Distinguishing between Lévy walks and strong alternative models: comment

Marie Auger-Méthé,[1] Michael J. Plank,[2] and Edward A. Codling[3,4,5]

Reynolds (2012) proposed that power spectra and the first-significant-digit law of Benford (1938) can be used to distinguish between movement data arising from a Lévy walk (LW) and data from a strong alternative model: the composite correlated random walk (CCRW). Under the CCRW model, the animal typically switches between two behavioral phases, one with a relatively large mean step length, representing movement between patches of food for example, and one with a smaller mean step length, representing movement within a patch (Benhamou 2007). These are sometimes referred to as extensive and intensive movement phases (Knell and Codling 2012). The results of Reynolds (2012) are important as CCRWs can produce movement patterns that appear similar to LWs, even though the underlying processes are quite different (Benhamou 2007, Plank and Codling 2009, Codling and Plank 2011, Gautestad 2012). In this comment, we show that the results of Reynolds (2012) are not robust to changes in the parameters of the CCRW model and that the proposed methods cannot always reliably distinguish between different movement models. We further discuss a number of methodological points related to the study of Reynolds (2012) that should be considered when testing movement models.

Reynolds (2012) used simulations to generate step lengths for LWs and for CCRWs. Similarly to others (Plank and Codling 2009, Codling and Plank 2011), he showed that simulated CCRW data can be misclassified as arising from a power law (i.e., as being from a LW)

using standard comparative tests based on Akaike weights (Edwards 2008). In addition, Reynolds (2012) claimed that a test of absolute fit (*G* test), which has been suggested as a method of resolving this issue (Auger-Méthé et al. 2011), does not correct this misidentification. Reynolds (2012) proposed two alternative approaches for distinguishing between a LW and a CCRW, based on (1) the power spectrum of the time series of turning points in the movement path and (2) Benford's law of first significant digits.

Power spectra

In the simulations of Reynolds (2012: Supplement), step lengths for the LW were drawn from a discrete truncated Pareto (also referred to as a discrete truncated power law) distribution with minimum step length $x_{min} = 1$, maximum step length $x_{max} = 1 \times 10^5$, and exponent $\mu = 2$. Reynolds (2012) did not generate full movement paths for the CCRW, which would require turning angles and transitions between states to be simulated (e.g., Plank and Codling 2009, Codling and Plank 2011). Instead, Reynolds (2012) only considered the step lengths that would be generated by this type of CCRW. Each step was randomly assigned as being in the intensive phase (e.g., with probability $p_I = 0.9$) or the extensive phase (with probability $p_E = 1 - p_I$), and the step length was then drawn from an exponential distribution with (for example) mean $s_I = 15$ for the intensive phase and $s_E = 74$ for the extensive phase. The theoretically expected spectrum of a LW is $S(f) \sim f^{-\beta}$ (where f is frequency) with $\beta = 3 - \mu$ (Viswanathan et al. 1996). On log-log axes, this power spectrum is linear with slope $-\beta$ and, for a LW with $\mu \approx 2$, the expected value of β is approximately 1. Reynolds (2012) claims that a CCRW will always produce noise close to being "white" ($\beta \approx 0$).

