Base Kernels	Encoding Function	$k(x, x')$
White Noise (WN)	Uncorrelated noise	$\sigma^2 \delta_(x, x')$
Constant (C)	Constant functions	σ^2
Linear (LIN)	Linear functions	$\sigma^2 (x - l)(x - l')$
Squared Exponential (SE)	Smooth functions	$\sigma^2 \exp(-\frac{(x-x')^2}{2l^2})$
Periodic (PER)	Periodic functions	$\sigma^2 \frac{\exp(\frac{\cos \frac{2\pi(x-x')}{p}}{l^2}) - I_0(\frac{1}{l^2})}{\exp(\frac{1}{l^2}) - I_0(\frac{1}{l^2})}$

Table 2: List of base kernels

A. Base kernels and search grammar in the ABCD framework

Table 2 contains base kernels described in (Lloyd et al., 2014).

The language of models (or kernels) is presented by a set of rules in the grammar:

$$\mathcal{S} \to \mathcal{S} + \mathcal{B} \qquad \qquad \mathcal{S} \to \mathcal{S} \times \mathcal{B}$$
$$\mathcal{S} \to \mathcal{B}$$

where $\mathcal{S}$ represents any kernel subexpression, $\mathcal{B}$ and $\mathcal{B}'$ are base kernels (Lloyd et al., 2014).

B. Compute $\mathbb{E}[\log p(\mathbf{Z})]$

In Doshi et al. (2009), the variational inference approximating $\mathbf{Z}$ considered two approaches: finite variational approach and infinite variational approach. We will take a brief review of estimating $\mathbb{E}[\log p(\mathbf{Z})]$. Readers may refer to (Doshi et al., 2009) to have more details. In the finite variational approach, sampling $\mathbf{Z}$ involves

$$\begin{aligned} \pi_k \sim & \text{Beta}(\alpha/K, 1), \\ z_{nk} \sim & \text{Bernoulli}(\pi_k). \end{aligned}$$

Here the generative procedure involves an additional random variable $\boldsymbol{\pi}$ which is omitted in the main text for simplicity. The variational inference requires to approximate the posterior distribution over $\boldsymbol{\pi}$ by $\prod_k q(\pi_k)$. Specifically, each $q(\pi_k)$ follows a Beta distribution Beta(τ_{k_1}, τ_{k_2}). Since $\mathbf{X}$ and $\boldsymbol{\pi}$ are conditionally independent given $\mathbf{Z}$, $\mathbb{E}[\log p(\mathbf{X}|\mathbf{Z})]$ discussed in the main text is independent to $\boldsymbol{\pi}$. We can compute $\mathbb{E}[\log p(\mathbf{Z})]$ as

$$\begin{aligned} &\mathbb{E}[\log p(\mathbf{Z})] \\ =&\sum_{k=1}^{K} \left[\log \frac{\alpha}{K} + \left(\frac{\alpha}{K} - 1\right)(\psi(\tau_{k_1}) - \psi(\tau_{k_1} + \tau_{k_2}))\right] + \sum_{k=1}^{K}\sum_{n=1}^{N} [\nu_{nk}\psi(\tau_{k_1}) + (1-\nu_{nk})\psi(\tau_{k_2}) - \psi(\tau_{k_1} + \tau_{k_2})], \end{aligned}$$

where $\psi(\cdot)$ is the digamma function.

While in the finite variational approach, stick breaking construction (Teh et al., 2007) is used to sample $\mathbf{Z}$ as

$$\begin{aligned} v_k \sim & \text{Beta}(\alpha, 1), \\ \pi_k = & \prod_{i=0}^{K} v_i, \\ z_{nk} \sim & \text{Bernoulli}(\pi_k), \end{aligned}$$

with $k = 1 \ldots \infty$. Similarly, the variational distribution $q(\boldsymbol{v})$ is proposed to approximate $p(\boldsymbol{v})$ by independent Beta(τ_{k_1}, τ_{k_2})s

$$\begin{aligned} \mathbb{E}[\log p(\mathbf{Z})] = & \sum_{k=1}^{K} [\log \alpha + (\alpha - 1)(\psi(\tau_{k_1}) - \psi(\tau_{k_1} + \tau_{k_2}))] \\ & + \sum_{k=1}^{K}\sum_{n=1}^{N} \left[\nu_{nk}\left(\sum_{m=1}^{k} \psi(\tau_{m_1}) - \psi(\tau_{m_1} + \tau_{m_2})\right) + (1-\nu_{nk})\mathbb{E}_{\boldsymbol{v}}\left[\log(1 - \prod_{m=1}^{k} v_m)\right]\right], \end{aligned}$$

