



HAL
open science

How to describe and measure phenology? An investigation on the diversity of metrics using phenology of births in large herbivores

Lucie Thel, Simon Chamailé-Jammes, Christophe Bonenfant

► To cite this version:

Lucie Thel, Simon Chamailé-Jammes, Christophe Bonenfant. How to describe and measure phenology? An investigation on the diversity of metrics using phenology of births in large herbivores. *Oikos*, 2022, 10.1111/oik.08917 . hal-03568554

HAL Id: hal-03568554

<https://hal.science/hal-03568554v1>

Submitted on 12 Feb 2022

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.

1 **How to describe and measure phenology? An investigation**
2 **on the diversity of metrics using phenology of births in**
3 **large herbivores**

4

5 Lucie Thel¹, Simon Chamaille-Jammes^{2, 3, 4} & Christophe Bonenfant^{1, 4}

6

7 **Authors details**

8 ¹ Laboratoire de Biométrie et Biologie Évolutive, Unité Mixte de Recherche 5558, Bâtiment
9 711, Université Lyon I, 43 Boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex,
10 France.

11 ² CEFE, Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier 3, Montpellier,
12 France.

13 ³ Mammal Research Institute, Department of Zoology & Entomology, University of Pretoria,
14 Pretoria, South Africa.

15 ⁴ LTSER France, Zone Atelier “Hwange”, Hwange National Park, Bag 62, Dete, Zimbabwe.

16

17 **Correspondence author:** Lucie Thel, lucie.mc.thel@gmail.com.

18 **Abstract**

19 Proposed in 1849 by Charles Morren to depict periodical phenomena governed by seasons,
20 the term “phenology” has spread in many fields of biology. With the wide adoption of the
21 concept of phenology flourished a large number of metrics with different meaning and
22 interpretation. Here, we first *a priori* classified 52 previously published metrics used to
23 characterise the phenology of births in large herbivores according to four biological
24 characteristics of interest: timing, synchrony, rhythmicity and regularity of births. We then
25 applied each metric retrieved on simulation data, considering normal and non-normal
26 distributions of births, and varying distributions of births in time. We then evaluated the
27 ability of each metric to capture the variation of the four phenology characteristics *via* a
28 sensitivity analysis. Finally, we scored each metric according to eight criteria we considered
29 important to describe phenology correctly. The high correlation we found among the many
30 metrics we retrieved suggests that such diversity of metrics is unnecessary. We further show
31 that the best metrics are not the most commonly used, and that simpler is often better.
32 Circular statistics with the mean vector orientation and mean vector length seems,
33 respectively, particularly suitable to describe the timing and synchrony of births in a wide
34 range of phenology patterns. Tests designed to compare statistical distributions, like Mood
35 and Kolmogorov-Smirnov tests, allow a first and easy quantification of rhythmicity and
36 regularity of birth phenology respectively. By identifying the most relevant metrics our study
37 should facilitate comparative studies of phenology of births or of any other life-history event.
38 For instance, comparative studies of the phenology of mating or migration dates are
39 particularly important in the context of climate change.

40

41 **Keywords:** regularity, rhythmicity, seasonality, synchrony, timing, ungulate

42 **Introduction**

43 In 1849, Charles Morren coined the term “phenology” to describe how periodical phenomena
44 such as plant growth and reproduction are governed by the course of seasons (Morren 1849,
45 see also Demarée 2011). With his observations he opened a new field of research and almost
46 two centuries later the concept of phenology has become a cornerstone of ecology (Begon *et*
47 *al.* 1986), used in plant and animal ecology simultaneously (Forrest and Miller-Rushing
48 2010). By describing when particular life-history events (*e.g.* flowering, parturition) occur in
49 relation to the characteristics or states of the individual (*e.g.* size, age) as well as to
50 environmental factors (*e.g.* photoperiod, predation risk) the concept of phenology is key to
51 understanding the temporal cycles in the life history of species (Forrest and Miller-Rushing
52 2010). Nowadays, the term phenology is commonly employed to describe the temporal
53 occurrence of many aspects of a species biology (*e.g.* moulting, migration, diapause in
54 animals), but the phenology of reproduction (*e.g.* Sinclair *et al.* 2000, Rubenstein and
55 Wikelski 2003, van den Hoff 2020) has attracted most interest. Reproductive phenology is an
56 integral part of life history theory as it is at the heart of inter-generational trade-offs (*i.e.*
57 between parents and offspring) and is a key factor of the reproductive success and fitness of
58 the individuals (Stearns 1989, Forrest and Miller-Rushing 2010). On the one hand, the time
59 of the year when most births occur is often linked to seasonal variations in food resources so
60 that the flush of food resources matches the energetic needs of breeding, which ultimately
61 improves the reproductive success of parents and the fitness of offspring (Plard *et al.* 2015).
62 While on the other hand, the spread of birth dates in a year is supposed to reflect anti-predator
63 strategies to reduce the mortality associated with predation (Darling 1938, Gosling 1969), but
64 also many other social and biological mechanisms (Ims 1990), such as avoidance of male
65 harassment undergone by females (Boness *et al.* 1995) or intra-specific competition between
66 offspring (Hodge *et al.* 2011).

67 In most ecological studies, measurements and observations of phenology are
68 frequently performed at the population level by characterising the temporal distribution of
69 biological events ([Visser *et al.* 2010](#)). These rather complex and variable patterns are reduced
70 to two main components: “timing”, the date at which the event of interest occurs, and
71 “synchrony”, the spread of the dates at which the event occurs, *i.e.* the variability between
72 individuals (Fig. 1). Stimulated by research on the effects of climate change on biodiversity
73 (*e.g.* [Crick and Sparks 1999](#), [Parmesan 2007](#), [Sarkar *et al.* 2019](#)), the question of whether
74 phenology is consistent or varies in time, both at individual and population levels, has
75 received increased interest in recent years (*e.g.* [Renaud *et al.* 2019](#)). We therefore need to
76 quantify two underappreciated properties of phenology: the consistency of the timing and
77 synchrony (at the population scale) of the events from one reproductive season to the next. As
78 these characteristics of phenology are not described by specific words yet, we suggest using
79 “rhythmicity” and “regularity” to describe the consistency of timing and synchrony
80 respectively (Fig. 1), in line with Newstrom’s terminology coined for tropical plants
81 ([Newstrom *et al.* 1994](#)).

82 Despite appearing simple, the concept of phenology carries a lot of confusion in
83 literature, both from a semantic and a descriptive point of view ([Visser *et al.* 2010](#)). Previous
84 studies have explored phenology using a vast diversity of mathematical descriptors, many of
85 which remain specific to a single study. This is problematic as well-defined, comparable and
86 reliable descriptors of the temporal distribution of biological events are key to achieving
87 meaningful comparisons of phenology patterns within or across species. English and
88 colleagues reassessed the most influential factors of reproductive synchrony in large
89 herbivores using the existing literature, but had to narrow their original data set because there
90 was no standardised way of measuring and comparing synchrony across the studies ([English
91 *et al.* 2012](#)). This large diversity of metrics is associated with a lack of widely accepted

92 definitions or divergent definitions for the same word (see “seasonality” *sensu* [Skinner et al.](#)
93 [2002](#) and [Heideman and Utzurrum 2003](#)), which further limits our ability to make meaningful
94 comparisons (*e.g.* [Ryan et al. 2007](#), [Heldstab et al. 2018](#)). As experimental studies are
95 logistically challenging or virtually impossible to conduct with large species, the comparison
96 of phenology patterns within a species living in contrasting environments or across species
97 ([Clauss et al. 2020](#)) is of major importance to assess the role of explanatory factors
98 accounting for the often marked variability in phenology reported in empirical studies
99 ([Rutberg 1987](#)). Such comparative approaches (*sensu* [Felsenstein 1985](#)) indeed shed light on
100 the ecological and evolutionary causes shaping the main stages of the life cycle of organisms
101 ([Bronson 1989](#)).