We repeated the simulations in Fig. 2 of Reynolds (2012) and obtained similar results: the power spectrum of the turning point time series for the CCRW with these parameter values is very different from that of the LW (Fig. 1a, b). The LW power spectrum appears approximately linear on log-log axes with a slope of $\beta \approx 0.63$, while the CCRW appears relatively flat with $\beta \approx 0$. However, for other choices of parameter values for the CCRW, the difference between the power spectra of the CCRW and of the LW is much less pronounced. As an example, consider the CCRW with $p_I = 0.9$, $s_I = 1.2$, $s_E = 12$. This CCRW has a greater relative difference between the mean step lengths of the two phases (s_E/s_I) than that considered by Reynolds (2012). Like the example of Reynolds (2012), this CCRW would be misidentified as a LW with $\mu \approx 2$ using Akaike weights and maximum-likelihood estimation (MLE). However, the power
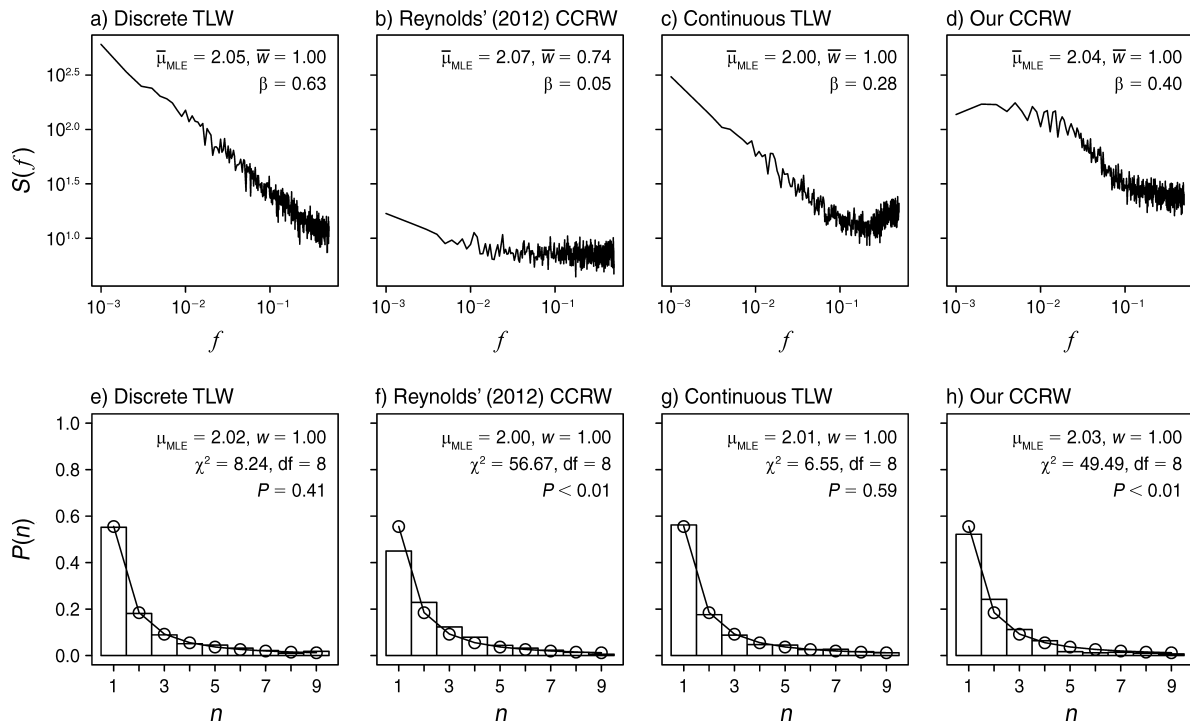
Fig. 1.   Example simulation output for (a, e) a discrete truncated Lévy walk (TLW) with exponent $\mu = 2$, minimum step length $x_{min} = 1$, maximum step length $x_{max} = 100\,000$, minimum step length cutoff $a = 10$; (b, f) a composite correlated random walk (CCRW) with the parameter values considered by Reynolds (2012), intensive phase probability $p_I = 0.9$, intensive phase step length $s_I = 15$, extensive phase step length $s_E = 74$, $x_{min} = 0$, $a = 10$; (c, g) a continuous TLW with $\mu = 2$, $x_{min} = 1$, $x_{max} = 100\,000$, $a = 1$; (d, h) a CCRW with $p_I = 0.9$, $s_I = 1.2$, $s_E = 12$, $x_{min} = 0$, $a = 1$. Panels a–d show the average power spectrum ($S(f)$, where f is frequency) of an ensemble of 50 movement paths, each having a turning point time series $u(t)$ of length 1000. For consistency with Reynolds (2012), we show the results for the first 500 frequencies. The exponent β was estimated using linear regression on log-log axes. The mean maximum-likelihood estimation (MLE) value of μ, $\bar{\mu}_{MLE}$, and the mean Akaike weight for the power law, $\bar{w}$, for the each set of simulations are presented. Panels e–h show the first-significant-digit distribution $P(n)$, the relative frequency of the 1000 sampled step lengths whose first significant digit is n; rectangles represent observed frequencies and dots represent expected frequencies for a LW with $\mu = 2$. Results of a χ^2 test comparing these two distributions are shown. The MLE value of μ and the Akaike weight for the power law, w, are presented for each simulation.

spectrum for this CCRW is not flat and, at least for some frequencies, is approximately linear (Fig. 1d).

