

1 Original research

2

3 Site-scale variability of calling activity of insectivorous bats:
4 implications for passive acoustic monitoring

5

6 Alexander Bruckner¹, Lisa Maria Greis²

7

8

9 ¹ Institute of Zoology, Department of Integrative Biology and Biodiversity Research, University of
10 Natural Resources and Life Sciences, Gregor-Mendel-Strasse 33, A-1180 Vienna, Austria. ORCID:
11 <https://orcid.org/0000-0002-2453-4352>

12 ² Sternwartestrasse 21/8, A-1180 Vienna, Austria. ORCID: <https://orcid.org/0000-0003-4003-4259>

13

14

15 Correspondence: Alexander Bruckner, Institute of Zoology, Department of Integrative Biology and
16 Biodiversity Research, University of Natural Resources and Life Sciences, Gregor-Mendel-Strasse
17 33, A-1180 Vienna, Austria. Tel.: +43 1 47654 83 333, alexander.bruckner@boku.ac.at

18

19

20 8,635 words: 321 Abstract, 6 Keywords

21 10 figures

22 2 tables

23

Abstract

Detectors for the passive acoustic monitoring (PAM) of bats have become invaluable research tools, especially for surveys, monitoring programs and environmental impact assessments. However, little is known about the small-scale (within-site) variability of PAM recordings and especially about the influence of detector identity and distance, and of microphone orientation on the statistical confidence of activity estimates and species detection probabilities. We recorded vocalizations in a homogeneous meadow with no trees, bushes or tall ground vegetation. Eight detector pairs were arranged in an octagon, the two detectors of a pair facing in opposite directions. The call sequences of eight species were analyzed. The deviations of individual detectors from the overall mean were generally small, but large outliers occurred both at the file (temporal resolution: five seconds) and the night (resolution: one night) scale. All devices detected the main temporal patterns of calling activity in the study period, but three devices deviated systematically from the others and the sensitivity of two devices deteriorated over time. Detector orientation and distance were significant, yet small, sources of variability. The probability of detecting the presence of species correlated with species' activity and ranged on average from 100 % for bats in total to only 18.8 % for the least active *Myotis myotis*. The sample sizes necessary to achieve 90 % statistical confidence of activity estimates ranged from 7 to 16 detectors and from 5 to 12 nights, depending on taxon. Increasing the number of nights resulted in much higher confidence than increasing the number of detectors. We recommend PAM studies of bats to frequently calibrate detector sensitivity; deploying at least three detectors per study site; sampling longer periods instead of deploying more detectors; randomly assigning and swapping detectors among sites, treatments, strata, etc.; and statistically scrutinizing the sample data, especially for outliers.

Keywords: Chiroptera, survey design, sampling optimization, detection probability, bat detector, microphone

50 1 Introduction

51 Passive acoustic monitoring (PAM, Stowell & Sueur 2020) units autonomously record and store
52 animal sounds to a tape or digital file. They offer a convenient way to measure animal activity in
53 the field: using these units is often more time and cost efficient than alternative assessment methods
54 because, once assembled, operating personnel is not required for their performance. When deployed
55 for a prolonged period, the devices can sample vast amounts of vocalizations, and thus have a high
56 probability of detecting the presence of a study taxon. PAMs are used for a broad variety of faunal
57 groups, notably birds (e.g. Darras et al. 2019, Frommolt 2017), anurans (e.g. Duarte et al. 2019,
58 Madalozzo et al. 2017), even invertebrates (e.g. Görres & Chesmore 2019, Mankin et al. 2011), and
59 most often for bats (Sugai et al. 2019).

60 Beginning with low-tech applications in the mid-1990s (e.g. Krusic et al. 1996), PAM quickly rose
61 to popularity among batworkers to detect, quantify the activity and calculate the density of their
62 elusive objects (Milchram et al. 2019). PAM is especially valuable for remote areas and large-scale
63 studies where mist-netting and telemetry are unable to reach robust conclusions over large spatial
64 scales (Kaiser & O'Keefe 2015). Thus, acoustic detectors have today become the "new normal"
65 (Nocera et al. 2019) in bat monitoring programs and related applications, not least because modern
66 software packages now also help to (semi-) automatically process and analyze the often quickly
67 amassed volume of sound files (see Fraser et al. 2020 for a current overview).

68 However, PAM detectors sample only a small fraction of the acoustical activity in their
69 surroundings (Adams et al. 2012, Britzke et al. 2013). A growing body of literature provides
70 empirical evidence that the size of this fraction depends on the interplay between the technical
71 specifications of the equipment, the modalities of detector field deployment, and the call
72 characteristics of the bat community under study.

73 For example, the probability of recording echolocation calls of specific frequencies may vary
74 significantly between detector brands (Adams et al. 2012, Kaiser & O'Keefe 2015), and even
75 between detectors of the same brand and type, if not properly calibrated (Larson and Hayes 2000).

76 Kaiser and O'Keefe (2015), O'Keefe et al. (2014), and Weller and Zabel (2002) indicated that
77 detector position relative to vegetation clutter or to flyways may affect the quantity and probability
78 of bat species detection. Kubista & Bruckner (2017) found less recording variability among three
79 detectors simultaneously deployed in sites with no woody vegetation than in sites with dense
80 vegetation structure. Height of detectors above ground and orientation relative to the horizontal axis
81 are further often reported influencing factors on the number of recorded call sequences and species
82 (e.g. Britzke et al. 2010, Menzel et al. 2005, Staton & Poulton 2012, Weller and Zabel 2002).
83 Altogether, many intricacies of detector deployment are critical for obtaining unbiased estimates of
84 bat calling activity at monitoring sites, and yielding rigorous data from PAM campaigns is much
85 less straightforward than it may at first seem.

86 A different, but closely related, question concerns sample size: how many PAM units, sites, or
87 nights are needed to adequately sample the bat population or community under study? Here, PAM
88 needs to balance the trade-off between high replicability (many sites investigated) and reliability
89 (sites, and potentially sub-plots and strata, are investigated intensely to get a detailed picture of the
90 resident fauna, Fischer et al. 2009, Froidevaux et al. 2014, Law et al. 2015). Since high-quality bat
91 detectors are still expensive and the number of available devices is limited in many projects, we
92 suspect that batworkers may sometimes be tempted to sacrifice reliability for replicability,
93 especially in environmental impact assessments and other practical applications.