102 Despite the increasing diversity of approaches to describe phenology, we found only a
103 few attempts to compare phenology metrics and to provide advice on which one should be
104 used preferentially according to the context of the study ([Moussus et al. 2010](#), [Landler et al.](#)
105 [2018](#)). These initiatives are rare and we currently lack a comprehensive comparison of the
106 metrics previously used to characterise phenology. The extent to which the different metrics
107 capture the desired characteristics of the temporal distribution of events, or the sensitivity of
108 those metrics to actual changes in phenology remain to be adequately assessed. Here, we
109 propose such a comparison of metrics based on a literature survey of reproductive phenology
110 in large herbivore species. We focus on the taxonomic group of the large herbivores as it has
111 been studied in a number of species and at different locations ([Rutberg 1987](#)). As a result, we
112 expect to find a wide variety of patterns of births and a wide diversity of metrics to describe
113 them. We first clarify and formally define the four main terms describing phenology: timing,
114 synchrony, rhythmicity and regularity, using our knowledge from the existing literature. We
115 then conduct a comparative analysis of 52 metrics that have been used to quantify the
116 different characteristics of phenology of births in large herbivores, highlighting their

117 strengths and weaknesses. To conclude, we recommend one metric for each of the four main
118 characteristics of phenology.

119

120 **Materials and methods**

121 We conducted a quantitative comparison of a wide range of metrics used to analyse
122 phenology in six steps. In Step 1, we recorded all metrics employed to measure phenology in
123 a selection of papers that we considered representative of the study of phenology of births in
124 large herbivores. In Step 2, we simulated contrasting phenology by varying independently the
125 four parameters that determine timing, synchrony, rhythmicity and regularity of phenology of
126 births (see details below). In Step 3, we calculated all metrics on the simulated phenology to
127 understand how they compare and what characteristic of phenology they measure. In Step 4,
128 we explored the similarities between metrics from a correlation matrix, and identified
129 categories of metrics capturing the same characteristic of phenology. In Step 5, we evaluated
130 the sensitivity of each metric to changes in the estimated parameter. In Step 6, we ranked
131 each metric based on eight criteria that we considered important to identify robust and
132 efficient metrics, but also meaningful from an ecological point of view (see Table 1 for a
133 description of each criterion).

134

135 *Step 1: Retrieving and coding the different phenology metrics*

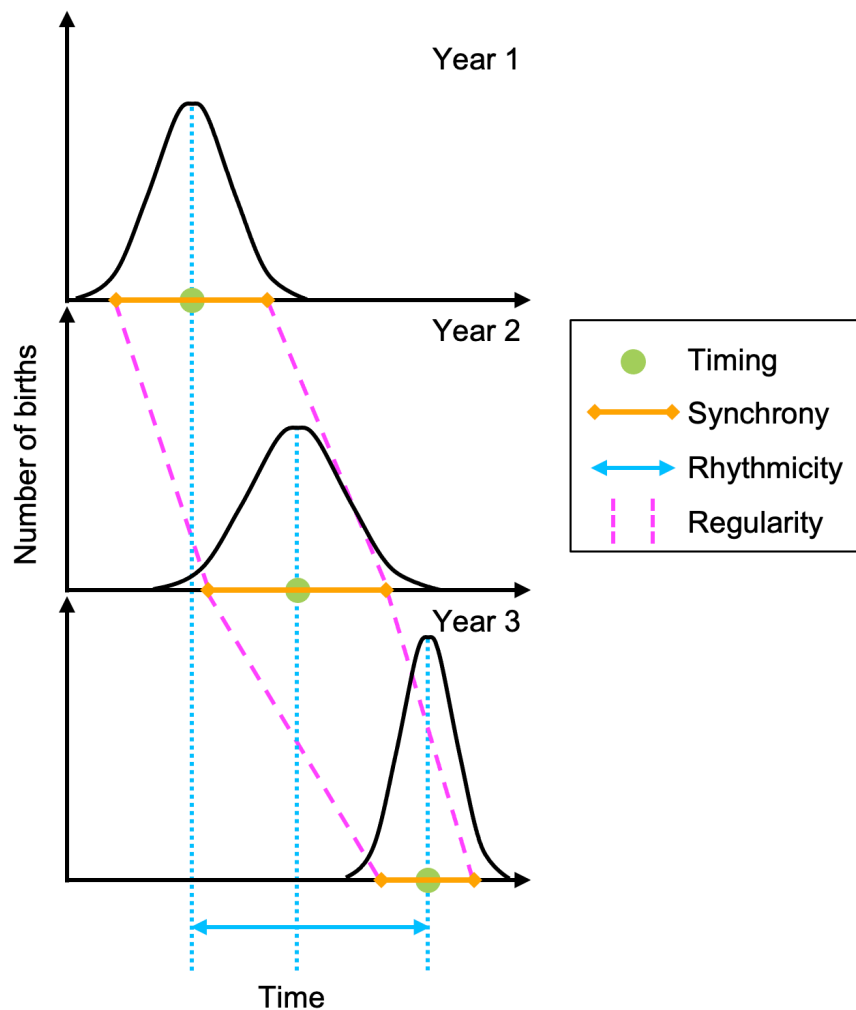
136 We opportunistically searched the literature for articles focusing on the distribution of births
137 in large herbivores using keywords such as “phenology”, “timing”, “synchrony”,
138 “seasonality”, “period” or “season”, and using various sources such as search engines and the
139 references in previously found articles. From these articles, published between 1966 and
140 2019, we recorded the metrics used to describe phenology of births at the population level.

141 We stopped our search once the rate at which we discovered new metrics with additional
142 papers became negligible.

143 We *a priori* classified each metric into one out of four categories based on our
144 understanding of the original description and formula of the metric (Fig. 1): (1) *timing*
145 metrics, defining when within the year most births occur, (2) *synchrony* metrics, defining
146 whether females tend to give birth at the same time in a population in a given year, (3)
147 *rhythmicity* metrics, defining the consistency of timing between years, (4) *regularity* metrics,
148 defining the consistency of synchrony between years. In the literature, the term “seasonality”
149 can be used to describe the location of births in the year (*i.e.* timing, *e.g.* in [Sinclair et al.](#)
150 [2000](#)), the duration of birth period (*i.e.* synchrony, *e.g.* in [Zerbe et al. 2012](#)), and even the fact
151 that births occur at the same period of the year every year (*i.e.* rhythmicity and/or regularity,
152 *e.g.* in [Heideman and Utzurrum 2003](#)). However, this term is initially used to describe the
153 cyclical nature of the environment in a wider range than the study of birth phenology ([Visser](#)
154 [et al. 2010](#)). Thus, it should be used to describe organisms’ phenology only when a direct
155 relationship between periodic environmental phenomena and the cycle of the organism at
156 stake has been demonstrated, which is not always the case in phenology studies. For this
157 reason, we suggest using the term “seasonality” only to describe the cyclicity of the
158 environment and prefer the use of neutral terms such as those we introduced in this paper to
159 describe phenology of births: rhythmicity and regularity.

