Automatic Decomposition of Self-Triggering Kernels of Hawkes Processes

Rafael Lima    Jaesik Choi
Ulsan National Institute of Science and Technology, Ulsan, 44919, Korea
{rafael_glima,jaesik}@unist.ac.kr

July 19, 2017

Abstract

Hawkes Processes capture self- and mutual-excitation between events when the arrival of one event makes future ones more likely to happen in time-series data. Identification of the temporal covariance kernel can reveal the underlying structure to better predict future events. In this paper, we present a new framework to represent time-series events with a composition of self-triggering kernels of Hawkes Processes. Our automatic decomposition procedure is composed of three main steps: (1) discretized kernel estimation through frequency domain inversion equation associated with the covariance density when a direct parameter optimization is not possible, (2) greedy kernel decomposition through four base kernels and their combinations (addition and multiplication), and (3) automated report generation. In addition, we report the first multiplicative kernel compositions along with stationarity conditions for Hawkes Processes. We demonstrate that the new automatic kernel decomposition procedure performs better to predict future events than the existing framework in real-world data.

1 Introduction

Point Processes \cite{4} have been used for modeling time-events series data. Hawkes Processes (HP) \cite{9} are point processes for modeling self-exciting behavior, i.e., when the arrival of one event makes future events more likely to happen. This type of behavior has been observed in various domains, such as earthquakes, financial markets, web traffic patterns, crime rates \cite{14,16} and social media \cite{22}.

As an example, in high-frequency finance, buyers and sellers of stocks demonstrate herding behavior \cite{7,2}. After the main earthquake, several aftershocks follow according to a time-clustered pattern \cite{18}. In web data, hyperlink proliferation
across pages exhibit self- and mutual-excitation [8]. In criminology, gang-related retaliatory crime patterns are grouped in time [14]. In social media, the ‘infectiousness’ of posts can be shown to be modeled through a self- and mutual-excitement assumption [22].

In Hawkes Processes analysis, some parametric kernels capture intra-domain typical behaviors, e.g., quick time-decaying exponential excitation in the case of finance and web data [1] 3; slower power-law decay in earthquake-related data [18]; and periodicity-inducing sinusoidal kernel in TV-watching data [21].

Choosing a proper kernel type may significantly contribute to learning the model parameters and better predicting future events. Kernel parameters may be fitted in the data through the gradient descent method over a likelihood function penalized by a regularization criterion (e.g., Akaike Information Criterion) on the number of parameters [19]. Another method of kernel estimation is through the use of the power spectrum of the second order statistics of the process: covariance density and normalized covariance [9]. These are well defined when the self-triggering function induces what is called stationary behaviour.

In this paper, we present an automatic kernel selection and learning algorithm for Hawkes Processes. Given four types of base kernels (EXP,PWL,SQR and SNS), our algorithm finds the best fitting kernel. For handling composite types of events, we develop a new kernel decomposition method which represents the composition (sum and product) of different kernels. For verifying the stationarity property of each composite kernel, we also derived analytical expressions for the stationarity conditions. To our best knowledge, this is the first multi-class kernel decomposition framework for HPs.

The main steps of the automatic framework, which will be thoroughly explained in the following sections, are then: discretized kernel estimation and greedy kernel decomposition. An automatic framework for extracting relevant typical features of self-excitement from raw data would likely make the Hawkes Processes analysis faster and conveniently expressed through a human-readable report. Thus, the automated generation of such a report is also discussed.

Automatic analysis frameworks for Gaussian Processes (GPs) are proposed in [5] and [10]. However, due the fundamental distinctions between GPs and HPs (such as stationarity conditions and causality assumptions for the latter), the techniques proposed for GPs can not be extended to HPs in a straightforward manner. We include additional references in the appendix.
2 Hawkes Processes

A point process representation of a sequence of n time-events is expressed by a vector of the form \((t_1, t_2, \ldots, t_n)\). Treating the real line as a time axis, the vector can be intuitively associated with a so-called Counting Process \(N(t)\), s.t.:

- \(dN(t) = 1\), if there is an event at time \(t\);
- \(dN(t) = 0\), otherwise.

An example of a Counting Process is shown in Figure 1.

A point process can be described through its intensity function \(\lambda(t)\), which can be understood as the instantaneous expected rate of arrival of events, or the expectation of derivative of the Counting Process \(N(t)\):

\[
\lambda(t) = \lim_{h \to 0} \frac{\mathbb{E}[N(t+h) - N(t)]}{h}
\]

This intensity function, if existing, uniquely characterizes the finite-dimensional distributions of the point process [4]. A simple example of this function would be the constant mean rate of arrival, \(\mu\), in the case of a homogeneous Poisson process.

Hawkes processes model the intensity function in terms of the self-excitement: when the arrival of an event makes subsequent arrivals more likely to happen [11]; and can be described through the following conditional intensity function \(\lambda(t)\):

\[
\lambda(t) = \lim_{h \to 0} \frac{\mathbb{E}[N(t+h) - N(t)|\mathcal{H}(t))]}{h} = \mu + \int_{-\infty}^{t} \phi(t-u)dN(u),
\]

where:

- \(\mathcal{H}(t)\) is the history of the process, the set containing all the events up to time \(t\);
- \(\mu\) is called background rate, or exogenous intensity, which is fixed as the mean of a Homogeneous Poisson Process;
- \(\phi(t)\) is denominated self-triggering kernel, or excitation function;

From this function, one may notice that the intensity at time \(t\) will likely be affected by events which happened before the time \(t\), described by the history of the process.
From \cite{9}, we have that, if:

\[ ||\phi|| = \int_0^\infty \phi(t)dt \leq 1, \]

(3)

then the corresponding process will reach wide-sense stationary behavior, from which the asymptotic steady arrival rate:

\[ \Lambda = \frac{\mu}{(1 - ||\phi||)}, \]

(4)

can be obtained, along with its covariance function, which is independent of t:

\[ \nu(\tau) = E(dN(t)dN(t + \tau)). \]

(5)

Estimating \( \Lambda \) and \( \nu(\tau) \), also referred to as first- and second-order statistics, respectively, requires wide-sense stationarity assumptions which, besides being analytically convenient, are also connected to the fact that, in real data, the chain of self-excitedly induced further events will always be of finite type, or without ‘blowing up’, what corroborates the practicality of the estimated model.

