1 # Challenging the Raunkiaeran shortfall and the consequences of using
2 # imputed databases

3

4 Lucas Jardim[1]*, Luis Mauricio Bini[1], José Alexandre Felizola Diniz-Filho[1], Fabricio

5 Villalobos[1,2]

6 [1]Departamento de Ecologia, Universidade Federal de Goiás, Goiânia, Goiás, Brazil;

7 [2]Red de Biología Evolutiva, Instituto de Ecología, A.C., Carretera antigua a Coatepec

8 351, El Haya, 91070 Xalapa, Veracruz, Mexico

9 *Corresponding author: Lucas Jardim, Departamento de Ecologia, Universidade Federal

10 de Goiás, Goiânia GO. E-mail = lucas.ljardim9@gmail.com

11 Running title: *Missing data and trait imputed databases*

12 Word count: 7830

13

Summary

1. Given the prevalence of missing data on species' traits – Raunkiaeran shorfall — and its importance for theoretical and empirical investigations, several methods have been proposed to fill sparse databases. Despite its advantages, imputation of missing data can introduce biases. Here, we evaluate the bias in descriptive statistics, model parameters, and phylogenetic signal estimation from imputed databases under different missing and imputing scenarios.

2. We simulated coalescent phylogenies and traits under Brownian Motion and different Ornstein-Uhlenbeck evolutionary models. Missing values were created using three scenarios: missing completely at random, missing at random but phylogenetically structured and missing at random but correlated with some other variable. We considered four methods for handling missing data: delete missing values, imputation based on observed mean trait value, Phylogenetic Eigenvectors Maps and Multiple Imputation by Chained Equations. Finally, we assessed estimation errors of descriptive statistics (mean, variance), regression coefficient, Moran's correlogram and Blomberg's K of imputed traits.

3. We found that percentage of missing data, missing mechanisms, Ornstein-Uhlenbeck strength and handling methods were important to define estimation errors. When data were missing completely at random, descriptive statistics were well estimated but Moran's correlogram and Blomberg's K were not well estimated, depending on handling methods. We also found that handling methods performed worse when data were missing at random, but phylogenetically structured. In this case adding phylogenetic information provided better estimates. Although the error caused by imputation was

38   correlated with estimation errors, we found that such relationship is not linear

39   with estimation errors getting larger as the imputation error increases.

40   4. Imputed trait databases could bias ecological and evolutionary analyses. We

41   advise researchers to share their raw data along with their imputed database,

42   flagging imputed data and providing information on the imputation process.

43   Thus, users can and should consider the pattern of missing data and then look for

44   the best method to overcome this problem. In addition, we suggest the

45   development of phylogenetic methods that consider imputation uncertainty,

46   phylogenetic autocorrelation and preserve the level of phylogenetic signal of the

47   original data.

48

49   **Key-words**: bias, Multiple Imputation, trait databases, Phylogenetic Eigenvector

50   Maps, phylogenetic signal, Phylogenetic Comparative Methods.

51

Introduction

Missing data are a ubiquitous feature of real-world datasets (Nakagawa & Freckleton 2008). Lack of information may limit the application of statistical analysis and can lead to biased estimates and conclusions on the phenomena of interest. In 1976, Donald B. Rubin proposed a missing-data theory to allow analysis of incomplete datasets (Rubin 1976), explaining how unbiased parameters could be estimated with missing data by considering the mechanisms causing missing data. These mechanisms were classified in three categories: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). They mean, respectively, that missing values are equally probable across a dataset, probability of missing data is correlated with other variables rather than to the variable with missing data (target variable), and probability of missing data is itself correlated to the target variable and dependent on the missing data (Rubin 1976; Nakagawa & Freckleton 2008; Enders 2010; van Buuren 2012) (Fig.1).

When dealing with missing data, the above mechanisms need to be taken into account before analysis (Rubin 1976). This is because different methods that handle missing data assume different mechanisms, so using them indiscriminately may bias parameters estimates (Rubin 1976; Enders 2010; van Buuren 2012). Multiple Imputation and Full Information Maximum Likelihood methods are currently regarded as the most appropriate methods to handle missing data, because they work under MAR and MCAR scenarios and provide unbiased estimates (Enders 2010). In contrast, it is very difficult to model missing data under a MNAR scenario. This is so due to the need of considering a model that represents the probability of missing values to occur and because the shape of the probability density function is not known (Enders 2010; van Buuren 2012).

Research in ecology and evolutionary biology usually requires data about species and their traits to answer different questions from community assembly and

4

77   ecogeographical rules to correlated evolution, diversification rates and extinction

78   probability, among others (Purvis et al. 2000; Webb et al. 2002; Gaston et al. 2008;

79   Goldberg et al. 2010; Lukas & Clutton-Brock 2013; Jetz & Freckleton 2015). Thus, to

80   facilitate research and make it reproducible and data more accessible (Reichman et al.

81   2011), ecologists and evolutionary biologists usually create databases that include

82   information on huge amounts of species and their traits (e.g., (Jones *et al.* 2009; Kattge

83   *et al.* 2011; Wilman *et al.* 2014). However, as databases become larger, the probability of

84   having all the necessary data for all species rapidly decreases. This lack of knowledge

85   about species' traits and their ecological functions was recently defined as the

86   Raunkiaeran shortfall (Hortal et al. 2015) or Eltonian shortfall (Rosado *et al.* 2015).

87   Owing to the ubiquity of the Raunkiaeran shortfall, some researchers are

88   interested in filling such gaps in their databases for their own analyses but also to make

89   them available for other researchers (Swenson 2014; Schrodt *et al.* 2015). To do so, recent

90   studies suggest the use of phylogenetic information in the imputation process (Guénard

91   *et al.* 2013; Swenson 2014; Schrodt *et al.* 2015). Phylogenetic information is important

92   in imputation because closely related species resemble, on average, each other more than

93   distantly related species. Such phenomenon is commonly known as phylogenetic signal

94   (Blomberg *et al.* 2003). Consequently, knowing the phylogenetic position of species

95   could, in principle, be used to perform a good estimation of missing trait values. However,

96   the relationship between trait divergence and phylogenetic distance may be more complex

97   (due to distinct evolutionary models) than usually assumed (Hansen & Martins 1996;

98   Münkemüller *et al.* 2012). For instance, under Ornstein-Uhlenbeck evolutionary model

99   traits may evolve under selection restrictions where species track a trait optimum, causing

100   phenotypic resemblance even among phylogenetically distant species (Hansen & Martins

101   1996). Alternatively, under Early- burst model traits may show evolutionary rates early

102   in species history and later the rates slow down, resulting in phylogenetically closely

103   related species having different trait values (Blomberg *et al.* 2003; Harmon *et al.* 2010).

104   Finally, trait evolution may happen under a drift process (e.g., Brownian motion) where

105   species trait differences are directly correlated with time since divergence (Felsenstein

106   1985; Hansen & Martins 1996; Freckleton *et al.* 2002). Therefore, imputation methods

107   should explicitly consider or assume a trait evolutionary model determining the

108   relationship between species resemblance and phylogenetic proximity (Guénard *et al.*

109   2013).

110   Nowadays, large, imputed databases already exist that used taxonomic, ecological

111   or allometric relationships to fill in missing values (Jones *et al.* 2009; Wilman *et al.* 2014).

112   This highlights the need to critically evaluate the use of imputed databases given that the

113   reliability of statistical analysis under missing data is dependent on how much values

114   were missing in the original data, what mechanism caused data to be missing and which

115   methods were used in the imputation process (Schafer & Graham 2002; Enders 2010; van

116   Buuren 2012). Moreover, other problems can also arise when testing for phylogenetic

117   signal (Cavender-Bares *et al.* 2009; Münkemüller *et al.* 2012). In such cases, if analysis

118   were to be conducted on phylogenetically imputed data, results could be misleading given

119   that missing values would have been already filled based on their phylogenetic structure,

120   thus potentially inflating the level of phylogenetic signal. This potential issue can have

121   important consequences for studies evaluating, for example, niche conservatism, trait

122   lability, community assembly and diversification (Blomberg *et al.* 2003; Wiens &

123   Graham 2005; Cavender-Bares *et al.* 2009; Goldberg *et al.* 2010).

124   Considering the current need for complete databases and the use of imputation

125   methods to accomplish this, we evaluate how the estimation of descriptive statistics,

126   regression coefficients and phylogenetic signal can be misled by the percentage of

127   missing data, the particular mechanism of missing data, the model of trait evolution and

128   the choice of methods used to handle missing values. To accommodate all of these

129   scenarios, we use simulated phylogenies and traits under different combinations of such

130   conditions. In addition, to address imputation accuracy, we evaluated the relationship

131   between error caused by imputation and statistical estimation errors.

132

133   **Methods**

134   *Phylogeny simulation*

135   To evaluate the effect of imputing missing values into sparse databases (i.e. with

136   missing data), we first simulated 100 coalescent phylogenies with 200 species using the

137   function *rcoal* from the R package *ape* (Paradis *et al.* 2004). We focused on this

138   phylogeny size because it has been considered appropriate to evaluate power and

139   accuracy of phylogenetic analysis (Davis *et al.* 2013; Cooper *et al.* 2015), and it represents

140   a conservative approximation to database size (e.g. several hundreds to thousands of

141   species).

142

143   *Trait simulation*

144   For each phylogeny, we simulated two traits: a target trait and an auxiliary trait.

145   The first trait represented the one that would be imputed (i.e. missing-value trait), whereas

146   the second trait represented an auxiliary trait that would be used to impute values for the

147   target trait.

