Technical appendices for “Disentangling social interactions and environmental drivers in multi-individual wildlife tracking data”

Justin M. Calabrese[1, *], Chris H. Fleming[1, 2], William F. Fagan[2], Martin Rimmler[3], Petra Kaczensky[4], Sharon Bewick[2], Peter Leimgruber[1], and Thomas Mueller[5, 6]

[1]Smithsonian Conservation Biology Institute, Front Royal, VA , USA
[2]Dept. of Biology, University of Maryland, College Park, MD, USA
[3]Dept. of Biology, University of Stuttgart
[4]Research Institute of Wildlife Ecology, Vienna, Austria
[5]Biodiversity and Climate Research Centre, Frankfurt, Germany
[6]Dept. of Biological Sciences, University Frankfurt, Germany
[*]Corresponding author; Email: CalabreseJ@si.edu

1 Maximum-likelihood analysis

The Gaussian log-likelihood function for S steps of our models is given by

$$\ell(\mathbf{r}|\boldsymbol{\nu},\mathbf{D}) = C - S\,\mathrm{tr}\log\mathbf{D} - \frac{1}{2}\sum_{i=1}^{S}\sum_{\substack{n=1\\m=1}}^{N}\left[\Delta\mathbf{r}_n(t_i)-\boldsymbol{\nu}\Delta t_i\right]^{\mathrm{T}}\frac{[\![\mathbf{D}^{-1}]\!]_{nm}}{\Delta t_i}\left[\Delta\mathbf{r}_m(t_i)-\boldsymbol{\nu}\Delta t_i\right]$$

$$C = -N\sum_{i=1}^{S}\log(2\pi\Delta t_i)\,, \tag{1}$$

where T denotes the matrix transpose and the double square brackets $[\![\cdots]\!]_{nm}$ denote that we take entry (n,m) of the inverse matrix $\mathbf{D}^{-1}$ and not the inverse of the entry. In the following derivations we will use the fact that $\mathbf{D}$ is circulant. Circulant matrices can be diagonalized with a discrete Fourier transform. In this case, the eigenvector representing aggregate motion $\boldsymbol{\phi}_0$ and corresponding eigenvalue representing bulk diffusion λ_0 are given by

$$\boldsymbol{\phi}_0 = \frac{1}{\sqrt{N}}(1,1,1,\cdots)\,, \qquad\qquad \lambda_0 = \sigma + (N-1)\rho\,, \tag{2}$$

and all other eigenvectors (deviations from the aggregate) are degenerate with common eigenvalue

$$\lambda_+ = \sigma - \rho\,. \tag{3}$$

The diffusion correlation parameter ρ is therefore constrained

$$-\frac{\sigma}{N-1} < \rho < \sigma\,. \tag{4}$$

If individuals are perfectly correlated with $\rho = \sigma$, then there is only aggregate diffusion and no individual deviations from the aggregate. On the other hand, if individuals are maximally anti-correlated with $\rho = -\sigma/(N-1)$, then there is no aggregate diffusion and only individual deviations.

With the eigensystem solved, the spectral decomposition of $\mathbf{D}$ can be expressed

$$\mathbf{D} = \lambda_0\,\boldsymbol{\phi}_0\,\boldsymbol{\phi}_0^{\mathrm{T}} + \lambda_+\left(\mathbf{I}_N - \boldsymbol{\phi}_0\,\boldsymbol{\phi}_0^{\mathrm{T}}\right), \tag{5}$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix. Matrix inversion and logarithm operations can then be applied directly to the eigenvalues, so that we obtain the useful relations

$$\operatorname{tr}\log\mathbf{D} = \log\lambda_0 + (N-1)\log\lambda_+\,, \qquad \mathbf{D}^{-1} = \lambda_0^{-1}\,\boldsymbol{\phi}_0\,\boldsymbol{\phi}_0^{\mathrm{T}} + \lambda_+^{-1}\left(\mathbf{I}_N - \boldsymbol{\phi}_0\,\boldsymbol{\phi}_0^{\mathrm{T}}\right). \tag{6}$$

1.0.1 Uncorrelated diffusion

In the case of uncorrelated diffusion, where $\rho = 0$, we have the simplified relations

$$\lambda_0 = \lambda_+ = \sigma\,, \qquad \operatorname{tr}\log\mathbf{D} = N\log\sigma\,, \qquad \mathbf{D}^{-1} = \sigma^{-1}\,\mathbf{I}_N\,. \tag{7}$$

1.1 ML & MVU drift

Maximizing the likelihood with respect to uniform drift vector $\boldsymbol{\nu}$, after some simplification we obtain the expression

$$\hat{\boldsymbol{\nu}} = \frac{\sum_{n=1}^{N}\sum_{m=1}^{N}[\![\mathbf{D}^{-1}]\!]_{nm}\sum_{i=1}^{S}\Delta\mathbf{r}_m(t_i)}{\sum_{n=1}^{N}\sum_{m=1}^{N}[\![\mathbf{D}^{-1}]\!]_{nm}\sum_{i=1}^{S}\Delta t_i}\,. \tag{8}$$