2.1 Self-Exciting Kernels

From the definition of the conditional intensity function, the self-excitement of the process is expressed through the kernel function \( \phi(t) \). For the kernel decomposition,
Figure 2: An example of an Intensity Function of a self-triggering Point Process $\lambda(t)$.

Four base kernels will be used for identifying and estimating typical triggering behaviors:

- **EXP($\alpha$, $\beta$):** The decay exponential kernel is parameterized by the amplitude $\alpha$ and decay rate $\beta$, and is useful for modeling quick influence decay, such as in finance or web data [8], in which initial transactions/hyperlinks have a lot of impact initially but gradually reduce their influence over time:
  \[ \text{EXP}(\alpha, \beta) = \alpha e^{-\beta t} \]  
  \[ (6) \]

- **PWL($K,c,p$):** The power law kernel is parameterized by the amplitude $K$, the exponent $p$, and the constant $c$, such as in Equation (7), and has been prevalent in earthquake [18] and social media-related data [22], modeling a slower decaying trend than the exponential:
  \[ \text{PWL}(K, c, p) = \frac{K}{(t + c)^p} \]  
  \[ (7) \]

- **SQR($B,L$):** The pulse kernel is described by the amplitude $B$ and the length $L$. Being a trivial, steady, self-exciting dynamics on its own, it may also work as an offset level for the combined triggering with other kernel types, in the case of addition, and as a horizon truncation, in the case of multiplication $^1$:
  \[ \text{SQR}(B, L) = B(u(t) - u(t - L)) \]  
  \[ (8) \]

$^1u(t)$ is the step function
• **SNS(\(A, \omega\)):** A truncated sinusoidal, parameterized by the amplitude \(A\) and the angular velocity \(\omega\). This type of kernel base function captures well the self-excitement of periodic events, such as TV watching-related data (IPTV) [21], in which watching one episode of a TV program makes the viewer more likely to watch further ones. Since these shows are usually broadcasted weekly, the TV-watching behavior will likely demonstrate the weekly self-excitement. In addition, according to [14], homicide rates show a pronounced seasonal effect, peaking in the summer and tapering in the winter:

\[
\text{SNS}(A, \omega) = A \sin(\omega t),
\]

for \(t \in [0, \frac{\pi}{\omega}]\), and 0, otherwise.

![Figure 3: The four base kernel types.](image)

| Type              | Equation                                      |
|-------------------|-----------------------------------------------|
| Exponential (EXP(\(\alpha, \beta\))) | \(\alpha e^{-\beta t}\)                     |
| Power-Law (PWL(K,c,p)) | \(K \frac{c}{(c+t)^p}, \ (p > 1)\)          |
| Pulse (SQR(B,L))   | \(B(u(t) - u(t-\text{L}))\)                 |
| Sinusoidal (SNS(A,\(\omega\))) | \(A \sin(\omega t), \ t \in [0, \frac{\pi}{\omega}]\) |

Table 1: Four base kernels


3 Automatic Kernel Decomposition for HPs

This section presents two steps for the automatic kernel identification. The steps are (1) nonparametric kernel estimation from discretized covariance and (2) parametric kernel search through our new kernel decomposition scheme.

Here, the discretized kernel estimation is optional when a direct optimization of kernel structure is possible. Unfortunately, discontinuous functions (SQR, SNS) do not allow such optimization (e.g., gradient descent). In this paper, we use the discretized kernel estimation as an unified method for both continuous (EXP, PWL), discontinuous kernels (SQR, SNS) and, most importantly, their combinations.

Furthermore, another great advantage of this step, compared with traditional sequential methods, is the fact that the value of $\nu$ for each value of $\tau$ can be calculated independently, while, in gradient descent, the value of the parameters at step $t$ must be obtained before the values for step $t + 1$. When combined with parallelization of the loops, in our case, this step greatly improved the speed of obtainment of the most likely parametric representations of the sample processes.

3.1 Discretized Kernel Estimation

This step is fully described in [1], and basically consists of building an estimator of $\phi(t)$ from empirical measurements of $\nu(\tau)$, the stationary covariance.

Given a finite sequence of ordered time-events in $[0, T]$, we fix a window size of $h$, and estimate $\nu(\tau)$ as:

$$\nu^{(h)}(\tau) = \frac{1}{h}E\left((\int_{0}^{h} dN_{s} - \Lambda h)(\int_{\tau}^{\tau+h} dN_{s} - \Lambda h)\right)$$

(10)

In practice, this estimation is done in discrete time steps $\delta$:

$$\nu^{(h)}(\tau, \delta) = \frac{1}{T}\sum_{i=1}^{\lfloor T/\delta \rfloor} (dN_{i\delta}^{(h)} - dN_{(i-1)\delta}^{(h)})(dN_{i\delta+\tau}^{(h)} - dN_{(i-1)\delta+\tau}^{(h)}),$$

(11)

where $dN_{i\delta}^{(h)}$ is the total number of events happening between $t = i\delta$ and $t = i\delta + h$.

From Theorem 1 in [1], we have that, given $g_{l}^{(h)} = (1 - \frac{|h|}{T})^{+}$, i.e., a triangular kernel density estimator with bandwidth “$h$”, we have the following relation in Laplace domain:

$$\hat{\nu}_{z}^{(h)} = \hat{g}_{z}^{(h)}(1 + \hat{\psi}_{z}^{*})\Lambda(1 + \hat{\psi}_{z}^{*})^{\dagger},$$

(12)
Figure 4: Discretized covariance estimate from a sequence generated with (a) EXP, (b) SQR, and (c) SNS kernels.

where\(^2\):

\[ \hat{\psi}_z = \sum_{n=1}^{+\infty} \hat{\phi}_n^2 = \frac{\hat{\phi}_z}{(1 - \hat{\phi}_z)}. \]  

(13)

Working with the Fourier transform restriction, i.e., \((z = i\omega, \text{ with } \omega \in \mathbb{R})\) and given that:

\[ \hat{g}_{i\omega}^{(h)} = \frac{4}{\omega^2 h} \sin^2 \left(\frac{\omega h}{2}\right), \]  

(14)

we get to:

\[ (1 + \hat{\psi}_z^*) \Lambda (1 + \hat{\psi}_z^*)^\dagger = \frac{\hat{\nu}_z^{(h)}}{\hat{g}_z^{(h)}}, \]  

(15)

where we fix \(h = \delta\), so we do not bother with cancellations of \(\hat{g}_z^{(h)}\). And then, from:

\[ |1 + \hat{\psi}_i\omega|^2 = \frac{\hat{\nu}_z^{(h)}}{\Lambda \hat{g}_z^{(h)}}, \]  

(16)

we get to the discretized estimation of \(\hat{\phi}_t\) by taking the inverse Fourier transform of:

\[ \hat{\phi}_{i\omega} = 1 - e^{-\log |1 + \hat{\psi}_i\omega| + iH(\log |1 + \hat{\psi}_i\omega|)}, \]  

(17)

in which the operator \(H(\cdot)\) refers to the Hilbert transform.