148   The target trait was simulated using the *rTraitCont* function from the *ape* package

149   (Paradis *et al.* 2004). We modeled this trait under a Ornstein-Uhlenbeck evolutionary

7

150   process (OU) (Gillespie 1996), because it allowed us to simulate trait evolution within a

151   continuum from evolutionary drift (i.e. Brownian motion) to weak and strong levels of

152   selection strength on trait evolution (Hansen & Martins 1996; Hansen 1997). Thus, we

153   could evaluate the performance of imputation methods under different levels of

154   phylogenetic signal. We fixed the target trait's optimum (Ɵ) to zero and the trait

155   interspecific variation (σ) equal to one. Also, we simulated different selection strengths

156   by varying α (selective strength) from 0 to 2, in 0.5 steps (0, 0.5, 1, 1.5 and 2). Such values

157   covered evolutionary scenarios from Brownian motion (OU α = 0) to strong selective

158   strength (OU α = 2).

159   The auxiliary trait represented a variable used to impute values into the target trait.

160   We simulated auxiliary traits in two ways: (i) correlated with the phylogeny and (ii)

161   correlated with the target trait but uncorrelated with phylogeny. For (i), we simulated the

162   trait following Liam Revell (pers. comm.):

163
$$x = ry + \sqrt{1 - r^2}\, MVN(0, \sigma^2 \textstyle\sum) \qquad\qquad \text{eqn 1}$$

164   where *y* is the target trait, *x* the auxiliary trait, and *r* the correlation coefficient between

165   both traits. We set *r* equal to 0.6 and 0.9 to explore the sensibility of our results to the

166   strength of trait correlation. $\sum$ is the species covariance matrix (Felsenstein 1985; Revell

167   *et al.* 2008) and σ² the target trait variation rate calculated as the mean of squared

168   phylogenetic independent contrasts (Freckleton & Jetz 2009), which was estimated using

169   the *pic* function from *ape* (Paradis *et al.* 2004). MVN means Multivariate Normal

170   Distribution and it was simulated using the *fastBM* function from the *phytools* R package

171   (Revell 2012). This auxiliary trait was later used when simulating the MCAR (Missing

172   Completely at Random) and MAR.PHYLO (Missing at Random correlated with

173   phylogeny) (see below).

174   For the second scenario, where the auxiliary trait is correlated with the target trait

175   but uncorrelated with phylogeny, the auxiliary trait was simulated using equation 1 with

176      $\sum$ having off-diagonal entries equal to zero (i.e. no covariance among species) and

177   diagonal entries representing, for each species, the sum of all branch lengths from the root

178   to the tip. We simulated MVN using the *mvrnorm* function in the R package MASS

179   (Venables & Ripley 2002). When using this auxiliary trait to impute target trait values,

180   we expected that using the phylogeny into the imputation methods would not improve

181   our analysis (i.e. provide no information on missing data) since the probability of missing

182   values would only be correlated with the auxiliary trait and not with the phylogeny.

183   *Missing data scenarios*

184   To create missing data, we used the target trait simulated above and deleted

185   different percentages of its values following three scenarios of missing data: Missing

186   Completely at Random (MCAR), Missing at Random but phylogenetically structured

187   (MAR.PHYLO), and Missing at Random but correlated with another phylogenetically

188   unstructured trait (MAR.TRAIT). We created the MCAR scenario by randomly sampling

189   a percentage (see below) of species along each phylogeny and replacing their trait values

190   with missing values. For the MAR.PHYLO scenario, we sampled a species in each

191   phylogeny and selected a percentage of its closest species to replace their trait values with

192   missing values, allowing a strong missing data pattern that was phylogenetically

193   structured. For the last scenario, MAR.TRAIT, we used the auxiliary trait (see above) to

194   replace values in the target trait. We ordered the values of the auxiliary trait in ascending

195   order and replaced the first percentage of values of the target trait with missing values.

196   This represented a missing data pattern correlated with another trait, different to the target

197   one. For each scenario, we simulated different percentages of missing values in the target

198   trait: 5, 10, 20, 50, 70 and 90% of missing data. These percentages were chosen to

199 represent common proportions of missing data present in highly used databases such as

200 PanTHERIA (Jones et al., 2009) and EltonTraits (Wilman *et al.* 2014) (Fig. S1, Appendix

201 S2).

202 *Imputation methods*

203   We evaluated four methods often applied by researchers to handle missing data:

204 imputation based on averaging values (MEAN), no imputation and simply deleting

205 missing values (LISTWISE), phylogenetic eigenvector maps (PEM), and multiple

206 imputation by chained equations (MICE).

207   We used the MEAN method to impute missing values by filling them with the

208 average of the observed values of the target trait. Under the LISTWISE method, we did

209 not impute values but simply deleted those species with missing values in the phylogenies

210 before the analyses. The PEM method uses both phylogenetic eigenvectors (Diniz-Filho

211 *et al.* 1998) and traits to impute data considering different OU processes (Guénard *et al.*

212 2013). We applied this method in two ways: first, using only the phylogenetic

213 eigenvectors (PEM.notrait) and, second, using these eigenvectors and the auxiliary trait

214 (PEM.trait). By applying the PEM method in these two ways allowed us to evaluate

215 whether phylogenetic information alone could impute data well or auxiliary traits were

216 necessary. Eigenvector selection and fitting of trait evolutionary models were performed

217 using the *MPSEM* R package (Guénard *et al.* 2013) using forward selection based on the

218 second-order Akaike Information Criterion. The MICE method simulates several possible

219 values for missing data from a posterior predictive distribution, then runs analysis and

220 pools results over all simulated data (van Buuren *et al.* 2006). We chose this method

221 because it is flexible and allows imputing categorical, continuous, and non-normally

222 distributed data (van Buuren *et al.* 2006). We applied MICE by creating 10 datasets to

223 run our analysis over them and pooled the results. The quantity of datasets created by

224   MICE is dependent on the percentage of missing data and more datasets can provide

225   higher accuracy and power in the analyses (Graham *et al.* 2007; Enders 2010; van Buuren

226   2012). However, because our objective was simply to estimate statistical bias instead of

227   inference power, 10 datasets can be considered appropriate (Graham *et al.* 2007). As with

228   the PEM method, we applied MICE in two ways: only considering the auxiliary trait

229   (MICE) and using this trait plus the phylogenetic eigenvectors selected as in PEM

230   (MICE.phylo). We imputed data with MICE using the *mice* R package (van Buuren &

231   Groothuis-Oudshoorn 2011).

232   We simulated 540 scenarios representing each combination of missing data

233   percentage, mechanism, OU selection strength, and imputation methods. For each

234   scenario, we simulated 100 replicates, thus producing 54000 independent results. Finally,

235   we averaged 10 replicates for each scenario and ended up with 5400 simulations to

236   analyze.

237   *Estimating Phylogenetic Signal*

238   We calculated the phylogenetic signal (PS) in our simulated phylogenies using

239   two metrics: Blomberg's K (Blomberg *et al.* 2003) calculated with the *phylosig* function

240   of *phytools* (Revell 2012) and Moran's *I* correlograms (Gittleman & Kot 1990; Diniz-

241   Filho 2001). For these correlograms, we created a phylogenetic distance matrix per

242   phylogeny using the *cophenetic* function of *ape* (Paradis *et al.* 2004) and built the

243   correlograms with the *lets.correl* function of the *letsR* R package (Vilela & Villalobos

244   2015). Then, based on the correlogram, we used the intercept of the following linear

245   model as indicative of PS:

246

247    $$\text{PS}_{\text{Moran}} = \alpha_1 - \frac{cov(\alpha, \beta)}{var(\beta)} * \beta \qquad \text{eqn 2}$$

11

248   where *cov* is the covariance between the mean within-class distance and Moran's Index,

249   *var* is the variance of the mean within-class distance, α is the value in each correlogram

250   distance class, and α_1 is the value in the first distance class.

251   *Imputation effects on phylogenetic signal*

252   To evaluate the effect of using imputed trait data for estimating phylogenetic

253   scenario. In particular, we estimated PS for the original target trait values before they

254   were deleted by the missing data mechanisms and estimated PS again after they were

255   filled by the imputation methods. Such PS delta was defined as:

256    $$PS_{delta} = (PS_{imputed} - PS_{original}) / PS_{original} \qquad\qquad eqn\ 3$$

257   where $PS_{imputed}$ is the PS calculated after imputing missing data, $PS_{original}$ is the observed

258   PS. If PS_{delta} is positive, there is a gain in PS (i.e. more PS than the original data), meaning

259   that imputing data increased the phylogenetic structure of the target trait. Conversely, if

260   PS_{delta} is negative, there is a decrease in PS (i.e. less PS than the original data), meaning

261   that imputing data decreased phylogenetic structure and made species' trait values seem

262   more phylogenetically independent than they originally were. PS_{delta} equal to zero

263   represents no change in trait phylogenetic structure (i.e. no imputation effect).

264   *Imputation effects on descriptive statistics*

265   Traditionally, performance evaluation of imputation methods have focused on

266   common descriptive statistics such as (mean, variance, regression coefficient) (Collins *et*

267   *al.* 2001; van Buuren *et al.* 2006; Penone *et al.* 2014) instead of phylogenetic patterns.

