An efficient Gaussian process framework for analysis of oscillations in nonstationary time series

Andrew H. Song [1] **Demba Ba** [2] **Emery N. Brown** [3]

Abstract

We propose **P**iecewise **L**ocally **S**tationary **O**scillation (PLSO) state-space model for decomposing nonstationary time series with slowly time-varying spectra into several oscillatory, piecewise-stationary processes. PLSO combines piecewise stationarity in signal processing and stationary Gaussian process kernels, effectively addressing the drawbacks of these ideas, such as inefficient inference and discontinuous/distorted estimates across stationary interval boundaries.

1. Introduction

With the collection of long time series now common in neuroscience and biological sciences, it is important to develop an *efficient* inference framework that does not rely on the restrictive stationarity assumption. We focus on data 1) with spectral dynamics that change slowly over time and 2) for which decomposition into several oscillatory components is warranted for interpretation, such as electroencephalogram or electrophysiology recordings.

Can we use Gaussian Process (GP) (Rasmussen & Williams, 2005)? We can indeed model the data as samples from a GP with spectral mixture (SM) kernel (Wilson & Adams, 2013; Solin & Särkkä, 2014; Wilkinson et al., 2019), which consists of the superposition of *localized* and *frequency-modulated* kernels. However, these are fundamentally stationary kernels that assume constant spectra across time, and thus cannot capture evolving spectral dynamics. Although nonstationary kernels (Paciorek & Schervish, 2003; Remes et al., 2017) offer a potential solution, these are expensive - The computational complexity cubic in the number of samples K, making them prohibitive for hours of biological time series with $> 10^3$ Hz sampling rate.

Another line of work from signal processing literature segments the data into small piecewise stationary (PS) intervals and fits stationary models *independently* to each interval. The short-time Fourier transform (STFT) (Oppenheim et al., 2009) for time-frequency analysis is a representative example. This philosophy can also be found in GP literature (Gramacy & Lee, 2008; Solin et al., 2018). The major drawback is that the estimates in different intervals are independent, conditioned on the data, and do not reflect the dependence across intervals. This causes discontinuity/distortion of the time-domain estimates near the interval boundaries.

We propose **P**iecewise **L**ocally **S**tationary **O**scillation (PLSO) framework, a generative framework for nonstationary time series that combines the paradigm of piecewise stationarity and stationary GP kernels, which addresses these shortcomings[1]. PLSO is a time-domain state-space model, with time-varying parameters across different intervals. In each interval, PLSO yields a spectra that is a superposition of localized and frequency-modulated stationary kernels. The state-space model offers an efficient inference, with computational complexity linear in K, and ensures stochastic continuity/dependence across different intervals.

2. PLSO and its mathematical properties

2.1. Notation

We use $j \in \{1, \dots, J\}$ and $k \in \{1, \dots, K\}$ to denote frequency and discrete-time sample index. We use $\omega \in [-\pi, \pi]$ for normalized frequency. The j^{th} latent process centered at $\omega = \omega_j$ is denoted as $\mathbf{z}_j \in \mathbb{C}^K$, with $\mathbf{z}_{j,k} \in \mathbb{C}$ denoting the k^{th} sample of $\mathbf{z}_j$ and $\mathbf{z}_{j,k}^{\Re}$, $\mathbf{z}_{j,k}^{\Im}$ its real and imaginary parts. We also represent $\mathbf{z}_{j,k}$ as $\widetilde{\mathbf{z}}_{j,k} = [\mathbf{z}_{j,k}^{\Re}, \mathbf{z}_{j,k}^{\Im}]^{\mathrm{T}} \in \mathbb{R}^2$.

2.2. Generative model

PLSO is based on the Theorem 1 of Adak (1998), which states that the widely-used piecewise locally stationary (PLS) process, which includes local stationary (Dahlhaus,

[1]Department of electrical engineering and computer sciences, Massachusetts Intitute of Technology (MIT), Cambridge, USA [2]School of engineering and applied sciences, Harvard university, Cambridge, USA [3]Picower Institute of Learning and Memory, MIT, Cambridge, Massachusetts, USA. Correspondence to: Andrew H. Song <andrew90@mit.edu>.

Time series Workshop. Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

[1]The code is available at https://github.com/andrewsong90/plso

1997) processes, can be approximated as a piecewise stationary (PS) process. To this end, we partition the nonstationary data $\mathbf{y} \in \mathbb{R}^K$ into M non-overlapping intervals, indexed by $m \in \{1, \dots, M\}$, of length N, indexed by $n \in \{1, \dots, N\}$, such that $K = MN$. We model data in each segment as a superposition of *oscillatory* PS processes $\{\mathbf{z}_j\}_j$, with $\mathbf{z}_j$ centered at frequency ω_j (Matsuda & Komaki, 2017),

$$\begin{aligned} \widetilde{\mathbf{z}}_{j,mN+n} &= \rho_j \overbrace{\begin{pmatrix} \cos(\omega_j) & -\sin(\omega_j) \\ \sin(\omega_j) & \cos(\omega_j) \end{pmatrix}}^{\mathbf{R}(\omega_j)} \widetilde{\mathbf{z}}_{j,mN+(n-1)} \\ &\quad + \sigma_{j,m}(1-\rho_j^2)^{1/2} \cdot \varepsilon_{j,mN+n} \\ \mathbf{y}_{mN+n} &= \sum_{j=1}^{J} \mathbf{z}^{\Re}_{j,mN+n} + \sigma_\nu \cdot \nu_{mN+n}, \end{aligned} \tag{1}$$