### 3.2 Greedy Kernel Decomposition

For expressing the discretized estimation in terms of the four base kernels, the following steps are executed:

\(^2\)Given a function \(f_t\), \(\hat{f}_z\) is its Laplace Transform, and the “\(^*\)” symbol corresponds to its conjugate
Figure 5: A diagram for the Kernel Decomposition Algorithm.

Table 2: Stationarity conditions of the four base kernels.

| EXP(α,β) | α/β | PWL(K,c,p) | K^{1-p}/p - 1, (p > 1) | SQR(B,L) | BL | SNS(A,ω) | 2A/ω |
|-----------|-----|------------|-----------------|---------|----|----------|------|

1. Calculate residues (L^1-error) w.r.t the four basic kernels \{EXP, PWL, SQR, SNS\};

2. Check whether the estimated parameters of the kernels satisfy the stationarity condition, by using the closed-form expressions from table 2;

3. Among those estimated kernels which satisfy the said condition, select the kernel with the minimum residue MR_1 (denominated K_1);

4. Calculate residues w.r.t. a total of 8 kernel expansions, resulting from 2 operations (addition and multiplication) per base kernel \{+EXP, *EXP, +PWL, *PWL, +SQR, *SQR, +SNS, *SNS\}, while fixing the optimized parameters for K_1, in the case of Additive Combination, and recalculating all the parameters, in the case of Multiplicative Combination;

5. Among those estimated combinations which satisfy the spectral radius condition (calculated in closed form from Table 3), select the kernel with minimum residue MR_2, denominated K_2;

6. If MR_1 < MR_2/η (η would act as a regularization parameter), pick K_1. Else, pick K_2.

The pseudocode of our algorithm is included in Algorithm 1.
Algorithm 1 Automatic Decomposition of HP Kernels

1: \textit{input} = \textit{k}_{est}
2: \textit{fit}_1 = \text{fit}(\textit{k}_{est}; \emptyset, \{\text{EXP, PWL, SQR, SNS}\})
3: \textit{K}_1 = \text{index_of_kernel}(\text{min_residue}(%fit_1))
4: \text{MR}_1 = \text{min_residue}(%fit_1)
5: \textit{fit}_2 = \text{fit}(\textit{k}_{est}; \textit{K}_1, \{+\text{EXP}, *\text{EXP}, +\text{PWL}, *\text{PWL},
\text{+SQR}, *\text{SQR}, +\text{SNS}, *\text{SNS}\})
6: \text{MR}_2 = \text{min_residue}(%fit_2)
7: \textit{K}_2 = \text{index}(\text{min_residue}(%fit_2))
8: \textit{output} = \text{None}
9: \textbf{if} ||\phi_{K_1}|| < 1 \textbf{then}
10: \textit{output} = \textit{K}_1
11: \textbf{if} ||\phi_{K_2}|| < 1 \textbf{then}
12: \textbf{if} \textit{output} \neq \text{None} \textbf{then}
13: \textbf{if} \text{MR}_1 \geq \frac{1}{\eta} \text{MR}_2 \textbf{then}
14: \textit{output} = \textit{K}_2
15: \textbf{else}
16: \textit{output} = \textit{K}_2

| Base Kernel | Base Kernel | Condition |
|-------------|-------------|-----------|
| EXP(\alpha,\beta) | EXP(\alpha,\beta) | \alpha_1\alpha_2/(\beta_1 + \beta_2) (closed under multiplication) |
| EXP(\alpha,\beta) | PWL(K,c,p) | \alpha K^{\beta-1} e^{\beta c} \Gamma(1-p, \beta c) |
| EXP(\alpha,\beta) | SQR(B,L) | (KB(1-e^{-\beta L})) / \beta |
| EXP(\alpha,\beta) | SNS(A,\omega) | (A_{1\omega}(1+e^{-\beta})) / (\omega^2 + \beta^2) |
| PWL(K_1,c_1,p_1) | PWL(K_2,c_2,p_2) | (K_1 K_2) / ((c_1 + p_2 - 1) \min(c_1,c_2)^{p_1+p_2-1}) (upper bound) |
| PWL(K_1,c_1,p_1) | SQR(B,L) | (KB(1-e^{-\beta L})) / (p-1) |
| PWL(K_1,c_1,p_1) | SNS(A,\omega) | \leq KA((c + \frac{p}{2})^{1-p} - c^{1-p}) / (1-p) (upper bound) |
| SQR(B,L) | SQR(B,L) | BL |
| SQR(B,L) | SNS(A,\omega) | 2AB/\omega |
| SNS(A,\omega) | SNS(A,\omega) | \pi A / (2\omega) |

Table 3: Stationarity Condition for Multiplicative Combination of the four Base Kernels.

3.3 Stationarity Conditions

As a quick way of evaluating the validity of the estimated models regarding the stationarity criterion, we developed closed-form expressions, either in the form of equality or as an upper bound, which are shown in Table 2 for the case of a single kernel, and Table 3 for multiplicative combinations of two kernels. The conditions for additive combinations can be derived from the conditions for single

\[ \Gamma(a, y) = \int_y^\infty e^{-t}e^{-t}dt \]
kernels in a straightforward manner.

The kernel is said to induce stationarity if the result of the expression calculated using the estimated parameters belongs to the interval \([0,1)\). This can be justified both from the point-of-view of Hawkes Process as a branching process, also called *immigrant-birth representation* [15], and of the boundedness of the spectral radius (largest absolute value among the eigenvalues) of the excitation matrix. \(^4\)

4 Analysis of Higher-order Kernel Decomposition

A sequential additive decomposition of the discretized estimation vector is rather straightforward, since one may just set the residual vector from the previous stages as the input of the next ones.

