

Supplementary Information for 'Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer' by Edwards, A. M., Phillips, R. A., Watkins, N. W., Freeman, M. P., Murphy, E. J., Afanasyev, V., Buldyrev, S., da Luz, M. G. E., Raposo, E. P., Stanley, H. E. & Viswanathan, G. M.

Correspondence should be addressed to A.M.E. (EdwardsAnd@pac.dfo-mpo.gc.ca).

### **Summary of Supplementary Information**

**Supplementary Methods 1. Derivation of likelihood function for flights in 2004.** This derivation takes into account the sampling protocol of the loggers used in 2004.

**Supplementary Methods 2. Derivation of likelihood function for flights in 1992.** This differs from Supplementary Methods 1 because of the different sampling protocol of the loggers. Both likelihood functions are applicable to other data sets that are collected with similar protocols.

**Supplementary Methods 3. Calculations of the expected number of misclassified flight durations for 1992 data.** This shows the importance of considering

the sampling protocol. For example, we would expect 143 flights in the range 1-2 h, but 78 of these would not be directly detected by the loggers.

**Supplementary Methods 4. Likelihood function for a power law over a bounded range for prebinned data.** This extends the methodology of Box 1 to allow analysis of data that are only available already prebinned, as in the deer and bumblebee data. Both bounded and unbounded power laws are considered.

**Supplementary Methods 5. Likelihood function for an exponential distribution over a bounded range for prebinned data.** As for Supplementary Methods 4, but for the exponential distribution.

**Supplementary Methods 6. Analysis of deer data.** Full details of the analysis of the deer data.

**Supplementary Methods 7. Analysis of bumblebee data.** Full details of the analysis of the bumblebee deer data, including, sensitivity to the cut-off point ( $b$ ) of the tail of the bounded distributions.

**Supplementary Table 1. Sensitivity to  $b$  for bumblebee data in high-food scenario.** Statistical analyses for different values of  $b$ .

**Supplementary Table 2. Sensitivity to  $b$  for bumblebee data in low-food scenario.** Statistical analyses for different values of  $b$ , showing for the unbounded range ( $b \rightarrow \infty$ ) the exponential is definitively better supported than the power law.

**Supplementary Figure 1. Original log-log histogram of 1992 wandering albatross flight durations (Fig. 3a of ref. 7) computed here from the original raw data.** This is the original figure that then gets corrected in Fig. 3a.

**Supplementary Figure 2. The 2004 data as a log-log histogram (LBN method).** This shows the 2004 data in LBN form for comparison with Supplementary Fig. 1 for the 1992 data.

**Supplementary Figure 3. Rank/frequency plot for individual birds 1-9 from 2004.** The data for each individual bird do not appear to be consistent with coming from individual exponential distributions, suggesting that the pooled gamma distribution (2) does not arise as a mixture of exponentials.

**Supplementary Figure 4. Rank/frequency plot for individual birds 10-18 from 2004.** Continues Supplementary Fig. 3.

**Supplementary Figure 5. Rank/frequency plot for individual birds 19-20 from 2004.** Continues Supplementary Figs. 3 and 4.

**Supplementary Figure 6. The 1992 data as in Fig. 3a, but unbinned and on linear axes.** The raw data from 1992 as standard linear histograms, to clearly show how adjusting the first and last dry sequences removes the longest ones.

**Supplementary Methods 1. Derivation of likelihood function for flights in 2004**

For the 2004 data, the wet/dry status of each of 20 birds was recorded every 10 s by the salt-water loggers. Locations were recorded by GPS devices, at a resolution of approximately 1 h. The GPS data were used to calculate when each bird departed the colony. These departure times were used to determine the true initial flight durations, eliminating the long initial series of dry readings that do not represent flights (e.g. Fig. 2 for the 1992 data set). Such dry readings commence when the logger is switched on at a computer, and include the time taken to attach the logger to a bird, and the time that the bird spends on the nest before finally flying.

Durations of final flights for each trip were also calculated in a similar manner.

The data set for each bird therefore consists of a wet/dry reading every 10 s. We thus have discrete data from which we want to infer the continuous probability density function (pdf) of flight durations. Here we develop a multinomial<sup>31</sup> likelihood approach to this, that takes into account the fact that the (continuous) true flight durations have been recorded in a discrete manner. This is more important for the shorter flights than the longer ones of, say, 1 h. We then take a similar approach to derive the likelihood function for the 1992 data set, for which all recorded flights are  $\geq 1$  h and ignoring the sampling protocol does profoundly impact the results.

An example sequence of wet/dry readings is wet-dry-dry-dry-dry-wet, each 10 s apart – the bird was on the water and then undertook a short flight before landing. We define a record to be the number of consecutive dry readings in-between two wet readings; thus the example gives a record of 4. This record corresponds to a true flight duration that may be anywhere in the range 30-50 s. If the bird took off just before the first dry reading, and then landed just after the final dry reading then the flight lasted just over 30 s. Whereas at the other extreme, if the bird took off just after the first wet reading and landed just before the final

wet reading then the flight lasted just under 50 s.

We consider 30 s to be the minimum dry time that represents a true flight between food sources. Shorter dry times may simply be a bird lifting its leg out of the water to scratch, or could correspond to an abandoned take off. Thus we ignore records of 1, 2 and 3, as these would include such short flights. Our conclusions would be unaffected by using, say, 60 s instead of 30 s as the minimum.

Following ref. 7, the data for all birds were pooled. The resulting data set consists of a set of records (separate flights)  $\mathbf{r} = \{r_i\}$ , with  $i = 1, 2, 3, \dots, 1416$ , and each  $r_i$  being an integer. The full data set is shown in the rank/frequency plot in Fig. 1 of the main paper, where for simplicity a record of, say, 6 is plotted as a flight of duration  $t = 60$  s (even though this record may represent a true flight of duration 50-70s – but this is clearly only an issue for the very shortest flights). Note that there can be multiple flights that have the same value of  $t$ , and so these give multiple circles with consecutive ranks on the ordinate of Fig. 1. Because the fitted distribution, to be calculated shortly, reaches 30 s, we also show records of 3 in Fig. 1, even though, as just discussed, these were not included in the statistical analysis. The rank/frequency plot is clearly not a straight line, indicating that the data are not power-law distributed. We now develop the statistical analysis to test

whether the data are consistent with coming from a shifted gamma distribution.

Define the pdf of flight durations (which is what we want to determine) as  $f(x; \theta)$ , where  $x$  is the random variable representing a flight duration (in s), and  $\theta$  is a vector of parameters. We will find the maximum likelihood estimates (MLEs) for the parameters, by maximizing the log-likelihood function<sup>18,19</sup>  $l(\theta|\mathbf{r})$  defined by

$$l(\theta|\mathbf{r}) = \log[L(\theta|\mathbf{r})] \quad (11)$$

$$= \log[\mathbf{P}(\mathbf{r}|\theta)] \quad (12)$$

$$= \log \left[ \prod_{i=1}^n \mathbf{P}(r_i|\theta) \right] \quad (13)$$

$$= \sum_{i=1}^n \log [\mathbf{P}(r_i|\theta)], \quad (14)$$

where  $L(\theta|\mathbf{r})$  is the likelihood function for  $\theta$  given the data  $\mathbf{r}$ ,  $n = 1416$  is the total number of flights and  $\mathbf{P}(r_i|\theta)$  is the probability of obtaining a record  $r_i$  when the parameters in the pdf are given by  $\theta$ .

