



HAL
open science

Adequate statistical modelling and data selection are essential when analysing abundance and diversity trends

Marion Desquilbet, Pierre-André Cornillon, Laurence Gaume, Jean-Marc Bonmatin

► To cite this version:

Marion Desquilbet, Pierre-André Cornillon, Laurence Gaume, Jean-Marc Bonmatin. Adequate statistical modelling and data selection are essential when analysing abundance and diversity trends. *Nature Ecology & Evolution*, 2021, 5, pp.592-594. 10.1038/s41559-021-01427-x . hal-03191511v1

HAL Id: hal-03191511

<https://hal.inrae.fr/hal-03191511v1>

Submitted on 18 Nov 2021 (v1), last revised 8 Dec 2021 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Title page

Title. Inconsistencies in data selection and issues in statistical modelling invalidate the results of Crossley et al., 2020.

Matters arising. Arising from Crossley *et al.*, No net insect abundance and diversity declines across US Long Term Ecological Research sites, *Nature Ecology & Evolution* (2020) doi:10.1038/s41559-020-1269-4.

Authors. Marion Desquilbet^{1*†}, Pierre-André Cornillon^{2†}, Laurence Gaume³, Jean-Marc Bonmatin⁴

Affiliations

¹ Toulouse School of Economics, INRAE, University of Toulouse Capitole, Toulouse, France.

² Univ Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France.

³ AMAP, University of Montpellier, CNRS, CIRAD, INRAE, IRD, Montpellier, France.

⁴ Centre de Biophysique Moléculaire, CNRS, 45071 Orléans, France.

*Correspondence to: marion.desquilbet@inrae.fr.

†These authors contributed equally to this work.

Acknowledgments

Doyle McKey (University of Montpellier) made suggestions to improve the paper. This work was publicly funded by the French National Research Agency (grant ANR-17-EURE-0010 under the "Investissements d'avenir" program and grant ANR-15-CE21-0006 to M.D.).

Author contributions

M.D. and P.A.C. performed both the detailed and overall analysis of the article and wrote the original draft. P.A.C. examined and re-programmed the R code. J.M.B. contributed to the analysis of data selection in the article. L.G. extensively edited the manuscript. All authors contributed to the general comment and reviewed the manuscript.

1 **Inconsistencies in data selection and issues in statistical modelling invalidate the results**
2 **of Crossley et al., 2020**

3 Matters arising. Arising from Crossley *et al.*, No net insect abundance and diversity declines
4 across US Long Term Ecological Research sites, *Nature Ecology & Evolution* (2020)
5 doi:10.1038/s41559-020-1269-4.

6

7 **Abstract**

8 In a meta-analysis of a large number of time series on arthropod abundance in natural and
9 managed areas of the United States, Crossley et al. reported no evidence of an overall decline
10 in insect abundance and diversity¹. We found inconsistencies in the selection of data and
11 major concerns in the statistical analysis, which, we argue, invalidate their conclusions. We
12 call for rigorous methodologies in meta-analyses of biodiversity trends because relevant
13 information is crucial for stakeholders and policy makers.

14

15 **Matters arising**

16 The extent of the decline of insect populations worldwide is much debated²⁻⁴, with major
17 implications for public policy investment for the protection of biodiversity. Crossley et al.
18 conducted a meta-analysis of 5,375 geographically and taxonomically varied time series on
19 arthropod abundance during 4 to 36 years across the United States¹. They suggested that the
20 broad representation of taxa, habitats, feeding guilds and sampling methods made their data
21 well suited to detect any overall decline in insect biodiversity. They concluded that there was
22 no significant change in insect populations, a result that was consistent across insect feeding
23 groups and between heavily disturbed and relatively natural sites. However, we argue that
24 serious inconsistencies in the selection of data and issues in the statistical analysis invalidate

25 their conclusions.

