

Smoothing and Interpolating Noisy GPS Data with Smoothing Splines

Jeffrey J. Early^{*} and Adam M. Sykulski[†]

^{*}*Corresponding author address:* NorthWest Research Associates, 4118 148th Ave NE, Redmond,
WA 98052, USA

E-mail: jearly@nwra.com

[†]Current affiliation: Data Science Institute and the Department of Mathematics and Statistics, Lancaster University, Bailrigg LA1 4YW, UK (email: a.sykulski@lancaster.ac.uk)

ABSTRACT

8 A comprehensive methodology is provided for smoothing noisy, irregularly
9 sampled data with non-Gaussian noise using smoothing splines. We demon-
10 strate how the spline order and tension parameter can be chosen *a priori* from
11 physical reasoning. We also show how to allow for non-Gaussian noise and
12 outliers which are typical in GPS signals. We demonstrate the effectiveness
13 of our methods on GPS trajectory data obtained from oceanographic floating
14 instruments known as drifters.

1. Introduction

In 2011 an array of floating ocean surface buoys (drifters) were deployed in the Sargasso Sea to assess the lateral diffusivity of oceanic processes (Shcherbina et al. 2015). Each drifter was equipped with a global positioning system (GPS) receiver recording locations every 30 minutes. Addressing the primary goal of understanding the processes controlling lateral diffusivity requires significant processing of the drifter positions, including removing mean flow, accounting for the large scale strain field, and analyzing the residual spectra for hints of a dynamical process. However, it quickly became clear that the GPS position data, which can have accuracies as low as a few meters (WAAS T&E Team 2016), was contaminated by outliers with position jumps of hundreds of meters or more. Prior to analysis, the position data requires removing outliers, and interpolating gaps to keep the position data synchronized in time across the drifter array.

The basic problem is ubiquitous: observations from GPS receivers return observed positions x_i at times t_i that differ from the true positions $x_{\text{true}}(t_i)$ by some noise $\varepsilon_i \equiv x_i - x_{\text{true}}(t_i)$ with variance σ^2 . The goal of *smoothing* is to find the true position $x_{\text{true}}(t_i)$ not contaminated by the noise, while the goal of *interpolating* is to find the true position $x_{\text{true}}(t)$ between observation times. The approach taken here is to use smoothing splines. This approach is relatively broad (Handcock et al. 1994; Nychka 2000), and is related to the methodologies in Yaremchuk and Coelho (2015) and Elipot et al. (2016) for smoothing drifter trajectories, as discussed later.

Our model for the ‘true’ path $x(t)$ is specified using interpolating b-splines $X^K(t)$ such that

$$x(t) = \sum_{i=1}^N \xi_i X_i^K(t), \quad (1)$$

where K is the order (degree $S = K - 1$) of the spline. For N observations we construct N b-splines such that $x(t_i) = x_i$ for appropriately chosen coefficients ξ_i . To smooth the data we choose new

36 coefficients $\bar{\xi}_i$ that minimize the penalty function

$$\phi = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - x(t_i)}{\sigma} \right)^2 + \frac{\lambda_T}{t_N - t_1} \int_{t_1}^{t_N} \left(\frac{d^T x}{dt^T} \right)^2 dt, \quad (2)$$

37 for some tension parameter $\lambda_T \geq 0$. If $\lambda_T = 0$ then $\phi = 0$ and $\xi_i = \bar{\xi}_i$ because $x(t_i) = x_i$, but if
38 $\lambda_T \rightarrow \infty$ then this forces $x(t)$ to a T -th order polynomial (e.g., when $T = 2$, the model is forced
39 to be a straight line because it has no second derivative). The resulting path $x(t)$ is known as a
40 smoothing spline and was first introduced in modern form by Reinsch (1967), but according to
41 De Boor (1978) the idea dates back to Whittaker (1923). Once S and T are chosen, the smoothing
42 spline has one free parameter (λ_T) and its optimal value can be found by minimizing the expected
43 mean square error when the true value of σ is known (Craven and Wahba 1979).

44 Three issues must be addressed before smoothing splines are applied to GPS data:

- 45 1. how to choose S and T —and how do these choices affect the recovered power spectrum?
- 46 2. how to modify the spline fit to accommodate the non-Gaussian errors of GPS receivers?
- 47 3. how to identify and remove outliers?

48 To address these issues, but also serve as a practical guide to other practitioners, we review B-
49 splines in section 2 and introduce the canonical interpolating spline as the underlying model for
50 path $x(t)$ in (1). We demonstrate the effect choosing S has on the high-frequency slope of the
51 power spectrum of the interpolated fit.

52 Section 3 takes a broad look at smoothing splines and the assumptions they make on the underly-
53 ing process. Many of the ideas presented in this section are known to the statistics community, so
54 here we present these ideas from a more physical perspective. We show that the penalty function
55 in (2) can be formulated as a maximum likelihood problem, and applying tension is equivalent to
56 assuming a Gaussian distribution on the tensioned derivative of the underlying process.

57 Section 4 uses ensembles from synthetic data that mimic the oceanographic data to test a number
58 of choices that must be made. We establish that setting $T = S$ is a reasonable choice. We show how
59 the tension parameter can be chosen *a priori* (without optimization of the mean square error) when
60 the *effective sample size* (which we define later) can be estimated from the data. This estimate for
61 effective sample size can be used to reduce the coefficients, ξ^i , in the spline fit without increasing
62 mean square error.

63 The second half of the manuscript addresses issues specific to GPS position errors. In section 5
64 we discuss the assumptions of stationarity and isotropy required for bivariate smoothing splines.
65 In section 6 we show that GPS errors are not Gaussian distributed, but t -distributed, and we show
66 how to modify our methodology for a t -distribution. Section 7 addresses how to modify our
67 methodology to make smoothing splines robust to outliers. We compare with alternative methods
68 and conclude in sections 8 and 9 respectively.

69 A major outcome of this work is the implementation of Matlab classes for generating b-splines,
70 interpolating splines, smoothing splines, and a class specific to smoothing GPS data¹. These
71 classes are highlighted throughout in relevant sections.

72 2. Interpolating Spline

73 Assume we are given N observations of a particle position (t_i, x_i) with no errors. The simplest
74 form of interpolation is a nearest neighbor method that assigns the position of the particle to the
75 nearest observations in time. The resulting interpolated function $x(t)$ is a polynomial of *order*
76 $K = 1$ (piecewise constant), shown in the top row of Fig. 1. The next level of sophistication is
77 to assume a constant velocity between any two observations and use that to interpolate positions
78 between observations, second row of Fig. 1. This means we now have a piecewise constant

¹<https://github.com/JeffreyEarly/GLNumericalModelingKit>

79 function $\frac{dx}{dt}$ that represents the velocity of the particle, shown in the second row, second column of
 80 Fig. 1. This is a polynomial function of order $K = 2$.

81 It is less obvious how to proceed to a polynomial of order $K = 3$. With N data points we can
 82 construct a piecewise constant acceleration (the second derivative) using the $N - 2$ independent
 83 accelerations computed from finite differencing, but where to place *knot points* that define the
 84 boundaries of the regions and how to maintain continuity is less clear. The approach taken here is
 85 to use B-splines.

86 *a. B-Splines*

87 A B-spline (or basis spline) of *order* K (*degree* $S = K - 1$) is a piecewise polynomial that main-
 88 tains nonzero continuity across S knot points. The knot points are a nondecreasing collection of
 89 points in time denoted by τ_i . The basic theory is well documented in De Boor (1978), but here we
 90 present a reduced version tailored to our needs.

91 The m -th B-spline of order $K = 1$ is defined as

$$X_m^1(t) \equiv \begin{cases} 1 & \text{if } \tau_m \leq t < \tau_{m+1}, \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

92 This is the rectangle function as shown in the first row, first column of Fig. 2. Given P knot points
 93 we can construct $P - 1$ B-splines of order $K = 1$, although if a knot point is repeated this results in
 94 a spline that is zero everywhere. To represent an interpolating function $x(t)$ for the N observations
 95 of a particle position (t_i, x_i) we define $N + 1$ knot points as

$$\tau_m = \begin{cases} t_1 & m = 1, \\ t_{m-1} + \frac{t_m - t_{m-1}}{2} & 1 < m \leq N, \\ t_N & m > N. \end{cases} \quad (4)$$

96 This creates N independent basis functions that provide support for the region $t_1 \leq t \leq t_N$ (provided
 97 the last spline is defined to include the last knot point). The interpolating function $x(t)$ is defined
 98 as $x(t) \equiv X_m^1(t)\xi^m$ where the coefficients ξ^m are found by solving $X_m^1(t^i)\xi^m = x^i$. The result of
 99 this process is shown in Fig. 1 for 7 irregularly spaced data points.

100 All higher order ($K > 1$) B-splines are defined by recursion,

$$X_m^K(t) \equiv \frac{t - t_m}{t_{m+K-1} - t_m} X_m^{K-1}(t) + \frac{t_{m+K} - t}{t_{m+K} - t_{m+1}} X_{m+1}^{K-1}(t). \quad (5)$$

101 This creates splines that span across one additional knot point at each order, and maintain conti-
 102 nuity across one more derivative. Examples are shown in Fig. 2.

103 Any knot points that are repeated T times result in a total of $T - 1$ splines of order one that are
 104 everywhere zero. This has the effect of introducing discontinuities in the derivatives for higher
 105 order splines. For our purposes, we use this feature to prevent higher order splines from crossing
 106 boundaries. For $K = 2$ order splines we use $N + 2$ knot points at locations

$$\tau_m = \begin{cases} t_1 & m \leq 2, \\ t_{m-1} & 2 < m \leq N, \\ t_N & m > N. \end{cases} \quad (6)$$

107 This creates a knot point at every observation point, but repeats the first and last knot point. This
 108 has the effect of terminating the first and last spline at the boundary and creating N second order B-
 109 splines, $X_m^2(t)$. The interpolating function $x(t)$ is defined as $x(t) \equiv X_m^2(t)\xi^m$ where the coefficients
 110 ξ^m are found by solving $X_m^2(t^i)\xi^m = x^i$. The second row of Fig. 1 shows an example.

111 This process can be continued to higher order B-splines. For splines that are of *even* order, we
 112 create $N + K$ knots points with

$$\tau_m^{K\text{-even}} = \begin{cases} t_1 & m \leq K, \\ t_{m-K/2} & K < m \leq N, \\ t_N & m > N, \end{cases} \quad (7)$$

113 and for splines that are *odd* order, we create $N + K$ knot points with

$$\tau_m^{K\text{-odd}} = \begin{cases} t_1 & m \leq K, \\ t_{m-\frac{K+1}{2}} + \frac{t_{m+1-\frac{K+1}{2}} - t_{m-\frac{K+1}{2}}}{2} & K < m \leq N, \\ t_N & m > N. \end{cases} \quad (8)$$

114 The knot points are chosen to create N splines for the N data points such that the interpolated
 115 function $x(t)$ crosses all N observations (t_i, x_i) . The path $x(t)$ is the *canonical interpolating spline*
 116 *of order K* . Examples are shown in Fig. 1.

