Reviewer #1 Evaluations:   
  
Q1: It is not clear whether the datasets can belong to regions in phase space which have different low-dimensional manifold. The datasets resented in the work has same low-dimensional manifold. Instead, if three datasets are created using short MD trajectories, one belonging to basin A of Fig 6a, another belonging to the basin B, while third in between A and B. In such a case, how would the method behave?

R1: The fundamental question is, to what extent the density of sampling different regions affects the coordinates we obtain. The scenario suggested by the reviewer (sample one basin OR another basin) is the extreme case; usually, we would sample both basins "a little", but not enough to establish the corresponding equilibrium densities.

We added a detailed analysis in Section II-D to describe how the embedding relates to the propagation of the intrinsic variable. This analysis shows that the NIV embedding depends on the potential field. Thus, the constructed embedding is consistent as long as the subsets of observations are sampled from the same common region.

In order to merge observations from different regions, we need to account for the potential field. In principal, we can decouple diffusion dynamics from geometry by normalizing the diffusion kernel with sample density (this is now described in detail in the paper).

Since local coordinates are based on the noise term in the governing SDE, sets of measurements from different regions will essentially detect the same system of coordinates (the same variable ``axes"), yet the different sampling densities arising from the drift term will result in different scalings of these axes. Applying the alternative kernel normalization in the NIV context will eliminate the sampling density effects. An alternative approach would be to estimate directly the drift term for each intrinsic variable by finding the optimal transformation that would align the common regions.

We are conscious of the importance of this research direction, but we respectfully consider it beyond the contributions of this paper; it is the subject of current work in our group.

Q2: Overall, I find the authors present the results without providing enough insights to allow the reader to fully appreciate the method. For instance, it is not clear how would the method behave if a smaller dataset was created with the SSA simulation. How can one determine whether the data set is rich enough (as the authors have mentioned).

R2: To address this critical point, we have performed a few more computational experiments and included a new figure in the paper (Figure 5, also attached here).

To determine whether a sufficient amount of data is available to obtain an accurate embedding, we compute the NIV embeddings for different numbers of data points (arising from different numbers of trajectories initialized randomly in the full concentration space). Figure 5 shows the two-dimensional NIV embeddings and corresponding eigenvalue spectra of the kernel for the Gillespie SSA example computed from 10, 100, 500, and 1500 data points. We observe that the eigenvalue spectra appear converged for 100 data points and above. This implies that a sufficient amount of data is available merely to discover the intrinsic variables. However, we observe that 100 points are insufficient to construct a self-consistent embedding *in the two variables discovered*. In order to account for the different scaling of the NIV coordinate axes, we require at least 500 data points. This last convergence is exemplified in Figure 5 (c) and (d).

The convergence of the spectrum of the kernel and the ultimate self-consistency of the NIV embedding is our empirical indicator that a sufficient amount of data is available. This is now stated in the paper along with the new Figure 5.

Q3: How were the 3000 random initial conditions chosen for the SSA simualtion. Are they spread over large range of concentrations, or were they close together?

R3: We now describe in the paper how the initial conditions were chosen. We added that “We generate 3000 initial conditions uniformly at random from the region of state space where all concentrations are non-negative.”

Reviewer #2 Evaluations:

Q1: The introduction of the paper is very short. The problem definition, approach and applications are not clear to a general reader. Perhaps the authors can distinguish their study from other multiscale studies/approaches. 

R1: We rewrote and expanded the introduction to include

1. A general description of the dimensionality reduction problem.
2. Specific examples of dimensionality reduction in chemical systems, including references to other studies of the same problem.
3. An outline of our proposed methodology.

To emphasize the problem setting, we renamed Section II-A to “Problem Formulation”.

Q2: The authors can give some physical insight on (a) lack of good quantitative agreement between true and reconstructed configurations in figure 10, and (ii) why the LP outperforms the nearest neighbor approach in figure 11.

R2: The example described in Figure 2 demonstrates the appropriateness of the LP algorithm to signals with an intrinsic multiscale structure. We performed a similar analysis for the data from the alanine dipeptide example, and showed that the signal exhibits a similar error structure, with large errors in some regions of NIV space and small errors in other regions; these results are shown in Figure 13. We therefore concluded that the alanine dipeptide data does contain a measure of multiscale behavior, and, therefore, that LP is an appropriate algorithm for reconstruction.

The discrepancies between the true and reconstructed positions of atoms 9 and for some of the molecular structures is not surprising. From Figure 9, one can see that the first three NIV describe the flipping of atoms 1 and 3 and the dihedral angles phi and psi. However, atoms 9 and 10 do not participate in any of these physical quantities, and so there is little information about their positions contained in the first three NIV. We suspect that using more NIV would result in a better description of atoms 9 and 10 and lead to more accurate structure reconstructions.