B Kernel-density estimation

Here we provide a derivation of the conventional, multivariate Gaussian reference-function KDE in B.1 and a derivation of the analogous autocorrelated KDE in B.2 that relaxes the IID assumption of conventional KDE, but otherwise has the same structure and follows the same sequence of steps as the derivation in B.1. For those unfamiliar with this kernel-density estimator, the autocorrelated derivation in B.2 is much easier to follow after understanding the simpler, conventional derivation in B.1.

B.1 Conventional KDE derivation

Given observed locations $\mathbf{r}(t_i)$ at times t_i with $i \in \{1, \cdots, n\}$, and in q dimensions with the location vector $\mathbf{r}$, the kernel-density estimate $\hat{p}(\mathbf{r})$ of the true probability-density function (PDF) of $\mathbf{r}$, $p(\mathbf{r})$, is taken to be

$$\hat{p}(\mathbf{r}) = \frac{1}{n}\sum_{i=1}^{n} \kappa(\mathbf{r}-\mathbf{r}(t_i))\,, \tag{B.1}$$

where $\kappa(\mathbf{r})$ is a kernel that smooths the empirical distribution of the data, which we will choose to be Gaussian with zero mean and $q \times q$ covariance matrix $\boldsymbol{\sigma}_\mathrm{B}$.

$$\kappa(\mathbf{r}) = \mathcal{N}(\mathbf{r}; \mathbf{0}, \boldsymbol{\sigma}_\mathrm{B}) = \frac{e^{-\frac{1}{2}\mathbf{r}^\mathrm{T}\boldsymbol{\sigma}_\mathrm{B}^{-1}\mathbf{r}}}{\sqrt{\det(2\pi\boldsymbol{\sigma}_\mathrm{B})}}\,. \tag{B.2}$$

The smoothing kernel's covariance $\boldsymbol{\sigma}_\mathrm{B}$ is also referred to as the 'bandwidth' or 'smoothing' parameter (Silverman, 1986). Traditionally, one seeks to solve for the optimal bandwidth $\hat{\boldsymbol{\sigma}}_\mathrm{B}$ that minimizes the mean integrated squared error (MISE), hereafter denoted $\varepsilon(\boldsymbol{\sigma}_\mathrm{B})$, with respect to the true density function $p(\mathbf{r})$.

$$\varepsilon(\boldsymbol{\sigma}_\mathrm{B}) = \left\langle \int d^q\mathbf{r}\, |\hat{p}(\mathbf{r})-p(\mathbf{r})|^2 \right\rangle = \overbrace{\int d^q\mathbf{r}\, \langle \hat{p}(\mathbf{r})\, \hat{p}(\mathbf{r})\rangle}^{1^\mathrm{st}} - 2\overbrace{\int d^q\mathbf{r}\, \langle p(\mathbf{r})\, \hat{p}(\mathbf{r})\rangle}^{2^\mathrm{nd}} + \overbrace{\int d^q\mathbf{r}\, \langle p(\mathbf{r})\, p(\mathbf{r})\rangle}^{3^\mathrm{rd}}, \tag{B.3}$$

where $\langle \cdots \rangle$ denotes the average over realizations of the time series, $\int d^q\mathbf{r}$ denotes the q-dimensional volume integral over $\mathbb{R}^q$, which is an integral over area in two dimensions. Without knowledge of the true density function, the MISE cannot be calculated exactly. In the reference-function approximation, $p(\mathbf{r})$ is approximated in the second term of Eq. (B.3) with a model of appropriate mean and covariance

$$p(\mathbf{r}) \approx \mathcal{N}(\mathbf{r}; \boldsymbol{\mu}_0, \boldsymbol{\sigma}_0) = \frac{e^{-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_0)^\mathrm{T}\boldsymbol{\sigma}_0^{-1}(\mathbf{r}-\boldsymbol{\mu}_0)}}{\sqrt{\det(2\pi\boldsymbol{\sigma}_0)}}\,, \tag{B.4}$$

where the mean $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\sigma}_0$ are estimated from the observed locations. Alternatively, one can also estimate the MISE by cross validation (Silverman, 1986; Izenman, 1991; Turlach, 1993). However, cross validating autocorrelated data is complicated, as different partitions of the data are exposed to different autocorrelation structure. Therefore, we

choose a method that easily generalizes to autocorrelated data. The third term of the MISE (B.3) has no effect upon optimization and so the degree to which it is well approximated by the reference function is irrelevant, but it serves to keep the MISE positive.

To calculate the MISE (B.3), we use the reference-function approximation (B.4) and perform all spatial integration by "completing the square", which, after some simplification, results in the following relations for the three terms of Eq. (B.3)

$$1^{\text{st}} \qquad \int d^q\mathbf{r}\,\langle \hat{p}(\mathbf{r})\,\hat{p}(\mathbf{r})\rangle = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\langle \frac{e^{-\frac{1}{2}(\mathbf{r}(t_i)-\mathbf{r}(t_j))^{\mathrm{T}}\,(2\,\boldsymbol{\sigma}_{\mathrm{B}})^{-1}\,(\mathbf{r}(t_i)-\mathbf{r}(t_j))}}{\sqrt{(2\pi)^q\det(2\,\boldsymbol{\sigma}_{\mathrm{B}})}}\right\rangle, \tag{B.5}$$

$$2^{\text{nd}} \qquad \int d^q\mathbf{r}\,\langle p(\mathbf{r})\,\hat{p}(\mathbf{r})\rangle = \frac{1}{n}\sum_{j=1}^{n}\left\langle \frac{e^{-\frac{1}{2}(\mathbf{r}(t_j)-\boldsymbol{\mu})^{\mathrm{T}}\,(\boldsymbol{\sigma}_0+\boldsymbol{\sigma}_{\mathrm{B}})^{-1}\,(\mathbf{r}(t_j)-\boldsymbol{\mu})}}{\sqrt{(2\pi)^q\det(\boldsymbol{\sigma}_0+\boldsymbol{\sigma}_{\mathrm{B}})}}\right\rangle, \tag{B.6}$$

$$3^{\text{rd}} \qquad \int d^q\mathbf{r}\,\langle p(\mathbf{r})\,p(\mathbf{r})\rangle = \frac{1}{\sqrt{(2\pi)^q\det(2\,\boldsymbol{\sigma}_0)}}, \tag{B.7}$$

where in the second term (B.6) we used the Woodbury matrix identity (Woodbury, 1950)

$$(\boldsymbol{\sigma}_0+\boldsymbol{\sigma}_{\mathrm{B}})^{-1} = \boldsymbol{\sigma}_{\mathrm{B}}^{-1} - \boldsymbol{\sigma}_{\mathrm{B}}^{-1}\,(\boldsymbol{\sigma}_0^{-1}+\boldsymbol{\sigma}_{\mathrm{B}}^{-1})^{-1}\,\boldsymbol{\sigma}_{\mathrm{B}}^{-1}. \tag{B.8}$$