Reynolds (2012) did not give a criterion to assess whether a given power spectrum is significantly different from linear, nor provide an algorithm to estimate β. The method therefore relies on a subjective assessment of whether the power spectrum is approximately linear and whether the observed slope β is significantly different from the expected slope of $3 - \mu$. In addition, a power spectrum from a real-world data set can only be linear over a finite frequency range. In such cases, where the model parameters are unknown, it is not clear how to define the frequency range over which the power spectrum should be assessed. For the power-spectrum method to be useful, an objective, quantitative method for resolving these issues needs to be defined, and the resulting rejection rate for data from LW and from CCRW tested.

To illustrate that this is not a trivial exercise, we estimated β using simple linear regression on log-log

axes, for a range of parameter values for the CCRW. We restricted attention to CCRW where the animal spends more time in the intensive phase than the extensive phase (i.e., $p_I > 0.5$) because this region of parameter space (1) is more biologically relevant and (2) contains most of the cases where the CCRW would be mistaken for a LW with $\mu \approx 2$ using MLE-based tests (Fig. 2a, b). For most of these cases, β is estimated to be much greater than 0 and, in some cases, close to the value of 0.63 estimated for a $\mu = 2$ LW (Fig. 2d). We acknowledge that simple linear regression over a fixed frequency range is an extremely crude method of characterizing the power spectrum, especially if the data are nonlinear. Nevertheless, our results highlight the potential difficulties in correctly assigning movement data to candidate models using power spectra and demonstrate that a more sophisticated criterion is needed.
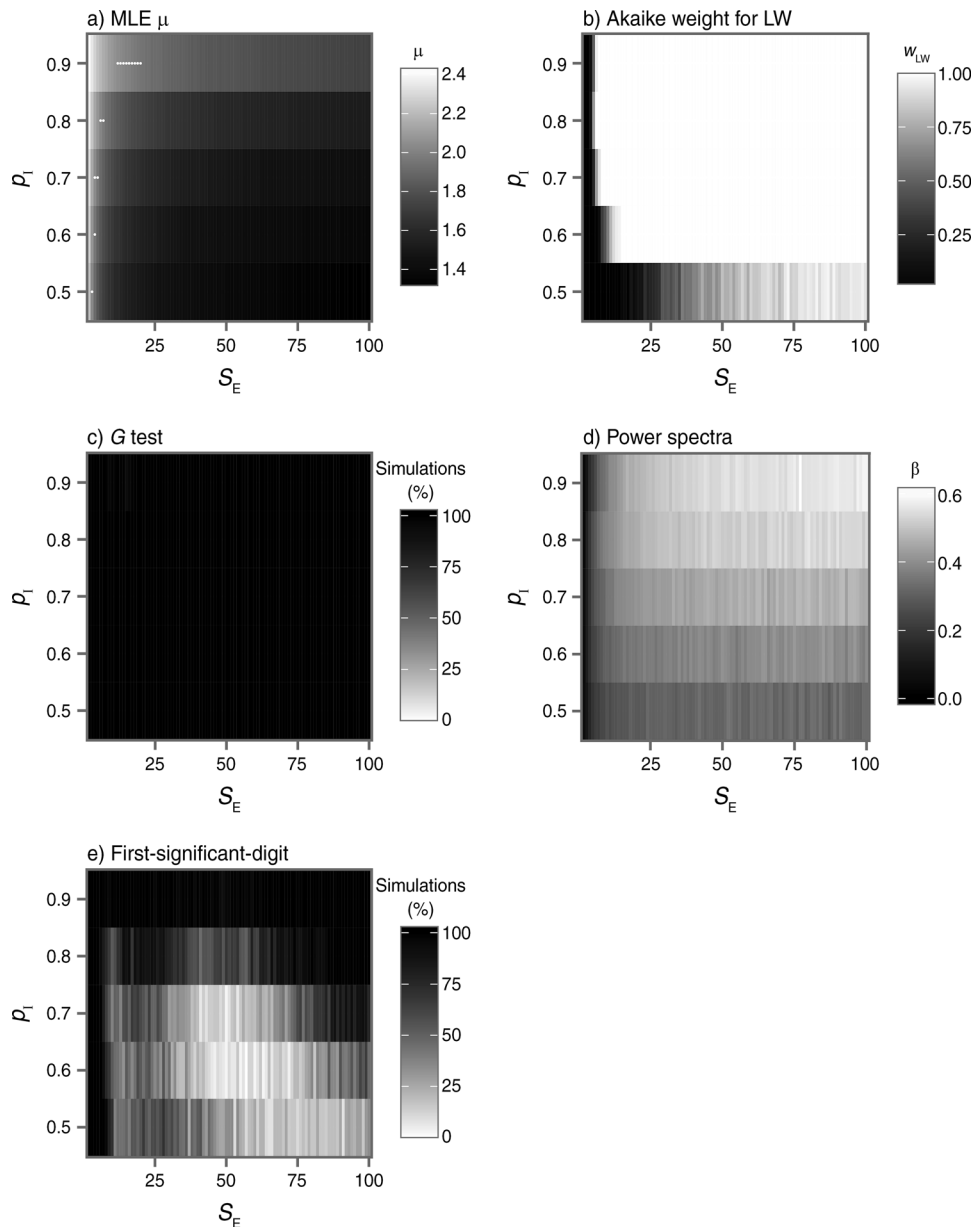
Fig. 2. Test results for CCRWs simulated as in Reynolds (2012) across a range of values of the parameters s_E and $p_I = 1 - p_E$ with $s_I = 1.2$ and $x_{min} = 0$: (a) the mean MLE value of μ, white dots represent $1.95 \leq \mu \leq 2.05$; (b) the mean Akaike weight for the power law, with 1 indicating strong support for the LW; (c) the percentage of simulations for which the G test indicated that the step length distribution was significantly different ($P < 0.05$) from that of a LW with $\mu = \mu_{MLE}$; (d) power-spectrum exponent β, estimated using linear regression on log-log axes of the ensemble average power spectrum when using the first 500 frequencies; (e) the percentage of simulations whose first-significant-digit distribution is significantly different ($P < 0.05$) from that of a LW with $\mu = \mu_{MLE}$. For panels (b–e), the lighter the shading, the more closely the CCRW resembles a LW under the proposed test. For the power-spectrum method, the $u(t)$ time series was of length 1000; for all other methods, the step length sample size was 1000 steps (after removing steps smaller than the cutoff $a = 1$). All results show the average of 50 replicate simulations.

