Statement of Revision

We thank the associate editor and reviewers for their useful comments and suggestion. We have significantly revised the manuscript to reflect the reviewer comments and concerns.

In the following, we provide an account of the changes that have been made. The nature of the review for this paper is such that the comments were provided both in paragraph form and in bulleted form. When addressing the paragraph form, we have identify the comment using the notation **(Rx:Py)** where **x** is the Reviewer number and **y** is paragraph referred to. When addressing the bullet form, we utilize the form **(Rx:z)** where **x** identifies the Reviewer and **z** identifies the bulleted comment. Each bulleted comment included in the listing is followed by a description of modifications to the paper addressing the comment. Furthermore, changes in the paper have been marked with a color coding. We hope that this method is sufficiently clear. We would like to thank the reviewers for their time and effort at providing feedback in order to improve the exposition of the paper.

An overview of the major changes

In response to reviewer comments, we have undertaken some major revisions for this paper. Here we summarize some of those major revisions, further details are provided in response to individual reviewers:

1. We have included proofs of all main theorems in the appendix.
2. We have included new sections on generating path for a single mobile agents to observe a dynamically varying phenomena.
3. We have added theoretical insights into links with Koopman operator theory
4. We have significantly re-focused the control problems in agriculture sidebar.
5. We have included several references in the related works section to better outline the contributions of this work in the context of the vast literature out there on Gaussian Processes, spatiotemporal modeling, and distributed sensing.

Response to comments by Associate Editor

In this paper the authors provide an overview of methods and algorithms to build data-driven models that may be used for estimation and prediction of spatio-temporal processes.

Four reviews were obtained for this paper. In general, the reviewers agree that the topic of this paper is of interest to the control community, and that the proposed approach appears sound. However, the reviewers point out a number of serious concerns, in particular:

Comparison with the state of the art: the comparison with the state of the art appears insufficient (see Reviewers 5, 6, and 7). This makes the contribution of this paper unclear (see Reviewer 6).

Presentation style and rigor: the presentation style is quite poor (see Reviewers 5, 6, 7, and 9). In particular, the paper contains several typos and the presentation of the results is rather dense. Most importantly, the authors refer to an Appendix for the presentation of the proofs of the theoretical results; however, such an Appendix is missing. Collectively, these shortcomings make this paper very difficult to read and evaluate – see, in particular, Reviewer 7. Furthermore, note that CSM papers do not have appendices, so the location of these proofs will have to be changed.

Tutorial contribution: the insufficient comparison with the state of the art, the poor presentation style, and the lack of proofs for the theoretical results make the tutorial value of this paper quite weak. In other words, it would be very difficult for a practitioner to use this paper as a guide for developing estimation and prediction algorithms for spatio-temporal processes (when should one use the methods presented in this paper? are there implementation guidelines? etc.). The authors do reference a software library - in my opinion, this paper should more explicitly reference such a library in order to better explain the practical and implementation aspects of the proposed methodology.

I concur with the reviewers comments and I recommend that the authors prepare a significantly revised version addressing all the concerns and suggestions, with a key focus on strengthening the tutorial value of the paper. In addition, the authors should address the following comments:

(AE:1) We thank the associate editor for the opportunity to submit a revision. We have prepared a significant revision that includes the complete proofs of all of the results presented in this paper.

The sidebar "Key control problems in agriculture," while interesting, appears excessively long. The authors should shorten it, and make its relevance to the topic of this paper clearer.

(AE:2) We have reduced the key control problems sidebar. We have significantly reduced the length of the sidebar by reducing the general discussion about agriculture. We have focused instead the sidebar on the challenges in spatiotemporal estimation for advancing agricultural robotics.

The quality of many of the figures is quite low (see, for example, Figure 1, Figure 8, and Figure S8). The authors should make all figures clearly readable.

