A. Technical Proofs

A.1. Proof of Theorem 1

We first start by showing that a random translation of an inhomogeneous Poisson process is another a Poisson process (Daley & Jones, 2003; Doob, 1953). Subsequently, we will use this result to compute the cumulants of a randomly translated MHP.

Lemma 1. *(Daley & Jones, 2003) Let $N(\cdot)$ be an inhomogeneous Poisson process on $\mathbb{R}_+$ with rate $\lambda(t)$. The resulting process after a random translation with distribution function $F(\cdot)$ is another Poisson process with rate*

$$\widetilde{\lambda}(t) \coloneqq \int_{-\infty}^{t} \lambda(t-x)F(dx).$$

Proof. We prove this lemma by showing that the number of events within an arbitrary interval of the randomly translated Poisson is Poisson with the specified rate. Let $I_\theta = [0, \theta]$ be an interval in $\mathbb{R}$, then

$$\mathbb{P}(N(I_\theta) = k) = \frac{(\Lambda(I_\theta))^k}{k!} e^{-\Lambda(I_\theta)},$$

where $\Lambda(I_\theta)$ is $\int_0^\theta \lambda(x)dx$. The probability that an arrival at time $t \in I_\theta$ is translated to an arbitrary interval $I \coloneqq [a, b]$ is

$$\mathbb{P}(a \le x + t \le b) = F(b-t) - F(a-t) \coloneqq F(I-t)$$

The probability that only m of k arrivals within I_θ are translated to I is equal to

$$\binom{k}{m} \left(\int_0^\theta F(I-t)\lambda(t)dt/\Lambda(I_\theta) \right)^m \left(\int_0^\theta (1-F(I-t))\lambda(t)dt/\Lambda(I_\theta) \right)^{k-m}.$$

Therefore, the probability that m events are observed in I after translation is

$$\begin{aligned}
&\lim_{\theta\to\infty} \sum_{k\ge m} \frac{(\Lambda(I_\theta))^k}{k!} e^{-\Lambda(I_\theta)} \binom{k}{m} \left(\int_0^\theta F(I-t)\lambda(t)dt/\Lambda(I_\theta) \right)^m \left(\int_0^\theta (1-F(I-t))\lambda(t)dt/\Lambda(I_\theta) \right)^{k-m} \\
&= \lim_{\theta\to\infty} \sum_{k\ge m} \frac{e^{-\Lambda(I_\theta)}}{m!(k-m)!} \left(\int_0^\theta F(I-t)\lambda(t)dt \right)^m \left(\int_0^\theta (1-F(I-t))\lambda(t)dt \right)^{k-m} \\
&= \lim_{\theta\to\infty} \frac{e^{-\Lambda(I_\theta)}}{m!} \left(\int_0^\theta F(I-t)\lambda(t)dt \right)^m \sum_{k\ge 0} \frac{1}{k!} \left(\int_0^\theta (1-F(I-t))\lambda(t)dt \right)^k \\
&= \lim_{\theta\to\infty} \frac{e^{-\Lambda(I_\theta)}}{m!} \left(\widetilde{\Lambda}(I_\theta) \right)^m e^{\left(\Lambda(I_\theta) - \widetilde{\Lambda}(I_\theta)\right)},
\end{aligned}$$

where $\widetilde{\Lambda}(I_\theta) \coloneqq \int_0^\theta F(I-t)\lambda(t)dt$. Letting θ to go to infinity and setting $\{a = v, b = v + dv\}$, we obtain

$$\begin{aligned}
\mathbb{P}(d\tilde{N}(v) = 1) = \widetilde{\lambda}(v)dv &= \left(\int_0^\infty f(v-u)\lambda(u)du \right) dv \\
&= \left(\int_{-\infty}^{v} \lambda(v-x)F(dx) \right) dv,
\end{aligned}$$

where $F(dv) = f(v)dv$. □

To establish the result in Theorem 1, it is also useful to prove the following lemma.