Benford's first-significant-digit law

The first-significant-digit method is more robust than the power-spectrum method. Both the example CCRWs considered above have first-significant-digit distribu-

tions that are significantly different ($P < 0.05$) from a LW with $\mu = 2$ (Fig. 1f, h). Fig. 2e shows the proportion of simulated CCRWs whose first-significant-digit distribution is significantly different ($P < 0.05$) from that of a
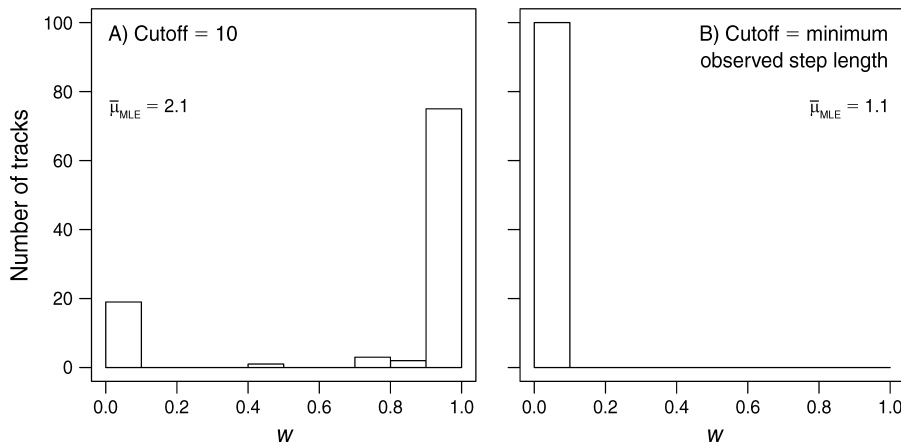
FIG. 3. Histogram of the Akaike weights value for the LW for 100 CCRW simulations with $s_I = 15$, $s_E = 74$, $p_E = 0.9$ (a) with a cutoff of $a = 10$ and (b) with a cutoff equal to the minimum observed step length (mean observed $x_{min} = 0.015$). An Akaike weight of 1 indicates strong support for the LW. The mean μ_{MLE} ($\bar{\mu}_{MLE}$) for each set of simulations is displayed. Note that applying the cutoff effectively removes step lengths from the simulation. We ensured that the comparison was made on step length distributions of the same sample size by simulating enough steps to have 1000 remaining after the cutoff was applied.