26 Regarding data selection, the article is intended to be a meta-analysis of insect abundance
27 trends in US Long Term Ecosystem Research (LTER) sites, but it departs from this
28 description in three ways. First, 39.5% of time series are from the Suction Trap Network
29 (STN). One suction trap of the STN is located in the Kellogg LTER site, and all STN data,
30 encompassing ten US states, were incorporated into the Kellogg LTER dataset up to 2014⁵.
31 But the dataset used in the meta-analysis, spanning up to 2019, is not linked to a LTER. The
32 inclusion of this dataset may bias results because the STN exclusively provides data on
33 aphids, and was primarily aimed at documenting pest aphids⁵, which benefit from intensive
34 agriculture. Second, 9% of the time series were of non-insect arthropods or included both
35 insects and other arthropods, and therefore should have been excluded (Table 1). Third,
36 several LTER datasets were not included in the analysis, without justification (e.g. refs ⁶⁻⁹).
37 These inconsistencies add to other criticisms of this article¹⁰ regarding unaccounted-for
38 changes in sampling location and sampling effort at LTER sites and the unaccounted-for
39 impact of experimental conditions on insect populations in experimental datasets.

40 The statistical approach in Crossley et al. reveals further problems. These authors modelled
41 separately abundance trends of each species in each considered ‘locale’ of each LTER (in the
42 R script, a locale could be an arthropod group, a location or a collection method)¹¹. For each
43 species of each locale of each LTER (hereafter, LLS), changes in abundance were estimated
44 by performing a univariate regression of the scaled logarithm of abundance on scaled time,
45 with a Gaussian auto-regressive error of order 1 (box 1). This statistical modelling raises four
46 issues.

47 Insert box 1.

48 First, most individual time series were too short to provide reliable estimations of the four
49 unknown parameters for each LLS: the intercept and the slope of the linear regression, the
50 autocorrelation coefficient and the variance of the error term. Indeed, 75% of LLS time series
51 had 4 to 14 years of data. While no simple threshold number of years exists, 16 years of
52 continuous monitoring are often necessary to identify long-term changes in abundance¹².
53 Although some limited sensitivity analysis was provided with a better sub-dataset involving a
54 minimum of 15 years of data, this strict dataset only included less than 6% of time series.

55 Second, the analysis was performed at a very fine taxonomic level, implying that a high
56 proportion of abundance counts was equal to zero (the full dataset contained 49% of zeros and
57 the strict dataset, 30.5%; moreover, 71% of the series in the full dataset, and 84% in the strict
58 dataset, contained at least a zero). In this context, the log-transformation of the data carried
59 out in the article was inadequate. As the logarithm of 0 is undefined, all zero abundance
60 values were shifted upwards before being log-transformed by adding an arbitrary value. Such
61 rudimentary log-transformation is to be proscribed because results depend on the chosen
62 value¹³. Given the high occurrence of zero counts in the dataset, zero-inflated models,
63 classically used for a long time now, would have allowed dealing appropriately with this
64 problem¹⁴.

65 Third, the model corrected for scale differences between abundance series without accounting
66 for imperfect detection, which can be of particular concern for rare species. This problem may
67 be illustrated by the case of *Aphis asclepiadis*, NEPAC locale, Midwest STN LTER site
68 (external database S1 in ref ¹¹). In the ten years of record included in the meta-analysis, its
69 abundance was 0 for the nine first years, and 1 for the last year. This time series (like the
70 others in the dataset) was not composed of abundance levels, but estimations of abundance.
71 Due to imperfect detection, a shift from an estimated abundance from 0 to 1 provides poor
72 information on the real abundance trend. After scaling log-abundances (see Box 1), this

73 uninformative *A. asclepiadis* data series was erroneously modelled as having the highest
74 abundance increase of all the time series (external database S2 in ref ¹¹), while it could just
75 reflect the rarity of the species or its poor detection. The same slope could have been obtained
76 with a time series actually reflecting a significant abundance change, with for example a
77 hundred insects in all years except one year with a thousand insects. Simple models of
78 occurrence and abundance have already been developed to cope with this problem.¹⁵ Not
79 using them is a serious shortcoming of the article by Crossley et al.

80 A fourth major flaw lies in the drawing of the violin plots in figure 2 of the article. For each
81 LTER, or, depending on cases, each LTER locale, this figure gathered all estimated slopes
82 (one per LLS) which were used to draw a violin plot. However, a violin plot, a median or a
83 mean are statistical estimators built to estimate their counterparts in the population. These
84 estimators are based on the assumption that observations in the sample are independent and
85 identically distributed. But in Crossley et al., each slope was estimated separately from the
86 others and thus had its own variability, which depended on the design of data in the series.
87 Therefore, the slopes were not identically distributed, meaning that violin plots, average
88 abundance trends and their confidence intervals were not relevant and should not have been
89 presented. Moreover, the authors transformed all time units using a common scale varying
90 between 0 (the first year of the LLS abundance time series) and 1 (its last year). Obviously,
91 since time series had different lengths, the time scale varied depending on series and thus
92 slopes cannot be compared. Figure 2 by Crossley et al. and all related discussion are thus
93 meaningless.