94 In the research presented here, we ignored between-site variability to instead focus on the sources of
95 within-site variability of bat call activity. With an intentionally simple experimental setting, we
96 aimed at providing a baseline for PAM studies operating in more complex situations and
97 implementing more complex designs. We excluded potential effects of detector brand, vegetation
98 clutter, and deployment height and angle, which we feel are well enough covered in literature.
99 Rather, unshielded detectors of the same manufacturer and type were set in a spatially
100 homogeneous hay meadow with no trees, bushes or tall ground vegetation. No sub-plots or strata
101 were designated within the site. We placed eight pairs of devices in an octagon of 13 m diameter,

with the devices of each pair facing in opposite directions (inwards and outwards of the octagon). This setup allowed us to assess the effects of (i) varying sensitivity among detectors, (ii) the horizontal orientation (direction) of microphones, and (iii) the spatial placement of detectors (= the distance between each two detectors), and (iv) temporal patterns on the variability of activity recordings and on the detection probability of species. Finally, we assessed the influence of sample size (number of nights sampled and number of detectors, respectively) on the statistical confidence of calling activity estimates at the site scale.

109

110

111 2 Materials and Methods

112 2.1 Study Site and Experimental Setup

The experiment was conducted in an enclosure fenced off within a 6 ha meadow situated in the deer compound of the Wildlife Ecology Research Institute of the University of Veterinary Medicine, Vienna, Austria (16°16' E, 48°13' N). The meadow was surrounded by mature oak (*Quercus spp.*) and European beech (*Fagus sylvatica*) forest.

We deployed ultrasound detectors on 2.5 m high tent poles in a regular octagon. The vertices of the octagon corresponded with the cardinal directions of the compass rose and lay 6.5 m from the center and at a minimum distance of 4.96 m from each other (Fig. 1). During the recording period, the poles were not moved, thus excluding any data variability due to translocation.

Each pole carried a detector pair that consisted of two devices attached together back to back. One detector of each pair faced towards the octagon's center, the other in the opposite (outward) direction. The detector positions were randomly switched each night, making sure each device held each position only once. To prevent the detector pairs from swinging on the pole and thus changing direction, they were secured with wire. The detectors were not inclined relative to the horizontal, and covered with artificial fur to avoid ultrasound reflections.

127

128 2.2 Detector Specifications and Settings

129 We used 16 automated devices (batcorders, ecoObs GmbH, Nürnberg, Germany,
 130 <http://www.ecoobs.de>) of two different generations (1, 2.0) for recording and storing bat call
 131 sequences. The results derived from the different generations are supposed not to differ, since they
 132 use identical microphones and were checked and calibrated by the manufacturers prior to the study.
 133 Recordings were made using the default settings of the 'Auto+Timer' function (-27dB threshold
 134 level, 400 ms post trigger, real time recordings at 500kHz and 16bit; see Adams et al. 2012 for a
 135 comparison of batcorder specifications and performance with other popular devices).
 136 The devices were deployed during 16 nights with a good weather forecast (minimum night
 137 temperature $> 12^{\circ}\text{C}$, no rain, wind ≤ 4 Beaufort) between May and July, 2012. All recordings were
 138 started at 6 pm and ended at 7 am the following day to include possible activity before sunset and
 139 after sunrise. To ensure comparability among the recorded vocalizations, all detectors were time
 140 synced every recording night to the nearest second.

141

142 2.3 Call identification and selection of taxa for analysis

143 The recorded files were transferred to a computer and analyzed with bcAdmin 2.18 and BatIdent
 144 1.03 (ecoObs, Nürnberg, Germany). Supposed call sequences of zero or excessive length (ratio
 145 sequence length [s] / number of calls < 0.5) were removed after manual inspection because they did
 146 not contain any or only orthopteran vocalizations.

147 Species-specific analyses were done using verified species only. A species was regarded verified
 148 after manually validating all its call sequences (see Fritsch & Bruckner 2014 for details). The data
 149 of each night and detector were inspected blindly and in random order to ensure that the operator's
 150 decisions were unbiased with respect to these factors. The validations were performed by the
 151 second author.

152 Due to the plasticity of orientation calls, the acoustical identification of many bat species is
 153 questionable and has been profoundly criticized (Barclay 1999, Ratcliffe and Jakobson 2018, Russo

154 & Voigt 2016, Rydell et al. 2017). We therefore restricted the data analysis to a set of phonic
 155 groups (operational taxonomic units = OTUs) the calls of which could be identified with high
 156 confidence. The very similar sequences of *Pipistrellus kuhlii* (Kuhl, 1817) and *P. nathusii*
 157 (Keyserling & Blasius, 1839) were pooled and labelled "*P. kuhlii/nathusii*". For the same reason,
 158 the sequences of *Eptesicus nilssonii* (Keyserling & Blasius, 1839), *E. serotinus* (Schreber, 1774),
 159 *Nyctalus leisleri* (Kuhl, 1817), *N. noctula* (Schreber, 1774), and *Vespertilio murinus* Linnaeus, 1758
 160 were pooled to form the "nyctaloid" OTU.

161

162 2.4 Statistical Analyses

163 We used the length of the call sequences (= length of passes, in seconds) as measure of the
 164 recording performance of bat detectors. To cater for any inaccuracies of the time syncing of the
 165 detectors, the sequences were summed in 5-seconds intervals. Because of incorrect labelling of
 166 recordings in the field, we had to delete one night from the data set, hence the data analysis
 167 comprised 15 recording nights (15 nights x 16 detectors = 240 detector nights in total).
 168 If applicable, data variability was expressed on a fine and a coarse temporal scale: the fine scale had
 169 a minimum resolution of seconds (= the individual files of the recordings, summed in 5-seconds
 170 intervals, dubbed "file scale" in the following) and aimed at investigating the sources of data
 171 variability. The coarse scale was that of sampling nights (dubbed "night scale") and aimed at
 172 providing recommendations for the practice of bat work.

173

174 We conducted the following analyses:

175 (i) To characterize the influence of detector sensitivity on data variability, we computed Δ call
 176 sequence length as the absolute difference in call sequence length between every detector and the
 177 arithmetic mean of all 16 simultaneously recording detectors. We plotted frequency histograms of
 178 this measure, pooling the data of all sampling nights, and expected the shape of the histograms to
 179 correlate with the accuracy of the detectors: wide histograms for large recording differences

180 between individual devices and their average, and narrow ones for small differences. Bin width of
 181 these and the following histograms was manually set to best display the data, using the Sturges and
 182 Freedman-Diaconis rules as a guide (procedures `nclass.Sturges()` and `nclass.FD()` in base R).

183 (ii) To assess the effects of detector orientation (that is, microphones pointing in- or outwards of
 184 the octagon), Δ call sequence length was calculated as the absolute difference in call sequence
 185 length between two detectors deployed in the same position of the octagon but oriented in opposite
 186 directions, and histograms were produced from the results. Prior to the analysis, double-zero
 187 pairings (that is, neither of the detectors of a position pair had recorded a file) were removed from
 188 the data set so as not to confound them with pairs recording files of the same length.