Next we introduce an important theorem: if the covariance matrix has a uniform eigenvector $\boldsymbol{\phi}_0$ (2), then the weighted mean in eqn. (8) reduces to an unweighted mean. To prove this note that eigenvectors of the covariance matrix are also eigenvectors of its inverse. Furthermore, as $\boldsymbol{\phi}_0 = \boldsymbol{\phi}_0^*$, this eigenvector is both a right and left eigenvector after transposition. Next we can identity the left sums as contractions with the (left) uniform eigenvector

$$\sum_{n=1}^{N} \left[\!\left[\mathbf{D}^{-1} \right]\!\right]_{nm} \propto \sum_{n=1}^{N} \left[\!\left[\boldsymbol{\phi}_0^{\mathrm{T}} \right]\!\right]_{n} \left[\!\left[\mathbf{D}^{-1} \right]\!\right]_{nm} = \lambda_0^{-1} \left[\!\left[\boldsymbol{\phi}_0^{\mathrm{T}} \right]\!\right]_{m} , \tag{9}$$

where λ_0 is the eigenvalue of the uniform eigenvector $\boldsymbol{\phi}_0$. Applying this result to eqn. (8) results in the simplified expression

$$\hat{\boldsymbol{\nu}} = \frac{1}{NT} \sum_{n=1}^{N} \sum_{i=1}^{S} \Delta \mathbf{r}_n(t_i) \,, \qquad\qquad T \equiv \sum_{i=1}^{S} \Delta t_i \,, \tag{10}$$

which is an unweighted mean and independent of the diffusion parameters. Differentiating the log-likelihood function twice, we have the covariance

$$\mathrm{COV}[\hat{\boldsymbol{\nu}}] = \frac{\mathbf{I}_2}{NT \boldsymbol{\phi}_0^{\mathrm{T}} \, \mathbf{D}^{-1} \, \boldsymbol{\phi}_0} = \frac{\lambda_0}{NT} \, \mathbf{I}_2 \,, \tag{11}$$

where $\hat{\nu}_x$ and $\hat{\nu}_y$ are uncorrelated, as indicated by the 2×2 spatial matrix $\mathbf{I}_2$.

1.2 ML diffusion

1.2.1 ML correlated diffusion

Applying the spectral decomposition of $\mathbf{D}$ from eqns (2)-(3), the log-likelihood becomes

$$\begin{aligned} \ell(\mathbf{r}|\boldsymbol{\nu},\boldsymbol{\lambda}) = C - S \log \lambda_0 - S(N-1) \log \lambda_+ - \frac{1}{2\lambda_+} \sum_{i=1}^{S} \frac{1}{\Delta t_i} \sum_{n}^{N} \left| \Delta \mathbf{r}_n(t_i) - \boldsymbol{\nu} \Delta t_i \right|^2 \\ + \frac{1}{2} \left(\frac{1}{\lambda_+} - \frac{1}{\lambda_0} \right) \sum_{i=1}^{S} \frac{1}{\Delta t_i} \frac{1}{N} \left| \sum_{n=1}^{N} \left[\Delta \mathbf{r}_n(t_i) - \boldsymbol{\nu} \Delta t_i \right] \right|^2 . \end{aligned} \tag{12}$$

Next we identify the sufficient statistics $\hat{\boldsymbol{\nu}}$, s, and r, given by

$$s = \frac{1}{2}\frac{1}{NS}\sum_{i=1}^{S}\sum_{n=1}^{N}\frac{|\Delta\mathbf{r}_n(t_i) - \hat{\boldsymbol{\nu}}\Delta t_i|^2}{\Delta t_i}\,, \tag{13}$$

$$r = \frac{1}{2}\frac{1}{S}\sum_{i=1}^{S}\frac{1}{\Delta t_i}\left|\frac{1}{N}\sum_{n=1}^{N}\left[\Delta\mathbf{r}_n(t_i) - \hat{\boldsymbol{\nu}}\Delta t_i\right]\right|^2 . \tag{14}$$

Finally, the mean-profiled log-likelihood function can be expressed

$$\ell(\mathbf{r}|\hat{\boldsymbol{\nu}},\boldsymbol{\lambda}) = C - \frac{n_0}{2}\left(\log\lambda_0 + \frac{\hat{\lambda}_0}{\lambda_0}\right) - \frac{n_+}{2}\left(\log\lambda_+ + \frac{\hat{\lambda}_+}{\lambda_+}\right), \tag{15}$$

where the degrees of freedom and ML eigenvalues are given by

$$n_0 = 2\,S \qquad \hat{\lambda}_0 = N\,r\,, \qquad n_+ = 2\,(N-1)\,S \qquad \hat{\lambda}_+ = \frac{N}{N-1}(s-r)\,. \tag{16}$$

From the inverse relations, the ML correlations are given by

$$\sigma = \frac{\lambda_0 + (N-1)\lambda_+}{N}\,, \qquad \hat{\sigma} = s\,, \tag{17}$$

$$\rho = \frac{\lambda_0 - \lambda_+}{N}\,, \qquad \hat{\rho} = \frac{N\,r - s}{N-1}\,. \tag{18}$$

which is the ordinary variance statistic for the ML σ, but the result is non-trivial for the ML cross-correlation ρ. The MLEs are (asymptotically) independent chi-squared random variables, obeying

$$n_0\frac{\hat{\lambda}_0}{\lambda_0} \sim \chi^2_{n_0}\,, \qquad n_+\frac{\hat{\lambda}_+}{\lambda_+} \sim \chi^2_{n_+}\,, \tag{19}$$

$$\mathrm{VAR}[\hat{\lambda}_0] = \frac{2\,\lambda_0^2}{n_0}\,, \qquad \mathrm{VAR}[\hat{\lambda}_+] = \frac{2\,\lambda_+^2}{n_+}\,, \tag{20}$$

as can be seen by the structure of the likelihood function. Finally, to propagate independent uncertainties in (λ_0, λ_+) into our parameters of interest (σ, ρ), by linear transformation, we can apply the Jacobian

$$\frac{\partial(\sigma,\rho)}{\partial(\lambda_0,\lambda_+)} = \frac{1}{N}\begin{bmatrix}1 & N-1\\ 1 & -1\end{bmatrix}. \tag{21}$$

1.2.2 ML uncorrelated diffusion

In the case of uncorrelated diffusion ($\rho = 0$) we have the mean-profiled log-likelihood function

$$\ell(\mathbf{r}|\hat{\boldsymbol{\nu}}, \sigma) = C - \frac{n}{2}\left(\log\sigma + \frac{\hat{\sigma}}{\sigma}\right), \tag{22}$$

to within a constant, and where

$$n = n_0 + n_+ = 2NS\,, \qquad \hat{\sigma} = s\,. \tag{23}$$

This corresponds to a simpler, univariate chi-squared distribution with n degrees of freedom.

