We are appreciative of the considerable time and thought that reviewer #1 took with the manuscript. The manuscript is greatly improved as the result of their suggestions, especially with regard to calculation of $P_{\xi 1}$. Below we outline our responses to each of the comments.

Our responses to reviewer's primary concerns:

<u>Concern 1:</u> Reviewer #1 points out that there are many ways in which kinematics might change without any influence on the summary statistics we used to measure complexity ($\xi_{95\%}$ and $P_{\xi 1}$), and we agree. This arises because we measured just one component of complexity, a component that we call *dimensional complexity*. The reviewer notes that "higher POD modes may be operating at different frequencies." Indeed, changes in frequency would be missed by $\xi_{95\%}$ and $P_{\xi 1}$, as would several other on-linear changes in the motions of markers or joint angles.

Our definition of dimensional complexity is quite specific, but it clearly needs to be improved, since this led Reviewer #1 to question the appropriateness of our metrics ($\xi_{95\%}$ and $P_{\xi 1}$) for measuring it, when in fact our attempt was to phrase the definition to fit exactly the information that these metrics reveal. We have changed our definition slightly, and addressed the fact that our indices are not sensitive to all kinds of changes in complexity (section 1.1, second paragraph).

Reviewer #1 offers a different possible index of dimensional complexity, for example, measuring how many joint angles "operate independently to some specified level of tolerance." This is exactly how we came to delineate our three functional groups, by setting a tolerance of 0.7 (figure 6).

<u>Concern 2 (part 1):</u> The reviewer suggested that $P_{\xi 1}$ decreased as the number of markers used decreased, as the result of an "artificial 1/n dependence" introduced by the way we calculated that metric. The decrease in $P_{\xi 1}$ as the number of kinematic markers increased was therefore not truly representative of differences in complexity among the marker sets.

After careful analysis, we have come to agree with the reviewer, and have adjusted our analyses and the manuscript to follow their suggestion that $P_{\xi 1}$ be calculated simply as the normalized first eigenvalue that results from the POD analysis, and to *not* include the number of degrees of freedom in its calculation. Figure 4 now has a very different appearance as the result of the new formula for calculating $P_{\xi 1}$.
As the result of this change, the decreases in $P_{\xi 1}$ values that accompany increased numbers of markers are more subtle than they had been when the previous metric was used. In the current version of our paper, the distribution of $P_{\xi 1}$-values is overlapping for most numbers of wing markers, but the median values decrease systematically as the number of markers is increased. We feel that this change to the manuscript is a significant improvement, and we gratefully acknowledge the insights of the reviewer on the topic.

The reviewer points out that in figure 3, the $\xi_{95\%}$ values are similar for the marker matrix (46 input dimensions) and the joint angle matrix (20 input dimensions). The reviewer states that this is to be expected, since they are measurements of the same system, so complexity should come out the same no matter how data are treated before

POD. However, this is not the case. As we point out in our discussion, different ways of treating the data before POD is performed can alter the $\xi_{95\%}$ values. Below, we show the $\xi_{95\%}$ values that result, depending on whether standard deviations are standardized for each dimension or not standardized, whether linear or spherical coordinates are used, whether a local or global frame of reference is used, or whether joint angles or the marker positions are used.

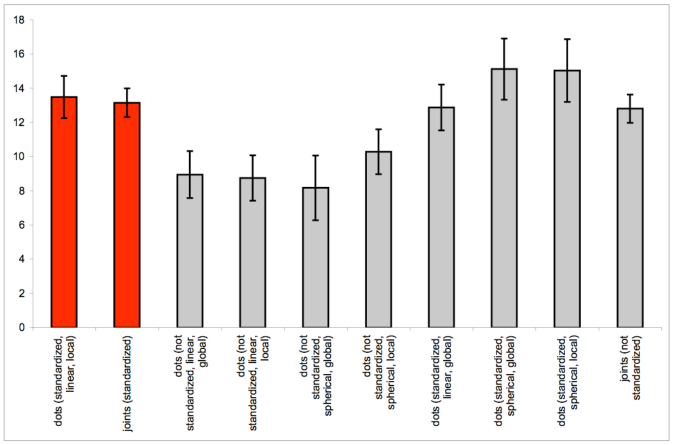

Figure 1: mean $\xi_{95\%}$ values (n = 9 trials) that result from POD differ, depending how the data were prepared before singular value decomposition was performed. The two bars in red are those represented by figure 3 in the paper. These two values happen to be similar to one another, but it is important to note that values can range markedly, from as low as 8.16 ± 1.9 to 15.1 ± 1.7.