268   Therefore, we also evaluated the effect of imputed data on the estimation of such

269   descriptive statistics. We calculated the mean and variance of the target trait as well as

270   the regression coefficient (Ordinary Least Square) between the target trait and the

271   auxiliary trait, before producing missing data and after imputing such data. Next, we

12

272   measured the estimation error for these statistics as the mean squared error (MSE), as

273   below:

$$\text{MSE}_i = \frac{\sum_1^n (\tau 1 - \tau 0)^2}{n} \qquad \text{eqn 4}$$

274

275   where $\tau 1$ represents the statistics calculated over imputed traits, $\tau 0$ is the statistics

276   calculated from original traits, *n* means the number of simulations averaged to result ith

277   MSE value.

278   *Imputation error*

279   To measure the potential error introduced by imputation methods, that is the

280   deviation between imputed and original data, we followed Penone *et al.* (2014) and used

281   the normalized root mean squared error (NRMSE):

$$\text{NRMSE} = \sqrt{\frac{mean((y - y imputed)^2)}{\max(y) - \min(y)}} \qquad \text{eqn 5}$$

282

283   where *y* is the original trait value, $y_{imputed}$ is the imputed value, $\max(y)$ and $\min(y)$ are the

284   maximum and minimum values of the original trait, respectively. NRMSE varies between

285   0, no estimation error, and 1, maximum error (Oba *et al.* 2003).

286   *Overall analyses*

287   We were also interested on evaluating the effects of percentage of data missing,

288   missing data mechanism, OU selection strength, and imputation methods as factors

289   influencing the abovementioned effects of imputation (estimation errors: PS_{delta} and MSE

290   of descriptive statistics). To do so, we built linear models with these factors (e.g.

291   percentage of data missing) and their interactions as predictors and estimation errors,

292   separately, as individual responses. We specified the models using the *dredge* function

293   from the *MuMIn* R package (Bartón 2016) and ranked the different models using delta

13

294   AICc (Burnham & Anderson 2002). In addition, given concerns on the accuracy of

295   imputation methods (Guénard *et al.* 2013; Penone *et al.* 2014), we also evaluated the

296   relationship between imputation error (NRSME) and estimation errors (PS_{delta} and MSE)

297   caused by imputation. All simulations and analysis were run in R 3.2.2 (R Core Team

298   2015).

Results

300   In our simulations we found that differences in estimation errors were dependent

301   on missingness mechanism, imputation method, evolutionary model, percentage of

302   missing data and statistics being estimated. Moreover, imputation errors showed different

303   results between trait correlations (target vs. auxiliary trait; *r*) of 0.6 and 0.9, but

304   descriptive statistics and phylogenetic signal errors did not show different results

305   concerning this correlation. Therefore, we present all results for *r* = 0.6 and only those

306   for *r* = 0.9 corresponding to *Imputation error* (see below). Full results for *r* = 0.9 can be

307   found in the Appendix S1.

308   Not surprisingly, our results showed a clear tendency of increasing error in

309   estimating phylogenetic signal and descriptive statistics as the percentage of missing data

310   gets larger (Fig. 2). We did not identify a clear threshold in the amount of missing data

311   that would guarantee lower statistical errors. For the best imputation methods (MICE.phy,

312   PEM.trait, PEM.notrait; see below) lower errors were possible for as low as 30% and up

313   to 70% of missing data in the target trait.

314   When data were missing completely at random (MCAR), most imputation

315   methods showed good performance (Fig. 3-5; Fig. S2 and S3, in Appendix S2), except

316   the MEAN method. Nevertheless, when estimating Blomberg's K only LISTWISE and

317   PEM.trait resulted in low proportional changes (Fig. 3). For mean, variance and

14

318   regression coefficient MSE, imputation methods worked better when data were missing

319   at random but correlated with another trait (MAR.TRAIT) than when data were missing

320   and phylogenetically structured (MAR.PHYLO) (Fig. 5; Fig. S2 and S3, Appendix S2).

321   Nevertheless, the lowest proportional changes in Blomberg's K and PS_{Moran} were

322   observed under the MAR.PHYLO scenario (Fig. 3 and 4).

323   The level of selection strength on trait evolution under the OU process was also

324   important for the performance of imputation methods (Table 1). Accordingly, we found

325   a tendency PS_{delta} and MSE to decrease as the selection strength increased from pure

326   evolutionary drift (i.e. OU alpha = 0; Brownian motion) to strong selection (OU alpha =

327   2) (Fig. 3-5; Fig. S2 and S3, Appendix S2).

328   The less sensitive methods were those that considered phylogenetic information

329   in the imputation process (Fig. 3-5). PEM.trait, PEM.notrait, and MICE.phylo showed

330   results less sensitive over different mechanisms of missing data (Fig. 3-5; Fig. S2 and S3,

331   Appendix S2). From these three methods, PEM.trait was the less sensitive. The MEAN

332   method was the most sensitive, similarly to MICE under MAR.PHYLO scenario (Fig. 3-

333   5; Fig. S2 and S3, Appendix S2). The LISTWISE method caused the lowest changes in

334   Blomberg's K under all missing data mechanisms (Fig. 3) and for descriptive statistics

335   only under the MCAR mechanism (Fig. 5; Fig. S2 and S3, Appendix S2).

336   Phylogenetic signal metrics (Blomberg's K and PS_{Moran}) were lower than the

337   original (before imputation) when using MEAN and MICE methods (Fig. 3 and 4). All

338   other methods estimated PS_{Moran} correctly under most simulated scenarios (Fig. 4),

339   whereas the estimation of Blomberg's K showed different patterns (Fig. 3). Blomberg's

340   K was overestimated by PEM.trait and PEM.notrait and underestimated by MICE, even

341   under the MCAR missing mechanism (Fig. 3). Nevertheless, Blomberg's K estimation

342   errors decreased when phylogenetic eigenvectors were used in MICE.phylo (Fig. 3).

15

343   Descriptive statistics (mean, variance, and regression coefficient) were well

344   estimated by all imputation methods (except MEAN) under MCAR. MAR.TRAIT and

345   MAR-PHYLO generated biased estimations, but these biases were higher under MAR-

346   PHYLO (Fig. 5, Fig. S2 and S3, Appendix S2). Nonetheless, variance had high

347   estimations errors in MAR-PHYLO and MAR-TRAIT, independent of the imputation

348   methods (Fig. S3, Appendix S2). For all descriptive statistics, considering phylogenetic

349   structure improved estimations in MAR.PHYLO (Fig. S5, Fig. 2 and 3, Appendix S2).

350   Imputation error was lower when correlation (r) between the target and auxiliary

351   traits was 0.9 than r equal 0.6. When traits were moderately correlated ($r = 0.6$), the lowest

352   imputation error was found under missing completely at random (MCAR) scenarios (Fig.

353   S4, Appendix S2) and when using imputation methods that considered phylogenetic

354   information (PEM.trait, PEM.notrait, and MICE.phylo) (Fig. S4, Appendix S2).

355   Moreover, all imputation methods performed better under the MAR.TRAIT than under

356   MAR.PHYLO missing mechanism, but still poorly than under MCAR (Fig. S4, Appendix

357   S2). When traits were strongly correlated ($r = 0.9$), the MICE methods presented lower

358   imputation errors than when these traits were correlated moderately ($r = 0.6$). The

359   PEM.notrait method increased its imputation errors when trait correlation was strong (r

360   $= 0.9$) and PEM.trait was not influenced by correlation strength.

361   We found that estimation errors of descriptive statistics (MSE), Blomberg's K,

362   PS_{Moran} and imputation error (NRMSE) were influenced by all factors individually and

363   their interactions (Table 1). Despite some differences among selected models in respect

364   to the two- and three-way interactions, all models had interactions among all factors in

365   some level (Table 2). Finally, we found that imputation errors were correlated with

366   estimation errors (PS_{delta} and MSE) (Fig. 6). In addition, the imputation error and

367 estimation error relationship evaluated here was asymptotic in log-scale, thus as

368 imputation error increases the estimation error increases faster (Fig. 7).

369 **Discussion**

370   Ecologists and evolutionary biologists are increasingly creating, using, and

371 sharing large trait databases that are inevitably sparse and often completed by imputing

372 missing values (Guénard *et al.* 2013; Swenson 2014; Schrodt *et al.* 2015). Here we argue

373 that we should be extremely careful when using imputed databases, even for the

374 estimation of simple parameters (i.e. means, variances and regression coefficients). Our

375 findings revealed that estimations based on imputed data depends on every aspect of data

376 property and strategy of analysis, as percentage of missing data, source/mechanism of

377 absence, trait evolution, methods for gap filling, and statistics or parameters to be

378 estimated. This has commonly been acknowledged in statistical research (Rubin 1976;

379 Enders 2010) and should begin to be so in the ecological and evolutionary research as

380 claimed by Nakagawa & Freckleton (2008). Based on our results, we can infer that the

381 large changes in the estimations, due to different analytical choices, may also be an

382 important cause of irreproducibility in our field (Borregaard & Hart 2016).

383   The most pervasive obstacle for deriving conclusions from large datasets is simply

384 the proportion of those species lacking data. Previous studies found that reliable

385 estimations from imputed data can be made when up to 60% of the values were missing

386 (Barzi 2004; Penone *et al.* 2014). However, in our results, the effect of missing data

387 percentage was not direct, but rather interacted with all of the other aspects evaluated

388 here. Thus, there is no simple way of deriving a threshold on how much missing data

389 would be allowed to be imputed and still make reliable estimations.