2 Minimum-Variance Unbiased analysis

The drift MLEs are unbiased, normally distributed, and independent of the diffusion estimates, and therefore they are already MVU. In the case of zero drift, the diffusion MLEs are also MVU, whereas more generally there is small-sample-size bias in the diffusion estimates that can be substantial if either the number of individuals or steps are small. Here we construct unbiased estimates for our non-zero drift model with a minimal number of sufficient statistics: $\hat{\boldsymbol{\nu}}, s, r$, which should be complete under the normal distribution. Therefore, according to the Lehmann–Scheffé theorem all of these unbiased estimates should be minimum-variance unbiased estimates (MVU).

2.1 MVU correlated diffusion with uniform drift

To derive unbiased estimates of σ and ρ, we will consider the expectation values of s and r. For the components of s we have

$$\begin{aligned}
\langle s\rangle &= \frac{1}{2}\frac{1}{NS}\sum_{i=1}^{S}\sum_{n=1}^{N}\frac{1}{\Delta t_i}\left\langle\left|\Delta\mathbf{r}_n(t_i) - \frac{\Delta t_i}{NT}\sum_{m=1}^{N}\sum_{j=1}^{S}\Delta\mathbf{r}_m(t_j)\right|^2\right\rangle, & (24)\\
&= \frac{1}{2}\frac{1}{NS}\sum_{i=1}^{S}\sum_{n=1}^{N}\frac{\langle\mathbf{w}_n(t_i)^2\rangle}{\Delta t_i} - \frac{1}{2}\frac{1}{N^2ST}\sum_{\substack{i=1\\j=1}}^{S}\sum_{\substack{n=1\\m=1}}^{N}\left\langle\mathbf{w}_n(t_i)^{\mathrm{T}}\mathbf{w}_m(t_j)\right\rangle, & (25)
\end{aligned}$$

after decomposing the steps into their stochastic and deterministic components, according to eqn. (2) in the main text, whereupon the deterministic contributions cancel. We then have

$$\langle s\rangle = \frac{1}{NS}\sum_{i=1}^{S}\sum_{n=1}^{N} D_{nn} - \frac{1}{N^2ST}\sum_{i=1}^{S}\sum_{\substack{n=1\\m=1}}^{N} D_{nm}\,\Delta t_i, \tag{26}$$

$$= \frac{1}{N}\sum_{n=1}^{N}\sigma - \frac{1}{N^2S}\sum_{n=1}^{N}\sigma - \frac{1}{N^2S}\sum_{n\neq m}^{N}\rho\,, \tag{27}$$

$$= \frac{NS-1}{NS}\sigma - \frac{N-1}{NS}\rho\,, \tag{28}$$

after separating the diffusive variance from the diffusive covariance, according to eqn. (10) in the main text. Similarly, for the components of r we have

$$\langle r\rangle = \frac{1}{2}\frac{1}{S}\sum_{i=1}^{S}\frac{1}{\Delta t_i}\left\langle\left|\frac{1}{N}\sum_{n=1}^{N}\Delta\mathbf{r}_n(t_i) - \frac{\Delta t_i}{NT}\sum_{n=1}^{N}\sum_{j=1}^{S}\Delta\mathbf{r}_n(t_j)\right|^2\right\rangle, \tag{29}$$

$$= \frac{1}{2}\frac{1}{N^2S}\sum_{i=1}^{S}\sum_{\substack{n=1\\m=1}}^{N}\frac{\left\langle\mathbf{w}_n(t_i)^{\mathrm{T}}\mathbf{w}_m(t_i)\right\rangle}{\Delta t_i} - \frac{1}{2}\frac{1}{N^2ST}\sum_{\substack{i=1\\j=1}}^{S}\sum_{\substack{n=1\\m=1}}^{N}\left\langle\mathbf{w}_n(t_i)^{\mathrm{T}}\mathbf{w}_m(t_j)\right\rangle, \tag{30}$$

where again the drift dependence cancels, and further we have

$$\langle r\rangle = \frac{1}{N^2S}\sum_{i=1}^{S}\sum_{\substack{n=1\\m=1}}^{N} D_{nm} - \frac{1}{N^2ST}\sum_{i=1}^{S}\sum_{\substack{n=1\\m=1}}^{N} D_{nm}\,\Delta t_i, \tag{31}$$

$$= \frac{S-1}{N^2S}\sum_{\substack{n=1\\m=1}}^{N} D_{nm}, \tag{32}$$

$$= \frac{S-1}{N^2S}\sum_{n=1}^{N}\sigma + \frac{S-1}{N^2S}\sum_{n\neq m}^{N}\rho\,, \tag{33}$$

$$= \frac{S-1}{NS}\sigma + \frac{(N-1)(S-1)}{NS}\rho\,. \tag{34}$$

Altogether we have the linear relation

$$\begin{bmatrix}\langle s\rangle\\ \langle r\rangle\end{bmatrix} = \frac{1}{NS}\begin{bmatrix}NS-1 & -(N-1)\\ S-1 & (N-1)(S-1)\end{bmatrix}\begin{bmatrix}\sigma\\ \rho\end{bmatrix}. \tag{35}$$