Clearly, some values of  $r_i$  will be repeated. So we define  $d_j$  to be the number of records that equal  $j$ , where  $j = 1, 2, 3, \dots, J$ , and  $J = \max_i(r_i)$  is the maximum record. Note that  $\sum_{j=1}^J d_j = n$ . Since only records of 4 and above are considered, we have  $d_1 = d_2 = d_3 = 0$ .

So the multinomial log-likelihood function<sup>31</sup> is

$$l(\theta|\mathbf{r}) = \sum_{j=1}^J d_j \log [\mathbf{P}(j|\theta)]. \quad (15)$$

For a single flight, if  $R$  is the random variable representing the value of the record, then

$$\mathbf{P}(R = j|\theta) = \int_0^\infty \mathbf{P}(R = j|\text{flight duration is } x)f(x; \theta)dx, \quad (16)$$

for integer values of  $j = 1, 2, 3, \dots, J$ . Next, define the random variable  $U$  to be the time from a bird taking off from the water to the time of the next dry reading by the salt-water logger. Then, for each flight,  $U$  is uniformly distributed on the interval  $[0, 10)$  seconds, such that its pdf is

$$g_U(u) = \begin{cases} \frac{1}{10}, & u \in [0, 10) \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

Then to obtain a record  $j$ , the  $x-U$  interval that commences at the first dry reading must last long enough to get a further  $j-1$  dry readings, but not last long enough to get  $j$  more. Thus

$$\mathbf{P}(R = j|\text{flight duration is } x) = \mathbf{P}(10(j-1) \leq x - U < 10j) \quad (18)$$

$$= \mathbf{P}(10(j-1) - x \leq -U < 10j - x) \quad (19)$$

$$= \mathbf{P}(x - 10j < U \leq x - 10(j-1)) \quad (20)$$



$$= \int_{x-10j}^{x-10(j-1)} g_U(u) du, \quad (21)$$

where the final line comes from the known pdf of  $U$ .

The two limits of the integral are 10 s apart, and so given  $x$  there are two cases:

i) The top limit of the integral is in  $[0, 10)$  and therefore the bottom limit is not, i.e.  $x - 10(j - 1) \in [0, 10)$ , such that  $x \in [10(j - 1), 10j)$ . Then, from (17),

$$\int_{x-10j}^{x-10(j-1)} g_U(u) du = \int_0^{x-10(j-1)} \frac{1}{10} du \quad (22)$$

$$= \frac{x - 10(j - 1)}{10}. \quad (23)$$

ii) The bottom limit of the integral is in  $[0, 10)$  and therefore the top limit is not, i.e.  $x - 10j \in [0, 10)$ , such that  $x \in [10j, 10(j + 1))$ . Then, from (17),

$$\int_{x-10j}^{x-10(j-1)} g_U(u) du = \int_{x-10j}^{10} \frac{1}{10} du \quad (24)$$

$$= \frac{10 - x + 10j}{10}. \quad (25)$$

These two cases simplify to

$$P(R = j | \text{flight duration is } x) = \begin{cases} 1 + \frac{x}{10} - j, & x \in [10(j - 1), 10j) \\ 1 - \frac{x}{10} + j, & x \in [10j, 10(j + 1)) \\ 0, & \text{otherwise.} \end{cases} \quad (26)$$

So, say that a bird flies for  $x$  seconds, with  $x \in [60, 70)$ . Then  $P(R = 7|x) = 1 + x/10 - 7$ , from setting  $j = 7$  in (26), and  $P(R = 6|x) = 1 - x/10 + 6$ , from setting  $j = 6$  in (26). As required, these two probabilities sum to 1 and no other records are possible for this flight.

Now, (16) becomes

$$P(R = j|\theta) = \int_{10(j-1)}^{10j} \left(1 + \frac{x}{10} - j\right) f(x; \theta) dx + \int_{10j}^{10(j+1)} \left(1 - \frac{x}{10} + j\right) f(x; \theta) dx \quad (27)$$

$$= \int_{10(j-1)}^{10(j+1)} f(x; \theta) dx - \int_{10(j-1)}^{10j} \left(j - \frac{x}{10}\right) f(x; \theta) dx - \int_{10j}^{10(j+1)} \left(\frac{x}{10} - j\right) f(x; \theta) dx. \quad (28)$$

The first term in (28) represents all flights in the range  $10(j-1)$  to  $10(j+1)$ , and the second and third terms then reduce this to account for the sampling protocol.

A formulation that is useful numerically is

$$P(R = j|\theta) = (1 - j) \int_{10(j-1)}^{10j} f(x; \theta) dx + (1 + j) \int_{10j}^{10(j+1)} f(x; \theta) dx + \frac{1}{10} \left[ \int_{10(j-1)}^{10j} x f(x; \theta) dx - \int_{10j}^{10(j+1)} x f(x; \theta) dx \right]. \quad (29)$$

Since flights of less than 30 s are not considered to be part of the food-searching mechanism that we are modelling, we have  $f(x; \theta) = 0$  for  $x < 30$ .

Therefore,

$$\begin{aligned} P(R = 1|\theta) &= 0, \\ P(R = 2|\theta) &= 0, \\ P(R = 3|\theta) &= 4 \int_{30}^{40} f(x; \theta) dx - \frac{1}{10} \int_{30}^{40} x f(x; \theta) dx. \end{aligned} \quad (30)$$

Now, we have records of  $R = 3$  in the data, but some of these will represent dry times in the range 20-40 s. In the model, and thus (30), we have taken  $f(x; \theta) = 0$  for  $x < 30$  s (see Methods). Since we exclude records of 3 from the data (but the probability of getting such a record is non-zero), to make the remaining  $P(R = j|\theta)$  sum to 1 in the multinomial log-likelihood function we need to divide each probability by  $1 - P(R = 3|\theta)$ . For the log-likelihood function, this translates to inserting the definitions  $P(R = j|\theta)$  from (28) into (15), and then subtracting  $n \log(1 - P(R = 3|\theta))$  from the log-likelihood function.

The multinomial log-likelihood function (15) then becomes

$$\begin{aligned} l(\theta|\mathbf{r}) &= \sum_{j=1}^J d_j \log \left\{ \int_{10(j-1)}^{10(j+1)} f(x; \theta) dx - \int_{10(j-1)}^{10j} \left( j - \frac{x}{10} \right) f(x; \theta) dx \right. \\ &\quad \left. - \int_{10j}^{10(j+1)} \left( \frac{x}{10} - j \right) f(x; \theta) dx \right\} - n \log(1 - P(R = 3|\theta)). \end{aligned} \quad (31)$$

This is the function to be maximized with respect to the parameters in  $\theta$ . It is valid

for any pdf  $f(x; \theta)$  for flight durations longer than 30 s.

For a gamma distribution of flight durations (in seconds), we have two parameters, the shape,  $s$ , and the rate,  $\tilde{r}$  (with units  $\text{s}^{-1}$ ); thus,  $\theta = (\tilde{r}, s)$ . The pdf is

$$f(x; \theta) = \begin{cases} \frac{\tilde{r}^s}{\Gamma(s)} (x - 30)^{s-1} e^{-\tilde{r}(x-30)}, & x > 30 \\ 0, & x \leq 30. \end{cases} \quad (32)$$

In the main text we discuss flight durations  $t$  in h, given by  $t = 3600x$ , and so report the rate  $r$  in units  $\text{h}^{-1}$ , given by  $r = \tilde{r}/3600$ .