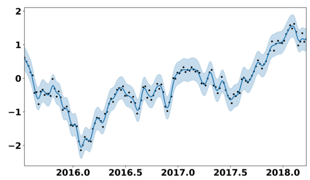
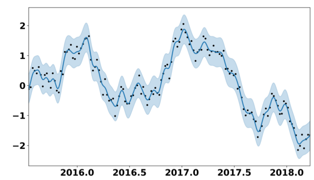
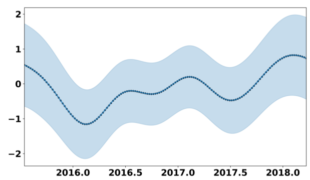
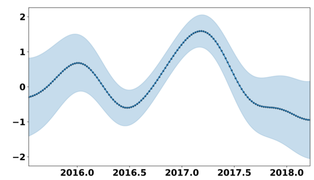
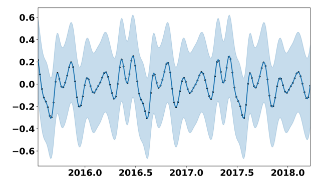
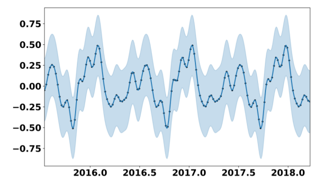
Description	Oil	USD Index
Plot of posterior mean and variance		
•This component is a smooth function with a typical lengthscale of 3.3 months		
•This component is periodic with a period of 1.0 years.The shape of this function within each period has a typical lengthscale of 7.2 days		

Figure 8: Comparing Oil and USD index. This is extracted from the pairwise comparison of GONU data set.

with $\mathbb{E}_{\boldsymbol{v}}\left[\log(1-\prod_{m=1}^{k} v_m)\right]$ is further approximated by Taylor expansion.

C. Comparison of search space in PE with LKM and CKL

We emphasize that PSE with LKM considers a larger number of kernel structures than those in CKL. Suppose that CKL and our search algorithm have the same found structure at a depth d. While the CKL's structure is $\mathcal{S}_d = \mathcal{S}_d^{(1)} + \cdots + \mathcal{S}_d^{(K)}$, PSE represents as a set $\{\mathcal{S}_d^{(1)}, \dots, \mathcal{S}_d^{(K)}\}$. Let us examine the cardinality of kernel spaces after performing an expansion to the next depth. The procedure is to extract substructures from the current structure, then apply grammar rules on the structure. In CKL, substructures consist of all structures generated from the combinations of $\mathcal{B}_d^{(k_l)}$ in each individual $\mathcal{S}_d^{(k)}$ and ones generated by the combination of all $\mathcal{S}_d^{(k)}$. The former has $O(K\sum_l \binom{L}{l}) = O(K2^L)$ substructures where L is the largest number of base kernels in $\mathcal{S}_d^{(k)}$. The latter creates $O(\sum_k \binom{K}{k}) = O(2^K)$ combinations. When the maximum number of grammar rules per substructure is R, the total number of candidates at the depth $d+1$ is $O(RK2^L + R2^K)$.

Our approach only applies expansion on individual structure $\mathcal{S}_d^{(k)}$ via the combinations of $\mathcal{B}_d^{(k_l)}$. However, the search space still includes all the cases when substructures are extracted from a combination of $\mathcal{S}_d^{(k)}$. For instance, the generation from LIN+PER+SE to (LIN+PER)×SE+SE in CKL is equivalent to the generation from {LIN, PER, SE} to {LIN×SE, PER×SE, SE} in our approach. For the case of PE, the additive kernel set will be expanded into a new one having the number of elements $R2^L + K$. With the flexible binary indications (on/off) of $\mathbf{Z}$, the number of all possible kernels is $O(K2^{R2^L+K})$ when all structures are visited to be expanded.

D. Pairwise comparison between Oil and USD index

As we discussed in the main text, our model can recognize the inverse correlation by looking at the first component in Figure 8. The second component in Figure 8 is another example of the posterior $\boldsymbol{f}_k|\mathbf{x}_n$ realized differently given different time series. This observation is found in a real-world data set.