Next we average over samples. The first term of the MISE (B.5) can be expanded into sums over $i=j$ and $i\neq j$ to yield

$$\int d^q\mathbf{r}\,\langle \hat{p}(\mathbf{r})\,\hat{p}(\mathbf{r})\rangle = \frac{1}{n^2}\sum_{i}\left\langle \frac{1}{\sqrt{(2\pi)^q\det(2\,\boldsymbol{\sigma}_{\mathrm{B}})}}\right\rangle + \frac{1}{n^2}\sum_{i\neq j}\left\langle \frac{e^{-\frac{1}{2}(\mathbf{r}(t_i)-\mathbf{r}(t_j))^{\mathrm{T}}\,(2\,\boldsymbol{\sigma}_{\mathrm{B}})^{-1}\,(\mathbf{r}(t_i)-\mathbf{r}(t_j))}}{\sqrt{(2\pi)^q\det(2\,\boldsymbol{\sigma}_{\mathrm{B}})}}\right\rangle. \tag{B.9}$$

Then, after further simplification we obtain simplified expressions for the first two terms of the MISE (B.3)

$$1^{\text{st}} \qquad \int d^q\mathbf{r}\,\langle \hat{p}(\mathbf{r})\,\hat{p}(\mathbf{r})\rangle = \frac{\frac{1}{n}}{\sqrt{(2\pi)^q\det(2\,\boldsymbol{\sigma}_{\mathrm{B}})}} + \frac{\frac{n-1}{n}}{\sqrt{(2\pi)^2\det(2\,\boldsymbol{\sigma}_0+2\,\boldsymbol{\sigma}_{\mathrm{B}})}}, \tag{B.10}$$

$$2^{\text{nd}} \qquad \int d^q\mathbf{r}\,\langle p(\mathbf{r})\,\hat{p}(\mathbf{r})\rangle = \frac{1}{\sqrt{(2\pi)^q\det(2\,\boldsymbol{\sigma}_0+\boldsymbol{\sigma}_{\mathrm{B}})}}, \tag{B.11}$$

where in the $i\neq j$ sum of Eq. (B.9) we integrated over $\mathbf{r}(t_i)$ before $\mathbf{r}(t_j)$ and then applied the Woodbury matrix identity

$$\boldsymbol{\sigma}_0^{-1} + (\boldsymbol{\sigma}_0+2\,\boldsymbol{\sigma}_{\mathrm{B}})^{-1} = \boldsymbol{\sigma}_0^{-1} + (2\,\boldsymbol{\sigma}_{\mathrm{B}})^{-1} - (2\,\boldsymbol{\sigma}_{\mathrm{B}})^{-1}\,(\boldsymbol{\sigma}_0^{-1}+(2\,\boldsymbol{\sigma}_{\mathrm{B}})^{-1})^{-1}\,(2\,\boldsymbol{\sigma}_{\mathrm{B}})^{-1}. \tag{B.12}$$

Altogether, by combining relations (B.10), (B.11), and (B.7), we finally have

$$\varepsilon(\boldsymbol{\sigma}_{\mathrm{B}}) = \frac{1}{(2\pi)^{\frac{q}{2}}}\left(\frac{\frac{1}{n}}{\sqrt{\det(2\,\boldsymbol{\sigma}_{\mathrm{B}})}} + \frac{\frac{n-1}{n}}{\sqrt{\det(2\,\boldsymbol{\sigma}_0+2\,\boldsymbol{\sigma}_{\mathrm{B}})}} - \frac{2}{\sqrt{\det(2\,\boldsymbol{\sigma}_0+\boldsymbol{\sigma}_{\mathrm{B}})}} + \frac{1}{\sqrt{\det(2\,\boldsymbol{\sigma}_0)}}\right). \tag{B.13}$$

Note that if we linearly transform the input data so that $\boldsymbol{\sigma}_0 \to \mathbf{I}$, the bandwidth matrix can only be a function of the identity matrix, which we can represent as $h^2\,\mathbf{I}$, for some unknown scalar value h. Linearly transforming this result back so that $\mathbf{I} \to \boldsymbol{\sigma}_0$, we then have a bandwidth matrix of the form $\boldsymbol{\sigma}_\mathrm{B} = h^2\,\boldsymbol{\sigma}_0$, so that the MISE (B.13) reduces to a function of only h

$$\varepsilon(h) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\sigma}_0)}}\left(\frac{\frac{1}{n}}{(2h^2)^{q/2}} + \frac{\frac{n-1}{n}}{(2+2h^2)^{q/2}} - \frac{2}{(2+h^2)^{q/2}} + \frac{1}{2^{q/2}}\right). \tag{B.14}$$

The optimal bandwidth provides a consistent estimator in the sense that in the limit of infinite data, an MISE of zero is obtained with a vanishingly small bandwidth

$$\lim_{h\to 0}\lim_{n\to\infty} \varepsilon(h) = 0\,, \tag{B.15}$$

and from the above limit, we can expand the MISE (B.14) in a Laurent series to obtain the approximate solution for large yet finite amounts of data

$$\hat{h} \approx \left(\frac{4}{2+q}\frac{1}{n}\right)^{\frac{1}{4+q}}, \qquad\qquad \hat{\boldsymbol{\sigma}}_\mathrm{B} \approx \left(\frac{4}{2+q}\frac{1}{n}\right)^{\frac{2}{4+q}}\boldsymbol{\sigma}_0\,, \tag{B.16}$$

which is Silverman's rule of thumb (Silverman, 1986). This is an asymptotic result for large amounts of data, in that the finite n corrections are higher order in $1/n$. Plugging this asymptotic result back into the MISE (B.14) gives us a second asymptotic result: the MISE scales asymptotically according to $\varepsilon(\hat{h}) = \mathcal{O}(n^{-4/(4+q)})$. This is an important limit, because, in terms of the order of dependence upon $1/n$, no KDE method can produce better asymptotic convergence (Izenman, 1991), and so all of the best bandwidth-optimization methods will share the same asymptotic bandwidth and error scaling.