We numerically found the values of  $\tilde{r}$  and  $s$  that maximize the log-likelihood. To test for goodness-of-fit we used the G-test (likelihood-ratio test) with Williams's correction<sup>29</sup>. The 95% CIs were obtained by the profile likelihood-ratio test<sup>18</sup>. The statistic  $R' = 2(\tilde{l}_p - \tilde{l}_{MLE})$  has a chi-square distribution with 1 df, where  $\tilde{l}_p$  is the negative log likelihood of the data given parameter value  $p$  (i.e. shape or rate), and  $\tilde{l}_{MLE}$  is the negative log likelihood of the data at the MLE values<sup>18</sup>. The 95% confidence interval for shape, for example, is then given by values of shape for which  $R' < \chi_{0.95[1]}^2$ , where  $\tilde{l}_p$  is minimized (with respect to rate) for each value of shape. All computations were performed using R, version 2.2.0 ([www.r-project.org](http://www.r-project.org)).

## Supplementary Methods 2. Derivation of likelihood function for flights in 1992

First, in Supplementary Fig. 6 we show the raw 1992 logger data, before and after adjustments of the first and last dry sequences. These are the same data as in Fig. 3a, but plotted on linear axes and not binned. Plotting this way clearly identifies the individual long sequences that get removed. In Fig. 3c we show the same data as a rank-frequency plot, as in Fig. 1 for the 2004 data.

The 1992 data were obtained from salt-water loggers, but the sampling protocol was different to that for the 2004 data. In 1992, the devices took a wet/dry reading every 3 s. For every 15 s interval, a bird was considered to be on the water if the logger was wet for 9 s or more. Due to data storage limitations, the logger only saved the total number of 15 s intervals in each hour for which the bird was considered wet. Thus the time series for each bird consisted of a number from 0 to 240 every hour (e.g. Fig. 1 of ref. 7). Thus, flights shorter than 1 h could not be directly inferred from the data, because a wet count of 60, for example, could imply one flight of 45 minutes ( $(240-60)/240$  of an hour), or several shorter flights interspersed by landings. So in ref. 7, consecutive hourly wet counts of 0, in between non-zero hours, were used to give flight times of 1, 2, 3, ... hours. A

sequence of hours that goes wet-dry-wet was then assumed to be a flight of 1 h.

However, such a 1 h record of dry readings could come from a true flight anywhere in the range 1-3 h. For example, consider the situation where the hourly counts started at 10:00, 11:00, 12:00 and 13:00, and a bird took off at 10:15 for a 2.5 h flight, landing again at 12:45. Then only the 11:00-12:00 hourly record would be completely dry, and so in ref. 7 this 2.5 h flight would have been considered to be a 1 h flight. Similarly, some flights of duration 1-2 h will not give a completely dry hour, and so will not get recorded at all in the data. Here we modify the approach developed above for the 2004 data to take both of these factors into account, and obtain the relevant likelihood function to infer the distribution of actual flight durations.

We define the pdf of true flight durations as  $g(t; \theta)$ , where  $t$  is the random variable representing a flight duration (in hours), and  $\theta$  is again a vector of parameters. The data set is the set of records  $\mathbf{v} = \{v_i\}$ ,  $i = 1, 2, 3, \dots, N$ , with each record  $v_i$  being a number of consecutive dry hours recorded between two wet hours, and  $N = 335$  is the number of flight records (pooled for all birds) in the data set.

Analogous to (15), we want to determine and maximize the multinomial log-likelihood function  $l(\theta|\mathbf{v})$  given by

$$l(\theta|\mathbf{v}) = \sum_{j=1}^{J_v} c_j \log [\mathbf{P}(j|\theta)], \quad (33)$$

where  $c_j$  is the number (count) of records that equal  $j$ ,  $J_v = \max_i(v_i)$  is the maximum record, and  $\mathbf{P}(j|\theta)$  is the probability of obtaining a record  $j$  when the parameters are  $\theta$ .

For a single flight, if  $V$  is the random variable representing the value of the record, then

$$\mathbf{P}(V = v|\theta) = \int_0^\infty \mathbf{P}(V = v|\text{flight duration is } t)g(t; \theta)dt. \quad (34)$$

We will later substitute this into (33), replacing  $v$  by  $j$ .

Considering only flights  $\geq 1$  h, we define the random variable  $U_h$  to be the time (in h) between a bird taking off and the start of the next hourly block of 240 measurements. Then,  $U_h$  is uniformly distributed on the interval  $[0, 1)$  hours, such that its pdf is simply

$$g_{U_h}(u_h) = \begin{cases} 1, & u_h \in [0, 1) \\ 0, & \text{otherwise.} \end{cases} \quad (35)$$

Then, analogous to (18)-(21),

$$P(V = v | \text{flight duration is } t) = P(v \leq t - U_h < v + 1) \quad (36)$$

$$= P(v - t \leq -U_h < v - t + 1) \quad (37)$$

$$= P(t - v - 1 < U_h \leq t - v) \quad (38)$$

$$= \int_{t-v-1}^{t-v} g_{U_h}(u_h) du_h. \quad (39)$$

The two limits of the integral are 1 h apart, and so given  $t$  there are two cases:

i) The top limit of the integral is in  $[0, 1)$  and therefore the bottom limit is not, i.e.

$t - v \in [0, 1)$ , such that  $t \in [v, v + 1)$ . Then

$$\int_{t-v-1}^{t-v} g_{U_h}(u_h) du_h = \int_0^{t-v} 1 du_h \quad (40)$$

$$= t - v. \quad (41)$$

ii) The bottom limit is in  $[0, 1)$  and therefore the top limit is not, i.e.  $t - v - 1 \in$

$[0, 1)$ , such that  $t \in [v + 1, v + 2)$ . Then

$$\int_{t-v-1}^{t-v} g_{U_h}(u_h) du_h = \int_{t-v-1}^1 1 du_h \quad (42)$$

$$= 2 - t + v. \quad (43)$$



These two cases simplify to

$$P(V = v | \text{flight duration is } t) = \begin{cases} t - v, & t \in [v, v + 1) \\ 2 - t + v, & t \in [v + 1, v + 2) \\ 0, & \text{otherwise,} \end{cases} \quad (44)$$

for  $v > 0$ . So say a bird flies for  $t$  hours with  $t \in [5, 6)$ . Then  $P(V = 5 | t) = t - 5$ ,

from setting  $v = 5$  in (44), and  $P(V = 4 | t) = 6 - t$ , from setting  $v = 4$  in (44).

As required, these two probabilities sum to 1 and no other records are possible.

For  $v = 0$ ,

$$P(V = 0 | \text{flight duration is } t) = \begin{cases} 1, & t \in [0, 1) \\ 2 - t, & t \in [1, 2) \\ 0, & \text{otherwise,} \end{cases} \quad (45)$$

i.e. all flights  $< 1$  h yield a record of 0, and so do some flights of duration 1-2h.

These latter flights would not be present in the data set.

