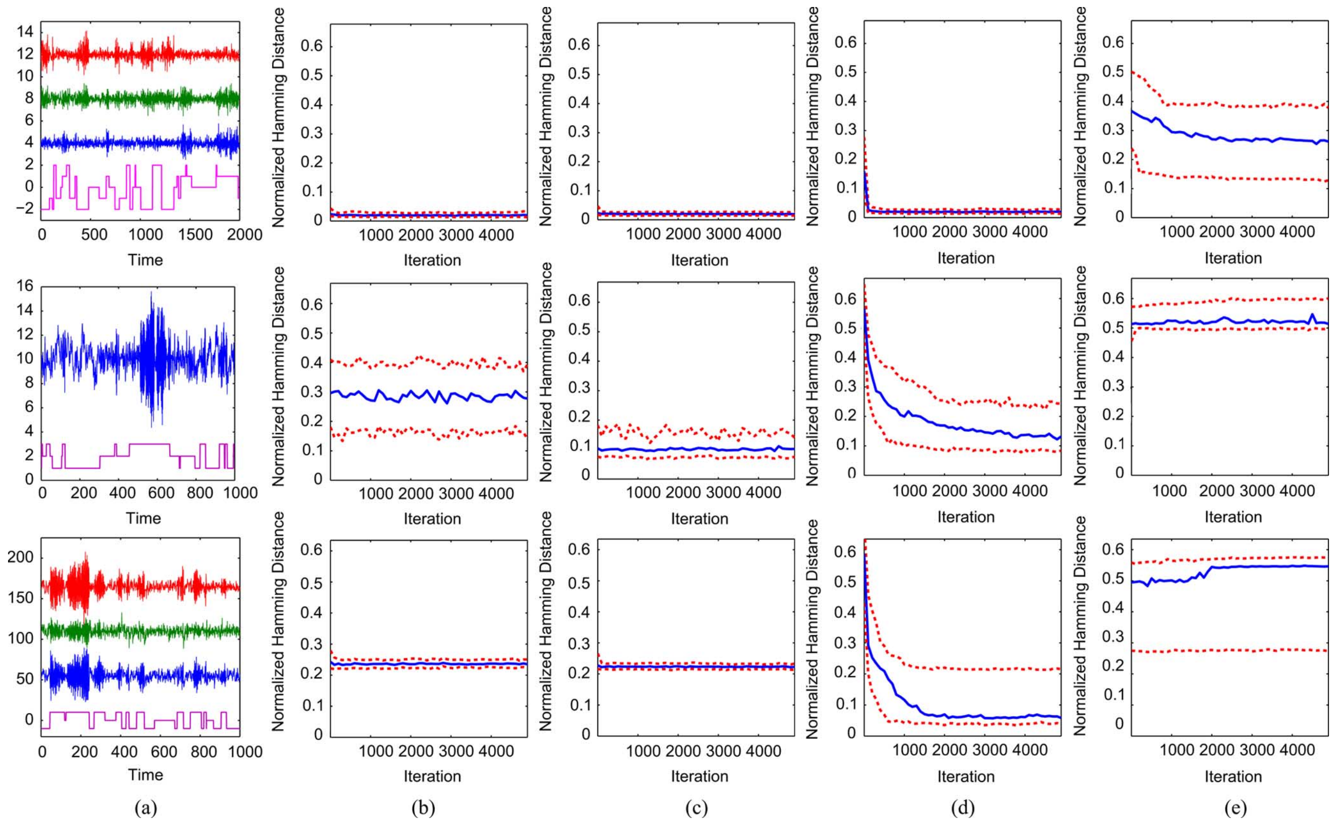

Fig. 2. (a) Observation sequence (blue, green, red) and associated mode sequence (magenta) for a 5-mode switching VAR(1) process (top), 3-mode switching AR(2) process (middle) and 3-mode SLDS (bottom). The components of the observation vector are offset for clarity. The associated 10th, 50th, and 90th Hamming distance quantiles over 100 trials are shown for the (b) HDP-VAR(1)-HMM, (c) HDP-VAR(2)-HMM, (d) HDP-SLDS with $C = I$ (top and bottom) and $C = [1 \quad 0]$ (middle), and (e) sticky HDP-HMM using first difference observations.

IV. RESULTS

A. MNIW Prior

We begin by examining a set of three synthetic datasets displayed in Fig. 2(a) in order to analyze the relative modeling power of the HDP-VAR(1)-HMM, HDP-VAR(2)-HMM and HDP-SLDS using the MNIW prior. Here, we use the notation HDP-$\mathrm{VAR}(r)$-HMM to explicitly denote an order r HDP-AR-HMM with vector observations. We compare to a baseline sticky HDP-HMM using first difference observations, imitating a HDP-VAR(1)-HMM with $A^{(k)} = I$ for all k. In Fig. 2(b)–2(e), we display Hamming distance errors that are calculated by choosing the optimal mapping of indices maximizing overlap between the true and estimated mode sequences.

We place a $\mathrm{Gamma}(a, b)$ prior on the sticky HDP-HMM concentration parameters $\alpha + \kappa$ and γ and a $\mathrm{Beta}(c, d)$ prior on the self-transition proportion parameter $\rho = \kappa/(\alpha + \kappa)$. We choose the weakly informative setting of $a = 1$, $b = 0.01$, $c = 10$ and $d = 1$. The details on setting the MNIW hyperparameters from statistics of the data are discussed in the Appendix.

For the first scenario [Fig. 2 (top)], the data were generated from a five-mode switching VAR(1) process with a 0.98 probability of self-transition and equally likely transitions to the other modes. The same mode-transition structure was used in the subsequent two scenarios, as well. The three switching linear dynamical models provide comparable performance since both the HDP-VAR(2)-HMM and HDP-SLDS with $C = I_3$ contain the class of HDP-VAR(1)-HMMs. In the second scenario [Fig. 2 (middle)], the data were generated from a 3-mode switching AR(2) process. The HDP-AR(2)-HMM has significantly better performance than the HDP-AR(1)-HMM while the performance of the HDP-SLDS with $C = [1 \quad 0]$ performs similarly, but has greater posterior variability because the HDP-AR(2)-HMM model family is smaller. Note that the HDP-SLDS sampler is slower to mix since the hidden, continuous state is also sampled. The data in the third scenario [Fig. 2 (bottom)] were generated from a three-mode SLDS model with $C = I_3$. Here, we clearly see that neither the HDP-VAR(1)-HMM nor HDP-VAR(2)-HMM is equivalent to the HDP-SLDS. Note that all of the switching models yielded significant improvements relative to the baseline sticky HDP-HMM. This input representation is more effective than using raw observations for HDP-HMM learning, but still much less effective than richer models which switch among learned LDS. Together, these results demonstrate both the differences between our models as well as the models' ability to learn switching processes with varying numbers of modes.

B. ARD Prior

We now compare the utility of the ARD prior to the MNIW prior using the HDP-SLDS model when the true underlying dynamical modes have sparse dependencies relative to the assumed model order. That is, the HDP-SLDS may have dynamical regimes reliant on lower state dimensions, or the HDP-AR-HMM may have modes described by lower order VAR processes. We generated data from a two-mode SLDS