160 Forty-seven articles (Supporting information 1) presented at least one mathematically-
161 defined phenology metric yielding 52 different metrics. In order to compare metrics
162 quantitatively, we slightly tweaked some of them: when the metric was a boolean (true/false)
163 variable based on the significance of a statistical test ($n = 9$ metrics), we used the value of the
164 test statistic as output metric, thereby allowing us to investigate how the statistic was
165 influenced by the value of phenology parameters (see details in Supporting information 2).

166 All metrics could be coded in R software ([R Core Development Team 2019](https://www.r-project.org/)) except one, for
167 which Perl was used (www.perl.org).



168
169 *Figure 1: Four characteristics of phenology of births can be explored to fully describe phenology at*
170 *the population scale: timing, synchrony, rhythmicity and regularity. Timing describes when within the*
171 *year most births occur, synchrony illustrates whether females tend to give birth at the same time in a*
172 *population in a given year, rhythmicity defines the consistency of timing between years, regularity*
173 *refers to the consistency of synchrony between years. Green = timing, orange = synchrony, blue =*
174 *rhythmicity, pink = regularity.*

175

176 *Step 2: Simulating phenology of births*

177 We simulated phenology of births from statistical distributions with known parameters

178 (Supporting information 2) to assess what characteristic of phenology (timing, synchrony,

179 rhythmicity, regularity) each metric would capture, their sensitivity to changes into these four
180 key characteristics of interest, and the correlation between the 52 metrics. We simulated the
181 distributions of births over a year as most large herbivores breed once a year. This choice
182 does not limit the generality of our results: for species breeding more than once per year (*e.g.*
183 small species with short gestation length such as dikdik *Rynchotragus (Madoqua) kirki*,
184 [Sinclair et al. 2000](#)), the same metrics may be applied on sub-periods of time, each displaying
185 only one birth peak (see [Heideman and Utzurrum 2003](#) for a similar approach in bats).

186 Each simulated phenology was generated by randomly distributing births in time,
187 following a normal distribution. We distributed $n = 1000$ births within a year of 365 days,
188 repeated over 10 years (see why in “Material and Methods” section, step 3). We changed four
189 parameters independently to modify the distribution of births: the mean day of birth for a
190 given year (*mean*), the standard deviation of the distribution of births for a given year (*sd*),
191 the range over which the mean birth date can vary across years ($\Delta mean$), and the range over
192 which the standard deviation can vary across years (Δsd). Each parameter varied in a range
193 from a minimum to a maximum value and was incremented with a constant step (Supporting
194 information 2). Choosing the value of these parameters allowed us to simulate changes in the
195 timing, synchrony, rhythmicity and regularity of the phenology of births independently. As
196 the simulated phenology of births relied on random draws, the actual values of parameters in
197 the simulated distribution of births could differ from the theoretical values used in the
198 simulation algorithm. We used the realised values of the distribution parameters in the
199 following analyses. Note that we replicated the same analyses using non-normal distributions
200 of births (*i.e.* skewed normal, bimodal, Cauchy, and random distributions) to cover the
201 variety of empirical distributions of births observed in *natura* and assess robustness to non-
202 normality (Supporting information 4). We performed all simulations using the R software and
203 made the code available on GitHub (<https://github.com/LucieThel/phenology-metrics>).

204

205 *Step 3: Computing the phenology metrics from simulated patterns of births*

206 Among the 52 phenology metrics we analysed, most applied to a single year, but others
207 required two or more years of data to be computed (see the complete list in Supporting
208 information 3). As we aimed to compute all 52 metrics, we chose to simulate annual
209 distributions of births over 10 consecutive years by default. For each simulation, we used data
210 from the first year to compute metrics requiring only one year of data ($n = 33$ metrics), data
211 from the first two years for metrics requiring two years of data ($n = 9$ metrics), and data from
212 the whole simulation for the other metrics ($n = 10$ metrics).

213

214 *Step 4: Comparing the metrics*

215 With the results from Step 3, we computed the global correlation matrix between all pairs of
216 metrics using Pearson correlations. We then identified groups of strongly correlated metrics
217 from the pairwise correlation coefficients and assigned each metric to one or several of the
218 four characteristics of phenology it was best related to. We compared this categorisation with
219 our *a priori* classification of the metrics. This step enabled us to check our intuitive
220 classification of the metrics in addition to revealing whether some metrics could incidentally
221 capture several aspects of the distribution of births at once.

222

223 *Step 5: Estimating the sensitivity of the metrics*

224 For each metric, we performed a sensitivity analysis by quantifying the observed variation of
225 each metric with a fixed variation in the characteristic of phenology it was previously
226 associated with in Step 4. We did this by computing, for each possible pair of simulations
227 within the set of all simulations performed, the proportional difference between the realised
228 values of the phenology parameter of interest of the two simulations, and the proportional

229 difference between the values of the metric of interest of the same two simulations. In each
230 case the proportional difference was calculated as $[(Value_{max} - Value_{min}) / Value_{min}] * 100$.

231 This formulation allowed us to work with positive values only as we were interested in the
232 amplitude but not in the direction of the differences.

233

234 *Step 6: Scoring metrics*

235 Finally, as there were too many different metrics, we were unable to discuss the pros and
236 cons for each of them. We chose instead to provide guidance about the usefulness of the
237 different metrics by scoring them according to a set of eight criteria that we considered as
238 important behaviour for a metric to be relevant (Table 1). Having systematic criteria helped
239 us to minimise the subjectivity of the scoring so we ranked the metrics from 0 (not advised)
240 to 8 (strongly advised) according to the number of criteria they fulfilled. The proposed
241 criteria (Table 1) consisted in verifying if 1) the metric varied according to the phenology
242 characteristic it was supposed to measure, 2) the variation of the metric according to the
243 phenology characteristic was monotonous, 3) the relationship with the characteristic of
244 phenology was strong (visual assessment of the association between the computed statistic
245 and the phenology characteristic), 4) the metric did not saturate within a biologically realistic
246 range of distributions of births. We considered that metrics with scores < 4 for which the first
247 four essential criteria were not validated should not be advised. If those four criteria were
248 satisfied, we evaluated an additional set of four criteria (normality, independence of the
249 temporal origin, linearity and unicity of the output, see Table 1 for a detailed description). All
250 criteria were scored from visual inspection of the results by one of us (LT).