where $\rho_j = \exp(-\Delta/l_j)$ with sampling interval Δ, $\varepsilon_{j,mN+n} \sim \mathcal{N}(0, \mathbf{I}_{2\times 2})$, and $\nu_{mN+n} \sim \mathcal{N}(0,1)$. The rotation matrix $\mathbf{R}(\omega_j)$ rotates $\widetilde{\mathbf{z}}_{j,mN+(n-1)}$ by ω_j for each transition, and is equivalent to the operation $\exp(-i\omega_j)\mathbf{z}_{j,mN+(n-1)}$. This dynamics is what makes a $\mathbf{z}_j$ latent oscillatory process. The parameterization of ρ_j and $\varepsilon_{j,mN+n}$ establishes connection between the PLSO autocovariance and the exponential GP kernels, as explained later. We define $\mathbf{P}^j_{m,n} = \mathbb{E}[\widetilde{\mathbf{z}}_{j,mN+n}(\widetilde{\mathbf{z}}_{j,mN+n})^{\mathrm{T}}]$ as the covariance of $\widetilde{\mathbf{z}}_{j,mN+n}$, with $\mathbf{P}^j_{1,1} = \sigma^2_{j,1}\mathbf{I}_{2\times 2}, \forall j$.

Hyperprior for $\sigma^2_{j,m}$ For a given ω_j, we impose stochastic continuity on $\log \sigma^2_{j,m}$, as $\log(\sigma^2_{j,m}) = \log(\sigma^2_{j,m-1}) + \eta_{j,m}$, with $\eta_{j,m} \sim \mathcal{N}(0, 1/\sqrt{\lambda})$. This hyperprior pools estimates of $\{\sigma^2_{j,m}\}_m$ to 1) prevent overfitting to the noisy spectra and 2) estimate smooth dynamics of $\{\sigma^2_{j,m}\}_m$, the smoothness of which is controlled by λ. We use $\log(\cdot)$ for nonnegativity of $\sigma^2_{j,m}$. Fig. 1 depicts the graphical model for PLSO.

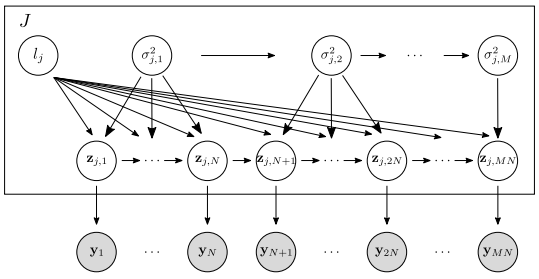

Figure 1. The graphical model for PLSO.

Connection to time-frequency domain Given Eq. 1, we can express the spectra of PLSO in each interval, through the autocovariance. The autocovariance of $\mathbf{z}_j$ for m^{th} interval is $\mathbb{E}[\mathbf{z}^{\Re}_{j,mN}\mathbf{z}^{\Re}_{j,mN+n'}] = \sigma^2_{j,m}\cos(\omega_j n')\exp(-n'\Delta/l_j)$, which is an exponential kernel, frequency-modulated by ω_j. The spectra for $\mathbf{z}_j$ is $S^{(m)}_j(\omega) = \varphi^{(m)}_j(\omega) + \varphi^{(m)}_j(-\omega)$, obtained by taking the Fourier transform (FT) of the autocovariance, where $\varphi^{(m)}_j(\omega) = \frac{\sigma^2_j(1-\exp(-2\Delta/l_j))}{1+\exp(-2\Delta/l_j)-2\exp(-\Delta/l_j)\cos(\omega-\omega_j)}$. The power spectral density (PSD) of the entire process is given as $\gamma^{(m)}(\omega) = \sum_{j=1}^J S^{(m)}_j(\omega)$, leveraging the fact that 1) $\mathbf{z}_j$ is independent across j and 2) FT is a linear operator.

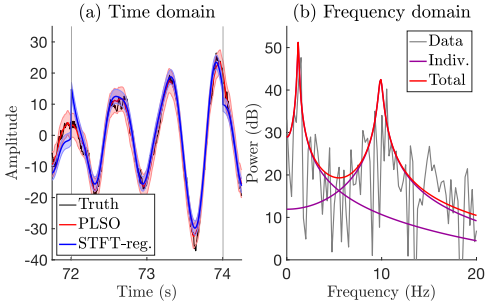

Figure 2. A simulated example. (a) Time domain. Data (black) around boundaries (gray) and inferred time series using PLSO (red) and regularized STFT (blue). (b) Frequency domain. Spectrum of the data (gray), PLSO for $J = 2$ (purple) and their sum (red).

PLSO thus provides a *parameterized* spectrogram, which captures the spectral dynamics of the data. The lengthscale l_j controls the bandwidth of the j^{th} process, with *large* l_j corresponding to *narrower* bandwidth. The time-varying variance $\sigma^2_{j,m}$ results in time-varying $S^{(m)}_j(\omega)$ and $\gamma^{(m)}(\omega)$.

Fig. 2 shows a simulated example. In the time domain, the oscillation inferred by the regularized STFT (Kim et al., 2018), which imposes stochastic continuity on the STFT coefficients, suffers from discontinuity/distortion near boundaries, whereas that inferred by PLSO does not. In the frequency domain, the PSD $\gamma^{(m)}(\omega)$ comprised of PLSO components $\{S^{(m)}_j(\omega)\}_j$, is fit to the data STFT $I^{(m)}(\omega_n) = |\sum_{n'=1}^N \exp(-2\pi i(n'-1)(n-1)/N)\,\mathbf{y}_{mN+n'}|^2$ for discrete frequency $\omega_n = 2\pi n/N, 1 \le n \le N$.

2.3. Mathematical properties

We now analyze two important properties of PLSO: 1) stochastic continuity and 2) piecewise stationarity. The proofs to the propositions are in the **Appendix**.

Stochastic continuity The state-space model for $\mathbf{z}_j$ provides stochastic continuity across different intervals. We define this rigorously in the following proposition.