LW with $\mu = \mu_{MLE}$ (where μ_{MLE} is the maximum-likelihood estimate of the LW exponent) across a range of parameter values for the CCRW. This confirms that this method rarely misidentifies CCRWs for LW with $1.95 < \mu < 2.05$ (indicated by the white dots in Fig. 2a).

Although the optimum exponent in simple search models is $\mu \approx 2$ (Viswanathan et al. 1999), theoretical mechanisms have been proposed for LW with different values of μ (Reynolds 2010a, b, 2013). Also, some of the studies providing empirical evidence for LW have estimated exponents outside the range 1.95–2.05 (e.g., Sims et al. 2008, Humphries et al. 2010, 2012). It is therefore relevant to investigate CCRWs with a wider range of μ_{MLE}. Widening the range only slightly to $1.8 < \mu_{MLE} < 2.2$ increases the potential misidentification of CCRWs for LWs using this test. Extending the analysis to all CCRW that are classified by Akaike weights as a LW with $1 < \mu_{MLE} < 3$ shows that a broad range of parameter combinations leads the first-significant-digit method to misidentify CCRWs for a LW (Fig. 2e). In contrast, the G test, a test of absolute fit, shows that, across the same parameter range, almost all the CCRW samples have a step-length distribution that is significantly different ($P < 0.05$) from that of a LW with $\mu = \mu_{MLE}$ (Fig. 2c).

The misidentification of CCRW for LW can be partly mitigated by increasing the number of random walk steps in the simulation (the results in Fig. 2e are for 1000 steps per simulation). However, with empirical data, the sample size is usually limited by practical considerations. Even if the sample size can be increased, this does not guarantee a better fit because, unlike simulated data that are drawn from simple statistical distributions, empirical data are typically subject to substantial sources of variation that are not included in the model.

Simulations and data processing

Akaike weights and the value of μ_{MLE} depend on the chosen value of the minimum step length cutoff, a. This cutoff represents the start of the tail of the power-law distribution and all observed step lengths smaller than the cutoff are removed from the analysis. Some authors advocate that the complete distribution should be explored, in which case the cutoff value is set to be the minimum observed step length (Edwards et al. 2012). Others advocate that only part of the distribution should be explored (Clauset et al. 2009), in which case the choice of this cutoff value is either arbitrary, as mentioned by Reynolds (2012), or relies on methods that identify the part of the empirical distribution that most closely resembles a power law (Clauset et al. 2009). Fig. 3 shows that, to reproduce Fig. 1 of Reynolds (2012), it is necessary to use a cutoff value much greater than the minimum observed step length. The choice of cutoff can therefore impact the level of evidence for a LW with $\mu \approx 2$. Again, this highlights the need to use an objective method for selecting the cutoff based on the observed data alone (e.g., Edwards 2008, Clauset et al. 2009), not on the model parameters, which would in reality be unknown.

Reynolds (2012: Supplement) simulated discrete LWs, which allow only integer step lengths, rather than the more commonly used continuous LWs, which allow a continuous range of step lengths. Many animals do not necessarily make turns at discrete time intervals and would, therefore, be better described by a continuous LW than by a discrete LW. The power spectrum of a continuous LW (Fig. 1c) visually resembles the power spectrum of a CCRW (Fig. 1d) and has a comparable estimate for β. The value of β for the continuous LW is

affected by the points in the nonlinear, high-frequency, end of the spectrum, a problem that could be avoided by analyzing the spectrum over a different frequency range. Nevertheless, these results emphasize the importance of not relying on subjective comparisons and of the need for a robust method for quantifying how different an observed power spectrum is from the expected power spectrum for a LW.

Conclusion

We investigated whether the two methods proposed by Reynolds (2012) can be used to distinguish between a LW and a strong alternative model (a CCRW). The power-spectrum method lacks an objective criterion to identify whether a movement path is consistent with a LW. Visually assessing the fit of a linear model to data on log-log axes is notoriously inaccurate (Stumpf and Porter 2012) and using basic regression metrics such as r^2 can also be highly misleading (Edwards 2008, Clauset et al. 2009). Our results show that, without an objective criterion, it is not clear how to consistently differentiate the power spectrum of a CCRW from that of a LW with $\mu = 2$.