117 The knot placements in (7) and (8) are equivalent to the *not-a-knot* boundary conditions de-
 118 scribed in De Boor (1978) and used in the cubic spline implementation in Matlab. In the usual
 119 formulation of the not-a-knot boundary condition, the knot positions do not change as a function
 120 of spline order, and therefore additional constraints must be added at each order—especially the
 121 requirement that the highest derivative maintain continuity near the boundaries. In the formulation
 122 here, these constraints are implicit in (7) and (8).

123 *b. Numerical implementation*

124 The root class in our suite of Matlab classes is the BSpline class, which evaluates a complete
 125 B-spline basis set given a set of knot points. This class was used to generate Fig. 2.

The interpolating spline used to generate Fig. 1 is implemented in the `InterpolatingSpline` class—a subclass of `BSpline`. This class generates interpolating splines of arbitrary order given a set of data points (t_i, x_i) , thus generalizing the cubic spline command built in to Matlab.

c. *Synthetic Data*

Throughout this manuscript we generate synthetic data for both the signal and the noise. The velocity of the signal is generated from a bivariate Gaussian process known as the Matérn (Lilly et al. 2017). The spectrum of the Matérn is given by

$$S(\omega) = \frac{A^2}{(\omega^2 + \lambda^2)^{p/2}}, \quad (9)$$

where A sets the amplitude, p sets the high frequency slope, and λ sets the frequency below which the signal looks white. When $p > 1$, this spectrum has finite amplitude at low frequencies and power-law fall off at high frequencies, two physically realistic properties observed in ocean surface drifters (Sykulski et al. 2016). In our experiments the parameter A is chosen such that the square root of velocity variance in each direction is $u_{\text{rms}} = 0.20$ m/s and the damping scale is $\lambda^{-1} = 30$ minutes. Values of p are varied with $p = 2, 3, 4$ so that the high frequency spectrum is proportional to ω^{-2} , ω^{-3} , ω^{-4} . Velocities are sampled every minute, and integrated to get positions. Fig. 3 shows an example velocity spectrum of the signal with $p = 2$.

The position data is contaminated with (white) Gaussian noise with $\sigma = 10$ meters, a value chosen to resemble GPS errors. For all experiments we use a range of *strides*, that is, subsampled versions of the underlying process as input into the spline fits. A stride of 100 indicates that the signal is subsampled to 1 in every 100 data points. This lets us evaluate the quality of fit against different strides. In analyzing the quality of fits, we use velocities when computing the power spectrum, but report mean square errors from positions.

147 *d. Spline degree, S*

148 We first examine a synthetic signal *uncontaminated* by noise, to examine the role of the spline
149 degree, S , on the interpolated fit. As noted in Craven and Wahba (1979), the degree of the spline
150 sets its roughness. In terms of the power spectrum, this corresponds to the high frequency slope as
151 can be seen in Fig. 3 which shows fits with $S = 1..4$. Setting $S = 1$ produces a high frequency fall
152 off in the spline fit of ω^{-2} . Although this appears to be a desirable feature when fitting to a process
153 with true slope ω^{-2} , the mean square error is consistently higher (as indicated in the legend of Fig.
154 3).

155 The bottom panel of Fig. 3 shows the coherence between the spline fit and the true signal.
156 A coherence of 1 indicates that the signals are perfectly matched at a given frequency, while
157 a coherence of 0 indicates that the signals are unrelated. There is no discernible difference in
158 coherence between spline fits with $S = 1..4$. The coherence quickly drops to near zero at the same
159 frequency in all three cases. The implication here is that the spline fits are essentially producing
160 noise at frequencies above the loss-of-coherence. This is why the spline fits with shallower slopes
161 (with more variance at high, incoherent frequencies) produce a larger overall mean square error
162 than those with steeper slopes (with less variance at high, incoherent frequencies). The conclusion
163 here is that smoother is better: it is better to use an unnecessarily high order spline to avoid adding
164 extra noise at high frequencies.

165 **3. Smoothing Spline**

166 A typical starting point for maximum likelihood is to establish the probability distribution func-
167 tion (PDF) of the errors, $\epsilon_i \equiv x_i - x_{\text{true}}(t_i)$. The canonical example in one-dimension (e.g., Press
168 et al. (1992)) is to assume errors are independently drawn from a Gaussian with the following

169 probability distribution

$$p_g(\varepsilon|\sigma_g) = \frac{e^{-\frac{1}{2} \frac{\varepsilon^2}{\sigma_g^2}}}{\sigma_g \sqrt{2\pi}}, \quad (10)$$

170 where σ_g is the standard deviation. This assumption alone places no assumptions on the signal,
171 only on the structure of the noise.

172 The probability of the observed data given model $x(t)$ is

$$P = \frac{1}{\sigma \sqrt{2\pi}} \prod_{i=1}^N \exp \left[-\frac{1}{2} \left(\frac{x_i - x(t_i)}{\sigma} \right)^2 \right], \quad (11)$$

173 where we have taken $\sigma = \sigma_g$. Maximizing the probability function in (11) is the same as minimiz-
174 ing its argument—this is the log likelihood (up to a constant), called the penalty function

$$\phi = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - x(t_i)}{\sigma} \right)^2. \quad (12)$$

175 Stated in this way this is the same as asking for the ‘least-squares’ fit of the errors.

176 *a. Smoothing spline penalty function*

177 The model used here is the canonical interpolating spline of order K described in section 2.
178 We have chosen our knot points such that the model intersects the observations and this certainly
179 maximizes (11) (and minimizes (12)) because all the errors are zero, but the resulting distribution
180 of errors (a delta function at zero) does not resemble the assumed Gaussian distribution. Thus,
181 additional constraints are required if the assumed error distribution is to be recovered.

182 The smoothing spline augments the penalty function of (12) by adding a global constraint on the
183 T -th derivative of the resulting function as in (2). If $\lambda_T \rightarrow 0$ then this reduces to the least-squares
184 fit in (12), but if $\lambda_T \rightarrow \infty$ then this forces the model to an T -th order polynomial.

185 To interpret the first term of (2), consider a motionless particle at true position x_0 . Using the
186 N relevant observations x_i , the *sample mean* $\bar{x} = \frac{1}{N} \sum x_i$ estimates the particle’s position x_0 . The

187 unbiased *sample variance* estimates the variance of the noise, σ^2 , and is given by $\hat{\sigma}^2 = \frac{1}{N-1} \sum (x_i -$
 188 $\bar{x})^2$, the expected value of which is $\langle \hat{\sigma}^2 \rangle = (1 - \frac{1}{N}) \sigma^2$.

189 Now consider the opposite extreme where the particle is moving so fast (or the observations are
 190 so sparse) that each observation is independent of its neighbors. In this case, each observation
 191 must be considered separately, so the sample mean at time t_i is $\bar{x}_i = x_i$ (i.e., we are summing over
 192 the single relevant observation). In this scenario we cannot produce a sample variance, because
 193 there is only a single relevant observation at time t_i .

194 In practice, the number of relevant observations is anywhere between 1 and N . Here we use the
 195 term *effective sample size*, denoted by n_{eff} , to describe the typical number of observations being
 196 used to estimate either the particle's position or the variance of the noise at any given time. In this
 197 context, the first term of (2) is proportional to an ensemble of multiple estimates of the sample
 198 variance

$$\hat{\sigma}^2 \equiv \frac{1}{N} \sum_{i=1}^N (x_i - x(t_i))^2, \quad (13)$$

199 which is expected to scale as

$$\langle \hat{\sigma}^2 \rangle = \left(1 - \frac{1}{n_{\text{eff}}^{\text{var}}} \right) \sigma^2, \quad (14)$$

200 where $1 < n_{\text{eff}}^{\text{var}} \leq N$ is our definition of the effective sample size of the sample variance. Revisiting
 201 the limiting cases, as $n_{\text{eff}}^{\text{var}} \rightarrow N$ the sample variance matches the true variance, but as $n_{\text{eff}}^{\text{var}} \rightarrow 1$, the
 202 sample variance vanishes.

203 There is a simple physical interpretation for the second term in (2). Consider the case $T = 1$ so
 204 that the smoothing spline is a constraint on velocity. When averaged over the integration time, the
 205 integral produces the root mean square velocity, u_{rms} , such that the second term scales as u_{rms}^2 . In
 206 general, where $x_{\text{rms}}^{(T)}$ is the root-mean-square of the T -th derivative, this means λ_T scales like

$$\lambda_T = \left(1 - \frac{1}{n_{\text{eff}}^{\text{var}}} \right) \frac{1}{\left(x_{\text{rms}}^{(T)} \right)^2}. \quad (15)$$

The interpretation of the smoothing spline is that the two terms are balanced by a relative weighting of the sample variance of the noise and mean square of the T -th derivative of the physical process. As discussed in section 4, both $x_{\text{rms}}^{(T)}$ and $n_{\text{eff}}^{\text{var}}$ can be estimated *a priori* such that a good initial estimate for λ_T can be made.

b. Smoothing spline maximum likelihood

The penalty function in (2) can be restated in terms of maximum likelihood under some conditions (see chapter 3.8 in Green and Silverman (1994)). Assume that in addition to knowing the distribution of errors as in (11), we know how the velocity of the underlying physical process is distributed. For example, in geophysical turbulence the velocity probability distribution function is like a Laplace distribution (Bracco et al. 2000). To recover the smoothing spline, we consider the case where the velocity PDF is Gaussian. Stated as maximum likelihood, this means at *any given instant* (not just the times of observation) we expect the model velocity to be Gaussian. We discretize the problem by sampling the velocity Q times $t_q = t_1 + q\Delta t_q$, where $\Delta t_q = \frac{t_N - t_1}{Q-1}$ and $q = 0..Q-1$. The maximum likelihood is thus stated as

$$P = \prod_{i=1}^N \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x_i - x(t_i)}{\sigma} \right)^2 \right] \cdot \prod_{q=1}^Q \frac{\sqrt{\gamma}}{x_{\text{rms}}^{(T)}\sqrt{2\pi}} \exp \left[-\frac{\gamma}{2} \left(\frac{x^{(T)}(t_q)}{x_{\text{rms}}^{(T)}} \right)^2 \right], \quad (16)$$

which is the joint probability of the error distribution from (11) and the velocity distribution of the underlying physical process. We include parameter γ to set the relative weighting between the two distributions, although it could be absorbed into the definition of $x_{\text{rms}}^{(T)}$. Writing (16) as a penalty function (after converting the product of exponentials into exponentials of sums), we have

$$-\log P = \frac{1}{2} \sum_{i=1}^N \left(\frac{x_i - x(t_i)}{\sigma} \right)^2 + \frac{\gamma}{2} \sum_{q=1}^Q \left(\frac{x^{(T)}(t_q)}{x_{\text{rms}}^{(T)}} \right)^2 + C, \quad (17)$$

where C is a constant. Setting $\gamma = \frac{N}{Q}$ and renormalizing the penalty function by $\frac{2}{N}$ (which has no effect on the location of its minimum), (17) can be written as

$$\phi = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - x(t_i)}{\sigma} \right)^2 + \frac{1}{t_N - t_1} \sum_{q=1}^Q \left(\frac{x^{(T)}(t_q)}{x_{\text{rms}}^{(T)}} \right)^2 \Delta t_q. \quad (18)$$

Apart from the discretization of the integral, (18) is the same as the penalty function in (2).