Proposition 1. *For a given m, as $\Delta \to 0$, the samples on either side of the interval boundary, which are $\widetilde{\mathbf{z}}_{j,(m+1)N}$ and $\widetilde{\mathbf{z}}_{j,(m+1)N+1}$, converge to each other in mean square,*

$$\lim_{\Delta\to 0} \mathbb{E}[\Delta\widetilde{\mathbf{z}}_{j,(m+1)N}\Delta\widetilde{\mathbf{z}}^T_{j,(m+1)N}] = 0,$$

where we use $\Delta\widetilde{\mathbf{z}}_{j,(m+1)N} = \widetilde{\mathbf{z}}_{j,(m+1)N+1} - \widetilde{\mathbf{z}}_{j,(m+1)N}$.

This matches our intuition that the adjacent samples from the same continuous process should converge as $\Delta \to 0$.

Piecewise stationarity For the m^{th} interval to be piecewise stationary, the initial covariance $\mathbf{P}^j_{m,1}$ should equal *steady-state* covariance, denoted as $\mathbf{P}^j_{m,\infty}$. However, ensuring $\mathbf{P}^j_{m+1,1} = \mathbf{P}^j_{m+1,\infty}$ is not possible, when the covariance of the previous window is different, i.e., $\mathbf{P}^j_{m,N} = \mathbf{P}^j_{m,\infty} \neq \mathbf{P}^j_{m+1,\infty}$. For this to hold, the *variance* of $\varepsilon_{j,(m+1)N+1}$ has to be $\mathbf{P}^j_{m+1,\infty} - \exp\left(-2\Delta/l_j\right)\mathbf{P}^j_{m,\infty}$, which violates the Proposition 1, as it does not converge to zero as $\Delta \to 0$. Consequently, the trajectory is discontinuous.

In summary, there exists a trade-off between maintaining *piecewise stationarity* and *continuity* across intervals. PLSO strikes the balance by maintaining continuity while ensuring a quick transition to the steady-state covariance. The following proposition quantifies the speed of transition.

Proposition 2. *Assume $l_j \ll N\Delta$, such that $\mathbf{P}^j_{m,N} = \mathbf{P}^j_{m,\infty}$. The difference between $\mathbf{P}^j_{m,\infty}$ and $\mathbf{P}^j_{m+1,\infty}$ decays exponentially fast as a function of n, i.e., $\mathbf{P}^j_{m+1,n} = \mathbf{P}^j_{m+1,\infty} + \exp\left(-2n\Delta/l_j\right) \cdot (\mathbf{P}^j_{m,\infty} - \mathbf{P}^j_{m+1,\infty})$.*

This implies that, except for the transition at the beginning of each window, we can assume stationarity.

2.4. Related works

The spectral mixture (SM) kernel (Wilson & Adams, 2013) uses a superposition of frequency-modulated *squared-exponential* kernels, while PLSO uses *exponential* kernels. PLSO has the equivalent time-domain state-space model for efficient inference, unlike the SM approach. The infinite horizon Gaussian Process (IHGP) (Solin et al., 2018) also offers complexity linear in K. For nonstationary data, however, it uses the PS assumption and fits the model *independently* to each interval, which leads to boundary artifacts.

3. Inference

We perform inference on the posterior distribution

$$p(\{\mathbf{z}_j\}_j, \{\sigma^2_{j,m}\}_{j,m} \mid \mathbf{y}, \theta) = \underbrace{p(\{\sigma^2_{j,m}\}_{j,m} \mid \mathbf{y}, \theta)}_{\text{window-level posterior}} \cdot \underbrace{p(\{\mathbf{z}_j\}_j \mid \{\sigma^2_{j,m}\}_{j,m}, \mathbf{y}, \theta)}_{\text{sample-level posterior}}, \quad (2)$$

with $\theta = \{\lambda, \sigma^2_\nu, \{l_j\}_j, \{\omega_j\}_j\}$ to be learned. We choose J that minimizes the Akaike Information Criterion (AIC) (Akaike, 1981). We break the inference into two stages.

Stage 1 We minimize the *window-level* negative log-posterior, with respect to θ and $\{\sigma^2_{j,m}\}_{j,m}$. Specifically, we obtain maximum likelihood (ML) estimate $\hat{\theta}_{\text{ML}}$ and maximum a posteriori (MAP) estimate $\{\widehat{\sigma}^2_{j,m,\text{MAP}}\}_{j,m}$. We drop subscripts ML and MAP for notational simplicity.

Stage 2 We perform inference on the *sample-level* posterior. This includes the mean $\widehat{\mathbf{z}}_j = \mathbb{E}[\mathbf{z}_j | \{\widehat{\sigma}^2_{j,m}\}_{j,m}, \mathbf{y}, \hat{\theta}]$ and credible intervals (CI), or any statistical quantity of interest.

3.1. Stage 1: Optimization of $\{\sigma^2_{j,m}\}_{j,m}$ and θ

We have $p(\{\sigma^2_{j,m}\}_{j,m} \mid \mathbf{y}, \theta) \propto p(\mathbf{y} \mid \{\sigma^2_{j,m}\}_{j,m}, \theta) \times p(\{\sigma^2_{j,m}\}_{j,m} \mid \theta)$. Due to intractability, we minimize the *nonconvex* negative log-posterior. This is an *empirical Bayes* approach (Casella, 1985), since we estimate $\{\sigma^2_{j,m}\}_{j,m}$ with respect to the marginal likelihood $p(\mathbf{y} \mid \{\sigma^2_{j,m}\}_{j,m}, \theta)$. The smooth prior provides the MAP estimate for $\{\sigma^2_{j,m}\}_{j,m}$. We use the *Whittle likelihood* (Whittle, 1953), $\log p(\mathbf{y} \mid \{\sigma^2_{j,m}\}_{j,m}, \theta) = -\frac{1}{2}\sum_{m,n=1}^{M,N} \log(\gamma^{(m)}(\omega_n) + \sigma^2_\nu) + \frac{I^{(m)}(\omega_n)}{(\gamma(\omega_n)+\sigma^2_\nu)}$, in frequency domain. The log-prior is $\log p(\{\sigma^2_{j,m}\}_{j,m} \mid \theta) = -\frac{\lambda}{2}\sum_{j,m=1}^{J,M} \left(\log \sigma^2_{j,m} - \log \sigma^2_{j,m-1}\right)^2$. We optimize θ and $\{\sigma^2_{j,m}\}_{j,m}$ with block coordinate descent.

