# Hidden Markov Modeling of X-Ray Light Curves

Robert Zimmerman
University of Toronto + Imperial College London

Joint work with David A. van Dyk (Imperial College London), Vinay L. Kashyap (Harvard & Smithsonian), and Aneta Siemiginowska (Harvard & Smithsonian)

CHASC AstroStatistics Center
Harvard & Smithsonian
November 13, 2024

# Short-Duration Flares

- Certain stars produce sporadic short-duration flares which emanate from their coronae
- Interest lies in understanding the proportion of time a star spends in flaring and quiescent states
- The sun is close enough to be directly observable and we have plenty of "continuous" information to work with (e.g., Stanislavsky et al., 2020)
- However, for distant stars that emit X-rays, all we have are light curves computed from lists of photons recorded by X-ray telescopes (just time and energy)

# *EV Lac*

Figure 1: *EV Lac* unleashing a monster flare (image source: NASA)

# Bivariate *EV Lac* Light Curves

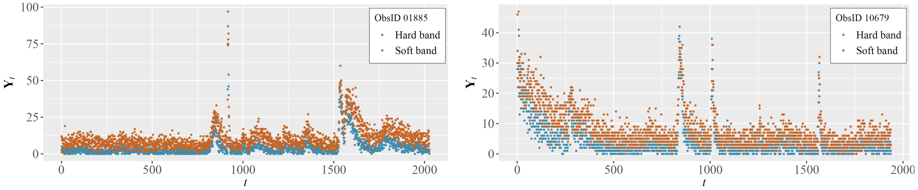


Figure 2: Bivariate light curves: time series plots of *EV Lac* count data based on event lists (September 2001 left; March 2009 right) split into hard (1.5–8.0 keV) and soft (0.3–1.5) keV passbands

- Previous work on flaring state estimation mostly apply ad-hoc rules or black-box/model-free learning methods (e.g., neural networks)
- For the *EV Lac* data above, the best guesses so far for flaring state proportions are 39% (September 2001) and 29% (March 2009) (Huenemoerder et al., 2010)

# The Plan

- In this project, we use *hidden Markov models (HMMs)* to model flaring and quiescence
- We take a two-stage approach in our analysis
- In Stage 1, we use HMMs to predict the values of a continuous latent process that stochastically induces the observations in the data
- In Stage 2, we use a finite mixture model (a special case of an HMM) to approximate the distribution of the predicted states and use it to estimate the proportion of time *EV Lac* spends flaring

# (Discrete-Time) Hidden Markov Models

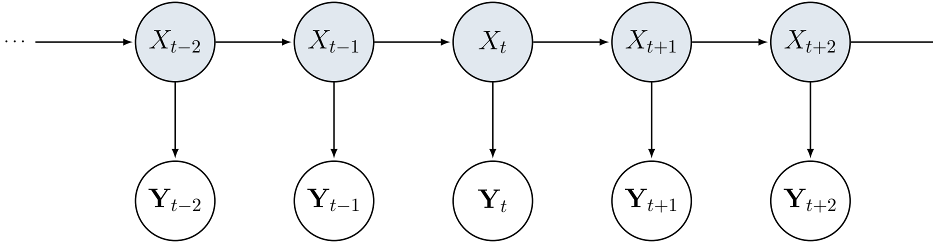


Figure 3: A graphical model representing the standard discrete-time HMM dependence structure

- An HMM consists of an unobserved Markov chain $X_{1:T} \subset \mathcal{X}$ and an observed time series $\mathbf{Y}_{1:T}$ such that...
  - $X_t$ determines the distribution of $\mathbf{Y}_t$, and
  - $\mathbf{Y}_s$ and $\mathbf{Y}_t$ are conditionally independent given $X_{1:T}$

# Discrete-Space HMMs...

- When the state space $\mathcal{X}$ is finite — say $\mathcal{X} = \{1, \ldots, K\}$ — the HMM is a *discrete-space HMM*

- These are characterized by initial probabilities $\delta_i$, transition probabilities $\gamma_{i,j}$, and state-dependent densities/mass functions $h_i(\cdot \mid \boldsymbol{\lambda}_i)$ for $i, j \in \mathcal{X}$

- The likelihood is given by

$$L(\boldsymbol{\eta} \mid \mathbf{y}_{1:T}) = \sum_{x_1=1}^{K} \cdots \sum_{x_T=1}^{K} \left( \delta_{x_1} \cdot h_{x_1}(\mathbf{y}_1 \mid \boldsymbol{\lambda}_{x_1}) \prod_{t=2}^{T} \left( \gamma_{x_{t-1},x_t} \cdot h_{x_t}(\mathbf{y}_t \mid \boldsymbol{\lambda}_{x_t}) \right) \right)$$

- This can be calculated efficiently ("forward algorithm", etc.)!