94 As all analyses of diversity (richness, evenness and β diversity) in the article relied on these
95 estimations of abundance and on the same modelling, they shared similar methodological
96 problems.

97 We also point out that we had to re-program the R script provided by the authors using their

98 external databases S1 and S2 (ref¹¹) to make it run.

99 Overall, we argue that these methodological shortcomings invalidate the results of the meta-
100 analysis by Crossley et al. The analysis of their dataset would have required an adequate
101 global model for all data, considering all our criticisms and those of ref.¹¹. We call for the
102 application of rigorous standards for meta-analyses on global change, especially because
103 results could have significant impact on policy decision-making and the fate of biodiversity.

104 Insert Table 1.

105

106 **References**

- 107 1 Crossley, M. S. *et al.* No net insect abundance and diversity declines across US Long
108 Term Ecological Research sites. *Nat Ecol Evol* **4**, 1368-1376, doi:10.1038/s41559-
109 020-1269-4 (2020).
- 110 2 Seibold, S. *et al.* Arthropod decline in grasslands and forests is associated with
111 landscape-level drivers. *Nature* **574**, 671-674, doi:10.1038/s41586-019-1684-3 (2019).
- 112 3 Montgomery, G. A. *et al.* Is the insect apocalypse upon us? How to find out.
113 *Biological Conservation* **241**, 108327,
114 doi:<https://doi.org/10.1016/j.biocon.2019.108327> (2020).
- 115 4 Cardoso, P. *et al.* Scientists' warning to humanity on insect extinctions. *Biological*
116 *Conservation* **242**, doi:10.1016/j.biocon.2020.108426 (2020).
- 117 5 Lagos-Kutz, D. and D. Voegtlin. Midwest Suction Trap Network. *Iowa State Research*
118 *Farm Progress Reports*, 2203 (2015). http://lib.dr.iastate.edu/farms_reports/2203.
- 119 6 Landis, D. Insect population dynamics on the main cropping system experiment at the
120 Kellogg Biological Station, Hickory Corners, MI (1989 to 2019) ver 30.

- 121 Environmental Data Initiative (2020).
122 <https://doi.org/10.6073/pasta/f0776c1574808b08c484c1f7645a7357> (Accessed 2020-
123 10-15).
- 124 7 Lightfoot, D. Small Mammal Exclosure Study (SMES) ant data from Chihuahuan
125 desert grassland and shrubland at the Sevilleta National Wildlife Refuge, New Mexico
126 (1995-2005) ver 122263. Environmental Data Initiative (2016).
127 <https://doi.org/10.6073/pasta/c4963aa3363d18ac99bd40307db2249d> (Accessed 2020-
128 10-15).
- 129 8 Grimm, N. & Fisher, S. Sycamore Creek macroinvertebrate collections after flooding
130 event ver 7. Environmental Data Initiative (2013).
131 <https://doi.org/10.6073/pasta/3c1e4c267477ba1b0742c39b049c94ea> (Accessed 2020-
132 10-15).
- 133 9 Georgia Coastal Ecosystems LTER Project and S.C. Pennings. Long-term mid-marsh
134 grasshopper abundance and species diversity at eight GCE-LTER sampling sites ver 5.
135 Environmental Data Initiative (2017).
136 <https://doi.org/10.6073/pasta/ec9d524b8f13c000e8e8a225d5a23c7b> (Accessed 2020-
137 10-15).
- 138 10 Welti, E. A. R. *et al.* Meta-analyses of insect temporal trends must account for the
139 complex sampling histories inherent to many long-term monitoring efforts,
140 *EcoEvoRxiv Preprints* (2020). doi:10.32942/osf.io/v3sr2
- 141 11 Crossley, Michael et al. (2020), No net insect abundance and diversity declines across
142 US Long Term Ecological Research sites, Dryad, Dataset,
143 <https://doi.org/10.5061/dryad.cc2fqz645>
- 144 12 White, E. R. Minimum time required to detect population trends: the need for long-

145 term monitoring programs. *BioScience* **69**, 40-46, doi:10.1093/biosci/biy144 (2019).