In contrast to the power-spectrum method, the first-significant-digit method is based on a well defined statistical hypothesis test. This test can differentiate a CCRW from a LW when attention is restricted to CCRW whose MLE exponent lies in the narrow range 1.95 to 2.05. However, including CCRW whose MLE exponent lies in a wider range reveals that the first-significant-digit method becomes less reliable. These results may also be sensitive to the subjective choice of cutoff value, as changing this value alters the results of the Akaike test and the maximum-likelihood estimate for the LW exponent μ.

Following Reynolds (2012), we have investigated the rate of type I error of the power-spectrum and first-significant-digit methods, i.e., the proportion of samples from the CCRW model that the methods fails to distinguish from a LW. An effective test should also have a low rate of type II error, i.e., it should not reject too many samples that are generated from a LW model. Placing tight restrictions on, for example, the MLE exponent of the data reduces the type I error rate, as we have seen in the first-significant-digit method. However, this is likely to come at the expense of a higher type II error rate. Until objective criteria for determining whether a given sample is consistent with a LW are provided, and the rates of type I and type II error are shown to be acceptable, the results of these methods should be treated with caution.

All of our analyses were done in R (R Development Core Team 2012) and the code used to create all figures can be found in the Supplement.

Literature cited

Auger-Méthé, M., C. C. St. Clair, M. A. Lewis, and A. E. Derocher. 2011. Sampling rate and misidentification of Lévy and non-Lévy movement paths: comment. Ecology 92:1699–1701.

Benford, F. 1938. The law of anomalous numbers. Proceedings of the American Philosophical Society 78:551–572.

Benhamou, S. 2007. How many animals really do the Lévy walk? Ecology 88:1962–1969.

Clauset, A., C. R. Shalizi, and M. E. J. Newman. 2009. Power-law distributions in empirical data. SIAM Review 51:661–703.

Codling, E. A., and M. J. Plank. 2011. Turn designation, sampling rate and the misidentification of power-laws in movement path data using maximum likelihood estimates. Theoretical Ecology 4:397–406.

Edwards, A. M. 2008. Using likelihood to test for Lévy flight search patterns and for general power-law distributions in nature. Journal of Animal Ecology 77:1212–1222.

Edwards, A. M., M. P. Freeman, G. A. Breed, and I. D. Jonsen. 2012. Incorrect likelihood methods were used to infer scaling laws of marine predator search behaviour. PLoS ONE 7(10):e45174.

Gautestad, A. O. 2012. Animal space use: distinguishing a two-level superposition of scale-specific walks from scale-free Lévy walk. Oikos 122:612–620.

Humphries, N. E., N. Queiroz, J. R. M. Dyer, N. G. Pade, M. K. Musyl, K. M. Schaefer, D. W. Fuller, J. M. Brunnschweiler, T. K. Doyle, and J. D. R. Houghton. 2010. Environmental context explains Lévy and Brownian movement patterns of marine predators. Nature 465:1066–1069.

Humphries, N. E., H. Weimerskirch, N. Queiroz, D. J. Southall, and D. W. Sims. 2012. Foraging success of biological Lévy flights recorded in situ. Proceedings of the National Academy of Sciences USA 109:7169–7174.

Knell, A. S., and E. A. Codling. 2012. Classifying area-restricted search (ARS) using a partial sum approach. Theoretical Ecology 5:325–339.

Plank, M. J., and E. A. Codling. 2009. Sampling rate and misidentification of Lévy and non-Lévy movement paths. Ecology 90:3546–3553.

R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Reynolds, A. M. 2010a. Animals that randomly reorient at cues left by correlated random walkers do the Lévy walk. American Naturalist 175:607–613.

Reynolds, A. M. 2010b. Bridging the gulf between correlated random walks and Lévy walks: autocorrelation as a source of Lévy walk movement patterns. Journal of the Royal Society Interface 7:1753–1758.

Reynolds, A. M. 2012. Distinguishing between Lévy walks and strong alternative models. Ecology 93:1228–1233.

Reynolds, A. M. 2013. Beating the odds in the aerial lottery: passive dispersers select conditions at takeoff that maximize their expected fitness on landing. American Naturalist 181:555–561.