251

252 *Table 1: Ordered list of the criteria used to evaluate the relevance of each metric describing*
 253 *phenology of births. Each criterion can be individually fully (score of 1) or partially (score of 0.5)*
 254 *validated or no (score of 0) by each metric. The value for the first four criteria (in bold type) should*
 255 *be > 0 to consider a metric to be possibly worthwhile and evaluate the remaining criteria. The sum of*
 256 *the value obtained for each criterion gives the relevance index of the metric (range between 0 and 8*
 257 *points).*

Criterion	Description	Score
goodness	Measures the parameter it is expected to measure	true = 1 false = 0
monotony	Varies monotonically with the value of the parameter it is expected to measure, <i>i.e.</i> the sign of the slope coefficient is constant	true = 1 false = 0
saturation	Does not saturate at the upper or lower boundary in a biological range of values (<i>e.g.</i> if a synchrony metric returned the same value when all births occurred during periods of various durations such as fifteen or thirty days, it was considered to saturate within a biologically realistic range of birth distributions because such distributions of births can be found in the wild)	true = 1 false = 0
strength¹	Is characterised by a strong relationship with the parameter it is expected to measure, <i>i.e.</i> is the scatter plot not too dispersed around the general trend of the relationship between the metric and the phenology characteristic, as an empirical approach of the predictive power?	high = 1 medium = 0.5 low = 0 (strength of the association)
normality	1- Does not assume normally distributed birth dates; if false (assumes normality): 2- Is it robust to deviations to normality? <i>i.e.</i> , is the relationship between the metric and the parameter it is expected to measure conserved when births are not normally distributed	true = 1 false-true = 1 false-false = 0
origin	Does not depend on the temporal origin set by the investigator	true = 1 false = 0
linearity ²	Is characterised by a linear relationship with the parameter it is expected to measure	type 1 = 1 type 2 and 3 = 0.5 type 4 = 0
unicity	Gives a unique result	true = 1 false = 0

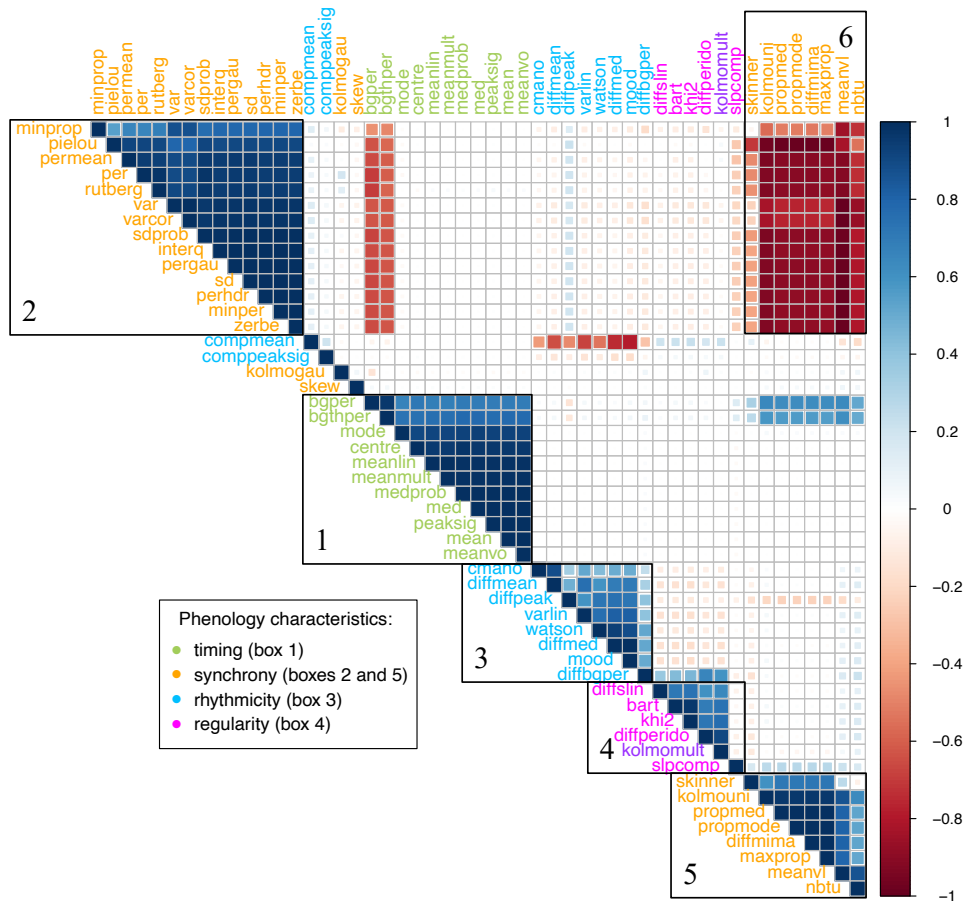
258 ¹ *high association: very small dispersion of points, medium association: small dispersion of points*
 259 *that does not prevent from detecting a trend, low association: dispersion of points large enough to*
 260 *prevent from detecting any trend, whatever the shape of the relationship (linear, but also sigmoid or*
 261 *quadratic for instance). ² type 1 is a linear relationship, type 2 is a sigmoid-like relationship, type 3 is*
 262 *a quadratic-like relationship, type 4 is a binary relationship.*

263

264 **Results**

265 The mean number of metrics used in each paper was 3.8 ± 2.1 sd (*range* = 1 - 8). Eleven
266 metrics were *a priori* associated with timing, 25 with synchrony, 10 with rhythmicity and five
267 with regularity. We did not classify one metric because it could either be a rhythmicity or
268 regularity metric *a priori*. Those metrics were based on descriptive statistics, circular
269 statistics, statistical tests or statistical modelling such as general linear models. The unit of the
270 metrics were date, duration, counts (*e.g.* a number of births), binary classification (*i.e.* if a
271 given condition was satisfied or not), or unitless indices (Supporting information 3).

272 The correlation matrix (Step 4) revealed groups of metrics that were highly correlated
273 and thus reflected the same characteristic of phenology (Fig. 2). Five groups were clearly
274 identifiable, representing timing metrics (Fig. 2 - box 1), synchrony metrics (Fig. 2 - boxes 2
275 and 5), rhythmicity metrics (Fig. 2 - box 3), and regularity metrics (Fig. 2 - box 4). The two
276 groups of metrics measuring synchrony had highly but negatively correlated values (Fig. 2 -
277 box 6). This indicated that all metrics of the two groups captured synchrony correctly,
278 however, in an opposing way. Three metrics were singular and were associated with neither
279 of the five groups. The metric which compares the slope coefficients of linear models
280 describing the log percent of cumulative births ("*splcomp*") should measure regularity, but it
281 rather correlated better with synchrony metrics. The metric which evaluates the duration
282 between the first birth dates of two reproductive cycles ("*diffbgper*"), an assessment of
283 rhythmicity, correlated well with both rhythmicity and regularity metrics. Seven other metrics
284 had a detectable relationship with at least one of the three remaining phenology
285 characteristics in addition to the relationship with the phenology characteristic they were
286 supposed to quantify (Supporting information 3 and 5).



287

288 *Figure 2: Correlation matrix between all pairs of metrics, using Pearson correlations ($n = 51$,*
 289 *“rayleigh” removed because because of no observed variation). It was not possible to classify*
 290 *“kolmomult” a priori in rhythmicity or regularity metrics, as it compares the complete distribution of*
 291 *births between two years. Box 6 highlights the high but negative correlation between the two groups*
 292 *of metrics measuring synchrony (boxes 2 and 5). Green = timing metrics, orange = synchrony*
 293 *metrics, blue = rhythmicity metrics, pink = regularity metrics. Note the high negative correlation*
 294 *between “compmean” and the other rhythmicity metrics, highlighting that it is also a rhythmicity*
 295 *metric.*

