Supplementary Material for “Optimal estimation of drift and diffusion coefficients in the presence of observation noise”

J. Devlin, D. Husmeier and J.A. Mackenzie

1 Mean and variance of the squared displacement

Since the particles are assumed to follow the drift-diffusion stochastic differential equation, the probability density function (PDF) for the observed trajectory \( p(x, t) \) being at location \( x = (x(t), y(t)) \) at time \( t \) is given by

\[
p(x, t) = \frac{1}{2\pi(2Dt + 2\eta^2)} \exp \left( \frac{-|x - \alpha t|^2}{2(2Dt + 2\eta^2)} \right),
\]

where \( D \) is the diffusion coefficient, \( \alpha = \alpha(\cos(\theta_d), \sin(\theta_d)) \) is the drift velocity with drift direction \( \theta_d \), and \( \eta^2 \) is the variance of the observational noise.

The mean-square displacement (MSD) is calculated as

\[
\rho(t) \equiv \mathbb{E}(|X_t|^2) = \int_{\mathbb{R}^2} |x|^2 p(x, t) \, dx.
\]

Substituting in the expression for \( p(x, t) \) gives

\[
\rho(t) = \frac{1}{2\pi(2Dt + 2\eta^2)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x^2 + y^2) \exp \left( \frac{-|x - \alpha t|^2}{2(2Dt + 2\eta^2)} \right) \, dx \, dy.
\]

With \( (\alpha_1, \alpha_2) = (\alpha \cos(\theta_d), \alpha \sin(\theta_d)) \), we use the change of variables \( x = \alpha_1 t + r \cos \theta \) and \( y = \alpha_2 t + r \sin \theta \) for \( 0 \leq r < \infty \) and \( 0 \leq \theta \leq 2\pi \), then

\[
d \, dx \, dy = \begin{vmatrix} x_r & x_y \\ y_r & y_y \end{vmatrix} \, dr \, d\theta = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} \, dr \, d\theta = r \, dr \, d\theta,
\]

and

\[ x^2 + y^2 = \alpha_1^2 t^2 + 2\alpha_1 t r \cos \theta + 2\alpha_2 t r \sin \theta + r^2. \]

When these expressions are substituted into (1), any terms involving \( \cos \theta \) or \( \sin \theta \) will vanish since \( \int_0^{2\pi} \cos \theta \, d\theta = \int_0^{2\pi} \sin \theta \, d\theta = 0 \). For simplification let \( \gamma = 2Dt + 2\eta^2 \), giving

\[
\rho(t) = \alpha^2 t^2 \left( \frac{1}{2\pi \gamma} \int_0^{2\pi} \int_0^\infty r \exp \left( \frac{-r^2}{2\gamma} \right) \, dr \, d\theta \right) + \frac{1}{2\pi \gamma} \int_0^{2\pi} \int_0^\infty r^3 \exp \left( \frac{-r^2}{2\gamma} \right) \, dr \, d\theta.
\]

The two integrals evaluate to give 1 and \( 2\gamma \) respectively and so substituting back in the original expression for \( \gamma \) we obtain

\[
\rho(t) = \alpha^2 t^2 + 4Dt + 4\eta^2.
\]

The variance of the square displacement is given by

\[
\text{Var}(|X_t|^2) \equiv \mathbb{E}(|X_t|^4) - (\mathbb{E}(|X_t|^2))^2.
\]
The latter term is simply the square of the MSD while the former term is calculated as

\[ \mathbb{E}(|X_t|^4) = \int_{\mathbb{R}^2} |x|^4 p(x, t) \, dx \]

\[ = \frac{1}{2\pi^\gamma} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x^2 + y^2)^2 \exp \left( -\frac{|x - \alpha t|^2}{2\gamma} \right) \, dx \, dy. \]

Using the same change of variables as before, let \( x = \alpha_1 t + r \cos \theta \) and \( y = \alpha_2 t + r \sin \theta \) for \( 0 \leq r < \infty \) and \( 0 \leq \theta \leq 2\pi \), then \( dx \, dy = r \, dr \, d\theta \) and

\[(x^2 + y^2)^2 = \alpha_1^2 t^4 + 2\alpha_1^2 \alpha_2 t^3 r \cos \theta + 2\alpha_1^2 \alpha_2 t^3 r \sin \theta + \alpha_2^2 t^2 r^2 + 2\alpha_1 \alpha_2 t^2 r \cos \theta \sin \theta + \alpha_1 \alpha_2 t^2 r \sin \theta \cos \theta + 2\alpha_2 t^3 r \sin \theta + \alpha_2 t^3 r \cos \theta + \alpha_2 t^3 r \sin \theta + r^4.\]

As previously, any terms involving \( \cos \theta, \sin \theta \) and additionally \( \sin \theta \cos \theta \) will be equal to zero once integrated. This leaves

\[ \mathbb{E}(|X_t|^4) = \alpha_1^4 t^4 \left( \frac{1}{2\pi^\gamma} \int_0^{2\pi} \int_0^\infty r \exp \left( -\frac{r^2}{2\gamma} \right) \, dr \, d\theta \right) \]

\[ + 2\alpha_1^2 t^2 \left( \frac{1}{2\pi^\gamma} \int_0^{2\pi} \int_0^\infty r^3 \exp \left( -\frac{r^2}{2\gamma} \right) \, dr \, d\theta \right) \]

\[ + 4\alpha_1^2 t^2 \left( \frac{1}{2\pi^\gamma} \int_0^{2\pi} \int_0^\infty \cos^2 \theta r^3 \exp \left( -\frac{r^2}{2\gamma} \right) \, dr \, d\theta \right) \]

\[ + 4\alpha_2^2 t^2 \left( \frac{1}{2\pi^\gamma} \int_0^{2\pi} \int_0^\infty \sin^2 \theta r^3 \exp \left( -\frac{r^2}{2\gamma} \right) \, dr \, d\theta \right) \]

\[ + \frac{1}{2\pi^\gamma} \int_0^{2\pi} \int_0^\infty r^5 \exp \left( -\frac{r^2}{2\gamma} \right) \, dr \, d\theta. \]

These integrals are equal to 1, 2\( \gamma \), \( \gamma \), \( \gamma \) and \( 8\gamma^2 \), respectively. Therefore, evaluating these integrals we obtain

\[ \mathbb{E}(|X_t|^4) = \alpha_1^4 t^4 + 4\alpha_1^2 t^2 \gamma \alpha_2 + 4\alpha_1^2 t^2 \gamma + 4\alpha_2^2 t^2 \gamma + 8\gamma^2 \]

\[ = \alpha_1^4 t^4 + 8\alpha_2 t^2 \gamma + 8\gamma^2. \]

Therefore, the variance of the square displacement is given by

\[ \text{Var}(|X_t|^2) = \mathbb{E}(|X_t|^4) - (\mathbb{E}(|X_t|^2))^2 \]

\[ = \alpha_1^4 t^4 + 8\alpha_2 t^2 \gamma + 8\gamma^2 - (\alpha_2^2 t^2 + 2\gamma)^2 \]

\[ = 4\alpha_2^2 t^2 \gamma + 4\gamma^2 \]

\[ = 4\alpha_2^2 t^2 (2Dt + 2\eta^2) + 4(2Dt + 2\eta^2)^2. \] (2)
2 Derivation of the variance of the MSD

The calculation of the variance of the MSD is trickier and uses properties of Weiner processes. First of all, we calculate the variance of the MSD for a single particle, denoted by \( \sigma^2_n(\sigma) \). Then, due to the independence of the \( N_S \) trajectories, the variance of the MSD for an ensemble of particles is given by

\[
\sigma^2_n = \frac{\sigma^2_n(\sigma)}{N_S}.
\]

By denoting

\[
\rho^{(1)}_n = \frac{1}{N + 1 - n} \sum_{i=1}^{N+1-n} |x_{i+n} - x_i|^2, \quad n = 1, \ldots, N,
\]

as the overlapping MSD for a single particle, then the variance of the MSD is defined as

\[
\sigma^2_n(\sigma) = \mathbb{E}((\rho^{(1)}_n)^2) - (\mathbb{E}(\rho^{(1)}_n))^2.
\]

Letting \( K = N + 1 - n \) be the number of samples of squared displacements of length \( n\Delta t \), then, due to the correlation between overlapping displacements, we have that

\[
\sigma^2_n(\sigma) = \frac{\text{Var}(|X_{i}|^2)(n\Delta t)}{K} + \frac{2}{K^2} \sum_{i=1}^{K} \sum_{j=i+1}^{K} \text{Cov}(i, j), \quad (i < j)
\]

where \( \text{Var}(|X_{i}|^2)(n\Delta t) \) is the variance of the squared displacement at time point \( n\Delta t \) and

\[
\text{Cov}(i, j) = \mathbb{E}(|x_{i+n} - x_i|^2 | x_{j+n} - x_j|^2) - \mathbb{E}(|x_{i+n} - x_i|^2) \mathbb{E}(|x_{j+n} - x_j|^2)
\]

is the covariance between squared displacements. The first term in (3) is known and so in order to calculate the variance of the MSD we need only to calculate the covariance of the squared displacements. For this we have to consider the different ways the displacements \( |x_{i+n} - x_i|^2 \) and \( |x_{j+n} - x_j|^2 \) can overlap. This can be split into two cases as shown in Figure 1, where \( \mathbf{R}_1, \mathbf{R}_2 \) and \( \mathbf{R}_3 \) are the displacements between the given vertices. We can then write the covariance as

\[
\begin{align*}
(a) \quad \text{Cov}(i, j) &= \mathbb{E}(\mathbf{R}_2^2 \mathbf{R}_3^2) - \mathbb{E}(\mathbf{R}_1^2) \mathbb{E}(\mathbf{R}_3^2), \quad j > i + n \\
(b) \quad \text{Cov}(i, j) &= \mathbb{E}((\mathbf{R}_1 + \mathbf{R}_2)^2 (\mathbf{R}_2 + \mathbf{R}_3)^2) - \mathbb{E}((\mathbf{R}_1 + \mathbf{R}_2)^2) \mathbb{E}((\mathbf{R}_2 + \mathbf{R}_3)^2), \quad j \leq i + n
\end{align*}
\]