3.2. Stage 2: Inference for $p(\{\mathbf{z}_j\}_j \mid \{\sigma^2_{j,m}\}_{j,m}, \mathbf{y}, \theta)$

As the posterior is a Gaussian distribution, the mean trajectories $\{\widehat{\mathbf{z}}_j\}_j$ and the credible intervals can be computed analytically. These are computed with Kalman filter/smoother of computational complexity $O(J^2K)$, since Eq. 1 is a linear Gaussian state-space model. We can also obtain posterior samples for Monte Carlo (MC) inference. To generate the MC trajectory samples, we use the forward-filter backward-sampling (FFBS) algorithm (Carter & Kohn, 1994).

How does the complexity compare to that of the piecewise-stationary GP (GP-PS), the previous practice where the kernels are independently fit to each interval? This requires the inversion of $N \times N$ matrix in each interval, yielding the complexity of $O(MN^3) = O(N^2K)$. Considering that for typical biological time series analysis $N \sim 10^3$ and $J \sim 10$, PLSO is much more efficient than GP-PS.

4. Experiments

We apply PLSO to 1) a simulated dataset and 2) local-field potential (LFP) from the rat hippocampus. We use PLSO with λ chosen by cross-validation, λ_{CV}. As baselines, we use 1) regularized STFT (STFT-reg.) and 2) GP-PS. For GP-PS, we use $\{\widehat{\sigma}^2_{j,m}\}_{j,m}$ and $\widehat{\theta}$ from PLSO with λ_{CV}, and focus on the time-domain estimates.

4.1. Simulated dataset

We simulate $\mathbf{y}_k = 10\left(\frac{K-k}{K}\right)\mathbf{z}^{\Re}_{1,k} + 10\cos^4(2\pi\omega_0 k)\mathbf{z}^{\Re}_{2,k} + \nu_k$, for $1 \leq k \leq K$, where $\mathbf{z}_{1,k}$, $\mathbf{z}_{2,k}$ are stationary harmonic oscillators centered at $(\omega_1, \omega_2) = (1, 10)$ Hz and $l_1 = l_2 = 1$, with $\omega_0 = 0.04$ Hz, $f_s = 200$ Hz, $T = 100$

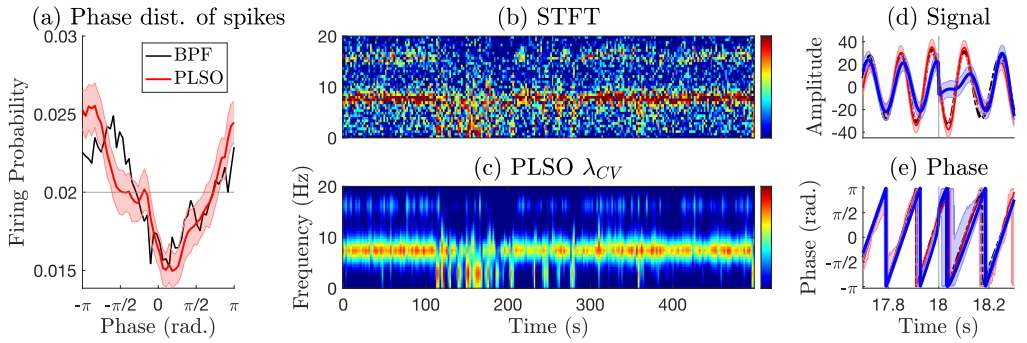

Figure 3. Hippocampal data. (a) Theta phase distribution of population neuron spikes, computed with bandpass-filtered LFP (black), PLSO estimate of $\hat{\mathbf{z}}_2$ with credible intervals from 200 posterior samples (red). Horizontal gray line indicates the uniform distribution. (b-c) Spectrogram (in dB) (b) STFT (c) PLSO with λ_{CV}. Learned frequencies are $(\widehat{\omega}_1, \widehat{\omega}_2, \widehat{\omega}_3) = (2.99, 7.62, 15.92)$ Hz, with $\widehat{\omega}_4 \sim \widehat{\omega}_5 > 25$ Hz. (d-e) Time-domain results. (d) Reconstructed signal (e) phase for $\widehat{\mathbf{z}}_2$ and interval boundary (vertical gray), with bandpass-filtered data (dotted black), STFT-reg. (blue), and PLSO (red). Shaded area represents 95% credible interval from 200 sample trajectories.

seconds, and $\nu_k \sim \mathcal{N}(0, 25)$. We simulate 20 realizations, assuming 2-second PS intervals. For PLSO, we use $J = 2$.

Results We use two metrics: 1) Mean squared error (MSE) between the mean estimate and the ground truth and 2) $\text{jump}(\mathbf{z}_j)$. The averaged results are shown in Table 1. We define $\text{jump}(\mathbf{z}_j) = \frac{1}{M-1}\sum_{m=1}^{M-1} |\hat{\mathbf{z}}_{j,mN+1} - \hat{\mathbf{z}}_{j,mN}|$ to be the level of discontinuity at the interval boundaries. If $\text{jump}(\mathbf{z}_j)$ greatly exceeds $\text{jump}(\mathbf{z}_j^{\text{True}})$, this implies the existence of large discontinuities. For GP-PS and STFT-reg., $\text{jump}(\mathbf{z}_j)$ exceeds $\text{jump}(\mathbf{z}_j^{\text{True}})$, indicating discontinuities at the boundaries. PLSO produces a similar jump metric as the ground truth, indicating the absence of discontinuities.