189 (iii) To investigate the influence of the spatial detector distance on data variability at the file scale,
 190 we computed Δ call sequence length as the absolute difference in call sequence length between each
 191 two detectors deployed at the eight positions of the octagon. Distances varied between 4.96 and
 192 13.00 m, and the data of the distances 9.18 and 9.19 m, and 12.98 and 13.00 m, respectively, were
 193 pooled. The calculations were restricted to inwardly directed detectors ($n = 8$) to reduce any effects
 194 of microphone orientation on the results. Double-zero pairings were removed from the data set prior
 195 to the calculations.

196 (iv) To test the influences of the experimental factors, we computed linear mixed effects models
 197 (LMM, gaussian family) (a) for the factors detector and orientation, and (b) for spatial distance. In
 198 standard R (Wilkinson-Rogers) notation, the detector/orientation models had the form:

199

$$\text{length sequence} \sim \text{detector} + \text{orientation} + (1|\text{time interval}) + (1|\text{position/orientation}) \quad (1)$$

200

201 thus in their random parts accounting for the longitudinal structure of the study and for the nested
 202 position of orientation within position (= the octagon vertices). Inevitably, the models had to ignore
 203 potential distance effects (but these were anyway small, see below). In addition, the total, nyctaloid
 204 and *P. kuhlii/nathusii* models each included an overdispersion and zero inflation term (dispformula

205 = detector + orientation, $ziformula = \sim.$), while the models for *P. pipistrellus* and *H. savii* included a
 206 zero inflation term only, and the model for *B. barbastellus* included no additional terms. We failed
 207 to find an appropriate model for the detector/orientation effects of the *M. myotis* data, presumably
 208 because of the low number of records for this species. Likewise, models testing potential
 209 deterioration of microphone sensitivity over time (by including detector:night as a fixed effect) did
 210 not converge, hence we had to examine this aspect with graphical methods only.

211 The detector/orientation models had the form:

212

$$\Delta \text{ sequence length} \sim \text{distance} + (1|\text{time interval}) + (1|\text{detector 1}) + (1|\text{detector 2}) \quad (2)$$

213 where "detector" referred to the two devices each used for calculating Δ sequence length.

214 The models were fitted using function `glmmTMB()` in package `glmmTMB` 1.0.1 (Brooks et al.

215 2017). During model development, we selected between alternative models by running likelihood

216 ratio tests (function `anova()` in base R) and by comparing AIC (Akaike information criterion) and

217 deviance values. Model residuals and homoscedasticity of factor levels were inspected using the

218 diagnostic tools of package `DHARMa` 0.3.2.0 (Hartig 2020), and were found to be adequate. We

219 tested for the significance of fixed effects (the experimental factors) using likelihood ratio tests

220 between null (comprise random part only) and final (comprise random and fixed part) models.

221 The marginal importance of factor distance was calculated using `r2_nakagawa()` in package

222 `performance` 0.4.5 (Lüdtke et al. 2020, Nakagawa et al. 2017). This was not possible for the

223 detector and orientation models because, to the best of our knowledge, there is as yet no accepted

224 way to calculate the marginal importance of `glmmTMB` models including zero inflation and

225 overdispersion terms. As a workaround, we calculated the difference between the conditional

226 coefficients of determination R^2 of the final and the null models using `r2_xu()` in package

227 `performance` 0.4.5 (Lüdtke et al. 2020, Xu 2003).

(v) For all nights with at least one call sequence registered by any of the devices, we calculated the fraction of devices with a record and used this as a proxy for the detection probability of the OTUs under investigation.

(vi) To characterize the influence of sample size (number of nights sampled or number of detectors deployed in parallel) on the statistical confidence of calling activity estimates, we bootstrapped (resampled with replacement) the file scale data 9,999 times with increasing number of nights (1 to 15) and of devices (1 to 16). Thus, there were 9,999 replicates of each sample size for both night and device. Assuming that researchers would usually average the activity recorded in multiple nights or with multiple devices, we calculated the median activity of each bootstrapped sample size, and used the 95 % interquantil range (the range between the 5 % and 95 % quantil) around the medians as a measure of confidence. Interquantil ranges are distribution-free and were selected because we did not find a frequency distribution which adequately fitted the bootstrapped data of all sample sizes and OTUs. As a breakpoint from which an increase in sample size would have only marginally increased statistical confidence, we arbitrarily set 90 % of the best available activity estimate - that is, the narrowest interquantil range.

All analyses were computed and figures produced in R 3.6.2 (R Core Team 2019) under Mac OS X Catalina 10.15.5.

245

246

247 3 Results

248 3.1 Data overview

15,502 s of vocalizations were recorded during this study, 12,859 s of which were from bats. The software automatically identified 21 bat species and genera, respectively. From these, we selected 8 operational taxonomic units (OTUs) for further analysis: total (= sum of all bat vocalizations; 12,859 s of calls), nyctaloid (an amalgamation of the software output for *Eptesicus nilssonii*, *E. serotinus*, *Nyctalus leisleri*, *N. noctula*, and *Vespertilio murinus*; 7,432 s), *Pipistrellus pygmaeus*

(1,022 s), *P. kuhlii/nathusii* (736 s), *P. pipistrellus* (434 s), *Hypsugo savii* (405 s), *Barbastella*
barbastellus (85 s), and *Myotis myotis* (35 s). Thus, the nyctaloids alone accounted for 57.8 % of
total call sequence length.

Activity differences between the selected OTUs and between nights contributed most to the
variability in the data set. The frequency of occurrence of OTUs during the study period differed
widely and ranged from highly common (total: 100 %, nyctaloids: 90.8 % of the detector nights
with recordings) to rare (*B. barbastellus*: 17.9 %, *M. myotis*: 6.3 %) (Figure 4). Total nightly
activity varied considerably, with an approximately 18-fold difference between the least and the
most active night (nights 14 and 5, respectively). However, this pattern was dominated by the
nyctaloid species, and the other OTUs exhibited different peaks of activity in the study period
(Figure 4).

265

3.2 Variability among detectors - overall patterns (file and night scale)

On the file scale, the deviations of individual detectors from the overall means of all 16 detectors
were generally small. The frequency distributions of absolute Δ sequence length were strongly
skewed to the right - that is, large deviations occurred, but only rarely (Figure 2). On average, 91.6
% of the detectors did not deviate ($\Delta = 0$ s) or deviated less than one second ($\Delta < 1$ s) from the
overall means (range: 84.4 % *M. myotis* to 97.0 % *B. barbastellus*). The 95 % data quantile was
between one and two seconds and the 99 % quantile close to two seconds for most OTUs; however,
outlier Δ values reached approximately three to six seconds (Figure 2).

A very similar picture emerged at the night scale in that detectors generally deviated little from the
overall means. The frequency distributions of Δ absolutes were again heavily right-skewed and
long-tailed for all OTUs, with outlying values far at their extreme ends (Figure 3).