(AE:3) XX Josh, Harshal can you handle this? XX

Response to comments by Referee 1. reviewer ID 4665

(R1:P1) This paper describes an approach, developed by the authors over the previous few years, to build data-driven models that may be used for estimation and prediction within spatiotemporal monitoring. The model is based on predicting a process that evolves in an RKHS, where finite dimensional measurements are available. Given these measurements and a choice of kernel, approximation methods for the feature space may be developed (via e.g. a dictionary of atoms, random fourier features, etc). Then, given the measurements, a linear dynamical system may be constructed in the feature space. This finite dimensional feature evolution may be analyzed from the point of view of linaer systems theory, and theoretical results on observability and estimation are stated.

Response: The authors thank the anonymous reviewer for his succinct and accurate evaluation of this paper.

(R1:P2) While I believe this paper is coherent and reasonably strong, there are a few ways in which it could be improved.

First, the technique of finding an alternate space in which the state evolves linearly is not novel (nor is it claimed to be). The paper would benefit from an extended discussion and comparison to alternative methods. In particular there has recently been considerable interest in methods inspired by Koopman analysis, which the authors mention only extrememly briefly, in passing. Since the authors present experiments on fluid flow problems, they should present comparisons to dynamic mode decomposition, which has been a popular technique and has many similarities to the method presented in this work. See, e.g., Schmid, Peter J. "Dynamic mode decomposition of numerical and experimental data." Journal of fluid mechanics 656 (2010): 5-28. and the many articles that cite the above. Moreover, there is a broad literature extending dynamic mode decomposition and other Koopman-inspired techniques. In particular, one may refer to the work of Steve Brunton and Nathan Kutz, such as Brunton, Steven L., Joshua L. Proctor, and J. Nathan Kutz. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems." Proceedings of the National Academy of Sciences 113.15 (2016): 3932-3937.

While these works do not explicitly discuss the estimation problem, comparisons between the chosen model and those discussed above would greatly improve the paper and allow readers to better appreciate and situate the proposed methods within the literature.

(Response) We have included an extensive discussion of this method in view of Koopman operator theory. We have also included a comparison of the modes discovered by this method to those obtained via dynamic mode decomposition. We have also included recent research into how the modes can be used to generate best paths for moving agents seeking to infer the current state of the system.

(R1:P3) In addition the above, I was confused by the presentation of the theoretical results. I could not find the appendix in which proof of the theoretical results was presented, and so have not been able to review them.

(Response)

(R1:P4) The evolving GP formulation, in which dynamics are shared across dynamical systems should not behave similarly, was confusing. In particular, the definition of similar fluid systems seems ad hoc, and this section must be better motivated and explained in general.

(Response)

(R1:P5) Minor comments XX when you have addressed the comment, remove the red lining XX ———————

1. page 1 line 10, ”spatiotemporally” misspelled
2. page 4 line 24 ”interpretability” misspelled
3. text very small in figure 8, figure 10, figure S8. Generally, the text size of the figures needs to be fixed throughout. XX Josh, Harshals XX

(Response) Thank you for pointing out these typos; they have been fixed.

Response to comments by Referee 6. reviewer ID 4667

(R6:P1) In this paper, the authors study how to perform state estimation of time-varying spatio-temporal processes, and how to design sensors placement. The topic is of interest for the community, however, the paper needs significant changes and work, both on the scientific side and writing style.

(Response) We thank the reviewer for their interest. We have significantly revised the paper to address the concerns raised by the reviewers.

(R6:P2) The authors claim that one of the two main contributions is

”we demonstrate that spatiotemporal functional evolution can be modeled using stationary kernels with a linear dynamical systems layer on their mixing weights”,

however, to me, this is not a novelty and there is a huge literature about the estimation of time-varying spatio-temporal processes and I would recommend the authors to compare their work with the following papers, highlighting the differences

