General Comments

The authors present a novel kernel density estimation approach that combines a diffusion-based method with a pilot step to produce better estimates of the bandwidth - a classically difficult problem. The authors also develop a Python package, *diffKDE*, for easy use of their method and highlight some of the package's capabilities within the manuscript. diffKDE is benchmarked against pre-existing KDE methods using synthetic data and applied to real marine data, highlighting its applicability. However, the material is organized and presented in such a manner that detracts from its scientific contribution. Additionally, as noted in the points below, the verbiage, lack of technical definitions in multiple locations throughout the text, and sectioning of the ideas muddle the presentation of the material. I recommend the authors revise the manuscript and give special attention to the structure of the manuscript and the presentation clarity. I believe this work has potential and I look forward to the reading the revised manuscript.

Specific Comments

1. The organization of the paper detracts from the scientific contribution. Here are a couple examples:

   (a) Section 2.1 gives background on the general kernel density estimator *and* the proposed diffusion-based kernel density estimator with pilot study. Then section 2.2 and 2.3 discuss discretization of the diff-KDE and implementation, respectively. This feels out of order. I would assume general background would be its own section (e.g., make section 2.1 its own section) and the proposed method and *all* associated ideas would be grouped together (e.g., make sections 2.1.2 through 2.4 its own section, say section 3, and order these appropriately).

   (b) Regarding the order material is presented, there is no initial explanation as to why a pilot estimation is needed. It feels as though section 2.1.4 needs some prior context or motivation.

   (c) Section 3 contains the simulation example and real world example. I would makes these distinct sections (e.g., following the above point, sections 4 and 5). To this point, the first paragraph of section 3 is confusion. There is no clear distinction between the simulation explanation and the real world example. By separating these, that should mitigate any confusion.

   I strongly recommend reorganizing the paper by grouping similar ideas by sections with appropriate sub-sections.

2. Some of the verbiage and notation used in this paper may preclude the general audience of GMD from understanding its scientific merit. There are some examples included in the list below, but I recommend the authors check the manuscript and make sure *all* mathematical terms are defined and avoid using overly technical notation if possible.

3. Abstract: Some of the sentences need some work. For example:

(a) Line 1: A PDF is a function defining the probability of a random variable taking on a specific value. I don't like the use of observed or simulated variables or the phrase, "comprise basic information", as I think this is a little too weak of wording.

(b) Line 6: Starting with "A diffusion-based KDE..." Why is this the case? Maybe rephrase to say, "Diffusion-based KDEs have been shown to provide a useful approach ..."

(c) Line 10: What is boundary close data and what details are suppressed? Do you mean your approach produces a smooth surface that is robust to noise and outliers? Also, there are various other places in the manuscript that use "details", but it is unclear what is being referred to.

4. The introduction falls a short. There is a nice body of literature on diffusion-based KDE and it would be beneficial to this manuscript if it is put into context with the current literature.

5. Line 19: Any citations to support the claim for the need in other model applications? This seems like a great place to include more references.

6. Lines 24-27: Your definition of a PDF does not seem correct. You need to explicitly define each element of you probability space $\Omega, \mathcal{A}, P$. Also, I assume $\mathcal{A}$ is a σ-algebra and Ω is the sample space. In which case I think you mean to have $f : \Omega \to [0, \infty)$ (also, open bracket on the right). Also, the sentence, "By definition, ..." should be reworded.

7. Line 41: I like the list, but are there any more current (last 5 or so years) papers that could be included? A cursory literature search returned quite a few recent papers on the topic.

8. Line 48: Define δ-distribution, or say something like, "which will be formally defined in section xxx.".

9. Line 75: Are all X_j assumed to come from the same distribution or are they completely independent? Should they be independent or independent and identically distributed?

10. Line 78: Is n the same as N? If so, verify the use of n and N throughout the paper.

11. Line 78 and generally: This is a personal choice, but I feel your paper would be more accessible to the general GMD audience if you change your function notation. Consider

$$\hat{f}_n(x; h) = \frac{1}{nh} \sum_{j}^{n} K\left(\frac{x - X_j}{h}\right) \quad \text{where } \hat{f} : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}.$$

12. Line 86: Should it be $\mathbb{R}_{>0}$ or $\mathbb{R}_{\geq 0}$?

13. Line 91: It is not clear where equation (6) comes from. It would be helpful if this result, or derivation, is included in the appendix.

14. Lines 98 and 100: see comment 11.

15. Line 109: Is the diffusion based approach an alternative method? Better? Also, different approach in what sense? Make this more clear.

16. Line 110: What is “It”? The KDE? Also, “progresses up to an estimate at the final time,” is unclear. What progresses and what estimate?

17. Line 113: see comment 11.

18. Line 121: What is the importance of acting inversely proportional to the diffusion quotient? Expand on this point.

19. Line 136: How do you set p properly? It is not clear at this point. (E.g., Line 166 starting with “Choosing p to be ...” could be incorporated here)

20. Line 147: Similar to comment 13, it is not clear where this equation comes from, consider adding its derivation to the appendix. Also, define all the terms in the equation (e.g., ... where $\|\cdot\|_{L^2}$ is the L^2-norm...).