Notice that when \( j > i + n \) (case (a) in Figure 1) the trajectories do not overlap and so \( \mathbf{R}_2 = 0 \). Since the remaining displacements \( \mathbf{R}_1 \) and \( \mathbf{R}_3 \) are independent then \( \text{Cov}(i, j) = 0 \). Therefore, we need only consider the case when \( j \leq i + n \) (case (b) in Figure 1). In this case, we can write the displacement vectors as
Note that when \( \eta \) is the observational noise at time \( t_i \), etc. Note that when \( j = i + n \), we have that \( \mathbf{R}_2 = 0 \). Expanding the right hand side of equation (4b) results in

\[
\text{Cov}(i,j) = \mathbb{E}(|\mathbf{R}_1|^2|\mathbf{R}_2|^2) + 2\mathbb{E}(|\mathbf{R}_1|^2(\mathbf{R}_2 \cdot \mathbf{R}_3)) + \mathbb{E}(|\mathbf{R}_1|^2|\mathbf{R}_3|^2)
+ 2\mathbb{E}(|\mathbf{R}_2|^2(\mathbf{R}_1 \cdot \mathbf{R}_2)) + 4\mathbb{E}((\mathbf{R}_1 \cdot \mathbf{R}_2)(\mathbf{R}_2 \cdot \mathbf{R}_3)) + 2\mathbb{E}(|\mathbf{R}_3|^2(\mathbf{R}_1 \cdot \mathbf{R}_2))
+ \mathbb{E}(|\mathbf{R}_2|^4) + 2\mathbb{E}(|\mathbf{R}_2|^2(\mathbf{R}_2 \cdot \mathbf{R}_3)) + \mathbb{E}(|R_2|^2|R_3|^2)
- \left( \mathbb{E}(|\mathbf{R}_1|^2)\mathbb{E}(|\mathbf{R}_2|^2) + 2\mathbb{E}(|\mathbf{R}_1|^2)\mathbb{E}(\mathbf{R}_2 \cdot \mathbf{R}_3) + \mathbb{E}(|\mathbf{R}_1|^2)\mathbb{E}(|\mathbf{R}_3|^2) \right)
+ 2\mathbb{E}(|\mathbf{R}_2|^2)\mathbb{E}(\mathbf{R}_1 \cdot \mathbf{R}_2) + 4\mathbb{E}(\mathbf{R}_1 \cdot \mathbf{R}_2)\mathbb{E}(\mathbf{R}_2 \cdot \mathbf{R}_3) + 2\mathbb{E}(|\mathbf{R}_3|^2)\mathbb{E}(\mathbf{R}_1 \cdot \mathbf{R}_2)
+ (\mathbb{E}(|\mathbf{R}_2|^2))^2 + 2\mathbb{E}(|\mathbf{R}_2|^2)\mathbb{E}(\mathbf{R}_2 \cdot \mathbf{R}_3) + \mathbb{E}(|\mathbf{R}_2|^2)\mathbb{E}(|\mathbf{R}_3|^2)).
\]

The first term is calculated as

\[
\mathbb{E}(|\mathbf{R}_1|^2|\mathbf{R}_2|^2)
= \mathbb{E}(|\alpha \, dt_1 + \sqrt{2D} \, dW_1 + \psi_1|^2|\alpha \, dt_2 + \sqrt{2D} \, dW_2 + \psi_2|^2)
= \mathbb{E}((|\alpha|^2 dt_1^2 + 2\sqrt{2D} dt_1(\alpha \cdot dW_1) + 2 dt_1(\alpha \cdot \psi_1) + 2D|dW_1|^2 + 2\sqrt{2D}(dW_1 \cdot \psi_1) + |\psi_1|^2)
\times (|\alpha|^2 dt_2^2 + 2\sqrt{2D} dt_2(\alpha \cdot dW_2) + 2 dt_2(\alpha \cdot \psi_2) + 2D|dW_2|^2 + 2\sqrt{2D}(dW_2 \cdot \psi_2) + |\psi_2|^2)).
\]
Multiplying out the brackets and taking the expectation of each term gives
\[
\mathbb{E}(|R_1|^2|R_2|^2) = \alpha^2 \mathbb{E}(dt_1^2 dt_2^2) + 2\sqrt{2}D \alpha^2 \mathbb{E}(dt_1^2 dW_1^2) + 2\alpha^2 \mathbb{E}(dt_2^2 (\alpha \cdot dW_2)) + 2D \alpha^2 \mathbb{E}(dt_1^2 (\alpha \cdot dW_2)) + 4\alpha^2 \mathbb{E}(dt_1^2 |\psi_2|^2) + 2\sqrt{2}D \alpha^2 \mathbb{E}(dt_1 dt_2^2 (\alpha \cdot dW_1)) + 8D \mathbb{E}(dt_1 dt_2 (\alpha \cdot dW_2)(\alpha \cdot dW_2)) + 4\sqrt{2}D \mathbb{E}(dt_1 dt_2 (\alpha \cdot dW_1)(\alpha \cdot \psi_2)) + 4D \sqrt{2}D \mathbb{E}(dt_1 dW_2^2 (\alpha \cdot dW_1)) + 8D \mathbb{E}(dt_1 (\alpha \cdot dW_1)(dW_2 \cdot \psi_2)) + 2\sqrt{2}D \mathbb{E}(dt_1 |\psi_2|^2 (\alpha \cdot dW_1)) + 2\alpha^2 \mathbb{E}(dt_1 dt_2^2 (\alpha \cdot \psi_1)) + 4\sqrt{2}D \mathbb{E}(dt_1 dt_2^2 (\alpha \cdot \psi_1)(\alpha \cdot dW_2)) + 4\alpha^2 \mathbb{E}(dt_1 dt_2 (\alpha \cdot \psi_1)(\alpha \cdot \psi_2)) + 4D \mathbb{E}(dt_1 dW_2^2 (\alpha \cdot \psi_1)) + 4\sqrt{2}D \mathbb{E}(dt_1 (\alpha \cdot \psi_1)(dW_2 \cdot \psi_2)) + 2\mathbb{E}(dt_1 |\psi_2|^2 (\alpha \cdot \psi_1)) + 2D \alpha^2 \mathbb{E}(dt_2^2 |dW_1|^2) + 4\sqrt{2}D \mathbb{E}(dt_2 dW_1^2 (\alpha \cdot dW_2)) + 4D \mathbb{E}(dt_2 dW_1^2 (\alpha \cdot \psi_2)) + 4D \mathbb{E}(|dW_1|^2 |dW_2|^2) + 4\sqrt{2}D \mathbb{E}(dW_1^2 |dW_2^2 |\psi_2)^2) + 2D \mathbb{E}(dW_1^2 |dW_2^2 |\psi_2)^2) + 2\sqrt{2}D \mathbb{E}((\alpha \cdot \psi_1)(\alpha \cdot \psi_2)) + 2D \mathbb{E}(|\psi_1|^2 |\psi_2|^2) + 2\sqrt{2}D \mathbb{E}(|\psi_1|^2 |\psi_2|^2) + \mathbb{E}(|\psi_1|^2 |\psi_2|^2).
\]

From here, we note that \(\mathbb{E}(dt_i) = t_i \) for \(i \in \{1, 2, 3\}\), and since the Wiener processes are independent, then \((dW_i \cdot dW_j) = 0 \) for \(i \neq j\). Furthermore, since the Wiener processes and noise terms are independent then \((dW_i \cdot \psi_j) = 0 \) for \(i, j \in \{1, 2, 3\}\). However, due to the noise terms containing variances of the observational noise which can share a common vertex, then \((\psi_i \cdot \psi_j) \neq 0 \) for \(|i - j| = 1\). Using these properties, many of the terms in (6) will be zero. Therefore, we are left with
\[
\mathbb{E}(|R_1|^2|R_2|^2) = \alpha^4 t_1^2 t_2 + 2D \alpha^2 t_1^2 dW_1^2 + \alpha^2 t_1^2 |\psi_2|^2 + 4t_1 t_2 \mathbb{E}((\alpha \cdot \psi_1)(\alpha \cdot \psi_2)) + 2D \alpha^2 t_2^2 |dW_1|^2 + 4D \mathbb{E}(|dW_1|^2 |dW_2|^2) + 2D \mathbb{E}(|dW_1|^2 |\psi_2|^2) + \alpha^2 t_2^2 |\psi_1|^2 + 2D \mathbb{E}(|dW_2|^2 |\psi_1|^2) + \mathbb{E}(|\psi_1|^2 |\psi_2|^2).
\]
We have that $dW_i = (dW_i^1, dW_i^2)$ for $i \in \{1, 2, 3\}$, where $dW_i^1, dW_i^2$ are independent and identically distributed (i.i.d.) normal random variables with zero mean and variance $\sqrt{\eta_i}$, i.e. $dW_i^1, dW_i^2$ are of the form $\mathcal{N}(0, \sqrt{\eta_i})$. Similarly for the noise term, we have, for example, $\psi_j = \eta_{ij} - \eta_i$, where $\eta_i = (\eta_i^1, \eta_i^2)$ and $\eta_i^1, \eta_i^2$ are also i.i.d. random variables of the form $\mathcal{N}(0, \eta^2)$. These can be used to calculate the final expectation terms left in (7). This gives

$$
\mathbb{E}(|R_1|^2|R_2|^2) = \alpha^4t_1^2t_2^2 + 4\alpha^2\eta^2t_1^2
- 4\alpha^2\eta^2t_1t_2 + 4\alpha^2\eta^2t_1^2 + 16D^2t_1t_2 + 16D\eta^2t_1
+ 4\alpha^2\eta^2t_2^2 + 16D\eta^2t_2 + 20\eta^4.
$$

The rest of the expectations will not be shown here but use similar algebra. Instead we will simply present the results.