To better understand the contribution of individual detectors to the tails of the frequency
distributions, we plotted, at the file scale and for all detectors and OTUs, the absolutes of Δ
exceeding the 95 % data quantile. A few devices stood out here, for example detector 83 which

280 generated many extreme values in the total and nyctaloid data, and detector 40 which contributed
 281 conspicuously little to the tails of all other OTUs (Figure 5). But apart from these exceptions, the
 282 outliers in the long tails of the frequency distributions could not be attributed clearly to individual
 283 detectors.

284

285 3.3 Variability among detectors - temporal patterns (night scale)

286 All detectors performed comparably well in tracing the temporal peaks of calling activity on the
 287 night scale (e.g. the peak of total and nyctaloid activity in night 5, and that of *H. savii* in night 1;
 288 Figure 4A and F). There were, however, evident cases of high within-night variability among
 289 devices, for example night 10 in *P. pygmaeus* and night 16 in *P. kuhlii/nathusii* (Figure 4C and D).
 290 At least three devices deviated systematically from the nightly overall means: detectors 25 and
 291 especially 40 performed worse than average throughout the study and for all OTUs, and their
 292 sensitivity seemed to deteriorate, especially at the end of deployment. Detector 83, in contrast,
 293 recorded consistently above average (Figure 4).

294

295 3.4 Variability due to detector orientation (file scale)

296 On the file scale the majority of position-pair differences were small, and, averaged over all OTUs,
 297 73.0 % of the Δ sequence length values were below 1 s (range: 35.7 % *M. myotis* to 75.1 %
 298 nyctaloids). The frequency distributions of Δ exhibited a strong right skew for all OTUs, with the
 299 95 % quantiles mostly falling between two and three, and the 99 % quantiles between two and four
 300 seconds (Figure 6).

301

302 3.5 Variability due to detector distance (file scale)

303 The medians of Δ sequence length of all OTUs became slightly larger with increasing spatial
 304 distance, while the shape of the frequency distributions of Δ remained almost unchanged (Figure 7).

305

3.6 Mixed effects models (file scale)

Detectors were a significant source of call sequence variability in the models of all OTUs except of *H. savii*. Models containing both detector and orientation as fixed effects were significantly different from detector-only models for *P. pygmaeus*, but not for the other OTUs (Table 1). Irrespective of statistical significances, including detector and/or orientation added no more than 1.2 to 5.2 % to the coefficient of determination, as compared to the null models. These minor improvements were also reflected in small differences of AIC and deviance between null, detector-only and detector/orientation models (Table 1).

Distance was a significantly influential factor on Δ sequence length between neighboring detectors of all OTUs; however, the marginal coefficients of determination were even smaller than those of detector/orientation (range 0.2 to 2.6 % of total variability, Table #2).

317

3.7 Detection probability of OTUs (night scale)

At the night scale, the probability of detecting the presence of bats varied widely among OTUs. It was 100 % for total bat activity, that is, if any one detector made a record during a study night, all other detectors also registered at least one call sequence. The probability was lower for all other OTUs: It dropped to an average (median) of 18.8 % for *M. myotis* and several minima were as low as 6.3 - 12.5 % (Figure 8). Across all nights and OTUs, the detection probability correlated monotonously with the sums of nightly activity (Kendall's $\tau = 0.79$), that is, more actively calling OTUs had a greater chance of being detected by any of the devices than rarer OTUs.

326

3.8 Sample size and confidence of activity estimates (night scale)

As a measure of data confidence, the 95 % interquantil ranges of bootstrapped call activity data decreased exponentially with increasing number of detectors for all OTUs. The same was true for the number of nights. Since the ends of all curves paralleled the abscissae, further increasing the sample sizes beyond the limits of this experiment (that is, more than 16 detectors / 15 nights) would

332 not have resulted in higher statistical confidence (Figures 9, 10). The sample sizes necessary to
333 achieve 90 % confidence varied widely among OTUs and approx. ranged from 7 - 16 detectors and
334 5 - 12 nights. Increasing the number of nights resulted in much higher data confidence (narrower
335 interquantil ranges) than increasing the number of detectors (compare the scales of the ordinates of
336 Figures 9 and 10).

337

338

339 4 Discussion

340 In this study, we found only small sensitivity differences among the 16 detectors we ran in parallel;
341 both at the file (seconds) and at the night scale, the lengths of recorded sequences showed no, or
342 only slight, deviation. We did demonstrate a significant overall influence of detectors on data
343 variability for most OTUs, however, we attribute these significances to the large number of
344 observations (mostly 10^3 to 10^4 , depending on model and OTU) and the high power of well-
345 constructed mixed effects models (Bolger 2015). Thus, the tests were likely to find even small
346 contributions of the detectors to the overall variability in our data sets, and identify them as
347 significant.

348 Whether researchers can cope with similarly small differences between detectors in future
349 investigations will depend on their study aims and data. For example, little additional variability
350 (the marginal R^2 we found for "detector") may be acceptable for studies of environmental impacts
351 of highways or wind parks, where the presence of particular species is more important for the
352 assessment than their activity. Conversely, even little additional variability due to detector
353 differences may contribute to missing the level of significance in experiments on the vocational
354 biology of species or, in monitoring programs, to failing to detect temporal trends of endangered
355 populations.

356 We were unfortunately not able to quantitatively compare the marginal importance of the factor
357 "detector" to that of "species" and "night" because tentative mixed effect models including all

358 experimental factors were overparameterized and failed to converge. However, inspecting the
359 variability among species and among nights, we recognized the latter two factors as far more
360 important for total data variability than the detectors. They are therefore also more urgent to
361 consider when designing monitoring studies and when analyzing activity data (Hayes 1997, Skalak
362 et al. 2012).

363 In contrast to the generally small among-detector variability, single call sequences were overly long
364 and produced extremely high values in our data set. These far outlying call sequences occurred in
365 all investigated OTUs (except in the sparse data of the two very rare species) and reached maxima
366 of 5.6 s (*P. kuhlii/nathusii*) and 6.7 s (total activity). We have also seen extremes of this magnitude
367 in other studies we have conducted and suppose they are a common phenomenon in PAM data.
368 Unfortunately, outlying values were not only apparent at the file, but also on the night scale. They
369 were therefore not a result of just a few and random events of particularly high calling activity.
370 They did not "average out" when, as in most studies, nightly sums were calculated before data
371 analysis. Although we could not cleanly attribute the extreme values to individual detectors because
372 they occurred in the data of all devices, there was at least a tendency of several detectors to
373 accumulate outlying values that then also showed up at the night scale.

374 Whatever the biological reasons for outliers in PAM data, they can introduce serious problems to
375 the analysis because if not properly accounted for, they may strongly bias results and mislead
376 conclusions. We therefore strongly recommend routinely including a thorough exploratory data
377 analysis step in PAM studies, and especially inspecting outlier plots and frequency distributions of
378 the data. If extreme values are found, the responsible call sequences need to be reviewed for
379 correctness (orthopteran calls? two sequences wrongly added together?) and possibly removed. If
380 found correct, the data sets should to be summarized using robust (distribution-free) statistics and/or
381 characterized with models that properly account for overdispersion (e.g. Bolker 2015, Brooks et al.
382 2017).