There is an important special case when tension is applied at the same order as the spline, $T = S$. In this case the spline is piecewise constant for $x^{(T)}$ with exactly $N - T$ unique values. The parameter $\gamma = \frac{N}{N-T} \approx 1$ and (16) can be simplified. This case is appealing because only the $N - T$ unique values of the derivative $x^{(T)}$ that can be computed from N data points are being used for tension, which is not the case when $T < S$.

This maximum likelihood perspective shows that adding tension to the penalty function is equivalent to assuming a higher order derivative in the model (e.g., velocity if $T = 1$) is Gaussian. This is therefore making an assumption about the underlying *physical process* of the model. This is in contrast to the first term which is entirely a statement about *measurement noise*.

Writing the smoothing spline as a maximum-likelihood condition (16), suggests that if the underlying physical process has a non-zero mean value in tension, the fit will not behave as expected. However, smoothing splines can be easily modified to accommodate a mean value in tension, as shown in appendix A.

c. Optimal parameter estimation

For a given choice of T and λ_T , the minimum solution to (2) can be found analytically (see Teanby (2007) and our appendix A). Once the solution is found the smoothing matrix \mathbf{S}_λ is defined as the matrix that takes observations \mathbf{x} and maps them to their smooth values, $\hat{\mathbf{x}} = \mathbf{S}_\lambda \mathbf{x}$.

245 The free parameter λ_T is a relative weighting between the two terms in (2). Choosing its optimal
 246 value can be done by minimizing the expected mean square error (Craven and Wahba 1979),

$$\text{MSE}(\lambda) = \frac{1}{N} \|(\mathbf{S}_\lambda - \mathbf{I})\mathbf{x}\|^2 + \frac{2\sigma^2}{N} \text{Tr} \mathbf{S}_\lambda - \sigma^2, \quad (19)$$

247 where $\|\cdot\|^2$ is the Euclidean norm, Tr indicates the trace, and \mathbf{I} is the identify matrix.

248 A significant amount of the literature on smoothing splines is devoted to minimizing the mean
 249 square error when the variance, σ^2 , is *not* known. Craven and Wahba (1979) and Wahba (1978) use
 250 cross-validation to estimate σ and minimize mean square error. Recent work comparing different
 251 estimators shows no single technique to be optimal (Lee 2003). For our application however, the
 252 errors in GPS data can be relatively easily established, as shown in section 6.

253 The mean square error in (19) is a combination of the sample variance and the variance of the
 254 mean. As already discussed in the context of the penalty function ϕ in section 3a, the first term in
 255 (19) is an ensemble of sample variances, and therefore by combining (13), (14) and (19) we obtain

$$\left(1 - \frac{1}{n_{\text{eff}}^{\text{var}}}\right) \sigma^2 = \frac{1}{N} \|(\mathbf{I} - \mathbf{S}_\lambda)\mathbf{x}\|^2. \quad (20)$$

256 The second term in (19) is proportional to twice the squared standard error, i.e., the variance of
 257 the sample mean. As discussed in Teanby (2007), the quantity $\mathbf{S}_\lambda \Sigma$ is the covariance matrix with
 258 the squared standard error along the diagonal and thus the mean squared standard error is given
 259 by $\frac{1}{N} \text{Tr}(\mathbf{S}_\lambda \Sigma)$. The variance of the sample mean is known to scale inversely with the number of
 260 samples being used to estimate the mean. We use this to define the effective sample size of the
 261 variance of the mean, $n_{\text{eff}}^{\text{SE}}$ with

$$\frac{\sigma^2}{n_{\text{eff}}^{\text{SE}}} = \frac{1}{N} \text{Tr}(\mathbf{S}_\lambda \Sigma). \quad (21)$$

262 Taking the measures of effective sample size as functions of λ , the mean square error can be
 263 expressed by combining (19)–(21):

$$\text{MSE}(\lambda) = 2 \frac{\sigma^2}{n_{\text{eff}}^{\text{SE}}} - \frac{\sigma^2}{n_{\text{eff}}^{\text{var}}}. \quad (22)$$

264 If one assumes $n_{\text{eff}}^{\text{var}} = n_{\text{eff}}^{\text{SE}}$, then the expected mean square error from (19) is equal to σ^2/n_{eff} . Al-
 265 though not shown here, in an empirical analysis we find that $n_{\text{eff}}^{\text{var}}$ and $n_{\text{eff}}^{\text{SE}}$ are approximately equal,
 266 although $n_{\text{eff}}^{\text{var}}$ becomes highly variable when $n_{\text{eff}}^{\text{SE}}$ approaches 1. These measures of effective sam-
 267 ple size can be used to estimate the value of λ_T necessary for optimal tension without minimizing
 268 the expected mean square error.

269 The definition of effective sample size used here is related to, but not the same as, the notion of
 270 degrees-of-freedom used in Cantoni and Hastie (2002) and references therein.

271 4. Spline order, tension order, and the spectrum

272 With a model path (1), a penalty function (2), and a minimization condition (19), we have all the
 273 primary pieces to create a smoothing spline interpolant to the data. However, a number of choices
 274 still must be made. In this section we use synthetically generated data to represent our physical
 275 process, and contaminate the process with Gaussian noise as described in section 2c. We test our
 276 ability to recover the signal and examine the effects of changing the spline and tension order on
 277 the mean square error and the resulting spectrum.

278 The results of this section are empirical, and we acknowledge upfront that any conclusions
 279 reached *may* depend on our particular choice of physical model generating the signal. Neverthe-
 280 less, our expectation is the conclusions are ‘O(1)’ correct, and applicable, at least, to our GPS-
 281 tracked drifter dataset.

282 *a. Tension degree, T*

283 Given a smoothing spline of degree S , the tension in the penalty function (2) can be applied at
 284 any degree $T \leq S$. We use the synthetic data for the three different slopes to empirically establish
 285 the relationship between the tension degree, T and the spline degree, S .

286 For $S = 1 \dots 5$ and all $T \leq S$ we minimize the mean square error against the true values. The
 287 minimization is performed for 200 ensembles of noise and signal with three slopes (ω^{-2} , ω^{-3} ,
 288 ω^{-4}) and 5 different strides. For a given slope, stride, and realization of noise, we identify the
 289 minimum mean square error across S and T and compare all values of S and T as a percentage
 290 increase relative to that minimum. After aggregating across slopes, strides, and ensembles, the
 291 68% confidence range is shown in table 1. The table shows while setting $T = S$ is not always
 292 optimal, it is never significantly worse than the optimal choice. Thus for the remainder of the
 293 manuscript we set $T = S$.

294 *b. Loss of coherence*

295 The loss-of-coherence defines the time scale below which the smoothing spline is not providing
 296 useful information. A reasonable hypothesis is that this scale is related to the effective sample size,
 297 n_{eff} because this indicates how many points are being used to estimate the true value. Therefore
 298 the loss-of-coherence occurs at the *effective Nyquist* which we define as

$$f_s^{\text{eff}} \equiv \frac{1}{2n_{\text{eff}}\Delta t}. \quad (23)$$

299 In practice, we use $n_{\text{eff}}^{\text{SE}}$ because it is less variable than $n_{\text{eff}}^{\text{var}}$ for values near 1 and is the more direct
 300 measure of how many points are being used to estimate the model path. Fig. 4 shows the power
 301 spectrum and coherence of optimal tension fits for three different strides of the data. In each case
 302 (23) indicates the approximate value where the coherence drops below 0.5.

303 *c. Reduced spline coefficients*

304 One practical consideration when working with large datasets is the computational cost of creat-
305 ing the spline fit, which is limited by the rate of solving for the spline coefficients. It is beneficial
306 to reduce knot points (and therefore total splines) where possible. A reasonable strategy is that
307 when the effective sample size is large, as measured by (21), we avoid placing a knot point at
308 every data point—essentially ‘skipping’ data points.

309 To test this idea, we find the optimal fit over a range of different strides (which varies the effective
310 sample size) and increase the number of skipped knot points until the mean square error starts to
311 rise. We find we can skip $\max(1, \text{floor}(2n_{\text{eff}}/3))$ knot points without sacrificing precision. The
312 column labeled ‘optimal mse’ in table 2 indicates the optimal fit where one knot point is created
313 for every observation point, whereas the ‘reduced dof’ (reduced degrees of freedom) indicates a fit
314 where the number of knot points is reduced. In some cases the optimal mean square error improves
315 with fewer knot points. This means when handling large datasets, we can reduce the number of
316 splines being used if the effective sample size is large, and we can ‘chunk’ the data (split into
317 multiple independent pieces) when the effective sample size is small.

318 *d. Interpolation condition*

319 To estimate λ_T from (15), we estimate the mean square value of a derivative of the process,
320 $x_{\text{rms}}^{(T)}$ (see appendix C) and the effective sample size, n_{eff} . We argue that effective sample size
321 should vary based on the relative size of the measurement errors to the speed of motion. For
322 example, if the position errors are only 1 meter, but a particle typically travels 10 meters between
323 measurements, then it is hardly justifiable to increase the tension so that the smoothing spline
324 misses the observation points by 1 meter. There is not enough statistical evidence to suggest that
325 the particle did not go right through the observation point. On the other hand, if the position

errors are 1 meter, but the particle typically travels 10 centimeters between measurements, nearby measurements provide more information about the particle's true position during that time, so our estimate of the particle's true position is closer to a mean of the nearby observations.