$$
\mathbb{E}(R_i \cdot R_j) = \alpha^2t_it_j - 2\eta^2, \quad |i - j| = 1
$$

$$
\mathbb{E}(R_1 \cdot R_3) = \alpha^2t_1t_3
$$

$$
\mathbb{E}(|R_i|^2) = \alpha^2t_i^2 + 4Dt_i + 4\eta^2, \quad i \in \{1, 2, 3\}
$$

$$
\mathbb{E}(|R_2|^4) = \alpha^4t_2^4 + 16D\alpha^2t_2^2 + 16\alpha^2\eta^2t_2
+ 32D^2t_2^2 + 64D\eta^2t_2 + 32\eta^4,
$$

$$
\mathbb{E}(|R_i|^2|R_j|^2) = \alpha^4t_i^2t_j^2 + 4\alpha^2\eta^2t_i^2
- 4\alpha^2\eta^2t_it_j + 4\alpha^2\eta^2t_j^2 + 16D^2t_it_j + 16D\eta^2t_i
+ 4\alpha^2\eta^2t_i^2 + 16D\eta^2t_j + 20\eta^4 \quad |i - j| = 1
$$

$$
\mathbb{E}(|R_1|^2|R_3|^2) = \alpha^4t_1^2t_3^2 + 4\alpha^2\eta^2t_1^2
+ 4\alpha^2\eta^2t_1t_3 + 16D^2t_1t_3 + 16D\eta^2t_1 + 4\alpha^2\eta^2t_3
+ 16D\eta^2t_3 + 16\eta^4 + (4\eta^4 - 4\alpha^2\eta^2t_1t_3)\delta_{j,i+n},
$$

$$
\mathbb{E}(|R_i|^2(R_i \cdot R_j)) = \alpha^4t_i^2t_j - 4\alpha^2\eta^2t_i^2 + 8D\alpha^2t_i^2t_j
- 16D\eta^2t_i + 8\alpha^2\eta^2t_it_j - 16\eta^4, \quad i \neq j
$$

$$
\mathbb{E}(|R_i|^2(R_j \cdot R_k)) = \alpha^4t_i^2t_jt_k + 4\alpha^2\eta^2tjit_k + 4\alpha^2\eta^2t_jt_k - 8D\eta^2t_i
- 8\eta^4 - 2\alpha^2\eta^2t_i^2 - 2\alpha^2\eta^2t_itk, \quad i \neq j \neq k, i \neq 2
$$

$$
\mathbb{E}(|R_2|^2(R_1 \cdot R_3)) = \alpha^4t_1^2t_3^2 + 4\alpha^2\eta^2t_1t_3 + 4\alpha^2\eta^2t_1t_3
+ 4\eta^4 - 2\alpha^2\eta^2t_1t_3 - 2\alpha^2\eta^2t_2t_3,
$$

$$
\mathbb{E}((R_1 \cdot R_2)(R_2 \cdot R_3)) = \alpha^4t_1^2t_2^2t_3 - 3\alpha^2\eta^2t_1t_2 - 3\alpha^2\eta^2t_2t_3
+ 2\alpha^2\eta^2t_1t_3 + 2D\alpha^2t_1t_2t_3 + 6\eta^4. \quad (8)
$$
Here,

\[ \delta_{j,i+n} = \begin{cases} 
0 & j \neq i + n \\
1 & j = i + n
\end{cases} \]

represents the Kronecker delta function. In the absence of drift we recover the same expressions as found in the supplementary material of Michalet’s paper [?]. Substituting these results into (5) and simplifying yields

\[ \text{Cov}(i,j) \equiv \text{Cov}(l) = 16D^2((n-l)\Delta t)^2 + 8D\alpha^2n^2(n-l)(\Delta t)^3 + (4\eta^4 - 4(\alpha\eta\Delta t)^2)\delta_{j,i+n}, \quad (9) \]

where \( l = j - i \). Now that the covariance has been calculated, to obtain the variance of the MSD we must sum all the non-zero terms appearing in the double sum in (3). This can be split into two cases. When \( n \leq K \) we sum the non-zero terms along the diagonals \( l = 1, \ldots, n - 1 \) of the covariance matrix. The \( l^{th} \) diagonal contains \( K - l \) identical elements, and therefore the non-zero terms are

\[ \sum_{l=1}^{n-1} (K-l)\text{Cov}(l), \quad (10) \]

where \( \text{Cov}(l) = 16D^2((n-l)\Delta t)^2 + 8D\alpha^2n^2(n-l)(\Delta t)^3 \). When \( l = n \) this corresponds to the case when \( j = i + n \) and so the Kronecker delta term will now be non-zero. This diagonal will contain \( K-n = N + 1 - 2n \) elements, and so will equal \((N + 1 - 2n)(4\eta^4 - 4(\alpha\eta\Delta t)^2)\). When \( n > K \) we don’t have any segments with no overlaps. Therefore, we only have overlapping contributions to the covariance matrix. This time the upper limit in the sum (10) is \( N - n \) rather than \( n - 1 \). So when \( n > K \) we have

\[ \sum_{l=1}^{N-n} (K-l)\text{Cov}(l), \]

where \( \text{Cov}(l) = 16D^2((n-l)\Delta t)^2 + 8D\alpha^2n^2(n-l)(\Delta t)^3 \). Summing all these terms in either case \((n \leq K \text{ and } n > K)\) as above allows us to calculate \( \sigma_n^2 \). It can be shown that the final result is

\[ \sigma_n^2 = \begin{cases} 
\left[ \frac{n}{6K^2}(4n^2K + 2K - n^3 + n)16D^2(\Delta t)^2 + 8\alpha^2D(\Delta t)^3 \left( \frac{n^3}{3K} (3Kn + 1 - n^2) \right) \right] / N_s & n \leq K \\
\left[ \frac{1}{6K}(6n^2K - 4nK^2 + 4n + K^3 - K)16D^2(\Delta t)^2 + 8\alpha^2D(\Delta t)^3 \left( \frac{n^2}{3K} (3nK - K^2 + 1) \right) \right] / N_s & n > K
\end{cases} \quad (11) \]

In Figure 2 we plot the theoretical variance of the MSD (11) with an empirical estimate averaged over 1000 and 10,000 samples for the parameter values \( D = 2 \mu m^2/s, \alpha = 1 \mu m/s, \eta = 2 \mu m, N_S = 10, N = 100 \) and \( T = 100 \) s. We can see that as we increase the number of samples, the empirical estimate gets closer to the theoretical expression (11).
3 Derivation of the covariance of the MSD

A theoretical expression for the covariance is needed to calculate the uncertainties in the regression parameters. As before, we consider the covariance of the MSD for a single particle, denoted by $\sigma_{nm}^2(\text{nm})$. Then, the covariance of the MSD for an ensemble of particles will be

$$\sigma_{nm}^2 = \frac{\sigma_{nm}^2(\text{nm})}{N_S}.$$

From the definition of the covariance, we have that

$$\sigma_{nm}^2(\text{nm}) = \mathbb{E}(\rho_n^{(1)}\rho_m^{(1)}) - \mathbb{E}(\rho_n^{(1)})\mathbb{E}(\rho_m^{(1)}).$$

The process of calculating the covariance is similar to that of deriving the formula for the variance of the MSD. First, we assume that $m > n$, then letting $K = N+1-n$ and $P = N+1-m$ we have

$$\sigma_{nm}^2 = \frac{1}{KP} \sum_{i=1}^{K} \sum_{j=1}^{P} \text{Cov}(i, j),$$

where

$$\text{Cov}(i, j) = \mathbb{E}(|x_{i+n} - x_i|^2|x_{j+m} - x_j|^2) - \mathbb{E}(|x_{i+n} - x_i|^2)\mathbb{E}(|x_{j+m} - x_j|^2).$$

As before, we have to consider the different ways the trajectories can overlap. This time we consider the five cases as shown in Figure 3. The covariance for each case can be re-written as

(a) $\text{Cov}(i, j) = \mathbb{E}(R_1^2R_3^2) - \mathbb{E}(R_1^2)\mathbb{E}(R_3^2) \quad i < i + n < j < j + m$

(b) $\text{Cov}(i, j) = \mathbb{E}((R_1 + R_2)^2(R_2 + R_3)^2) - \mathbb{E}((R_1 + R_2)^2)\mathbb{E}((R_2 + R_3)^2). \quad i < j \leq i + n < j + m$

(c) $\text{Cov}(i, j) = \mathbb{E}(R_2^2|R_1 + R_2 + R_3)^2 - \mathbb{E}(R_2^2)\mathbb{E}((R_1 + R_2 + R_3)^2) \quad j \leq i < i + n < j + m$

(d) $\text{Cov}(i, j) = \mathbb{E}((R_1 + R_2)^2(R_2 + R_3)^2) - \mathbb{E}((R_1 + R_2)^2)\mathbb{E}((R_2 + R_3)^2) \quad j < i < j + m \leq i + n$

(e) $\text{Cov}(i, j) = \mathbb{E}(R_1^2R_3^2) - \mathbb{E}(R_1^2)\mathbb{E}(R_3^2) \quad j < j + m \leq i < i + n$
These expectations can be calculated using the results of (8). From this we obtain

(a) \( \text{Cov}(l_1) = 0 \)

(b) \( \text{Cov}(l_2) = 16D^2((n - l_2)\Delta t)^2 + 8D\alpha^2 mn(n - l_2)(\Delta t)^3 \\
+ (4\eta^4 - 4\alpha^2\eta^2 mn(\Delta t)^2)\delta_{j,i+n} \), where \( l_2 = j - i \).

(c) \( \text{Cov}(l_3) = 16D^2(n\Delta t)^2 + 8D\alpha^2 n^2 m(\Delta t)^3 \\
+ (16D\eta^2 n\Delta t + 4\eta^4 + 4\alpha^2\eta^2 mn(\Delta t)^2)\delta_{i,j} \).

(d) \( \text{Cov}(l_4) = 16D^2(l_4\Delta t)^2 + 8D\alpha^2 mn l_4(\Delta t)^3 \\
+ (16D\eta^2 n\Delta t + 4\eta^4 + 4\alpha^2\eta^2 mn(\Delta t)^2)\delta_{i+n,j+m} \), where \( l_4 = j + m - i \).

(e) \( \text{Cov}(l_5) = (4\eta^4 - 4\alpha^2\eta^2 mn(\Delta t)^2)\delta_{i,j+m} \).