296

297

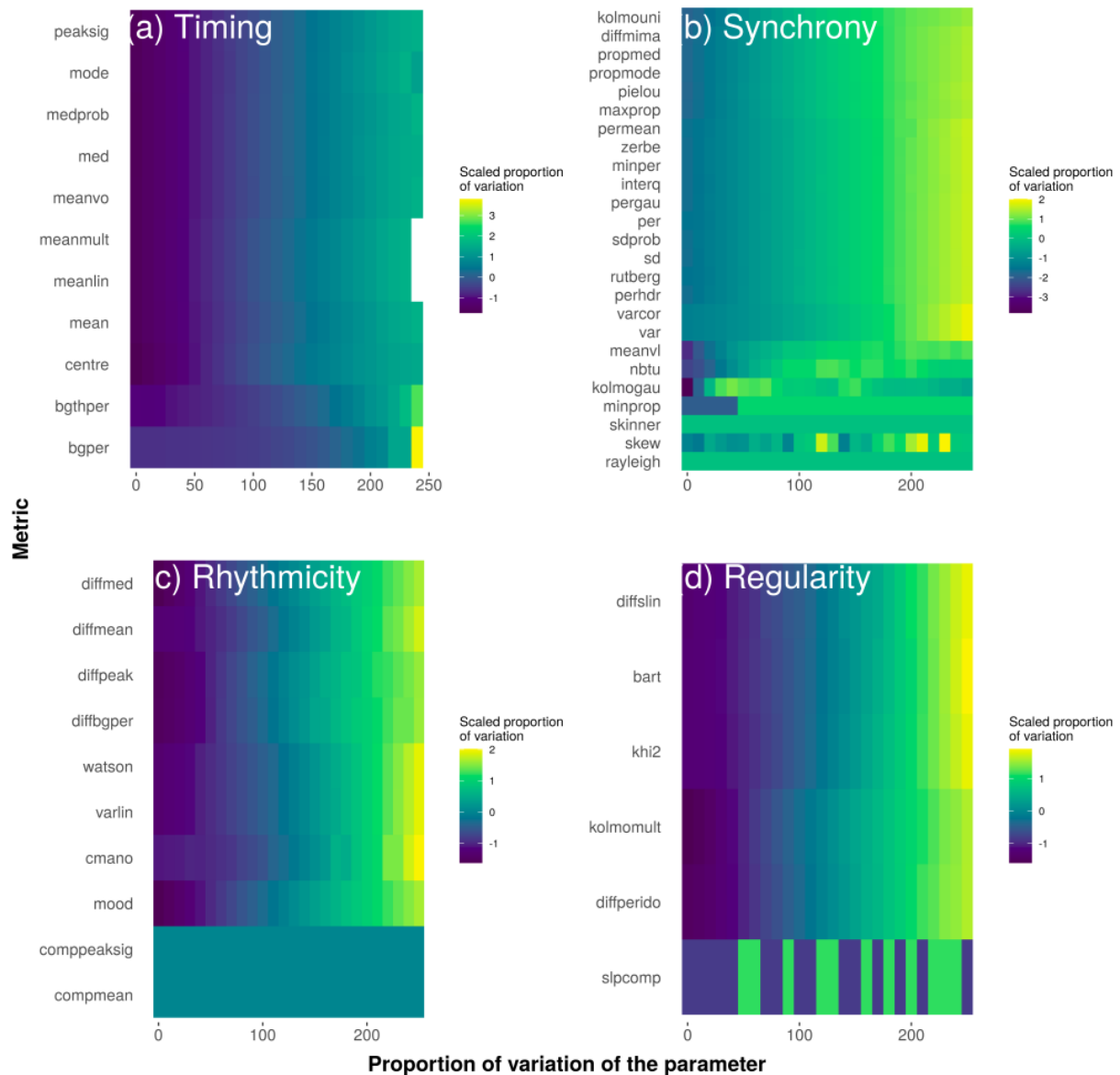
298

299

300

The sensitivity of the metrics to the simulated variation of the phenology characteristics (Step 5) differed markedly between metrics, especially in synchrony and regularity metrics (Fig. 3 and Supporting information 5). The proportion of variation of the metrics for a 10 % variation of the associated parameter ranged from 14 % to 33 % for timing

301 metrics, from 0 % to 139 % for synchrony metrics, from 0 % to 471 % for rhythmicity
302 metrics and from 0 % to 138 % for regularity metrics. The variation of almost all timing,
303 rhythmicity and regularity metrics according to variations of their associated parameter was
304 highly homogeneous. Synchrony metrics were less homogeneous, certainly due to the fact
305 that those metrics were the most numerous and based on more diverse methods (proportion of
306 variation, integrative indexes or moments of the distribution of births, for instance). The
307 metrics that were singular in the correlation matrix were clearly visible in the heat maps,
308 characterised by erratic or non-existent variations (*e.g.* skewness of the birth distribution
309 “*skew*”, and comparison of mean date of births “*compmean*”).



310

311 *Figure 3: Heat maps representing the (scaled) proportion of variation of the metric in relation to the*

312 *proportion of variation of the parameter of phenology (sensitivity analysis): a) timing metrics*

313 *according to the mean birth date for a given year (mean, $n = 11$), b) synchrony metrics according to*

314 *the standard deviation of the distribution of births for a given year (sd , $n = 25$), c) rhythmicity metrics*

315 *according to the range over which the mean birth date can vary across years ($\Delta mean$, $n = 10$), d)*

316 *regularity metrics according to the range over which the standard deviation of the distribution of*

317 *births can vary across years (Δsd , $n = 6$).*

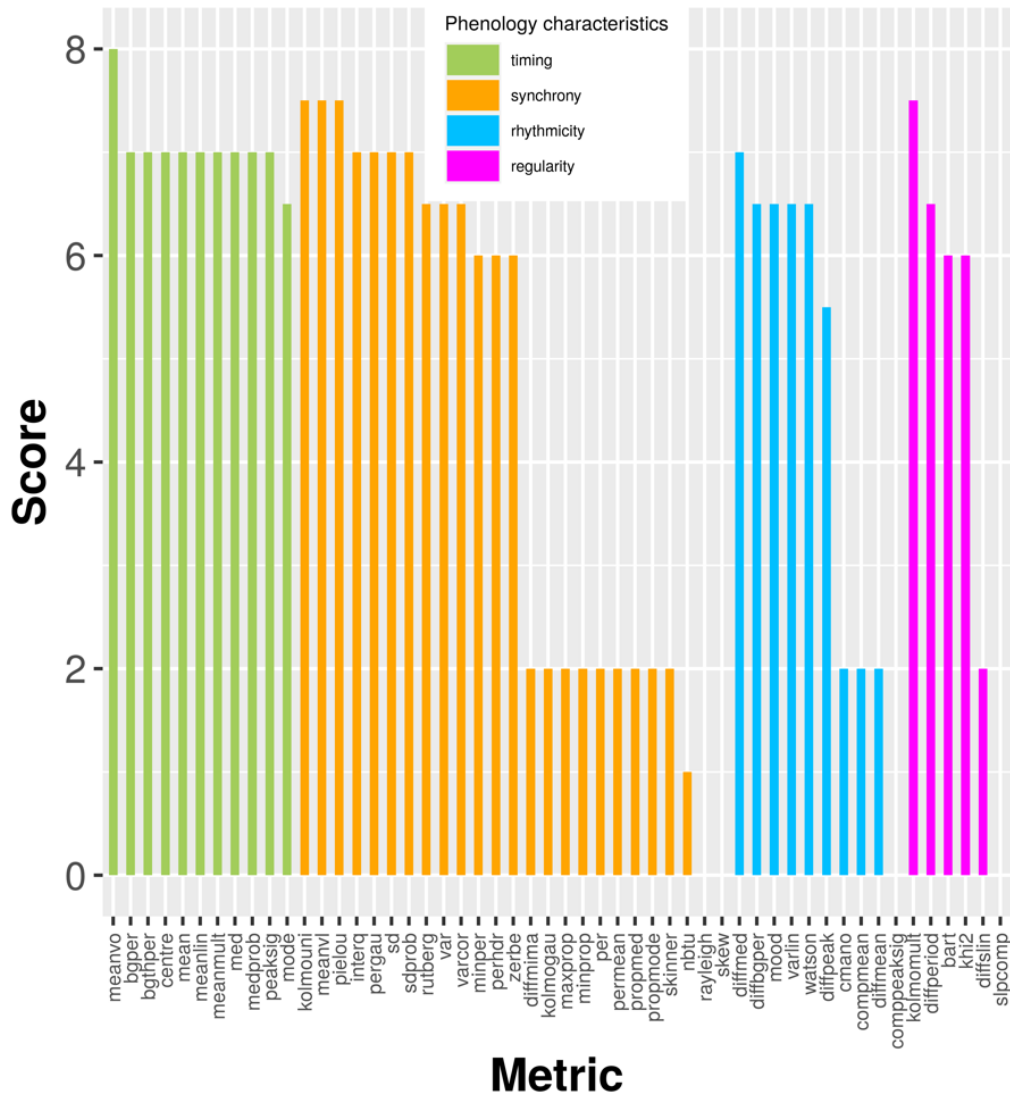
318 *Colours in the heat maps reflect the proportion of variation of each metric according to the*
319 *proportion of variation of the phenology parameter, normalised for each metric using all values of the*
320 *metric obtained across all simulations. We normalised the sensitivity of each metric individually to*
321 *prevent the representation of the large variation of some metrics to hide the smaller but meaningful*
322 *variations of other metrics to be visible. Metrics characterised by a large colour gradient vary widely*
323 *in response to the variation of the parameter of phenology they measure. Metrics with a smoothed*
324 *colour transition vary regularly in response to the variation of the parameter of phenology they*
325 *measure. To the contrary, metrics characterised by sudden and/or random colour transitions vary*
326 *inconsistently in response to the variation of the parameter of phenology we changed.*