Therefore, the MVU diffusion estimates are given by matrix inversion to be

$$\begin{bmatrix} \hat{\sigma} \\ \hat{\rho} \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{S-1} \\ \frac{-1}{N-1} & \frac{NS-1}{(N-1)(S-1)} \end{bmatrix} \begin{bmatrix} s \\ r \end{bmatrix}, \qquad \begin{bmatrix} \hat{\lambda}_0 \\ \hat{\lambda}_+ \end{bmatrix} = \begin{bmatrix} 0 & \frac{NS}{S-1} \\ \frac{N}{N-1} & \frac{-N}{N-1} \end{bmatrix} \begin{bmatrix} s \\ r \end{bmatrix}. \tag{36}$$

In line with expectations, the MVU estimate of $\hat{\lambda}_+$ is the ML estimate, while the MVU estimate of $\hat{\lambda}_0$ is $n_0/(n_0-2)$ times the ML estimate, indicating that estimation of the two-dimensional mean has cost two degrees-of-freedom's worth of information. The distributions of our MVU diffusion estimates are given by

$$(n_0-2)\frac{\hat{\lambda}_0}{\lambda_0} \sim \chi^2_{n_0-2}\,, \qquad n_+\frac{\hat{\lambda}_+}{\lambda_+} \sim \chi^2_{n_+}\,, \tag{37}$$

$$\mathrm{VAR}[\hat{\lambda}_0] = \frac{2\,\lambda_0^2}{n_0-2}\,, \qquad \mathrm{VAR}[\hat{\lambda}_+] = \frac{2\,\lambda_+^2}{n_+}\,, \tag{38}$$

where these four statistics are independent, $n_0 = 2S$ and $n_+ = 2(N-1)S$. The degrees of freedom for the two diffusion coefficients are finally given by

$$d_0 = \mathrm{DOF}[\hat{\lambda}_0] = 2(S-1)\,, \qquad d_+ = \mathrm{DOF}[\hat{\lambda}_+] = 2(N-1)S\,. \tag{39}$$

Finally, with Jacobian (35), we can transform the variances of λ_0, λ_+ into the covariances of σ, ρ.

2.2 MVU uncorrelated diffusion with uniform drift

In this case the diffusion matrix has one eigenvalue σ, with one sufficient statistic s. The MVU estimator and chi-squared degrees of freedom are easily found to be

$$\hat{\sigma} = \frac{NS}{NS-1}s\,, \qquad d = \mathrm{DOF}[\hat{\sigma}] = 2(NS-1)\,, \tag{40}$$

$$d\,\hat{\sigma} \sim \chi^2_d\,, \qquad \mathrm{VAR}[\hat{\sigma}] = \frac{2\,\sigma^2}{d}\,, \tag{41}$$

effectively summing the degrees of freedom in eqn. (39).

2.3 MVU uniform drift correlation

While eqn. (8) provides an unbiased estimator of the drift, because it is a non-linear operation, squaring this value does not result in an unbiased square drift estimator.

Instead, the square drift has expectation value

$$\langle \hat{\boldsymbol{\nu}}^2 \rangle = \frac{1}{N^2 T^2} \sum_{\substack{n=1 \\ m=1}}^{N} \sum_{\substack{i=1 \\ j=1}}^{S} \left\langle \Delta \mathbf{r}_n(t_i)^{\mathrm{T}} \Delta \mathbf{r}_m(t_j) \right\rangle , \tag{42}$$

$$= \boldsymbol{\nu}^2 + \frac{1}{N^2 T^2} \sum_{\substack{n=1 \\ m=1}}^{N} \sum_{\substack{i=1 \\ j=1}}^{S} \left\langle \mathbf{w}_n(t_i)^{\mathrm{T}} \mathbf{w}_m(t_j) \right\rangle , \tag{43}$$

$$= \boldsymbol{\nu}^2 + \frac{2}{N^2 T^2} \sum_{\substack{n=1 \\ m=1}}^{N} \sum_{i=1}^{S} D_{nm} \, \Delta t_i, \tag{44}$$

$$= \boldsymbol{\nu}^2 + \frac{2}{N^2 T} \sum_{\substack{n=1 \\ m=1}}^{N} D_{nm} \, , \tag{45}$$

$$= \boldsymbol{\nu}^2 + \frac{2}{NT} \left(\sigma + (N-1)\rho \right) , \tag{46}$$

$$= \boldsymbol{\nu}^2 + \frac{2}{NT} \lambda_0 \, . \tag{47}$$

The estimator $\hat{\boldsymbol{\nu}}^2$ is distributed according to a non-central chi-squared distribution. From the moment decomposition of the normal distribution, the variance of the ordinary (biased) square drift estimate is given by

$$\mathrm{VAR}[\hat{\boldsymbol{\nu}}^2] = 4 \, \boldsymbol{\nu}^2 \, \mathrm{VAR}[\hat{\nu}] + 4 \, \mathrm{VAR}[\hat{\nu}]^2 = 4 \, \boldsymbol{\nu}^2 \frac{\lambda_0}{NT} + 4 \left(\frac{\lambda_0}{NT} \right)^2 , \tag{48}$$

where the first term equates to the delta approximation. Finally, note that $\lambda_0 = \sigma$ for the model of uncorrelated diffusion.

3 Model selection

Here we derive corrected AIC values for all of our candidate models, without relying on maximum likelihood estimation. Most readers will be familiar with AIC-based model selection, which is thoroughly reviewed in Burnham and Anderson (2002). The basic AIC formula is (proportionally) an asymptotic estimator of the Kullback-Leibler (KL) divergence between the true model and a model estimated via maximum likelihood. Model-specific $\mathrm{AIC_C}$ formulas are unbiased estimators of the KL divergence between the true model and a specific class of model, though generally still estimated via maximum likelihood. As we are working in a regime where maximum likelihood is heavily biased, neither ordinary AIC nor even ML-based $\mathrm{AIC_C}$ may be adequate for our needs. Instead, we work from first principles and derive MVU estimators for the KL divergence between the true model and our class of models estimated via MVU parameter estimators, making our information criteria asymptotically equivalent to AIC, but more suitable than both AIC and ML-based $\mathrm{AIC_C}$ for small sample sizes. We emphasize that $\mathrm{AIC_C}$ formulas are both model and estimator specific.