- So why not assume a 2-state (or maybe 3-state) HMM for our data?

# ...Don't Really Work

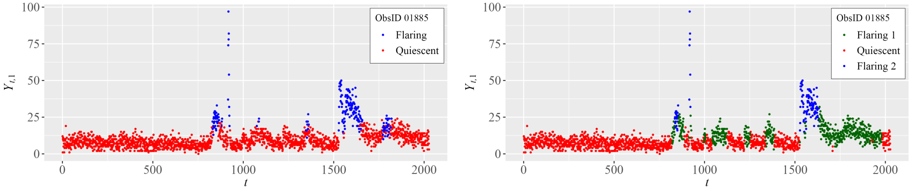


Figure 4: Soft-band curves colored with classifications based on 2-state (left) and 3-state (right) HMMs fit directly to the observed data from September 2001

- The conditional independence HMM assumption clearly fails for this data!
- There is some kind of continuous temporal trend driving the emissions
- So we need something else

# Continuous-Space HMMs and State-Space Models

- A *continuous-space HMM* has the same definition as a discrete-space HMM, except now the underlying chain $\mathbf{X}_{1:T}$ takes values in a continuum (we take $\mathcal{X} = \mathbb{R}^d$)

- When the dynamics of $\mathbf{X}_{1:T}$ are specified, the resulting continuous-space HMM is an example of a *state-space model*

- For example, a general *Poisson state-space model* is given by

$$\mathbf{Y}_t \mid \mathbf{X}_t \sim \text{Poisson}(w \cdot \beta_1 \cdot e^{X_{t,1}}) \cdot \text{Poisson}(w \cdot \beta_2 \cdot e^{X_{t,2}}),$$
$$\mathbf{X}_t = \mathbf{\Phi}\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t,$$
$$\text{where } \mathbf{\Phi} = \begin{bmatrix} \phi_1 & \phi_{12} \\ \phi_{21} & \phi_2 \end{bmatrix} \text{ and } \boldsymbol{\varepsilon}_t \overset{\text{iid}}{\sim} \mathcal{N}_2\left(\mathbf{0}, \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}\right)$$

# Three Nested State-Space Models for Flaring Coronae

- Model 1: AR(1) Process

$$\mathbf{Y}_t \mid X_t \sim \text{Poisson}(w \cdot \beta_1 \cdot e^{X_t}) \cdot \text{Poisson}(w \cdot \beta_2 \cdot e^{X_t}),$$
$$X_t = \phi X_{t-1} + \varepsilon_t, \text{ where } \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$$

- Model 2: Uncorrelated VAR(1) Process On a Line

$$\mathbf{Y}_t \mid \mathbf{X_t} \sim \text{Poisson}(w \cdot \beta_1 \cdot e^{X_{t,1}}) \cdot \text{Poisson}(w \cdot \beta_2 \cdot e^{X_{t,2}}),$$
$$\mathbf{X}_t = \mathbf{\Phi}\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t,$$
$$\text{where } \mathbf{\Phi} = \begin{bmatrix} \phi & 0 \\ 0 & \phi \end{bmatrix} \text{ and } \boldsymbol{\varepsilon}_t \overset{\text{iid}}{\sim} \lim_{\rho \to 1} \mathcal{N}_2\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix} \right)$$

- Model 3: Uncorrelated VAR(1) Process

$$\mathbf{Y}_t \mid \mathbf{X_t} \sim \text{Poisson}(w \cdot \beta_1 \cdot e^{X_{t,1}}) \cdot \text{Poisson}(w \cdot \beta_2 \cdot e^{X_{t,2}}),$$
$$\mathbf{X}_t = \mathbf{\Phi}\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t,$$
$$\text{where } \mathbf{\Phi} = \begin{bmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{bmatrix} \text{ and } \boldsymbol{\varepsilon}_t \overset{\text{iid}}{\sim} \mathcal{N}_2\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix} \right)$$