390   Knowing the causes of data absence is the first issue to be sorted out before any

391   analysis (van Buuren 2012). The most common assumption in ecological and

392   evolutionary studies is that data is missing completely at random (MCAR). This is evident

393   in the wide variety of functions of the most commonly used software (the R programming

394   language) allowing deleting missing values indiscriminately. Indeed, if data were under

395   MCAR, previous findings and ours showed that estimations based on deletions and

396   imputations could safely be made (Nakagawa & Freckleton 2010; Penone *et al.* 2014;

397   Taugourdeau *et al.* 2014). However, biological data are rarely missing completely at

398   random (Nakagawa & Freckleton 2008; Enders 2010). For instance, bias in ecological

399   data absence can be related to the fact that some taxa are most studied than others

400   (Gonzalez-Suarez *et al.* 2012). Moreover, such bias can stem from body mass differences

401   among species, where large species have a higher probability of being described first

402   (Vilela *et al.* 2014) and have their data collected (Gonzalez-Suarez *et al.* 2012) compared

403   to small species. Also, species present in easily accessible regions are better studied than

404   those occurring in regions that are hard to access (Reddy & Dávalos 2003). In our

405   simulations, higher biased estimates were found when data were missing at random but

406   correlated with other variable (MAR), especially phylogeny (MAR.PHYLO). Such

407   results differ from those found by Penone *et al.* (2014), who did not find significant

408   estimation differences among missing data mechanisms. This discrepancy could be

409   related to our way of simulating MAR.PHYLO, creating a stronger phylogenetic structure

410   than that simulated by them.

411   Our simulations revealed that imputation methods considering phylogenetic

412   structure (PEM.trait, PEM.notrait and MICE.phylo) performed better than methods not

413   doing so (MEAN, LISTWISE, and MICE) under all missing data mechanisms (MCAR,

414   MAR.PHYLO, and MAR.TRAIT). Such findings support previous claims favoring

18

415 "phylogenetic imputation" as a powerful tool in predicting missing species values

416 (Penone *et al.*, 2014; Swenson 2014). More interestingly, our results showed that some

417 phylogenetic imputation methods (PEM.notrait) perform better than non-phylogenetic

418 ones, even when missing data was uncorrelated with phylogeny but to an auxiliary trait

419 (MAR.TRAIT). This result was unexpected based on missing data theory, which suggests

420 that under MAR.TRAIT some variable correlated with missing data probability is

421 required to guarantee reliable estimations (Enders 2010).

422   Overall, PEM.trait performed best among all imputation methods tested. A

423 potential caveat of this method is the imputation of a single value for each missing datum,

424 thus not accounting for uncertainty of the imputed value. Consequently, PEM.trait (or

425 PEM in general) may underestimate standard errors and bias subsequent hypothesis

426 testing (i.e. increasing Type I error rates) (Enders 2010; van Buuren 2012). To avoid such

427 biases, the statistical literature suggests using multiple imputation methods (Schafer &

428 Graham 2002; Enders 2010; van Buuren 2012). However, our results did not show better

429 performance of MICE, even when including phylogenetic information, in estimating

430 descriptive statistics or phylogenetic signal compared to PEM. Despite multiple

431 imputation being one of the most suggested methods for handling missing data (van

432 Buuren 2012), additional research is necessary to evaluate its performance with

433 phylogenetically structured data.

434   Filling missing values by averaging the observed ones (MEAN) or simply deleting

435 species with missing values (LISTWISE) generated poor estimates, which is related to

436 the fact that both methods assume that data is MCAR. MEAN only worked satisfactorily

437 for estimating the trait average. LISTWISE disrupts the distribution of trait values, thus

438 results in biased estimates (Enders 2010). However, this method performed well when

439 estimating phylogenetic signal. This is encouraging, given that researchers interested in

19

440   trait phylogenetic signal usually delete missing values (Blomberg & Garland 2002;

441   Kamilar *et al.* 2013) thus guaranteeing potentially unbiased results.

442   Phylogenetic imputation is based on the assumption of target traits being

443   phylogenetically structured (i.e. showing phylogenetic signal; Swenson 2014). However,

444   phylogenetic structure is dependent on how traits evolved (Diniz-Filho 2001; Guénard *et*

445   *al.* 2013). Accordingly, trait evolution was an important issue in our study. Across our

446   simulated scenarios, estimation errors were higher when target traits were simulated

447   under Brownian motion (BM) than under OU processes, agreeing with previous study

448   (Guénard *et al.* 2013). Better estimates under OU than BM processes may result from

449   higher trait resemblance and lower variance among species generated when increasing

450   selection strength under OU processes (Hansen 1997; Butler & King 2004). Thus,

451   predicting missing values of target traits will benefit from knowing their particular

452   evolutionary model and will be more accurate if such traits evolved under strong selection

453   regimes. Again, this suggests that researchers need to find the appropriate evolutionary

454   model for their target traits before judging the need to use phylogenetic imputation

455   methods for handling missing data. It should be noted, however, that fitting evolutionary

456   models over incomplete data could itself be biased owing to the use of observed values

457   only and thus pruned phylogenies (Slater *et al.* 2012).

458   Despite we showed phylogenetic imputation may recover descriptive statistics,

459   phylogenetic imputation methods may produce bias when estimating phylogenetic signal.

460   More specifically, our findings suggest that such methods can actually alter the original

461   phylogenetic structure of the trait (i.e. the structure if data were complete). In fact, PS

462   may be incorrectly estimated even under MCAR. Moreover, when using Blomberg's K,

463   imputation by PEM overestimated the original phylogenetic signal of the target trait (i.e.

464   created when the trait was simulated) whereas MICE.phylo underestimated it.

465   In addition, PS estimation errors were dependent on the evaluated metric.

466   Regardless of the simulated scenario, estimation errors were lower for PS based on

467   Moran's I correlogram than Blomberg's K. Similarly, Münkemüller *et al.* (2012) showed

468   that Moran's I is less sensible than Blomberg's K to changes in trait phylogenetic

469   structure even when random noise is added. Blomberg's K measures a global pattern

470   along a phylogeny, based on observed and expected total trait variance under Brownian

471   motion (Blomberg *et al.* 2003), whereas Moran's I correlogram measures the correlation

472   of trait values within different phylogenetic distance classes (Gittleman & Kot 1990).

473   Therefore, changes in total trait variance caused by imputation may not have strong

474   impacts on within-class correlations, rendering Blomberg's K more sensitive than

475   Moran's I to such changes.

476   New proposed methods to fill sparse databases currently concerns about their

477   degree of imputation error, that is how much imputed values deviate from the original

478   trait values (Guénard *et al.* 2013; Penone *et al.* 2014; Schrodt *et al.* 2015). We found that

479   single and multiple phylogenetic imputation methods can be highly accurate, resulting in

480   small deviations between imputed and observed values, as suggested by other authors

481   (Guénard *et al.* 2013; Penone *et al.* 2014; Diniz-Filho *et al.* 2015; Schrodt *et al.* 2015). In

482   addition, we found that imputation error was positively correlated with estimation errors

483   but their relationship was not linear. That is, increasing imputation error causes estimation

484   errors to increase much more rapidly. This is particularly relevant if researchers were to

485   use imputed databases blindly −without correctly treating imputed values. Such practice

486   could create spurious results. This is because even if imputation is accurate, imputed

487   values simply represent one among several possibilities without providing information

488   on imputation uncertainty. In fact, using an accurately imputed database does not

489   necessarily mean that the original trait distribution and its relationship with other

490   variables will be recovered (van Buuren 2012).

491   *Concluding remarks*

492   Instead of providing imputed trait databases, we should focus on treating missing

493   values with appropriate methods. We have shown here that such methods should consider

494   phylogenetic information. With the increase of computational literacy among ecologists

495   and evolutionary biologists (Ram 2013), we encourage researchers to use simulations of

496   their data and methods to find the appropriate solution for their study goals. Furthermore,

497   researchers need to develop phylogenetic methods that consider imputation uncertainty

498   and preserve the original data's phylogenetic signal. Missing data is one of the most

499   pervasive features of trait databases and the only effective solution for this Raunkiaeran

500   shortfall is collecting more data. Nevertheless, acknowledging such shortfall instead of

501   ignoring it will effectively help guiding research towards solving it.

507

508 **References**

509 Bartón, K. (2016). MuMIn: multi-model inference. R package version 1.15.6.

510 Barzi, F. (2004). Imputations of Missing Values in Practice: Results from Imputations
511 of Serum Cholesterol in 28 Cohort Studies. *American Journal of Epidemiology*,
512 **160**, 34–45.

513 Blomberg, S.P. & Garland, T. (2002). Tempo and mode in evolution: phylogenetic
514 inertia, adaptation and comparative methods. *Journal of Evolutionary Biology*, **15**,
515 899–910.

516 Blomberg, S.P., Garland, T. & Ives, A.R. (2003). Testing for phylogenetic signal in
517 comparative data: behavioral traits are more labile. *Evolution*, **57**, 717–745.

518 Borregaard, M.K. & Hart, E.M. (2016). Towards a more reproducible ecology.
519 *Ecography*, **39**, 349–353.

520 Burnham, K.P. & Anderson, D.R. (2002). *Model Selection and Multimodel Inference*,
521 2nd edn. Springer-Verlag, New York, NY.

522 Butler, M.A. & King, A.A. (2004). Phylogenetic Comparative Analysis : A Modeling
523 Approach for Adaptive Evolution. *The American Naturalist*, **164**, 683–695.