This idea can be made more rigorous by stating that a change in position, Δx , is statistically significant if it exceeds the position errors σ by some factor. Assuming the physical process has a characteristic velocity scale, u_{rms} , we use this concept to define Γ as

$$\Gamma \equiv \frac{\sigma}{u_{\text{rms}} \Delta t}, \quad (24)$$

where Δt is the typical time between observations. This argument suggests effective sample size should be proportional to Γ , i.e.,

$$n_{\text{eff}}^{\Gamma} = \max(1, C \cdot \Gamma^m) \quad (25)$$

where C and m are unknown constants, and we prevent the effective sample size from dropping below 1. Intuitively this means as long as the particle does not move too far between observations, nearby observations help to estimate the true position of the particle.

To test the relationship between Γ and effective sample size, we compute the optimal smoothing spline for a range of values of Γ (created by sub-sampling the signal) for three different spectral slopes (ω^{-2} , ω^{-3} , ω^{-4}). The value $n_{\text{eff}}^{\text{SE}}$ is computed from the optimal solution for 50 ensembles and shown in Fig. 5. The fits are remarkably good, but depend on the slope. Processes with shallower slopes (rougher trajectories) yield a smaller effective sample size for a given value of Γ .

Using the interpolation condition Γ to estimate effective sample size, we set $n_{\text{eff}}^{\Gamma} = 14 \cdot \Gamma^{0.71}$, the empirically determined best fit for slope ω^{-3} . For all spline fits we use

$$\lambda_T^{\text{initial}} = \left(1 - \frac{1}{n_{\text{eff}}^{\Gamma}}\right) \frac{1}{\left(x_{\text{rms}}^{(T)}\right)^2} \quad (26)$$

as an initial estimate for the optimal smoothing parameter where $x_{\text{rms}}^{(T)}$ in (26) and u_{rms} in (24) are estimated using the method described in appendix C. The scaling law for n_{eff}^{Γ} can be estimated analytically. Let position observations be given by x_i where

$$x_i = u_{\text{rms}} i \Delta t + \varepsilon_i \text{ where } \varepsilon_i = \mathcal{N}(0, \sigma). \quad (27)$$

If the effective sample size is $\langle n \rangle$, then the particle changes position by $\langle n \rangle u_{\text{rms}} \Delta t$ between samples.

Applying the two-sample z -test, two positions will be considered different for $z > z_{\text{min}}$ where

$$z = \frac{\langle n \rangle u_{\text{rms}} \Delta t}{\sqrt{\frac{\sigma^2}{\langle n \rangle} + \frac{\sigma^2}{\langle n \rangle}}} \Rightarrow \langle n \rangle = \left(\frac{z \sigma \sqrt{2}}{u_{\text{rms}} \Delta t} \right)^{\frac{2}{3}}. \quad (28)$$

The power law in (28) is close to the empirically derived power laws shown in Fig. 5. This suggests the coefficient C in (25) can be related to z , a measure of statistical significance.

e. *Optimal fits*

Table 2 summarizes the key results of this section by applying a smoothing spline with $S = 3$ to 200 ensembles with three different slopes (ω^{-2} , ω^{-3} , ω^{-4}) and five different strides. The second and third columns show the effective sample size and average mean square error when the smoothing spline is applied using the true values (i.e., ‘unblinded’) to minimize the mean square error—this is the lower bound. The fourth column shows average increase in mean square error when reducing the number of spline coefficients as documented in section 4c. There is almost no change in mean square error and therefore all subsequent methods (whether blind or unblind) use this technique. The fifth column uses (26) from section 4d to provide a (blind) initial guess of the tension parameter. The results are mixed—a typical increase in mean square error is 30-50% when the effective sample size is large. While this seems large, this is a small fraction of the total noise variance, e.g., an optimal mean square error of 6m^2 increase to 8m^2 when the total variance is 100m^2 . Nearly optimal fits can be found using (19), as shown in the last column of the table.

364 *f. Numerical implementation*

365 The numerical implementation of the methods in this section are available in the
 366 SmoothingSpline class which subclasses BSpline. This class is initialized with three required
 367 parameters: a set of data points (t_i, x_i) and an error distribution.

368 **5. Bivariate smoothing splines and stationarity**

369 Up to this point we have considered univariate data, (t_i, x_i) , but GPS position data is fundamen-
 370 tally bivariate. The term ‘bivariate’ in the context of splines is often used to denote splines defined
 371 on two independent variables—however, in this context we define bivariate to mean two dependent
 372 variables (e.g., x and y) and one independent variable (e.g., t).

373 The trivial approach to working with bivariate data is to treat each direction independently—
 374 i.e., minimize λ_T^x and λ_T^y independently of each other. However, the underlying physical process
 375 is often isotropic. In the context of the maximum likelihood formulation of smoothing splines
 376 (18), this means we expect $x_{\text{rms}}^{(T)}$ (the rms value of the tensioned variable) to be the same in all
 377 directions (invariant under rotation). This however does *not* mean that λ_x should necessarily equal
 378 λ_y . To be explicit, if

$$\lambda_T^x = \left(1 - \frac{1}{n_{\text{eff}}^x}\right) \frac{1}{\left(x_{\text{rms}}^{(T)}\right)^2}, \quad \lambda_T^y = \left(1 - \frac{1}{n_{\text{eff}}^y}\right) \frac{1}{\left(y_{\text{rms}}^{(T)}\right)^2}, \quad (29)$$

379 then even if $x_{\text{rms}}^{(T)} = y_{\text{rms}}^{(T)}$, the effective sample sizes n_{eff}^x and n_{eff}^y may differ if there is any mean
 380 velocity because, as shown in section 4d, effective sample size depends on velocity.

381 Therefore to assume isotropy in λ_T and use a bivariate smoothing spline, the mean velocity from
 382 the underlying process must be removed. What qualifies as mean and fluctuation rarely has a clear
 383 answer, but a reasonable option is letting a polynomial of degree $T + 1$ define the mean. This has

the added benefit of removing a constant non-zero tension value, which as shown in section 3b, changes the problem formulation.

It is stationarity, not isotropy, that requires removing the mean velocity. The effective sample size is shown to be dependent on rms velocity, so if velocity varies in time, then the optimal effective sample size varies as well. This means not only do smoothing splines require stationarity in the tensioned variable $x^{(T)}$ as shown in section 3b, but they also require stationarity in the velocity $x^{(1)}$ to be effective. This last requirement can be solved by either removing the mean (as suggested here), or segmenting observations into locally-stationary chunks.

a. Assessing errors

Removing the mean or some other low-passed version of the data means the total smoothing matrix is a combination of the low-passed and high-passed smoothing matrices. Once this matrix is computed, it can be used to compute the standard errors.

We first create a low pass filter to capture the *mean* component of the flow using a simple polynomial fit $\bar{\mathbf{x}} = \bar{\mathbf{S}}\mathbf{x}$ and then define the residual as our stationary part, $\mathbf{x}' \equiv \mathbf{x} - \bar{\mathbf{x}}$. We now compute the smoothing spline as usual on the residual, $\mathbf{x}'_{\lambda} = \mathbf{S}_{\lambda}\mathbf{x}'$. So the total, smoothed path is

$$\hat{\mathbf{x}} = \bar{\mathbf{x}} + \mathbf{x}'_{\lambda} = \bar{\mathbf{S}}\mathbf{x} + \mathbf{S}_{\lambda}(\mathbf{x} - \bar{\mathbf{S}}\mathbf{x}) = (\bar{\mathbf{S}} + \mathbf{S}_{\lambda} - \mathbf{S}_{\lambda}\bar{\mathbf{S}})\mathbf{x} \equiv \mathbf{S}_T\mathbf{x} \quad (30)$$

From this we can compute the covariance matrix and the standard error.

b. Numerical implementation

The `BivariateSmoothingSpline` class is initialized with data (t_i, x_i, y_i) and a distribution. For a spline of degree $S = T$, a spline of degree $S + 1$ is used to remove the mean in each direction. With a Gaussian distribution this is simply a least squares polynomial fit. By assumption, the residual

404 data is stationary and isotropic, so the tension parameter λ_T is applied equally in each direction.
405 Minimization is performed on the sum of the expected mean square errors in each direction.

406 **6. GPS data set**

407 The primary dataset considered here are nine surface drifters deployed in the Sargasso Sea in
408 the summer of 2011 (Shcherbina et al. 2015). In the past, such drifters used the Argos positioning
409 system which has significantly poorer temporal coverage and position accuracy (Elipot et al. 2016),
410 but recently most surface drifters have employed GPS receivers and transmitted their data back
411 through Argos or Iridium satellites.

412 The GPS receiver sits on the surface drifter and collects position data, but because of atmospheric
413 conditions or ocean waves, the receivers are sometimes unable to obtain a position, or when they
414 do, it is highly inaccurate. Despite nominal accuracies of a few meters, it is often the case that
415 some positions are off by more than 1000 meters, as can be seen in Fig. 8. Applying a smoothing
416 spline fit using the methodology in section 3 produces an extremely poor fit, with clear overshoots
417 to bad data points.

418 *a. GPS error distribution*

419 We characterize the GPS errors by considering data from a motionless GPS receiver allowed to
420 run for 12 hours. The GPS receiver used in this test is not the same as the one used for the drifters
421 (because it was no longer available) but should produce errors similar enough for this analysis.

422 The position recorded by the motionless GPS are assumed to have isotropic errors with mean
423 zero, which means the positions themselves are the errors. The probability distribution function
424 (PDF) of the combined x and y position errors are shown in Fig. 6.

425 The error distribution is first fit to a zero-mean Gaussian PDF (10). The maximum likelihood fit
 426 is found by computing the standard deviation of the sample, which is found to be $\sigma \approx 10$ meters
 427 and shown as the gray line in Fig. 6. However, it is clear the error distribution shows much longer
 428 tails than the Gaussian PDF.

429 The Student t -distribution is a generalization of the Gaussian that produces longer tails and is
 430 defined as

$$p_s(\varepsilon|\nu, \sigma_s^2) = \frac{\Gamma(\frac{\nu+1}{2})}{\sigma_s \sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{\varepsilon^2}{\sigma_s^2 \nu}\right)^{-\frac{\nu+1}{2}}, \quad (31)$$

431 where the σ_s parameter scales the distribution width and the ν parameter sets the number of de-
 432 grees of freedom (as $\nu \rightarrow \infty$, equation 31 becomes a Gaussian). The variance is $\sigma^2 = \sigma_s^2 \frac{\nu}{\nu-2}$
 433 and only exists for $\nu > 2$. Minimizing the Anderson-Darling test to find the best fit t -distribution
 434 to the data, we find parameters $\sigma_s \approx 8.5$ meters and $\nu \approx 4.5$ shown as the black line in Fig. 6.
 435 Different choices in GPS receivers and using the Kolmogorov-Smirnoff test results in very similar
 436 parameters, i.e., $\sigma_s \approx 8 - 10$ meters and $\nu \approx 4 - 6$.