Now, for  $v > 0$  equation (34) becomes

$$P(V = v | \theta) = \int_v^{v+1} (t - v)g(t; \theta)dt + \int_{v+1}^{v+2} (2 - t + v)g(t; \theta)dt.$$

For  $v = 0$  we insert (45) into (34) to obtain

$$P(V = 0 | \theta) = \int_0^1 g(t; \theta)dt + \int_1^2 (2 - t)g(t; \theta)dt. \quad (46)$$

We do not have records of  $v = 0$  in the data. So analogous to the incorporation of (30) for the 2004 data, but for a different reason, for  $v > 0$  we need to divide each  $P(V = v|\theta)$  by  $1 - P(V = 0|\theta)$ . Then the multinomial probabilities sum to 1 in the likelihood function. Again, we insert the definitions  $P(V = v|\theta)$  from (46) into (33), replacing  $v$  by  $j$ , and subtract  $N \log(1 - P(V = 0|\theta))$  from the log-likelihood function.

The resulting multinomial log-likelihood function (33) is then

$$l(\theta|\mathbf{v}) = \sum_{j=1}^{J_v} c_j \log \left\{ \int_j^{j+1} (t-j)g(t;\theta)dt + \int_{j+1}^{j+2} (2-t+j)g(t;\theta)dt \right\} - N \log(1 - P(V = 0|\theta)). \quad (47)$$

This is the function to be maximized to find the MLEs and confidence intervals for the parameters in  $\theta$ , and to be used in goodness-of-fit tests.

To test for the shifted gamma distribution of flight durations, as for the 2004 data, we have

$$g(t; \theta) = \begin{cases} \frac{r^s}{\Gamma(s)} \left(t - \frac{1}{120}\right)^{s-1} e^{-r(t-1/120)}, & t > \frac{1}{120} \\ 0, & t \leq \frac{1}{120}. \end{cases} \quad (48)$$

where  $t$  is in hours,  $s$  is the shape parameter,  $r$  is the rate parameter in  $\text{h}^{-1}$ , and  $1/120 = 30/3600$  is the conversion of 30 s into hours.

### Supplementary Methods 3. Calculations of the expected number of misclassified flight durations for 1992 data

From the above calculations for the 1992 data, it is clear that errors can arise from assuming that all flights in the range 1-2 h yielded records of 1. Furthermore, flights in the range 2-3 h can yield records of 1. In ref. 7 it was implicitly assumed that a record of 1 represented a flight of 1-2 h, and no account was taken of the data sampling protocol. The results in Fig. 3b show that the data-sampling protocol and binning procedure do need to be properly accounted for to determine the expected distribution of flight durations.

Assuming that the pdf (48) with MLE values represents the true distribution of flight durations, we now determine the expected number of true flights that would have been in the range 1-2 h but were missed completely by the sampling protocol (as they would have ended up with a record of 0).

We first calculate  $N_f$ , the expected total number of flights (including those that did not get recorded) that would have been required to yield the 335 records. Thus  $N_f$  consists of the 335 recorded flights (that by definition were all  $\geq 1$  h), plus the flights of 1-2 h that did not get recorded, plus the flights  $< 1$  h (that were

not recorded).

The 335 records are those  $> 0$ , thus

$$[1 - \mathbf{P}(V = 0|\hat{\theta})]N_f = 335, \quad (49)$$

where  $1 - \mathbf{P}(V = 0|\hat{\theta})$  is the probability of getting a record  $> 0$ , given that the parameters take the MLE values  $\theta = \hat{\theta}$ . Using the known  $\hat{\theta}$ , we have

$$\mathbf{P}(V = 0|\hat{\theta}) = 0.534, \quad (50)$$

i.e. over half of the  $N_f$  flights are expected to give a record of 0, and thus will not explicitly show up in the data (many of them will be  $< 1$  h). Then

$$N_f = 719, \quad (51)$$

of which 335 are expected to give records  $\geq 1$  and thus be detected in the data.

The expected number of 0 records resulting from flights of duration  $t \in [1, 2)$ , which we will call  $E_0$ , is given by

$$E_0 = \mathbf{P}(V = 0|t \in [1, 2)) \times \mathbf{P}(t \in [1, 2)) \times N_f \quad (52)$$

$$= \int_1^2 \mathbf{P}(V = 0|t)g(t;\hat{\theta})dt \times N_f \quad (53)$$

$$= \int_1^2 (2 - t)g(t;\hat{\theta})dt \times N_f \quad (54)$$

$$= 78, \quad (55)$$

where we have used (45) with  $v = 0$  to obtain (54), and rounded to the nearest integer (and shall continue to do so where appropriate).

The expected total number of flights in the range  $[1, 2)$  is

$$\int_1^2 g(t; \hat{\theta}) dt \times N_f = 143. \quad (56)$$

Thus, we expect there to be 143 flights of duration 1-2 h, of which 78 will not be directly detected by the data. So simply considering records of 1 to represent all the 1-2 h flights is incorrect.

Finally, we calculate the expected number of flights of duration 2-3 h that will give a record of 1 h (call this  $E_1$ ). These would have been incorrectly considered to be flights in the range 1-2 h in ref. 7. Similarly to above, we have

$$E_1 = \int_2^3 \mathbf{P}(V = 1|t)g(t; \hat{\theta})dt \times N_f \quad (57)$$

$$= \int_2^3 (3 - t)g(t; \hat{\theta})dt \times N_f \quad (58)$$

$$= 47, \quad (59)$$

where we have used (44) with  $v = 1$  to obtain (58). Analogous to (56), we expect a total of 88 flights in the range 2-3 h. So of these 88 flights we expect 47 to yield a record of 1, and thus these 47 would have been incorrectly assumed in ref. 7 to represent flights of duration 1-2 h.

So in summary, in the data there were 102 records of 1, and 92 of 2, which would have been considered in ref. 7 to respectively represent flights in the ranges 1-2 h and 2-3 h. However, here we have shown that we would expect there to be (based on the MLE values) 143 real flights in the range 1-2 h, of which only 65 would yield a record of 1, the remaining 78 giving a record of 0 and thus not be detected by the data. And we would expect 88 true flights in the range 2-3 h, but that 47 of these would yield a record of 1, and therefore have been incorrectly considered to be of duration 1-2 h.

#### **Supplementary Methods 4. Likelihood function for a power law over a bounded range for prebinned data**

The bumblebee and deer data shown in ref. 10 were digitized from histograms presented in the original papers<sup>27,28</sup>. Here we develop a likelihood approach to analyze these data sets. The methods are general and so can be applied to any binned data set.

We want to test whether a power law occurs over a given range  $[a, b]$ . Thus, the pdf is

$$f(x) = Cx^{-\mu}, \quad x \in [a, b], \quad (60)$$

where  $C$  is the normalization constant given by  $C = (\mu - 1)/(a^{1-\mu} - b^{1-\mu})$ , which is obtained by solving  $\int_a^b f(x)dx = 1$ . We are only considering the distribution of data that lie in the range  $[a, b]$ . This formulation requires explicit specification of the range over which power-law behaviour is being tested. It is commonly stated that a power law occurs over ‘the tail of the distribution’, and a line drawn or fitted over this range, without qualification as to how the tail  $[a, b]$  is determined.

Consider the data to be counts  $d_j$  in bins indexed by  $j = 1, 2, 3, \dots, J$ , defining  $J$  to be the index of the final bin. Let  $w$  be the bin width (assumed equal for all bins), such that first bin covers the range  $a$  to  $a + w$ , and the right-hand side of the final bin is  $a + Jw$ . The maximum value that we consider attainable by the data,  $b$ , thus satisfies  $b \geq a + Jw$ . Simply taking  $b$  to equal  $a + Jw$  assumes that the maximum data point recorded represents the maximum possible attainable (to within  $w$ ). This is unlikely to occur in practice, particularly for small data sets, and especially in a power-law situation. This is a subtle point that has received little attention.

