



This is a repository copy of *Comments to “Persistent problems in the construction of matrix population models”*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/186299/>

Version: Accepted Version

Article:

Che-Castaldo, J., Jones, O.R., Kendall, B.E. et al. (13 more authors) (2020) Comments to “Persistent problems in the construction of matrix population models”. *Ecological Modelling*, 416. 108913. ISSN 0304-3800

<https://doi.org/10.1016/j.ecolmodel.2019.108913>

Article available under the terms of the CC-BY-NC-ND licence
(<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Title: Comments to “Persistent problems in the construction of matrix population models”

Authors: Judy Che-Castaldo* (1), Owen R. Jones (2), Bruce E. Kendall (3), Jean H. Burns (4), Dylan Z. Childs (5), Thomas H.G. Ezard (6), Haydee Hernandez-Yanez (1), David J. Hodgson (7), Eelke Jongejans (8), Tiffany Knight (9), Cory Merow (10), Satu Ramula (11), Iain Stott (12), Yngvild Vindenes (13), Hiroyuki Yokomizo (14), & Roberto Salguero-Gómez (15)

- (1) Alexander Center for Applied Population Biology, Conservation & Science Department, Lincoln Park Zoo, Chicago, IL 60614-4712 USA
- (2) Interdisciplinary Center on Population Dynamics & Department of Biology, Campusvej 55, University of Southern Denmark, 5230 Odense M, Denmark
- (3) Bren School of Environmental Science & Management, University of California Santa Barbara, Santa Barbara, CA 93106-5131 USA
- (4) Department of Biology, Case Western Reserve University, Cleveland, OH 44106-7080 USA
- (5) Department of Animal and Plant Sciences, University of Sheffield, Sheffield, S10 2TN, UK
- (6) Ocean & Earth Science, National Oceanography Centre Southampton, University of Southampton Waterfront Campus, Southampton SO14 3ZH, UK
- (7) Centre for Ecology & Conservation, College of Life and Environmental Sciences, University of Exeter, Cornwall, UK
- (8) Animal Ecology and Physiology Department, Radboud University, Nijmegen, The Netherlands
- (9) Institute of Biology, Martin Luther University Halle-Wittenberg, Am Kirchtor 1, 06108 Halle (Saale), Germany; Department of Community Ecology, Helmholtz Centre for Environmental Research-UFZ, Theodor-Lieser-Straße 4, 06120 Halle (Saale), Germany; German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany
- (10) Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA
- (11) Department of Biology, University of Turku, FI-20014, Turku, Finland
- (12) School of Life Sciences, University of Lincoln, Brayford Pool, LN6 7TS, Lincoln UK
- (13) Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Oslo, Norway
- (14) Center for Health and Environmental Risk Research, National Institute for Environmental Studies, Tsukuba, 305-8506, Japan
- (15) Department of Zoology, 11a Mansfield Rd, University of Oxford, OX1 3SZ, Oxford, UK

*Corresponding author: jchecastaldo@lpzoo.org; 301-351-8290

1 The use of matrix population models (MPMs*) to summarize and analyze the empirically
2 observed demography of both plant and animal populations has exhibited rapid growth in recent
3 decades (Salguero-Gómez et al. 2015, Salguero-Gómez et al. 2016). Comparative analyses
4 using independent collections of MPMs have generated valuable ecological insights (e.g.,
5 Franco & Silvertown 1990, Franco & Silvertown 2004, Ramula et al. 2008, Buckley et al. 2010,
6 Burns et al. 2010, Salguero-Gómez & Casper 2010, Stott et al. 2010, Crone et al. 2011, Zeigler
7 et al. 2013), leading to the recognition that a comprehensive MPM dataset would be a valuable
8 resource for the ecological research community. We therefore launched the COMPADRE Plant
9 Matrix Database (Salguero-Gómez et al. 2015) and the COMADRE Animal Matrix Database
10 (Salguero-Gómez et al. 2016; <http://compadre-db.com/>), which comprise 1204 plant and animal
11 species and 10949 MPMs as of October 2019.