<u>Concern 2 (part 2).</u> Reviewer #1 states that a major problem with the paper is that the relationship between $\xi_{95\%}$ and $P_{\xi 1}$ is not explored. We have added a paragraph to the discussion to address this issue (end of section 4.1). Below, we provide a more detailed response than we feel is necessary for the paper. If the editor feels that some portion of the discussion below should be included in the paper, we are happy to do so, but the paragraphs below are written in response to questions that the reviewer has brought up.

POD is, by definition, the optimal orthogonal linear transform to a new coordinate system such the first coordinate (called mode 1) minimizes the sum of squared errors when the original data (with mean subtracted and variance standardized) are projected onto that coordinate. Subsequent modes are computed analogously, with the restriction that the modes be orthogonal. Our goal is to describe how information gets distributed among the modes, since we contend this information is useful for the purposes outlined in the paper. We have used two different metrics to describe this. Based on the way in which $\xi_{95\%}$ and $P_{\xi 1}$ are calculated, we expect them to be negatively correlated (which they are in the case of our experimental data ($r^2 = 0.45$), as shown below), but one cannot be predicted entirely from the other.

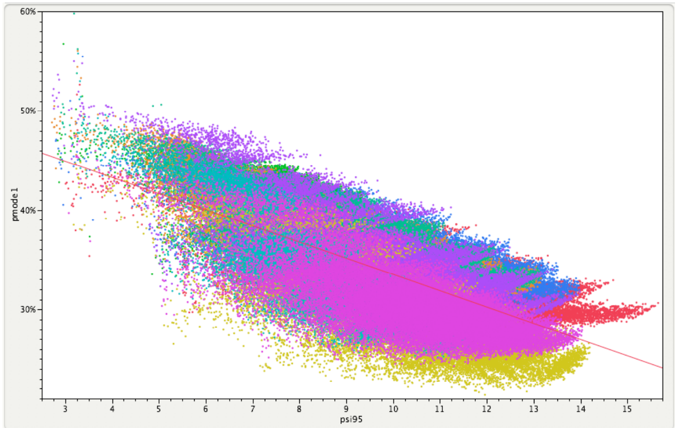

Figure 2: In response to reviewer #1's comments, we tested our prediction that the two measures we use of dimensional complexity should be inversely correlated. Using the 32,767 possible permutations of 1 to 15 markers for each of the nine trials to generate 294, 903 $\xi_{95\%}$ and $P_{\xi 1}$ values, we found that they were indeed negatively correlated (overall linear $r^2 = 0.45$). Data from each of the nine trials is coded in a different colour, and the overall best fit line is shown in red.

Because they are negatively correlated, as expected, we find no reason to question the utility of $\xi_{95\%}$ and $P_{\xi 1}$ for the purposes of this paper. In agreement with the reviewer's comments, we have changed the manuscript to be more clear about what those metrics measure, and about the fact that many changes in complexity might be missed by our analyses. In the earlier version of the paper we had used "complexity" and "dimensional complexity" interchangeably, and we have now restricted our language to "dimensional complexity" every time, to avoid confusion.

Our responses to minor reviewer concerns:

1. We have added several citations to the manuscript, including Cappellini et al., 2006; Chau, 2001; Forner-Cordero et al., 2007; Ivanenko et al., 2008; Mason et al., 2001; Todorov et al., 2004; and Tripp et al., 2006. Our use of the technique is somewhat novel, so we cannot cite papers that use the technique in the same way we do. The main purpose of citing these articles is to show that POD is used widely to study animal movement. If the reviewer has specific papers in mind that should be added, we are happy to add these, but the manuscript now includes citations to 16 papers that use POD (though the technique is otherwise named in many of them).

2. We have now included a brief analysis of position error associated with our methods.

3. As pointed out by the reviewer, our transformation to a body-referenced coordinate system makes us unable to measure the influence of yaw changes, and the lack of markers on the left side of the body makes it impossible to measure body roll with this data set.