21. Line 153: How or why is the additional effort avoided in your approach?

22. Line 168: What is “resolves the unexpected structure”

23. Line 185: The second equal should be an approximate sign. Same with multiple places following. Eq 15, 16, ...

24. Line 201: Bold u?

25. Line 208: Why not an explicit? Does the solver have stability issues? Is this relevant?

26. Line 220: Define $\mathcal{B}_\rho(0)$.

27. Line 222: How is $\frac{|\Omega|}{n} = h$ calculated? Moreover, isn’t $\Omega = \mathbb{R}$?

28. Line 228: Is $L^1(\mathbb{R})$ defined?

29. Line 259: Define *iqr*. A non-statistical audience may not know what this is.

30. Line 271: Is the discretization of Ω the same as $\bar{\Omega}$? See line 180. Also, if this is the case, there are multiple instances throughout the paper where this appears.

31. Line 292: What do you mean by the variable Ω? Is Ω the spatial domain or something else?

32. Table 1: Why is the default number of spatial discretization intervals 1004 and not some value dependent on the dimension of Ω or the concentration and number of observed points?

33. Line 309: What is time forward?

34. Line 324: Why 20 and 10? Any justification?

35. Section 2.4: I think this section could use a major rework to make it more readable. Also, there are figures that relate to each of these functions. Why not relate the function to its corresponding figure? For example, "The function call *evol_plot* opens a plot showing the time evolution of the diffKDE (e.g., see Figure 2 for example output). An alternative would be to incorporate this paragraph into section 3.1 and use your simulation example to highlight some of your programs capabilities.

36. Figure 5: Are you able to include the true curve as a reference?

37. Figure 6: Can you report the MISE or AMISE for all of the curves so there is some numerical reference on how they perform? Also, what about a third column using 1000 points or so to highlight where diffKDE does *really* well. It would be nice to see how the other methods perform when diffKDE effectively perfectly captures the truth (e.g., are Botev, Gaussian, Epan still way off when diffKDE is nearly perfect).

38. Figure 9: I would either connect the dots with a line or have no line at all. I think the best fit line is odd here.

Technical Corrections

General: Consider using simpler, more precise, descriptions of the concepts. There are various cases, some that are included in the following list, where your message gets lost due to word choice. Below is a list of the some of the technical and grammatical issues, but not all. I recommend the authors verify the manuscript is void of technical and grammatical issues before re-submission.

1. Line 2: Comma after geoscience.

2. Line 5: No need for a comma, "but incomplete because of the..."

3. Line 7: This sentence feels odd. Consider rewording. Something like, "To make diffusion-based KDE accessible for general use, we designed and developed..."

4. Line 8: "We demonstrate our tool on simulated and real marine biogeochemical data individually, and compare our results against other popular KDE method". Also, be clear on if the simulated data is marine biogeochemical data or is not related to marine data.

5. Line 11: This sentence reads awkwardly. Consider breaking it into two pieces. "The convergence ... smaller error. This is most notable for ... "

6. Line 12: I don't understand the use of "exemplify".

7. Line 13: Is this related to the real-world example? If so, perhaps move it to earlier in the paragraph.

8. Line 15: Comma after geoscience.

9. Line 18: "Such necessity is not only..."

10. Line 19: "... applications such as social science and financial or ecological research."

11. Line 21: “... by some distance or divergence measure between...”
12. Lines 33-43: Some of the sentences in this paragraph read awkwardly. For example, the first three feel like they could be combined to read more clearly. Consider reworking this paragraph.
13. Line 45: No comma after “is possible”
14. Line 53: ”... choose between varying levels of smoothness by design.”
15. Line 55: ”... KDE with an accompanying Python package, *diffKDE* **cite or link package**.”
16. Line 59: Remove “, so called”
17. Line 61: Starting with ”Thus, ...” this sentence is confusion, consider rewording.
18. Line 72: You already defined PDFs, can remove ”probability density functions”.
19. Line 85: “integrated squared error (MISE) is ...”
20. Line 88: Define AMISE.
21. Line 117: “...Chaudhuri and Marron (2000) and its benefits were ...”
22. Line 119: What does “is“ refer to in “... and is extended ...”
23. Line 123: No comma after (diffKDE)
24. Line 128: “When regarded as a PDE, the Dirac δ-distribution puts all of the probability as the corresponding data point.”
25. Line 131: This sentence feels out of place, consider incorporating it into a paragraph so it is not a stand alone sentence.
26. Line 133: Obliterates does not feel like the correct word.
27. Line 144: Both sentences start with “in”.
28. Line 154: Unclear, consider rewording.
29. Line 171: What is Gaussian here?
30. Lines 178-179: Incorporate into the paragraph after it.
31. Line 199: “...division by $\boldsymbol{p}$ is applied column-wise.”
32. Line 221: Is fineness the right word? Maybe size? Or resolution?
33. Figure 1: “The function Φ_h depends on the ...”
34. Line 228: Is this supposed to be a paragraph? Odd indenting.
35. Line 240: Starting with “It is built...”, re-word this sentence.
36. Line 466: Chapter?
37. Line 474: Define the euphotic zone.