Sims, D. W., et al. 2008. Scaling laws of marine predator search behaviour. Nature 451:1098–1103.

Stumpf, M. P. H., and M. A. Porter. 2012. Critical truths about power laws. Science 335:665–666.

Viswanathan, G. M., V. Afanasyev, S. V. Buldyrev, E. J. Murphy, P. A. Prince, and H. E. Stanley. 1996. Lévy flight search patterns of wandering albatrosses. Nature 381:413–415.

Viswanathan, G. M., S. Buldyrev, S. Havlin, M. G. E. da Luz, E. P. Raposo, and H. E. Stanley. 1999. Optimizing the success of random searches. Nature 401:911–914.

Distinguishing between Lévy walks and strong alternative models: reply

ANDY M. REYNOLDS[1]

At the time of writing the *Note* (Reynolds 2012), composite correlated random walks (CCRW) were regarded as being competing models of movement patterns that resemble Lévy walks (LW), and even as potential replacements for LW (Benhamou 2007). Now CCRW are seen more as mechanistic models pointing at a way by which organisms could approximate LW, as evidenced in recent theoretical and empirical studies of mussels (*Mytilus edulis*) and the Australian desert ant *Melophorus bagoti* (de Jager et al. 2011, 2012, Jansen et al. 2012, Reynolds 2013, Reynolds et al. 2014; A. M. Reynolds, *unpublished manuscript*). This shift in emphasis addresses why CCRW could resemble LW and why differentiating between these two patterns can be so difficult. The required "fine-tuning" of the suite of parameters in a CCRW could be brought about by selection pressures for LW, in accordance with the LW foraging hypothesis (Viswanathan et al. 2008). Tuning is required because most CCRW do not resemble LW and fine-tuning is required if a CCRW is to resemble the particular kind of LW associated with optimal foraging. Rather than distinguishing between LW and strongly competing models, the tools that I advocated in my *Note* (Reynolds 2012)—power spectrum and significant first analyses—can thus help identify underlying processes giving rise to LW movement patterns. This is a significant change because, with the surge in empirical

support for LW (Korobkova et al. 2004, Reynolds et al. 2007*a*, *b*, Sims et al. 2008, Humphries et al. 2010, de Jager et al. 2011, Harris et al. 2012, Hays et al. 2012), the debate has now shifted from whether some organisms perform LW to when and why organisms perform LW (Ornes 2013). The critical reevaluation of the analysis tools by Auger-Méthé et al. (2014) is therefore timely. Auger-Méthé et al. (2014) suggested that the tools are not as reliable as previously thought. In particular, they note (1) that LW and CCRW do have different power spectra but that it is not clear how this difference can be used to distinguish between LW and CCRW, and (2) that significant first-digit analyses can reliably distinguish between LW and CCRW when the Lévy (power-law) exponent is close to 2 (1.95–2.05), but are less reliably when the Lévy exponent is not close to 2.

I address both of these criticisms and show that for the most part they stem from misunderstandings and misapplications of the tools.

Power spectra

LW are characterized by power spectra, S, that exhibit power-law scaling, $S \propto f^{-\beta}$, over an extended range of frequencies, f (Viswanathan et al. 1996). The scaling exponent β approaches $3 - \mu$ asymptotically for sufficiently long LW. In Reynolds (2012), I stated that this "is quite unlike the power spectra that characterize CCRW, which are always close to being 'white' ($\beta \approx 0.0$)" and that, as a consequence, power spectra can distinguish between LW and CCRW. Nonetheless, Auger-Méthé et al. (2014) rightly point out that I did not provide an algorithm for β and that power spectra for real-world data set can only exhibit power-law scaling over a finite frequency range. They add, that "in such cases, where the model parameters are unknown, it is not clear how to define the frequency range over which the power spectrum should be analyzed." The simplest and most used algorithm is linear regression on log-log scales. That is, the slope of a plot of $\log(S)/\log(t)$ (where t is time) should be $-\beta$ This is, in effect, the algorithm that Auger-Méthé et al. (2014) used; they report that the spectrum for a CCRW, which resembles a LW with Lévy exponent $\mu = 2$ (Fig. 1a), is not flat but is instead curvilinear around about one decade. They suggest that, as a consequence, the spectrum of a CCRW can be confounded with the spectrum of a LW. This is not