B.2 Autocorrelated KDE derivation

We start with the same form of data as in App. B.1, but now we more generally consider the animal movement as represented by a non-stationary process, which allows both the mean velocity and the autocorrelation function to vary in time (possibly resulting from spatial dependence). In this case, the reference-function approximation for the animal's probability of being at location $\mathbf{r}$ at time t is given by

$$p(\mathbf{r},t) \approx \mathcal{N}(\mathbf{r};\boldsymbol{\mu}(t),\boldsymbol{\sigma}(t,t)) = \frac{e^{-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}(t))^\mathrm{T}\boldsymbol{\sigma}(t,t)^{-1}(\mathbf{r}-\boldsymbol{\mu}(t))}}{\sqrt{\det(2\pi\boldsymbol{\sigma}(t,t))}}\,, \tag{B.17}$$

in terms of the mean and autocorrelation function (ACF)

$$\boldsymbol{\mu}(t) = \langle \mathbf{r}(t)\rangle\,, \qquad\qquad \boldsymbol{\sigma}(t,t') = \left\langle (\mathbf{r}(t)-\boldsymbol{\mu}(t))\,(\mathbf{r}(t')-\boldsymbol{\mu}(t'))^\mathrm{T}\right\rangle. \tag{B.18}$$

Just as the covariance must be estimated for the conventional KDE in App. B.1, here the mean and ACF must be estimated from the data (App. C).

As the data pertain to a particular set of times, $\{t_1,\cdots,t_n\}$, we consider the corresponding time-averaged density as our reference function

$$p(\mathbf{r}) \approx \frac{1}{n}\sum_{i=1}^{n}\mathcal{N}(\mathbf{r};\boldsymbol{\mu}(t_i),\boldsymbol{\sigma}(t_i,t_i)) = \sum_{i=1}^{n}\frac{e^{-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}(t_i))^\mathrm{T}\boldsymbol{\sigma}(t_i,t_i)^{-1}(\mathbf{r}-\boldsymbol{\mu}(t_i))}}{\sqrt{\det(2\pi\boldsymbol{\sigma}(t_i,t_i))}}\,, \tag{B.19}$$

though other alternatives can be considered within this formalism. If the movement process is stationary (App. B.2.1), then there is no ambiguity in what distribution determines the lifetime range of the animal, and the sum in Eq. (B.19) is one of all like terms. However, if the animal changes its movement behavior, perhaps by migrating between summer and winter ranges, then specific research objectives might be better addressed by weighting this sum to reflect the summer, winter, or annual range.

Again using a Gaussian kernel-density estimator (B.1) for the density estimate $\hat{p}(\mathbf{r})$ and aiming to minimize the MISE (B.3), we start resolving the MISE analytically by performing the spatial integrals. With some simplification, the three terms of the MISE (B.3) can be expressed

$$1^{\text{st}} \qquad \int d^q\mathbf{r}\,\langle \hat{p}(\mathbf{r})\,\hat{p}(\mathbf{r})\rangle = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\langle \frac{e^{-\frac{1}{2}(\mathbf{r}(t_i)-\mathbf{r}(t_j))^{\mathrm{T}}(2\,\boldsymbol{\sigma}_{\mathrm{B}})^{-1}(\mathbf{r}(t_i)-\mathbf{r}(t_j))}}{\sqrt{(2\pi)^q\det(2\,\boldsymbol{\sigma}_{\mathrm{B}})}}\right\rangle, \tag{B.20}$$

$$2^{\text{nd}} \qquad \int d^q\mathbf{r}\,\langle p(\mathbf{r})\,\hat{p}(\mathbf{r})\rangle = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\langle \frac{e^{-\frac{1}{2}(\boldsymbol{\mu}(t_i)-\mathbf{r}(t_j))^{\mathrm{T}}(\boldsymbol{\sigma}(t_i,t_i)+\boldsymbol{\sigma}_{\mathrm{B}})^{-1}(\boldsymbol{\mu}(t_i)-\mathbf{r}(t_j))}}{\sqrt{(2\pi)^q\det(\boldsymbol{\sigma}(t_i,t_i)+\boldsymbol{\sigma}_{\mathrm{B}})}}\right\rangle, \tag{B.21}$$

$$3^{\text{rd}} \qquad \int d^q\mathbf{r}\,\langle p(\mathbf{r})\,p(\mathbf{r})\rangle = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{e^{-\frac{1}{2}(\boldsymbol{\mu}(t_i)-\boldsymbol{\mu}(t_j))^{\mathrm{T}}(\boldsymbol{\sigma}(t_i,t_i)+\boldsymbol{\sigma}(t_j,t_j))^{-1}(\boldsymbol{\mu}(t_i)-\boldsymbol{\mu}(t_j))}}{\sqrt{(2\pi)^q\det(\boldsymbol{\sigma}(t_i,t_i)+\boldsymbol{\sigma}(t_j,t_j))}}, \tag{B.22}$$

after applying Woodbury matrix identities as in App. B.1. Next we average over samples, $\langle\cdots\rangle$, using the fact that all locations $\mathbf{r}(t)$ are approximated as having a joint Gaussian distribution with ACF given by Eq. (B.18). In particular, for the first term of the MISE (B.20), we use the two-time correlation matrix

$$\left\langle \begin{bmatrix} \mathbf{r}(t_i)-\boldsymbol{\mu}(t_i) \\ \mathbf{r}(t_j)-\boldsymbol{\mu}(t_j) \end{bmatrix} \begin{bmatrix} \mathbf{r}(t_i)-\boldsymbol{\mu}(t_i) \\ \mathbf{r}(t_j)-\boldsymbol{\mu}(t_j) \end{bmatrix}^{\mathrm{T}} \right\rangle = \begin{bmatrix} \boldsymbol{\sigma}(t_i,t_i) & \boldsymbol{\sigma}(t_i,t_j) \\ \boldsymbol{\sigma}(t_j,t_i) & \boldsymbol{\sigma}(t_j,t_j) \end{bmatrix}, \tag{B.23}$$

and the block-matrix representation

$$(\mathbf{r}(t_i)-\mathbf{r}(t_j))^{\mathrm{T}}\,(2\,\boldsymbol{\sigma}_{\mathrm{B}})^{-1}\,(\mathbf{r}(t_i)-\mathbf{r}(t_j)) = \begin{bmatrix} \mathbf{r}(t_i) \\ \mathbf{r}(t_j) \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} +\mathbf{1} \\ -\mathbf{1} \end{bmatrix} (2\,\boldsymbol{\sigma}_{\mathrm{B}})^{-1} \begin{bmatrix} +\mathbf{1} \\ -\mathbf{1} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{r}(t_i) \\ \mathbf{r}(t_j) \end{bmatrix}, \tag{B.24}$$

to complete the squares, wherein we apply the more general Woodbury matrix identity

$$\left(\mathbf{A}+\mathbf{U}\,\mathbf{B}\,\mathbf{U}^{\mathrm{T}}\right)^{-1} = \mathbf{A}-\mathbf{A}^{-1}\,\mathbf{U}\left(\mathbf{B}^{-1}+\mathbf{U}^{-1}\,\mathbf{A}^{-1}\,\mathbf{U}\right)^{-1}\mathbf{U}^{\mathrm{T}}\,\mathbf{A}^{-1}. \tag{B.25}$$