We define  $b = a + J'w$ , where integer  $J' \geq J$ . Therefore new bins  $J+1, J+2, \dots, J'$  have counts of 0, and so  $d_{J+1} = d_{J+2} = \dots = d_{J'} = 0$ . This allows the possibility that measurements could attain values higher than those that happen to

be sampled in the particular data set being studied.

For a single data value, the probability of being in bin  $j$  given the parameter  $\mu$  is

$$P(\text{being in bin } j|\mu) = \int_{a+(j-1)w}^{a+jw} Cx^{-\mu} dx \quad (61)$$

$$= \frac{C}{1-\mu} \left[ x^{1-\mu} \right]_{a+(j-1)w}^{a+jw} \quad (62)$$

$$= \frac{C}{1-\mu} \left[ (a+jw)^{1-\mu} - (a+(j-1)w)^{1-\mu} \right] \quad (63)$$

$$= \frac{(a+(j-1)w)^{1-\mu} - (a+jw)^{1-\mu}}{a^{1-\mu} - b^{1-\mu}}, \quad (64)$$

substituting  $C$  to obtain (64). Note that  $\sum_{j=1}^{J'} P(\text{being in bin } j|\mu) = 1$ , as required.

The log likelihood function, analogous to (15), is

$$l(\mu|\text{data}) = \sum_{j=1}^J d_j \log [P(\text{being in bin } j|\mu)] \quad (65)$$

$$= -n \log(a^{1-\mu} - b^{1-\mu}) + \sum_{j=1}^J d_j \log \left[ (a+(j-1)w)^{1-\mu} - (a+jw)^{1-\mu} \right], \quad (66)$$

which can be used to find MLEs and confidence intervals. The summation is taken to  $J$  instead of  $J'$  because  $d_{J+1} = d_{J+2} = \dots = d_{J'} = 0$ . Therefore, an increase in  $b$  does not change the summation term in (66), but only changes the first term; the first term is related to the normalization constant  $C$ . The likelihood function



for the idealized infinite power-law tail can be derived using the same approach, and the resulting function is equivalent to taking the limit  $b \rightarrow \infty$  in (66).

### Supplementary Methods 5. Likelihood function for an exponential distribution over a bounded range for prebinned data

The pdf for an exponential distribution over the range  $[a, b]$  is

$$f(x) = Ae^{-\lambda x}, \quad x \in [a, b], \quad (67)$$

where  $A$  is the normalization constant given by  $A = \lambda / (e^{-\lambda a} - e^{-\lambda b})$ , as obtained by solving  $\int_a^b f(x) dx = 1$ .

For a single data value, the probability of being in bin  $j$  given the parameter  $\lambda$  is

$$P(\text{being in bin } j | \lambda) = \int_{a+(j-1)w}^{a+jw} Ae^{-\lambda x} dx \quad (68)$$

$$= -\frac{A}{\lambda} \left[ e^{-\lambda x} \right]_{a+(j-1)w}^{a+jw} \quad (69)$$

$$= -\frac{A}{\lambda} \left[ e^{-\lambda(a+jw)} - e^{-\lambda(a+(j-1)w)} \right] \quad (70)$$

$$= -\frac{Ae^{-\lambda(a+jw)}}{\lambda} \left[ 1 - e^{\lambda w} \right] \quad (71)$$

$$= \frac{e^{-\lambda(a+jw)} \left[ e^{\lambda w} - 1 \right]}{e^{-\lambda a} - e^{-\lambda b}}. \quad (72)$$

The log likelihood function is then

$$l(\lambda|\text{data}) = \sum_{j=1}^J d_j \log [\text{P}(\text{being in bin } j|\lambda)] \quad (73)$$

$$= n \log \left( \frac{e^{\lambda w} - 1}{e^{-\lambda a} - e^{-\lambda b}} \right) - \lambda n a - \lambda w \sum_{j=1}^J d_j j. \quad (74)$$

The likelihood function for the idealized infinite tail (so testing a standard exponential distribution that has been shifted by  $a$ ) can be derived using the same approach. The resulting function is equivalent to taking the limit  $b \rightarrow \infty$  in (74), and can be solved analytically to obtain the MLE of  $\lambda$ .

### Supplementary Methods 6. Analysis of deer data

Figures 3c and 3d of ref. 10 show log-log histograms of foraging times of deer, for unfenced and fenced scenarios. The data were digitized from ref. 27. In ref. 10 the foraging times were assumed to represent time intervals between the deer finding food. However, the times actually represent time spent cropping and processing food at a particular foraging site (pages 608 and 610 in ref. 27), i.e. handling times, rather than time spent moving between stations. Nevertheless we re-analyzed the data in order to test the methods commonly used to demonstrate Lévy flight behaviour.

We digitized the original histograms from ref. 27 to obtain the data. In the log-log histograms in ref. 10 (our Fig. 4) these same values were plotted on a log-log scale, with no lumping of the original bins. Thus the log-log histograms show bins that are of equal width on a linear scale (namely a width of 20 s, as in the original histograms in ref. 27). However, for the original wandering albatross study<sup>7</sup>, the log-log histograms had bins that progressively doubled in width. And for the wandering albatross plot, the geometric means (midpoints on a log scale) of bins were plotted on the abscissa, whereas for the deer the midpoints on a linear scale were used. Our likelihood approach avoids these potential sources of error.

For both scenarios, the final bin of the original data was seemingly omitted in ref. 10. We do likewise, so that we perform our statistical studies on the same data as used in ref. 10. Also, the lumping of bins means that bins of zero count in the original data do not show up in any way on the log-log histograms (but they do get considered in our likelihood approach).

In Figures 3c and 3d of ref. 10 (our Fig. 4), the straight lines representing power laws were drawn across all points excluding the first point (i.e. the tails were considered to be the second points onwards). Thus, for the ranges  $[a, b]$  needed for our likelihood analyses, we use the outside endpoints of these bins. Specifically,

for the unfenced scenario we have  $[a, b] = [20, 180]$ , and for the fenced scenario  $[a, b] = [20, 120]$ .

The Lévy flight hypothesis, as conventionally stated, relates to power-law tails with no mention of an upper limit (i.e.  $b \rightarrow \infty$ ). We tested both data sets under the assumption of no upper limit, and both sets gave  $p < 10^{-4}$  for the power law, and Akaike weights  $< 10^{-8}$  for the power law compared to the exponential. Thus, a true Lévy flight with an infinite tail is completely inconsistent with the data.

### Supplementary Methods 7. Analysis of bumblebee data

Figure 3a of ref. 10 shows two linear histograms of inter-flower distances flown by bumblebees. The data were digitized from ref. 28. One histogram is for a high-food scenario and the other is for a low-food scenario. In Fig. 3b of ref. 10 these data sets were plotted on a log-log histogram, and lines of slope  $\mu \approx 2$  and  $\mu \approx 3.5$  were shown to indicate power-law behaviour in the tails.