437 The *position* error distributions imply a combined *distance* error distribution by computing $\varepsilon_d =$
 438 $\sqrt{\varepsilon_x^2 + \varepsilon_y^2}$ and is shown in the lower panel of Fig. 6. With Gaussian noise this results in a Rayleigh
 439 distribution,

$$p_r(\varepsilon_d|\sigma_g) = \frac{\varepsilon_d}{\sigma_g^2} e^{-\frac{1}{2} \frac{\varepsilon_d^2}{\sigma_g^2}}, \quad (32)$$

440 as shown by the gray line. The distance error distribution from t -distributed noise is computed
 441 numerically and is shown by the black line. Around 95% of distance errors are within 30 meters.

442 Fig. 7 shows the autocorrelation function of the GPS position errors and the 99% confidence in-
 443 tervals. We find a rough empirical fit to be $\rho(\tau) = \exp(\max(-\tau/t_0, -\tau/t_1 - 1.35))$ where $t_0 = 100$
 444 seconds and $t_1 = 760$ seconds, which reflects an initially rapid fall off in correlation, followed by
 445 a slower decline. The smallest sampling interval of the GPS drifters in question is 30 minutes and

the correlation indistinguishable from zero according to Fig. 7. It is therefore safe to assume the errors are uncorrelated for our real data example. Although the drifter sampling rate allows us to avoid further discussion of the autocorrelation function of GPS errors, accounting for autocorrelation is a relatively easy extension (and is implemented in the code).

The smoothing spline algorithms described in section 3 are modified to use the t -distribution as described in appendix B. Table 3 shows the conclusions reached for Gaussian data in section 3 still apply with t -distributed data.

7. Minimization with Outliers

The goal here is to find a smooth solution in the presence of outliers—points that do not appear to be of the known error distribution for the GPS receiver shown in section 6a. These points are obviously problematic as can be seen in Fig. 8, where individual data points jump hundreds of meters and even several kilometers away from its neighbors. Errors of this size are inconsistent with the noise analysis of the preceding section, so the goal here is to find a model path $x(t)$ robust to this uncharacterized noise. What makes outliers ‘obvious’ to the eye is they appear as unexpectedly large motions, inconsistent with the other motion for that path. The smoothing spline formulation is therefore useful, as it assumes the motion at some order (e.g., acceleration) is Gaussian, as shown in section 3b. In the nine drifters we are analyzing here, one drifter shows no obvious outliers, suggesting the issue may be related to how the antennae are configured. This particular drifter serves as a useful point of comparison.

Minimizing with the expected mean square error (19) produces a fit so poor it is not worth showing. Because outliers add enormous amounts of variance, the expected mean square error vastly underestimates the spline tension—essentially chasing every outlier shown in Fig. 8. Because some of the noise is uncharacterized, this suggests using a method such as cross-validation might

be effective. The orange line in Fig. 8 uses a smoothing spline fit, assuming Student t-distributed errors, but minimized with cross-validation. This fit performs relatively well, but compared with the drifter 7, it is clear it still chases some outliers. The goal in this section is to develop a method robust to outliers in cases where we know something about the noise.

The basic problem formulation is as follows: we define a new ‘robust distribution’, p_{robust} , that includes the known noise distribution, p_{noise} , plus an unknown (or assumed) form of an outlier distribution, p_{outlier} ,

$$p_{\text{robust}}(\varepsilon) = (1 - \alpha) \cdot p_{\text{noise}}(\varepsilon) + \alpha \cdot p_{\text{outlier}}(\varepsilon). \quad (33)$$

We consider a t -distribution for p_{noise} with parameters found from the GPS errors in section 6a. The distribution of p_{outlier} is also set to be a t -distribution, but with $\nu = 3$ and $\sigma = 50\sigma_{\text{gps}}$ which roughly matches the total variance of the observed outliers. In our tests we varied α from 0 up to 0.25, approximately the range of observed outliers from the drifter data sets.

Throughout our attempts to smooth the noisy GPS data we tried many different approaches to modifying smoothing splines for robustness to outliers, but ultimately found enormous gains are made by simply discarding outliers while minimizing the expected mean square error (19). The results of this approach are shown in section 7a, and we document our methodology to reliably estimate the outlier distribution in section 7b.

a. Robust minimization

The challenge with outliers is we do not know their distribution, so minimizing the expected mean square error using (19) with the expected variance from the robust distribution defined in (33) does not work. Outliers add extra variance, and therefore cause the spline to be under tensioned (λ_T too small). Our method excludes the outliers from the calculation of (19), where outliers are defined as points unlikely to arise with the known noise distribution. The *ranked expected mean*

491 *square error* replaces σ^2 with,

$$\sigma_\beta^2 = \int_{\text{cdf}^{-1}(\beta/2)}^{\text{cdf}^{-1}(1-\beta/2)} z^2 p_{\text{noise}}(z) dz \quad (34)$$

492 and discards all rows (and columns) of \mathbf{S}_λ where $(\mathbf{S}_\lambda - I)\mathbf{x} < \text{cdf}^{-1}(\beta/2)$ or $(\mathbf{S}_\lambda - I)\mathbf{x} >$
 493 $\text{cdf}^{-1}(1 - \beta/2)$.

494 To test this approach we generated data as before, but allowed a certain percentage of outliers
 495 (α) to be generated with an outlier distribution following (33). We considered five values of $\beta =$
 496 $[\frac{1}{50}, \frac{1}{100}, \frac{1}{200}, \frac{1}{400}, \frac{1}{800}]$ as well as $\beta = 0$, which is just (19). Testing across a number of ensembles
 497 with outlier ratios $\alpha = [0.0, 0.05, 0.10, 0.25]$ we found $\beta = \frac{1}{100}$ is overall the best choice.

498 *b. Full tension solution and outlier distribution*

499 The *full tension* solution is defined as the maximum allowable value of λ given the known noise
 500 distribution. That is, the spline fit is pulled away from the observations so that the distribution
 501 of observed errors $(x_i - x(t_i))$ matches the expected distribution $p_{\text{noise}}(\epsilon)$. In cases where the
 502 effective sample size n_{eff} is large, the full tension solution approximately matches the optimal
 503 (minimal mean square error) solution. In cases where the effective sample size is small, the full
 504 tension solution is more akin to a low-pass solution (as increasing λ is equivalent to decreasing
 505 $x_{\text{rms}}^{(T)}$).

506 In the simplest case where there are no outliers, the full tension solution can be found by requir-
 507 ing the sample variance match the variance of $p_{\text{noise}}(\epsilon)$. When outliers are present, a more robust
 508 method of estimation is required. After some experimentation, we found the most reliable method
 509 of achieving full tension is to minimize the Anderson-Darling test of $p_{\text{noise}}(\epsilon)$ on the interquartile
 510 range of observed errors. This method can be used to estimate the outlier distribution and further

511 refine both the full tension solution and the range over which the expected mean square error is
 512 computed.

513 The outlier distribution is estimated as follows. We first assume the outlier distribution follows
 514 a t -distribution with $\nu = 3$ and $\alpha < 0.5$. If the spline is in full tension, then the observed total
 515 variance can be used to find σ_o for the outlier distribution. From (33):

$$\text{var}_{\text{total}} = (1 - \alpha)\text{var}_{\text{noise}} + \alpha 3\sigma_o^2 \quad (35)$$

516 which, given some α , can be solved for σ_o . Our method uses 100 values of α logarithmically
 517 spaced from 0.01 to 0.5 and chooses the value which minimizes the Anderson-Darling test. With
 518 an estimate for $p_{\text{robust}}(\epsilon)$, the full tension solution can be refined by minimizing the Anderson-
 519 Darling test of $p_{\text{robust}}(\epsilon)$ on the interquartile range of observed errors. This iterative process con-
 520 verges quite quickly to a good estimate for the outlier distribution and the full tension solution.

521 *c. Extension to bivariate data*

522 The strategies in this section are relatively easily extended to bivariate data. All error distribu-
 523 tions are assumed isotropic, and the outlier distribution can be estimated by including the errors
 524 from both independent directions. The ranged expected mean square error calculation defined in
 525 section 7a uses the *distance* of the error for its cutoff to remain invariant under rotation.

526 Application of this method to one of the GPS drifters (drifter 6) is shown in Fig. 8. Although it
 527 is impossible to know exactly how well the spline fit performed, comparison with drifter 7 (with
 528 no apparent outliers) suggests our methodology successfully avoids chasing outliers.

529 *d. Numerical implementation*

530 The GPSSmoothingSpline inherits from the BivariateSmoothingSpline class and assumes
531 errors follow a t -distribution found in section 6a. The class projects latitude and longitude using a
532 transverse Mercator projection with the central meridian set to the center of the dataset.

533 **8. Discussion**

534 The methods discussed in this manuscript are related to other methodologies used to smooth and
535 interpolate drifter trajectories.

536 Yaremchuk and Coelho (2015) formulate a cost function, their equation 9, based on PDFs of
537 the drifter accelerations and the GPS errors. Setting their $\mu = 1$ (they choose $\mu = 0.9$) this is
538 equivalent to the special case of equation 18 when $S = T = 2$, where they have implicitly chosen
539 λ_T by assuming an infinite effective sample size, n^{eff} . Their methodology for isolating outliers is
540 nearly equivalent to the iteratively reweighted least squares method detailed in appendix B using
541 a weight function similar to Tukey’s biweight, equation B6.

542 Elipot et al. (2016) apply their methodology to the ARGOS tracked surface drifter, which are
543 significantly noisier positions than GPS errors, but also follow a t -distribution. They assume a
544 linear model for positions, equivalent to assuming $S = T = 2$ with $\lambda_T \rightarrow \infty$. In the numerical
545 implementation of this manuscript, this special case is implemented in the ConstrainedSpline
546 class. The time dependent weight function used in Elipot et al. (2016) requires manually specifying
547 a weight for each point used, and this method is therefore somewhat different than the approach
548 taken here.

549 Another technique used for smoothing and interpolating drifter positions is kriging (Hansen and
550 Poulain 1996), however, its relationship to smoothing splines is less clear. In response to a study
551 empirically comparing kriging to smoothing splines (Laslett 1994), Handcock et al. (1994) point

552 out that kriging and smoothing splines are just two specific parameter choices of a more general
553 class of splines defined by their covariance functions. In the context of the maximum likelihood
554 equation for smoothing splines, equation 18, this generalization could be modeled by including a
555 covariance structure on the physical process.

556 9. Conclusions

557 The methodology in this manuscript solves our problem of finding smoothed, interpolated posi-
558 tions from a noisy GPS drifter dataset with outliers. More generally, for signals with second-order
559 structure similar to a Matérn process we found:

- 560 1. the spline degree S should be set to a value higher than the high frequency spectral slope of
561 the process (section 2),
- 562 2. the optimal tension parameter can be estimated *a priori* (section 4).