146 13 O'Hara, R. B. & Kotze, D. J. Do not log-transform count data. *Methods in Ecology*
147 *and Evolution* **1**, 118-122, doi:10.1111/j.2041-210X.2010.00021.x (2010).

148 14 Lambert, D. Zero-inflated Poisson regression, with an application to defects in
149 manufacturing. *Technometrics* **34**, 1-14, doi:10.2307/1269547 (1992).

150 15 Royle, J. A., Nichols, J. D. & Kéry, M. Modelling occurrence and abundance of
151 species when detection is imperfect. *Oikos* **110**, 353-359, doi:10.1111/j.0030-
152 1299.2005.13534.x (2005).

153

154

Table 1. Times series with non-insect arthropods

Arthropod subphylum or class	Number of time series	Information in the “order” and/or “species” column of external database S2 ¹¹	Involved LTER sites
Time series of only non-insect arthropods			
Arachnida	423	<p>“Amblypygida”, “Anactinotrichidea”, “Araneae”, “Ixodida”, “Mesostigmata”, “Opiliones”, “Oribatida”, “Pseudoscorpiones”, “Scorpiones”, “Solifugae”, “Solpugida” or “Trombidiformes” in the “order” column;</p> <p>“Various” in the “order” column and “Acari” (described as “Arachnida” in the corresponding North Temperate Lakes LTER dataset) in the “species” column;</p> <p>No “order” provided; the “species” column indicates “VAE_COA” or “VAE_RUS”, for the <i>Vaejovis coahuilae</i> and <i>Vaejovis russelli</i> scorpion species</p>	Cedar Creek, Central Arizona Phoenix, Harvard Forest, North Temperate Lakes, Sevilleta
Crustacea	29	<p>“Amphipoda”, “Decapoda”, “Harpacticoida” or “Isopoda” in the “order” column;</p> <p>“Various” in the “order” column and “Copep”, “Clado” or “Ostra” (respectively described as “Maxillopoda”, “Branchiopoda” and “Ostracoda” in the corresponding North Temperate Lakes LTER dataset) in the “species” column</p>	Central Arizona Phoenix, Georgia Coastal Ecosystems, North Temperate Lakes
Entognatha	18	“Collembola” in the “order” column	Central Arizona Phoenix, North Temperate Lakes
Myriapoda	14	<p>“Chilopoda”, “Geophilomorpha”, “Lithobiomorpha”, “Scolopendromorpha”, “Scutigleromorpha” or “Spirobolida” in the “order” column;</p> <p>No “order” provided; “Diplopoda” in the “species” column</p>	Central Arizona Phoenix
Time series including both insects and other arthropods			
No information provided in the dataset	10	No “order” provided; the “species” column, rather than individual species, indicates aquatic invertebrate communities in terms of the following functional feeding groups: mixed substrate filterers, gatherers, predators, scrapers or shredders; or rockface filterers, gatherers, predators, scrapers or shredders	Coweeta

157 **Box 1. Model used by Crossley et al.**

158 Each time series i is composed of abundance levels A_{it} for LLS i and for years t_{i1} to t_{iT_i} .

159 The scaled logarithm of abundance of LLS i in year t , y_{it} , is obtained in two steps. First, the

160 abundance value A_{it} is replaced either by its logarithm, $\log A_{it}$, or, if $A_{it} = 0$, by the logarithm

161 of a constant, $\log c_i$, where c_i is half the minimum non-zero abundance in time series i , to

162 obtain a series of log-transformed abundances a_{it} . Second, the empirical mean \bar{a}_i of log-

163 transformed abundances of the series is subtracted to each a_{it} and this difference is divided by

164 the empirical standard deviation s_i of log-transformed abundances, which yields

165
$$y_{it} = (a_{it} - \bar{a}_i) / s_i.$$

166 The scaled year x_{it} is obtained by transforming the first year of time series t_{i1} to 0, and its last

167 year t_{iT_i} to 1, and scaling all years accordingly as $x_{it} = (t_{it} - t_{i1}) / (t_{iT_i} - t_{i1})$.

168 The proposed modelling is: $y_{it} = \beta_{i1} + \beta_{i2} x_{it} + \varepsilon_{it}$, where $\varepsilon_{it} = \rho_i \varepsilon_{i,t-1} + \eta_{it}$, with $\eta_{it} \sim N(0, \sigma_i^2)$.

169

170

171 **Competing interests.**

172 The authors declare no competing interests.

173