Starting essentially from Burnham and Anderson (2002) eqn. (7.20), we have for the KL divergence and target expectation value

$$\mathrm{KL}(\boldsymbol{\theta}) = \langle \ell(\boldsymbol{\theta}|\mathbf{r}) \rangle_{\mathbf{r}} + \text{constant}\,, \tag{49}$$

$$\langle \mathrm{AIC_C} \rangle = -2 \left\langle \mathrm{KL}(\hat{\boldsymbol{\theta}}) \right\rangle = -2 \, \langle\!\langle \ell(\boldsymbol{\theta}|\mathbf{r}) \rangle\!\rangle_{\mathbf{r},\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\,, \tag{50}$$

where first we marginalize the log-likelihood of the data $\mathbf{r}$, given some not yet specified parameters $\boldsymbol{\theta}$, to obtain the KL divergence at $\boldsymbol{\theta}$, and then we take the expectation value of the KL divergence with respect to our estimators $\hat{\boldsymbol{\theta}}$, after evaluating $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Operationally, we can take the two expectation values simultaneously if we treat the data and parameter estimators as independent, even though that is not the case. From hereon, we will use the double brackets $\langle\!\langle \cdots \rangle\!\rangle$ to denote this double average.

First we note that the quadratic terms in our log-likelihood function are the squares of mean-zero Gaussain random variables under the double expectation values, independent of their denominators, and easily resolve to

$$\langle\!\langle s \rangle\!\rangle = \frac{1}{NS} \sum_{i=1}^{S} \sum_{n=1}^{N} \left(\sigma + \Delta t_i \frac{1}{2} \mathrm{tr}\, \mathrm{COV}[\hat{\boldsymbol{\nu}}] \right), \tag{51}$$

$$= \sigma + \frac{T}{S}\frac{1}{2} \mathrm{tr}\, \mathrm{COV}[\hat{\boldsymbol{\nu}}], \tag{52}$$

$$\langle\!\langle r \rangle\!\rangle = \frac{1}{N^2 S} \sum_{i=1}^{2} \sum_{\substack{n=1 \\ m=1}}^{N} \left(D_{nm} + \Delta t_i \frac{1}{2} \mathrm{tr}\, \mathrm{COV}[\hat{\boldsymbol{\nu}}] \right), \tag{53}$$

$$= \frac{\sigma + (N-1)\rho}{N} + \frac{T}{S}\frac{1}{2} \mathrm{tr}\, \mathrm{COV}[\hat{\boldsymbol{\nu}}], \tag{54}$$

$$= \frac{\lambda_0}{N} + \frac{T}{S}\frac{1}{2} \mathrm{tr}\, \mathrm{COV}[\hat{\boldsymbol{\nu}}], \tag{55}$$

$$\langle\!\langle s-r \rangle\!\rangle = \frac{N-1}{N}(\sigma - \rho), \tag{56}$$

$$= \frac{N-1}{N} \lambda_+ . \tag{57}$$

The remaining calculations are model specific.

3.1 Uniform drift and correlated diffusion

From the double expectation value of the log-likelihood function, we have the target expectation value

$$\langle \mathrm{AIC_C} \rangle = n_0 \left(\left\langle \log \hat{\lambda}_0 \right\rangle + N \left\langle\!\!\left\langle \frac{r}{\hat{\lambda}_0} \right\rangle\!\!\right\rangle \right) + n_+ \left(\left\langle \log \hat{\lambda}_+ \right\rangle + \frac{N}{N-1} \left\langle\!\!\left\langle \frac{s-r}{\hat{\lambda}_+} \right\rangle\!\!\right\rangle \right) - 2\,C, \tag{58}$$

and from relations (51)-(57) the double expectation values of s and r resolve to

$$\langle\!\langle r \rangle\!\rangle = \frac{\lambda_0}{N} + \frac{\lambda_0}{NS}, \qquad \langle\!\langle s-r \rangle\!\rangle = \frac{N-1}{N} \lambda_+, \tag{59}$$

and are independent of their corresponding denominators. Next we only have to evaluate some inverse chi-squared mean values

$$\langle \mathrm{AIC_C} \rangle = n_0 \left(\left\langle \log \hat{\lambda}_0 \right\rangle + \frac{S+1}{S} \left\langle \frac{\lambda_0}{\hat{\lambda}_0} \right\rangle \right) + n_+ \left(\left\langle \log \hat{\lambda}_+ \right\rangle + \left\langle \frac{\lambda_+}{\hat{\lambda}_+} \right\rangle \right) - 2\,C, \tag{60}$$

$$= n_0 \left(\left\langle \log \hat{\lambda}_0 \right\rangle + \frac{S+1}{S} \frac{d_0}{d_0 - 2} \right) + n_+ \left(\left\langle \log \hat{\lambda}_+ \right\rangle + \frac{d_+}{d_+ - 2} \right) - 2\,C, \tag{61}$$

plus a constant. Therefore, the unbiased estimator is to within a constant

$$\mathrm{AIC_C} = n_0 \left(\log \hat{\lambda}_0 + \frac{n_0\!+\!2}{n_0}\frac{n_0\!-\!2}{n_0\!-\!4}\right) + n_+ \left(\log \hat{\lambda}_+ + \frac{n_+}{n_+\!-\!2}\right) - 2\,C\,. \tag{62}$$