To obtain the data for our analyses, we digitized the original histograms from ref. 28 as well as the reproduced linear histograms from ref. 10, so that

we could minimize any discrepancies due to digitization. Consider the low-food case, namely the second diagrams in Fig. 3a of ref. 10 and Fig. 4 of ref. 28. The original bin width is  $w = 37.5$ , and the values plotted are percentages of the total 352 flights. To obtain the counts we utilized this latter fact because each count must be a multiple of  $100/352$ . (The units of distance were given in the original Fig. 4 of ref. 28 as cm, but should actually be mm, as in Figures 1 and 2 of ref. 28; B. Heinrich, pers. comm. This does not affect our conclusions, or those of ref. 10). The final bin in the tail of the original ref. 28 plot was omitted in the ref. 10 plot. We do likewise as our interest is in comparing methods for evaluating power laws, and so wish to be consistent with the data used in ref. 10.

For the log-log histograms in ref. 10, the data were smoothed using running averaging and then lumped<sup>10</sup> (but not in the manner described for the albatross data in ref. 7). We note that a linear regression fit to the seven highest lumped bins for the low-food scenario gives a slope of  $\mu = 2.33$ , larger than the value of 2 shown in Fig. 3b of ref. 10. The likelihood approach presented here avoids any issues concerning the lumping of bins, and uses the original raw data as binned in ref. 28.

For both food scenarios we need a range of  $[a, b]$  over which to test the

power-law and exponential models. We used the ranges implied by the straight lines drawn in Fig. 3b of ref. 10. For both scenarios we take  $a = 150$ , which is the only value that is consistent with being both a break point of the original bins of ref. 28, and with lying between the third and fourth lumped bins in Fig. 3b of ref. 10 (which is where the power-law lines are drawn to). For the high-food scenario we take  $b = 412.5$  (the endpoint of the corresponding line on Fig. 3b of ref. 10), and for the low-food scenario we take  $b = 937.5$ , which is the right endpoint of the range of data included in the final lumped bin in Fig. 3b of ref. 10. These values, as used in the main text, assume that the highest data recorded is the highest data possible, and are the most favourable for the power law compared to the exponential.

In Supplementary Table 1, for the high-food scenario, we show the effect of increasing  $b$  above the default value of 412.5, to allow for the (very likely) possibility that the true value of  $b$  will be larger than that based on the largest observed data. This possibility is particularly likely to occur for power laws. We also test the models under the idealized Lévy flight assumption of an infinite tail (i.e.  $b \rightarrow \infty$ ). Supplementary Table 1 demonstrates that the computed MLE value for  $\mu$  is sensitive to  $b$ , and that neither model is strongly favoured for any value of

$b$ , even for the idealized infinite situation.

In Supplementary Table 2 we test the sensitivity to  $b$  for the low-food scenario. Extending  $b$  to just one or two bins beyond the default shows that the mild support in favour of the power law (as given by the Akaike weights) is not robust to the value of  $b$ . For the idealized assumption of an infinite tail, the exponential distribution is overwhelmingly favoured. In reality, the ‘true’ value of  $b$  in any situation will lie somewhere between these extremes. This is a general issue that requires further attention.

31. Lawless, J. F. *Statistical Models and Methods for Lifetime Data* (2nd ed., Wiley series in Probability and Statistics, Wiley, New Jersey, 2003).

**Supplementary Table 1.** Sensitivity to  $b$  for high-food bees data

Final bin	$b$	MLE for $\mu$	MLE for $\lambda$	Akaike weights	
		(95% CI)	(95% CI)	Power law	Exponential
		$G, p$	$G, p$		
11	412.5	3.68 (2.14, 5.42) 0.16, 0.69	0.0153 (0.0086, 0.0234) 0.59, 0.44	0.59	0.41
12	450	3.85 (2.38, 5.55) 0.10, 0.75	0.0159 (0.0094, 0.0237) 0.50, 0.48	0.55	0.45
13	487.5	3.97 (2.54, 5.63) 0.075, 0.78	0.0162 (0.0099, 0.0239) 0.46, 0.50	0.52	0.48
25	937.5	4.29 (3.06, 5.84) 0.063, 0.80	0.0166 (0.0108, 0.0241) 0.41, 0.52	0.41	0.59
none	$\infty$	4.34 (3.18, 5.87) 0.076, 0.78	0.0166 (0.0108, 0.0241) 0.41, 0.52	0.40	0.60

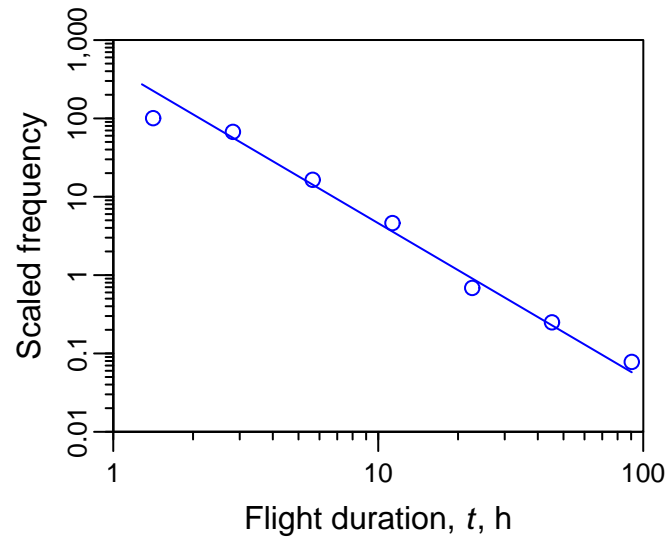


**Supplementary Table 1. Sensitivity to  $b$  for bumblebee data in high-food scenario.** Final bin values represent the index of the right-most original bin<sup>28</sup> that is included;  $b$  is then the right endpoint of this bin. A final bin of 11 was used in the main text, and the value of 25 is also used here as that is the default for the low-food scenario (so it could be argued that here we have counts of 0 in bins 12 to 25). Setting  $b \rightarrow \infty$  represents the idealized Lévy flight assumption of no maximum value. The MLE for  $\mu$  is sensitive to  $b$ . For all goodness-of-fit tests,  $n = 25$  and  $df=1$ . The small sample size and rapid decay of the distributions mean that for all the values of  $b$ , neither model conclusively has the most support.

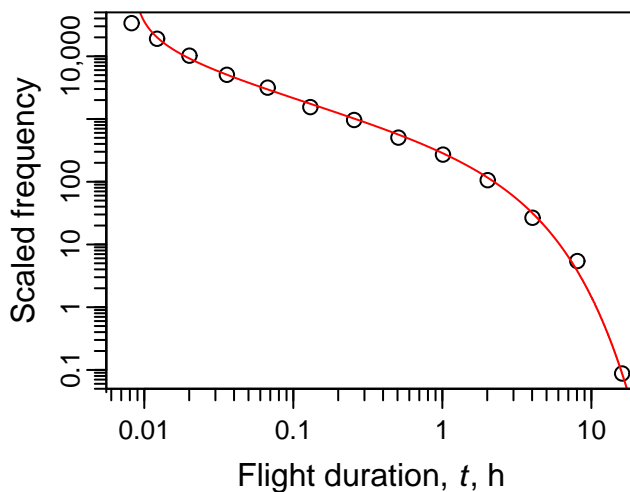
**Supplementary Table 2.** Sensitivity to  $b$  for low-food bees data