12
13 Our initial goal with COMPADRE and COMADRE was to present MPMs as intended by the
14 authors of the original publications, relying on the authors and the peer review process to
15 ensure that the published population projection matrix correctly described the intended life
16 history. Thus, our error-checking procedure to date has focused for the most part on eliminating
17 typographical errors introduced during digitization, rather than errors of MPM construction made
18 by the original authors. However, Kendall et al. (2019) have shown that a substantial fraction of
19 animal MPMs in peer-reviewed papers contain errors in construction, such that model
20 components are missing and/or the model does not accurately represent the described life
21 history of the species. Plant MPMs are also prone to errors in construction, including dormant
22 life stages (most commonly a seed bank) that are missing (Doak et al. 2002, Nguyen et al.
23 2019) or incorrectly incorporated, leading to a one-year delay in the life cycle (Caswell 2001, p.
24 60). Although some construction errors may not dramatically affect estimates of demographic
25 parameters in some cases, they can be crucial in others. Estimates of generation time and
26 population growth rate, in particular, can be sensitive to missing components or life stages

*MPMs = matrix population models

27 (Kendall et al. 2019, Nguyen et al. 2019). Construction errors may also lead to directional
28 biases, such as underestimating the elasticities of population growth rate to survival when
29 fertility coefficients do not include survival (Kendall et al. 2019), and these biases may affect
30 results in comparative analyses. Moving forward, there is a clear need for better training
31 materials to help ecologists and conservation scientists (many with little experience in
32 mathematical modeling) construct MPMs from their data. But how can we ensure that previously
33 published MPMs, including those already compiled in COMPADRE and COMADRE and those
34 remaining to be digitized, correctly represent the intended life histories, and thus support
35 meaningful synthetic analyses?

36
37 We believe that an important principle in big data analyses can be summed up as “*Just because*
38 *you can, does not mean you should*”. That is, although large volumes of data and tools that
39 enable their fast analysis are available, they should not be used without careful consideration
40 and evaluation. There are many possible sources of error whenever large datasets are
41 compiled, and we have previously discussed some of these (including errors in matrix
42 construction by authors) as a cautionary note to data users (Salguero-Gómez et al. 2016). We
43 also maintain an open, frequently updated digitization protocol, and an extensive user guide
44 (both available on our GitHub repository, <https://github.com/jonesor/compadreDB>). These open,
45 transparent policies are essential for users to fully understand the data and choose the subsets
46 of MPMs that are most appropriate for addressing their specific research questions.

47
48 In our collective cases, comparative analyses of demographic data from COMPADRE and
49 COMADRE (e.g., Vindenes and Engen 2017, Beckman et al. 2018, Che-Castaldo et al. 2018,
50 Davidson et al. 2019, Healy et al. 2019, Jones et al. 2019, Nguyen et al. 2019, Roper et al.
51 2019, Pistón et al. 2019) have each resulted in a separate sample size because the research
52 questions required a different set of selection criteria. For instance, for some questions related

*MPMs = matrix population models

53 to estimating various moments of longevity, the subcomponents of fertility are not necessary (if
54 one assumes no trade-offs between reproduction and survival), and so MPMs that do not
55 include those (indicated by the metadata field “MatrixFec”) can be used. Similarly, to calculate
56 the population growth rate from an MPM, the subcomponents of survival, sexual reproduction
57 and clonality do not need to be separated (indicated by the metadata field “MatrixSplit”), but to
58 estimate life history traits and vital rate sensitivities/elasticities, it is fundamental that they are
59 separated. Thus, the set of selection criteria used for a specific study must be carefully
60 designed before the analysis begins. Moreover, the availability of digitized data should not be a
61 substitute for examining the original sources, which may provide important details of interest for
62 the user’s specific goals. For this reason (and to facilitate the citation of original sources),
63 COMPADRE and COMADRE include bibliographic information for the source studies.

64
65 Nevertheless, we agree that addressing the newly-identified MPM construction errors in
66 COMPADRE and COMADRE would be a useful service for the scientific community. In fact, in
67 recent years we have started to include information on data quality in the metadata. For
68 example, there is a field (“SurvivalIssue”) for noting when the summed survival for a given
69 age/stage exceeds 1.0, which may indicate a typographical error or a sensitivity matrix being
70 incorrectly identified as an MPM in the original source. Other issues, such as when GPS
71 coordinates are approximate rather than accurate, or environmental conditions that, while
72 naturally occurring, may not be representative of typical conditions for a given population (e.g.,
73 fires, herbivory, droughts), are noted in the unstructured “Observations” field. In the coming
74 months, we will take further steps including updating the metadata to record (1) whether the
75 MPM is based on a pre-breeding or post-breeding census whenever the information is available,
76 and (2) whether the MPM in the database has been altered from the original published one and
77 why [e.g., typo in the original source; inclusion of an unnecessary seed stage where no
78 permanent seed bank exists (Caswell 2001, p. 60)], or whether the MPM was calculated from

