**Review of the manuscript JTECH-D-19-0087-R1, “Smoothing and Interpolating Noisy GPS Data with Smoothing Splines”, by Jeffrey J. Early and Adam M. Sykulski.**

The submitted manuscript presents methods for smoothing and interpolating noisy, irregularly sampled data using smoothing splines with variable tension. In particular, the authors demonstrate how to obtain optimum choices for adjustable parameters of the spline fit, notably the spline degree *S* and tension *T*, and how to deal with the non-Gaussian noise intrinsic to GPS position fixes obtained by surface ocean drifters. The work provides an extensive and careful description of the proposed methods, with numerical implementations also being made available by the authors. With the increasing number of deployments of GPS-tracked drifters in recent years, I believe that this study is of interest for the oceanography community and will be a valuable contribution to JTECH pending a Major Revision.

The manuscript rushes through concepts that I believe are key for the interpretation of the results, and many statements along the text need rephrasing or clarification. I also believe that it lacks a discussion about potential advantages/drawbacks of the presented methods relative to other procedures available in literature, which would be useful for informing readers about which methods are better suited for their needs. Finally, it is noted that the manuscript in its present form has 34 double-spaced pages, exceeding the 26-page limit stablished by the journal and thus requiring preliminary approval from the Chief Editor for publication with the page overage (personally, I have no objections to the paper length).

Please find below my major and minor comments.

1. **Major comments**
2. **The manuscript exceeds the 26-page limit for JTECH submissions:**

Excluding references, tables, figures, and the title and abstract pages, the manuscript currently has 34 double-spaced pages, thus exceeding the approximate 26-page limit stablished for JTECH submissions. According to the journal’s formatting instructions, the authors should prepare a justification for the length of the manuscript on the submission’s cover letter, and request the Chief Editor’s approval for exceeding the page limit.

1. **Discussion of advantages/caveats of proposed approach relative to other methods:**

The manuscript currently lacks a discussion on the potential advantages and drawbacks of using smoothing splines following the proposed procedure relative to other processing methods described in literature, such as those applied to Global Drifter Program (GDP) drifters (Hansen and Poulain, 1996), to recent massive drifter deployments in the Gulf of Mexico (Yaremchuk and Coelho, 2015), and even for the generation of an GDP drifter dataset set at an hourly resolution in a previous paper by the authors (Elipot et al. 2016). I think that including such discussion could help informing readers of which methods are more suitable for their needs, and thus add value to this work. For example, it seems to me that the GPS position jumps described by the authors should probably be more severe in regions with intense wind and wave activity such as the Southern Ocean. Does that mean that their methods would outshine others for data collected in such regions? Please elaborate.

I also believe that the drifter dataset described in the study has been under-explored. The manuscript mentions that it contains position observations from 9 drifters, and yet results for only 2 of them (drifters 6 and 7) are presented. If within the reach of the authors, I suggest using the remaining observations to evaluate the proposed methods against other procedures described in the literature.

* Elipot, S., R. Lumpkin, R. C. Perez, J. M. Lilly, J. J. Early, and A. M. Sykulski (2016), A global surface drifter data set at hourly resolution, *J. Geophys. Res. Oceans*, 121
* Hansen, D., and Poulain, P.M. (1996), Quality control and interpolations of WOCE-TOGA drifter data. *J. Atmos. Ocean. Technol.*, 900–909.
* Yaremchuk, M., and E. F. Coelho (2015), Filtering drifter trajectories sampled at submesoscale resolution, *IEEE J. Oceanic Eng.*, 40(3), 497–505.

1. **Practical examples of the numerical implementation of the described methods:**

I downloaded the Matlab classes developed by the authors at <https://github.com/JeffreyEarly/GLNumericalModelingKit>. I think that the readers of this study would appreciate if practical examples for processing the set of drifter observations described in the study were also made available.