1. J. Hartikainen and S. Sarkka. Kalman filtering and smoothing solutions to temporal Gaussian process regression models. In Machine Learning for Signal Processing (MLSP), 2010 IEEE International Workshop on, pages 379384. IEEE, 2010.
2. J. Hartikainen, J. Riihimaki, and S. Sarkka. Sparse spatio-temporal gaussian processes with general likelihoods. In Artificial Neural Networks and Machine LearningICANN 2011, pages 193200. Springer, 2011.
3. A. Carron, M. Todescato, R. Carli, L. Schenato, and G. Pillonetto. Machine learning meets Kalman filtering. In 55th Conference on Decision and Control. IEEE, December 2016.
4. J. Hartikainen. Sequential Inference for Latent Temporal Gaussian Process Models. PhD thesis, Aalto University, 2013.
5. S. Sarkka and R. Piche. On convergence and accuracy of statespace approximations of squared exponential covariance functions. In Machine Learning for Signal Processing (MLSP), 2014 IEEE International Workshop on, pages 16. IEEE, 2014.
6. S. Sarkka, A. Solin, and J. Hartikainen. Spatiotemporal learning via infinite-dimensional bayesian filtering and smoothing: A look at Gaussian process regression through kalman filtering. Signal Processing Magazine, IEEE, 30(4):5161, 2013.
7. TT Ho et.al., Multiresolution stochastic models for the efficient solution of large-scale space time estimation problems, Proc. IEEE ICASSP, 1996

8. Noh J, Solo V, Testing for Space-time Separability in functional MRI. Proc IEEE Int Symposium on Biomedical Imaging, pp 412-415,Washington DC, USA, April 2007.
9. J. Noh and V. Solo, Space-Time Separability in FMRI: Asymptotic Power Analysis and Cramer- Rao Lower Bounds, IEEE Trans. Sig. Proc., (2013), 61(1), pp 148-153.
10. JR Stroud et al, Dynamic models for spatio-temporal data, JRSSB,63, pp673-689, 2001.
11. RN Miller, Toward the application of the Kalman filter to regional open ocean modelling, Jl Phys Oceanog., 16, 72-86, 1986.
12. F Lindgren and Havard Rue and J Lindstrom, An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach, Jl Royal Stat. Soc. B, 2011.

(Response): The reviewer is indeed correct, there is a very large body of work on modeling spatiotemporally varying dynamic systems in the literature. Our work is inspired from this literature, and our main contribution is in realizing that the Hilbert spaces produced by Kernel methods can embed dynamical systems, and then utilizing those dynamical systems for doing estimation and control. The prior literature spans from that of statistical modeling to modern machine learning methods such as LSTMs. In particular, there is also a very large body of work on kernel methods, and Gaussian processes to modeling spatiotemporally varying functions. A large number of those papers have been cited in the related work. We have now made sure that all of the papers suggested by reviewer have also been cited. With respect to those papers, and many similar ones, what is new with E-GP is its ability to learn end-to-end complex spatiotemporal dynamic patterns such as solutions to Navier Stokes equations, and its ability to approximate the eigenmodes of Koopman decomposition (please see response to reviewer 1).

We have revised the paper to address the reviewer concern in the following manner:

We have added the below statement at the beginning of related work section:

There is a very large amount of literature on Gaussian Processes and spatiotemporal modeling, a complete survey of this literature is beyond the scope of this paper. Since our contributions are in the area of creating a feedback based observer in the feature spaces of GP models, we discuss here related work in three related areas: Spatiotemporal modeling with GPs, Kalman filtering and GP connection, and sensor placement for inference in spatiotemporal domains.

We then proceed to discuss the various modeling methods mentioned by the reviewer:

Other approaches that utilize hierarchy or evolution of kernels have also been used for modeling spatiotemporal functions [2, 7, 4].

A more clever approach is to use state-space representations of time-varying GPs [12, 2]. In this view point, each GP instance is viewed as a snapshot of an evolving set of weights. We follow in a similar vein here, with added emphasis on understanding the mathematical structures leading to observability and controllability.

This approach has been utilized previously in MRI imaging [10, 11].

Kalman filtering in the context of Gaussian processes and Kernel models has also been quite widely studied [1, 3, 13, 14, 9]. There is clearly a direct link between the Bayesian approach to inference taken in GPs and its natural extension to Kalman Filters. Our contributions here are in creating explicit connections between feedback observers and inference by deriving conditions of observability in the kernel space. This leads to explicit conditions on the number of sensors required and where to place them.

specific comparisons, discuss the work on reservoir computing, cite Brockett

(R6:P2) Also, the second contribution is not clear to me, what are the differences with the previous work [19], [20],[21]?