383 Similar to the differences among detectors, we detected only minor variability within most position
384 pairs, and orientation was a significant factor in the model of only one of the investigated OTUs.
385 This was expected since batcorder microphones are small and, as directionality increases with
386 microphone size (Ratcliffe & Jakobson 2018), are considered omnidirectional (Adams et al. 2012).
387 Despite this, a number of extremely high Δ values of several seconds magnitude emerged in our
388 data, resulting from occasionally big differences between paired devices.
389 In a study conducted in a forested area and using directional microphones, Weller & Zabel (2002)
390 found great variation of bat detections due to microphone orientation relative to vegetation clutter.
391 They concluded that standardizing orientation is a necessary requirement to make meaningful
392 comparisons between sites possible. We agree in principle, but add that the orientation of
393 omnidirectional microphones in open field settings without vegetation clutter may safely be
394 ignored. We have not tested this, but assume that the same holds true for the inclination (angle to
395 horizontal) of microphones, as long as the detectors are sufficiently raised from the ground.
396 With few exceptions, the temporal patterns of calling activity in the study period were more or less
397 well traced by all 16 detectors. As above, we regard this adequate for monitoring situations where
398 minor differences among devices do not have much impact for study outcomes.
399 A more disturbing finding was that the sensitivity of two detectors seemed to decrease over time.
400 This is especially remarkable since the devices were checked and calibrated by the manufacturer
401 prior to exposure, were deployed for only a short time (16 nights in a 10-week period) and solely
402 under conditions of fair weather. Thus, microphone damage due to precipitation, condensation or
403 freezing could be excluded. No defects were apparent in the field, and we only recognized the
404 malfunctions during data analysis. Likewise, detector and microphone age did not offer a plausible
405 explanation, as both batcorders belonged to the newest generation used in this study (batcorder 2.0).
406 These findings corroborate Fischer et al. (2009) and Larson & Hayes (2000), who stressed the
407 importance of calibration when using multiple detectors in parallel. Before this experiment, we
408 were ignorant of the fact that microphone deterioration may occur so suddenly and without obvious

reason. In all preceding studies, we were confident in having our microphones inspected and calibrated once a year. But apparently, batworkers are well advised to additionally, and repeatedly, check devices themselves during the field season so as not to miss any signs of declining sensitivity. Calibration once a year (as is sometimes recommended, e.g. Loeb et al. 2015) is evidently not enough. We do not advise against having detectors regularly checked by the manufacturers, but it is not advisable to trust that one service can equilibrate the recording performance of individual devices for a full field season. Standardized testing and calibrating equipment is available for this purpose, but unfortunately only for a segment of the detectors currently on the market.

At least for the spatial distances among detectors in this study (5 to 13 meters), we found significant, but only unsubstantial effects. Although small microphones generally have limited sensitivity (Ratcliffe & Jakobson 2018) and the performance of batcorders is poor in comparison to that of other brands (Adams et al. 2012), the distances here were not sufficient to considerably contribute to data variability. For studies in more heterogeneous settings however, even short-scale distance may have an effect, and implementing spatial stratification can be a key factor to optimize sampling (Fischer et al. 2009, Meyer et al. 2010), for example by deploying detectors at several locations within sites.

Except for the very abundant OTUs (total, nyctaloids, and *P. pygmaeus*), the average probability of detection in this study was only 50 % or lower - that is, at least half or more of the detectors failed to register the presence of a calling OTU. For similar monitoring situations and using a single detector per site, we therefore expect the odds of missing an OTU calling less than ~ 45 s per night (the nightly average of *P. kuhlii/nathusii*) to be larger than 50 % - much higher than expected for our structurally simple experimental site. Similarly low detection rates have been found in other studies, but at much larger spatial scales and in more cluttered situations (Duchamp et al. 2006, Kaiser and O'Keefe 2015, Skalak et al 2012).

In contrast to the variability of recorded calling activity due to detector, orientation and distance effects, which we consider acceptably small for monitoring purposes, this variability in detection

435 probability offers very poor prospects, indeed. The often large volumes of recordings returned by
436 PAM campaigns may give batworkers the deceptive feeling of having produced exhaustive species
437 inventories with little effort – yet this is obviously not true, not even at the very small spatial scale
438 of this study.

439 An important result of this study is that sampling more nights at an investigation site produces more
440 confident estimates of calling activity than sampling with more detectors in parallel. This supports
441 Fischer et al. (2009) who, generalizing across several species, estimated that 10 % of overall
442 variability of activity occurred due to within-site differences, in contrast to 20 % that they attributed
443 to between-night differences. Although this study is not fully comparable to ours because of
444 divergent vegetation structure and spatial scale, it also suggests emphasizing temporal over spatial
445 replication to characterize the calling activity of bats at investigations sites. For monitoring
446 campaigns, this is an encouraging result because usually detectors, and not so much nights, are in
447 short supply.

448 Given the wide range of activity among the OTUs in this study, we expected that actively calling
449 OTUs should be more confidently estimated than rarer species of which only few call sequences
450 were recorded. In other words, actively calling OTUs would need fewer detectors / sampling nights
451 to level out to a confident estimate. Surprisingly, we did not find this in the data, nor was there any
452 other apparent relationship with biological characteristics of the OTUs (e.g. maximum frequency,
453 bandwidth, sound pressure level of calls, Ratcliffe & Jakobsen 2018; compare to Kubista &
454 Bruckner 2017). We have no clear explanation to offer for this observation, only that the number of
455 OTUs in this study was probably too small to find such expected relationships, and that some OTUs
456 (total and nyctaloid activity, possibly also *P. kuhlii/nathusii*) were not biological entities, but
457 compounds of more than one species.

458

459

460 5 Recommendations

- 461 (1) Check your PAM detectors frequently, adjust the sensitivity and replace damaged microphones.
462 Do this more than once or twice per field season since microphones may decline suddenly. At least
463 a rough-and-dirty check with a chirp means little effort.
- 464 (2) If you aim to characterize the bat assemblages of sites (for example for detecting changes over
465 time due to anthropogenic impacts), prefer sampling longer periods over sampling sites with more
466 detectors. Given our finding that microphone sensitivity may decrease without any visual signs of
467 damage, the logical minimum is three detectors per site - this enables you to differentiate damaged
468 from properly working detector data. This recommendation also applies to studies stratifying sites
469 into subplots - also here, each subplot should be equipped with a minimum of three detectors.
- 470 (3) Throughout the field program, regularly and randomly change the assignment of detector
471 individuals to sites, treatments, study strata, etc. This prevents confusing potential microphone bias
472 with activity declines in the experimental units of the program.
- 473 (4) Check your data after the field period. Especially look for outliers and re-inspect the raw data of
474 suspicious values. Use robust statistics to summarize the data, and models that account for
475 overdispersion.

