Statement of Revision

We thank the associate editor and reviewers for their useful comments and suggestion. We have made significant revisions to the manuscript to reflect the reviewer comments and concerns.

In the following, we provide an account of the changes that have been made. Once again, we will be responding to the comments both in their paragraph form and bulleted form, using the notation **(Rx:Py)** where **x** is the Reviewer number and **y** is paragraph referred to. As before, wherever a full sentence or more has been added in the manuscript, we have marked the changes in **blue**.. We would like to thank the reviewers for their time and effort in providing feedback for the improvement of the paper.

An overview of the major changes

Here we provide a summary of the most major revisions in response to reviewer comments. Further details will be provided in the sections dedicated to each individual reviewer.

1. We have moved the materials in the appendix to the main text.
2. We have improved the quality of the figures in the document.
3. We have overhauled the language of the paper for greater readability and tutorial value.
4. We have removed some material which was insufficiently developed in order to shorten the overall length of the paper.

Response to comments by Associate Editor

(AE:1) The paper has been re-reviewed by four reviewers. In general, the reviewers found the revised paper significantly improved, but still point out a number of issues. Specifically:

- Presentation style: the paper still contains several typos (Reviewer 5, Reviewer 9). Even the statement of revision has several clumsy sentences. Most importantly, some parts (in particular, the newly added part on Koopman operator theory) are very hard to follow (Reviewer 5, Reviewer 6, Reviewer 7, and Reviewer 9).
- Comparison with the state of the art: the comparison with the state of the art still requires improvement (Reviewer 6 and Reviewer 7).
- Tutorial contribution: the unusual length of this paper, along with the fact that some sections are difficult to follow, severely limit the tutorial contribution of this paper (Reviewer 7).

I concur with the reviewers' comments. The authors should prepare a revised version addressing *all* the concerns and suggestions. In particular, the authors should significantly strengthen the tutorial value of the paper, and carefully proofread the manuscript. Finally, note that CSM papers do not have appendices, so the location of the proofs needs to be changed.

Response: We thank the associate editor for the opportunity to submit a revision. We have prepared a revision that addresses each and every one of these concerns. Care has been taken to review the entire paper sentence-by-sentence to ensure complete readability and freedom from typos. We have diligently reviewed, clarified, and added nuance to the claims our paper makes about our contribution to the literature. We have shortened the paper by 15% while also restructuring multiple sections to better communicate the driving motivations and important outcomes of our work.

Response to comments by Referee 5. Reviewer ID 4949

(R1:P1) Overall, I wish to thank the authors for thoughtfully addressing the previous comments of the reviewers. My comments will primarily address the added content.

I think the added discussion of the connections between Koopman operator methods (both in theory and practice) and the proposed approach is very helpful. I think it could be structured slightly better: while the first paragraph discusses that connections will be presented, I think the results of Theorem 4 (either the explicit theorem, or a rough outline) could be presented earlier. I realize that the theorem relies on machinery developed in the preceding propositions— I'm not entirely sure how to restructure this section, but a better roadmap to the main results would be helpful for the reader.

Response: Thank you for your encouraging feedback. We have restructured this section in order to declare the main theoretical results more clearly at the beginning, while still maintaining the logical structure of the proofs.

(R1:P2) I think the discussion of scoring paths could be better motivated. In particular, providing more intuition for the measurement values (items 1-4) would be helpful for the reader.

(Response) We have chosen to remove the section on path planning for mobile agents. While we believe that this work has novel contributions to the problem at hand, we believe that to do it justice in comparing it to the work cited above would only lengthen a paper that the Associate Editor has mentioned is too long. We hope to publish this late-breaking work soon in another paper, but at this time we are convinced that this paper is strongest without it.

(R1:P3) Generally, the captions on figures need to be made more descriptive. For example, the caption on figure 20 is simply "more results" which isn't adequate for a reader.

(Response) The captions on the figures have been reviewed and updated per request.