In the case of multiplicative decomposition, it is nontrivial to find the result of intraclass decomposition. To the best of our knowledge, no analysis on multiplicative HP kernel decomposition is reported yet.

In this paper, we provide a new upper bound over an interclass kernel product of unknown degree, as in:

\[
[\text{EXP}]^{k_1} \times [\text{PWL}]^{k_2} \times [\text{SQR}]^{k_3} \times [\text{SNS}]^{k_4}
\]

for \(k_i \in \mathbb{Z}^*\), where the operator “\([\cdot]^{k}\)” corresponds to the set of functions which can be decomposed into a \(k\)-th order product of kernels, e.g:

\[
[\text{EXP}]^k = \underbrace{\alpha_1 e^{-\beta_1 x} \cdot \alpha_2 e^{-\beta_2 x} \cdot \ldots \cdot \alpha_k e^{-\beta_k x}}_{k \text{ terms}}.
\]

By deriving the four possible intraclass kernel products, one may observe that the typical self-exciting behavior features of each kernel type are preserved, as in the following:

4.1 EXP

\([\text{EXP}]^{k_1}\) reduces to the case of a single exponential with \(\alpha = \prod_{i=1}^{k_1} \alpha_i\) and \(\beta = \sum_{i=1}^{k_1} \beta_i\), thus still accounting for its ‘quick-decay’ behavior:

\[
[\text{EXP}]^{k_1} \subset [\text{EXP}] ;
\]

\(^4\)For the Univariate HP case, the excitation matrix has dimension one, being only the excitation function, \(\phi(t)\).
4.2 PWL

\[ [\text{PWL}]^{k_2} \text{ is lower bounded by a single PWL kernel with } K = \prod_{i=1}^{k_2} K_i, c = \max(c_1, ..., c_{k_2}) \text{ and } p = \sum_{i=1}^{k_2} p_i, \text{ thus still accounting for its ‘slow-decay’ behavior.} \]

4.3 SQR

\[ [\text{SQR}]^{k_3} \text{ reduces to a single SQR kernel with } B = \prod_{i=1}^{k_3} B_i \text{ and } L = \min(L_1, ..., L_{k_3}), \text{ thus still accounting for its ‘steady-triggering’ behavior: } [\text{SQR}]^{k_3} \subset [\text{SQR}]; \]

4.4 SNS

\[ [\text{SNS}]^{k_4} \text{ has } A = \prod_{i=1}^{k_4} A_i \text{ and a ‘spikier’ aspect (higher bandwidth), thus still accounting for its ‘periodicity-inducing’ behavior.} \]

Thus, on deepening the decomposition algorithm by overly increasing the number of levels above 2, we may be, in fact, adding little information on the qualitative aspect of the self-exciting behavior analysis of the data while making it more prone to overfitting on the noisiness of the discretized estimation vectors.

4.5 Upper Bound

Furthermore, regarding the boundedness of the higher-order decompositions, from the exact results for EXP and SQR intraclass decompositions and the upper bounds for the PWL and SNS ones, we have that:

\[
[\text{EXP}]^{k_1} \times [\text{PWL}]^{k_2} \times [\text{SQR}]^{k_3} \times [\text{SNS}]^{k_4} \leq \alpha e^{-\beta x} \frac{K}{(x + c_{\text{upper}})^p} B \sin(\omega x) \\
\leq \frac{\alpha B K e^{-\beta x}}{(x + c_{\text{upper}})^p} = \text{EXP}(\alpha, \beta) \times \text{PWL}(K, c_{\text{upper}}, p)^{k_2} \times \text{SQR}(B, L) \times A,
\]

for \(0 \leq x \leq \min(L, \frac{\pi}{\omega})\), and 0 otherwise.

5 Automated Report Generation

After estimating the likelihood-maximizing base kernel combination, a LaTeX file report, a sample of which is included in the Appendix, is generated with a description of the corresponding parameters and its most relevant aspects, such as:
- **Decay-Rate**, which would be associated with decaying kernels, EXP and PWL. A exponential decay rate is described as ‘a quick decaying triggering influence’, and the power-law decay may be described as ‘a slowly decaying triggering influence’, with its respective parameters being put in evidence;

- **Steady-Rate Influence**, equivalent to an offset level from the SQR kernel type. It is described as ‘a steadily triggering influence’, with the parameter B put into evidence;

- **Triggering Horizon**, associated with discontinuous kernels, SNS and SQR. It is described as ‘a horizon of triggering influence’, with the parameter L, in the SQR case, or the value of $\left(\frac{\pi}{\omega}\right)$, in the SNS case, put into evidence;

- **Induced Periodicity**, in the case of the SNS kernel type. It is described as ‘a periodicity-inducing triggering influence’, with the value of $\left(\frac{2\pi}{\omega}\right)$, from the SNS case, put into evidence.

6 Related Work

A spectral analysis approach to a one-dimensional self-exciting point process was introduced in [9]. Recently, the spectral method was extended to non-parametric kernel estimation for symmetrically networked HPs [11]. A Likelihood Maximization method was used on a class of parametric kernels for the excitation matrices in [19], while a series of works for likelihood maximization methods for power-law shaped kernels on seismology and earthquake data modeling were compiled in [18]. However, in [2], it is argued that, being designed for univariate Hawkes Processes, these methods can hardly be used to handle large amounts of data where the kernel function is not well localized compared to the exogenous inter-events time. Subsequently, an extensive analysis of spectral methods for non-parametric kernel estimation on networked HPs is performed.

In [14], Networked HPs are explored for analysis and modeling of stock-trading and crime data. A variant of the spectral method for networked HPs, with excitation matrix composed by linear combinations of decaying-exponentials, there referred to as ‘excitation modes’, is developed in [8]. An iterative log-likelihood maximization for non-parametric kernel estimation is presented in [23].

In [21], a maximum-likelihood estimator with a Sparse-Group-Lasso regularizer is introduced. In [15], an extensive mathematical treatment of the so-called ‘genuine multivariate’ Hawkes Processes, along with some considerations on the case of large-scale and networked data analysis, is provided. In [13], a nonparametric Expectation-Maximization algorithm for Multi-scale Hawkes Processes under the
assumption of exponential triggering kernels is proposed, while, in [12], a large-scale inference algorithm for kernels modeled after a summation of exponentials is introduced.