1. **Minor comments:**

* It is often unclear which Sections are referenced along the text (e.g. lines 146, 286, 405, 410, 416, 486, 493, 514, 515, 583, etc), specifically because the Subsection letter is cited unaccompanied from the Section number. I think that this may be caused by a bug on the AMS Latex package, considering that this issue is absent in the arXiv version of the paper. Please proof-read the manuscript, correcting this issue where needed;
* A similar problem is observed on how the Appendices are referenced (e.g. lines 248, 251, 335, etc), which are shown with a number instead of a capital letter. Please correct all occurrences of this issue.
* Caption of Figure 1: Please expand the first period as “An example of interpolating between 7 data points using spline functions of order *K*”;
* Line 101: For clarity, please modify this sentence as “All higher order (*K*1) B-splines are defined by recursion”;
* Equation (9): Please explicitly define all the equation’s variables (, , *p* and *A*) in the text immediately following the equation.
* Lines 140-141: It is mentioned that the Matérn spectrum is used to the generate the velocity of the signal. Please provide further details on how this is done.
  + In particular, Equation (9) for me suggests that the spectrum varies solely as a function of frequency , using prescribed magnitude *A* and damping scale . Is that so, or is *A* instead obtained stochastically? Also, is the signal speed obtained simply by taking the inverse Fourier transform of (9)? Please elucidate.
* Line 140: It is also mentioned that the Matérn spectrum is used to obtain the *velocity* of the signal. However, considering that you are representing the evolution of a particle along a single positive spatial dimension, wouldn’t the term *speed* thus be more suitable, in this case?
* Line 141: It is mentioned that the velocities are integrated to obtain the positions. Please state at which time increment the positions are computed;
* Lines 148-150: It is unclear for me how the quality of the fit is assessed. Is the interpolated versions of the subsampled data evaluated against the original (not-subsampled) signal by computing the RMS error between them? If so, please rephrase these lines to make this message go through more clearly.
* Figure (3):
  + Top panel y-axis label: I think that stating the speed power spectral density units as (m2/s) is a bit counter-intuitive. I suggest changing it to [(m/s)2/cpm];
  + Bottom panel legend: The RMSE units is set to meters, in contrast with the top panel, that shows the power spectral density for speed. Please double-check what time-series is used in the coherency analysis. I suggest using the velocity time-series, for consistency with the top panel;
  + Caption: “…shows the velocity spectrum of the signal (black).” To make the caption self-explanatory, please explain that the signal correspond to the synthetic Lagrangian velocities built using the Matérn spectrum.
* Lines158-165: I don’t think that many of your readers will be immediately familiar with magnitude-square coherency estimates. Please briefly explain how to interpret this statistic, after introducing it in the text.
* Lines 161-163: “This is why the shallower slopes (with more variance at high, incoherent frequencies) have a larger mean square error than the steeper slopes (with less variance at high, incoherent frequencies)”. I was confused by this sentence, as it initially lead me to believe that the errors were computed using the interpolated signals preliminarily high-pass filtered to isolate frequencies higher than the Nyquist frequency of the strided data (!!). Please reformulate it, for clarity.
* Line 169: The i.i.d acronym is used, which refers to an *independent and identically distributed random variable*. While this acronym might be common place for the statistics community, I think that a large fraction of the JTECH audience will not be familiar with it. Please reformulate the sentence on this line avoiding the use of this acronym.
* Lines 175-176: “– up to a constant this is the log likelihood”. Unclear (I think there is a missing comma after “constant”). Please reformulate, for clarity.
* Lines 184-185: The tone of this sentence seems excessively informal, and its construction can obscure the message the authors want to convey. Please reformulate, for clarity.
* Equations (19) and (20): the variable “**I**” is undefined in text;
* Caption of Figure 4: Please state what the vertical dashed lines correspond to;
* Line 315: If the vertical dashed lines in Fig. (4) correspond to *f*eff, than stating that this parameter “indicates almost exactly where the coherence drops below 0.5” is a stretch, as it visually intercepts values closer to 0.6 for the yellow and orange lines and to 0.7 for the blue. Please reformulate this sentence;
* Line 326: Table 2 does not include a “Full dof” column. Please double-check both the text and the information provided in Table 2;
* Line 326: The acronym “dof” is undefined. I assume it refers to “degrees of freedom”, but I think this should be explicitly stated along the text;
* Lines 355-356: Please add “spectral” to the sentence “… for three different spectral slopes …”;
* Line 378 and Table 2: It is unclear what is meant by “blind” and “unblind” methods. Please define this along the text;
* Lines 437-441: The text here described the drifter GDP dataset: I believe it should further include information on (a) the sampling rates of the GPS receivers, (b) the length of the used records, and (c) the temporal increment the data was interpolated to.
* Lines 471-473: Here, it is mentioned that two *t*-distributions are shown in the bottom panel of Fig. 6. However, this contrasts with Fig. 6 caption, which mentions that the gray line correspond to a Gaussian distribution, and the black to a *t*-distribution. Please double-check both the caption and the text, and remove the ambiguities.
* Lines 474-478: From the information provided by this paragraph alone, it is unclear for me why one can conclude that it is safe to assume that the 30-min sampling interval leads to statistically independent observations. Please rephrase this sentence for clarity, perhaps further mentioning that the autocorrelation function becomes statistically undistinguishable from zero within 30 min;
* Line 498: I suggest substituting the sentence “vastly under tensions the spline” to “vastly underestimates the spline tension”;
* Line 499: “… this suggests that using a method…”;
* Lines 572-573: For clarity, I suggest substituting the sentence “For signals similar to the Matérn process …” to “For signals with spectral characteristics compatible with those of the Matérn process”.