We need to sum these different covariance values to obtain the overall covariance of the MSD in the two cases \( m + n \leq N \) and \( m + n > N \). As before, this is done by summing the non-zero terms along diagonals in the covariance matrix. Also, note the Kronecker delta terms must be
considered separately. The sums are done as follows

\[ m + n \leq N \]

(a) \( 0. \)

(b) \( \left( \sum_{l_2=1}^{n-1} (P - l_2)\text{Cov}(l_2) \right) + (P - n)(4\eta^4 - 4\alpha^2\eta^2mn(\Delta t)^2). \)

(c) \( \left( \sum_{l_3=1}^{m-n+1} \text{PCov}(l_3) \right) + P(16\eta^2n\Delta t + 4\eta^4 + 4\alpha^2\eta^2mn(\Delta t)^2). \)

(d) \( \left( \sum_{l_4=1}^{n-1} (P - (n - l_4))\text{Cov}(l_4) \right) + P(16\eta^2n\Delta t + 4\eta^4 + 4\alpha^2\eta^2mn(\Delta t)^2). \)

(e) \( (P - n)(4\eta^4 - 4\alpha^2\eta^2mn(\Delta t)^2). \)

\[ m + n > N \]

(a) \( 0. \)

(b) \( \left( \sum_{l_2=1}^{P-1} (P - l_2)\text{Cov}(l_2) \right) + (P - n)(4\eta^4 - 4\alpha^2\eta^2mn(\Delta t)^2). \)

(c) \( \left( \sum_{l_3=1}^{m-n+1} \text{PCov}(l_3) \right) + P(16\eta^2n\Delta t + 4\eta^4 + 4\alpha^2\eta^2mn(\Delta t)^2). \)

(d) \( \left( \sum_{l_4=n+1-P}^{n-1} (P - (n - l_4))\text{Cov}(l_4) \right) + P(16\eta^2n\Delta t + 4\eta^4 + 4\alpha^2\eta^2mn(\Delta t)^2). \)

\( \sigma_{nm}^2 = \left\{ \begin{array}{ll}
\frac{16nD^2(\Delta t)^2}{6KP} \left( -n^3 - 2Pn^2 + (1 - 6m^2 + 6(N + 1)m)n + 2P \right) \\
+ \frac{8\alpha^2(\Delta t)^3mn^2D}{3KP} \left( -n^2 + 3(-m^2 + (N + 1)m + 1/3) \right) + \frac{32\eta^2nD\Delta t}{K} \\
+ \frac{8\eta^4(-n+2P)}{KP} + \frac{8\alpha^2(\Delta t)^2mn^2\eta^2}{KP} \right/m + n \leq N \\
\frac{8D^2(\Delta t)^2}{3K} \left( -m^3 + (3 + 3N - 4n)m^2 + (8(N + 1)n - 2 - 3N^2 - 6N)m \\
- 6n^3 + 6(N + 1)n^2 - (4N^2 + 8N)n + N(N + 2)(N + 1) \right) \\
+ \frac{8\alpha^2D(\Delta t)^3mn}{3K} \left( m^2 - 2(N + 1)m + 3n^2 - 3(N + 1)nN^2 + 2N \right) \\
+ \frac{8\eta^4}{K} \left( \alpha^2(\Delta t)^2mn + 4Dn\Delta t + \eta^2 \right) \right/m + n > N \\
\end{array} \right. \\
(12) \]
Figure 4 shows the comparison between the theoretical covariance of the MSD (12) with its empirical estimate, along with a cross section of the covariance of the MSD averaged over 1000 and 10,000 samples, for the parameters values $D = 2 \mu m^2/s$, $\alpha = 1 \mu m/s$, $\eta = 2 \mu m$, $N_S = 10$, $N = 100$ and $T = 100 \, s$. Note that we only provide one plot of the full covariance because the plots for the two different number of particles are very similar. Again, we see good agreement between the theoretical covariance of the MSD (12) and its empirical estimate.

Figure 4: A plot comparing the theoretical covariance of the MSD (12) (top left) and its empirical estimate (top right) averaged over 1000 samples, along with a plot of a cross section along $m = 40$ of the theoretical covariance of the MSD (— solid black line) with its empirical estimate (- - - dashed red line) averaged over 1000 (bottom left) and 10,000 (bottom right) samples for $D = 2 \mu m^2/s$, $\alpha = 1 \mu m/s$, $\eta = 2 \mu m$, $N_S = 10$, $N = 100$ and $T = 100 \, s$. Note that the jump in the bottom plots are a consequence of only calculating the upper triangular part of the covariance matrix.
4 Further fitting points experiments

4.1 Existence of $p_{opt}$ for different parameter values

In the main paper, all ensemble-averaged calculations used $N_S = 10$ trajectories calculated at $N = 100$ time points. To test the robustness of our calculations, we now test with $N_S = 100$ and $N = 10$. Figure 5 shows the theoretical and simulated uncertainty $\sigma_b/b + \sigma_c/c$ as a function of the number of fitting points. These experiments were for $D = 2 \mu m^2/s$, $\alpha = 1 \mu m/s$, $\eta = 2 \mu m$, $N_S = 100$ and $N = 10$, with $\Delta t = 1 s$ giving $T = 10 s$ for the left plot, while $\Delta t = 10 s$ giving $T = 100 s$ for the right plot. Here we see a similar result as that in Figure 2 in the main paper. We have that, for a small value of $\Delta t$, the optimal estimate of the parameters comes from fitting with all the MSD points. As $\Delta t$ is increased, we see that the optimal inference comes from fitting with a subset of points; specifically, for all the cases tested, $p_{opt} = 9$.

![Figure 5: A plot of the theoretical value of $\sigma_b/b + \sigma_c/c$ (— solid lines) and its empirically estimated value using 1000 samples (- - - dashed lines) when fit with the first $p$ MSD points for $\eta = 0.5 \mu m$, $2 \mu m$ and $8 \mu m$. These experiments were for $D = 2 \mu m^2/s$, $\alpha = 1 \mu m/s$, $N_S = 100$ and $N = 10$, for $\Delta t = 1 s$ giving $T = 10 s$ (left), and $\Delta t = 10 s$ giving $T = 100 s$ (right). For $\eta = 0.5 \mu m$, $2 \mu m$ and $8 \mu m$, the optimal number of fitting points are 10 for each (left) and 9 for each (right).](image)

We also looked for the existence of an optimal number of fitting points when inferring the diffusion coefficient $D$, the drift magnitude $\alpha$ and the observational noise $\eta$. Analogous to the two parameter case, we look to minimise the uncertainty $\sigma_a/a + \sigma_b/b + \sigma_c/c$ since the noise term is related to the regression coefficient $a$. Figure 6 shows the theoretical and simulated value of $\sigma_a/a + \sigma_b/b + \sigma_c/c$ as a function of the number of fitting points $p$ for $D = 2 \mu m^2/s$, $\alpha = 1 \mu m/s$, $\eta = 2 \mu m$, $N_S = 10$ and $N = 100$, with $\Delta t = 1 s$ giving $T = 100 s$ for the left plot, while $\Delta t = 10 s$ giving $T = 1000 s$ for the right plot. Here we see that the level of uncertainty, as well as the optimal number of fitting points, cannot be easily predicted by the value of $\eta$. 

![Figure 6: A plot of the theoretical value of $\sigma_a/a + \sigma_b/b + \sigma_c/c$ (— solid lines) and its empirically estimated value using 1000 samples (- - - dashed lines) when fit with the first $p$ MSD points for $\eta = 0.5 \mu m$, $2 \mu m$ and $8 \mu m$. These experiments were for $D = 2 \mu m^2/s$, $\alpha = 1 \mu m/s$, $N_S = 100$ and $N = 100$, with $\Delta t = 1 s$ giving $T = 100 s$ (left), and $\Delta t = 10 s$ giving $T = 1000 s$ (right). For $\eta = 0.5 \mu m$, $2 \mu m$ and $8 \mu m$, the optimal number of fitting points are 10 for each (left) and 9 for each (right).](image)
Figure 6: A plot of the theoretical value of $\sigma_a/a + \sigma_b/b + \sigma_c/c$ (solid lines) and its empirically estimated value using 1000 samples (dashed lines) when fit with the first $p$ MSD points for $\eta = 0.5 \mu m, 2 \mu m$ and $8 \mu m$. These experiments were for $D = 2 \mu m^2/s, \alpha = 1 \mu m/s, N_S = 10$ and $N = 100$, for $\Delta t = 1 s$ giving $T = 100 s$ (left), and $\Delta t = 10 s$ giving $T = 1000 s$ (right). For $\eta = 0.5 \mu m, 2 \mu m$ and $8 \mu m$, the optimal number of fitting points are 11, 100 and 100, respectively (left) and 6, 8 and 100, respectively (right).

### 4.2 MATLAB routine to calculate optimal number of fitting points

```matlab
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code to calculate the optimal number of fitting points for the %
% overlapping time-averaged mean-square displacement (MSD). %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Input:  D — Estimate of the diffusion coefficient.
%         al — Estimate of the drift magnitude.
%         eta — Estimate of the observational noise.
%         dt — The time interval between frames.
%         N — The number of time points.
% %
% Output: popt — The optimal number of fitting points for the given %
%          parameter values.
% %
% Notes — This code works for the overlapping time-averaged MSD %
%          calculated both from a single particle and an ensemble of %
%          particles as they have the same optimal number of fitting %
%          points.
```

13
function [popt] = optimal_preg(D,al,eta,dt,N)

T=N*dt;
alp=4*D*dt;
ep=4*eta^2;

% Calculate the variance of the MSD
%
for n=1:N
    t(n)=n*dt;
    K=N+1−n;
    if n<=K
        varmsd(n)=(n*(4*n^2*K+2*K−n^3+n)/(6*K^2))*(4*D*dt)^2+...
                  8*al^2*D*(dt)^3*(n^2*(3*K*n+1−n^2)/(3*K^2));
        varmsd(n)=varmsd(n)+((8*eta^2)/(K^2))*((K−n)*(eta^2−al^2*(n*dt)^2)+...
                                    K*(al^2*(n*dt)^2+4*D*n*dt+2*eta^2));
    else
        varmsd(n)=((6*n^2*K−4*n*K^2+4*n+K^3−K)/(6*K))*(4*D*dt)^2+...
                   8*al^2*D*(dt)^3*(n^2*(3*n*K+1−K^2))/(3*K);
        varmsd(n)=varmsd(n)+8*eta^2*(al^2*(n*dt)^2+4*D*n*dt+2*eta^2)/K;
    end
end