(R1:P4) Small comments: There remain many typos. The authors must do a thorough proofread. Some that I noticed:

- Mezic (2013) in page 21, pg 2, missing reference
- Broken references in line 13 of page 24
- On page 25, Riemann is mispelled
- Figure 22, "mean" is written ",ean"

(Response) Thank you for pointing out these typos. A thorough proofread has been performed, and these errors have been corrected.

Response to comments by Referee 6. Reviewer ID 4945

(R6:P1) The authors tried to improve the quality of the paper, in particular, they added novel material, references, and the appendix. However, I think it is still not enough for publication. In particular, they should better revise the introduction and in general the writing style. The quality of images 3,7,14,16,18,19,20,22,S1 is quite poor and must be improved.

(Response): The introduction has been thoroughly revised to improve the tutorial value of the paper, along with the writing style of the rest of the paper. The aforementioned images have been updated for better quality.

(R6:P2) I also do not think that a link between GPs and dynamical systems (and then using dynamical systems for estimation) is a contribution since it has already been studied in [42-44] and also deeply discussed here (this reference is missing in the paper): Todescato et al., Efficient Spatio-Temporal Gaussian Regression via Kalman Filtering, ArXiv, 2017

(Response) We thank the reviewer for pointing out the missing reference, we have included this in the paper among the other references to methods which use a nonstationary covariance kernel. While this work shares with ours a similar ability to predict-and-correct using a Kalman filter, it lacks the guarantees on observability granted by our informed sensor-placing strategies. We have reviewed our introduction and contributions section carefully to ensure that we are clear stating, not overstating, the contributions of the paper.

We would like to be completely clear about the concern about papers 42-44. For reference, these papers are

42 Carron et al. Machine learning meets kalman filtering. In 2016 IEEE 55th Conference on Decision and Control (CDC)

43 Jouni Hartikainen and Simo Sarkka Kalman filtering and smoothing solutions to temporal gaussian process regression models. In 2010 IEEE International Workshop on Machine Learning for Signal Processing

44 Simo Sarkka, Arno Solin, and Jouni Hartikainen. Spatiotemporal learning via infinite dimensional bayesian filtering and smoothing: A look at gaussian process regression through kalman filtering. IEEE Signal Processing Magazine.

We are certainly aware of the seminal work in this area done by Sarkka's group and other groups in the area of bringing together Kalman filters with Gaussian Processes. This work has led to efficient ways of estimating GP parameters. We certainly agree that our paper is not the first one to establish the link, and neither have we made such claim. In fact, we have cited other such papers, including the very first time the link between Kalman filters and GPs was established, as we have noted, in the work in Kriged KFs.

To clarify our contributions over the work of Sarkka et al., and specifically 44, we bring the reviewer's attention to equations (14) in Sarkka et al., note that the kernel uses both X and t. The formulation in (15) in Sarkka et al. further establishes this by showing that both the linear evolution and the spatial prediction equations have kernels over (x, t). As such, the kernel dictionary in these models need to have kernels over both x and t increasing drastically the number of kernels. This is the first point of departure and our contribution, in the E-GP model we enforce that the kernel of the GP is only a function of x, all of the temporal evolution is restricted to the linear part of the model. This approach is designed specifically to model large scale spatiotemporal dynamical systems. As such, we do not assume the knowledge of the A matrix, and provide methods for estimating it, which is different than 42-44. In addition, we present results on approximation of the Koopman operator and provide results that show that the method works well in predicting complex spatiotemporal PDEs such as Navier-Stokes.

In summary, we agree that this is not the first paper to link GPs and dynamical systems, indeed we have cited many others. Our contributions however are in strengthening that link and establishing a theory of observability and controllability in feature spaces, as well as using it to design sensing schemes.