7 Experimental Results

For testing the validity of the kernel decomposition framework, we conducted experiments with synthetic, financial and earthquake data. Synthetic data sequences were used for evaluating the effectiveness of a scale-independence strategy for setting the maximum value of $\tau$ in Equation 11. In the financial data experiments, quantitative comparisons were made among the first and second levels of the decomposition framework over a set of time sequences, and also among our decomposition framework and an estimation method largely used in financial applications, described in [19]. In the earthquake data experiments, we evaluate the overall triggering behavior of earthquakes through the relative frequency of each estimated kernel type, out of the four base kernels.

For real-world data sets, no prior information about the kernel (type and parameters) is available. Thus, we use the log-likelihood of the kernel function over the time sequence as a quality criterion.

Given a realization $(t_1, t_2, \ldots, t_k)$ of some regular point process on $[0,T]$, its log-likelihood ($l$) is expressed as:

$$ l = \sum_{i=1}^{k} \log(\lambda(t_i)) - \int_0^T \lambda(u)du. \tag{18} $$

7.1 Synthetic Data

As a means of evaluating the overall precision and efficiency of the kernel decomposition framework, experiments with synthetic data from the corresponding four basic kernel types were performed.

Simulation algorithms related to Hawkes Processes, from which synthetic time sequences data may be obtained, are divided into two main categories: cluster-based [17] and intensity-based [20]. For the proposed experiments, the “Thinning” algorithm, an intensity-based one, was used, mainly because the parametric representations of the kernel types make it very convenient to accurately calculate the value of the intensity function throughout the whole simulation horizon.

The referred algorithm, whose full denomination is “Ogata’s Modified Thinning Algorithm”, basically consists of simulating time sequences using a exceedingly high constant value for the intensity function and then thinning out the generated
Table 4: Confusion matrices among the four basic kernel types for original horizon length (left) and histogram-based horizon length (right).

For an automatic decomposition framework willing to perform effective analysis of several domains of data, scale-independence is indispensable, as time sequences of disjoint datasets may occur in time scales differing by several orders of magnitude. As an example, earthquake events’ occurrences in a sequence are spaced by intervals of monthly and yearly scales. Thus, setting a horizon of a few months as the maximum value of $\tau$ in Equation (11) might result in a satisfactory discrete estimation grid, but using the same time length for estimating the triggering behavior of a finance-related sequence would require an impractically large grid resolution for making itself effective.

A histogram of all the time intervals between events in a sequence may be readily generated, and is an indicator of the overall magnitude of the spacing among the events. Thus, as a rule of thumb, the horizon length for $\tau$ may be set as the smaller time interval strictly larger than 95% of the sequence’s intervals. In practice, this value of horizon length is obtained with the help of a histogram composed by 100 bins.

First, we generated 10 sequences with the ‘Thinning algorithm’ [20], in the time range of 0 to 100000, for each of the four basic kernel types, with predefined kernel parameters and influence horizon of 6.6. Then, the decomposition framework was used for building confusion matrices over the sets of sequences for both the original horizon length (i.e., 6.6) and the histogram-based horizon length. The confusion matrices are shown in Table [4]

---

5By influence horizon, we mean that the excitation effect of each kernel is considered null for $t > 6.6$. 

7.2 Financial Data

Previous empirical studies have provided evidence that price impact has, on many aspects, some universal properties and is the main source of price variations, what reinforces the notion of an endogenous nature, in other words, an internal feedback mechanism, of price fluctuations. This is contrasting with the classical notion of an exogenous flow of information driving the prices towards a fundamental value [3].

In this endogeneity scenario, Hawkes Processes have become more and more present in the high-frequency finance domain, due to its structure being naturally adapted to model systems in which the discrete nature of the jumps in $N_t$ is relevant, making the model remarkably suited to modeling high-frequency data [6]. Moreover, Hawkes Processes’ parameters, along with their corresponding straightforward interpretations, lead to a notably simple and flexible interpretation of the complex intra-day dynamics of modern electronic markets.

For the experimental setting, we picked 19 technology companies in the NASDAQ and NYSE lists of Google Finance: CSCO, GOOGL, HPQ, INTC, IBM, MSFT, ORCL, TXN, XRX, AAPL, ANGI, LITH, IAC, FB, LOCMQ, YELP, GRPN, AMZN and GCI. We extracted tick data from every two minutes of 30 business days (04/07/2017 to 05/18/2017). Whenever a stock price changed by some magnitude higher than some threshold, an event was logged in the corresponding time sequence. Ten different percentual thresholds, increasing at equally spaced intervals from 0.03% to 0.3%, were applied. This procedure resulted in 70 valid sequences, since the remaining ones did not contain enough points for the splitting between training and validation subsequences.

The 80% of the first elements for each sequence were then used as training data, and the remaining 20% were used for validation, i.e., we estimated the parameters of the kernel using the first 24 days and then calculated the log-likelihood on the last 6 days of each sequence. When comparing the log-likelihoods of first and second level decompositions, we observed that the second level, with composite kernels, resulted in a higher log-likelihood in 69 out of the 70 sequences (98.57%), what corroborates
Figure 7: Comparison of the log-likelihood of the Kernel Decomposition Algorithm and the gradient descent based Exponential Hawkes Process over Stock Data.

| Level   | K1     | K2     |
|---------|--------|--------|
| Average Loglikelihood | -2096.95 | -1894.61 |

Table 5: Average Loglikelihood, over the 70 sequences, for the two levels of decomposition.

the idea of a more flexible model of the kernel providing a more accurate description of the underlying dynamics of the process. Along the 70 sequences, the average value of the log-likelihood of the K1 Decomposition is \(-2096.95\), while, for the K2 Decomposition, it is equal to \(-1894.61\). The average Loglikelihood for each level is shown in Table 5. The percentual difference from each sequence is shown in Figure 8.

When comparing the performance of the best estimation among the two levels and the usual Exponential HP model used in financial analysis, fitted through the gradient-based method from [19], it is possible to see that the Decomposition Algorithm exhibited a much more robust performance. Although the Exponential HP performed well in some sequences, it tended to get stuck in local maxima with very poor performance, usually leading to unstable or negative combinations of parameters. The Decomposition Algorithm performs better in 67.14% (47 out of 70) of the sequences. The comparison for each individual sequence is shown in Figure 7.