327

328 The same analyses conducted on the basis of non-normal distributions led to similar
329 observations in the case of asymmetric distributions (skewed normal, bimodal and Cauchy
330 distributions). The correlation matrices showed similar patterns of correlations between the
331 metrics, and the metrics varied analogously according to the variation of the *mean*, *sd*, Δ *mean*
332 and Δ *sd* of the distributions for normal and asymmetric distributions either (see Supporting
333 information 4 for a detailed analysis). Nevertheless, it is worth noting that a very limited
334 number of metrics depending on the skewness of the distribution did not perform as well with
335 the normal distribution than with asymmetric distributions. On the contrary, metrics
336 depending on the presence of a period without any birth did not perform as well with non-
337 normal distributions than with a normal distribution. In the case of a random distribution, no
338 clear correlations between metrics nor relationships between the metrics and the four
339 parameters of the distribution were detectable, except for some rare synchrony and timing
340 metrics (Supporting information 4).

341 The relevance score of the metrics (step 6) varied between 0 and 8, covering the
342 complete range of variation possible (Fig. 4) and we list, for each phenology characteristic,
343 the metrics we identified as “best” (Table 2). Our classification also revealed what could be

344 considered as ineffective ($score = 0$, $n = 4$) and poor metrics ($score \in [0; 4[$, $n = 14$). All the
345 timing metrics reached excellent scores above 6. Nevertheless, the mean vector orientation
346 (“*meanvo*”) was the best metric, fulfilling all our criteria with a score of 8 (Fig. 4). Three
347 metrics provided a very good assessment of the synchrony of births with a score of 7.5: the
348 evenness index (“*pielou*”), the mean vector length (“*meanvl*”) and the comparison of the
349 distribution of births to a uniform distribution (“*kolmouni*”) (Fig. 4). The best metric to
350 quantify rhythmicity measured the time elapsed between the median birth dates of two years
351 (“*diffmed*”), with a score of 7 (Fig. 4). It is worth noting that the non-parametric Mood test
352 (“*mood*”) provides a statistical assessment of whether “*diffmed*” differs from 0. The non-
353 parametric Mood test (“*mood*”) obtained a marginally lower score (6.5, Fig. 4) than “*diffmed*”
354 only because of a slight non-linearity in the relationship between simulation parameter values
355 and the metric’s statistics. Altogether, we therefore considered that “*mood*” could be very
356 useful to measure rhythmicity. One metric quantifying regularity stood out from the others
357 according to our criteria: the non-parametric Kolmogorov-Smirnov test (“*kolmomult*”), which
358 compares two birth distributions ($score = 7.5$, Fig. 4).



359

360 *Figure 4: Score obtained by each phenology metric (n = 52) according to the eight criteria used to*
 361 *assess its relevance to characterise the four main characteristics of birth phenology (goodness,*
 362 *monotony, saturation, strength, normality, origin, linearity and unicity, as defined in Table 1). Green*
 363 *= timing metrics, orange = synchrony metrics, blue = rhythmicity metrics, pink = regularity metrics.*

364 *Table 2: List of the metrics considered as the best metric, for each characteristic of the phenology of*
365 *births (timing, synchrony, rhythmicity, regularity).*

Phenology characteristic	Metric	Complete name	Description	Reference
Timing	<i>meanvo</i>	mean vector orientation	evaluates mean vector orientation of the birth distribution	Paré <i>et al.</i> 1996
Synchrony	<i>meanvl</i>	mean vector length	evaluates mean vector length of the birth distribution	Paré <i>et al.</i> 1996
Rhythmicity	<i>mood</i>	Mood test	compares median birth dates between two years	Berger and Cain 1999
Regularity	<i>kolmomult</i>	Two-sample Kolmogorov-Smirnov test	compares birth distributions between two years	Green and Rothstein 1993

366

367 **Discussion**

368 With more than fifty metrics used to describe and analyse the distribution of births in large
369 herbivores since 1966, our survey of the literature clearly illustrates the diversity of
370 approaches, even when focusing on a specific taxonomic group. Although the choice of a
371 metric is most of the time justified, either to answer a specific ecological question or on
372 statistical grounds, the lack of consensual methods to quantify phenology makes comparisons
373 across species or populations difficult at best, if possible at all. Our simulation study suggests
374 that such a diversity of metrics may cause confusion and be unnecessary as we were able to
375 identify a reduced set of simple metrics that works well to measure the different
376 characteristics of phenology. Moreover, we believe our work can also provide insights into
377 how to analyse phenology of other traits than birth dates, such as migration dates of birds or
378 flowering dates ([Moussus *et al.* 2010](#)).

379 Many of the metrics we retrieved can be organised into four main categories, each one
380 capturing a particular characteristic of phenology: timing, synchrony, rhythmicity and
381 regularity. Of course, metrics belonging to the same category are not perfectly equivalent and

382 interchangeable (Fig. 2, see also a discussion comparing “*zerbe*” and “*rutberg*” metrics in
383 [Zerbe et al. 2012](#)). For instance, the correlations between timing metrics range between 0.68
384 and 1.00. The difference among metrics is more pronounced in the synchrony category with
385 correlations ranging from 0.05 to 1.00 (excluding “*kolmogau*” and “*skew*” metrics that appear
386 as singularities in the correlation matrix, Fig. 2). How different characteristics of phenology
387 are measured can lead to dependency between one another and this could explain the
388 confusions found in the literature between timing and synchrony through terms such as
389 “period” or “season” of births. Indeed, several of the metrics we tested vary not only
390 according to the phenology characteristic they were used to measure, but also according to
391 other characteristics of the phenology ($n = 8$ metrics). For instance, we show a strong
392 correlation between metrics that evaluate the start of the birth period (*i.e.* timing metrics
393 “*bgper*” and “*bgthper*”) and the synchrony metrics in general. This association between
394 different types of metrics arises when the standard deviation of the simulated distributions of
395 births increases (while the mean is fixed), leading to earlier births (Fig. 2).