Lemma 2. *For every multi-index $\boldsymbol{i} = (i_1, ..., i_n)$ and every vector $\tilde{\boldsymbol{t}} = (\tilde{t}_1, ..., \tilde{t}_n)$, we have*

$$K_{\boldsymbol{i}}(\tilde{\boldsymbol{t}})d\tilde{\boldsymbol{t}} = \mathbb{P}(E^{\boldsymbol{i}}_{\tilde{\boldsymbol{t}}} \cap C^{\boldsymbol{i}}_{\tilde{\boldsymbol{t}}}),$$

where $E^{\boldsymbol{i}}_{\tilde{\boldsymbol{t}}}$ denotes the event that for every k there is a type i_k events at time $\tilde{t}_k$ and $C^{\boldsymbol{i}}_{\tilde{\boldsymbol{t}}}$ is the event that there exists a cluster such that all the events in $\tilde{\boldsymbol{t}}$ belong to that cluster.

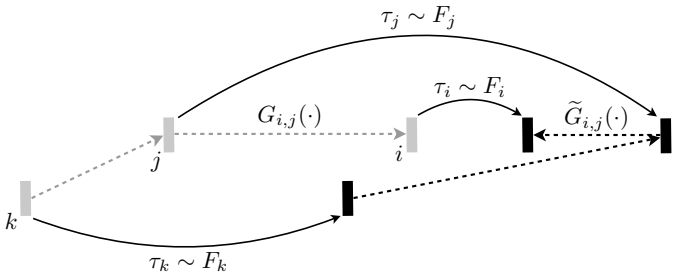


Figure 8: The evolution of a part of a cluster after random translation. Nodes are labeled by their types.

Proof. This can be seen by the facts that

$$\mathbb{E}[dN_{i_1}(\tilde{t}_1)...dN_{i_n}(\tilde{t}_n)] = \mathbb{P}\left(\forall k \in \{1,...,n\}, \text{ there is a type } i_k \text{ event at } \tilde{t}_k\right) = \mathbb{P}(E^{\boldsymbol{i}}_{\tilde{\boldsymbol{t}}}),$$
$$\mathbb{P}(E^{\boldsymbol{i}}_{\tilde{\boldsymbol{t}}}) = \mathbb{P}(E^{\boldsymbol{i}}_{\tilde{\boldsymbol{t}}} \cap C^{\boldsymbol{i}}_{\tilde{\boldsymbol{t}}}) + \mathbb{P}(E^{\boldsymbol{i}}_{\tilde{\boldsymbol{t}}} \cap \bar{C}^{\boldsymbol{i}}_{\tilde{\boldsymbol{t}}}),$$

where $\bar{C}^{\boldsymbol{i}}_{\tilde{\boldsymbol{t}}}$ denotes the complement of the event $C^{\boldsymbol{i}}_{\tilde{\boldsymbol{t}}}$. The rest follows similarly to the proof of Equation (24) in Appendix A of (Jovanović et al., 2015). □

Lemma 3. *Consider the setting in Theorem 1, and define*

$$\widetilde{R}^{(t)}_{i,j}dt \coloneqq \mathbb{P}\left(\textit{type } j \textit{ event at 0 causes type } i \textit{ event at } t\right),$$

then $\widetilde{R}^{(t)}_{i,j} = \left[\sum_{n\geq 0} \widetilde{\boldsymbol{G}}^{*n}(t)\right]_{i,j}$.

Proof. Suppose in a cluster C of the original (before translation) MHP, dimension j at some time y triggers an arrival in dimension i. Based on the definition of the clusters, the arrival times in dimension i are distributed as an inhomogeneous Poisson with rate $G_{i,j}(t-y)$. See Figure 8.

Suppose that nodes j and i are translated by τ_j and τ_i, respectively. Then $\tilde{t}_j = y + \tau_j$ and $\tilde{t}_i = y + x + \tau_i$, where $x \sim \text{Exp}(G_{i,j}(t-y))$, and $\tilde{t}_i - \tilde{t}_j = x + \tau_i - \tau_j$.