Table 1. Simulation results. For $\text{jump}(\mathbf{z}_j)$ and MSE, left/right metrics correspond to $\mathbf{z}_1/\mathbf{z}_2$, respectively.

	$\text{jump}(\mathbf{z}_j)$	MSE
TRUTH	0.95/12.11	0/0
$\lambda = \lambda_{\text{CV}}$	0.25/10.21	**2.88/3.91**
STFT-REG.	49.59/81.00	6.89/10.68
GP-PS	16.99/23.28	3.00/4.04

4.2. LFP data from the rat hippocampus

We use LFP collected from the rat hippocampus (Mizuseki et al., 2009), with $T = 1,600$ seconds and $f_s = 1,250$ Hz. The theta oscillation ($5 \sim 10$ Hz) is believed to be important for understanding the local neural circuit computation. We fit PLSO with $J = 5$, with 2-second PS interval. We bandpass-filter the data in the theta band as a reference.

Spike-phase coupling Fig. 3(a) shows the theta phase distribution of population neuron spikes. The PLSO-estimated distribution confirms the original results analyzed with bandpass-filtered signal (Mizuseki et al., 2009)–the hippocampal spikes show a strong preference for a specific phase, π for this dataset, of the theta band. Since PLSO provides posterior sample-trajectories, we can compute as many realizations of the phase distribution as desired. The resulting CI excludes the uniform distribution, suggesting the statistical significance of strong phase preference.

Time domain discontinuity Fig. 3(d-e) show that while the estimates from STFT-reg. (blue) and PLSO (red) follow the bandpass-filter (BPF) result closely, the STFT-reg. estimates exhibit discontiunity/distortion near the boundary. For further analysis, we also computed $\text{jump}(\phi_2)$ in degrees/sample for the phase. Considering that the theta band roughly progresses $2.16\,(= 7.5(\text{Hz}) \times 360/1250\,(\text{Hz}))$ degrees/sample, we observe that BPF (2.23), as expected, and PLSO (λ_{CV} : 2.40) are not affected by the boundary. This is not the case for STFT-reg. (26.83) and GP-PS (25.91).

Computational complexity We compare the runtime of PLSO and GP-PS for the mean trajectory estimation of varying lengths (50, 100, 200 seconds corresponding to $K = 6.25 \times 10^4$, 1.25×10^5, 2.5×10^5 samples), with the identical set of hyperparameters. As expected, the runtime of both methods are linear in K, as shown in table 2. More importantly, PLSO is much more efficient, indicating that PLSO is suitable for analyzing increasingly long time series.

Table 2. Runtime (s) for PLSO and GP-PS for varying length

	$T = 50$	$T = 100$	$T = 200$
PLSO	1.7	3.1	6.5
GP-PS	346.8	700.6	1334.0

5. Conclusion

We presented the **P**iecewise **L**ocally **S**tationary **O**scillatory (PLSO) framework to model nonstationary time series as the superposition of piecewise stationary (PS) oscillatory components. We show that PLSO is an efficient inference framework and strikes a balance between stochastic continuity of the data across PS intervals and piecewise stationarity.

References

Adak, S. Time-dependent spectral analysis of nonstationary time series. *Journal of the American Statistical Association*, 93(444):1488–1501, 1998.

Akaike, H. Likelihood of a model and information criteria. *Journal of Econometrics*, 16(1):3–14, 1981.

Carter, C. K. and Kohn, R. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

Casella, G. An introduction to empirical bayes data analysis. *The American Statistician*, 39(2):83–87, 1985.

Dahlhaus, R. Fitting time series models to nonstationary processes. *Ann. Statist.*, 25(1):1–37, 02 1997.

Gramacy, R. B. and Lee, H. K. H. Bayesian treed gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, 103(483): 1119–1130, 2008.

Kim, S.-E., Behr, M. K., Ba, D., and Brown, E. N. State-space multitaper time-frequency analysis. *Proceedings of the National Academy of Sciences*, 115(1):E5–E14, 2018.

Matsuda, T. and Komaki, F. Time series decomposition into oscillation components and phase estimation. *Neural Computation*, 29(2):332–367, 2017.

Mizuseki, K., Sirota, A., Pastalkova, E., and Buzsaki, G. Theta oscillations provide temporal windows for local circuit computation in the entorhinal-hippocampal loop. *Neuron*, 64:267–280, 2009.

Oppenheim, A. V., Schafer, R. W., and Buck, J. R. *Discrete-time Signal Processing (3rd Ed.)*. Prentice-Hall, Inc., 2009.

Paciorek, C. J. and Schervish, M. J. Nonstationary covariance functions for gaussian process regression. In *Proceedings of the 16th International Conference on Neural Information Processing Systems*, pp. 273–280, 2003.

Rasmussen, C. E. and Williams, C. K. I. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

Remes, S., Heinonen, M., and Kaski, S. Non-stationary spectral kernels. In *Advances in Neural Information Processing Systems 30*, pp. 4642–4651. 2017.

Solin, A. and Särkkä, S. Explicit Link Between Periodic Covariance Functions and State Space Models. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, volume 33 of *Proceedings of Machine Learning Research*, pp. 904–912. PMLR, 2014.

Solin, A., Hensman, J., and Turner, R. E. Infinite-horizon gaussian processes. In *Advances in Neural Information Processing Systems 31*, pp. 3486–3495. 2018.

Whittle, P. Estimation and information in stationary time series. *Ark. Mat.*, 2(5):423–434, 08 1953.

Wilkinson, W. J., Riis Andersen, M., Reiss, J. D., Stowell, D., and Solin, A. Unifying probabilistic models for time-frequency analysis. In *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3352–3356, May 2019.

Wilson, A. G. and Adams, R. P. Gaussian process kernels for pattern discovery and extrapolation. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, pp. III–1067—1075, 2013.