396 We attempted to identify what metrics could be the most suitable for measuring
397 timing, synchrony, regularity and rhythmicity of phenology by scoring them according to
398 what we subjectively considered as the main suitable properties. We considered that a good
399 metric should not be restricted to one kind of pattern (*e.g.* unimodal) as the distribution of
400 births is not necessarily known a priori and may change between years due to ecological
401 factors (see [Adams and Dale 1998](#) for instance). Slightly more than 10 % of the metrics
402 theoretically require normally distributed dates of birth to work well (based on the metrics for
403 which this criterion was evaluated, Supporting information 3). We showed these metrics are
404 generally robust to deviations from normality so this assumption does not limit their
405 application to most data. The metrics should also be independent of the temporal origin set by
406 the investigator, as the favourable periods for reproduction cycle differ between species and

407 populations (*e.g.* mountain sheep *Ovis spp.* inhabiting desert and alpine ecosystems, [Bunnell](#)
408 [1982](#)). Using the calendar year would be biologically meaningless and will create artificial
409 patterns of births by splitting the distribution around the end of the year. We identified six
410 metrics independent of the temporal origin: the day with the highest number of births
411 (“*mode*”), the evenness index (“*pielou*”), the mean vector orientation and length from the
412 circular statistics (“*meanvl*” and “*meanvo*” respectively), and the non-parametric
413 Kolmogorov-Smornov test comparing a birth distribution to a uniform distribution or another
414 birth distribution (“*kolmouni*” and “*kolmomult*” respectively). Circular statistics could be
415 favoured to answer the difficulties linked to the selection of temporal origin, as it is
416 frequently done in primate literature (*e.g.* [Di Bitetti and Janson 2000](#)). Notwithstanding such
417 limitations, we found several metrics that met our expectations of a good metric for each
418 phenology characteristic (Table 2 and Figure 4).

419 On the other side a few metrics should not be recommended to describe phenology of
420 births. The evaluation of rhythmicity describing the evolution of the mean dates of births of
421 several years with a linear regression (“*diffmean*”), or the quantification of synchrony through
422 the duration of the period gathering at least a certain percent of births (“*nbtu*”) are not to be
423 advised. In addition to undesirable statistical properties, these metrics fail to capture the
424 changes in the phenology parameter adequately. The metric “*nbtu*” varied non-monotonously
425 with the level of synchrony of the birth phenology. Similarly, the duration between first and
426 last birth to measure synchrony (“*per*”) plateaued for a range of biologically realistic values,
427 what limits its usability in a wide range of ecological conditions.

428 Overall, some phenology characteristics have been more consistently evaluated across
429 studies, a fact illustrated by the number of metrics of each category used in more than two
430 papers ($n = 5, 7, 2$ and 0 for timing, synchrony, rhythmicity and regularity respectively, see
431 Supporting information 6). If timing and synchrony of births are the easiest and most frequent

432 characteristics of phenology estimated and compared, only a handful of metrics evaluates
433 rhythmicity and regularity of the phenology of births across the years. Sound analysis of
434 rhythmicity and regularity indeed requires many years of data which may not be available as
435 such data is costly and time-consuming to collect (Kharouba and Wolkovich 2020).
436 Moreover, scientists are less interested in timing and synchrony consistency *per se* than in the
437 relationship between timing and synchrony, and ecological or environmental factors such as
438 temperature, rainfall or spring snow cover (Paoli *et al.* 2018). Our study shows that the
439 rhythmicity and regularity metrics currently available are only moderately correlated,
440 particularly when they are used to describe birth distributions that are not normally
441 distributed (Supporting information 4). Capturing the temporal variation of phenology across
442 years appears difficult and requires thoughtful selection and interpretation of the used metric.
443 Standardised and relevant statistical tools are needed to quantify regularity and rhythmicity of
444 phenology, and to test their hypothetical responses to global changes. This study should help
445 in this.

446 Although we show that the assumption of a normal distribution or another bell-shaped
447 (asymmetric or not) distribution mimicking those found in *natura* (*e.g.* skewed normal,
448 bimodal or Cauchy distribution) has no major consequences on our conclusions (Supporting
449 information 4), this is not true when there is no clear pattern in the distribution of births.
450 Indeed, most metrics give inconsistent and unreliable results when applied to birth dates
451 randomly distributed within the year (Supporting information 4), a pattern that has been
452 documented in some populations of large herbivores living in the southern hemisphere
453 (Sinclair *et al.* 2000). Describing random patterns using the metrics presented here is unlikely
454 to be useful because biologically meaningless: when births occur year-round, the timing and
455 rhythmicity are meaningless as they cannot reduce to one or two summarising statistics.

456 Using evenness indexes such as “*pielou*” could at least provide a quantification of the
457 heterogeneity of the distribution of births.

458 In conclusion, we recommend using the circular mean vector orientation (“*meanvo*”)
459 to describe timing and the circular mean vector length (“*meanvl*”) to describe synchrony,
460 because both are not influenced by the temporal origin set by the investigator. We
461 recommend using the underused Mood test which statistically compares the median birth
462 dates (“*mood*”) to describe rhythmicity and the Kolmogorov-Smirnov test which statistically
463 assesses if two birth distributions are similar to describe regularity (“*kolmomult*”, see Table 2
464 and Supporting information 3 for a formal description of those metrics). Being non-
465 parametric tests, they are applicable in a wide range of distributions as frequently observed in
466 large herbivore populations.

467

468 **Declarations**

469 **Acknowledgments:** The authors thank Simon Penel for his help with Perl Programming
470 Language, and Marcus Clauss for providing the code of one of the metrics. The authors thank
471 two anonymous reviewers and the subject editor for their comments on a previous version of
472 the manuscript, and Marion Valeix for her comments on early drafts of the manuscript. They
473 deeply appreciated the insightful comments, editorial suggestions and proofreading by Anna
474 Cryer, Leif Egil Loe and Primaëlle Fusto. This work was performed using the computing
475 facilities of the CC LBBE/PRABI and the IFB Cloud.

476 **Funding:** This work was supported by a grant from the "Ministère Français de
477 l'Enseignement Supérieur, de la Recherche et de l'Innovation" through the "Ecole Doctorale
478 E2M2" of the "Université Claude Bernard Lyon 1".

479 **Authors' contributions:** LT, CB and SCJ conceived the ideas and designed the
480 methodology; LT performed data collection and analysed the data; LT wrote the first draft of

481 the manuscript, which was then edited by all authors. All authors gave final approval for
482 publication.

483 **Data availability statement:** Code available on GitHub:

484 <https://github.com/LucieTheI/phenology-metrics>.