7.3 Earthquake Data

In lists regularly published by seismological services of most countries habituated to the occurrences of earthquakes, one can obtain relevant information, such as the epicenter of each shock, focal depth, instrumental magnitude and origin time of each earthquake occurrence. Statistical analyses of earthquake catalogues and
Figure 8: Percentual difference, for each sequence, among loglikelihood from first- and second-level decompositions. A positive value means that K2 decomposition results in a higher log-likelihood than the K1 decomposition.

Figure 9: Comparison, for each sequence, among log-likelihood from first- and second-level decompositions.
assessment of earthquake risk in a geophysical area can then be performed through the use of parametric models on the sequences of origin times, obtained by largely ignoring the remaining information. On further analyzing these fitted models, one can then identify and decompose components such as evolutionary trend, periodicity and clustering [18].

Again, the need to insert scale-independence, in the form of the aforementioned histogram-related heuristics, in the decomposition framework makes itself imperative, as earthquakes events are separated by time intervals of monthly or yearly scales. Thus, the estimation horizon for financial data, for example, lasting usually only a few seconds, would hardly capture the overall aspect of the triggering behavior in this case.

The data considered for the experiment was a set of 100 time sequences extracted from the USGS NCSN Catalog (NCEDC database), from the day of 1966/Jan/01 to 2015/Jan/01. The Latitude range was [30,55], and the Longitude range was [-140,-110]. Different length intervals and resulting areas were considered. Whenever the magnitude of an event exceeded some threshold, its time coordinate was added to the corresponding input time sequence. The magnitude thresholds were varied among 2.5, 3.0, 3.5 and 4.0; and the grid resolution was set to 20 and 100 points. For 20-point grid resolution, the relative frequency of each kernel was \((\text{EXP},\text{PWL},\text{SQR},\text{SNS}) = (0,97,2,1)\). For the 100-point grid resolution, the relative frequency was \((0,99,1,0)\). The results of first and second level decompositions are shown in Tables 6 and 7.

The results indicate a strong agreement with the long standing assumption of a power-law shaped kernel for the intensity of aftershocks’ occurrences (‘Omori’s Law’ (1894)). Q-Q plots from the estimated models are shown in Figure [10].
Figure 10: Q-Q Plots from Kernel Decomposition estimation over the Earthquake Dataset sequences.
8 Conclusion

Hawkes processes are temporal point processes which capture self-exciting discrete events in time series data. To predict future events with HPs, an appropriate kernel is selected by hands, previously. In this paper, we proposed a new temporal covariance-based kernel decomposition method to represent various self-exciting behaviors. We also presented a model (structure/parameter) learning algorithm to select the best HP kernel given the temporal discrete events. The stationarity conditions are derived to guarantee the validity of the kernel learning algorithm. In experiments, we demonstrated that the proposed algorithm performs better than existing methods to predict future events by automatically selecting kernels.

References

[1] E. Bacry, K. Dayri, and J. F. Muzy. Non-parametric kernel estimation for symmetric hawkes processes. application to high frequency financial data. The European Physical Journal B, 85(5):1–12, 2012.

[2] E. Bacry and J. Muzy. First- and second-order statistics characterization of hawkes processes and non-parametric estimation. IEEE Transactions on Information Theory, 62(4):2184–2202, 2016.

[3] J. P. Bouchaud, J. D. Farmer, and F. Lillo. How markets slowly digest changes in supply and demand. Handbook of Financial Markets: Dynamics and Evolution, pages 57–156, 2008.

[4] D. Daley and D. Vere-Jones. An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods. Springer, 2003.

[5] D. K. Duvenaud, J. R. Lloyd, R. B. Grosse, J. B. Tenenbaum, and Z. Ghahramani. Structure discovery in nonparametric regression through compositional kernel search. In Proceedings of the International Conference on Machine Learning, pages 1166–1174, 2013.

[6] E. Bacry, I. Mastromatteo, and J. Muzy. Hawkes processes in finance. arXiv, 2015.

[7] P. Embrechts, T. Liniger, and L. Lin. Multivariate hawkes processes: an application to financial data. Applied Probability Trust, 2011.

[8] J. Etesami, N. Kiyavash, K. Zhang, and K. Singhal. Learning network of multivariate hawkes processes: A time series approach. In Proceedings of the Conference on Uncertainty in Artificial Intelligence, 2016.

[9] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. Biometrika, (1):201–213, 1971.
[10] Y. Hwang, A. Tong, and J. Choi. Automatic construction of nonparametric relational regression models for multiple time series. In Proceedings of the International Conference on Machine Learning, pages 3030–3039, 2016.

[11] P. Laub, T. Taimre, and P. Pollett. Hawkes processes. 7 2015.

[12] R. Lemmonier, K. Scaman, and A. Kalogeratos. Multivariate hawkes processes for large-scale inference. In Proceedings of the Conference on Artificial Intelligence, 2017.

[13] E. Lewis and G. Mohler. A nonparametric em algorithm for multiscale hawkes processes. Journal of Nonparametric Statistics, 1(1):1–20, 2011.

[14] S. W. Linderman and R. P. Adams. Discovering latent network structure in point process data. In Proceedings of the International Conference on Machine Learning, pages 1413–1421, 2014.

[15] T. Liniger. Multivariate Hawkes Processes. PhD thesis, ETH Zurich, 2009.

[16] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita. Self-exciting point process modelling of crime. Journal of the American Statistical Association, 106(493):100–108, 2012.

[17] J. Moller and J. G. Rasmussen. Perfect simulation of hawkes processes. Advances in Applied Probability, 37(3):629–646, 2010.

[18] Y. Ogata. Seismicity analysis through point-process modelling: A review. Pure and Applied Geophysics, 155(5):471–507, 1999.

[19] T. Ozaki. Maximum likelihood estimation of hawkes’ self-exciting point processes. Annals of the Institute of Statistical Mathematics, (31):145–155, 1979.

[20] J. G. Rasmussen. Temporal point processes: the conditional intensity function. 2009.

[21] H. Xu, M. Farajtabar, and H. Zha. Learning granger causality for hawkes processes. In Proceedings of the International Conference on Machine Learning, pages 1717–1726, 2016.

[22] Q. Zhao, M. A. Erdogdu, H. Y. He, A. Rajaraman, and J. Leskovec. Seismic: A self-exciting point process model for predicting tweet popularity. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 1513–1522, 2015.