Then, upon integration we apply the matrix determinant lemma

$$\det(\mathbf{A}+\mathbf{U}\,\mathbf{B}\,\mathbf{U}^{\mathrm{T}}) = \det(\mathbf{A})\det(\mathbf{B})\det\left(\mathbf{U}^{\mathrm{T}}\,\mathbf{A}^{-1}\,\mathbf{U}+\mathbf{B}^{-1}\right). \tag{B.26}$$

Finally, we obtain compact expressions for the first two terms of the MISE (B.20)-(B.21)

and represent the last term (B.22) similarly

$$1^{\text{st}} \qquad \int d^q\mathbf{r}\,\langle \hat{p}(\mathbf{r})\,\hat{p}(\mathbf{r})\rangle = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{e^{+\frac{1}{2}(\boldsymbol{\mu}(t_i)-\boldsymbol{\mu}(t_j))^{\mathrm{T}}\boldsymbol{\sigma}_{(1)}^{-1}(t_i,t_j)(\boldsymbol{\mu}(t_i)-\boldsymbol{\mu}(t_j))}}{\sqrt{(2\pi)^q\det(\boldsymbol{\sigma}_{(1)}(t_i,t_j))}}\,, \tag{B.27}$$

$$2^{\text{nd}} \qquad \int d^q\mathbf{r}\,\langle p(\mathbf{r})\,\hat{p}(\mathbf{r})\rangle = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{e^{+\frac{1}{2}(\boldsymbol{\mu}(t_i)-\boldsymbol{\mu}(t_j))^{\mathrm{T}}\boldsymbol{\sigma}_{(2)}^{-1}(t_i,t_j)(\boldsymbol{\mu}(t_i)-\boldsymbol{\mu}(t_j))}}{\sqrt{(2\pi)^q\det(\boldsymbol{\sigma}_{(2)}(t_i,t_j))}}\,, \tag{B.28}$$

$$3^{\text{rd}} \qquad \int d^q\mathbf{r}\,\langle p(\mathbf{r})\,\hat{p}(\mathbf{r})\rangle = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{e^{+\frac{1}{2}(\boldsymbol{\mu}(t_i)-\boldsymbol{\mu}(t_j))^{\mathrm{T}}\boldsymbol{\sigma}_{(3)}^{-1}(t_i,t_j)(\boldsymbol{\mu}(t_i)-\boldsymbol{\mu}(t_j))}}{\sqrt{(2\pi)^q\det(\boldsymbol{\sigma}_{(3)}(t_i,t_j))}}\,, \tag{B.29}$$

in terms of the matrices

$$\boldsymbol{\sigma}_{(1)}(t,t') \equiv 2\,\boldsymbol{\gamma}(t,t') + 2\,\boldsymbol{\sigma}_{\mathrm{B}}\,, \tag{B.30}$$

$$\boldsymbol{\sigma}_{(2)}(t,t') \equiv \boldsymbol{\sigma}(t,t) + \boldsymbol{\sigma}(t',t') + \boldsymbol{\sigma}_{\mathrm{B}}\,, \tag{B.31}$$

$$\boldsymbol{\sigma}_{(3)}(t,t') \equiv \boldsymbol{\sigma}(t,t) + \boldsymbol{\sigma}(t',t')\,. \tag{B.32}$$

where the semi-variance function (SVF, Fleming et al., 2014) is given by

$$2\,\boldsymbol{\gamma}(t,t') \equiv \Big\langle (\mathbf{r}(t)-\mathbf{r}(t'))\,(\mathbf{r}(t)-\mathbf{r}(t'))^{\mathrm{T}}\Big\rangle = \boldsymbol{\sigma}(t,t) + \boldsymbol{\sigma}(t',t') - \boldsymbol{\sigma}(t,t') - \boldsymbol{\sigma}(t',t)\,. \tag{B.33}$$

B.2.1 Stationary processes

In the case of a stationary process, where an animal's movement has the same mean and autocorrelation structure throughout time, we have the mean location and stationary autocorrelation function

$$\boldsymbol{\mu}(t) = \boldsymbol{\mu}_0 \qquad\qquad \boldsymbol{\sigma}(t,t) = \boldsymbol{\sigma}_0 \qquad\qquad \boldsymbol{\sigma}(t,t') = \boldsymbol{\sigma}(t-t')\,, \tag{B.34}$$

which must be first estimated from the data (App. C). Given this lack of time dependence, the MISE terms in relations (B.27), (B.28), and (B.29) simplify and combine to yield

$$\varepsilon(\boldsymbol{\sigma}_{\mathrm{B}}) = \frac{1}{(2\pi)^{\frac{q}{2}}}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{\sqrt{\det(2\,\boldsymbol{\gamma}(t_i-t_j)+2\,\boldsymbol{\sigma}_{\mathrm{B}})}} - \frac{2}{\sqrt{\det(2\,\boldsymbol{\sigma}_0+\boldsymbol{\sigma}_{\mathrm{B}})}} + \frac{1}{\sqrt{\det(2\,\boldsymbol{\sigma}_0)}}\right), \tag{B.35}$$

where the semi-variance function is given by

$$2\,\boldsymbol{\gamma}(\tau) \equiv \Big\langle (\mathbf{r}(t+\tau)-\mathbf{r}(t))\,(\mathbf{r}(t+\tau)-\mathbf{r}(t))^{\mathrm{T}}\Big\rangle = 2\,\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}(+\tau) - \boldsymbol{\sigma}(-\tau)\,, \tag{B.36}$$

which is effectively estimated when the ACF is estimated. For evenly scheduled data, where $t_i = t_0 + i\,dt$ with sampling interval dt, but permitting missing observations, the computational cost of evaluating MISE (B.35) can be reduced from $\mathcal{O}(n^2)$ to $\mathcal{O}(n\log n)$ by summing over unique lags

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{\sqrt{\det\left(2\,\boldsymbol{\gamma}(t_i-t_j)+2\,\boldsymbol{\sigma}_{\mathrm{B}}\right)}} = \sum_{\tau}\frac{n(\tau)}{\sqrt{\det(2\,\boldsymbol{\gamma}(\tau)+2\,\boldsymbol{\sigma}_{\mathrm{B}})}}\,, \tag{B.37}$$

where $n(\tau)$ denotes the number of location pairs $\mathbf{r}(t)$ and $\mathbf{r}(t')$ with time lag $\tau = t - t'$ between them. This sum can be computed with an $\mathcal{O}(n \log n)$ computational cost fast Fourier transform (FFT) (Marcotte, 1996), and the lag sum $\sum_\tau$ runs over all sampled lags, including the negative lags. This expression for the MISE (B.35) can be used to show the superiority of the AKDE over the conventional KDE across a range of observation periods and sampling frequencies (Fig. B.1), in that if the data are autocorrelated then the conventionally estimated bandwidth will be far from optimal.