(Response) This paper represents a single unifying body of work bringing together out previous work and linking it together. In addition, there are significant new contributions introduced in this paper, including results relating to random sensor placement, connection with Koopman operators, and sensor placement with mobile sensors. To clarify the relationship with our prior work, we have added a new section titled: Outline and relationship with author's prior work. In that section, the following statement is added:

Elements of the work presented in this paper first appeared in Neural Information Processing Systems (NIPS 2016) ([5, 16]), IEEE CDC 2015 conference [6], the Conference on Robot Learning (CoRL 2017) [15] and IEEE ACC 2018 conference [8]. This paper presents a comprehensive set of results and fills in the missing links in a single encompassing publication, and introduces new results on observability in the presence of random sensor placement, connection to Koopman operators, and generating paths for single observing agents. As such, we have focused in this article mostly on the fundamental theory and practical algorithms for modeling, estimation, and control, while the excruciating details of how to optimally implement the presented algorithms are omitted. Instead an open-source code-base is made available in MATLAB at `http://daslab.illinois.edu/software.html` or at `https://github.com/hkingravi/FunctionObservers` and in Python on GitHub at `https://github.com/hkingravi/funcobspy`.

(R6:P2)Additionally, I think that the paper has not been carefully proofread since it is plenty of mistakes, repetitions, and sentence that are not appropriate for a journal publication. For example:

1. The comma after the last name in the authors' list
2. first page: spatiotermpoarlly -¿ spatiotemporally
3. first page: this very challenging problem -¿ this challenging problem
4. first page: There are many parallel (I would remove parallel) examples in other fields -¿ can you provide references?
5. first page: computational packages?
6. second page: has traditionally been in the province?
7. second page: very high amount of variability -¿ high variability
8. second page, line 13-24: can the literature review be structured?
9. second page: (for the matter) -¿ remove
10. second page, line 32: can you provide a reference about deep learning models?
11. fourth page, line 5: machine learning machine learning
12. fourth page, line 5: fused?
13. fourth page, line 25: This is mostly because -¿ This is one of the reasons
14. fourth page, line 27: I do not think this is one of the major challenges
15. fifth page, line 2-6: here the text is too informal
16. page 7, line 10: time-varying distribution?
17. page 7, line 17: taxing?
18. page 7, line 17: realizable? -¿ feasible

The list is however way longer. I recommend the authors to clarify their contributions, extend the literature review, and carefully proofread the paper.

(Response) Thank you for these comments. We have proofread the paper more carefully, and have addressed these issues.

Response to comments by Referee 7. reviewer ID 4669

(R7:P1) My main concern is that there are no proofs of the formal statements. The authors do say that the proofs are included in the appendix, but I could not find any appendix, nor any pointer to any online technical report for them. Therefore, I cannot recommend this paper for publication until I get a chance to review the proofs.

(Response) Appendix is now included

(R7:P2) For example, the notion of shadedness seems to too strong, i.e., it appears that the sensor matrix needs to ensure that it "sees" all states directly! On the other hand, the classical notion of observability leverages the dynamics matrix A to provide a crisp, necessary and sufficient rank condition.

(Response) The definition of shadedness is a simple way of defining coverage of the basis with respect to the sensors in a geometric way with respect the kernel function used. At first sight, it may seem to be too strong; however, if you look at the statement of Proposition 2, it becomes clear that even shadedness is not enough if there is even a single repeated eigenvalue (see our response to (R7:P3)). In that case, either a measurement map needs to be constructed (Proposition 3) or an ℓ-shaded matrix is required (Theorem 1).

(R7:P3) Proposition 2 is only a sufficient condition. I think the discussion just prior to Proposition 2 will benefit with the example of a K which is shaded and yet the system is not observable.

(Response) Please see the proof of Proposition 2, which gives such an example for the Gaussian kernel.