Final bin	$b$	MLE for $\mu$	MLE for $\lambda$	Akaike weights	
		(95% CI)	(95% CI)	Power law	Exponential
		$G, p$	$G, p$		
25	937.5	2.20 (1.84, 2.60) 11.3, 0.19	0.00609 (0.00497, 0.00732) 13.1, 0.11	0.78	0.22
26	975	2.23 (1.88, 2.61) 11.2, 0.19	0.00614 (0.00503, 0.00736) 12.8, 0.12	0.68	0.32
27	1012.5	2.26 (1.91, 2.63) 11.2, 0.19	0.00618 (0.00508, 0.00739) 12.5, 0.13	0.58	0.42
none	$\infty$	2.66 (2.39, 2.96) 14.3, 0.07	0.00635 (0.00531, 0.00751) 11.7, 0.16	0.001	0.999

**Supplementary Table 2. Sensitivity to  $b$  for bumblebee data in low-food scenario.** A final bin of 25 was used in Table 1, but adding just one or two extra bins of zero count (bins 26 and 27) shows that the weights are sensitive to  $b$ , and that no model is conclusively preferred. Allowing an infinite tail overwhelmingly favours the exponential. For all goodness-of-fit tests,  $n = 129$  and  $df = 8$ .

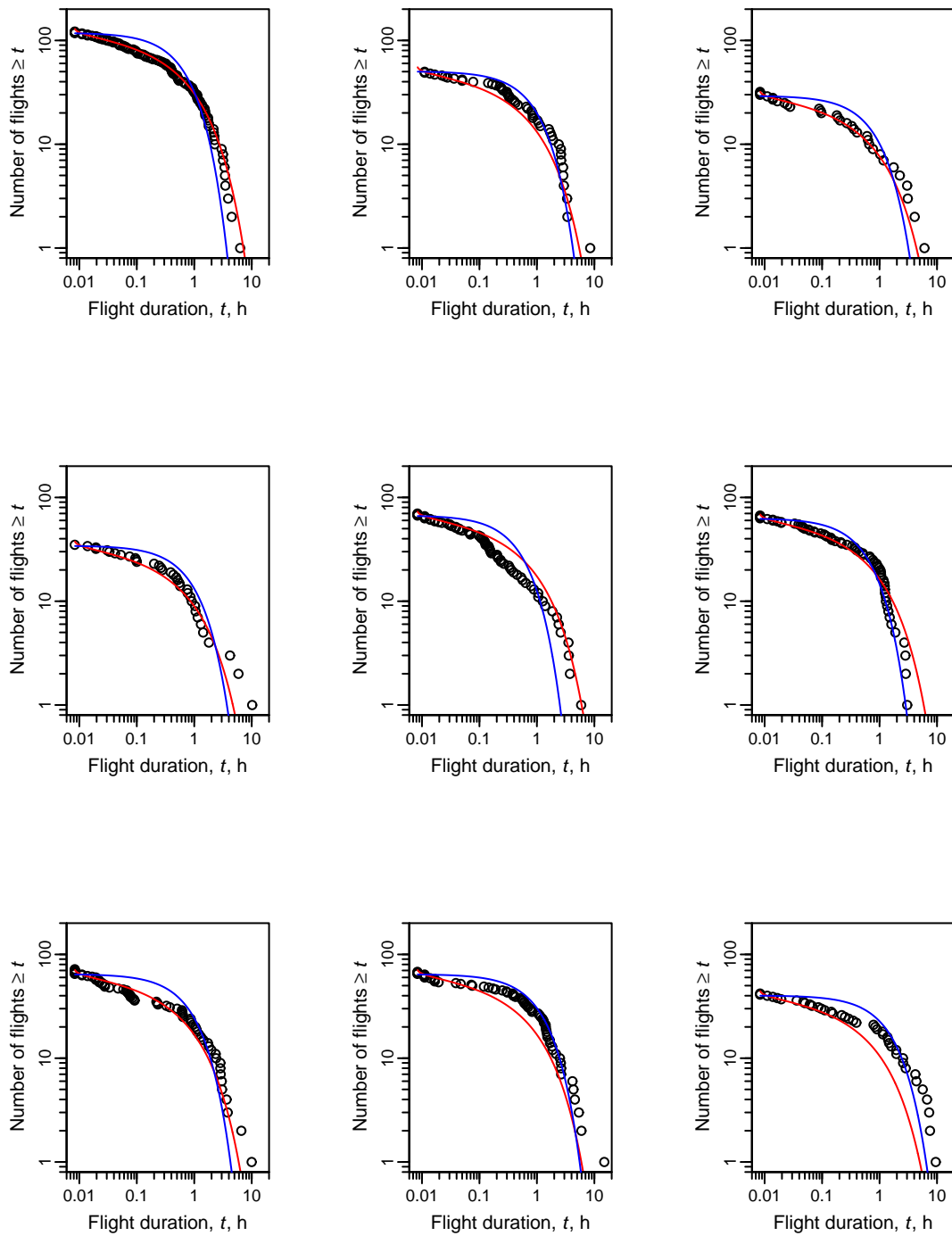


**Supplementary Figure 1. Original log-log histogram of 1992 wandering albatross flight durations (Fig. 3a of ref. 7) computed here from the original raw data.** Breakpoints of bins are at 1, 2, 4, 8, 16, 32, 64 and 128 h (with bin intervals  $1 \leq t < 2$ ,  $2 \leq t < 4$ , etc.), and results are plotted at the geometric means. The frequencies are each normalized by their respective bin widths to yield frequency densities that compensate for the increasing bin widths<sup>30</sup> (termed logarithmic binning with normalization, LBN, in ref. 24). The straight line indicates a power law of exponent  $\mu = 2$  (ref. 7). This is the original figure that then gets corrected in Fig. 3a.