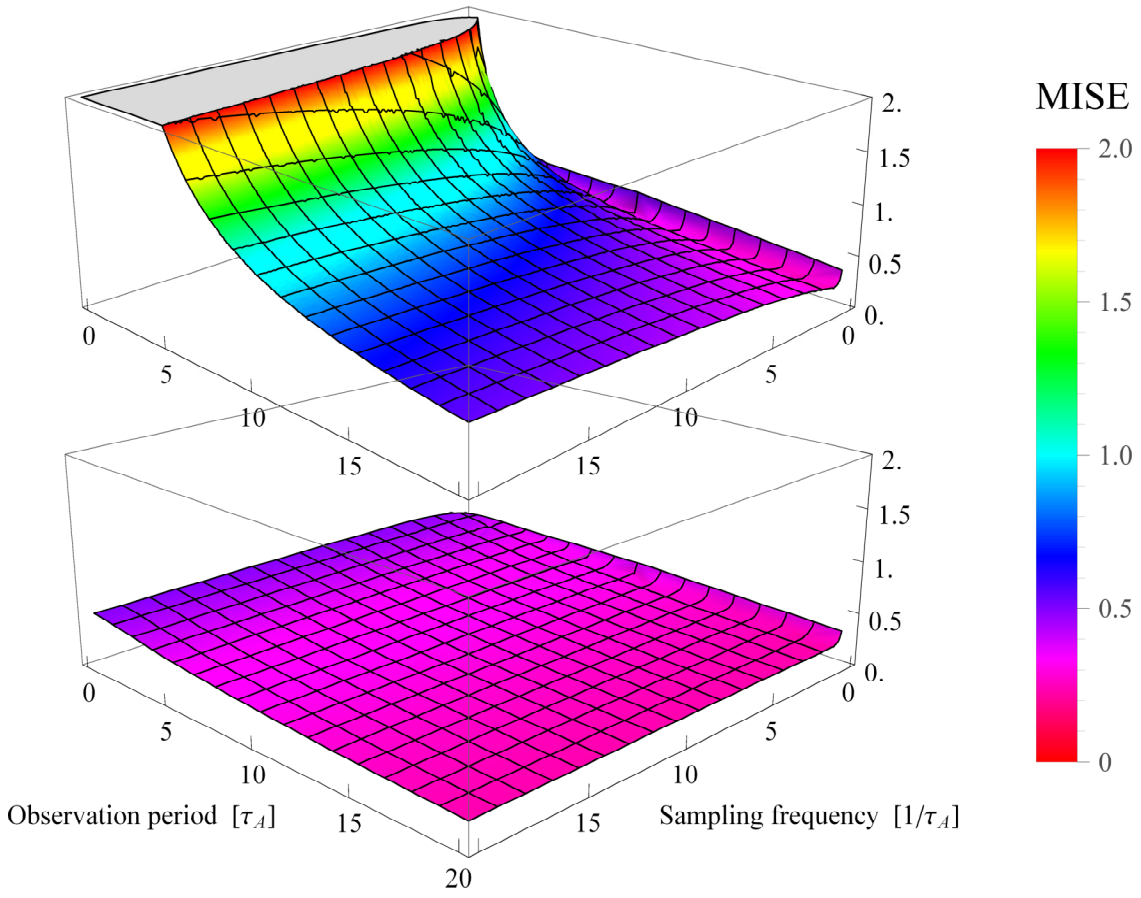

Figure B.1: Differences in mean integrated square error (MISE) for conventional KDE bandwidth optimization (**Top**) and autocorrelated KDE bandwidth optimization (**Bottom**), as a function of sampling frequency and observation period, relative to the autocorrelation timescale of the movement process τ_A. In both cases the MISE is scaled by a factor of $\sqrt{\det 4\pi\boldsymbol{\sigma}_0}$ so that a value of 1–2 on the z-axis denotes a high degree of failure.

If there is no autocorrelation, then by definition $\boldsymbol{\sigma}(\tau)$ vanishes for $\tau \neq 0$, and the autocorrelated MISE term described by Eq. (B.37) reduces to the previous, conventional result in Eq. (B.10). More generally, for autocorrelated movement processes $\boldsymbol{\sigma}(\tau)$ decays only gradually to zero for $\tau \neq 0$. As will be demonstrated, the optimal bandwidth of an autocorrelated process with an ACF that decays gradually is larger than that of an uncorrelated process with an ACF—of the same initial covariance $\boldsymbol{\sigma}(0) = \boldsymbol{\sigma}_0$—that decays instantaneously. This comparison is important because it captures some of the difference between accounting for

autocorrelation and ignoring autocorrelation in the data. Ignoring autocorrelation is tantamount to underestimating the optimal bandwidth.

To see that the optimal bandwidth increases with autocorrelation, first note that the effective sample size n_A of autocorrelated data is necessarily smaller than the total sample size n and that the optimal bandwidth for IID data decreases monotonically with n, as can be approximately seen in Eq. (B.16) or exactly by plotting Eq. (B.14) for different values of n. So, when including the effects of autocorrelation, the optimal bandwidth should be larger than the conventional result. Second, in Figs. B.2 and B.3 we demonstrate numerically how the bandwidth is underestimated with increasing autocorrelation for a given process. Finally, this effect can be proven more generally by inspecting the behaviors of each term in MISE. Comparing relations (1) and (3), for a stationary process the only contributions to the MISE that differ between uncorrelated and autocorrelated processes of the same covariance can be expressed similarly as

$$\underbrace{\frac{1}{(2\pi)^{\frac{q}{2}}}\frac{1}{n^2}\sum_{\tau\neq 0}\frac{n(\tau)}{\sqrt{\det(2\,\boldsymbol{\sigma}_0+2\,\boldsymbol{\sigma}_\mathrm{B})}}}_{\text{uncorrelated}} < \underbrace{\frac{1}{(2\pi)^{\frac{q}{2}}}\frac{1}{n^2}\sum_{\tau\neq 0}\frac{n(\tau)}{\sqrt{\det(2\,\boldsymbol{\sigma}_0+2\,\boldsymbol{\sigma}_\mathrm{B}-2\,\boldsymbol{\sigma}(\tau))}}}_{\text{autocorrelated}}, \tag{B.38}$$

which are both positive and decrease monotonically with increasing bandwidth. Because the presence of autocorrelation shrinks the denominator of the autocorrelated term, this term increases with autocorrelation, particularly at small bandwidths. The effect of this modification to the MISE is to increase the $\varepsilon(\boldsymbol{\sigma}_\mathrm{B})$ curve overall, but more so for smaller values of bandwidth, which then pushes the location of the minimum $\hat{\boldsymbol{\sigma}}_\mathrm{B}$ to a larger value.