Response to comments by Referee 7. reviewer ID 4669

(R7:P1) This is a revised version of the previously submitted manuscript on the use of data driven methods such as Kernel observers for learning and control of spatially varying dynamical systems. In my previous review, I had highlighted several missing aspects such as inclusion of formal mathematical proofs of the claims, use of an algorithmic format to explain the proposed placement methods, etc. The authors have addressed some aspects satisfactorily, but in an effort to respond to other reviewer comments, they have added significant new text that can confuse a reader. Below are some comments on the revised version:

(Response) We are glad that the appendix satisfied your desire for formal mathematical proofs and algorithms. On the advice of the Associate Editor, we have moved the appendix materials directly to the relevant sections of the text.

(R7:P2) 1. The new section on Koopman operator methods introduces new concepts and in order to compare the kernel based method. However, this comparison and the new description is confusing for a non-practitioner and needs a careful writing. For instance,

  (i) in the proof of Theorem 4, why is $w(\cdot)$ suddenly an operator, whereas in the earlier sections, w_τ was a vector.

  (ii) the moving agent version of the problem needs further discussion – how does the motion model of the agent play a role? There are a couple of missing references in the new section. There are several works addressing this problem in oceanic environments (e.g., the works by Sukhatme et al, Leonard et al) that one would need to compare with.

 (iii) The moving agent problem now needs a fresh comparison with spectral methods for coverage (e.g., the works of Mezic et al).

(Response)

(i) In the proof of Theorem 4, we note that an operator defined by $f_\tau \mapsto w_\tau$, or $w(f_\tau) = w_\tau$, is an observable. The output of the operator has the same value as the vector. This becomes useful in defining the eigenfunctions. In order to clear up this confusion, we have used a different symbol P for the projection operator.

(ii) We have chosen to remove the section on path planning for mobile agents for reasons given in the response to (R1:P2).

(iii) Nevertheless, to answer the questions posed about the now-removed section: 1. We assume that the system is large-scale enough that the motion models of the agent(s) do not play an important role in determining trajectories. 2. The problem we pose is fundamentally different than the problem posed in other works on coverage, most of which seek uniform coverage or at least a static weighted coverage. Since we seek to accurately predict the spatio-temporal evolution the system, our coverage needs change continuously.

(R7:P3) 2. The proofs of the original results seem to be fairly clear. However a few questions come up – Regarding the sampling section, it is unclear as to what is the measure ν selected in the experiments. Also how useful is the theoretical result? What would be the measures under which p_ϵ would tend to zero?

(Response) Measure ν relates to the domain space of the function, for one dimensional domain it corresponds to length, for two dimensional space to area, and so on. Our synthetic dataset experiments considers one dimensional domain space, hence the measure ν would correspond to length that each kernel center occupies given an epsilon. The useful part of these results is Theorem 3 which dictates a definite relationship between observability and number of random samples which generalizes to any domain, whereas Theorem 2 is a step that leads to theorem 3. p_ϵ would tend to zero only if ϵ is very small.

(R7:P3) Algorithms 3 and 9 talk about the use of a 'standard kernel inference procedure'. It would be helpful to explain what such a procedure is.

(Response) The step in those algorithms has been changed to read 'Solve for $\hat{w}_\tau$ by using least squares on $y_\tau = K\hat{w}_\tau$'.

(R7:P4) 3. The numerics are very useful and would be appealing for a reader. Overall, I think this paper needs to be condensed from its current state as I think there is too much content in this version to be considered for a CSM publication.

(Response) We agree with the reviewer that this was a very long paper, and so we have taken the effort to condense it by 15% from the previous version.

Response to comments by Referee 9. Reviewer ID 4943

(R9:P1) I have read the revised version of this paper and the response letter from the authors. We thank the authors for the efforts to make modifications and improvements based on the reviews. In particular, the authors make comparisons between their methods and Koopman operator theory. Additionally, the authors also provide an appendix to include the missing proofs from the first submission. Overall, the paper has been improved. There is still a potential to further improve the paper from both of the clarity of the theoretical analysis and the presentation style. Here are some comments:

(Response) The authors appreciate the helpful comments from the anonymous reviewer!