563 For GPS data, there appear to be three key steps for using smoothing splines:

- 564 1. using a t -distribution for the noise (section 6),
- 565 2. removing the mean velocity to make the bivariate data stationary (section 5), and
- 566 3. using the ranged expected mean square error for robustness to outliers (section 7).

567 *Acknowledgments.* Thanks to Miles Sundermeyer whose drifters were used in this analysis. This
568 work was funded by ONR through the Scalable Lateral Mixing and Coherent Turbulence Depart-
569 mental Research Initiative (LatMix) and National Science Foundation award 1658564.

570 APPENDIX A

571 Numerical implementation

572 The B-splines are generated using the algorithm described in De Boor (1978) with knot points
 573 determined by (7) and (8). The matrix \mathbf{X} with components X_m^i denotes the m -th B-spline at time
 574 t_i . The column vector ξ^m represents the coefficients of the splines such that positions at time t_i are
 575 given by \hat{x}^i where $\hat{x}^i = X_m^i \xi^m$.

576 The smoothing spline condition (16) can be augmented to include a nonzero mean tension, μ_u ,

$$\phi = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - x(t_i)}{\sigma_i} \right)^2 + \frac{1}{Q} \sum_{q=1}^Q \left(\frac{u(t_q) - \mu_u}{\sigma_u} \right)^2, \quad (\text{A1})$$

577 where we have taken $T = 1$ for this calculation. The discretized penalty function is

$$\phi = [\mathbf{x} - \mathbf{X}\xi]^T \Sigma^{-1} [\mathbf{x} - \mathbf{X}\xi] + \lambda_1 [\mathbf{V}\xi - \mu]^T [\mathbf{V}\xi - \mu], \quad (\text{A2})$$

578 where Σ denotes the covariance matrix describing the measurement errors and we absorbed several
 579 constants into λ_1 . To find the coefficients that minimize this function, we take the derivative with
 580 respect to ξ , set it to zero, and solve for ξ ,

$$\xi = [\mathbf{X}^T \Sigma^{-1} \mathbf{X} + \lambda_1 \mathbf{V}^T \mathbf{V}]^{-1} [\mathbf{X}^T \Sigma^{-1} \mathbf{x} + \mu \lambda_1 \mathbf{V}^T \mathbf{1}], \quad (\text{A3})$$

581 where $\mathbf{1}$ is a vector of 1s. The operation $\mathbf{V}^T \mathbf{1}$ essentially integrates the m -splines and results in a
 582 column vector with the integrated values.

583 We define the smoothing matrix as the linear operator that takes observations \mathbf{x} to their smoothed
 584 values $\hat{\mathbf{x}}$, $\hat{\mathbf{x}} = \mathbf{S}_\lambda \mathbf{x}$. From this definition and (A3),

$$\mathbf{S}_\lambda \equiv \mathbf{X} [\mathbf{X}^T \Sigma^{-1} \mathbf{X} + \lambda_1 \mathbf{V}^T \mathbf{V}]^{-1} \mathbf{X}^T \Sigma^{-1}, \quad (\text{A4})$$

585 when $\mu = 0$.

586 APPENDIX B

587 Iteratively reweighted least squares

588 Using the t -distribution is challenging because it does not result in a linear solution for the
 589 coefficients as in (A3). One solution is to use a search algorithm to directly look for maximum
 590 values. Alternatively, one can use iteratively reweighted least squares (IRLS).

591 The idea with IRLS is to reweight the coefficients of the Gaussian, σ_g in (10), so that the re-
 592 sulting distribution looks like the desired distribution, e.g., (31). Recalling $\varepsilon_i \equiv x_i - x(t_i, \xi)$, the
 593 minimization condition $\frac{dp_g}{d\xi} = 0$ implies that

$$\frac{\varepsilon_i}{\sigma_g^2} \frac{\partial x(t_i, \mathbf{x})}{\partial \xi} = 0, \quad (\text{B1})$$

594 for the Gaussian distribution, whereas for the t -distribution this implies

$$\frac{\varepsilon_i}{\sigma_s^2} \frac{\nu + 1}{\nu} \left(1 + \frac{\varepsilon_i^2}{\nu \sigma_s^2} \right)^{-1} \frac{\partial x(t_i, \mathbf{x})}{\partial \xi} = 0. \quad (\text{B2})$$

595 This means one can set

$$\sigma_g^2 = \sigma_s^2 \frac{\nu}{\nu + 1} \left(1 + \frac{\varepsilon_i^2}{\nu \sigma_s^2} \right), \quad (\text{B3})$$

596 to get a matching distribution. Of course, this is only true if ε_i is already known, which initially it
 597 is not. So the method becomes iterative—one starts with ε_i determined from the Gaussian fit, then
 598 determine a new ε_i after reweighting σ_g . This method iterates until σ_g stops changing. We can
 599 rewrite (B3) as a function of ε_i ,

$$w_s(\varepsilon_i) = \sigma_s^2 \frac{\nu + \frac{\varepsilon_i^2}{\sigma_s^2}}{\nu + 1}. \quad (\text{B4})$$

600 From (B4) it is clear if $\varepsilon_i < \sigma_s$ then it is reweighted to a smaller value, making the observation
 601 point more strongly weighted. On the other hand, if $\varepsilon_i > \sigma_s$, then its relative weighting decreases,
 602 and it is treated more as an outlier.

603 More generally, the weight function $w(z)$ for a pdf $p(z)$ is found by setting $-\partial_z \log p(z)$ equal to
 604 $-\partial_z \log p_g(z)$ of a Gaussian pdf where $w(z)$ replaces σ_g^2 , and then solving for $w(z)$. The result is

$$\frac{z}{w(z)} = -\frac{\partial_z p}{p} \Rightarrow w(z) = -z \frac{p}{\partial_z p}. \quad (\text{B5})$$

605 The same strategy could be used to reshape the pdf of a Gaussian to match the desired distribution,
 606 but here we simply match the minimization conditions of the pdfs.

607 As a point of reference, Tukey's biweight is given by

$$\psi(z) = \begin{cases} \frac{z}{\sigma_{tb}^2} \left(1 - \frac{z^2}{c^2 \sigma_{tb}^2}\right)^2 & |z| < c \cdot \sigma_{tb} \\ 0 & \text{else,} \end{cases} \quad (\text{B6})$$

608 which, as a weight function is

$$w_{tb}(\epsilon_i) = \frac{z}{\psi(z)}. \quad (\text{B7})$$

609 In a practical sense, Σ^{-1} in (A4) is replaced with the diagonal matrix $W \equiv \text{diag}(1/w(\epsilon_i))$ popu-
 610 lated with the reweighted values for each observation such that

$$\mathbf{S}_\lambda \equiv \mathbf{X} [\mathbf{X}^T \mathbf{W} \mathbf{X} + \lambda_1 \mathbf{V}^T \mathbf{V}]^{-1} \mathbf{X}^T \mathbf{W}. \quad (\text{B8})$$

611 This operator is used to compute the standard error from the variances, $\mathbf{S}_\lambda \Sigma$, where the variance is
 612 assumed to be $\sigma_s^2 \frac{v}{v-2}$ for each observation when using a t -distribution.

613 The smoothing spline solution *does* depend on the initial value of $w(\epsilon_i)$ used in the IRLS method.
 614 However, we find that for uniform initial weightings (e.g., all values start with the square root of
 615 the variance), the differences are not statistically significant from other initial values.

616 APPENDIX C

617 Estimating the variance of the signal

618 Our methodology requires good estimates of the root-mean-square velocity, u_{rms} , of the signal,
 619 to determine the effective sample size and variance of the tensioned derivative. Our approach is to
 620 compute the power spectrum of the signal at the derivative of interest, and sum the variance that is
 621 statistically significantly greater than the expected variance of the noise.

Given a process observed with values x_n at times $t_n = n\Delta$ where $n = 1..N$, we estimate the mean of its m -th derivative by performing a least squares fit to the polynomial $\bar{x}_n \equiv p_m t_n^m + p_{m-1} t_n^{m-1} + \dots + p_0$. The *detrended* time series is defined as $\tilde{x}_n \equiv x_n - \bar{x}_n$. The power spectrum of this time series is

$$S_{\text{signal}}(f_k) = \frac{\Delta}{N} \left| \sum_{n=0}^{N-1} x_n e^{-2\pi i f_k t_n} \right|^2, \quad (\text{C1})$$

where the frequencies f_k are given by $f_k = \frac{k}{N\Delta}$. By Plancherel's theorem,

$$\sum_{k=0}^{N-1} S(f_k) \cdot \frac{1}{N\Delta} = \frac{1}{N\Delta} \sum_{i=0}^{N-1} x_i^2 \Delta. \quad (\text{C2})$$

The power spectrum of the m -th derivative of the process is computed as

$$S_{\text{signal}}^{(m)}(f_k) = (2\pi f_k)^{2m} \cdot S(f_k). \quad (\text{C3})$$

It is important to detrend the signal prior to computing the derivative because, by assumption, the signal is periodic and has no secular trend.

The noise, ε_i , has total variance $\sigma^2 = \frac{1}{N} \sum_{i=1}^N \varepsilon_i^2$. Because the noise is assumed to be uncorrelated, the variance distributes evenly across all frequencies. The spectrum of the noise is therefore

$$S_{\text{noise}}(f_k) = \sigma^2 \Delta, \quad (\text{C4})$$

which immediately satisfies Plancherel's theorem (C2). The m -th derivative of the noise has power spectrum

$$S_{\text{noise}}^{(m)}(f_k) = \sigma^2 \Delta (2\pi f_k)^{2m}. \quad (\text{C5})$$

The technique used here sums the variance of the signal for a given frequency if it exceeds the expected variance of the noise at the frequency by some threshold. The estimate of power at each frequency follows a χ^2 distribution with 2 degrees-of-freedom, so we choose the threshold based on the 95-th percentile of the expected distribution. And thus,

$$x_{\text{std}}^{(m)} = \sum_{k=0}^{N-1} S_{\text{signal}}^{(m)}(f_k) \cdot \left(S_{\text{signal}}^{(m)}(f_k) > q S_{\text{noise}}^{(m)}(f_k) \right) \cdot \frac{1}{N\Delta}, \quad (\text{C6})$$

638 where $q \approx 20$ for the 95-percent confidence.

639 References

640 Bracco, A., J. H. LaCasce, C. Pasquero, and A. Provenzale, 2000: The velocity distribution of
641 barotropic turbulence. *Phys. Fluids*, **12** (10), 2478.

642 Cantoni, E., and T. Hastie, 2002: Degrees-of-freedom tests for smoothing splines. *Biometrika*,
643 **89** (2), 251–263.