In contrast to the uncorrelated case, the asymptotic expansion used for Eq. (B.16) cannot be used to derive a simple rule of thumb expression from Eq. (B.37). This can best be seen from the comparison of uncorrelated and autocorrelated MISE contributions in Eq. (B.38). The uncorrelated contribution can be expanded in a Taylor series about zero bandwidth, as asymptotically we have $\boldsymbol{\sigma}_\mathrm{B} \ll \boldsymbol{\sigma}_0$. However, an analogous Taylor-series expansion in the autocorrelated contribution requires $\boldsymbol{\sigma}_\mathrm{B} \ll \boldsymbol{\gamma}(\tau)$, which is not generally true because the bandwidth and SVF can be comparable at small time lags τ. Therefore, the MISE must be minimized numerically.

Finally, given constant anisotropy in the autocorrelation function, we can express the SVF and bandwidth in terms of the variance.

$$\boldsymbol{\gamma}(\tau) = g(\tau)\,\boldsymbol{\sigma}_0\,, \qquad\qquad \boldsymbol{\sigma}_\mathrm{B} = h^2\,\boldsymbol{\sigma}_0\,, \tag{B.39}$$

reducing the MISE (B.35) to

$$\varepsilon(h) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\sigma}_0)}}\left(\sum_\tau \frac{\frac{n(\tau)}{n^2}}{(2\,g(\tau)+2\,h^2)^{\frac{q}{2}}} - \frac{2}{(2+h^2)^{\frac{q}{2}}} + \frac{1}{2^{\frac{q}{2}}}\right). \tag{B.40}$$

B.3 Confidence-interval estimation

Errors associated with kernel-density estimation that comprise the MISE are of the order $\mathcal{O}(n_A^{-2/(4+q)})$, as determined by the the effective number of degrees of freedom $n_A \sim T/\tau_A$, where T is the total period of data and τ_A denotes the characteristic time lag it takes for

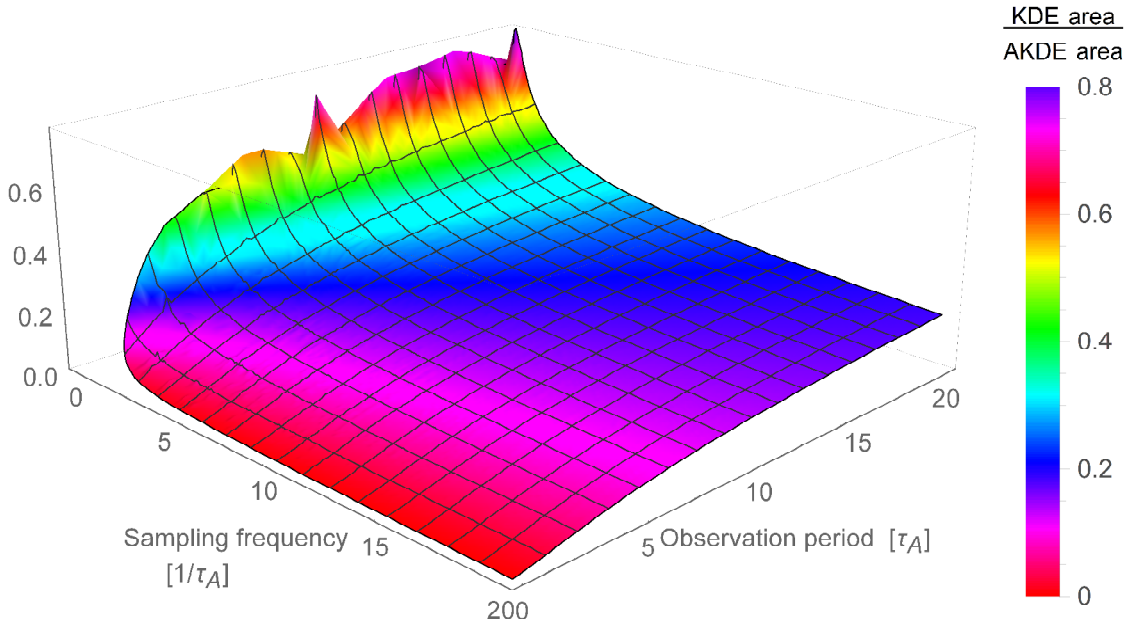

Figure B.2: The ratio of the KDE bandwidth area to the AKDE bandwidth area, as a function of sampling frequency and observation period, relative to the autocorrelation timescale of the movement process τ_A. Bias in the conventional KDE is exacerbated by brief observation periods and high sampling rates, while good conventional estimates can only be obtained with very long spans of coarsely sampled data.

autocorrelations to decay. These errors are asymptotically larger than $\mathcal{O}(n_A^{-1/2})$ — the order of the errors associated with autocorrelation estimation. However, errors in the estimated ACF propagate into errors in the estimated home-range area, in that an underestimated variance or underestimated τ_A will lead to an underestimated home-range area. In contrast, kernel-density errors are local—relative to the bandwidth. Therefore, to place confidence intervals on the home-range size, we will only consider ACF errors, while keeping in mind that the specific divvying of space use is subject to further uncertainty that we briefly consider in App. B.3.1.

For a maximum-likelihood ACF estimate with uncorrelated parameters θ_k, small variations in the ACF parameters $\delta\theta_k$ can easily be propagated into bandwidth variations, with which one can place confidence intervals on the bandwidth using the delta method.

$$\boldsymbol{\sigma}(\tau) + \frac{\partial \boldsymbol{\sigma}(\tau)}{\partial \theta_k}\delta\theta_k \to \boldsymbol{\sigma}_{\mathrm{B}} + \frac{\partial \boldsymbol{\sigma}_{\mathrm{B}}}{\partial \theta_k}\delta\theta_k\,. \tag{B.41}$$

Placing confidence intervals on the bandwidth then allows for confidence intervals to be placed on p that contain only valid PDFs and span the overall range of scale.