(R9:P2) One note about the set of time instances (which appears in Prop 1, Thm 1, and etc.). Could the authors give some explanations about the difference between this and the set of sampling locations? For example, we do not want the cardinality of the set of sampling locations (i.e., N) to be large, compared with M. Would the condition that the cardinality of the set of time instances (i.e., L) is greater than M be an issue?

(Response) The cardinality of the set of time instances is independent from that of the sampling locations. In the proof of Proposition 1, note that the interaction between the sampling locations and the time instances is relevant mainly to show observability. Therefore, $N \times L$ has to be greater than equal to M: since L is assumed to be at least M, the

number of sampling locations can remain small, as long as we have a shaded observation matrix.

(R9:P3) When proposing the random sensor placement scheme, it would be better to give some more details about how the random scheme compares with/differ from the measurement map approach, besides Figure 8. For example, the computational complexities of these two methods could be compared further and explain which method is suitable under which scenario.

(Response) The random scheme inherently requires no computational effort, but comes at the cost of requiring a large number of samples; on the other hand, the measurement map approach is computationally expensive and infeasible for budgeted kernel centers.

(R9:P4) One note about the value $\frac{\varsigma}{p_\epsilon}$. How does this value compare with the dimension of the matrix $\hat{A}$ in Eq. (3)?

(Response) For a kernel observers model generated by a set of time series data, we expect that the value $\frac{\varsigma}{p_\epsilon}$ should be independent of the number of centers chosen M (the dimension of the matrix $\hat{A}$ is $M \times M$). This is because the number of Jordan blocks ς in $\hat{A}$ should reflect the number of invariant subspaces represented in the dynamics. p_ϵ is pretty much entirely independent of M.

(R9:P5) A general suggestion about the section Spectral Analysis of Evolving Gaussian Process Model, and Resulting Algorithms and Applications. This section could be potentially better structured. One way could be to split it into several subsections, since the comparison with the Koopman method is also in this section.

(Response) We have restructured this section to better emphasize the main results, as well as remove less important results.

(R9:P6) There seem to be quite a few typos and unclear sentences throughout the paper, some of which could prevent the readers/reviewers from better understanding the content. In particular, the newly added sections (to the revised paper) need a careful proofread. A partial list of them is provided here for the authors reference:

1. In general, many equations and sentences in the text are missing a comma or period at the end.

2. The last sentence in the second paragraph on page 23 needs to be fixed.

3. When referring to a previous section or a defined algorithm, etc., later in the text, it would be better to add a number reference since the paper has many pages. For example, the k-invariant subspaces clustering algorithm mentioned in the first sentence in the third paragraph on page 23.

4. The references on page 24 need to be fixed. The sentence in line 20 of the same page needs to be restructured. Line 24 on the same page: quickest should be fastest.

5. Line 5 on page 25, duplicated to. The sentence in line 15 on this page should be restructured. Line 17, likeliest.

6. The Generating High-Scoring Random Paths subsection could be better explained with more details.

7. In Appendix, Remark 3 and the paragraph before it also appear in the main body of the paper. In the last subsection of Appendix containing the proofs of the Koopman method, the proof for Prop 4 or Prop 5 seems to be missing. The proof for Theorem 4 could be made clearer by splitting the proof according to the three results in the statement of Theorem 4.

(Response)

- Reviewed and corrected.
- Reviewed and corrected.
- This section actually first introduces said algorithm. We have made revisions to help avoid such confusion.
- The section on path planning for mobile agents has been removed, for reasons given in the response to (R1:P2).
- See above.
- See above.
- The Appendix has been merged with the main text in accordance with comments from the Associate Editor. Proposition 4, as has now been made more clear, follows directly from the definition of observables, so no proof

is provided. The proof for Theorem 4 has been structured as suggested.

Conclusion

Again, we thank the anonymous reviewers for their comments. We hope we addressed all concerns and improved the overall readability of the paper. We are happy to provide further clarification or revisions as requested.