644 Craven, P., and G. Wahba, 1979: Smoothing noisy data with spline functions. *Numer. Math.*,
645 **31** (4), 377–403.

646 De Boor, C., 1978: *A practical guide to splines*, Vol. 27. Springer-Verlag New York.

647 Elipot, S., R. Lumpkin, R. Perez, J. J. Early, and A. M. Sykulski, 2016: A global surface drifter
648 dataset at hourly resolution. *J. Geophys. Res. Oceans*.

649 Green, P. J., and B. W. Silverman, 1994: *Nonparametric Regression and Generalized Linear Mod-*
650 *els: A Roughness Penalty Approach*. Chapman & Hall, London.

651 Handcock, M. S., K. Meier, and D. Nychka, 1994: Kriging and Splines: An Empirical Compar-
652 ison of Their Predictive Performance in Some Applications: Comment. *Journal of American*
653 *Statistical Association*, **89** (426), 401–403.

654 Hansen, D. V., and P.-M. Poulain, 1996: Quality control and interpolations of WOCE-TOGA
655 drifter data. *J. Atmos. Ocean. Tech.*, **13**, 900–909.

656 Laslett, G. M., 1994: Kriging and Splines: An Empirical Comparison of Their Predictive Perfor-
657 mance in Some Applications. *Journal of American Statistical Association*, **89** (426), 391–400.

- 658 Lee, T. C. M., 2003: Smoothing parameter selection for smoothing splines: a simulation study.
659 *Comput. Stat. Data Anal.*, **42** (1-2), 139–148.
- 660 Lilly, J. M., A. M. Sykulski, J. J. Early, and S. C. Olhede, 2017: Fractional Brownian motion, the
661 Matérn process, and stochastic modeling of turbulent dispersion. *Nonlin. Processes Geophys.*,
662 **24** (3), 481–514.
- 663 Nychka, D., 2000: Spatial Process Estimates as Smoothers. *Smoothing and Regression. Ap-
664 proaches, Computation and Application*, M. G. Schimek, Ed., 393–424.
- 665 Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in C*.
666 2nd ed., The Art of Scientific Computing, Cambridge University Press.
- 667 Reinsch, C. H., 1967: Smoothing by spline functions. *Numer. Math.*, **10** (3), 177–183.
- 668 Shcherbina, A. Y., and Coauthors, 2015: The LatMix Summer Campaign: Submesoscale Stirring
669 in the Upper Ocean. *Bull. Amer. Meteor. Soc.*, **96** (8), 1257–1279.
- 670 Sykulski, A. M., S. C. Olhede, J. M. Lilly, and E. Danioux, 2016: Lagrangian time series models
671 for ocean surface drifter trajectories. *J. R. Statist. Soc. C*, **65** (1), 29–50.
- 672 Teanby, N. A., 2007: Constrained Smoothing of Noisy Data Using Splines in Tension. *Math Geol*,
673 **39** (4), 419–434.
- 674 WAAS T&E Team, 2016: Global Positioning System (GPS) Standard Positioning Service (SPS)
675 Performance Analysis Report. Tech. Rep. 92, William J. Hughes Technical Center.
- 676 Wahba, G., 1978: Improper priors, spline smoothing and the problem of guarding against model
677 errors in regression. *J. R. Statist. Soc. B*.

- 678 Whittaker, E. T., 1923: On a New Method of Graduation. *Proceedings of the Edinburgh Mathe-*
679 *matical Society*, **41 (01)**, 63–75.
- 680 Yaremchuk, M., and E. F. Coelho, 2015: Filtering Drifter Trajectories Sampled at Submesoscale
681 Resolution. *IEEE J. Oceanic Eng.*, **40 (3)**, 497–505.

682	LIST OF TABLES	
683	Table 1.	68th percentile range of increase in mean square error from the optimal fit 39
684	Table 2.	Mean square error and effective sample size for a range of strides and smooth-
685		ing spline methods. 40
686	Table 3.	Same as table 2, but with noise following a t distribution. 41

TABLE 1. 68th percentile range of increase in mean square error from the optimal fit

S	T				
	1	2	3	4	5
1	33.8-80.3%				
2	14.0-75.1%	0.8-12.1%			
3	17.1-77.5%	1.0-13.1%	0.0-4.5%		
4	22.8-81.9%	1.0-14.5%	0.0-4.6%	0.0-6.3%	
5	27.6-91.4%	0.8-15.4%	0.0-4.6%	0.0-6.1%	0.0-12.8%

TABLE 2. Mean square error and effective sample size for a range of strides and smoothing spline methods.

stride	n_{eff}	optimal mse	reduced dof	blind initial	expected mse
ω^{-2}					
1	8.6	11.5 m ²	0.1%	56.4%	7.4%
2	4.9	20.4 m ²	0.0%	36.3%	2.8%
4	2.9	34.2 m ²	0.1%	20.0%	1.7%
8	1.7	55.9 m ²	0.0%	5.6%	1.0%
16	1.2	81.8 m ²	0.0%	3.6%	0.5%
ω^{-3}					
1	12.5	7.64 m ²	-0.1%	38.6%	6.4%
2	7.1	13.4 m ²	-0.1%	20.4%	3.5%
4	4.1	23.5 m ²	-0.0%	9.8%	2.2%
8	2.3	41.8 m ²	0.0%	1.7%	1.2%
16	1.4	67.9 m ²	0.0%	9.6%	0.6%
ω^{-4}					
1	15.6	5.69 m ²	-0.1%	33.8%	7.9%
2	9.0	10.5 m ²	-0.1%	18.6%	5.1%
4	5.0	18.6 m ²	-0.0%	8.6%	2.4%
8	2.8	33.2 m ²	0.0%	3.2%	1.5%
16	1.6	57.6 m ²	0.0%	15.4%	0.8%

TABLE 3. Same as table 2, but with noise following a t distribution.

stride	n_{eff}	optimal mse	reduced dof	blind initial	expected mse
ω^{-2}					
1	8.2	11.8 m ²	0.3%	66.7%	7.7%
2	4.7	20.9 m ²	0.3%	47.3%	6.6%
4	2.8	38.0 m ²	0.1%	24.2%	4.4%
8	1.6	66.3 m ²	0.0%	8.2%	9.3%
16	1.2	101. m ²	0.0%	8.1%	3.7%
ω^{-3}					
1	12.1	7.51 m ²	-0.1%	36.2%	8.8%
2	6.8	13.4 m ²	-0.1%	22.8%	7.0%
4	3.9	26.0 m ²	-0.0%	11.5%	3.8%
8	2.2	47.5 m ²	0.0%	2.2%	3.2%
16	1.3	82.5 m ²	0.0%	12.6%	8.5%
ω^{-4}					
1	14.9	6.01 m ²	-0.2%	35.3%	9.0%
2	8.6	10.5 m ²	-0.2%	24.8%	7.0%
4	4.8	19.1 m ²	-0.1%	7.8%	4.6%
8	2.7	36.4 m ²	0.0%	3.2%	2.7%
16	1.6	69.1 m ²	0.0%	18.9%	11.5%

LIST OF FIGURES

687			
688	Fig. 1.	An example of interpolating between 7 data points using spline function of order K . The data points are shown as circles, and the interpolated function is shown as solid black lines. We show four different orders of interpolation $K = 1..4$ (rows) and their nonzero derivatives (columns). The thin vertical grey lines are the knot points.	43
689			
690			
691			
692	Fig. 2.	B-splines and derivatives (columns) for orders $K = 1..4$ (rows).	44
693	Fig. 3.	The upper panel shows the velocity spectrum of a synthetic Lagrangian velocity generated from the Matérn (black). The blue, red, and orange lines show the spectrum of the interpolating spline fit to the data with a stride of 100 for $S = 1..4$, respectively. The dashed vertical line denotes the Nyquist frequency of the strided data. The bottom panel shows the coherence between the smoothed velocities and the true velocity.	45
694			
695			
696			
697			
698	Fig. 4.	The upper panel shows the uncontaminated velocity spectrum of the signal (black) and velocity spectrum of the noise (red). The observed signal is the sum of the two. The blue, red, and orange lines show the spectrum of the smoothing spline best fit to the observations with all, 1/10th and 1/100th the data, respectively. The vertical dashed lines show the effective Nyquist computed using equation 23. The bottom panel shows the coherence between the smoothed signals and the true signal.	46
699			
700			
701			
702			
703			
704	Fig. 5.	Effective sample size from the standard error vs Γ	47
705	Fig. 6.	The top panel shows the position error distribution of the motionless GPS. The gray/black lines are the best fit Gaussian/ t -distributions respectively. The bottom panel shows the distance error distribution with the corresponding expected distributions from the Gaussian and t -distribution. The vertical line in the bottom panel shows the 95% error of the t -distribution.	48
706			
707			
708			
709	Fig. 7.	The autocorrelation function of the GPS positioning error with 99% confidence intervals shown in gray. The correlation at drifter sampling period of 30 minutes is indistinguishable from zero.	49
710			
711			
712	Fig. 8.	GPS position data for a 40 hour window from drifter 6. The points are the recorded positions and the black line is the optimal fit using the ranged expected mean square error. Data points with less than 0.01% chance of occurring are highlighted and deemed outliers. The light grey line is the is optimal smoothing spline fit for drifter 7, which has no apparent outliers and was released a few hundred meters from drifter 6. The orange line is the smoothing spline fit assuming t -distributed errors, but using cross-validation to minimize λ_T	50
713			
714			
715			
716			
717			

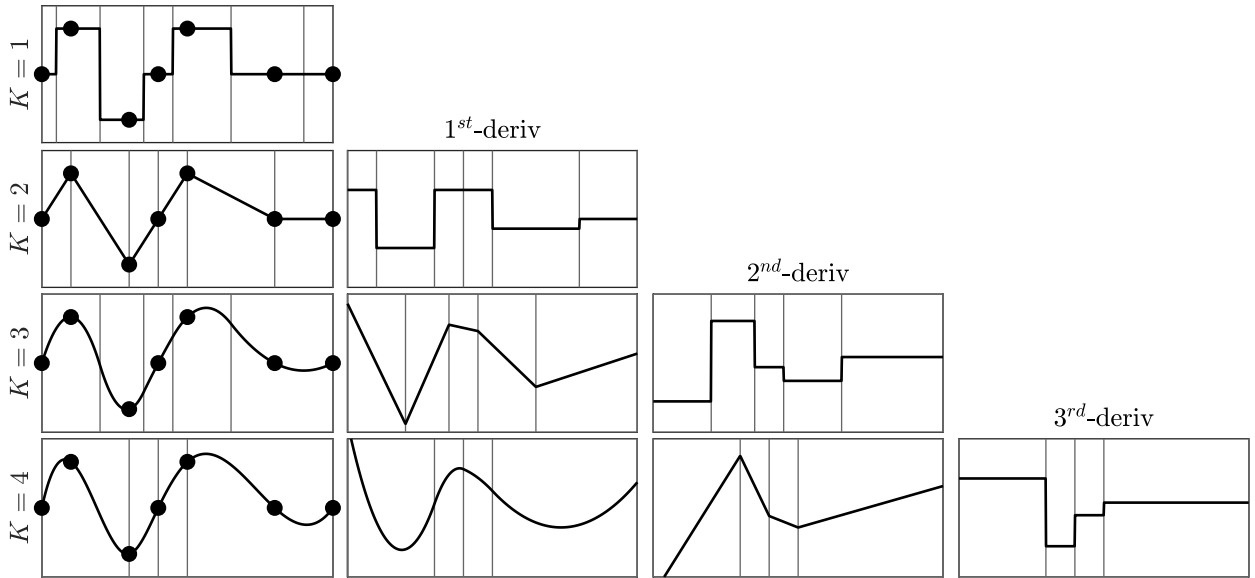


FIG. 1. An example of interpolating between 7 data points using spline function of order K . The data points are shown as circles, and the interpolated function is shown as solid black lines. We show four different orders of interpolation $K = 1..4$ (rows) and their nonzero derivatives (columns). The thin vertical grey lines are the knot points.