524 van Buuren, S. (2012). *Flexible Imputation of Missing Data*, 1st edn. Chapman and
525 Hall/CRC, Boca Raton, Fl.

526 van Buuren, S., Brands, J.P.L., Groothuis-Oudshoorn, K. & Rubin, D.B. (2006). Fully
527 conditional specification in multivariate imputation. *Journal of Statistical*
528 *Computation and Simulation*, **76**, 1049–1064.

529 van Buuren, S. & Groothuis-Oudshoorn, K. (2011). mice : Multivariate Imputation by
530 Chained. *Journal of Statistical Software*, **45**.

531 Cavender-Bares, J., Kozak, K.H., Fine, P.V. a & Kembel, S.W. (2009). The merging of
532 community ecology and phylogenetic biology. *Ecology letters*, **12**, 693–715.

533 Collins, L.M., Schafer, J.L. & Kam, C.M. (2001). A Comparision of Inclusive and
534 Restrictive Strategies in Modern Missing Data Procedures. *Psychological Methods*,
535 **6**, 330–351.

536 Cooper, N., Thomas, G.H., Venditti, C., Meade, A. & Freckleton, R.P. (2015). A
537 cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary.
538 *Biological Journal of the Linnean Society,* **118**, 64-77

539 Davis, M.P., Midford, P.E. & Maddison, W. (2013). Exploring power and parameter
540 estimation of the BiSSE method for analyzing species diversification. *BMC*
541 *evolutionary biology*, **13**, 38.

542 Diniz-Filho, J.A.F. (2001). Phylogenetic autocorrelation under distinct evolutionary
543 process. *Evolution*, **55**, 1104–1109.

544 Diniz-Filho, J.A.F., Sant'Ana, C.E.R. & Bini, L.M. (1998). An Eigenvector Method for
545 estimating Phylogenetic Inertia. *Evolution*, **52**, 1247–1262.

546 Diniz-Filho, J.A.F., Villalobos, F. & Bini, L.M. (2015). The best of both worlds :
547 Phylogenetic eigenvector regression and mapping. *Genetics and Molecular*
548 *Biology*, **38**, 396–400.

549 Enders, C.K. (2010). *Applied Missing Data Analysis*, 1st edn. New York, NY.

550 Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American*
551   *Naturalist*, **125**, 1–15.

552 Freckleton, R.P., Harvey, P.H. & Pagel, M. (2002). Phylogenetic Analysis and
553   Comparative Data : A Test and Review of Evidence. *The American naturalist*, **160**,
554   712–726.

555 Freckleton, R.P. & Jetz, W. (2009). Space versus phylogeny: disentangling
556   phylogenetic and spatial signals in comparative data. *Proceedings of the Royal*
557   *Society B*, **276**, 21–30.

558 Gaston, K.J., Chown, S.L. & Evans, K.L. (2008). Ecogeographical rules: elements of a
559   synthesis. *Journal of Biogeography*, **35**, 483–500.

560 Gillespie, D. (1996). Exact numerical simulation of the Ornstein-Uhlenbeck process and
561   its integral. *Physical Review E*, **54**, 2084–2091.

562 Gittleman, J.L. & Kot, M. (1990). Adaptation:Statistics and a Null model for estimating
563   phylogenetic effects. *Systematic Zoology*, **39**, 227–241.

564 Goldberg, E.E., Kohn, J.R., Lande, R., Robertson, K. a., Smith, S. a. & Igic, B. (2010).
565   Species Selection Maintains Self-Incompatibility. *Science*, **330**, 493–495.

566 Gonzalez-Suarez, M., Lucas, P.M. & Revilla, E. (2012). Biases in comparative analyses
567   of extinction risk: mind the gap. *The Journal of Animal Ecology*, **81**, 1211–22.

568 Graham, J.W., Olchowski, A.E. & Gilreath, T.D. (2007). How many imputations are
569   really needed? Some practical clarifications of multiple imputation theory.
570   *Prevention Science*, **8**, 206–213.

571 Guénard, G., Legendre, P. & Peres-Neto, P. (2013). Phylogenetic eigenvector maps: a
572   framework to model and predict species traits. *Methods in Ecology and Evolution*,
573   **4**, 1120–1131.

574 Hansen, T.F. (1997). Stabilizing Selection and the Comparative Analysis of Adaptation.
575   *Evolution*, **51**, 1341–1351.

576 Hansen, T.F. & Martins, E.P. (1996). Translating between microevolutionary process
577   and macroevolutionary patterns: correlation structure of interspecific data.
578   *Evolution*, **50**, 1404–1417.

579 Harmon, L.J., Losos, J.B., Jonathan Davies, T., Gillespie, R.G., Gittleman, J.L., Bryan
580   Jennings, W., Kozak, K.H., McPeek, M.A., Moreno-Roark, F., Near, T.J., Purvis,
581   A., Ricklefs, R.E., Schluter, D., Schulte, J.A., Seehausen, O., Sidlauskas, B.L.,
582   Torres-Carvajal, O., Weir, J.T. & Mooers, A.T. (2010). Early bursts of body size
583   and shape evolution are rare in comparative data. *Evolution*, **64**, 2385–2396.

584 Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J.
585   (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity.
586   *Annual Review of Ecology, Evolution, and Systematics*, **46**, 523–549.

587 Jetz, W. & Freckleton, R.P. (2015). Towards a general framework for predicting threat
588   status of data-deficient species from phylogenetic, spatial and environmental
589   information. *Philosophical transactions of the Royal Society of London. Series B,*
590   *Biological sciences*, **370**, 20140016.

591 Jones, K.E., Bielby, J., Cardillo, M., Fritz, S. a., O'Dell, J., Orme, C.D.L., Safi, K.,
592   Sechrest, W., Boakes, E.H., Carbone, C., Connolly, C., Cutts, M.J., Foster, J.K.,
593   Grenyer, R., Habib, M., Plaster, C. a., Price, S. a., Rigby, E. a., Rist, J., Teacher,
594   A., Bininda-Emonds, O.R.P., Gittleman, J.L., Mace, G.M. & Purvis, A. (2009).

PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, **90**, 2648–2648.

Kamilar, J.M., Cooper, N. & B, P.T.R.S. (2013). Phylogenetic signal in primate behaviour, ecology and life history. *Proceeding of the Royal Society B*, **368**, 20120341.

Kattge, J., Ogle, K., Bönisch, G., Díaz, S., Lavorel, S., Madin, J., Nadrowski, K., Nöllert, S., Sartor, K. & Wirth, C. (2011). A generic structure for plant trait databases. *Methods in Ecology and Evolution*, **2**, 202–213.

Lukas, D. & Clutton-Brock, T.H. (2013). The evolution of social monogamy in mammals. *Science*, **341**, 526–30.

Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K. & Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, **3**, 743–756.

Nakagawa, S. & Freckleton, R.P. (2008). Missing inaction: the dangers of ignoring missing data. *Trends in Ecology and Evolution*, **23**, 592–596.

Nakagawa, S. & Freckleton, R.P. (2010). Model averaging, missing data and multiple imputation: a case study for behavioural ecology. *Behavioral Ecology and Sociobiology*, **65**, 103–116.

Oba, S., Sato, M., Takemasa, I., Monden, M., Matsubara, K. & Ishii, S. (2003). A Bayesian missing value estimation method for gene expression profile data. *Bioinformatics*, **19**, 2088–2096.

Paradis, E., Claude, J. & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**, 289–290.

Penone, C., Davidson, A.D., Shoemaker, K.T., Marco, M. Di, Rondinini, C., Brooks, T.M., Young, B.E., Graham, C.H. & Costa, G.C. (2014). Imputation of missing data in life-history traits datasets: which approach performs the best? *Methods in Ecology and Evolution*, **5**, 961–970.

Purvis, A., Gittleman, J.L., Cowlishaw, G. & Mace, G.M. (2000). Predicting extinction risk in declining species. *Proceeding of the Royal Society B*, **267**, 1947–1952.

R Core Team. (2015). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*.

Ram, K. (2013). Git can facilitate greater reproducibility and increased transparency in science. *Source code for biology and medicine*, **8**, 7.

Reddy, S. & Dávalos, L.M. (2003). Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography*, **30**, 1719–1727.

Reichman, O.J., Jones, M.B. & Schildhauer, M.P. (2011). Challenges and opportunities of open data in ecology. *Science*, **331**, 703–705.

Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.

Revell, L.J., Harmon, L.J. & Collar, D.C. (2008). Phylogenetic signal, evolutionary process, and rate. *Systematic biology*, **57**, 591–601.

Rosado, B.H.P., de S. L. Figueiredo, M., de Mattos, E.A. & Grelle, C.E. V. (2015). Eltonian shortfall due to the Grinnellian view: functional ecology between the mismatch of niche concepts. *Ecography*,

639   http://onlinelibrary.wiley.com/doi/10.1111/ecog.01.

640   Rubin, D.. (1976). Inference and Missing Data. *Biometrika*, **63**, 581–592.

641   Schafer, J.L. & Graham, J.W. (2002). Missing Data: Our View of the State of the Art.
642   *Psychological Methods*, **7**, 147–177.

643   Schrodt, F., Kattge, J., Shan, H., Fazayeli, F., Joswig, J., Banerjee, A., Reichstein, M.,
644   Bönisch, G., Díaz, S., Dickie, J., Gillison, A., Karpatne, A., Lavorel, S., Leadley,
645   P., Wirth, C.B., Wright, I.J., Wright, S.J. & Reich, P.B. (2015). BHPMF - a
646   hierarchical Bayesian approach to gap-filling and trait prediction for macroecology
647   and functional biogeography. *Global Ecology and Biogeography*, **24**, 1510–1521.

