as follows. For the cited event dates, we form a "true" label sequence with labels that increment at each event. Then, for each inferred set of change points we form a separate label sequence in an analogous manner (i.e., with incrementing label numbers at each inferred change point.) We then compute the Hamming distance between the true and estimated label sequences after an optimal mapping between these labels. The resulting performances are summarized in the table of Fig. 8(c).

We also analyzed the performance of an HDP-SLDS as defined in Table I. We used raw daily-return observations and first preprocessed the data in the same manner as the honey bee data by centering the observations around 0 and scaling the data to be roughly within a $[-10, 10]$ dynamic range. We then took a MNIW prior on the dynamic parameters, as outlined in the Appendix. Overall, although the state of this HDP-SLDS does not have the interpretation of log-volatilities, we see are still able to capture regime-changes in the dynamics of this stock index and find change points that align better with the true world events than in the MSSV HDP-SLDS model. See Figs. 8(d)–8(h), which also provides a comparison with the change points inferred by an HDP-AR(1)-HMM[4] and a switching AR(1) product partition model (PPM) of Xuan and Murphy [15]. The PPM inferred change points align well with those of the HDP-AR(1)-HMM—we expect this similar performance in such low-dimensional, long time series where the penalty incurred (in terms of quality of parameter estimates) by not revisiting modes is minimal.

B. *Fixed Dynamic Matrix, Switching Driving Noise*

There are some cases in which the dynamical model is well-defined through knowledge of the physics of the system being observed, such as simple kinematic motion. More complicated motions can typically be modeled using the same fixed dynamical model, but with a more complex description of the driving force. A generic LDS driven by an unknown control input $\boldsymbol{u}_t$ can be represented as

$$\begin{aligned} \boldsymbol{x}_t &= A\boldsymbol{x}_{t-1} + B\boldsymbol{u}_t + \boldsymbol{v}_t \\ \boldsymbol{y}_t &= C\boldsymbol{x}_t + D\boldsymbol{u}_t + \boldsymbol{w}_t \end{aligned} \tag{41}$$

where $\boldsymbol{v}_t \sim \mathcal{N}(0, Q)$ and $\boldsymbol{w}_t \sim \mathcal{N}(0, R)$. It is often appropriate to assume $D = 0$, as we do herein.

1) Maneuvering Target Tracking: Target tracking provides an application domain in which one often assumes that the dynamical model is known. One method of describing a maneuvering target is to consider the control input as a random process [46]. For example, a *jump-mean* Markov process [47] yields dynamics described as

$$\begin{aligned} z_t \mid z_{t-1} &\sim \pi_{z_{t-1}} \\ \boldsymbol{x}_t &= A\boldsymbol{x}_{t-1} + B\boldsymbol{u}_t^{(z_t)} + \boldsymbol{v}_t \\ \boldsymbol{y}_t &= C\boldsymbol{x}_t + \boldsymbol{w}_t \\ \boldsymbol{u}_t^{(k)} &\sim \mathcal{N}\left(\boldsymbol{\mu}^{(k)}, \Sigma^{(k)}\right). \end{aligned} \tag{42}$$

[4]We do not compare to an HDP-AR(1)-HMM for the MSSV formulation since there is no adequate way to capture the complex MSSV observation model with an autoregressive process.

Classical approaches rely on defining a fixed set of dynamical modes and associated transition distributions. The state dynamics of (42) can be equivalently described as

$$\begin{aligned} \boldsymbol{x}_t &= A\boldsymbol{x}_{t-1} + \boldsymbol{e}_t^{(z_t)} \\ \boldsymbol{e}_t^{(k)} &\sim \mathcal{N}(B\boldsymbol{\mu}^{(k)}, B\Sigma^{(k)}B^T + Q). \end{aligned} \tag{43}$$

This model can be captured by our HDP-SLDS formulation of (37) with a fixed dynamic matrix (e.g., constant velocity or constant acceleration models [46]) and mode-specific, non-zero mean process noise. Such a formulation was explored in [9] along with experiments that compare the performance to that of standard multiple model techniques, demonstrating the flexibility of the Bayesian nonparametric approach. Fox *et al.* [9] also present an alternative sampling scheme that harnesses the fact that the control input may be much lower-dimensional than the state and sequentially block-samples $(z_t, \boldsymbol{u}_t)$ analytically marginalizing over the state sequence $\boldsymbol{x}_{1:T}$. Note that this variant of the HDP-SLDS can be viewed as an extension of the work by Caron *et al.* [20] in which the exogenous input is modeled as an independent noise process (i.e., no Markov structure on z_t) generated from a DP mixture model.

VI. Conclusion

In this paper, we have addressed the problem of learning switching linear dynamical models with an unknown number of modes for describing complex dynamical phenomena. We presented a Bayesian nonparametric approach and demonstrated both the utility and versatility of the developed HDP-SLDS and HDP-AR-HMM on real applications. Using the same parameter settings, although different model choices, in one case we are able to learn changes in the volatility of the IBOVESPA stock exchange while in another case we learn segmentations of data into *waggle*, *turn-right* and *turn-left* honey bee dances. We also described a method of applying automatic relevance determination (ARD) as a sparsity-inducing prior, leading to flexible and scalable dynamical models that allow for identification of variable order structure. We concluded by considering adaptations of the HDP-SLDS to specific forms often examined in the literature such as the Markov switching stochastic volatility model and a standard multiple model target tracking formulation.

The batch processing of the Gibbs samplers derived herein may be impractical and offline-training online-tracking infeasible for certain applications. Due both to the nonlinear dynamics and uncertainty in model parameters, exact recursive estimation is infeasible. One could leverage the *conditionally linear* dynamics and use *Rao-Blackwellized particle filtering* (RBPF) [48]. However, one challenge is that such particle filters can suffer from a progressively impoverished particle representation. A possible direction of future research is to consider building on the recent work of [49] and embedding a RBPF within an MCMC algorithm. Another interesting avenue of research is to analyze high-dimensional time series. Although there is nothing fundamentally different in considering such datasets, based on experiments in related models [21] we expect to run into mixing rate issues with the Gibbs sampler since the parameter associated with each new considered dynamical