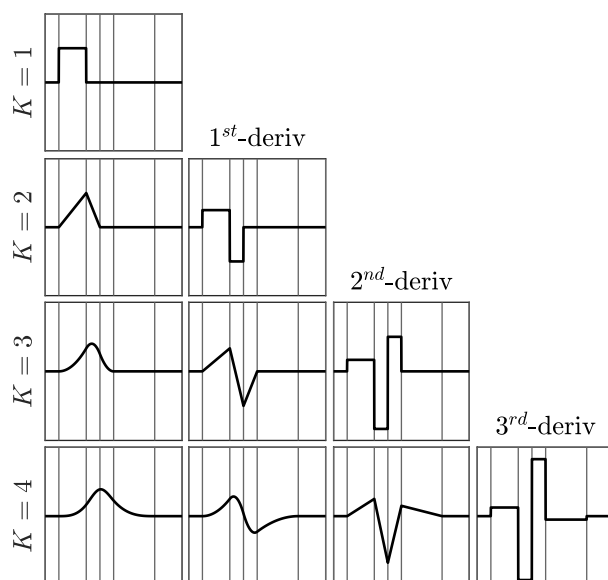
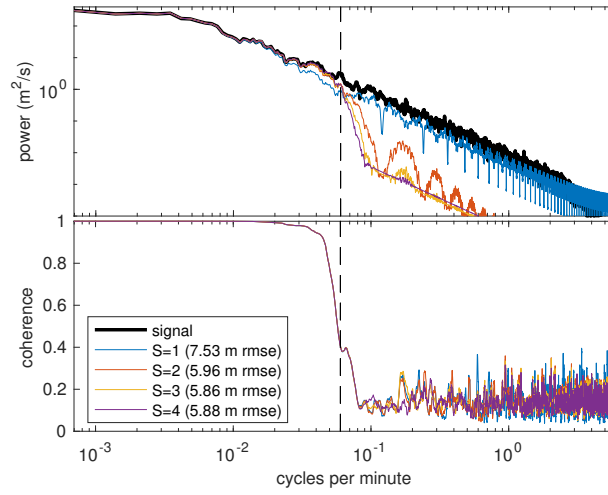
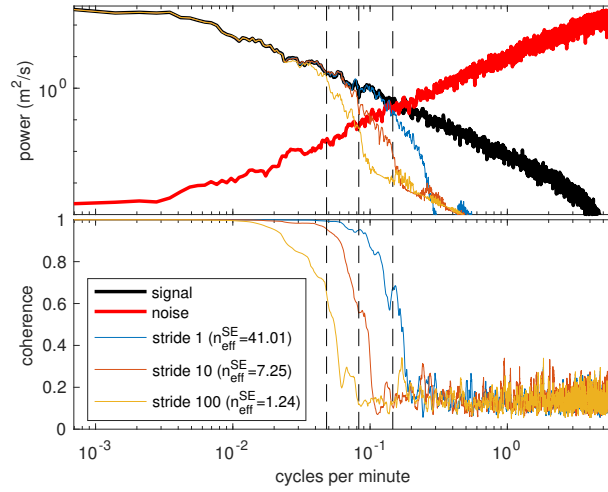


FIG. 2. B-splines and derivatives (columns) for orders $K = 1..4$ (rows).



722 FIG. 3. The upper panel shows the velocity spectrum of a synthetic Lagrangian velocity generated from the
 723 Matérn (black). The blue, red, and orange lines show the spectrum of the interpolating spline fit to the data with
 724 a stride of 100 for $S = 1..4$, respectively. The dashed vertical line denotes the Nyquist frequency of the strided
 725 data. The bottom panel shows the coherence between the smoothed velocities and the true velocity.



726 FIG. 4. The upper panel shows the uncontaminated velocity spectrum of the signal (black) and velocity
 727 spectrum of the noise (red). The observed signal is the sum of the two. The blue, red, and orange lines show the
 728 spectrum of the smoothing spline best fit to the observations with all, 1/10th and 1/100th the data, respectively.
 729 The vertical dashed lines show the effective Nyquist computed using equation 23. The bottom panel shows the
 730 coherence between the smoothed signals and the true signal.

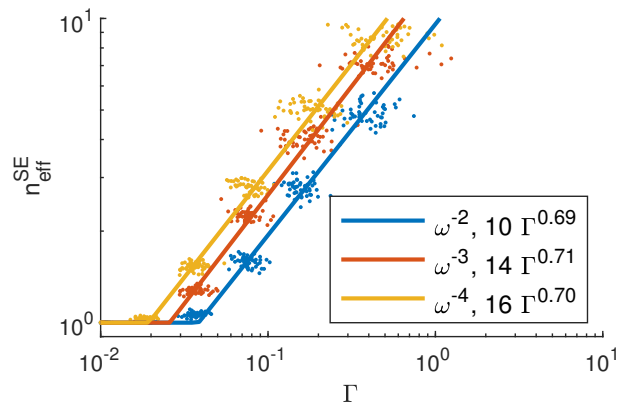


FIG. 5. Effective sample size from the standard error vs Γ

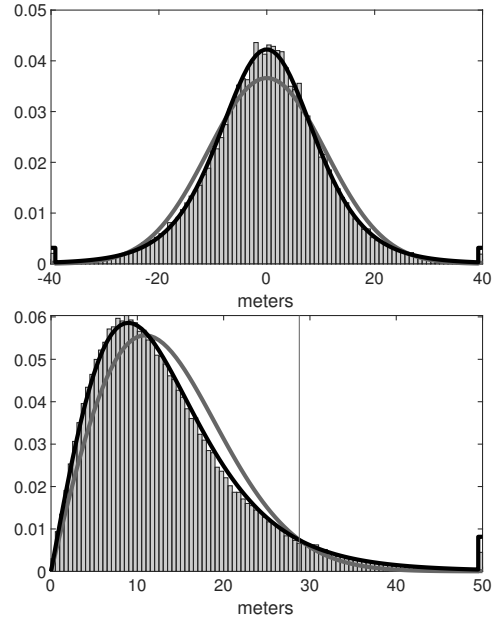


FIG. 6. The top panel shows the position error distribution of the motionless GPS. The gray/black lines are the best fit Gaussian/ t -distributions respectively. The bottom panel shows the distance error distribution with the corresponding expected distributions from the Gaussian and t -distribution. The vertical line in the bottom panel shows the 95% error of the t -distribution.

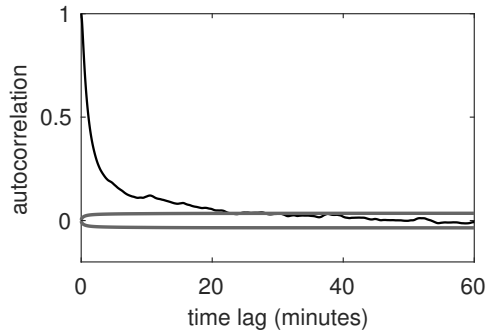
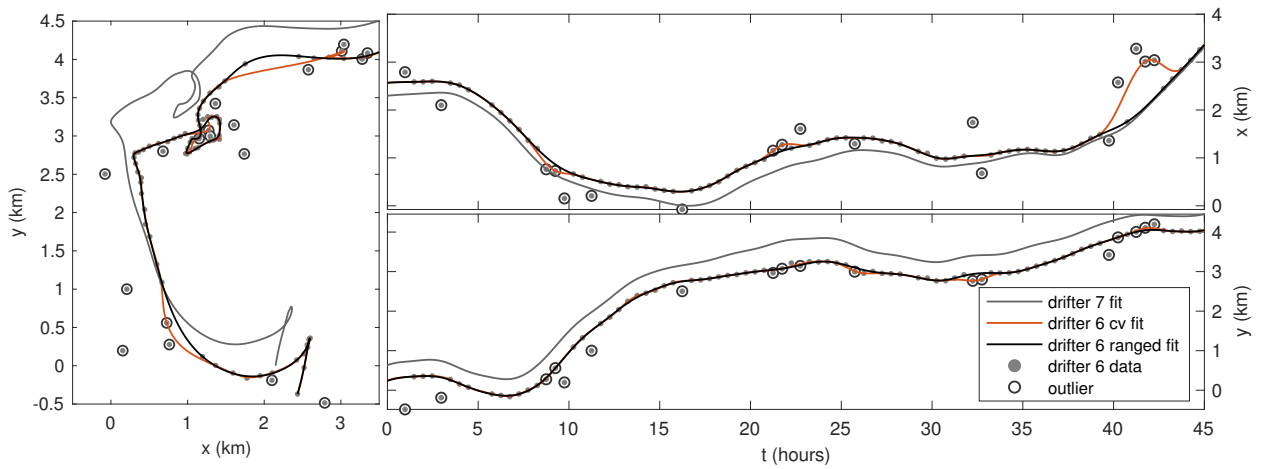


FIG. 7. The autocorrelation function of the GPS positioning error with 99% confidence intervals shown in gray. The correlation at drifter sampling period of 30 minutes is indistinguishable from zero.



737 FIG. 8. GPS position data for a 40 hour window from drifter 6. The points are the recorded positions and
 738 the black line is the optimal fit using the ranged expected mean square error. Data points with less than 0.01%
 739 chance of occurring are highlighted and deemed outliers. The light grey line is the is optimal smoothing spline
 740 fit for drifter 7, which has no apparent outliers and was released a few hundred meters from drifter 6. The orange
 741 line is the smoothing spline fit assuming t-distributed errors, but using cross-validation to minimize λ_T