(R7:P4) The second on random sampling is very vaguely written. There is no clear "algorithm" that summarizes the method. Further, how is the measure ν in equation 10 defined? What if the $p_\epsilon = 0$? Can that happen in the present set-up? Theorem 2 talks about an expected number of randomly placed sensors – what is this expectation over? There is a δ in Theorem 3, but that never appears anywhere in the result.

(Response) Algorithm 2 in the appendix summarizes the method. For two-dimensional domain, the measure ν would correspond to area, and for other high dimensional domains it would correspond to volume. Expectation is over the number of sensors (or samples). If the probability p_ϵ is 0, the solution would degenerate to infinite number of samples. Definition of δ is provided within Theorem 3's proof given in the appendix.

(R7:P5) Further, if one plugs in the value of N equal to the lower bound in Theorem 3, then the probability of unobservability appears to be a number that can be larger than 1 and therefore, the utility of this result appears to be unclear!

(Response) Thanks for pointing out this flaw. The condition on N in theorem 3 is now changed to $N > 2\varsigma/p_\epsilon$, with the proof remaining unaffected.

Response to comments by Referee 9. reviewer ID 4673

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.

References

[1] Andrea Carron, Marco Todescato, Ruggero Carli, Luca Schenato, and Gianluigi Pillonetto. Machine learning meets kalman filtering. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 4594–4599. IEEE, 2016.

[2] Jouni Hartikainen et al. Sequential inference for latent temporal gaussian process models. 2013.

[3] Jouni Hartikainen and Simo Särkkä. Kalman filtering and smoothing solutions to temporal gaussian process regression models. In *2010 IEEE International Workshop on Machine Learning for Signal Processing*, pages 379–384. IEEE, 2010.

[4] Terrence T Ho, Paul W Fieguth, and Alan S Willsky. Multiresolution stochastic models for the efficient solution of large-scale space-time estimation problems. In *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, volume 6, pages 3097–3100. IEEE, 1996.

[5] Hassan Kingravi, Harshal Maske, and Girish Chowdhary. Kernel observers: Systems theoretic modeling and inference of spatiotemporally varying processes. In *Advances in Neural Information Processing Systems (NIPS)*, Barcelona, Spain, 2016.

[6] Hassan A. Kingravi, Harshal Maske, and Girish Chowdhary. Kernel controllers: A systems-theoretic approach for data-driven modeling and control of spatiotemporally evolving processes. *arXiv*, abs/1508.02086, 2015.

[7] Finn Lindgren, Haavard Rue, and Johan Lindstrom. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

[8] Harshal Maske, Hassan Kingravi, and Girish Chowdhary. Sensor selection via observability analysis in feature space. In *American Control Conference*, 2018. accepted.

[9] Robert N Miller. Toward the application of the kalman filter to regional open ocean modeling. *Journal of Physical Oceanography*, 16(1):72–86, 1986.

[10] Joonki Noh and Victor Solo. Testing for space-time separability in functional mri. In *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 412–415. IEEE, 2007.

[11] Joonki Noh and Victor Solo. Space-time separability in fmri: Asymptotic power analysis and cramér-rao lower bounds. *IEEE Transactions on Signal Processing*, 61(1):148–153, 2012.

[12] Simo Särkkä and Robert Piché. On convergence and accuracy of state-space approximations of squared exponential covariance functions. In *2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE, 2014.

[13] 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*, 30(4):51–61, 2013.

[14] Jonathan R Stroud, Peter Muller, and Bruno Sanso. Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4):673–689, 2001.

[15] Joshua Whitman and Girish Chowdhary. Learning dynamics across similar spatiotemporally-evolving physical systems. In *Conference on Robot Learning*, pages 472–481, 2017.

[16] Joshua Whitman, Harshal Maske, Girish Chowdhary, Hassan Kingravi, Chen Lu, and Balaji Jayaraman. Modeling nonlinear dynamical fluid flows with evolving gaussian process models. In *NIPS workshop on Machine Learning in the Wild*, Barcelona, Spain, 2016. NIPS.