*MPMs = matrix population models

79 other published data along with reference to those data (e.g., a life table, integral projection
80 model, individual-based model). We have updated the digitization protocol to include Kendall et
81 al. (2019) as essential reading so that our team can avoid the identified common errors when
82 constructing MPMs from published life cycle data. Future releases of the database will include a
83 unique identification number for each existing and new MPM, so that users can look up a
84 specific MPM and all of its associated observations and metadata. They can also share with the
85 community any additional metadata they collect for their own analyses, which could be linked
86 back to the full dataset via the unique ID.

87

88 In the longer term, our plans include identifying MPMs that are likely to have construction issues
89 following the protocol presented in Appendix B of Kendall et al. (2019). If the publication
90 contains sufficient data, we will attempt to correct the MPMs and indicate why and how we have
91 done so. This way, users will be able to identify the modified matrices and determine whether
92 those modifications are appropriate for their use. In the last year, we have transitioned to a
93 relational database that will allow us to keep a record of changes to the MPM data, so that we
94 can record the MPMs as presented by original authors and also update them to fix known errors
95 such as those mentioned above. However, often the published paper and supplementary
96 materials do not include the necessary information (e.g., a separate life table or mean
97 development time for each life history stage) to help us contextualize the MPMs and determine
98 specific construction methods. Our digitization team invests a significant amount of time
99 contacting authors for clarification and additional information. Without the original researchers'
100 input, there may be many MPMs that we cannot correct or even flag with potential issues. This
101 highlights the need for authors to consider providing as much detail on the life cycles, data
102 collection, and biogeographic context for their species and populations as possible to enable
103 reproducible research.

104

*MPMs = matrix population models

105 Although finding errors like the ones reported by Kendall et al. (2019) is worrisome, we
106 emphasize that an evaluation of data quality issues and their extent (not to mention large-scale
107 assessments in general) would not be possible without the existence of a database for MPMs. A
108 database also allows us to implement solutions systematically across the entire set of
109 demographic studies represented in the database. As a centralized repository of MPM data,
110 COMPADRE and COMADRE can: (1) serve as a point of reference for recording whether a
111 particular publication may have miscalculated MPMs, (2) pool efforts from independent groups
112 or individuals to identify and/or fix the miscalculated MPMs, and (3) enable users to access and
113 benefit from previously corrected MPMs. The database will only improve as we continue to
114 aggregate data, assess data quality, and make corrections. Moreover, the mission of the
115 COMPADRE and COMADRE team goes beyond releasing large volumes of data, and includes
116 teaching workshops around the globe about MPMs and engaging the demographic community.
117 We can therefore use our existing platform, including educational materials on our GitHub
118 repository (see link above) and our regular user engagement workshops, to help raise
119 awareness and promote more careful construction of MPMs.

120

121 We encourage and welcome the community to help validate the MPM data in COMPADRE and
122 COMADRE, as the databases are currently supported by limited funding and volunteered
123 researcher time. Conducting a comprehensive check in the databases for construction errors
124 will require substantial effort, and we are only beginning to understand the extent of these
125 issues. We are developing website tools (such as the new function for reporting data errors and
126 corrections) to gather input and facilitate collaborative data validation. Participation from users
127 and original authors will be vital in this effort to help us deliver reliable, open-access MPM data
128 for demographic researchers and the broader scientific community.

129

130

*MPMs = matrix population models

131 **Literature Cited**

132 Buckley, Y. M. et al. 2010. Causes and consequences of variation in plant population growth
133 rate: a synthesis of matrix population models in a phylogenetic context. *Ecology Letters* 13,
134 1182–1197.

135

136 Burns, J. H. et al. 2010. Empirical tests of life-history evolution theory using phylogenetic
137 analysis of plant demography. *Journal of Ecology* 98, 334–344.

138

139 Beckman N, Bullock J, Salguero-Gómez R. 2018. High dispersal ability is related to fast life
140 history strategies. *Journal of Ecology* 4, 1349–1362.