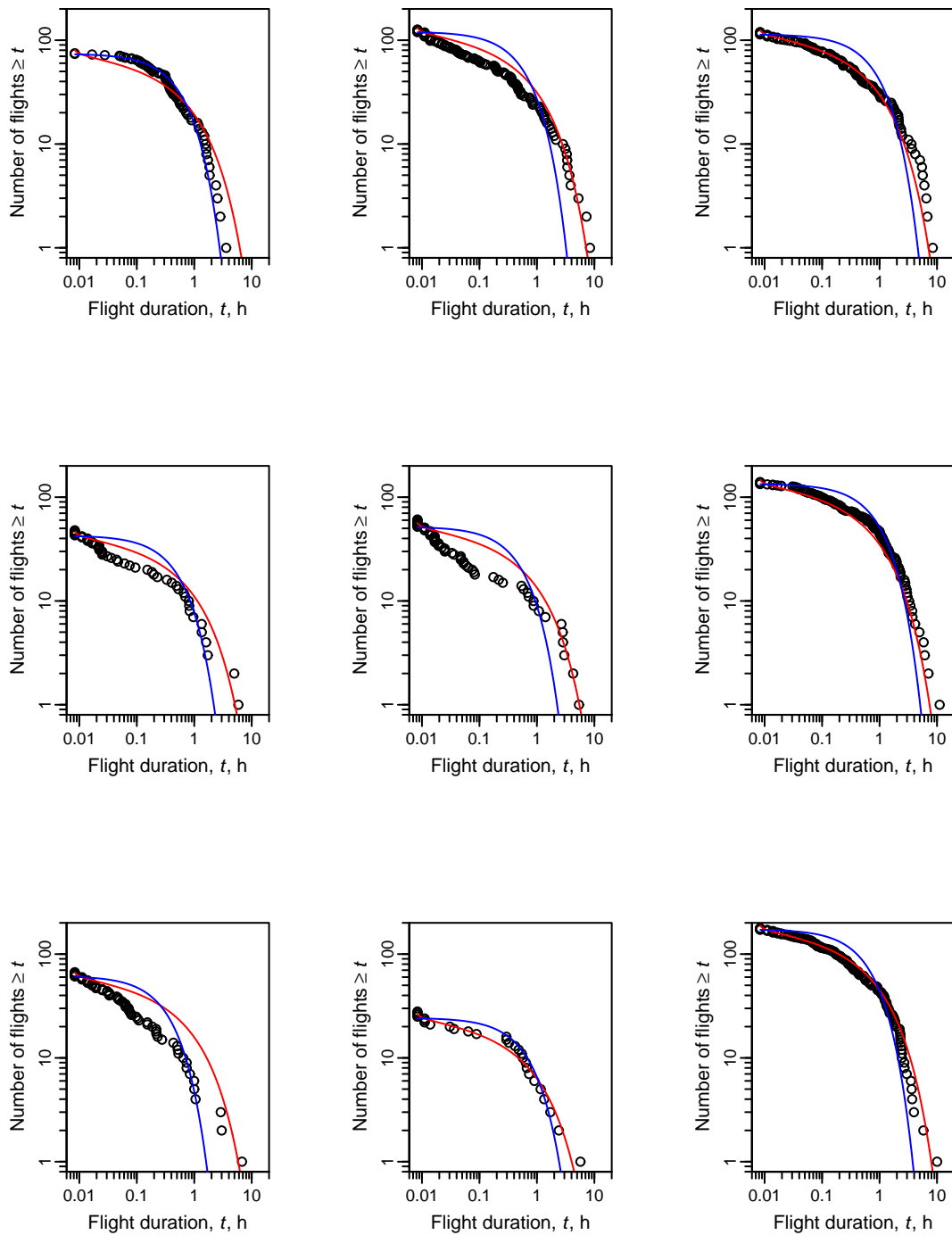


**Supplementary Figure 2. The 2004 data as log-log histogram (LBN method).**

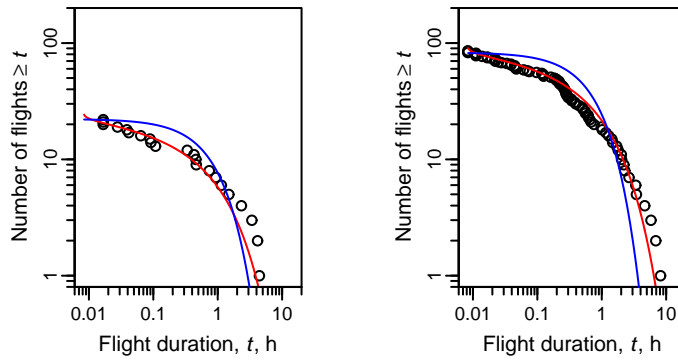
Red curve is the MLE fitted gamma distribution (2), shown in main text to be a good fit ( $p = 0.83$ ). The lowest bin contains just the records of 30 s, which were not used in the likelihood maximization (because they will include some true dry durations of 20-30 s, which are not considered flights, as outlined in Supplementary Methods 1). The large values of scaled frequency occur because the counts are scaled by the bin widths, which are small (in hours) for the shorter flights.



**Supplementary Figure 3. Rank/frequency plot for individual birds 1-9 from 2004.** Red curves are number of flights predicted from the MLE gamma distribution (2) as fitted to the pooled data (i.e. as in Fig. 1), scaled to the total number of flights for each bird. Blue curves are the MLE exponential fit individually to each bird. The individual data sets seem more consistent with coming from the pooled gamma distribution than from individual exponential distributions, suggesting

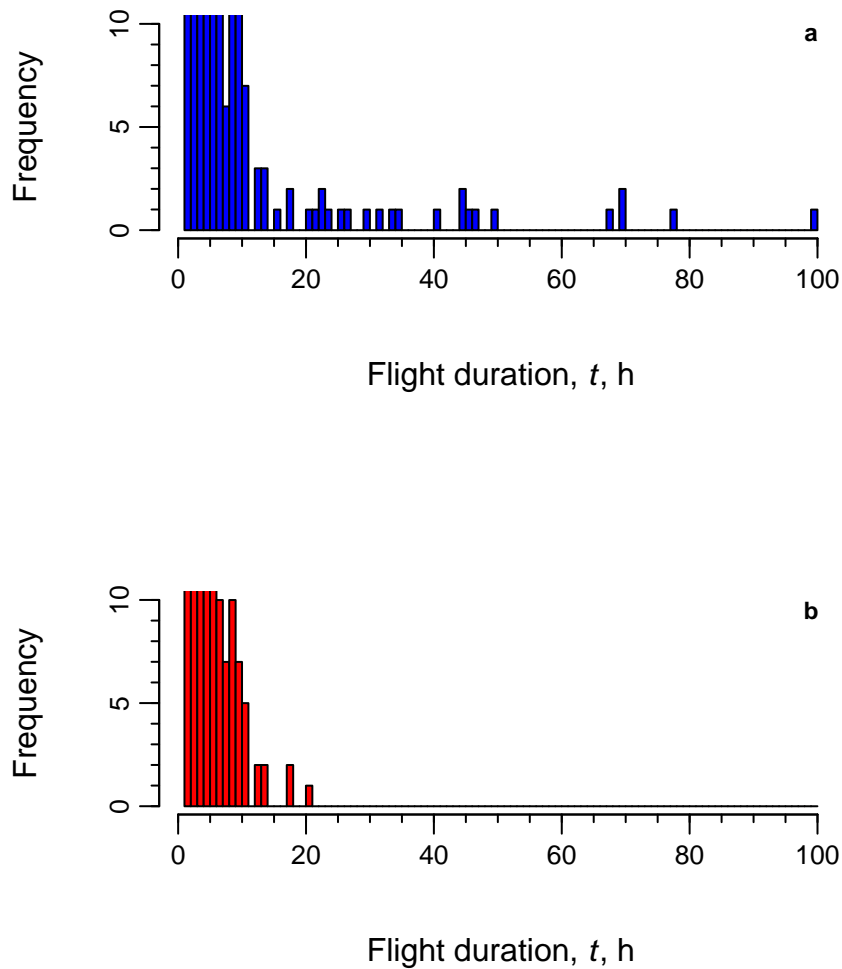


**Supplementary Figure 4. Rank/frequency plot for individual birds 10-18 from 2004.** As Supplementary Figure 3, for the next 9 birds.



**Supplementary Figure 5. Rank/frequency plot for individual birds 19-20 from 2004.** As Supplementary Figures 3 and 4, for the remaining 2 birds.





**Supplementary Figure 6. The 1992 data as in Fig. 3a, but unbinned and on linear axes. a** Raw logger data before adjusting for time spent on Bird Island. Values were binned to obtain the blue circles in Fig. 3a. Frequency axis is truncated at 10 to clearly show the individual counts in the tail of the histogram, and a 15 h dry sequence, for example, is counted in the 15-16 h bin. **b** As **a**, but after adjusting the first and last dry sequences. Values were binned to obtain the red circles in Fig. 3a. It is clear that the adjustment of first and last dry sequences removes the longest ones.