Working this expression into a form involving the log-likelihood function is possible, but not especially convenient, because the numerator and denominator of the quadratic terms do not cancel, in contrast to maximum likelihood analysis. Expanding in terms of $1/n$, the asymptotic limit of our MVU AIC value is

$$\lim_{n\to\infty} \mathrm{AIC_C} = n_0 \left(\log \frac{n_0}{n_0\!-\!2} \hat{\lambda}_0^{\mathrm{ML}} + 1 + \frac{4}{n_0}\right) + n_+ \left(\log \hat{\lambda}_+^{\mathrm{ML}} + 1 + \frac{2}{n_+}\right) - 2\,C\,, \tag{63}$$

$$= n_0 \left(\log \hat{\lambda}_0^{\mathrm{ML}} + \frac{2}{n_0} + 1 + \frac{4}{n_0}\right) + n_+ \left(\log \hat{\lambda}_+^{\mathrm{ML}} + 1 + \frac{2}{n_+}\right) - 2\,C\,, \tag{64}$$

$$= 2 \times 4 - 2\,\ell\!\left(\boldsymbol{\nu}^{\mathrm{ML}}, \hat{\lambda}_0^{\mathrm{ML}}, \hat{\lambda}_+^{\mathrm{ML}} \middle| \, \mathbf{r}\right)\,, \tag{65}$$

where we have used the $d_0 = n_0 - 2$ DOF relation between the MVU and ML estimates of λ_0, with all other estimators being equivalent.

3.2 No drift and correlated diffusion

In the case of zero drift, we can neglect the contribution from the drift covariance, leaving us with

$$\mathrm{AIC_C} = n_0 \left(\log \hat{\lambda}_0 + \frac{n_0}{n_0\!-\!2}\right) + n_+ \left(\log \hat{\lambda}_+ + \frac{n_+}{n_+\!-\!2}\right) - 2\,C, \tag{66}$$

which is equivalent to $\mathrm{AIC_C}$, as the MVU and ML estimates are equivalent in this case. In terms of the MLEs, the asymptotic value is easily found to be

$$\lim_{n\to\infty} \mathrm{AIC_C} = 2 \times 2 - 2\,\ell\!\left(\hat{\lambda}_0^{\mathrm{ML}}, \hat{\lambda}_+^{\mathrm{ML}} \middle| \, \mathbf{r}\right)\,, \tag{67}$$

consistent with conventional AIC, as it must be.

3.3 Uniform drift and uncorrelated diffusion

In the case of uncorrelated diffusion, we have the target expectation value

$$\langle \mathrm{AIC_C} \rangle = n \left(\langle \log \hat{\sigma} \rangle + \left\langle\!\!\left\langle \frac{s}{\hat{\sigma}} \right\rangle\!\!\right\rangle \right) - 2\,C\,, \tag{68}$$

$$= n \left(\langle \log \hat{\sigma} \rangle + \frac{NS{+}1}{NS} \frac{d}{d{-}2} \right) - 2\,C\,, \tag{69}$$

$$= n \left(\langle \log \hat{\sigma} \rangle + \frac{n{+}2}{n} \frac{n{-}2}{n{-}4} \right) - 2\,C\,, \tag{70}$$

and so our MVU AIC value is

$$\mathrm{AIC_C} = n \left(\log \hat{\sigma} + \frac{n{+}2}{n} \frac{n{-}2}{n{-}4} \right) - 2\,C\,, \tag{71}$$

which asymptotically resolves to

$$\lim_{n\to\infty} \mathrm{AIC_C} = 2 \times 3 - 2\,\ell\big(\boldsymbol{\nu}^{\mathrm{ML}}, \hat{\sigma}^{\mathrm{ML}} \big|\, \mathbf{r}\big)\,. \tag{72}$$

3.4 No drift and uncorrelated diffusion

With neither drift nor correlated diffusion, we have

$$\mathrm{AIC_C} = n \left(\log \hat{\sigma} + \frac{n}{n{-}2} \right) - 2\,C\,, \tag{73}$$

which is equivalent to the $\mathrm{AIC_C}$ value and asymptotically equal to

$$\lim_{n\to\infty} \mathrm{AIC_C} = 2 \times 1 - 2\,\ell\big(\hat{\sigma}^{\mathrm{ML}} \big|\, \mathbf{r}\big)\,. \tag{74}$$

4 Index estimators

Here we overview the analytic workflow for calculating well-behaved index estimates, with first-order bias corrections, and their corresponding confidence intervals. We start by expressing the indices in terms of their numerators and denominators

$$\eta_{\mathrm{dif}} = \frac{P_{\mathrm{dif}}}{M}\,, \qquad \eta_{\mathrm{dft}} = \frac{P_{\mathrm{dft}}}{M}\,, \qquad \eta_{\mathrm{tot}} = \eta_{\mathrm{dif}} + \eta_{\mathrm{dft}}\,, \tag{75}$$

where in $\hat{P}_{\mathrm{dft}}$ and $\hat{M}$, we use the biased estimator of $\boldsymbol{\nu}^2$, $\hat{\boldsymbol{\nu}}^2$. Making these terms unbiased would not make the indices unbiased, and it is more important at this stage to have $\boldsymbol{\nu}^2 > 0$. Next, we construct the Jacobian transformation matrix with which we can easily transform the variances of our sufficient statistics into the covariance matrix of $(\hat{P}_{\text{dif}}, \hat{P}_{\text{dft}}, \hat{M})$. Transforming this covariance into the index covariance can finally by performed with

$$\frac{\partial(\eta_{\text{dif}}, \eta_{\text{dft}}, \eta_{\text{tot}})}{\partial(P_{\text{dif}}, P_{\text{dft}}, M)} = \frac{1}{M} \begin{bmatrix} 1 & 0 & -\eta_{\text{dif}} \\ 0 & 1 & -\eta_{\text{dft}} \\ 1 & 1 & -\eta_{\text{tot}} \end{bmatrix}, \tag{76}$$

in the delta approximation.