141

142 Caswell, H. 2001. *Matrix Population Models: Construction, Analysis, and Interpretation*, 2nd
143 edn. Sinauer Associates, Inc, Sunderland, MA, USA

144

145 Che-Castaldo, J., Che-Castaldo, C., Neel, M. C. 2018. Predictability of demographic rates
146 based on phylogeny and biological similarity. *Conservation Biology* 32, 1290–1300.

147

148 Crone, E. E. et al. 2011. How do plant ecologists use matrix population models? *Ecology Letters*
149 14, 1–8.

150

151 Davison R., Stadman M., Jongejans E. 2019. Stochastic effects contribute to population fitness
152 differences. *Ecological Modelling* 408:108760.

153

154 Doak, D.F., Thomson, D., Jules, E.S. 2002. *Population viability analysis for plants:*

155 *Understanding the demographic consequences of seed banks for population health.*

156 in: Beissinger S.R., McCullough D.R. (eds.) *Population viability analysis*. The University

157 of Chicago Press, Chicago, Illinois, USA.

158

159 Franco, M. & Silvertown, J. 1990. Plant demography: What do we know? *Evolutionary Trends in*
160 *Plants* 4, 74–76.

161

162 Franco, M. & Silvertown, J. 2004. A comparative demography of plants based upon elasticities
163 of vital rates. *Ecology* 85, 531–538.

164

165 Healy K., Ezard T., Jones O., Salguero-Gómez R., Buckley Y. 2019. Animal life history is
166 shaped by the pace of life and the distribution of age-specific mortality and reproduction. *Nature*
167 *Ecology & Evolution* 3, 1217–1224.

168

169 Jones O.R., Ezard T.H.G., Dooley C., Healy K., Hodgson D.J., Mueller M., Townley S.,
170 Salguero-Gomez R. 2019. My family and other animals: Human demography under a
171 comparative cross-species lens. In "Human Evolutionary Demography", Burger O., Lee R., &
172 Sear R. (eds). Open Book Publishers. Cambridge. UK.

173

174 Kendall, B.E., Fujiwara, M., Diaz-Lopez, J., Schneider, S., Voigt, J., Wiesner, S. 2019.
175 Persistent problems in the construction of matrix population models. *Ecological Modelling* 406,
176 33–43 <https://doi.org/10.1016/j.ecolmodel.2019.03.011>

177

178 Nguyen V., Buckley Y.M., Salguero-Gómez R., Wardle G.M. 2019. Consequences of neglecting
179 cryptic life stages from demographic models. *Ecological Modelling* 408, 108723.

180

181 Pistón N., de Bello F., Dias A., Götzenberger L., de Mattos E., Rosado B., Salguero-Gómez R.,
182 Carmona C. 2019. Multidimensional ecological analyses demonstrate how interactions between
183 functional traits shape fitness and life history strategies. *Journal of Ecology* 107: 2317–2328.
184

185 Ramula, S., Knight, T. M., Burns, J. H. & Buckley, Y. M. 2008. General guidelines for invasive
186 plant management based on comparative demography of invasive and native plant populations.
187 *Journal of Applied Ecology* 45, 1124–1133.
188

189 Roper M., Capdevila P., Salguero-Gómez R. 2019. Senescence - still an unsolved problem of
190 biology. Biorxiv: DOI 10.1101/739730
191

192 Salguero-Gómez, R. & Casper, B. B. 2010. Keeping plant shrinkage in the demographic loop.
193 *Journal of Ecology* 98, 312–323.
194

195 Salguero-Gómez, R. et al. 2015. The COMPADRE Plant Matrix Database: an open online
196 repository for plant demography. *Journal of Ecology* 103:202–218
197

198 Salguero-Gómez, R. et al. 2016. COMADRE: a global data base of animal demography. *Journal*
199 *of Animal Ecology* 85:371–384
200

201 Stott, I., Franco, M., Carslake, D., Townley, S. Hodgson, D. 2010. Boom or bust? A comparative
202 analysis of transient population dynamics in plants. *Journal of Ecology* 98, 302–311.
203

204 Vindenes, Y., Engen, S. 2017. Demographic stochasticity and temporal autocorrelation in the
205 dynamics of structured populations. *Oikos* 126, 462–471.
206

*MPMs = matrix population models

207 Zeigler, S., Che-Castaldo, J. Neel, M. 2013. Actual and potential use of population viability
208 analysis in recovery of plant species listed under the U.S. Endangered Species Act.
209 *Conservation Biology* 27, 1265–1278.
210