485 **References**

- 486 Adams, L. G. and Dale, B. W. 1998. Timing and synchrony of parturition in Alaskan caribou.
487 - *J. Mammal.* 79: 287–294.
- 488 Begon, M. *et al.* 1986. *Ecology. Individuals, populations and communities.* - Blackwell
489 scientific publications.
- 490 Berger, J. and Cain, S. L. 1999. Reproductive synchrony in brucellosis-exposed bison in the
491 southern Greater Yellowstone Ecosystem and in noninfected populations. - *Conserv. Biol.*
492 13: 357–366.
- 493 Boness, D. J. *et al.* 1995. Does male harassment of females contribute to reproductive
494 synchrony in the grey seal by affecting maternal performance? - *Behav. Ecol. Sociobiol.*
495 36: 1–10.
- 496 Bronson, F. H. 1989. *Mammalian reproductive biology.* - University of Chicago Press.
- 497 Bunnell, F. L. 1982. The lambing period of mountain sheep: synthesis, hypotheses, and tests.
498 - *Can. J. Zool.* 60: 1–14.
- 499 Clauss, M. *et al.* 2020. Basic considerations on seasonal breeding in mammals including their
500 testing by comparing natural habitats and zoos. - *Mamm. Biol.:* 1–14.
- 501 Crick, H. Q. P. and Sparks, T. H. 1999. Climate change related to egg-laying trends. - *Nature*
502 399: 423.
- 503 Darling, F. 1938. *Bird flocks and the breeding cycle.* - Cambridge University Press.
- 504 Demarée, G. R. 2011. From “Periodical Observations” to “Anthochronology” and
505 “Phenology”- the scientific debate between Adolphe Quetelet and Charles Morren on the
506 origin of the word “Phenology.” - *Int. J. Biometeorol.* 55: 753–761.
- 507 Di Bitetti, M. S. and Janson, C. H. 2000. When will the stork arrive? Patterns of birth
508 seasonality in neotropical primates. - *Am. J. Primatol. Off. J. Am. Soc. Primatol.* 50: 109–
509 130.

- 510 English, A. K. *et al.* 2012. Reassessing the determinants of breeding synchrony in ungulates.
511 - PLoS One 7: e41444.
- 512 Felsenstein, J. 1985. Phylogenies and the comparative method. - Am. Nat. 125: 1–15.
- 513 Festa-Bianchet, M. 1988. Birthdate and survival in bighorn lambs (*Ovis canadensis*). - J.
514 Zool. 214: 653–661.
- 515 Forrest, J. and Miller-Rushing, A. J. 2010. Toward a synthetic understanding of the role of
516 phenology in ecology and evolution. - Philos. Trans. R. Soc. B Biol. Sci. 365: 3101–3112.
- 517 Gosling, L. M. 1969. Parturition and related behaviour in Coke's hartebeest, *Alcelaphus*
518 *buselaphus cokei* Günther. - J. Reprod. Fertil. Suppl. 6: 265–286.
- 519 Green, W. C. H. and Rothstein, A. 1993. Asynchronous parturition in bison: implications for
520 the lider-follower dichotomy. - J. Mammal. 74: 920–925.
- 521 Heideman, P. D. and Utzurrum, R. C. B. 2003. Seasonality and synchrony of reproduction in
522 three species of nectarivorous Philippines bats. - BMC Ecol. 3: 11.
- 523 Heldstab, S. A. *et al.* 2018. Geographical Origin, Delayed Implantation, and Induced
524 Ovulation Explain Reproductive Seasonality in the Carnivora. - J. Biol. Rhythms 33: 402–
525 419.
- 526 Hodge, S. *et al.* 2011. Reproductive competition and the evolution of extreme birth
527 synchrony in a cooperative mammal. - Biol. Lett. 7: 54-56.
- 528 Ims, R. A. 1990. The ecology and evolution of reproductive synchrony. - Trends Ecol. Evol.
529 5: 135–140.
- 530 Kharouba, H. M. and Wolkovich, E. M. 2020. Disconnects between ecological theory and
531 data in phenological mismatch research. - Nat. Clim. Chang.: 1–10.
- 532 Landler, L. *et al.* 2018. Circular data in biology: advice for effectively implementing
533 statistical procedures. - Behav. Ecol. Sociobiol. 72: 128.

- 534 Morren, C. 1849. Le globe, le temps et la vie. - Bulletins de l'Académie royale des Sciences,
535 des Lettres et des Beaux-Arts de Belgique 16: 660–684.
- 536 Moussus, J.-P. *et al.* 2010. Featuring 10 phenological estimators using simulated data. -
537 Methods Ecol. Evol. 1: 140–150.
- 538 Newstrom, L. E. *et al.* 1994. A new classification for plant phenology based on flowering
539 patterns in lowland tropical rain forest trees at La Selva, Costa Rica. - Biotropica: 141–
540 159.
- 541 Paoli, A. *et al.* 2018. Winter and spring climatic conditions influence timing and synchrony
542 of calving in reindeer. - PLoS One 13: e0195603.
- 543 Paré, P. *et al.* 1996. Seasonal reproduction of captive Himalayan tahrs (*Hemitragus*
544 *jemlahicus*) in relation to latitude. - J. Mammal. 77: 826–832.
- 545 Parmesan, C. 2007. Influences of species, latitudes and methodologies on estimates of
546 phenological response to global warming. - Glob. Chang. Biol. 13: 1860–1872.
- 547 Plard, F. *et al.* 2015. The influence of birth date via body mass on individual fitness in a long-
548 lived mammal. - Ecology 96: 1516–1528.
- 549 Renaud, L.-A. *et al.* 2019. Phenotypic plasticity in bighorn sheep reproductive phenology:
550 from individual to population. - Behav. Ecol. Sociobiol. 73: 50.
- 551 Rubenstein, D. R. and Wikelski, M. 2003. Seasonal changes in food quality: a proximate cue
552 for reproductive timing in marine iguanas. - Ecology 84: 3013–3023.
- 553 Rutberg, A. T. 1987. Adaptive hypotheses of birth synchrony in ruminants: an interspecific
554 test. - Am. Nat. 130: 692–710.
- 555 Ryan, S. J. *et al.* 2007. Ecological cues, gestation length, and birth timing in African buffalo
556 (*Syncerus caffer*). - Behav. Ecol. 18: 635–644.

- 557 Sarkar, U. K. *et al.* 2019. Climato-environmental influence on breeding phenology of native
558 catfishes in River Ganga and modeling species response to climatic variability for their
559 conservation. - *Int. J. Biometeorol.* 63: 991–1004.
- 560 Sinclair, A. R. E. *et al.* 2000. What determines phenology and synchrony of ungulate
561 breeding in Serengeti? - *Ecology* 81: 2100–2111.
- 562 Skinner, J. D. *et al.* 2002. Inherent seasonality in the breeding seasons of African mammals:
563 evidence from captive breeding. - *Trans. R. Soc. South Africa* 57: 25–34.
- 564 Stearns, S. C. 1989. Trade-offs in life-history evolution. - *Funct. Ecol.* 3: 259–268.
- 565 van den Hoff, J. 2020. Environmental constraints on the breeding phenology of Giant Petrels
566 *Macronectes* spp., with emphasis on Southern Giant Petrels *M. giganteus*. - *Mar. Ornithol.*
567 48: 33–40.
- 568 Visser, M. E. *et al.* 2010. Phenology, seasonal timing and circannual rhythms: towards a
569 unified framework. - *Philos. Trans. R. Soc. B Biol. Sci.* 365: 3113–3127.
- 570 Zerbe, P. *et al.* 2012. Reproductive seasonality in captive wild ruminants: implications for
571 biogeographical adaptation, photoperiodic control, and life history. - *Biol. Rev.* 87: 965–
572 990.