To calculate our first-order index biases, we first decompose our estimators into

$$\hat{A} = A + \text{BIAS}[\hat{A}] + \text{DEV}[\hat{A}], \quad \langle \text{DEV}[\hat{A}] \rangle = 0, \quad \text{VAR}\Big[\text{DEV}[\hat{A}]\Big] = \text{VAR}[\hat{A}], \tag{77}$$

where $\text{DEV}[\hat{A}]$ denotes the mean-zero random deviations of the estimator $\hat{A}$. With the above decomposition, our indices can be expressed and expanded

$$\hat{\eta} = \frac{P + \text{BIAS}[\hat{P}] + \text{DEV}[\hat{P}]}{M + \text{BIAS}[\hat{M}] + \text{DEV}[\hat{M}]}, \tag{78}$$

$$= \eta \frac{1 + \frac{\text{BIAS}[\hat{P}] + \text{DEV}[\hat{P}]}{P}}{1 + \frac{\text{BIAS}[\hat{M}] + \text{DEV}[\hat{M}]}{M}}, \tag{79}$$

$$= \eta \left(1 + \frac{\text{BIAS}[\hat{P}] + \text{DEV}[\hat{P}]}{P}\right) \sum_{k=0}^{\infty} \left(-\frac{\text{BIAS}[\hat{M}] + \text{DEV}[\hat{M}]}{M}\right)^k, \tag{80}$$

with which we calculate the first-order bias to be

$$\frac{\text{BIAS}[\hat{\eta}]}{\eta} = \frac{\text{BIAS}[\hat{P}]}{P} - \frac{\text{BIAS}[\hat{M}]}{M} - \frac{\text{COV}[\hat{P}, \hat{M}]}{PM} + \frac{\text{VAR}[\hat{M}]}{M^2} + \mathcal{O}\left(\frac{1}{n^2}\right), \tag{81}$$

Now, just as we do not recommend using the unbiased estimator for $\boldsymbol{\nu}^2$ in the indices, because the unbiased estimator is not positive definite, as the indices are generally constrained $-1 \leq \eta \leq 1$, we also do not recommend directly debiasing $\hat{\eta}$. Instead, we apply a correction to the Fisher-transformed index, $\hat{z} = \tanh^{-1} \hat{\eta}$, that produces the appropriate first-order correction in the untransformed index, $\hat{\eta}$. A small $\mathcal{O}(1/n)$ correction $d\hat{z}$ to the Fisher-transformed index will propagate as

$$\tanh(\hat{z} + d\hat{z}) = \tanh \hat{z} + \frac{dz}{\cosh^2 \hat{z}} + \mathcal{O}\left(\frac{1}{n^2}\right), \tag{82}$$

$$= \hat{\eta} + \frac{d\hat{z}}{1 - \hat{\eta}^2} + \mathcal{O}\left(\frac{1}{n^2}\right), \tag{83}$$

and so our desired Fisher-transformed correction is given by

$$d\hat{z} = -(1-\hat{\eta}^2)\,\mathrm{BIAS}[\hat{\eta}]\,. \tag{84}$$

As perfectly correlated movement is unlikely, it is also justified to calculate normal confidence intervals on the Fisher-transformed indices via

$$\mathrm{VAR}[\hat{z}+d\hat{z}] \approx \left(\frac{1}{1-\hat{\eta}^2} + 2\,\hat{\eta}\,\mathrm{BIAS}[\hat{\eta}]\right)^2 \mathrm{VAR}[\hat{\eta}]\,. \tag{85}$$

In the following sections we detail the model-specific components necessary for this analysis.

4.1 Uniform drift and correlated diffusion

For the model of correlated diffusion and uniform drift, the constituents of the index estimators are given by

$$\hat{P}_{\mathrm{dif}} = 2\,\hat{\rho}\,, \qquad \hat{P}_{\mathrm{dft}} = \Delta t\,\hat{\boldsymbol{\nu}}^2\,, \quad \hat{M} = 2\,\hat{\sigma} + \Delta t\,\hat{\boldsymbol{\nu}}^2\,, \tag{86}$$

$$\hat{P}_{\mathrm{dif}} = \frac{2}{N}\left(\hat{\lambda}_0 - \hat{\lambda}_+\right), \qquad \hat{M} = \frac{2}{N}\left(\hat{\lambda}_0 + (N-1)\lambda_+\right) + \Delta t\,\hat{\boldsymbol{\nu}}^2\,, \tag{87}$$

and their biases are

$$\mathrm{BIAS}\Big[\hat{P}_{\mathrm{dif}}\Big] = 0\,, \qquad \mathrm{BIAS}\Big[\hat{P}_{\mathrm{dft}}\Big] = \frac{2\Delta t}{NT}\lambda_0\,, \qquad \mathrm{BIAS}\Big[\hat{M}\Big] = \frac{2\Delta t}{NT}\lambda_0\,. \tag{88}$$

Finally, the Jacobian matrix for transforming variances in the sufficient statistics into covariances is

$$\frac{\partial(P_{\mathrm{dif}}, P_{\mathrm{dft}}, M)}{\partial(\lambda_0, \lambda_+, \boldsymbol{\nu}^2)} = \begin{bmatrix} \frac{2}{N} & -\frac{2}{N} & 0 \\ 0 & 0 & \Delta t \\ \frac{2}{N} & 2\frac{N-1}{N} & \Delta t \end{bmatrix}. \tag{89}$$