[23] K. Zhou, H. Zha, and L. Song. Learning triggering kernels for multi-dimensional hawkes processes. In Proceedings of the International Conference on Machine Learning, pages 1301–1309, 2013.
A Derivations of Stationarity Criteria for Multiplicative Combinations of Kernels

This appendix introduces the full derivations of stationarity criteria for the second order multiplicative compositions of the four base kernels.

A.1 EXP x EXP

For the combination “EXPxEXP”, we have that, for stationarity to be achieved:

\[
0 \leq \int_0^\infty EXP(\alpha_1, \beta_1)EXP(\alpha_2, \beta_2)dx < 1
\]

\[
0 \leq \int_0^\infty \alpha_1 e^{\alpha_1 x} \cdot \alpha_2 e^{\beta_2 x} dx < 1
\]

Thus:

\[
\int_0^\infty \alpha_1 e^{-\beta_1 x} \cdot \alpha_2 e^{-\beta_2 x} dx = \int_0^\infty (\alpha_1 \alpha_2) e^{-(\beta_1 + \beta_2) x} dx
\]

\[
= \int_0^\infty \alpha e^{-x} dx = \frac{\alpha}{\beta} = \frac{\alpha_1 \alpha_2}{\beta_1 + \beta_2}
\]

So, this case reduces to the case of a single exponential.

A.2 EXP x PWL

For the combination “EXPxPWL”, we have that, for stationarity to be achieved:

\[
0 \leq \int_0^\infty EXP(\alpha, \beta)PWL(K, c)dx < 1
\]

\[
0 \leq \int_0^\infty \alpha e^{-\beta x} \frac{K}{(x+c)^p} dx < 1
\]
Thus:
\[
\int_0^\infty \alpha e^{-\beta x} \frac{K}{(x+c)^p} \, dx = \alpha K \int_0^\infty (x+c)^{-p} e^{-\beta x} \, dx \\
= \alpha Ke^{\beta c} \int_0^\infty (x+c)^{-p} e^{-\beta (x+c)} \, dx \\
= \alpha Ke^{\beta c} \int_0^\infty (\beta(x+c))^{-p} e^{-\beta(x+c)} \, dx \\
= \alpha Ke^{\beta c} \int_{\beta c}^\infty t^{-p} e^{-t} \, dt \\
= \alpha Ke^{\beta c} \Gamma(1-p, \beta c),
\]
where $\Gamma(\cdot, \cdot)$ is the well-known Incomplete Gamma Function: $\Gamma(a, y) = \int_y^\infty t^{a-1} e^{-t} \, dt$.

A.3 EXP x SQR

For the combination “EXPxSQR”, we have that, for stationarity to be achieved:
\[
0 \leq \int_0^\infty EXP(\alpha, \beta)SQR(B, L) \, dx < 1
\]
\[
0 \leq \int_0^L \alpha Be^{-\beta x} \, dx < 1
\]
Thus:
\[
\int_0^L \alpha Be^{-\beta x} \, dx = \left[ \frac{\alpha Be^{-\beta x}}{\beta} \right]_0^L = \frac{\alpha B(1-e^{-\beta L})}{\beta}
\]
So, in the case of a multiplicative combination, the SQR kernel acts as a truncation horizon.

A.4 EXP x SNS

For the combination “EXPxSNS”, we have that, for stationarity to be achieved:
\[
0 \leq \int_0^\infty EXP(\alpha, \beta)SNS(A, \omega) \, dx < 1
\]
\[
0 \leq \int_0^\pi A\alpha e^{-\beta x} \sin(\omega x) \, dx < 1
\]
Where:

\[ \int_{0}^{\pi} A e^{-\beta x} \sin(\omega x) dx = \int_{0}^{\pi} A e^{-\beta x} \frac{e^{i\omega x} - e^{-i\omega x}}{2i} dx \]

\[ = \frac{A \alpha}{2i} \left[ \frac{e^\beta e^{i\omega x}}{-\beta + i\omega} - \frac{e^{-\beta} e^{-i\omega x}}{-\beta - i\omega} \right]^{\pi}_{0} \]

\[ = \frac{A \alpha}{2i} \left[ (\beta - i\omega) e^{(-\beta + i\omega)x} - (\beta + i\omega) e^{(-\beta - i\omega)x} \right]^{\pi}_{0} \frac{1}{\omega} \]

\[ = \frac{A \alpha}{2i} \left[ \frac{2i\omega \cos(\omega x) - 2\beta \sin(\omega x)}{\beta^2 + \omega^2} \right]^{\pi}_{0} \]

\[ = \frac{A \alpha - 2i\omega(e^{-\omega} - 1)}{2i(\beta^2 + \omega^2)} \]

\[ = \frac{A \alpha \omega(1 + e^{-\beta \pi})}{(\omega^2 + \beta^2)} \]

A.5 PWL x PWL

In the case of the combination “PWLxPWL”, an upper bound is derived as follows:

\[ 0 \leq \int_{0}^{\infty} PWL(K_1, c_1, p_1) \cdot PWL(K_2, c_2, p_2) dx < 1 \]

\[ 0 \leq \int_{0}^{\infty} \frac{K_1}{(x + c_1)^{p_1}} \frac{K_2}{(x + c_2)^{p_2}} dx < 1 \]

Then:

\[ \int_{0}^{\infty} \frac{K_1}{(x + c_1)^{p_1}} \frac{K_2}{(x + c_2)^{p_2}} dx \leq \frac{K_1 K_2}{(x + \min(c_1, c_2))^{p_1 + p_2}} dx \]

\[ = \frac{K_1 K_2}{(p_1 + p_2 - 1) \min(c_1, c_2)(p_1 + p_2 - 1)} \]
A.6 PWL x SQR

For the combination “PWLxSQR”, we have that, for stationarity to be achieved:

\[ 0 \leq \int_{0}^{\infty} PWL(K, c, p) SQR(B, L) dx < 1 \]

\[ 0 \leq \int_{0}^{L} \frac{KB}{(x + c)^p} dx < 1 \]

Where:

\[ \int_{0}^{L} \frac{KB}{(x + c)^p} dx = \left[ \frac{KB}{(1 - p)(x + c)^{(p-1)}} \right]_{0}^{L} = \frac{KB(c^{-(p-1)} - (c + L)^{-(p-1)})}{p - 1} \]

So, once again, the SQR kernel acts as a truncation horizon.