% Calculate the covariance of the MSD
%
for n=1:N
    for m=1:N
        K=N−n+1;
        P=N−m+1;
        if m+n<=N
            covarmsd(n,m)=((alp^2*n)/(6*K))*(−n^3−2*P*n^2+(1−6*m^2)+...
                              (6*N+6)*m)*n+2*P*ep*n*alp/K+...
                              ((8*al^2*(dt)^3*m*n^2*D)/(3*K))*((−n^2+3*−m^2+(N+1)*m+1/3)+...)
                              ((−n+2)*P*ep^2)/(2*K*P)+...
                              (2*al^2*(dt)^2*m*n^2*ep)/(K*P);
        else
            covarmsd(n,m)=((alp^2)/(6*K))*((−m^3+(3+3*N−4*n)*m^2)+...
                              ((8+8*N)*n−2−3*N^2−6*N)*m−6*n^3+(6+6*N)*n^2−...
                              (4*N^2+8*N)*n+N*(N+1)+(N+1))−...
                              ((8*al^2*(dt)^3*m*n*D)/(3*K))*((m^2−2*(N+1)*m+...
                              (3*n^2−3*(N+1)*n+N^2+2*N))+...
                              (2*(al^2*(dt)^2*m*n+n*alp)*ep+ep^2)/K;
for n=1:N
    for m=1:n
        covarsd(n,m)=0;
    end
end

% Calculate the variance of the regression coefficients
%
for p=3:N
    s0over=0; s1over=0; s2over=0; s3over=0; s4over=0;
    for n=1:p
        tn=(n*T/N);
        variover=varmsd(n);
        s0over=s0over+1/variover;
        s1over=s1over+tn/variover;
        s2over=s2over+tn^2/variover;
        s3over=s3over+tn^3/variover;
        s4over=s4over+tn^4/variover;
    end
    del_over=s0over*s2over*s4over-s0over*(s3over)^2-
    (s1over)^2*s4over+2*s1over*s2over*s3over-(s2over)^3;
    vara_over(p)=0;
    varb_over(p)=0;
    varc_over(p)=0;
    for n=1:p
        tn=(n*T/N);
        variover=varmsd(n);
        da_over(n)=(s2over*s4over*(s3over)^2-s1over*s4over*tn+...
                   s1over*s3over*tn^2+s2over*s3over*tn-s2over^2*tn^2)/(del_over*variover);
        db_over(n)=(s0over*s4over*tn-s0over*s3over*tn^2-s1over*s4over+s2over*s3over+s1over*s2over*tn^2+...
                   s2over^2*tn)/((del_over*variover);
        dc_over(n)=(s0over*s2over*tn^2-s0over*s3over*tn-s1over*s2over*tn+s1over*s3over-s2over^2)/(del_over*variover);
        vara_over(p)=vara_over(p)+variover*da_over(n)^2;
        varb_over(p)=varb_over(p)+variover*db_over(n)^2;
end
end
\texttt{varc\_over(p) = varc\_over(p) + vari\_over*dc\_over(n)^2;}

\texttt{if n > 1}
\texttt{\hspace{1cm} for m = 1:n-1}
\texttt{\hspace{2cm} vara\_over(p) = vara\_over(p) + 2*covarmd(m,n)*da\_over(n)*da\_over(m);}
\texttt{\hspace{2cm} varb\_over(p) = varb\_over(p) + 2*covarmd(m,n)*db\_over(n)*db\_over(m);}
\texttt{\hspace{2cm} varc\_over(p) = varc\_over(p) + 2*covarmd(m,n)*dc\_over(n)*dc\_over(m);}
\texttt{\hspace{1cm} end}
\texttt{end}
\texttt{end}

\texttt{\%}
\texttt{\% Calculate the optimal number of fitting points}
\texttt{\%}

\texttt{[~,I] = min(sqrt(varb\_over(3:N))/(4*D) + sqrt(varc\_over(3:N))/(al*al));}
\texttt{popt = I + 2;}

\texttt{end}

4.3 Additional iterative algorithm experiments for \( p_{opt} \)

The iterative algorithm was also tested for different values of \( D \) and \( \alpha \) to check the robustness of the algorithm. First we test for \( D = 2 \, \mu m^2/s \), \( \alpha = 7 \, \mu m/s \), \( \eta = 2 \, \mu m \), \( N_S = 10 \) and \( N = 1000 \), for \( \Delta t = 1 \, s \), \( \Delta t = 10 \, s \) and \( \Delta t = 100 \, s \). The results are shown in Figure 7. These are similar to the case shown in the main paper. We see that the value of \( \langle p_i \rangle \) converges to \( p_{opt} \) in no more than three iterations for all the cases shown. The value of \( \langle |\alpha_i/\alpha - 1| \rangle \) does not decrease between iterations but decreases as \( \Delta t \) increases. The value of \( \langle |D_i/D - 1| \rangle \) decreases when we fit with \( p_{opt} \) points rather than all the points, and the final value of \( \langle |D_i/D - 1| \rangle \) decreases very slightly from \( \Delta t = 1 \, s \) to \( \Delta t = 10 \, s \) and then increases again from \( \Delta t = 10 \, s \) to \( \Delta t = 100 \, s \).

The algorithm was also tested for the same parameter values given above but for \( D = 6 \, \mu m^2/s \) and \( \alpha = 1 \, \mu m/s \). The results are shown in Figure 8. Again, the value of \( \langle p_i \rangle \) converges quickly to \( p_{opt} \), in this case, after 2 iterations each time. The value of \( \langle |\alpha_i/\alpha - 1| \rangle \) remains roughly the same between iterations but again decreases as \( \Delta t \) increases. Here, the value of \( \langle |D_i/D - 1| \rangle \) decreases after one iteration in each case but the final value of \( \langle |D_i/D - 1| \rangle \) does not change much between \( \Delta t \) values. Although not shown here, if the value of \( \Delta t \) was further increased then the value of \( \langle |D_i/D - 1| \rangle \) would begin to increase again.
Figure 7: A plot of the value of $\langle p_i \rangle$ for each iteration with standard error bars (left plot in each row), along with a plot of the value of $\langle |D_i/D - 1| \rangle$ (red crosses) and $\langle |\alpha_i/\alpha - 1| \rangle$ (blue circles) for each iteration with standard error bars (right plot in each row). These experiments were for $D = 2 \mu m^2/s$, $\alpha = 7 \mu m/s$, $\eta = 2 \mu m$, $N_S = 10$ and $N = 1000$. The dashed line in each left plot corresponds to $p_{opt} = 22$ (top), $p_{opt} = 5$ (middle), and $p_{opt} = 4$ (bottom), while the dashed line in each right plot corresponds with the value $10^{-2}$, indicating a 1% error. These correspond with $\Delta t = 1 s$ (top), $\Delta t = 10 s$ (middle) and $\Delta t = 100 s$ (bottom).
Figure 8: A plot of the value of $\langle p_i \rangle$ for each iteration with standard error bars (left plot in each row), along with a plot of the value of $\langle |D_i/D - 1| \rangle$ (red crosses) and $\langle |\alpha_i/\alpha - 1| \rangle$ (blue circles) for each iteration with standard error bars (right plot in each row). These experiments were for $D = 6 \mu m^2/s$, $\alpha = 1 \mu m/s$, $\eta = 2 \mu m$, $N_S = 10$ and $N = 1000$. The dashed line in each left plot corresponds to $p_{opt} = 88$ (top), $p_{opt} = 20$ (middle), and $p_{opt} = 10$ (bottom), while the dashed line in each right plot corresponds with the value $10^{-2}$, indicating a 1% error. These correspond with $\Delta t = 1 s$ (top), $\Delta t = 10 s$ (middle) and $\Delta t = 100 s$ (bottom).
4.4 Additional single particle experiments for $P_{opt}$

We now test the iterative algorithm in the single particle case for the same parameters as in Section 4.2. First we look at $D = 2 \, \mu m^2/s$, $\alpha = 7 \, \mu m/s$, $\eta = 2 \, \mu m$, $N_S = 1$ and $N = 1000$, for $\Delta t = 1 \, s$, $\Delta t = 10 \, s$ and $\Delta t = 100 \, s$. The results are shown in Figure 9. We see that the results have similar dynamics to those from Figure 4 in the main paper. In all three cases, the value of $p_i$ converges close to $P_{opt}$, the value of $\langle |\alpha_i/\alpha - 1| \rangle$ only decreases significantly when $\Delta t$ is increased and the value of $\langle |D_i/D - 1| \rangle$ is significantly improved when fit with $P_{opt}$ points rather than all the points. Notice, however, for corresponding $\Delta t$ values, that the final value of $\langle |\alpha_i/\alpha - 1| \rangle$ are always smaller, while the final value of $\langle |D_i/D - 1| \rangle$ are always larger, than those given in the main paper. This is expected as the larger value of $\alpha$ will cause the MSD to increase quicker for larger times, and so a smaller value of $\Delta t$ would be required for better inference of $D$.