We know that yaw changes do not influence the major conclusions drawn from this study, from supplementary analyses of the same data that we performed using a global reference coordinate system. Using global coordinates, we still find that complexity ($\xi_{95\%}$) does not change with speed, and that the shoulder and hip move independently of the other markers. We do not know what effect roll might have. This is one element of kinematic complexity that our methods are not able to address.

Furthermore, while we did not examine differences between the two sides of the body in this study, the methods we have outlined could be used to, for example, quantify the kinematic symmetry of the two sides of the body.

4. The algorithm is a least-squares fit. We have adjusted the text to make this clear.

5. We did, indeed, use the normalized eigenvalues. The text has been updated to make this clear.

6. We assume the threshold to which the reviewer refers is the Butterworth cutoff frequency. We repeated these analyses using cutoff frequencies of 100 Hz and 50 Hz and got the same basic results.

7. We have included the mathematical formula used to construct the dendrogram.

8. The reviewer asks whether the changes in methodology discussed in section 4.1 affect the preferred marker locations. As stated in the text, "adjusting these user inputs on our data resulted in similar trends from POD – such as independence of complexity and speed, and anatomical marker motions at the shoulder and hip that were independent of other marker motions."

9. We have adjusted the sentence in question for clarity.

10. We have adjusted the section in question to address the reviewer's comments, and cited the work that the reviewer mentions.

11. We have added citations to the section in question.

12. The reviewer suggests examination of the phase and amplitude information present among joint angles. In particular, "the membranes between digits III, IV, and V may not remain flat" if the amplitudes of the corresponding joint angles are not similar.

The phase relationships of the joint angle groups to one another are discernable from Figure 7. The amplitude information, however, has been removed. Below, we reproduce figure 7, but without standardizing variances. Note that the amplitudes of joint angles in each group are of similar magnitude. The joint angles that should be of similar magnitude for the membrane to remain flat near the wrist are joint angles 8,9, and 10, which are of similar magnitude (group 2). Therefore, it is likely that the membrane between them remains somewhat flat during flight (we did not track markers in this study, but from the videos the membrane appears flat, as we state in the paper).

The figure below differs from the one in the paper. For the purposes of the paper, we have decided to leave figure 7 as it is (with variances standardized), because (1) standardized angles were used to define the groups, (2) differences in amplitude among joint angles might distract the readers from the point of the figure, that the phase relationships are similar, and (3), because scaling the fourth ("other joint angles") subplot to make room for joint angle 18 makes it difficult to see how the other joint angles change through time.

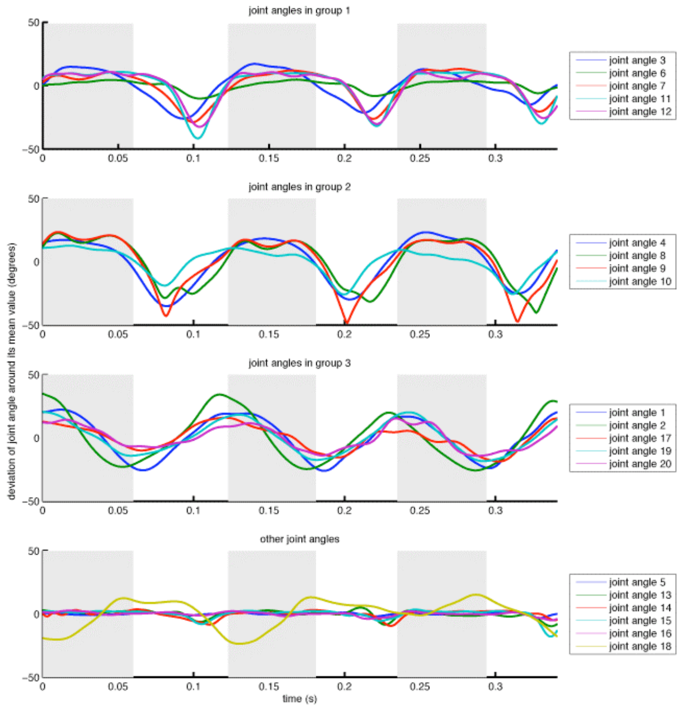

Figure 3: Joint angles deviations from their average values over the course of a trial. Note that this figure is similar to Figure 7 in the manuscript, but that the joint angles here have not been standardized to have the same standard deviations.

Daniel K. Riskin
Department of Ecology and Evolutionary Biology
Brown University
Box 6-B204, 171 Meeting St.
Providence RI 02912, U.S.A.