E. Full output of seizure data set

Figure 9 describes the output of our model.

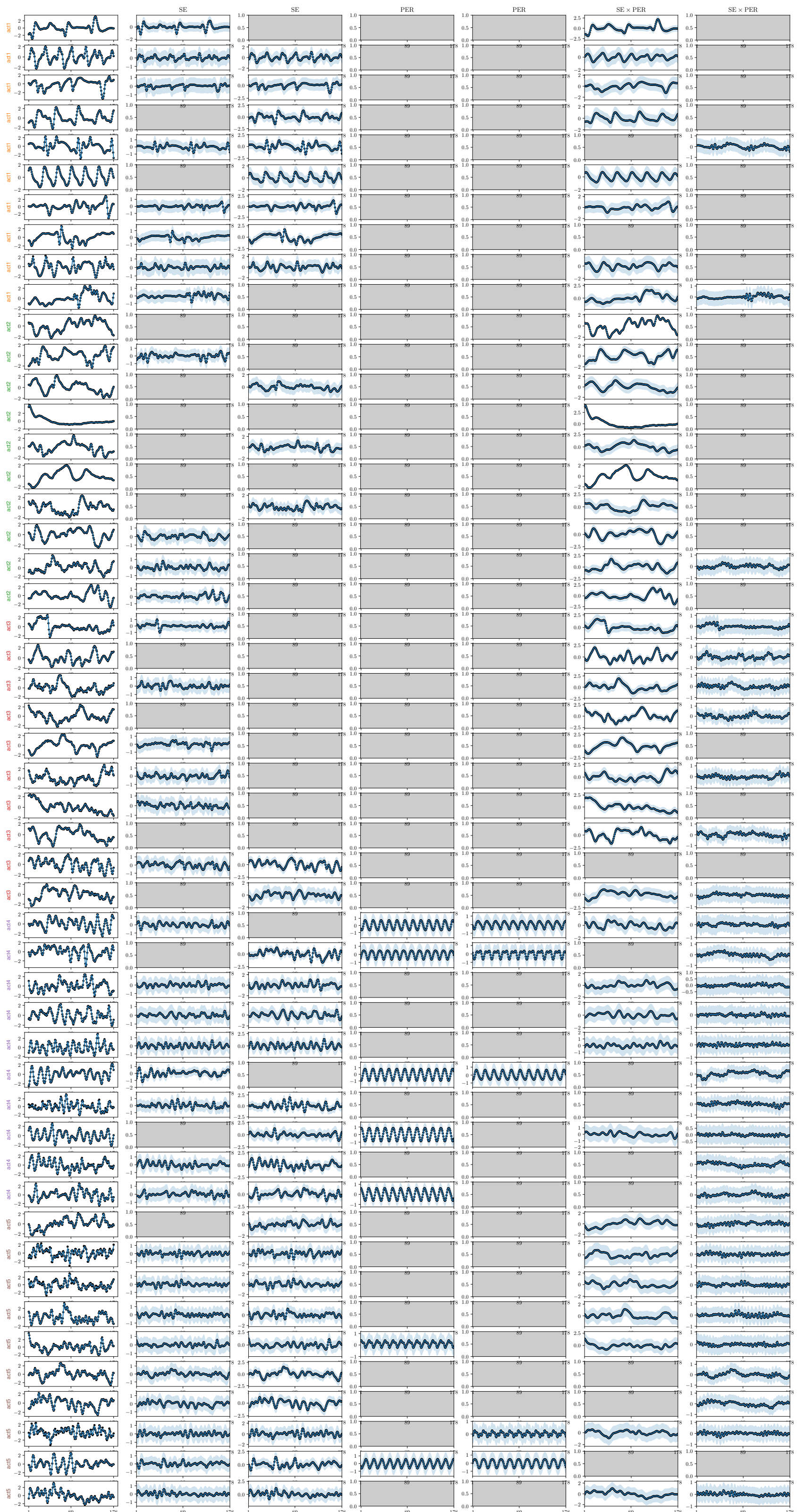


Figure 9: Full output of seizure data set. First column is the posterior plot of each time series. The remaining columns are decomposed components. Missing plot indicate there is no component w.r.t the corresponding time series.