Based on data from MW.

Figure S5. Relationships between true and estimated diversification rates for the crown group MS estimators when rates are variable over time. The results are the same as those in Fig. 2, after excluding 150 replicates with the lowest variation in diversification rates over time. Variation in diversification rates was determined by calculating the standard deviation of the mean used to infer the overall true diversification rate of each tree (i.e. the mean of the 100 evenly spaced time intervals; see Table S1).

Figure S6. Accuracy of crown-group MS estimators when rates are variable over time, based on mean percentage deviance and mean absolute error. The results are based on five sets of simulations (40 replicates each, one clade per replicate) with: (a) linear change in speciation rates over time and constant extinction rates, (b) exponential change in speciation and constant extinction, (c) constant speciation and linear change in extinction rates over time, (d) constant speciation and exponential change in extinction, and (e) linear change in both speciation and extinction rates over time.

Figure S7. Accuracy of crown-group MS estimators when rates are variable over time, based on mean percentage deviance and mean absolute error. The results are the same as those in Fig. S6, except that those values outside the 95% quantiles have been removed.

Supplementary File S1. Time-calibrated tortoise phylogeny used in BAMM analyses.

Supplementary File S2. Time-calibrated thraupid phylogeny used in BAMM analyses.

Supplementary File S3. Simulated trees with time-variable diversification rates.