Appendix

References to the sections in the section headings are made with respect to the sections in the main text. Below is the table of contents for the Appendix.

A. Continuous model interpretation of PLSO

We can establish the equivalent continuous model of the PLSO, the connection which is important for explaining some of the mathematical properties. For a given interval, we use stochastic different equation

$$\frac{d\tilde{z}_j(t)}{dt} = \underbrace{\left(\left(-\frac{1}{l_j}\right) \oplus \begin{pmatrix} 0 & -\omega_j \\ \omega_j & 0 \end{pmatrix}\right)}_{\mathbf{F}} \tilde{z}_j(t) + \varepsilon(t), \quad (3)$$

where $\tilde{z}_j(t) : \mathbb{R} \to \mathbb{R}^2$, $\oplus$ denotes the Kronecker sum and $\varepsilon(t) \sim \mathcal{N}(0, \sigma_j^2 \mathbf{I}_{2\times 2})$. Discretizing the solution of Eq. 3 at Δ, such that $\widetilde{\mathbf{z}}_{j,k} = \tilde{z}_j(k\Delta)$, yields PLSO for a signle window. Consequently, we obtain the following for $\Delta > 0$

$$\begin{aligned} &\exp(\mathbf{F}\Delta) = \exp(-\Delta/l_j)\mathbf{R}(\omega_j), \\ &\sigma_j^2 \int_0^\Delta \exp(\mathbf{F}(\Delta-\tau)) \exp(\mathbf{F}(\Delta-\tau))^{\mathrm{T}} d\tau \\ &\quad = \sigma_j^2 (1 - \exp(-2\Delta/l_j)) \mathbf{I}_{2\times 2}. \end{aligned} \quad (4)$$

This interpretation extends to the nonstationary PLSO. The corresponding continuous model for $\widetilde{\mathbf{z}}_{j,mN+n}$ in Eq. 1 is the same as Eq. 3, with different variance $\mathbb{E}[\varepsilon_j(t)\varepsilon_j^{\mathrm{T}}(t)] = \sum_{m=1}^{M} \sigma_{j,m}^2 \cdot \mathbf{1}\left(\left(\frac{m-1}{M}\right) T \le t < \left(\frac{m}{M}\right) T\right) \mathbf{I}_{2\times 2}$, where $T = K\Delta$.

B. PSD for complex AR(1) process

We compute the steady-state covariance denoted as $\mathbf{P}_\infty^j$. For notational simplicity, we drop references to the interval m. Since we assume $\mathbf{P}_1^j = \sigma_j^2 \mathbf{I}_{2\times 2}$, it is easy to show that $\mathbf{P}_k^j$ is a diagonal matrix from $\mathbf{R}(\omega_j)\mathbf{R}^{\mathrm{T}}(\omega_j) = \mathbf{I}_{2\times 2}$. Denoting $\mathbf{P}_\infty^j = \alpha \mathbf{I}_{2\times 2}$, we use the discrete Lyapunov equation

$$\begin{aligned} &\mathbf{P}_\infty^j \\ &= \exp(-2\Delta/l_j)\mathbf{R}(\omega_j)\mathbf{P}_\infty^j \mathbf{R}^{\mathrm{T}}(\omega_j) \\ &\quad + \sigma_j^2 (1 - \exp(-2\Delta/l_j)) \mathbf{I}_{2\times 2} \\ &\Rightarrow \alpha = \exp(-2\Delta/l_j)\alpha + \sigma_j^2 (1 - \exp(-2\Delta/l_j)) \\ &\Rightarrow \mathbf{P}_\infty^j = \sigma_j^2 \mathbf{I}_{2\times 2}, \end{aligned} \quad (5)$$

which implies that under the assumption $\mathbf{P}_1^j = \sigma_j^2 \mathbf{I}_{2\times 2}$, we are guaranteed $\mathbf{P}_k^j = \sigma_j^2 \mathbf{I}_{2\times 2}$, $\forall k$. To compute the PSD of the stationary process $\mathbf{z}_j$, we now need to compute the autocovariance. Since only $\mathbf{z}_{j,k}^{\Re}$ contributes to $\mathbf{y}_k$, we compute the autocovariance of $\mathbb{E}[\mathbf{z}_{j,k}^{\Re} \mathbf{z}_{j,k+n'}^{\Re}]$ as

$$\begin{aligned} \mathbb{E}[\mathbf{z}_{j,k}^{\Re} \mathbf{z}_{j,k+n'}^{\Re}] &= \mathbb{E}[\mathbf{z}_{j,k}^{\Re} \cdot \Re(\rho_j^{n'} \exp(i\omega_j n')\mathbf{z}_{j,k})] \\ &= \rho_j^{n'} \mathbb{E}[\mathbf{z}_{j,k}^{\Re} \mathbf{z}_{j,k}^{\Re} \cos\omega_j n'] \\ &= \rho_j^{n'} \cos\omega_j n' \cdot \mathbb{E}[\{\mathbf{z}_{j,k}^{\Re}\}^2] \\ &= \rho_j^{n'} \sigma_j^2 \cos w_j n', \end{aligned} \quad (6)$$

where $\Re(\cdot)$ denotes the operator that extracts the real part of the complex argument and we used the fact that $\mathbb{E}[\mathbf{z}_{j,k}^{\Re} \mathbf{z}_{j,k}^{\Im}] = 0$. The spectra for the j^{th} component, $S_j(\omega)$ can be written as

$$\begin{aligned} S_j(\omega) &= \sum_{n'=-\infty}^{\infty} \mathbb{E}\left[\mathbf{z}_{j,k}^{\Re} \mathbf{z}_{j,k+n'}^{\Re}\right] \exp(-i\omega n') \\ &= \sum_{n'=-\infty}^{\infty} \rho_j^{n'} \sigma_j^2 \cos w_j n' \exp(-i\omega n') \\ &= \sigma_j^2 \sum_{n=-\infty}^{\infty} \rho_j^{n'} \{\exp(i\omega_j n) + \exp(-i\omega_j n')\} \exp(-i\omega n'). \end{aligned} \quad (7)$$