The case where $D = 6 \, \mu m^2/s$, $\alpha = 1 \, \mu m/s$, $\eta = 2 \, \mu m$, $N_S = 1$ and $N = 1000$, for $\Delta t = 1 \, s$, $\Delta t = 10 \, s$ and $\Delta t = 100 \, s$ was also tested and is shown in Figure 10. Again, the results show the same dynamics as those in the main paper. This time, however, we see that the final value of $\langle |\alpha_i/\alpha - 1| \rangle$ are always larger for corresponding values of $\Delta t$. Again, this is is expected as the larger value of $D$ will in turn require a larger value of $\Delta t$ to optimise the inference of $\alpha$. 
Figure 9: A plot of the value of $\langle p_i \rangle$ for each iteration with standard error bars (left plot in each row), along with a plot of the value of $\langle |D_i/D - 1| \rangle$ (+ red crosses) and $\langle |\alpha_i/\alpha - 1| \rangle$ (○ blue circles) for each iteration with standard error bars (right plot in each row). These experiments were for $D = 2 \mu m^2/s$, $\alpha = 7 \mu m/s$, $\eta = 2 \mu m$, $N_S = 1$ and $N = 1000$. The dashed line in each left plot corresponds to $p_{opt} = 50$ (top), $p_{opt} = 16$ (middle), and $p_{opt} = 7$ (bottom), while the dashed line in each right plot corresponds with the value $10^{-1}$, indicating a 10% error. These correspond with $\Delta t = 1 s$ (top), $\Delta t = 10 s$ (middle) and $\Delta t = 100 s$ (bottom).
Figure 10: A plot of the value of $\langle p_i \rangle$ for each iteration with standard error bars (left plot in each row), along with a plot of the value of $\langle |D_i/D - 1| \rangle$ (+ red crosses) and $\langle |\alpha_i/\alpha - 1| \rangle$ (○ blue circles) for each iteration with standard error bars (right plot in each row). These experiments were for $D = 6 \mu m^2/s$, $\alpha = 1 \mu m/s$, $\eta = 2 \mu m$, $N_S = 1$ and $N = 1000$. The dashed line in each left plot corresponds to $p_{opt} = 50$ (top), $p_{opt} = 16$ (middle), and $p_{opt} = 7$ (bottom), while the dashed line in each right plot corresponds with the value $10^{-1}$, indicating a 10% error. These correspond with $\Delta t = 1 \text{ s}$ (top), $\Delta t = 10 \text{ s}$ (middle) and $\Delta t = 100 \text{ s}$ (bottom).
5 Further measurement time interval experiments

5.1 Existence of \( T_{opt} \) for different parameter values

As before, to test the robustness of our calculations, we now test the existence of an optimal measurement time interval with \( N_S = 100 \) and \( N = 10 \). Figure 11 shows the value of \( \sigma_b/b + \sigma_c/c \), again with a comparison between theoretical expressions and simulations, for \( \eta = 0.5 \, \mu m, 2 \, \mu m \) and \( 8 \, \mu m \). We see good agreement between the theory and simulations, particularly that both have their minimum’s at the same times. We also notice that the optimal measurement time intervals have decreased to \( T_{opt} \approx 189 \, s, 212 \, s \) and \( 445 \, s \). We have observed in general that as the value of \( N \) is increased the optimal measurement time interval \( T_{opt} \) also increases.

![Figure 11](image-url)

Figure 11: A plot of the theoretical value of \( \sigma_b/b + \sigma_c/c \) (--- solid lines) and its empirically estimated value using 1000 samples (--- dashed lines) against \( T \) for \( \eta = 0.5 \, \mu m, 2 \, \mu m \) and \( 8 \, \mu m \). These experiments were for \( D = 2 \, \mu m^2/s, \alpha = 1 \, \mu m/s, N_S = 100 \) and \( N = 10 \). For \( \eta = 0.5 \, \mu m, 2 \, \mu m \) and \( 8 \, \mu m \), the optimal measurement time intervals are \( T_{opt} \approx 189 \, s, 212 \, s \) and \( 445 \, s \), respectively.

Again, we also look for the existence of an optimal measurement time interval which minimises the uncertainty \( \sigma_a/a + \sigma_b/b + \sigma_c/c \) in order to infer the diffusion coefficient \( D \), the drift magnitude \( \alpha \) and the observational noise \( \eta \). The results are plotted in Figure 12. We can see good agreement between the theoretical and simulated uncertainties, which for \( \eta = 0.5 \, \mu m, 2 \, \mu m \) and \( 8 \, \mu m \), have corresponding optimal measurement time intervals \( T_{opt} \approx 32 \, s, 178 \, s \) and \( 708 \, s \), respectively.
Figure 12: A plot of the theoretical value of $\sigma_a/a + \sigma_b/b + \sigma_c/c$ (solid lines) and its empirically estimated value using 1000 samples (dashed lines) against $T$ for $\eta = 0.5 \mu m$, $2 \mu m$ and $8 \mu m$. These experiments were for $D = 2 \mu m^2/s$, $\alpha = 1 \mu m/s$, $N_S = 100$ and $N = 10$. For $\eta = 0.5 \mu m$, $2 \mu m$ and $8 \mu m$, the optimal measurement time intervals are $T_{opt} \approx 32 s$, $178 s$ and $708 s$, respectively.

5.2 MATLAB routine to calculate optimal time measurement interval

```matlab
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code to calculate the optimal measurement time interval for the %
% overlapping time–averaged mean–square displacement (MSD). %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Input: D — Estimate of the diffusion coefficient.
% al — Estimate of the drift magnitude.
% eta — Estimate of the observational noise.
% N — The number of MSD points.
%
% Output: Topt — An estimate of the optimal measurement time interval.
%
% Notes — This code works for the overlapping time–averaged MSD calculated both from a single particle and an ensemble of
```
% particles as they have the same optimal measurement time
% interval.
%

function [Topt] = optimal_over(D,al,eta,N)

NT=251;
Topt_init=100;

for i=1:NT
    T=Topt_init*(1.03)^(i-(NT+2)/2);
    time(i)=T;
    dt=T/N;
    alp=4*D*dt;
    ep=4*eta^2;

    % % Calculate the variance of the MSD
    %
    for n=1:N
        K=N+1-n;
        if n<=K
            varmsd(n)=(n*(4*n^2*K+2*K-n^3+n))/(6*K^2)*(4*D*dt)^2+
            8*al^2*D*(dt)^3*(n^3*(3*K*n+1-n^2))/(3*K^2);
            varmsd(n)=varmsd(n)+((8*eta^2)/(K^2))*((K-n)*(eta^2-
                al^2*(n*dt)^2)+K*(al^2*(n*dt)^2+4*D*n*dt+2*eta^2));
        else
            varmsd(n)=((6*n^2*K-4*n*K^2+4*n*K^3-K)/(6*K))*(4*D*dt)^2+
            8*al^2*D*(dt)^3*(n^2*(3*n*K+1-K^2))/(3*K);
            varmsd(n)=varmsd(n)+8*eta^2*(al^2*(n*dt)^2+4*D*n*dt+2*eta^2)/K;
        end
    end
%
    % % Calculate the covariance of the MSD
    %
    for n=1:N
        for m=1:N
            K=N+1-n;
            P=N+1-m;
            if m+n<=N+1
                covarmsd(n,m)=((al^2*n)/(6*K*P))*(-n^3-2*P*n^2+(1-
                    6*m^2+(6*N+6)*m)*n+2*P)+2*ep*n*alp/K+...
                ((8*al^2*dt^3*m*n^2*D)/(3*K*P))*(-n^2+3*(-m^2+
                    (N+1)*m+1/3))+((-n+2*P)*ep^2)/(2*K*P)+...
            end
        end
    end

end
\[
(2\alpha ^2 \omega \delta ^2 \tau m^2 n^2 \omega \varepsilon) / (K \Psi);
\]
else
\[
covarmsd(n, m) = \left(\frac{\alpha ^2}{6 K}\right) \left(-m^3 + \left(3 + 3N - 4n\right) m^2 + \ldots\right)
\]
\[
(8 + 8N)n - 3N^2 - 6N + m - 6n^3 + (6 + 6N)n^2 - \ldots
\]
\[
(4 + N^2 + 2 + 8N) + n + N + (N+2)(N+1) - \ldots
\]
\[
(8\alpha ^2 \omega \delta ^2 \tau m^2 n D) / (3 K) \left(m^2 - 2(N+1)m + \ldots \right)
\]
\[
(3n^2 - 3(N+1)n + N^2 + 2N) + \ldots
\]
\[
(2 + (\alpha ^2 \omega \delta ^2 \tau m n + n \alpha ^2 \omega ^2 \varepsilon) e + e^2 / 2) / K;
\]
end
end
end
end
end

for n=1:N
for m=1:n
covarmsd(n, m) = 0;
end
end

\%
\%
Calculate the variance of the regression coefficients
\%

s0=0; s1=0; s2=0; s3=0; s4=0;
for n=1:N
\[\text{tn} = \left(\frac{n \tau}{N}\right)\]
\[\text{vari} = \text{vararmsd}(n)\]
\[s0 = s0 + 1 / \text{vari}\]
\[s1 = s1 + \text{tn} / \text{vari}\]
\[s2 = s2 + \text{tn}^2 / \text{vari}\]
\[s3 = s3 + \text{tn}^3 / \text{vari}\]
\[s4 = s4 + \text{tn}^4 / \text{vari}\]
end

\[\text{del} = s0 * s2 * s4 - s0 * (s3)^2 - \ldots\]
\[\left(s1\right)^2 * s4 + 2 * s1 * s2 * s3 - (s2)^3\]

\[\text{var}_a(i) = 0;\]
\[\text{var}_b(i) = 0;\]
\[\text{var}_c(i) = 0;\]

for n=1:N
\[\text{tn} = \left(\frac{n \tau}{N}\right)\]
\[\text{vari} = \text{vararmsd}(n)\]
\[\text{da}(n) = (s2 * s4 - s3^2 * s1 * s4 * \text{tn} + \ldots\]
\[s1 * s3 + s2 * s3 + \text{tn} - s2 * s3 * \text{tn} + s2 * s3 * \text{tn}^2) / \ldots\]
\[\left(\text{del} * \text{vari}\right)\]
\[\text{db}(n) = (s0 * s4 * \text{tn} - s0 * s3 * \text{tn}^2 - s1 * s4 + \ldots\]
\[ s_2s_3 + s_1s_2 t_n^2 - s_2^2 t_n \]/... 
(del*vari); 
\[ dc(n) = (s_0s_2 t_n^2 - s_0s_3 t_n -... 
 s_1^2 t_n^2 + s_1s_2 t_n + s_1s_3 - s_2^2)/... 
(del*vari); \]
var_a(i)=var_a(i)+vari*da(n)^2; 
var_b(i)=var_b(i)+vari*db(n)^2; 
var_c(i)=var_c(i)+vari*dc(n)^2; 
for m=1:n-1 
\[ var_a(i) = var_a(i) + 2*\text{covarmd}(m,n)*da(n)*da(m); \] 
\[ var_b(i) = var_b(i) + 2*\text{covarmd}(m,n)*db(n)*db(m); \] 
\[ var_c(i) = var_c(i) + 2*\text{covarmd}(m,n)*dc(n)*dc(m); \] 
end 
end 
end