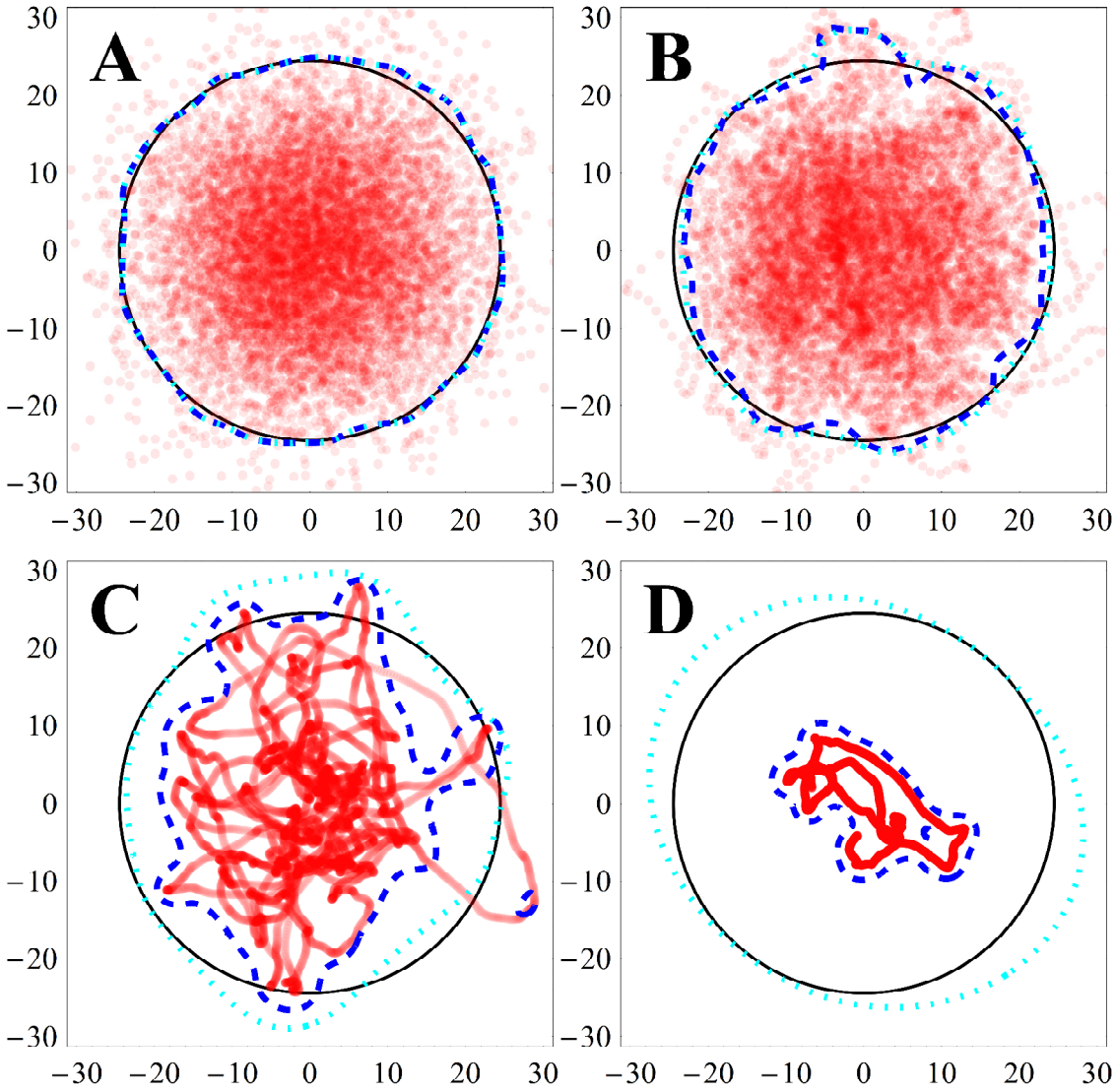

Figure B.3: The progressive degradation of conventional home-range estimates with equal amounts of data (red dots •) and increasing autocorrelation. Outlined are the true (black circle ○), conventional KDE (dashed dark blue –) and AKDE (dotted aqua --) home ranges. In panel **A**, the animal crosses its home range numerous times and the conventional estimate is good. Next in panel **B**, there is still sufficient data to observe the animal in most of its home-range area. The conventional method begins to noticeably underestimate space use, and its confidence intervals will not compensate for this (c.f., Fig. 2). Autocorrelation is further increased in panel **C**, where the limited observation period (relative to the amount of autocorrelation present) yields a poor representation of the entire home range. Finally in panel **D**, the conventional estimate has become uninformative, grossly underestimating the home-range area while the predictions from the AKDE approach continue to reasonably approximate the true range.

B.3.1 Local errors

In kernel-density estimation, confidence intervals are conventionally applied to the PDF estimate. For autocorrelated data, the covariance of $\hat{p}$ is given by

$$\mathrm{COV}[\hat{p}(\mathbf{r}), \hat{p}(\mathbf{r}')] = \langle \hat{p}(\mathbf{r})\, \hat{p}(\mathbf{r}') \rangle - \langle \hat{p}(\mathbf{r}) \rangle \langle \hat{p}(\mathbf{r}') \rangle \ , \tag{B.42}$$

$$= \frac{1}{n^2} \left\langle \sum_{ij} \kappa(\mathbf{r}-\mathbf{r}_i)\, \kappa(\mathbf{r}'-\mathbf{r}_j) \right\rangle - \frac{1}{n^2} \left\langle \sum_{i} \kappa(\mathbf{r}-\mathbf{r}_i) \right\rangle \left\langle \sum_{j} \kappa(\mathbf{r}'-\mathbf{r}_j) \right\rangle \ , \tag{B.43}$$

$$= \frac{1}{n^2} \sum_i \int d^q\mathbf{r}_i\, \kappa(\mathbf{r}-\mathbf{r}_i)\, \kappa(\mathbf{r}'-\mathbf{r}_i)\, p(\mathbf{r}_i) - \frac{1}{n^2} \sum_i \int d^q\mathbf{r}_i\, \kappa(\mathbf{r}-\mathbf{r}_i)\, p(\mathbf{r}_i) \int d^q\mathbf{r}'_i\, \kappa(\mathbf{r}'-\mathbf{r}'_i)\, p(\mathbf{r}'_i)$$

$$+ \frac{1}{n^2} \sum_{i \neq j} \iint d^q\mathbf{r}_i\, d^q\mathbf{r}_j\, \kappa(\mathbf{r}-\mathbf{r}_i)\, \kappa(\mathbf{r}'-\mathbf{r}_j)\, [p(\mathbf{r}_i, \mathbf{r}_j) - p(\mathbf{r}_i)\, p(\mathbf{r}_j)] \ , \tag{B.44}$$