# State-Space Model Estimation Is Non-Trivial

- Such state-space models are fully characterized by an initial density $\delta(\mathbf{x})$, a transition kernel $\gamma(\mathbf{x}, \mathbf{y})$, and state-dependent densities $h_{\mathbf{x}}(\cdot \mid \boldsymbol{\lambda}_{\mathbf{x}})$ for $\mathbf{x}, \mathbf{y} \in \mathcal{X}$
- The likelihood function for the general state-space model is

$$L(\boldsymbol{\eta} \mid \mathbf{y}_{1:T}) = \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} \delta(\mathbf{x}_1) \cdot h_{\mathbf{x}_1}(\mathbf{y}_1 \mid \boldsymbol{\lambda}_{\mathbf{x}_1}) \prod_{t=2}^{T} \gamma(\mathbf{x}_{t-1}, \mathbf{x}_t) \cdot h_{\mathbf{x}_t}(\mathbf{y}_t \mid \boldsymbol{\lambda}_{\mathbf{x}_t}) \, \mathrm{d}\mathbf{x}_{T:1}$$

  where $\mathrm{d}\mathbf{x}_{T:1} = \mathrm{d}\mathbf{x}_T \cdots \mathrm{d}\mathbf{x}_1$
- Even when $\mathcal{X} = \mathbb{R}$, this cannot be *computed*, let alone maximized

# A Discrete-Space HMM Approximation

- By reducing $\mathcal{X}$ to a bounded subset $A \subset \mathbb{R}^d$, partitioning it into $m$ hyperrectangles $A = \bigcup_{i=1}^m A_i$, and selecting a "representative point" $\mathbf{c}_i^* \in A_i$ within each, the likelihood can be approximated as

$$L(\boldsymbol{\eta} \mid \mathbf{y}_{1:T}) \approx \sum_{i_1=1}^{m} \cdots \sum_{i_T=1}^{m} \Bigg( \mathbb{P}(\mathbf{X}_1 \in A_{i_1}) \cdot h_{\mathbf{c}^*_{i_1}}(\mathbf{y}_1 \mid \boldsymbol{\lambda}_{\mathbf{c}^*_{i_1}}) \cdot \prod_{t=2}^{T} \Big( \mathbb{P}(\mathbf{X}_t \in A_{i_t} \mid \mathbf{X}_{t-1} = \mathbf{c}^*_{i_{t-1}}) \cdot h_{\mathbf{c}^*_{i_t}}(\mathbf{y}_t \mid \boldsymbol{\lambda}_{\mathbf{c}^*_{i_t}}) \Big) \Bigg)$$

- With a change of notation, this is

$$L(\boldsymbol{\eta} \mid \mathbf{y}_{1:T}) \approx \sum_{i_1=1}^{m} \cdots \sum_{i_T=1}^{m} \Bigg( \tilde{\delta}_{i_1} \cdot h_{\mathbf{c}^*_{i_1}}(\mathbf{y}_1 \mid \boldsymbol{\lambda}_{\mathbf{c}^*_{i_1}}) \prod_{t=2}^{T} \Big( \tilde{\gamma}_{i_{t-1}, i_t} \cdot h_{\mathbf{c}^*_{i_t}}(\mathbf{y}_t \mid \boldsymbol{\lambda}_{\mathbf{c}^*_{i_t}}) \Big) \Bigg)$$

which is (essentially) a discrete-space HMM likelihood!

# Model Selection for *EV Lac*

- With some computational tricks, we maximize the approximate likelihood numerically and use the parametric bootstrap for bias-correction and standard errors
- For Model 1 vs 2, the LRT works easily and clearly favors Model 2
- For Model 2 vs 3, the LRT fails! So we use other checks and criteria
- For example, the estimate of $\rho$ in Model 3 is $\hat{\rho} = 0.99999987$... with a standard error of $\approx 0$
- ...we go with Model 2

# Flaring State Interval Classification

- How do we perform inference on the hidden states?
- Once the state-space model is fit, we can make posterior state predictions $\hat{X}_t$ for each $X_t$ (this is *local decoding*)
  - Specifically: $\hat{X}_t = \underset{x_t \in \mathcal{X}}{\operatorname{argmax}} \, \mathbb{P}_{\hat{\boldsymbol{\theta}}}(X_t = x_t \mid \mathbf{Y}_{1:T} = \mathbf{y}_{1:T})$
- We view the $\hat{X}_1, \ldots, \hat{X}_T$ as fresh "data" and approximate their distribution by a 2-component mixture:

$$\hat{X}_1, \ldots, \hat{X}_T \overset{\text{iid}}{\sim} \alpha \cdot F_1 + (1 - \alpha) \cdot F_2$$

- If we assume that the distribution $F_2$ corresponds to "flaring", then $(1 - \alpha)$ is the overall proportion of time spent in this state

# Semi-Supervised Classification

- If we have a clear sustained period $[t_1, t_q]$ of quiescence at hand, we use $\hat{X}_{t_1:t_q}$ as training data for a KDE for the quiescent mixture component
- We approximate the flaring mixture component with a step function
- We fit the mixture using a custom-designed EM algorithm, which gives $100\% \cdot (1 - \hat{\alpha}) \approx 45\%$