Unpacking the infinite sum for one of the terms,

$$\begin{aligned} &\sum_{n=-\infty}^{\infty} \rho_j^n \exp(-i(\omega-\omega_j)n') \\ &= 1 + \sum_{n=1}^{\infty} \rho_j^{n'} \exp(-i(\omega-\omega_j)n') + \rho_j^{n'} \exp(i(\omega-\omega_j)n') \\ &= 1 + \frac{\rho_j \exp(-i(\omega-\omega_j))}{1-\rho_j \exp(-i(\omega-\omega_j))} + \frac{\rho_j \exp(i(\omega-\omega_j))}{1-\rho_j \exp(i(\omega-\omega_j))} \\ &= 1 + \frac{2\rho_j \cos(\omega-\omega_j) - 2\rho_j^2}{(1-\rho_j \exp(-i(\omega-\omega_j)))(1-\rho_j \exp(i(\omega-\omega_j)))} \\ &= \frac{1-\rho_j^2}{1+\rho_j^2 - 2\rho_j \cos(\omega-\omega_j)}. \end{aligned} \quad (8)$$

Using the relation $\rho_j = \exp(-\Delta/l_j)$ and unpacking the infinite sum for the other term, we have

$$\begin{aligned} &S_j(w) \\ &= \frac{\sigma_j^2(1-\exp(-2\Delta/l_j))}{1+\exp(-2\Delta/l_j)-2\exp(-\Delta/l_j)\cos(\omega-\omega_j)} \\ &\quad + \frac{\sigma_j^2(1-\exp(-2\Delta/l_j))}{1+\exp(-2\Delta/l_j)-2\exp(-\Delta/l_j)\cos(\omega+\omega_j)}. \end{aligned}$$

Since Fourier transform is a linear operator, we can conclude that $\gamma(\omega) = \sigma_\nu^2 + \sum_{j=1}^J S_j(\omega)$.

C. Proof for Proposition 1

Proposition 1. For a given m, as $\Delta \to 0$, the samples on either side of the interval boundary, which are $\widetilde{\mathbf{z}}_{j,(m+1)N}$ and $\widetilde{\mathbf{z}}_{j,(m+1)N+1}$, converge to each other in mean square,

$$\lim_{\Delta\to 0} \mathbb{E}[\Delta\widetilde{\mathbf{z}}_{j,(m+1)N}\Delta\widetilde{\mathbf{z}}_{j,(m+1)N}^{\mathrm{T}}] = 0,$$

where we use $\Delta\widetilde{\mathbf{z}}_{j,(m+1)N} = \widetilde{\mathbf{z}}_{j,(m+1)N+1} - \widetilde{\mathbf{z}}_{j,(m+1)N}$.

Proof. To analyze Eq. 1 in the limit of $\Delta \to 0$, we use the equivalent continuous model (Eq. 4). It suffices to show that $\lim_{\Delta\to 0}\exp(\mathbf{F}\Delta) = \mathbf{I}_{2\times 2}$ and $\lim_{\Delta\to 0}\mathbb{E}[\varepsilon_{j,(m+1)N+1}\varepsilon_{j,(m+1)N+1}^{\mathrm{T}}] = \mathbf{0}$. We have,

$$\begin{aligned} &\lim_{\Delta\to 0}\exp(\mathbf{F}\Delta) = \mathbf{I}_{2\times 2} + \lim_{\Delta\to 0}\sum_{k=1}^{\infty}\frac{\Delta^k}{k!}\mathbf{F}^k = \mathbf{I}_{2\times 2} \\ &\lim_{\Delta\to 0}\mathbb{E}[\varepsilon_{j,(m+1)N+1}\varepsilon_{j,(m+1)N+1}^{\mathrm{T}}]/\sigma_{j,m+1}^2 \\ &= \lim_{\Delta\to 0}\int_0^\Delta \exp\left(\mathbf{F}(\Delta-\tau)\right)\exp\left(\mathbf{F}(\Delta-\tau)\right)^{\mathrm{T}} d\tau = \mathbf{0}. \end{aligned}$$

Since this implies $\lim_{\Delta\to 0}\mathbb{E}[\Delta\widetilde{\mathbf{z}}_{j,(m+1)N}\Delta\widetilde{\mathbf{z}}_{j,(m+1)N}^{\mathrm{T}}] = 0$, we have convergence in mean square. □

D. Proof for Proposition 2

Proposition 2. Assume $l_j \ll N\Delta$, such that $\mathbf{P}_{m,N}^j = \mathbf{P}_{m,\infty}^j$. In Eq. 1, the difference between $\mathbf{P}_{m,\infty}^j = \sigma_{j,m}^2\mathbf{I}_{2\times 2}$ and $\mathbf{P}_{m+1,\infty}^j = \sigma_{j,m+1}^2\mathbf{I}_{2\times 2}$ decays exponentially fast as a function of n, for $1 \le n \le N$,

$$\mathbf{P}_{m+1,n}^j = \mathbf{P}_{m+1,\infty}^j + \exp\left(-2n\Delta/l_j\right)(\mathbf{P}_{m,\infty}^j - \mathbf{P}_{m+1,\infty}^j). \tag{9}$$