which only differs from the conventional relation by the presence of the final term in Eq. (B.44), which persists as $p(\mathbf{r}_i, \mathbf{r}_j) \neq p(\mathbf{r}_i)\, p(\mathbf{r}_j)$ due to the autocorrelation between $\mathbf{r}_i$ and $\mathbf{r}_j$. In contrast to the MISE calculation, there is no final spatial integration so the reference function approximation to p is illegitimate here. Asymptotically, $\kappa(\mathbf{r})$ naturally limits to a Dirac delta function, which is sufficient for all but the first term of Eq. (B.44). The first term has a pair of kernels under integration, and to evaluate this term asymptotically we represent the kernel product as

$$\kappa(\mathbf{r}-\mathbf{r}_i)\, \kappa(\mathbf{r}'-\mathbf{r}_j) = \frac{e^{-\frac{1}{4}(\mathbf{r}-\mathbf{r}')^{\mathrm{T}} \boldsymbol{\sigma}_{\mathrm{B}}^{-1} (\mathbf{r}-\mathbf{r}')}}{\sqrt{\det(4\pi\boldsymbol{\sigma}_{\mathrm{B}})}} \frac{e^{-\left(\mathbf{r}_i - \frac{\mathbf{r}+\mathbf{r}'}{2}\right)^{\mathrm{T}} \boldsymbol{\sigma}_{\mathrm{B}}^{-1} \left(\mathbf{r}_i - \frac{\mathbf{r}+\mathbf{r}'}{2}\right)}}{\sqrt{\det(\pi\boldsymbol{\sigma}_{\mathrm{B}})}} \ , \tag{B.45}$$

so that the second factor can be evaluated asymptotically under integration. With this, the leading order terms of the covariance are given by

$$\mathrm{COV}[\hat{p}(\mathbf{r}), \hat{p}(\mathbf{r}')] = \frac{1}{n} \frac{e^{-\frac{1}{4}(\mathbf{r}-\mathbf{r}')^{\mathrm{T}} \boldsymbol{\sigma}_{\mathrm{B}}^{-1} (\mathbf{r}-\mathbf{r}')}}{\sqrt{\det 4\pi\boldsymbol{\sigma}_{\mathrm{B}}}} p\left(\frac{\mathbf{r}+\mathbf{r}'}{2}\right) - \frac{1}{n}\, p(\mathbf{r})\, p(\mathbf{r}') + \frac{1}{n^2} \sum_{i \neq j} [p_{ij}(\mathbf{r}, \mathbf{r}') - p(\mathbf{r})\, p(\mathbf{r}')] \ , \tag{B.46}$$

where $p_{ij}(\mathbf{r}, \mathbf{r}')$ denotes the bi-variate density function with correlation $\boldsymbol{\sigma}(t_i, t_j)$. In principle, $p_{ij}(\mathbf{r}, \mathbf{r}')$ can be estimated via AKDE, though every such term would have different optimal bandwidths, giving the entire calculation an $\mathcal{O}(n^3)$ computational cost.

B.4 The limit of continuous monitoring

As previously noted in App. B.1, all optimal, conventional KDEs share the same asymptotic behavior of a bandwidth of $h \propto n^{-1/(4+q)}$ standard deviations in q spatial dimensions, which decreases with an increasing number of data points n. Thus, for any realistic, continuous-velocity process that is sampled continuously, any bandwidth optimization that neglects autocorrelation will predict a distribution singularly defined upon the sampled trajectory. To see this clearly, consider an evenly sampled and fixed period of data, with $T = t_n - t_1$, for a spatially isotropic process with covariance $\boldsymbol{\sigma}_0 = \sigma_0\, \mathbf{I}$. For any realistic, continuous-velocity movement process the animal could have only traveled a finite distance L within this finite amount of time T. The kernel widths around sampled locations asymptotically shrink

according to $w_n = \sqrt{\sigma_0}/n^{1/(4+q)}$, while on average the gaps between subsequent locations asymptotically shrink according to $\ell_n = L/n$. Given that $n^{-1} < n^{-1/(4+q)}$, for sufficiently large n we will have $\ell_n \ll w_n$ and so asymptotically the trajectory is completely filled in by the kernel-density estimator. Although the gaps between sampled locations are filled in by the increased sampling rate, there is nothing to fill in the space between path segments, because the kernels collapse to the sampled locations. Therefore, the density estimate collapses to the movement path and the estimated home-range area vanishes in this limit.

To demonstrate that this pathology does not occur in AKDE, without loss of generality we consider the autocorrelated MISE (B.35) in the case of evenly-sampled data with sampling interval dt, which can be expressed

$$\varepsilon(\boldsymbol{\sigma}_\mathrm{B}) = \frac{1}{(2\pi)^{\frac{q}{2}}}\left(\frac{1}{T^2}\sum_{i=1}^{n} dt \sum_{j=1}^{n} dt \frac{1}{\sqrt{\det(2\,\boldsymbol{\gamma}(t_i - t_j) + 2\,\boldsymbol{\sigma}_\mathrm{B})}} - \frac{2}{\sqrt{\det(2\,\boldsymbol{\sigma}_0 + \boldsymbol{\sigma}_\mathrm{B})}} + \frac{1}{\sqrt{\det(2\,\boldsymbol{\sigma}_0)}}\right), \tag{B.47}$$

and note that asymptotically the double sum limits to the double integral

$$\lim_{dt\to 0} \sum_{i=1}^{n} dt \sum_{j=1}^{n} dt \frac{1}{\sqrt{\det(2\,\boldsymbol{\gamma}(t_i - t_j) + 2\,\boldsymbol{\sigma}_\mathrm{B})}} = \int_0^T dt \int_0^T dt' \frac{1}{\sqrt{\det(2\,\boldsymbol{\gamma}(t - t') + 2\,\boldsymbol{\sigma}_\mathrm{B})}}\,. \tag{B.48}$$

Therefore, asymptotic improvements in the sampling rate have a limited effect on the MISE, and an infinite sampling rate will produce a home-range estimate that is only slightly better than that produced by a large but finite sampling rate.

References

Fleming, C. H., J. M. Calabrese, T. Mueller, K. A. Olson, P. Leimgruber, and W. F. Fagan. 2014. From fine-scale foraging to home ranges: A semi-variance approach to identifying movement modes across spatiotemporal scales. The American Naturalist 183:E154–E167.

Izenman, A. J. 1991. Recent developments in nonparametric density estimation. Journal of the American Statistical Association 86:205–224.

Marcotte, D. 1996. Fast variogram computation with FFT. Computers & Geosciences 22:1175–1186.

Silverman, B. W. 1986. Density estimation for statistics and data analysis. Chapman & Hall.

Turlach, B. A. 1993. Bandwidth selection in kernel density estimation: A review. *In* CORE and Institut de Statistique.

Woodbury, M. A. 1950. Inverting modified matrices. Memorandum report 42:106.