A.7 PWL x SNS

In the case of the combination “PWLxSNS”, an upper bound is derived as follows:

\[ 0 \leq \int_{0}^{\infty} PWL(K, c, p) SNS(A, \omega) dx < 1 \]

\[ 0 \leq \int_{0}^{\pi} \frac{KA \sin(\omega x)}{(x + c)^p} dx < 1 \]

Where:

\[ \int_{0}^{\pi} \frac{KA \sin(\omega x)}{(x + c)^p} dx \leq \int_{0}^{\pi} \frac{KA}{(x + c)^p} dx = \left[ \frac{KA}{(1 - p)(x + c)^{(p-1)}} \right]_{0}^{\pi} = \frac{KA((c + \frac{\pi}{2})^{1-p} - c^{1-p})}{1 - p} \]
A.8 SQR x SQR

For the combination “SQRxSQR”, we have that, for stationarity to be achieved:

\[
0 \leq \int_{0}^{\infty} SQR(B_1, L_1)SQR(B_2, L_2)dx < 1
\]

\[
0 \leq \int_{0}^{\min(L_1, L_2)} B_1B_2dx < 1
\]

Where:

\[
\int_{0}^{\min(L_1, L_2)} B_1B_2dx = B_1B_2\min(L_1, L_2) = BL
\]

So, the multiplicative combination of two SQR kernels may be reduced to the case of a single SQR kernel.

A.9 SQR x SNS

In the case of combinations of discontinuous kernels (SQR and SNS), we assume they have the same starting and ending points, i.e., \( L = \frac{\pi}{\omega} \). So, for the combination “SQRxSNS”, we have that, for stationarity to be achieved:

\[
0 \leq \int_{0}^{\infty} SQR(B, L)SNS(A, \omega)dx < 1
\]

\[
0 \leq \int_{0}^{\frac{\pi}{\omega}} AB\sin(\omega x)dx < 1
\]

Where:

\[
\int_{0}^{\frac{\pi}{\omega}} AB\sin(\omega x)dx = \frac{2AB}{\omega}
\]

A.10 SNS x SNS

In the case of combinations of discontinuous kernels (SQR and SNS), we assume they have the same starting and ending points. So, for the combination “SNSxSNS”, we have that, for stationarity to be achieved:
\[ 0 \leq \int_0^\infty SNS(A_1, \omega)SNS(A_2, \omega)dx < 1 \]

\[ 0 \leq \int_0^\infty A_1A_2\sin^2(\omega x)dx < 1 \]

Where:

\[ \int_0^\infty A_1A_2\sin^2(\omega x)dx = \int_0^\infty A(1 - \cos(2\omega x))dx = \frac{\pi A}{2\omega} \]

**B Derivation of the Log-likelihood formula for HPs**

This derivation follows the steps on [11]. Given a realization \((t_1, t_2, ..., t_k)\) of some regular point process observed over the interval \([0,T]\), the log-likelihood is expressed as:

\[ l = \sum_{i=1}^{k} \log(\lambda(t_i)) - \int_0^T \lambda(u)du \]

**Proof.** Let be the joint probability density of the realization:

\[ L = f(t_1, t_2, ..., t_k) = \prod_{i=1}^{k} f(t_i) \]

It can be written in terms of the Conditional Intensity Function. We can then find \(f\) in terms of \(\lambda\):

\[ \lambda(t) = \frac{f(t)}{1 - F(t)} = \frac{dF(t)}{1 - F(t)} = -\frac{d\log(1 - F(t))}{dt}, \]

where, given the history up to last arrival \(u, \mathcal{H}(u)\), \(F(t)\) is then defined as the conditional cumulative probability distribution of the next arrival time \(T_{k+1}^t\):

\[ F(t) = F(t|\mathcal{H}(u)) = \int_u^t f(s|\mathcal{H}(u))ds \]
Integrating both sides of Equation (19) over \((t_k, t)\):

\[- \int_{t_k}^{t} \lambda(u)du = \log(1 - F(t)) - \log(1 - F(t_k))\]

Given that the realization is assumed to have come from a so-called simple process, i.e., a process in which multiple arrivals cannot occur at the same time, we have that \(F(t_k) = 0\) as \(T_{k+1} > t_k\), which simplifies equation (19) to:

\[- \int_{t_k}^{t} \lambda(u)du = \log(1 - F(t))\]

Further rearranging the expression:

\[F(t) = \exp \left( - \int_{t_k}^{t} \lambda(u)du \right),\]

and

\[f(t) = \lambda(t) \exp \left( - \int_{t_k}^{t} \lambda(u)du \right)\]

Thus, the likelihood becomes:

\[L = \prod_{i=1}^{k} f(t_i) = \prod_{i=1}^{k} \lambda(t_i) \exp \left( - \int_{t_{i-1}}^{t_i} \lambda(u)du \right)\]

\[= \left[ \prod_{i=1}^{k} \lambda(t_i) \right] \exp \left( - \int_{0}^{t_k} \lambda(u)du \right)\]

Given that the process is observed on \([0,T]\), the likelihood must include the probability of seeing no arrivals in \((t_k, T)\):

\[L = \left[ \prod_{i=1}^{k} f(t_i) \right] \left( 1 - F(T) \right)\]

Through using the formulation of \(F(t)\), we have that:

\[L = \left[ \prod_{i=1}^{k} \lambda(t_i) \right] \exp \left( - \int_{0}^{T} \lambda(u)du \right)\]

Finally, getting the logarithm of the expression, we have the formula for \(l\):

\[l = \sum_{i=1}^{k} \log(\lambda(t_i)) - \int_{0}^{T} \lambda(u)du\]
C Automatic Report

Discretized Kernel Estimation

This part corresponds to the results from the Discretized Kernel Estimation step:
K1 Decomposition

Results from the first level of kernel decomposition (EXP,PWL,SQR,SNS):

Chosen K1: Sinousidal
The data most likely presents some Periodicity-Inducing self-triggering behavior with Amplitude (A) 0.053554413238203313 and Angular Frequency (omega) equal to 0.915916255363827
K2 Decomposition

Chosen K2: SNS x EXP

The data most likely presents a jointly quickly decaying and periodicity-inducing self-triggering behaviour of Amplitude (alphaA) of 0.09493520564312838, Decay Rate (beta) equal to 0.35200971347500182, and angular frequency (omega) of 0.90055454546505057.

Output Kernel

Considering the two levels of decomposition, the best kernel fitting is 'SNS'.