The above observation and Lemma 1 imply that

$$\frac{\mathbb{P}(dN_i(\tilde{t}_i) = 1|\tau_j)}{dt} = \int_{\mathbb{R}} f_i(\tilde{t}_i - s)G_{i,j}(s-y)ds = \int_{\mathbb{R}} f_i(\tilde{t}_i - \tilde{t}_j + \tau_j - s)G_{i,j}(s)ds,$$

where N_i denotes the number of arrivals in i-th dimension of the translated process. Therefore,

$$\widetilde{G}_{i,j}(\tilde{t}_i - \tilde{t}_j) \coloneqq \frac{\mathbb{P}(dN_i(\tilde{t}_i) = 1)}{dt} = \int_{\mathbb{R}}\int_{\mathbb{R}} f_i(\tilde{t}_i - \tilde{t}_j + \tau_j - s)G_{i,j}(s)f_j(\tau_j)dsd\tau_j.$$

This equation may be interpreted as follows: the cluster of an MHP after the random translation forms a new cluster in which dimension j at some time $\tilde{t}_j$ causes an offspring of type i by generating a realization of an inhomogeneous Poisson process with rate $\tilde{G}_{i,j}(t - \tilde{t}_j)$. Moreover, an immigrant from dimension k appears in the translated cluster with rate $\int_{\mathbb{R}} \mu_k F_k(dx) = \mu_k$.

Define $p^n_{i,j}(t)$ as the probability that an event of type j at 0, after n generations, causes a type i event at t in the translated cluster. Clearly, $p^0_{i,j}(t) = [\boldsymbol{I}\delta(t)]_{i,j}dt$. For $n > 0$, we have

$$p^1_{i,j}(t) = \widetilde{G}_{i,j}(t)dt,$$
$$p^2_{i,j}(t) = \sum_{k=1}^{d}\int_{\mathbb{R}} \tilde{G}_{k,j}(s)\tilde{G}_{i,k}(t-s)dsdt = [\widetilde{\boldsymbol{G}}^{*2}(t)]_{i,j}dt,$$
$$\vdots$$
$$p^n_{i,j}(t) = [\widetilde{\boldsymbol{G}}^{*n}(t)]_{i,j}dt.$$

This implies

$$\widetilde{R}_{i,j}^{(t)}dt = \sum_{n\geq 0} p_{i,j}^{n}(t) = \left[\sum_{n\geq 0} \widetilde{\boldsymbol{G}}^{*n}(t)\right]_{i,j} dt.$$

□

With Lemma 3 at hand, we are ready to prove Theorem 1.

For Equation (4): From the definition of the cumulant, we have

$$K_i(\tilde{t})d\tilde{t} = K(dN_i(\tilde{t})) = \mathbb{E}[dN_i(\tilde{t})].$$

The last term in the above expression is the probability that an immigrant, say from dimension j, generates an event in dimension i at time $\tilde{t}$, which equals

$$\sum_{j=1}^{d} \int_{\mathbb{R}} \mu_j \widetilde{R}_{i,j}^{(\tilde{t}-x)} dx = \sum_{j=1}^{d} \int_{\mathbb{R}} \mu_j \widetilde{R}_{i,j}^{(x)} dx = K_i.$$

For Equation (5): We have

$$K_{i,j}(\tilde{t}_1, \tilde{t}_2)d\tilde{t}_1 d\tilde{t}_2 = \mathbb{P}\left(\text{types } i,j \text{ events at times } \tilde{t}_1, \tilde{t}_2 \text{ within the same cluster}\right).$$