648   Slater, G.J., Harmon, L.J., Wegmann, D., Joyce, P., Revell, L.J. & Alfaro, M.E. (2012).
649   Fitting models of continuous trait evolution to incompletely sampled comparative
650   data using approximate bayesian computation. *Evolution*, **66**, 752–762.

651   Swenson, N.G. (2014). Phylogenetic imputation of plant functional trait databases.
652   *Ecography*, **37**, 105–110.

653   Taugourdeau, S., Villerd, J., Plantureux, S., Huguenin-Elie, O. & Amiaud, B. (2014).
654   Filling the gap in functional trait databases: use of ecological hypotheses to replace
655   missing data. *Ecology and Evolution*, **4**, 944–958.

656   Venables, W.N. & Ripley, B.D. (2002). *Modern Applied Statistics with S*, 4th edn.
657   Springer, New York.

658   Vilela, B. & Villalobos, F. (2015). letsR: a new R package for data handling and
659   analysis in macroecology. *Methods in Ecology and Evolution*, n/a–n/a.

660   Vilela, B., Villalobos, F., Rodríguez, M.Á. & Terribile, L.C. (2014). Body Size,
661   Extinction Risk and Knowledge Bias in New World Snakes. *PloS one*, **9**, e113429.

662   Webb, C.O., Ackerly, D.D., Mcpeek, M.A. & Donoghue, M.J. (2002). Phylogenies and
663   Community Ecology. 475–505.

664   Wiens, J.J. & Graham, C.H. (2005). Niche Conservatism: Integrating Evolution,
665   Ecology, and Conservation Biology. *Annual Review of Ecology, Evolution, and*
666   *Systematics*, **36**, 519–539.

667   Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M.M. & Jetz, W.
668   (2014). EltonTraits 1.0 : Species-level foraging attributes of the world 's birds and
669   mammals. *Ecology*, **95**, 2027.

670

Table 1. Model selection of descriptive statistic and phylogenetic signal errors. The values represent the ΔAICc of the three best models for each error.

Models	Blomberg's K	PS Moran	Imputation error	Mean	Variance	Regression Coefficient
Model 1	6.23	-	0.00	0.04	11.71	-
Model 2	0.00	-	4.71	0.00	0.00	8.28
Model 3	3.79	-	-	-	1.94	-
Model 4	-	0.00	-	15.45	-	0.00
Model 5	-	1.07	-	-	-	3.74
Model 6	-	-	34.85	-	-	-
Model 7	-	3.31	-	-	-	-

671

672

673

674

675

676

677

678

679

680

681

682

683

27

684

685

Table 2. Description of selected models explaining estimation error of descriptive statistics (mean, variance and regression coefficient) and phylogenetic signal (Blomberg's K and Moran Correlogram).

Terms	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Mec	x	x	x	x	x	x	x
Met	x	x	x	x	x	x	x
OU	x	x	x	x	x	x	x
Per	x	x	x	x	x	x	x
Mec:Met	x	x	x	x	x	x	x
Mec:OU	x	x	x	x	x	x	-
Mec:Per	x	x	x	x	x	x	x
Met:Per	x	x	x	x	x	x	x
OU:Per	x	x	x	x	x	x	x
Mec:Met:OU	x	x	x	-	-	x	-
Mec:Met:Per	x	x	x	x	x	x	x
Mec:OU:Per	x	x	-	x	-	x	-
Met:OU:Per	x	x	x	x	x	-	-
Mec:Met:OU:Per	x	-	-	-	-	-	-

Mec = missing data mechanism, Met = imputation method, OU = OU selection strength, Per = missing data percentage. ":" means interaction among variables.

686

687

688

689

690

691

692
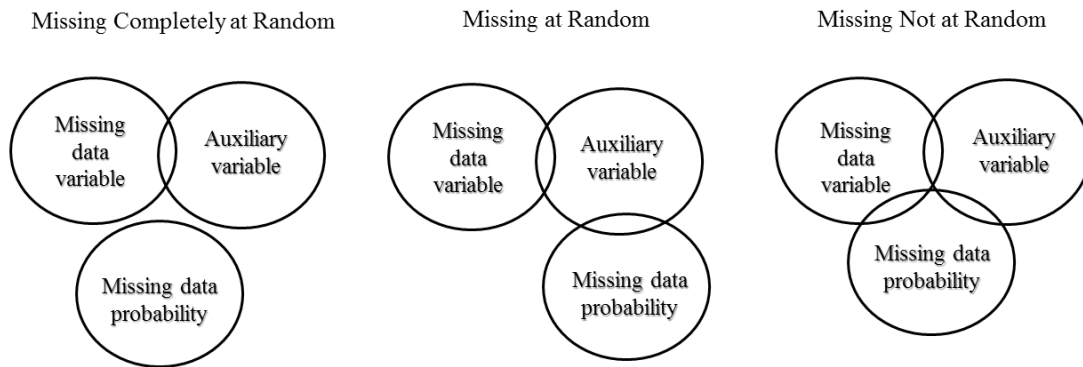
Figure 1. Correlation structure among variables in each missing data mechanism. Circles represent model components and their intersection represents correlation among them.

696
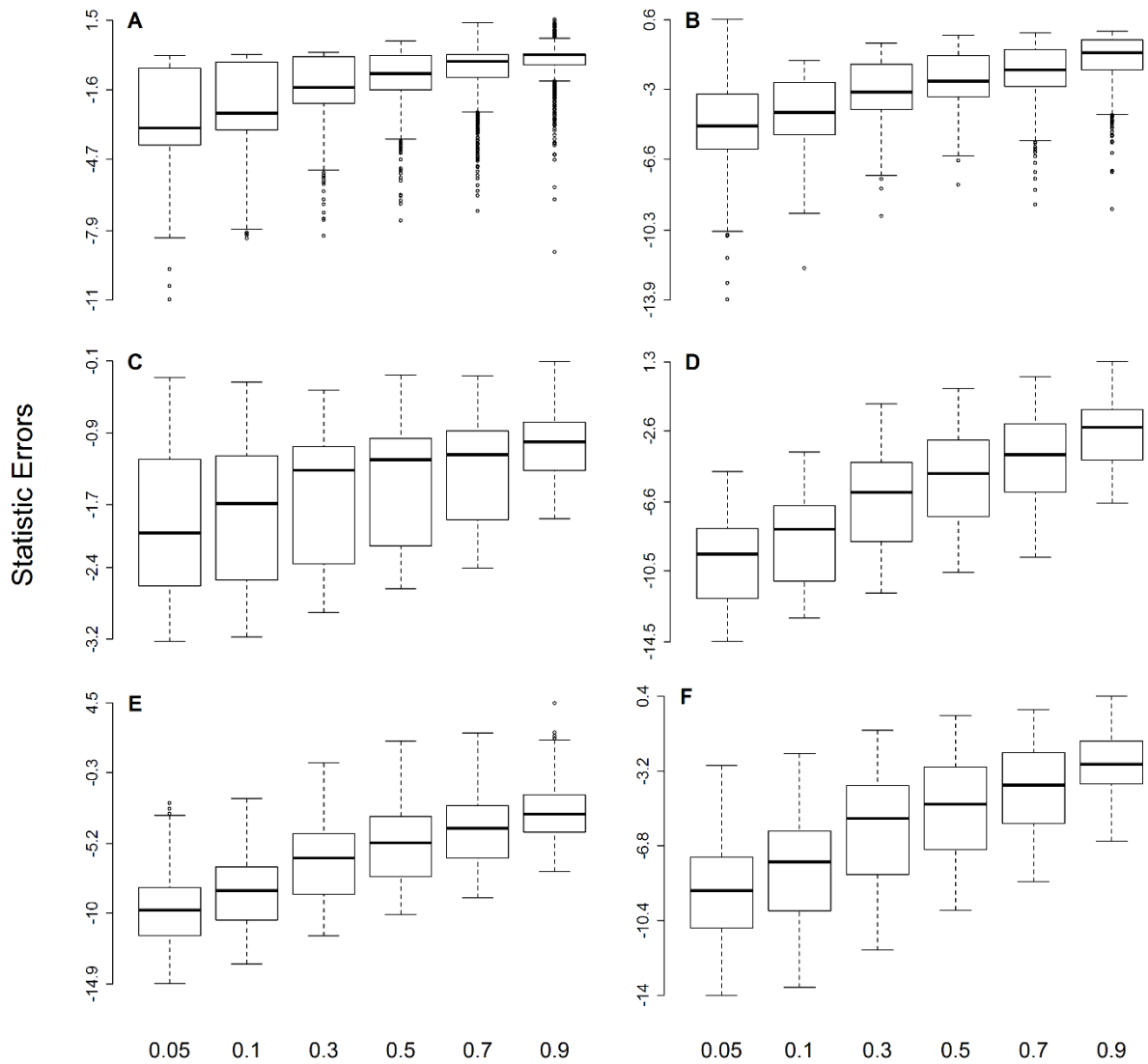
Figure 2. Missing data percentage and statistic estimation errors. (A) Logarithm of absolute average proportions of Blomberg's K change after imputation or deletion, (B) Logarithm of average proportion of Moran's Correlogram values change, (C) Imputation error measured as logarithm of average NRMSE, (D) Logarithm of MSE of trait mean, (E) Logarithm of MSE of trait variance and (F) Logarithm of MSE of regression coefficient.