476

477

478 Acknowledgements

479 We thank Claudia Bieber and the staff of the Research Institute of Wildlife Ecology for permitting
480 access to the study site, Claudia Kubista, Georg Fritsch, Barry Jones and Birgit Szabo for invaluable
481 help and support throughout the study, and Marcia Sittentaler and Lea Mirwald for assistance in the
482 field. Benjamin Seaman helped improve the English.

483

484 Conflict of interest

485 None declared.

486

487 Author contributions

488 Alexander Bruckner: Conceptualization (lead), formal analysis, writing original draft (lead), review
489 and editing.

490 Lisa Maria Greis: Conceptualization (supporting), field sampling, data organization, writing
491 original draft (supporting).

492

493 Data availability statement

494 All data are available in the Dryad online repository under [https:// ...](https://...) [data will be deposited upon
495 manuscript acceptance]

496

497 References

498 Adams, A.M., Jantzen, M.K., Hamilton, R.M. & Fenton, M.B. (2012). Do you hear what I hear?
499 Implications of detector selection for acoustic monitoring of bats. *Methods in Ecology and*
500 *Evolution*, 3, 992–998.

501 Barclay, R.M.R. (1999). Bats are not birds - a cautionary note on using echolocation calls to
502 identify bats: A comment. *Journal of Mammalogy*, 80, 290-296.

503 Bolger B.M. (2015). Linear and generalized linear mixed models. In G.A. Fox, S. Negrete-
504 Yankelevich & V.J. Sosa (Eds), *Ecological statistics: Contemporary theory and application* (pp.
505 309-333). Oxford, UK: Oxford University Press.

506 Britzke, E.R., Slack, B.A., Armstrong, M.P. & Loeb, S.C. (2010). Effects of orientation and
507 weatherproofing on the detection of bat echolocation calls. *Journal of Fish and Wildlife*
508 *Management*, 1, 136–141.

509 Britzke, E.R., Gillam, E.H. & Murray K.L. (2013). Current state of understanding of ultrasonic
510 detectors for the study of bat ecology. *Acta Theriologica*, 58, 109-117.

- 511 Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A.,
 512 Skaug, H. J., Mächler, M. & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among
 513 packages for zero-inflated generalized linear mixed modeling. *R Journal*, 9, 378–400.
- 514 Darras, K., Batáry, P., Furnas, B.J., Grass, I., Mulyani, Y.A. & Tschardtke, T. (2019). Autonomous
 515 sound recording outperforms human observation for sampling birds: a systematic map and user
 516 guide. *Ecological Applications*, 29, 1247-1265.
- 517 Duarte, M.H.L., Caliari, E.P., Viana, Y.P. & Nascimento, L.B. (2019). A natural orchestra: how are
 518 anuran choruses formed in artificial ponds in southeast Brazil? *Amphibia-Reptilia* 40, 373-382.
- 519 Duchamp, J.E.R., Yates, M., Muzika, R.-M. & Swihart, R.K. (2006). Estimating probabilities of
 520 detection for bat echolocation calls: An application of the double-observer method. *Wildlife Society*
 521 *Bulletin*, 34, 408-412.
- 522 Fischer, J., Stott, J., Law, B.S., Adams, M.D. & Forrester, R.I. (2009). Designing effective habitat
 523 studies: quantifying multiple sources of variability in bat activity. *Acta Chiropterologica*, 11, 127-
 524 137.
- 525 Fraser, E.E., Silvis, A., Brigham, R.M., Czenze, Z. (2020) *Bat echolocation research. A handbook*
 526 *for planning and conducting acoustic studies*. Second edition. Bat Conservation International,
 527 Austin, Texas, U.S.A.
- 528 Fritsch, G. & Bruckner, A. (2014). Operator bias in software-aided bat call identification. *Ecology*
 529 *and Evolution*, 4, 2703–2713.
- 530 Froidevaux, J.S.P., Zellweger, F., Bollmann & K. Obrist, M.K. (2014) Optimizing passive acoustic
 531 sampling of bats in forests. *Ecology and Evolution*, 4, 4690-4700.
- 532 Frommolt, K.-H. (2017). Information obtained from long-term acoustic recordings: applying
 533 bioacoustic techniques for monitoring wetland birds during breeding season. *Journal of Ornithology*
 534 158, 659-668.
- 535 Görres, C.-M. & Chesmore, D. (2019). Active sound production of scarab beetle larvae opens up
 536 new possibilities for species-specific pest monitoring in soil. *Scientific Reports*, 9, 10115.

- 537 Hartig, F. (2020). DHARMA: Residual Diagnostics for Hierarchical (Multi-Level / Mixed)
 538 Regression Models. R package version 0.3.2.0. <https://CRAN.R-project.org/package=DHARMA>
- 539 Hayes, J.P. (1997). Temporal variation in activity of bats and the design of echolocation-monitoring
 540 studies. *Journal of Mammalogy*, 78, 514-524.
- 541 Kaiser, Z.D.E. & O'Keefe, J.M.O. (2015). Factors affecting acoustic detection and site occupancy
 542 of Indiana bats near a known maternity colony. *Journal of Mammalogy*, 96, 344-360.
- 543 Krusic, R.A., Yamasaki, M., Neefus, C.D. & Pekins, P.J. (1996). Bat habitat use in the White
 544 Mountain National Forest. *Journal of Wildlife Management*, 60, 625-631.
- 545 Larson, D.J. & Hayes, J.P. (2000). Variability in sensitivity of Anabat II bat detectors and a method
 546 of calibration. *Acta Chiropterologica*, 2, 209-213.
- 547 Law, B., Gonsalves, L., Tap, P., Penman, T. & Chidel, M. (2015). Optimizing ultrasonic sampling
 548 effort for monitoring forest bats. *Austral Ecology*, 40, 886-897.
- 549 Loeb, S.C., Rodhouse, T.J., Ellison, L.E., Lausen, C.L., Reichard, J.D., Irvine, K.M., Ingersoll,
 550 T.E., Coleman, J.T.H., Thogmartin, W.E., Sauer, J.R., Francis, C.M., Bayless, M.L., Stanley, T.R.
 551 & Johnson, D.H. (2015). *A Plan for the North American Bat Monitoring Program (NABat)*.
 552 Department of Agriculture Forest Service, General Technical Report SRS-208, Asheville, N.C.,
 553 U.S.A.
- 554 Lüdecke, D., Makowski, D., Waggoner, P. & Patil, I. (2020). performance: Assessment of
 555 Regression Models Performance. R package version 0.4.5.
 556 <https://CRAN.R-project.org/package=performance>
- 557 Madalozzo, B., Santos, T.G., Santos, M.B., Both, C. & Cechin, S. (2017). Biodiversity assessment:
 558 selecting sampling techniques to access anuran diversity in grassland ecosystems. *Wildlife*
 559 *Research*, 44, 78-91.
- 560 Mankin, R.W., Hagstrum, D.W., Smith, M.T., Roida, A.L., Kairo, M.T. (2011). Perspective and
 561 promise: a century of insect acoustic detection and monitoring. *American Entomologist*, 57, 30-44.