4.2 Uniform drift and uncorrelated diffusion

In the case of uncorrelated diffusion, we have only drift correlation

$$\hat{P}_{\mathrm{dif}} = 0\,, \qquad \hat{P}_{\mathrm{dft}} = \Delta t\,\hat{\boldsymbol{\nu}}^2\,, \qquad \hat{M} = 2\,\hat{\sigma} + \Delta t\,\hat{\boldsymbol{\nu}}^2\,, \tag{90}$$

and their biases are

$$\mathrm{BIAS}\Big[\hat{P}_{\mathrm{dft}}\Big] = \frac{2\Delta t}{NT}\sigma_0\,, \qquad \mathrm{BIAS}\Big[\hat{M}\Big] = \frac{2\Delta t}{NT}\sigma_0\,. \tag{91}$$

The Jacobian matrix for transforming variances in the sufficient statistics into covariances is

$$\frac{\partial(P_{\mathrm{dft}}, M)}{\partial(\sigma, \boldsymbol{\nu}^2)} = \begin{bmatrix} 0 & \Delta t \\ 2 & \Delta t \end{bmatrix} . \tag{92}$$

4.3 No drift and correlated diffusion

Without drift, we have only diffusion correlation

$$\hat{P}_{\mathrm{dif}} = 2\,\hat{\rho}\,, \qquad \hat{P}_{\mathrm{dft}} = 0\,, \qquad \hat{M} = 2\,\hat{\sigma}\,, \tag{93}$$

$$\hat{P}_{\mathrm{dif}} = \frac{2}{N}\left(\hat{\lambda}_0 - \hat{\lambda}_+\right), \qquad \hat{M} = \frac{2}{N}\left(\hat{\lambda}_0 + (N-1)\lambda_+\right), \tag{94}$$

which are all unbiased. The Jacobian matrix for transforming variances in the sufficient statistics into covariances is

$$\frac{\partial(P_{\mathrm{dif}}, M)}{\partial(\lambda_0, \lambda_+)} = \begin{bmatrix} \frac{2}{N} & -\frac{2}{N} \\ \frac{2}{N} & 2\frac{N-1}{N} \end{bmatrix} . \tag{95}$$

5 Synopsis of analysis

As a broad outline, our analysis flow begins with the MVU parameter estimates and corresponding covariances of our four basic models. For each model, we have derived exact $\mathrm{AIC_C}$ values to facilitate model selection. With the selected model parameters in hand, we can then estimate our indices, η, however these naive estimates can be fairly biased. Removing the first-order bias directly from $\hat{\eta}$ can produce invalid estimates extending beyond the $(-1, 1)$ interval, so we debias $\hat{\eta}$ under a Fisher transformation. Note that we do not debias the Fisher transform of $\hat{\eta}$. We also calculate valid confidence intervals on the indices under a Fisher transformation, and back-transform them as well. Our more detailed guide with equation references follows.

1. The MVU mean drift estimator is given in eqn. (10), and the sufficient statistics s,r are given in eqns. (13)-(14), which can be evaluated with and without drift. The MVU diffusion parameter estimates $\hat{\lambda}_0, \hat{\lambda}_+$ (or $\hat{\sigma}$) and their associated degrees of freedom are given in sec. 1.2.2 for uncorrelated diffusion without drift, sec. 1.2.1 for correlated diffusion without drift, sec. 2.2 for uncorrelated diffusion with uniform drift, and sec. 2.1 for correlated diffusion with uniform drift. The sample numbers $n_0 = 2S$, $n_+ = 2(N-1)S$, and $n = n_0 + n_+$ are the same for all models.

2. With the help of relation (48), one then has for each model its MVU parameter estimates and their diagonal covariance matrix

$$\mathrm{COV}[(\lambda_0, \lambda_+, \boldsymbol{\nu}^2)] = \mathrm{diag}\big(\mathrm{VAR}[\lambda_0], \mathrm{VAR}[\lambda_+], \mathrm{VAR}[\boldsymbol{\nu}^2]\big)\,. \tag{96}$$

3. Model-specific $\mathrm{AIC_C}$ values are given under the respective sub-headings of section 3. In all cases, the likelihood function is given by eqn. 15.

4. Formulas for the indices are decomposed $\eta = P/M$ with point estimates, biases, and Jacobians provided in the respective sub-headings of section 4. The covariance of $\hat{P}, \hat{M}$ is given by the linear transformation

$$\mathrm{COV}[(\hat{P}_{\mathrm{dif}}, \hat{P}_{\mathrm{dft}}, \hat{M})] = \frac{\partial(P_{\mathrm{dif}}, P_{\mathrm{dft}}, M)}{\partial(\lambda_0, \lambda_+, \boldsymbol{\nu}^2)}\, \mathrm{COV}[(\lambda_0, \lambda_+, \boldsymbol{\nu}^2)]\, \frac{\partial(P_{\mathrm{dif}}, P_{\mathrm{dft}}, M)}{\partial(\lambda_0, \lambda_+, \boldsymbol{\nu}^2)}^{\mathrm{T}}, \tag{97}$$

   and one should then have the bias and covariance of each P, M.

5. The naive index esitmator $\hat{\eta} = \hat{P}/\hat{M}$ has first-order bias (81) and Jacobian (76), with which we can transform the covariance of $\hat{P}, \hat{M}$ into the covariance of the indices. We remove this bias under Fisher transformation $\hat{z} = \tanh^{-1}\hat{\eta}$, with the correction $d\hat{z}$ (84). Also under Fisher transformation, the estimator $\hat{z} + d\hat{z}$ has variance (85). Finally the point estimate $\hat{z} + d\hat{z}$ and associated normal confidence intervals can be transformed back into the index domain via hyperbolic tangent.

References

Burnham, K. P. and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer Science & Business Media.