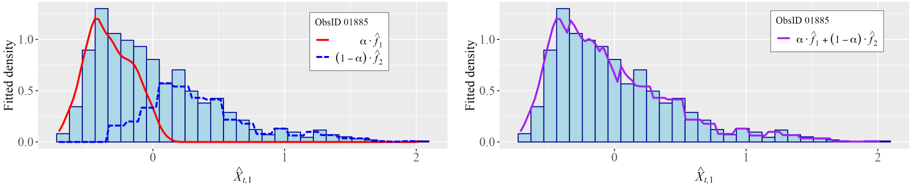


Figure 5: Fitted component densities (left) and mixture density (right) for September 2001, overlaid on a histogram of $\{\hat{X}_1, \ldots \hat{X}_{2027}\}$

# Unsupervised Classification

- If no sustained period of quiescence is available, we instead use a 3-component normal mixture,

$$\hat{X}_1, \ldots, \hat{X}_T \overset{\text{iid}}{\sim} \alpha_1 \cdot \mathcal{N}(\mu_1, \tau_1^2) + \alpha_2 \cdot \mathcal{N}(\mu_2, \tau_2^2) + \alpha_3 \cdot \mathcal{N}(\mu_3, \tau_3^2)$$

  where the "left" two components correspond to quiescence

- Also easily fit with an EM algorithm, which gives $100\% \cdot \hat{\alpha}_3 \approx 27\%$

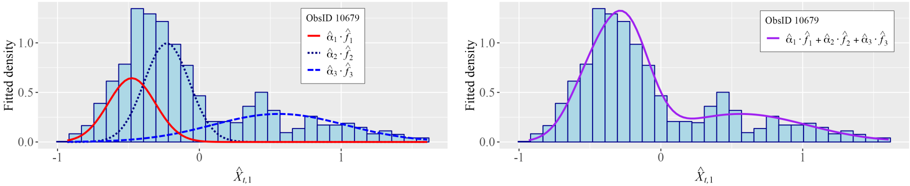


Figure 6: Fitted component densities (left) and mixture density (right) for March 2009, overlaid on a histogram of $\{\hat{X}_1, \ldots \hat{X}_{1937}\}$

# High Resolution Spectra

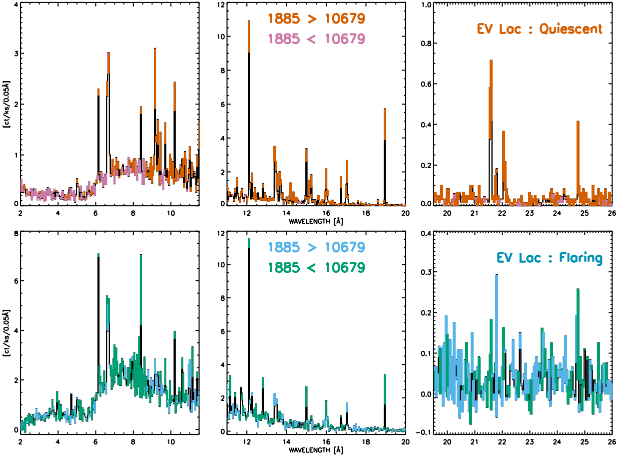


Figure 7: Spectra from both September 2001 and March 2009 epochs shown superposed for both flaring and quiescence; the overall brightness is higher, and the continuum is stronger and more prominent during the flaring state, signifying a different thermal signature

# Potential Future Work

- Consider more complex models for $Y_{t,1} \mid X_t$ and $Y_{t,2} \mid X_t$
  - Instead of assuming conditional independence, model dependence using state-dependent copulas
  - Model other temporal patterns (e.g., a flare in one band proceeds a flare in another)
- Enlarge the set of distributions of $\mathbf{X}_t \mid \mathbf{X}_{t-1}$ under consideration
  - Multivariate-$t$ for heavier tails
  - Mixture distributions for more complicated physical mechanisms
- Split the observations into more passbands and thus move $\mathbf{X}_{1:T}$ into a higher-dimensional state-space (many covariance structures available)
  - But even our "efficient" estimation technique suffers from the curse of dimensionality! Can we get around this?
- Etc...

Thank you!

# References

Huenemoerder, D. P., Schulz, N. S., Testa, P., Drake, J. J., Osten, R. A., and Reale, F. (2010). X-ray flares of EV Lac: Statistics, spectra, and diagnostics. *The Astrophysical Journal*, 723(2):1558.

Stanislavsky, A., Nitka, W., Małek, M., Burnecki, K., and Janczura, J. (2020). Prediction performance of hidden markov modelling for solar flares. *Journal of Atmospheric and Solar-Terrestrial Physics*, 208:105407.