- 562 Menzel, J.M., Menzel, M.A., Kilgo, J.C., Ford, W.M., Edwards, J.W. & McCracken G.F. (2005).
 563 Effect of habitat and foraging height on bat activity in the coastal plain of South Carolina. *Journal*
 564 *of Wildlife Management*, 69, 235-245.
- 565 Meyer, C.F.J., Aguiar, L.M.S., Aguirre, L.F., Baumgarten, J., Clarke, F.M., Cosson, J.-F., Estrada
 566 Villegas, S., Fahr, J., Faria, D., Furey, N., Henry, M., Hodgkison, R., Jenkins, R.K.B., Jung, K.G.,
 567 Kingston, T., Kunz, T.H., MacSwiney Gonzalez, M.C., Moya, I., Pons, J.-M., Racey, P.A., Rex, K.,
 568 Sampaio, E.M., Stoner, K.E., Voigt, C.C., Staden, D. von, Weise, C.D., Kalko, E.K.V. (2010).
 569 Long-term monitoring of tropical bats for anthropogenic impact assessment: Gauging the statistical
 570 power to detect population change. *Biological Conservation*, 143, 2797-2807.
- 571 Kubista, C.E., Bruckner, A. (2017). Within-site variability of field recordings from stationary,
 572 passively working detectors. *Acta Chiropterologica*, 19, 189-197.
- 573 Milchram, M., Suarez-Rubio, M., Schröder, A., Bruckner, A. (2019). Estimating population density
 574 of insectivorous bats based on stationary acoustic detectors: A case study. *Ecology and Evolution*,
 575 10, 1135-1144.
- 576 Milne, D.J., Armstrong, M., Fisher, A., Flores, T. & Pavey, C.R. (2004). A comparison of three
 577 survey methods for collecting bat echolocation calls and species-accumulation rates from nightly
 578 Anabat recordings. *Wildlife Research*, 31, 57-63.
- 579 Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and
 580 intra-class correlation coefficient from generalized linear mixed-effects models revisited and
 581 expanded. *Journal of The Royal Society Interface*, 14, 20170213.
- 582 Nocera, T., Ford, W.M., Silvis, A. & Dobony, C.A. (2019). Let's agree to disagree: Comparing
 583 auto-acoustic identification programs for northeastern bats. *Journal of Fish and Wildlife*
 584 *Management*, 10, 346-361.
- 585 O'Keefe, J.M., Loeb, S.C., Hill, H.S. Jr. & Lanham J. D. (2014). Quantifying clutter: A comparison
 586 of four methods and their relationship to bat detection. *Forest Ecology and Management*, 322, 1-9.

- 587 Ratcliffe, J.M. & Jakobson, L. (2018). Don't believe the mike: behavioural, directional, and
 588 environmental impacts on recorded bat echolocation call measures. *Canadian Journal of Zoology*,
 589 96, 283-288.
- 590 R Core Team (2019). R: A language and environment for statistical computing. R Foundation for
 591 Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- 592 Russo, D., Ancillotto, L. & Jones, G. (2018). Bats are still not birds in the digital era: echolocation
 593 call variation and why it matters for bat species identification. *Canadian Journal of Zoology*, 96,
 594 63-78.
- 595 Rydell, J., Nyman, S., Eklöf, J., Jones, G. & Russo, D. (2017). Testing the performances of
 596 automated identification of bat echolocation calls: A request for prudence. *Ecological Indicators*,
 597 78, 416-420.
- 598 Skalak, S.L., Sherwin, R.E. & Brigham, R.M. (2012). Sampling period, size and duration influence
 599 measures of bat species richness from acoustic surveys. *Methods in Ecology and Evolution*, 3, 490-
 600 502.
- 601 Staton T. & Poulton S. (2012). Seasonal variation in bat activity in relation to detector height: a
 602 case study. *Acta Chiropterologica*, 14, 401-408.
- 603 Sugai, L.S.M., Silva, T.S.F., Ribeiro, J.W. jr. & Lluisa, D. (2019), Terrestrial passive acoustic
 604 monitoring: Review and Perspectives. *BioScience*, 69: 15-25.
- 605 Weller, T.J. & Zabel, C.J. (2002). Variation in bat detections due to detector orientation in a forest.
 606 *Wildlife Society Bulletin*, 30, 922-930.
- 607 Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine*,
 608 22, 3527-3541.
- 609

610 Tables

611

612 TABLE 1 Results of likelihood ratio tests of the effects of detector and detector orientation on bat calling activity in mixed effects models. Null models
 613 contained only random effects, the other two models additionally contained detector or detector + orientation as fixed effects. No adequate model was
 614 found for *Myotis myotis*. The null model R² are the conditional (whole-model) coefficients of determination, the R² of the two final models are the
 615 additional R² relative to the nulls. OTU: operational taxonomic unit.

616

OTU, number of observations	model / fixed effects	AIC	log-likelihood	deviance	χ^2	df	P	R ²
total	null	53037	-26494	52987				0.427
n = 35 258	detector	52209	-26049	52099	888.85	30	< 2 * 10 ⁻¹⁶ ***	+ 0.019
	detector/orientation	52210	-26048	52096	2.55	2	0.280	+ 0.019
nyctaloid	null	25597	-12774	25547				0.532
n = 15 112	detector	24503	-12196	24393	1154.08	30	< 2 * 10 ⁻¹⁶ ***	+ 0.052
	detector + orientation	24506	-12196	24392	0.7564	2	0.685	+ 0.052
<i>Pipistrellus</i> <i>pygmaeus</i>	null	6134.5	-3042.2	6084.5				0.118

n = 6 126	detector	6118.7	-3004.3	6008.7	75.797	30	7.798 * 10 ⁻⁶ ***	+ 0.012
	detector + orientation	6108.2	-2997.1	5994.2	14.454	2	7.268 * 10 ⁻⁴ ***	+ 0.012
<i>P. kuhlii/nathusii</i>	null	3347.5	-1648.8	3297.5				0.127
n = 3337.4	detector	3337.4	-1614.7	3229.4	68.05145	29	5.750 * 10 ⁻⁰⁵ ***	+ 0.018
	detector + orientation	3335.9	-1610.9	3221.9	7.5792	3	0.056	+ 0.028
<i>P. pipistrellus</i>	null	2110.3	-1046.1	2092.3				0.199
n = 1 575	detector	2093.6	-1011.8	2023.6	68.6535	26	1.04 * 10 ⁻⁰⁵ ***	+ 0.030
	detector + orientation	2100.5	-1009.3	2018.5	5.1085	6	0.530	+ 0.032
<i>Hypsugo savii</i>	null	1775.5	-878.74	1757.5				0.301
n = 1 229	detector	1798.4	-860.19	1720.4	37.0880	30	0.175	+ 0.015
	detector + orientation	1800.2	-859.13	1718.2	2.1361	2	0.344	+ 0.012
<i>Barbastella barbastellus</i>	null	610.94	-300.47	600.94				0.117
n = 694	detector	613.07	-286.53	573.07	27.8747	15	0.022 *	+ 0.035
	detector + orientation	612.55	-285.28	570.55	2.5147	1	0.113	+ 0.038