\%
\%
Calculate the optimal measurement time interval
\%

\[ [~,I] = \min(\sqrt{\text{var}_b}/(4*\text{D}) + \sqrt{\text{var}_c}/(\alpha^2)); \]
Topt=time(I);

end

5.3 Additional iterative algorithm experiments for \( T_{opt} \)

The iterative algorithm was tested for other values of the adaptation parameter \( \psi \), as well as more values of \( D \) and \( \alpha \), to test the robustness of the algorithm. First, we look at testing the iterative algorithm, but changing the adaptation parameter (Algorithm 2, Step 10 in the main paper) to \( \psi = 0.5 \) and \( \psi = 0.2 \). Both of these experiments were for \( D = 2 \mu m^2/s \), \( \alpha = 1 \mu m/s \) and \( \eta = 2 \mu m \) with the two initial measurement time intervals, \( T_0 = 10^7 \) s and \( T_0 = 10^{-3} \) s. The results are shown in Figures 13 and 14. The reduction of the adaptation parameter does not appear to have affected the inference in either case. The value of \( \langle T_i \rangle \) still converges to a time close to \( T_{opt} \) and the final value of both \( \langle |D_i/D - 1| \rangle \) and \( \langle |\alpha_i/\alpha - 1| \rangle \) are under 10% in both cases.
Figure 13: A plot of the value of $\langle T_i \rangle$ for each iteration with standard error bars (left in each), along with a plot of the value of $\langle |D_i/D - 1| \rangle$ (+ red crosses) and $\langle |\alpha_i/\alpha - 1| \rangle$ (○ blue circles) for each iteration with standard error bars (right in each). These experiments were for $D = 2 \mu m^2/s$, $\alpha = 1 \mu m/s$, $\eta = 2 \mu m$, $N_S = 10$, $N = 100$ and $\psi = 0.5$, with a starting time of $T_0 = 10^7 s$ (top) and $T_0 = 10^{-3} s$ (bottom). The dashed line in each left plot corresponds to $T_{opt} \approx 780 s$, while the dashed line in each right plot corresponds with the value $10^{-1}$, indicating a 10% error.
Figure 14: A plot of the value of $\langle T_i \rangle$ for each iteration with standard error bars (left in each), along with a plot of the value of $\langle |D_i/D - 1| \rangle$ (red crosses) and $\langle |\alpha_i/\alpha - 1| \rangle$ (blue circles) for each iteration with standard error bars (right in each). These experiments were for $D = 2 \mu m^2/s$, $\alpha = 1 \mu m/s$, $\eta = 2 \mu m$, $N_S = 10$, $N = 100$ and $\psi = 0.2$, with a starting time of $T_0 = 10^7 s$ (top) and $T_0 = 10^{-3} s$ (bottom). The dashed line in each left plot corresponds to $T_{opt} \approx 780 s$, while the dashed line in each right plot corresponds with the value $10^{-1}$, indicating a 10% error.
We now test adapting the values of $D$ and $\alpha$. The first experiment is for $D = 2 \mu m^2/s$, $\alpha = 7 \mu m/s$ and $\eta = 2 \mu m$; for these parameters $T_{opt} \approx 32 s$. The second experiment is for $D = 6 \mu m^2/s$, $\alpha = 1 \mu m/s$ and $\eta = 2 \mu m$; for these parameters $T_{opt} \approx 2195 s$. The value of the adaptation parameter for both experiments were set back to $\psi = 0.8$. The results are plotted in Figure 15 and 16, respectively. As before, the results are very similar to what has been seen before. As both cases converge to $T_{opt}$ the relative errors are reduced, particularly for a smaller value of the measurement time interval where the reduction is more significant.

Figure 15: A plot of the value of $\langle T_i \rangle$ for each iteration with standard error bars (left in each), along with a plot of the value of $\langle |D_i/D - 1| \rangle$ (+ red crosses) and $\langle |\alpha_i/\alpha - 1| \rangle$ (○ blue circles) for each iteration with standard error bars (right in each). These experiments were for $D = 2 \mu m^2/s$, $\alpha = 7 \mu m/s$, $\eta = 2 \mu m$, $N_S = 10$ and $N = 100$, with a starting time of $T_0 = 10^7 s$ (top) and $T_0 = 10^{-3} s$ (bottom). The dashed line in each left plot corresponds to $T_{opt} \approx 32 s$, while the dashed line in each right plot corresponds with the value $10^{-1}$, indicating a 10% error.
Figure 16: A plot of the value of $\langle T_i \rangle$ for each iteration with standard error bars (left in each), along with a plot of the value of $\langle |D_i/D - 1| \rangle$ (+ red crosses) and $\langle |\alpha_i/\alpha - 1| \rangle$ (○ blue circles) for each iteration with standard error bars (right in each). These experiments were for $D = 6 \, \mu m^2/s$, $\alpha = 1 \, \mu m/s$, $\eta = 2 \, \mu m$, $N_S = 10$ and $N = 100$, with a starting time of $T_0 = 10^7 \, s$ (top) and $T_0 = 10^{-3} \, s$ (bottom). The dashed line in each left plot corresponds to $T_{opt} \approx 2195 \, s$, while the dashed line in each right plot corresponds with the value $10^{-1}$, indicating a 10% error.

5.4 Additional single particle experiments for $T_{opt}$

We look at using the iterative algorithm in the single particle case for different values of $D$ and $\alpha$. First we test for $D = 2 \, \mu m^2/s$, $\alpha = 7 \, \mu m/s$, $\eta = 2 \, \mu m$, $N_S = 1$ and $N = 100$, again for both $T_0 = 10^7 \, s$ and $T_0 = 10^{-3} \, s$. For these parameter values, we have that $T_{opt} \approx 32 \, s$. The results are given in Figure 17. The algorithm takes a few more iterations to converge compared with the ensemble of particles case. However, we still see convergence in a small number of iterations, especially for a smaller measurement time interval.
Figure 17: A plot of the value of $\langle T_i \rangle$ for each iteration with standard error bars (left in each), along with a plot of the value of $\langle |D_i/D - 1| \rangle$ (red crosses) and $\langle |\alpha_i/\alpha - 1| \rangle$ (blue circles) for each iteration with standard error bars (right in each). These experiments were for $D = 2 \, \mu m^2/s$, $\alpha = 7 \, \mu m/s$ and $\eta = 2 \, \mu m$, $N_s = 1$ and $N = 100$, with a starting time of $T_0 = 10^7 \, s$ (top) and $T_0 = 10^{-3} \, s$ (bottom). The dashed line in the left plots correspond to $T_{opt} \approx 32 \, s$, while the dashed line in the right plots correspond with the value $10^{-1}$, indicating a 10% error.

The final parameter values tested are $D = 6 \, \mu m^2/s$, $\alpha = 1 \, \mu m/s$, $\eta = 2 \, \mu m$, $N_s = 1$ and $N = 100$, again for both $T_0 = 10^7$ and $T_0 = 10^{-3}$. For these parameter values, we have that $T_{opt} \approx 2195 \, s$. The results are given in Figure 18. The algorithm continues to converge in a small number of iterations and reduce the errors significantly in both cases.
Figure 18: A plot of the value of $\langle T_i \rangle$ for each iteration with standard error bars (left in each), along with a plot of the value of $\langle |D_i/D - 1| \rangle$ (+ red crosses) and $\langle |\alpha_i/\alpha - 1| \rangle$ (o blue circles) for each iteration with standard error bars (right in each). These experiments were for $D = 6 \mu m^2/s$, $\alpha = 1 \mu m/s$ and $\eta = 2 \mu m$, $N_S = 1$ and $N = 100$, with a starting time of $T_0 = 10^7$ s (top) and $T_0 = 10^{-3}$ s (bottom). The dashed line in the left plots correspond to $T_{opt} \approx 2195$ s, while the dashed line in the right plots correspond with the value $10^{-1}$, indicating a 10% error.

6 Motion blur

It is of interest what effect motion blur has on the MSD. To answer this question we follow the approach of Goulian and Simon [?], who considered the effect of full-frame motion blur on the MSD of a purely diffusive motion. We assume that the particle position in the $n$th image, $x(t)$ with $t = n\Delta t$, is taken to be the combination of the average of the true position of the particle $\bar{x}(t)$ over the time frame and the addition of static Gaussian measurement noise. If the time interval $\Delta t$ is divided by $M$ smaller microsteps of size $\delta t = \Delta t/M$, then we assume that the measured position

$$x(t) = \frac{1}{M+1} \sum_{m=0}^{M} \bar{x}(t - m\delta t) + \eta_t \equiv \bar{x}(t) + \eta_t.$$
We then have for $t' \geq t + M \delta t$

$$
\mathbb{E}((\bar{x}(t) - \bar{x}(t'))^2) = \mathbb{E}((\bar{x}(t) + \eta_t - \bar{x}(t') - \eta_{t'})^2)
$$

$$
= \mathbb{E}((\bar{x}(t) - \bar{x}(t'))^2) + 2\mathbb{E}((\eta_t - \eta_{t'})(\bar{x}(t) - \bar{x}(t'))) + \mathbb{E}((\eta_t - \eta_{t'})^2)
$$

$$
= \mathbb{E}((\bar{x}(t) - \bar{x}(t'))^2) + 4\eta^2. \quad (13)
$$

Furthermore, we have

$$
\mathbb{E}((\bar{x}(t) - \bar{x}(t'))^2)
$$

$$
= \frac{1}{(M + 1)^2} \sum_{m_1,m_2=0}^{M} \mathbb{E}((\bar{x}(t - m_1 \delta t) - \bar{x}(t' - m_1 \delta t))(\bar{x}(t - m_2 \delta t) - \bar{x}(t' - m_2 \delta t))
$$

$$
= \frac{1}{2(M + 1)^2} \sum_{m_1,m_2=0}^{M} 2\mathbb{E}((\bar{x}(t - m_1 \delta t) - \bar{x}(t' - m_2 \delta t))
$$

$$
- \mathbb{E}((\bar{x}(t - m_1 \delta t) - \bar{x}(t - m_2 \delta t)) - \mathbb{E}((\bar{x}(t' - m_1 \delta t) - \bar{x}(t' - m_2 \delta t))
$$

$$
= \frac{1}{(M + 1)^2} \sum_{m_1,m_2=0}^{M} \left[ \alpha^2(t' - t - (m_2 - m_1) \delta t)^2 + 4D(t' - t - (m_2 - m_1) \delta t)
$$

$$
- (\alpha |m_2 - m_1| \delta t)^2 + 4D|m_2 - m_1| \delta t \right]
$$

$$
= \alpha^2(t' - t)^2 + 4D \left( t' - t - \frac{(M + 1)^2 - 1}{3(M + 1)} \right). \quad (14)
$$

For the MSD we take $t' - t = n \Delta t$, and using (13) and (14) and letting $M \rightarrow \infty$ we finally get

$$
MSD(n \Delta t) = \alpha^2(n \Delta t)^2 + 4Dn \Delta t + 4 \left( \eta^2 - \frac{D \Delta t}{3} \right). \quad (15)
$$

It is clear that motion blur only effects the offset of the MSD curve for zero time lags. In particular, if the MSD is fitted by a quadratic polynomial then the quadratic term can be used to estimate the drift velocity and the linear term used to estimate the diffusion coefficient. Note that the same MSD expression given by (15) can be obtained using the general formula for the MSD subject to dynamic and static localization errors derived in Savin and Doyle [? |].