703

704

705

706
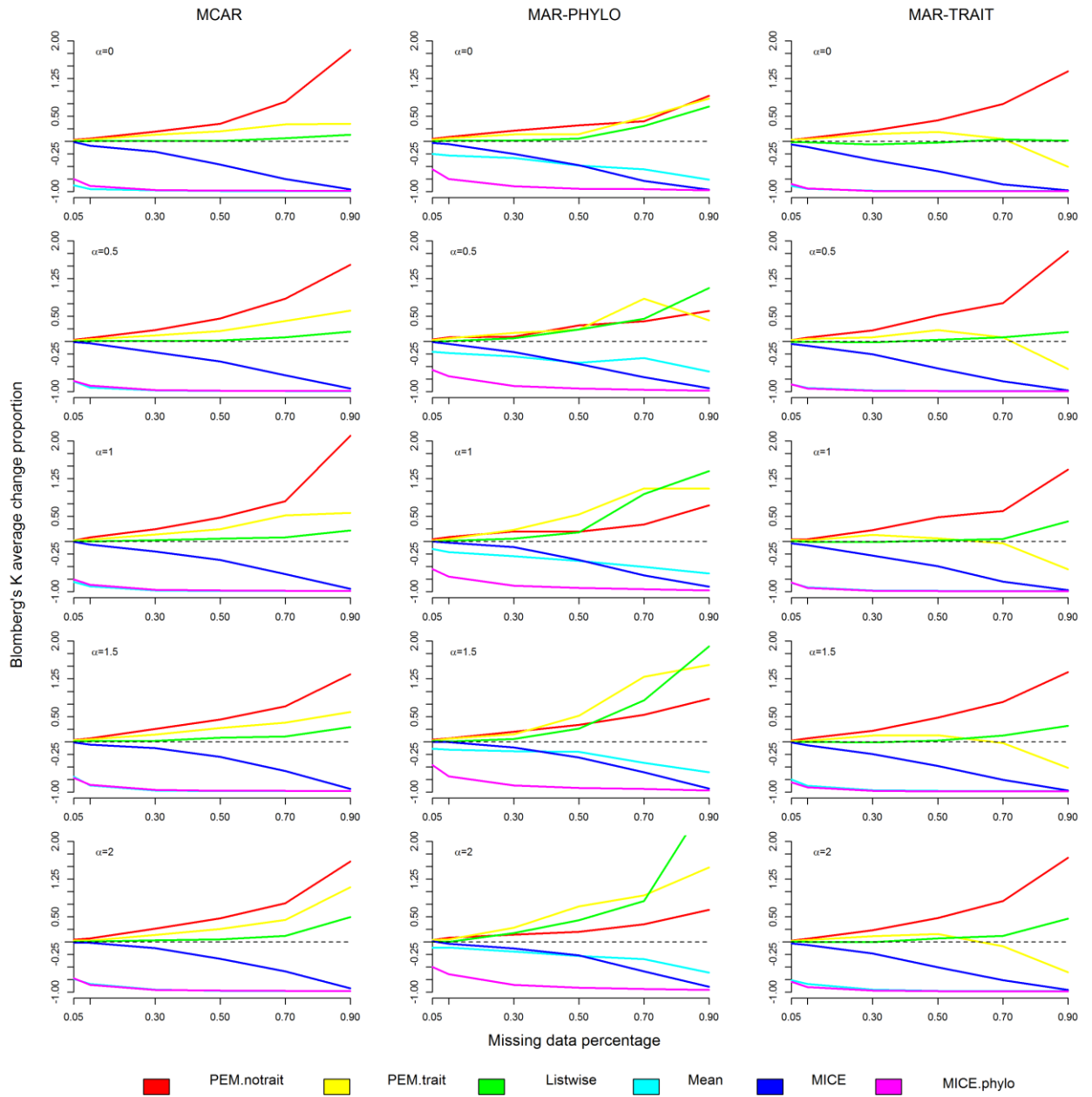
Figure 3. Blomberg's K average change proportion under different methods, OU

selective strength, missing data percentage and mechanisms.

709

710

711

712

713

714
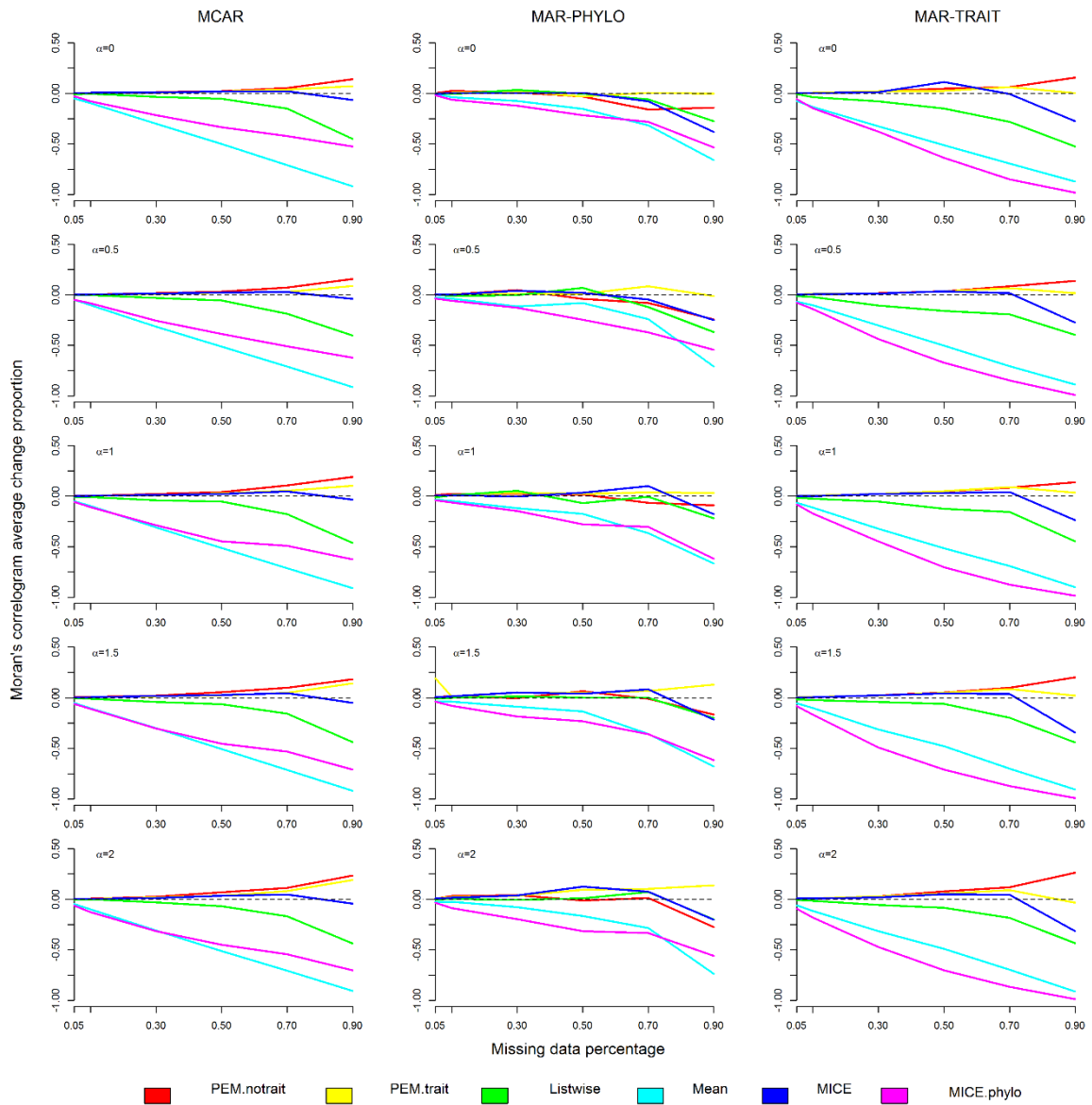
Figure 4. Moran's Correlogram average change proportion under different methods, OU selective strength, missing data percentage and mechanisms.
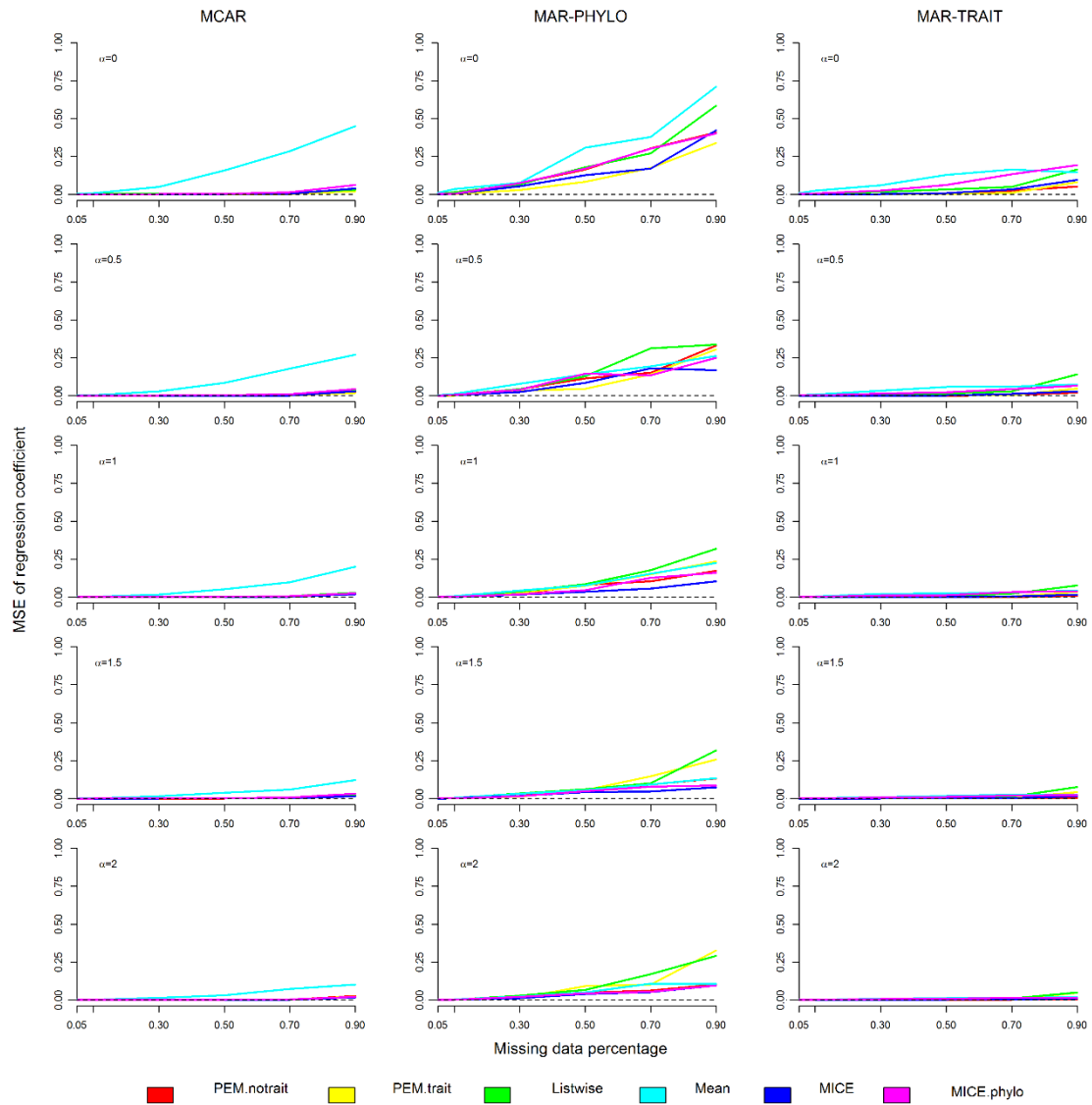
715