Proof. We first obtain the steady-state covariance $\mathbf{P}_{m,\infty}^j$, similar to **Appendix B**. Since we assume $\mathbf{P}_{1,1}^j = \sigma_{j,1}^2\mathbf{I}_{2\times 2}$, we can show that $\forall m, n$, $\mathbf{P}_{m,n}^j$ is a diagonal matrix, noting that $\mathbf{R}(\omega_j)\mathbf{R}^{\mathrm{T}}(\omega_j) = \mathbf{I}_{2\times 2}$. Denoting $\mathbf{P}_{m,\infty}^j = \alpha\mathbf{I}_{2\times 2}$, we now use the discrete Lyapunov equation

$$\begin{aligned} \mathbf{P}_{m,\infty}^j &= \exp(-2\Delta/l_j)\mathbf{R}(\omega_j)\mathbf{P}_{m,\infty}^j\mathbf{R}^{\mathrm{T}}(\omega_j) \\ &\quad + \sigma_{j,m}^2\left(1-\exp\left(-2\Delta/l_j\right)\right)\mathbf{I}_{2\times 2} \\ &\Rightarrow \alpha = \exp(-2\Delta/l_j)\alpha + \sigma_{j,m}^2\left(1-\exp\left(-2\Delta/l_j\right)\right) \\ &\Rightarrow \mathbf{P}_{m,\infty}^j = \sigma_{j,m}^2\mathbf{I}_{2\times 2}. \end{aligned} \tag{10}$$

We now prove the proposition by induction. For fixed j and m, and for $n = 1$,

$$\begin{aligned} &\mathbf{P}_{m+1,1}^j \\ &= \exp(-2\Delta/l_j)\mathbf{R}(\omega_j)\mathbf{P}_{m,N}^j\mathbf{R}^{\mathrm{T}}(\omega_j) \\ &\quad + \sigma_{j,m+1}^2\left(1-\exp\left(-2\Delta/l_j\right)\right)\mathbf{I}_{2\times 2} \\ &= \left\{\sigma_{j,m+1}^2 + \exp\left(-2\Delta/l_j\right)\left(\sigma_{j,m}^2 - \sigma_{j,m+1}^2\right)\right\}\mathbf{I}_{2\times 2}. \end{aligned}$$

Assuming the same holds for $n = n' - 1$, we have for $n = n'$,

$$\begin{aligned} &\mathbf{P}_{m+1,n'}^j \\ &= \exp(-2\Delta/l_j)\mathbf{R}(\omega_j)\mathbf{P}_{m,n'-1}^j\mathbf{R}^{\mathrm{T}}(\omega_j) \\ &\quad + \sigma_{j,m+1}^2\left(1-\exp\left(-2\Delta/l_j\right)\right)\mathbf{I}_{2\times 2} \\ &= \exp(-2\Delta/l_j)\sigma_{j,m+1}^2\mathbf{I}_{2\times 2} \\ &\quad + \exp\left(-2n'\Delta/l_j\right)\left(\sigma_{j,m}^2 - \sigma_{j,m+1}^2\right)\mathbf{I}_{2\times 2} \\ &\quad + \sigma_{j,m+1}^2\left(1-\exp\left(-2\Delta/l_j\right)\right)\mathbf{I}_{2\times 2} \\ &= \left\{\sigma_{j,m+1}^2 + \exp\left(-2n'\Delta/l_j\right)\left(\sigma_{j,m}^2 - \sigma_{j,m+1}^2\right)\right\}\mathbf{I}_{2\times 2}. \end{aligned}$$

By the principle of induction, Eq. 9 holds for $1 \le n \le N$. □

E. Initialization & Estimation for PLSO

E.1 Initialization

As noted in the main text, we use AIC to determine the optimal number of J. For a given number of components J, we first construct the spectrogram of the data using STFT and identify the frequency bands with prominent power, i.e., frequency bands whose average power exceeds pre-determined threshold. The center frequencies of these bands serve as the initial center frequencies $\{\omega_j^{\text{init}}\}_j$, which are either fixed throughout the algorithm or further refined through the estimation algorithm in the main text. If J exceeds the number of identified frequency bands from the spectrogram, 1) we first place $\{\omega_j^{\text{init}}\}_j$ in the prominent frequency bands and 2) we then place the remaining components uniformly spread out in $[0, \omega_c]$, where ω_c is a cutoff frequency to be further determined in the next section. As for $\{l_j^{\text{init}}\}_j$, we set it to be a certain fraction of the corresponding $\{\omega_j^{\text{init}}\}_j$. We then fit $\{\sigma_{j,m}^2\}_{j,m}$ and θ with $\lambda = 0$, through the procedure explained in Stage 1. We finally use these estimates as initial values for other values of λ.

E.2 Estimation for σ_ν^2

There are two possible ways to estimate the observation noise variance σ_ν^2. The first approach is to perform maximum likelihood estimation of $f(\{\sigma_{j,m}^2\}_{j,m}; \theta)$ with respect to σ_ν^2. The second approach, which we found to work *better* in practice and use throughout the manuscript, is to directly estimate it from the Fourier transform of the data. Given a cutoff frequency ω_c, informed by domain knowledge, we take the average power of the Fourier transform of $\mathbf{y}$ in $[\omega_c, f_s/2]$. For instance, it is widely known that the spectral content below 40 Hz in anesthesia EEG dataset is physiologically relevant and we use $\omega_c \simeq 40$ Hz.

F. Computation of phase

One quantity of interest is the *phase*. We obtain the phase as $\phi_{j,k} = \tan^{-1}(\mathbf{z}_{j,k}^{\Im}/\mathbf{z}_{j,k}^{\Re})$. Since $\tan^{-1}(\cdot)$ is a non-linear operation, we compute the mean and credible interval with MC samples through the FFBS algorithm. Given the posterior-sampled trajectories $\{\mathbf{z}_j^{(s)}\}_{j,s}$, where $s \in \{1, \ldots, S\}$ denotes MC sample index, we estimate $\widehat{\phi}_{j,k} = (1/S)\sum_{s=1}^{S} \tan^{-1}(\mathbf{z}_{j,k}^{\Im,(s)}/\mathbf{z}_{j,k}^{\Re,(s)})$, and use empirical quantiles for the associated credible interval.