The effect of motion blur and static measurement noise was simulated by first generating true particle trajectories using microsteps of size

$$
\delta t = \frac{\Delta t}{M},
$$

where $\Delta t$ is the time between frames and $M$ is the number of microsteps between frames. For the $j$th particle the displacement was updated as follows:

$$
\tilde{x}_{i+1}^{(j)} = \tilde{x}_i^{(j)} + \alpha \delta t + \sqrt{2D \delta t N}(0, 1), \quad i = 1, \ldots, NM + 1,
$$

with $\tilde{x}_1 = 0$. The measured displacement at $t = (n-1) \Delta t$ was obtained by averaging the true position of the particle over the previous $M$ microsteps to simulate full-frame motion blur, and a Gaussian static measurement noise was then added so that

$$
x_n^{(j)} = \frac{1}{M + 1} \sum_{i=(n-2)M+1}^{(n-1)M+1} \tilde{x}_i^{(j)} + \eta_n, \quad n = 2, \ldots, N + 1,
$$

33
and $\mathbf{x}_1^{(j)} = \eta_1$.

The following results are for calculating the optimal number of fitting points. First, we look for the existence of an optimal number of fitting points which minimises the value of $\sigma_b/b + \sigma_c/c$. Figure 19 shows a comparison between the theoretical value of $\sigma_b/b + \sigma_c/c$ assuming no motion blur and its simulated value which includes motion blur. We see that for small $\Delta t$, corresponding to using all the MSD points in the fitting, motion blur does not seem to have an effect on the results. However, when $\Delta t$ is increased, motion blur appears to effect the value of $\sigma_b/b + \sigma_c/c$. While the theoretical optimal number of fitting points is 8, the simulated optimal number of fitting points is 5.

![Figure 19: A plot of the theoretical value of $\sigma_b/b + \sigma_c/c$ assuming no motion blur (—– solid red line) and an empirically estimated value found using 1000 samples which include motion blur (- - - dashed black line) when fit with the first $p$ MSD points for $D = 2 \mu m^2/s$, $\alpha = 1 \mu m/s$, $\eta = 2 \mu m$, $N_S = 10$ and $N = 100$, for $\Delta t = 1 s$ giving $T = 100 s$ (left), and $\Delta t = 10 s$ giving $T = 1000 s$ (right). The optimal number of fitting points for the left plot are both 100, while for the right it is 8 (theory) and 5 (simulations).](image)

Since it can be seen that motion blur has an effect on the value of $p_{opt}$, we look to see whether using our theoretical value for $p_{opt}$ on data which includes motion blur leads to a change in results. To test this, we will use the iterative algorithm to see the effect of motion blur on the inference of $D$ and $\alpha$. For this, the simulated data will include motion blur but will be driven towards the theoretical $p_{opt}$ which does not include motion blur (see section 4.2). In the absence of motion blur, the regression coefficient $c = 4\eta^2$. However, when motion blur is present, the regression coefficient becomes $c = 4\eta^2 - \frac{4}{3}D\Delta t$ and so the value of $\eta$ will be inferred with this in mind. The results are shown in Figure 20. For this we have $D = 2 \mu m^2/s$, $\alpha = 1 \mu m/s$, $\eta = 2 \mu m$, $N_S = 10$ and $N = 1000$ for $\Delta t = 1 s$, $\Delta t = 10 s$ and $\Delta t = 100 s$. Note that these plots are comparable with Figure 3 from the main paper. Even though we are driving $p_i$ to the theoretical optimal number of fitting points which does not include motion blur, we do not see a significant difference in the inference of $D$ and $\alpha$ when motion blur is included. This suggests that our theory for the optimal number of fitting points can still be used when motion blur is present.

To further test the effects of motion blur, we also tested the iterative algorithm for a single particle. The parameters values were chosen to be the same as those above but for $N_S = 1$. The results are shown in Figure 21. Note that these plots are comparable with Figure 4 from
the main paper. Again, we do not see a significant difference with the inclusion of motion blur, further suggesting that our theoretical $p_{\text{opt}}$ value works when motion blur is present.

Figure 20: A plot of the value of $\langle p_i \rangle$ for each iteration with standard error bars (left plot in each row), along with a plot of the value of $\langle |D_i/D - 1| \rangle$ ( + red crosses) and $\langle |\alpha_i/\alpha - 1| \rangle$ (○ blue circles) for each iteration with standard error bars (right plot in each row). The simulated values include motion blur. These experiments were for $D = 2 \mu m^2/s$, $\alpha = 1 \mu m/s$, $\eta = 2 \mu m$, $N_S = 10$ and $N = 1000$. The dashed line in each left plot corresponds to $p_{\text{opt}} = 50$ (top), $p_{\text{opt}} = 16$ (middle), and $p_{\text{opt}} = 7$ (bottom), while the dashed line in each right plot corresponds with the value $10^{-2}$, indicating a 1% error. These correspond with $\Delta t = 1 s$ (top), $\Delta t = 10 s$ (middle) and $\Delta t = 100 s$ (bottom).
Figure 21: A plot of the value of $\langle p_i \rangle$ for each iteration with standard error bars (left plot in each row), along with a plot of the value of $\langle |D_i/D-1| \rangle$ (+ red crosses) and $\langle |\alpha_i/\alpha-1| \rangle$ (○ blue circles) for each iteration with standard error bars (right plot in each row). These experiments were for $D = 2 \mu m^2/s$, $\alpha = 1 \mu m/s$, $\eta = 2 \mu m$, $N_S = 1$ and $N = 1000$. The dashed line in each left plot corresponds to $p_{opt} = 50$ (top), $p_{opt} = 16$ (middle), and $p_{opt} = 7$ (bottom), while the dashed line in each right plot corresponds with the value $10^{-1}$, indicating a 10% error. These correspond with $\Delta t = 1 s$ (top), $\Delta t = 10 s$ (middle) and $\Delta t = 100 s$ (bottom).
7 Determination of the drift direction

Until now we have assumed that the main interest is the determination of the magnitude of the drift velocity as well as the diffusion coefficient. If the drift angle is also of interest then this can also be inferred using the trajectory data. To calculate the mean direction and measures of the spread about the mean requires the use of circular statistics, details of which can be found in, for example, Fisher [214] or Mardia and Jupp [345]. Following in the same vein as the estimation of the MSD, we first calculate the ensemble overlapping time-averaged quantities

\[ C_n = \sum_{j=1}^{N_S} \sum_{i=1}^{N+1-n} \cos \theta_{i,n}^{(j)} \quad \text{and} \quad S_n = \sum_{j=1}^{N_S} \sum_{i=1}^{N+1-n} \sin \theta_{i,n}^{(j)}. \]  

(16)

where

\[ \cos \theta_{i,n}^{(j)} = \frac{x_{i+n}^{(j)} - x_i^{(j)}}{|x_{i+n}^{(j)} - x_i^{(j)}|} \quad \text{and} \quad \sin \theta_{i,n}^{(j)} = \frac{y_{i+n}^{(j)} - y_i^{(j)}}{|x_{i+n}^{(j)} - x_i^{(j)}|}. \]  

(17)

The resultant vector using data with a time lag of \( n \Delta t \) is \( R_n = \sqrt{C_n^2 + S_n^2} \) from which we can define the cosine and sine of the average angle using a time lag of \( n \) by

\[ \cos \bar{\theta}_{d,n} = C_n/R_n \quad \text{and} \quad \sin \bar{\theta}_{d,n} = S_n/R_n. \]  

(18)

The average angle using displacements with a time lag of \( n \Delta t \) is given by

\[ \bar{\theta}_{d,n} = \begin{cases} \tan^{-1}(S_n/C_n) & S_n > 0, C_n > 0 \\ \tan^{-1}(S_n/C_n) + \pi & C_n > 0 \\ \tan^{-1}(S_n/C_n) + 2\pi & S_n < 0, C_n > 0. \end{cases} \]  

(19)

To help quantify the uncertainty in inferring the drift angle, the circular variance \( V_n \) is defined as

\[ V_n = 1 - \bar{R}_n \quad \text{where} \quad \bar{R}_n = R_n/(N_S(N + 1 - n)) \]

is the mean resultant length. Finally, the first and second central trigonometric moments

\[ m_{1,n} = \frac{1}{N_S(N + 1 - n)} \sum_{j=1}^{N_S} \sum_{i=1}^{N+1-n} \cos (\theta_{i,n}^{(j)} - \bar{\theta}_n), \]

and

\[ m_{2,n} = \frac{1}{N_S(N + 1 - n)} \sum_{j=1}^{N_S} \sum_{i=1}^{N+1-n} \cos (2(\theta_{i,n}^{(j)} - \bar{\theta}_n)), \]

can be used to define the sample circular dispersion

\[ \hat{\delta}_n = (1 - m_{2,n})(2m_{1,n}^2). \]

The error bars in the following plots use the circular standard error

\[ \hat{\sigma}_n = \sqrt{\hat{\delta}_n/(N_S(N + 1 - n))}. \]

Figure 22 shows the calculated circular variance and mean drift angle using \( N_S = 10 \) particles and \( N = 100 \) measurements, where we have assumed that the optimal measurement time has
been used. We can see that the variance in the estimation of the drift angle decreases rapidly over time. This would suggest that the optimal time to infer the drift angle would be at the maximal measurement time. However, due to the lack of samples in the time averaging using the large time lags we find that there is very little difference in the quality of the inferred drift angle other than when very short lags are used. We leave till further study a theoretical analysis to determine the optimal strategy of the drift direction in combination with the other parameters of the drift-diffusion model.