We thank the reviewers for their feedback. We will first respond to shared and then to individual comments.

All of the reviewers expressed interest in further comparisons between DCA and other methods, the subject of Section 4. We agree and will include several additional results: 1) comparisons with GPFA and the KF in the analyses of Fig. 4, 2) forecasting results for the neural datasets, 3) visualizations of the extracted DCA and PCA components and 4) a comparison of the loadings, i.e., the magnitudes of the components of the projection matrices, found by DCA and PCA.

Additionally, reviewers 2 and 3 requested clarification regarding the advantages of DCA over other methods. The key advantages of DCA over generative models such as the KF, GPFA and LFADS stem from the fact that DCA is a linear components-analysis method, i.e., it uses a time-independent linear projection for dimensionality reduction. The following specific advantages will be more clearly articulated in the revised paper. First, the loadings for DCA can be inspected to interpret the relationship between the high- and low-dimensional variables, as in the analysis we will include. For instance, one could attempt to correlate each neuron's contribution to the DCA subspace with single-neuron properties, such as its predictive information. Second, it is easy to augment DCA with an ℓ_1 penalty to achieve sparse feature selection, whereas sparse selection is much harder to incorporate in generative models. This could be used to select a small number of "important" neurons in population data, e.g., for BMI control. Third, one often wants to interpret the extracted components as being computed by a linear readout neuron. Finally, DCA can be kernelized to achieve nonlinear dimensionality reduction. Studying the behavior of Kernel DCA is a direction for future studies.

Additionally, we found and corrected a minor bug in Fig. 3A: the SFA and DCA lines are now blue and red, respectively.

Reviewer 1: While the GPFA model features a time-independent linear mapping from the latent space into the observation space, mapping each observation into the latent space (i.e., inference) uses the observations across all of time. Thus, dimensionality reduction using GPFA does not take the form of a time-independent linear projection as in DCA, PCA and SFA, and for this reason we do not lump GPFA in with these methods.

Reviewer 2: The dashed lines in Fig. 3 were intended to help visualization, but will be removed. The time complexity of DCA's fitting procedure does not scale in the total amount of data since DCA only needs spatiotemporal correlations which are computed prior to optimization. Each evaluation of the objective, or its gradient, requires computing $\mathcal{O}\left(T\right)$ quadratic products of the form $V^T C_{\Delta t} V$, each of which is $\mathcal{O}\left(n^2 d + n d^2\right)$, as well as the log determinants, which are $\mathcal{O}\left(T^3\right)$. Altogether, each evaluation is $\mathcal{O}\left(T n^2 d + T n d^2 + T^3\right)$. We will include this analysis.

Reviewer 3: We agree that DCA is not inherently superior to static methods, e.g., PCA and FA, but rather extracts a particular type of structure. Indeed, in neuroscience applications in which generic shared variability due to both dynamics and spatially structured noise is of interest, static methods are well-suited. We will clarify this.

It is true that DCA does not produce an explicit description of the dynamics. However, this is a potentially attractive property: while dynamical generative models such as the KF provide descriptions of the dynamics, they also assume a particular form of dynamics, biasing the extracted components toward this form. By contrast, DCA is formulated in terms of spatiotemporal correlations and, as result, can extract broad forms of (stationary) dynamics, be they linear or nonlinear. For example, the Lorenz attractor is a nonlinear dynamical system. That said, an interesting avenue for future work involves using DCA as a preprocessing step for methods which produce interpretable descriptions of the dynamics. For example, Burton et al. proposed a method for extracting nonlinear ODEs governing multidimensional time series, however this method scales poorly in the time series dimensionality (PNAS, 2016). To circumvent this challenge, DCA could be used to define a small number of dynamical degrees of freedom which are simple linear combinations of the original variables. Alternatively, the ℓ_1-regularized version of DCA could be used to select a small number of original variables which capture the majority of the temporal structure.

CCA differs from DCA in two key ways: 1) CCA finds distinct past and future subspaces and 2) CCA does not capture long-timescale dynamics. With regard to the first difference, note that in most applications, one is interested in finding a single subspace. Moreover, the two subspaces provided by CCA are only interpretable as a pair: the features in U are predictive of the features in V and vice versa. This is why we did not include CCA in our comparisons. With regard to the second difference, note that interesting dynamics are sometimes only discernible at long timescales (Fig. 3A).

We agree that GPFA and the KF do not have excessively many hyperparameters and will clarify this. With regard to computational efficiency, note that DCA does not scale in the total amount of data, whereas GPFA, the KF and LFADS do. Direct comparisons of fitting time are challenging since each method has a tolerance parameter, and it is unclear how to configure these parameters for a fair comparison.

For the DCA vs. PCA comparisons, the fact that the former outperforms the latter to a greater extent on some datasets and to a lesser extent on others reflects the underlying structure of the datasets, and this is potentially interesting. For the DCA vs. SFA comparisons, although the change in R^2 values between the two methods looks small, these values should be compared to the raw R^2 value for each method, which themselves are quite small due to the difficulty of the task and the fact that we are doing simple linear regression. The raw R^2 values will be included in the final manuscript.