617
618
619
620

621 TABLE 2 Results of likelihood ratio tests of the effects of detector distance on bat calling activity in mixed effects models. Null models contained only
 622 random effects, distance models additionally contained distance as fixed effect. Marginal R² is the contribution of the fixed effect to the total R² of the
 623 distance models. OTU: operational taxonomic unit.
 624

OTU, number of observations	model / fixed effects	AIC	log-likelihood	deviance	χ^2	df	P	marginal R ²
total	null	49094	-24542	49084				
n = 24 864	distance	48882	-24435	48870	213.62	1	< 2.2 * 10 ⁻¹⁶ ***	0.006
nyctaloid	null	27084	-13 537	27074				
n = 13 549	distance	27043	-13515	27031	42.845	1	5.926 * 10 ⁻¹¹ ***	0.002
<i>Pipistrellus</i>	null	3272.1	-1631.1	3262.1				
<i>pygmaeus</i>								
n = 2 588	distance	3217.7	-1602.8	3205.7	56.435	1	5.807 * 10 ⁻¹⁴ ***	0.013
<i>P. kuhlii/nathusii</i>	null	2765.9	-1378.0	2755.9				
n = 1 308	distance	2748.7	-1368.3	2736.7	19.254	1	1.144 * 10 ⁻⁰⁵ ***	0.007
<i>P. pipistrellus</i>	null	1568.1	-779.03	1558.1				
n = 916	distance	1536.4	-763.22	1526.4	31.614	0	< 2.2 * 10 ⁻¹⁶ ***	0.026

<i>Hypsugo savii</i>	null	1572.2	-781.09	1562.2				
n = 720	distance	1543.8	-765.89	1531.8	30.401	1	$3.514 * 10^{-08} ***$	0.020
<i>Barbastella barbastellus</i>	null	286.03	-138.01	276.03				
n = 270	distance	277.20	-132.60	265.20	10.83	1	$9.988 * 10^{-4} ***$	0.022
<i>Myotis myotis</i>	null	133.66	-61.830	123.66				
n = 65	distance	133.65	-61.823	123.65	0.0145	0	$< 2.2 * 10^{-16} ***$	0.000

625

626

627 Figure captions

628

629 FIGURE 1 Placement of devices in a study on site-scale variability of passively recording bat call
 630 detectors. Devices were arranged pairwise in the vertices of an octagon. One detector of each pair
 631 was directed towards the centre, the other towards the outside (labelled "in" and "out" at position
 632 E).

633

634 FIGURE 2 File-scale deviations of individual devices in a study on site-scale variability of
 635 passively recording bat detectors. Δ call sequence length is the absolute difference in call sequence
 636 length between individual detectors and the average of 15 neighboring detectors nearby, summed in
 637 5-second time intervals. A - total bat activity, B - nyctaloid species, C - *Pipistrellus pygmaeus*, D -
 638 *P. kuhlii/nathusii*, E - *P. pipistrellus*, F - *Hypsugo savii*, G - *Barbastella barbastellus*, H - *Myotis*
 639 *myotis*. Note the logarithmic scaling of the ordinates. The vertical lines are the 95 (light red) and
 640 99% (dark red) data quantiles, respectively.

641

642 FIGURE 3 Night-scale deviations of individual devices in a study on site-scale variability of
 643 passively recording bat detectors. Δ call sequence length is the absolute difference in call sequence
 644 length between individual detectors and the average of 15 neighboring detectors nearby, summed in
 645 1-night time intervals. See Figure 2 for a key to panel labels and line colors.

646

647 FIGURE 4 Temporal variability of device performance in a study on site-scale variability of
 648 passively recording bat detectors. Point sizes indicate calling activity and color intensities indicate
 649 the deviation of individual detectors from the arithmetic mean of all detectors (red: values above,
 650 blue: below mean). Point sizes are relative to the activity of each panel, therefore not comparable
 651 among panels. See Figure 2 for a key to panel labels.

652

FIGURE 5 Extremely long call sequences in a study on site-scale variability of passively recording bat detectors. Δ call sequence length is the absolute difference in call sequence length between individual detectors and the average of 15 neighboring detectors, summed in 5-second time intervals. To identify extreme values and outliers, only Δ values above the 95% data quantile are shown. To maximize visibility, points are scattered, and violins are plotted in front of points in panels A, B and C, and behind points in the other panels. See Figure 2 for a key to panel labels.

FIGURE 6 Effect of device orientation in a study on site-scale variability of passively recording bat detectors. Δ call sequence length is the absolute difference in call sequence length between two detectors oriented in opposite (180°) directions, summed in 5-second time intervals. Note the logarithmic scaling of the ordinates. The vertical lines are the 95 (light red) and 99% (dark red) quantiles of the data, respectively. See Figure 2 for a key to panel labels and line colors.

FIGURE 7 Effect of spatial distance in a study on site-scale variability of passively recording bat detectors. Δ call sequence length is absolute difference in call sequence length between two detectors deployed at different positions in an octagon, summed in 5-second time intervals. Horizontal lines inside the violins indicate the 25%, 50% (= median), and 75% quantiles, respectively. Since the number of detector pairs compared differed among the distance factor levels, the violins of every panel were scaled to have the same area. Note the square root scaling of the ordinates. See Figure 2 for a key to panel labels.

FIGURE 8 Detection probabilities of operational taxonomic units of bats in a study on site-scale variability of passively recording bat detectors. As a measure of nightly detection probability, the fraction of 16 neighboring devices registering a call sequence was computed. Note that the data set comprises only time intervals with at least one device recording, hence zero probability values are logically impossible.

679

680 FIGURE 9 Sample size (number of devices) and confidence of call activity estimates in a study on
681 site-scale variability of passively recording bat detectors. The 95 % interquantile range of
682 bootstrapped sampling data was used as a measure of statistical confidence. The dotted curve
683 summarizes the data using a loess smoothing function with $\alpha = 0.8$ (80% smoothing). The red
684 arrows indicate the number of detectors necessary to achieve 90 % of the smallest interquantil
685 range. See Figure 2 for a key to panel labels.

686

687 FIGURE 10 Sample size (number of nights) and confidence of call activity estimates in a study on
688 site-scale variability of passively recording bat detectors. See Figure 2 for a key to panel labels and
689 Figure 9 for computational details.

690