716

717

718

719

720

721

722

Figure 5. Regression coefficient MSE (mean squared error) under different methods, OU selective strength, missing data percentage and mechanisms.
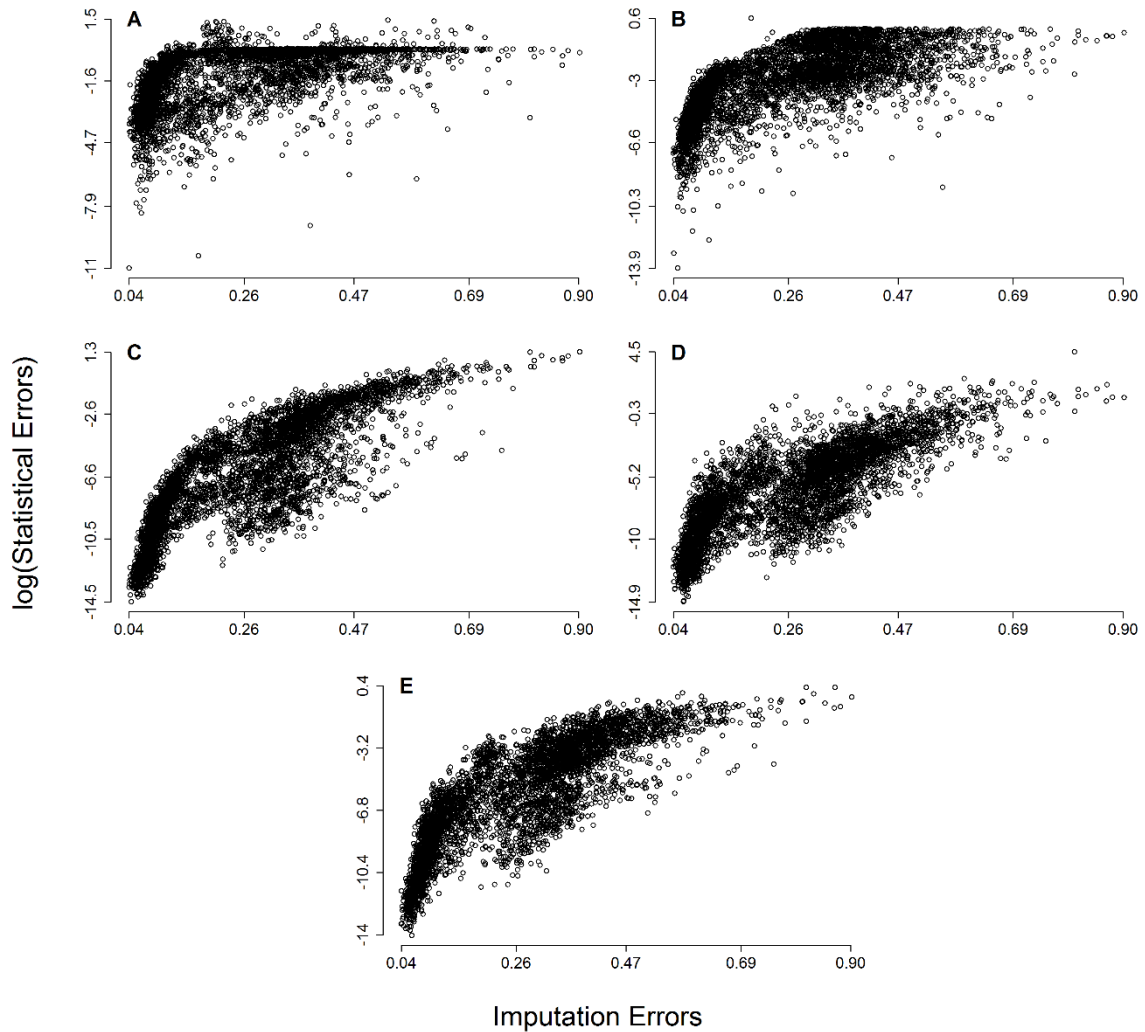
723
724
725
726
727
728
729
730
731
732
733

33

Figure 6. Scatterplot of imputation errors (average NRMSE) and statistical errors. (A) Logarithm of absolute average Blomberg's K change proportion, (B) Logarithm of absolute average Moran's Correlogram change proportion, (C) Logarithm of mean MSE, (D) Logarithm of variance MSE and (E) Logarithm of regression coefficient MSE.
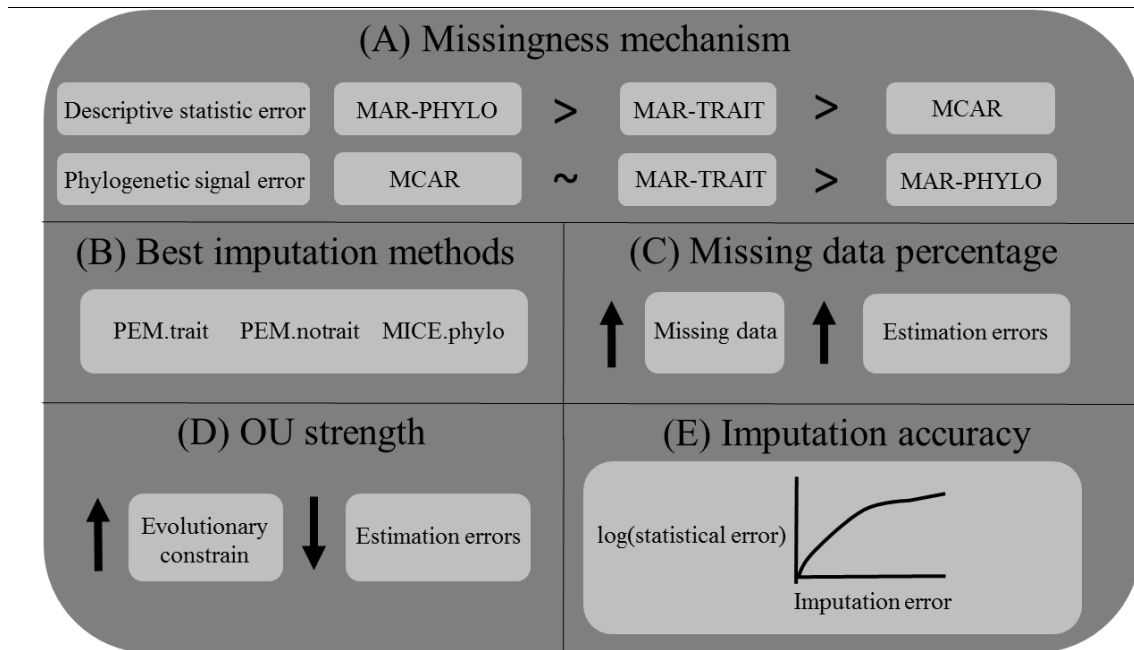
Figure 7. Summary of the main results showing (A) the differences on estimation errors among missing data mechanisms and estimated statistics; (B) highlighting the best imputation methods; (C) the effect of missing data percentage in statistical estimation; (D) OU selection strength; and (E) the non-linear relationship between imputation error and statistical estimation error logarithm.