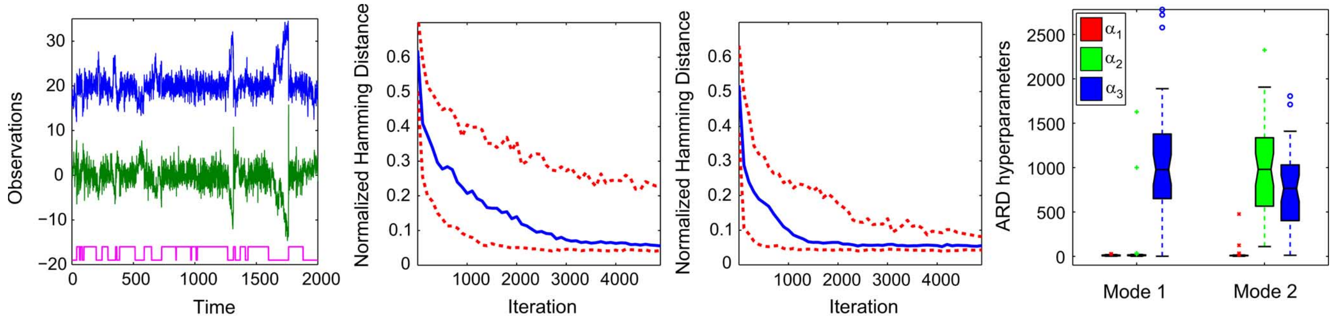

Fig. 3. (a) Observation sequence (green, blue) and mode sequence (magenta) of a 2-mode SLDS, where the first mode can be realized by the first two state components and the second mode solely by the first. The associated 10th, 50th, and 90th Hamming distance quantiles over 100 trials are shown for the (b) MNIW and (c) ARD prior (d) Box plots of inferred ARD precisions associated with the first and second dynamical modes at the 5000th Gibbs iteration. The center line indicates the median, edges the 25th and 75th quantiles and whiskers the range of data excluding outliers which are separately marked. Larger ARD precision values correspond to non-dynamical components.

with 0.98 probability of self-transition and

$$\mathbf{A}^{(1)} = \begin{bmatrix} 0.8 & -0.2 & 0 \\ -0.2 & 0.8 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$
$$\mathbf{A}^{(2)} = \begin{bmatrix} -0.2 & 0 & 0.8 \\ 0.8 & 0 & -0.2 \\ 0 & 0 & 0 \end{bmatrix},$$

with $C = [I_2 \quad 0]$, $\Sigma^{(1)} = \Sigma^{(2)} = I_3$ and $R = I_2$. The first dynamical process can be equivalently described by just the first and second state components since the third component is simply white noise that does not contribute to the state dynamics and is not directly (or indirectly) observed. For the second dynamical process, the third state component is once again a white noise process, but *does* contribute to the dynamics of the first and second state components. However, we can equivalently represent the dynamics of this mode as

$$x_{1,t} = -0.2x_{1,t-1} + \tilde{e}_{1,t}$$
$$x_{2,t} = 0.8x_{1,t-1} + \tilde{e}_{2,t}$$
$$\tilde{\mathbf{A}}^{(2)} = \begin{bmatrix} -0.2 & 0 & 0 \\ 0.8 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

where $\tilde{\boldsymbol{e}}_t$ is a white noise term defined by the original process noise combined with $x_{3,t}$ and $\tilde{\mathbf{A}}^{(2)}$ is the dynamical matrix associated with this equivalent representation of the second dynamical mode. Notice that this SLDS does not satisfy Criterion 3.1 since the second column of $\mathbf{A}^{(2)}$ is zero while the second column of C is not. Nevertheless, because the realization is in our canonical form with $C = [I_2 \quad 0]$, we still expect to recover the $\tilde{\boldsymbol{a}}_2^{(2)} = \tilde{\boldsymbol{a}}_3^{(2)} = 0$ sparsity structure. We set the parameters of the $\mathrm{Gamma}(a,b)$ prior on the ARD precisions as $a = |\mathcal{S}_\ell|$ and $b = a/1000$, where we recall the definition of $\mathcal{S}_\ell$ from (26). This specification fixes the mean of the prior to 1000 while aiming to provide a prior that is roughly equally informative for various choices of model order (i.e., sizes $|\mathcal{S}_\ell|$).

In Fig. 3, we see that even in this low-dimensional example, the ARD provides superior mode-sequence estimates, as well as a mechanism for identifying non-dynamical state components. The box plots of the inferred $\boldsymbol{\alpha}^{(k)}$ are shown in Fig. 3(d). From the clear separation between the sampled dynamic range of $\alpha_3^{(1)}$ and $\left(\alpha_1^{(1)}, \alpha_2^{(1)}\right)$ and between that of $\left(\alpha_2^{(2)}, \alpha_3^{(2)}\right)$ and $\alpha_1^{(2)}$, we see that we are able to correctly identify dynamical systems with $\boldsymbol{a}_3^{(1)} = 0$ and $\tilde{\boldsymbol{a}}_2^{(2)} = \tilde{\boldsymbol{a}}_3^{(2)} = 0$.

C. Dancing Honey Bees

Honey bees perform a set of dances within the beehive in order to communicate the location of food sources. Specifically, they switch between a set of *waggle*, *turn-right* and *turn-left* dances. During the waggle dance, the bee walks roughly in a straight line while rapidly shaking its body from left to right. The turning dances simply involve the bee turning in a clockwise or counterclockwise direction. We display six such sequences of honey bee dances in Fig. 4. The data consist of measurements $\boldsymbol{y}_t = [\cos(\theta_t) \quad \sin(\theta_t) \quad x_t \quad y_t]^T$, where (x_t, y_t) denotes the 2D coordinates of the bee's body and θ_t its head angle. Both Oh *et al.* [10] and Xuan and Murphy [15] used switching dynamical models to analyze these honey bee dances. We wish to analyze the performance of our Bayesian nonparametric variants of these models in segmenting the six sequences into the dance labels displayed in Fig. 4.

1) MNIW Prior—Unsupervised: We start by testing the HDP-VAR(1)-HMM using a MNIW prior. (Note that we did not see performance gains by considering the HDP-SLDS, so we omit showing results for that architecture.) We set the prior distributions on the dynamic parameters and hyperparameters as in Section IV-A for the synthetic data examples, with the MNIW prior based on a preprocessed observation sequence. The pre-processing involves centering the position observations around 0 and scaling each component of $\boldsymbol{y}_t$ to be within the same dynamic range. We compare our results to those of Xuan and Murphy [15], who used a change-point detection technique for inference on this dataset. As shown in Figs. 5(a) and 5(b), our model achieves a superior segmentation compared to the change-point formulation in almost all cases, while also identifying modes which reoccur over time. Example segmentations are shown in Fig. 6. Oh *et al.* [10] also presented an analysis of the honey bee data, using an SLDS with a fixed number of modes. Unfortunately, that analysis is not directly comparable to ours because Oh *et al.* [10] used their SLDS in