The above event happens if and only if node i and j have a common ancestor in a cluster, which happens with probability

$$\begin{aligned}
&\sum_{k=1}^{d} \int_{\mathbb{R}} \mathbb{P}\left(\text{an immigrant generates an event in dimension } k \text{ at time } x\right) \\
&\times \mathbb{P}\left(k \text{ generates } i \text{ and } j \text{ at times } \tilde{t}_1 \text{ and } \tilde{t}_2\right) d\tilde{t}_1 d\tilde{t}_2 dx \\
&= \sum_{k=1}^{d} \int_{\mathbb{R}} K_k \times \left(\widetilde{R}_{i,k}^{(\tilde{t}_1-x)} \widetilde{R}_{j,k}^{(\tilde{t}_2-x)} d\tilde{t}_1 d\tilde{t}_2\right) dx
\end{aligned}$$

For Equation (6): We use the above Lemmas 2 and 3, and the fact that i, j and k can all occur in one cluster if one of the followings cases happen: (1) $\{i,j\}$ and $\{k\}$, (2) $\{i,k\}$ and $\{j\}$, (3) $\{k,j\}$ and $\{i\}$, (4) $\{i,j,k\}$, where types within one set have a common ancestor and separate sets have a common ancestor. For example, in case of $\{i,j\}$ and $\{k\}$, i and j share a common ancestor and the common ancestor of $\{i,j\}$ and $\{k\}$ have their own common ancestor. Say the common ancestor of i and j be from type m at some time y, and assume m and k have a different common ancestor than $\{i,j\}$, say n at another time x. This case can be formally written as follows

$$\begin{aligned}
\text{case}(\{i,j\},\{k\}) &= \sum_{m,n=1}^{d} \int_{\mathbb{R}} K_n \widetilde{R}_{k,n}^{(\tilde{t}_3-x)} d\tilde{t}_3 \int_{\mathbb{R}} \widetilde{R}_{i,m}^{(\tilde{t}_1-y)} d\tilde{t}_1 \widetilde{R}_{j,m}^{(\tilde{t}_2-y)} d\tilde{t}_2 \widetilde{\Psi}_{m,n}(y-x) dy dx \\
&= \sum_{m=1,n=1}^{d} K_n \int_{\mathbb{R}} \int_{\mathbb{R}} \widetilde{R}_{k,n}^{(\tilde{t}_3-x)} \widetilde{R}_{i,m}^{(\tilde{t}_1-y)} \widetilde{R}_{j,m}^{(\tilde{t}_2-y)} \widetilde{\Psi}_{m,n}(y-x) dy dx d\tilde{t}_1 d\tilde{t}_2 d\tilde{t}_3.
\end{aligned}$$

In this expression, $\widetilde{\Psi}_{m,n}$ appears because nodes m and n cannot coincide, then there must be at least one generation difference from n to m. The probability of this event is

$$\sum_{r>0} p_{m,n}^{r}(y-x) = \left[\sum_{r>0} \widetilde{\boldsymbol{G}}^{*r}(y-x)\right]_{m,n} = \left[\widetilde{\mathbf{R}}^{(y-x)} - \boldsymbol{I}\delta(y-x)\right]_{m,n} \coloneqq \widetilde{\Psi}_{m,n}(y-x).$$

Similarly, one can compute the probability of the other partitions and conclude the result.

A.2. Proof of Corollary 1

Recall from Theorem 1 that $\widetilde{G}_{i,j}(t) = f_i * G_{i,j} * \underline{f}_j(t)$. Since functions $G_{i,j}$, f_i, and f_j are bounded, their convolutions is also bounded. We have

$$\overline{G}_{i,j} = \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}} f_i(t+x-s)G_{i,j}(s)f_j(x)dsdxdt = \int_{\mathbb{R}} G_{i,j}(s)ds.$$

The last equality is due to the fact that $\int_{\mathbb{R}} f_i(t)dt = \int_{\mathbb{R}} f_j(t)dt = 1$. By assumption, the matrix $\mathcal{L}[\mathbf{G}](0)$ has spectral radius less than one, and by the above equality so does the matrix $\overline{\boldsymbol{G}}$.

A.3. Proof of Corollary 2

The result is immediate from the definition of the covariance density matrix of a randomly translated MHP,

$$\widetilde{\Sigma}_{i,j}(\tilde{t}_1, \tilde{t}_2) \coloneqq K_{i,j}(\tilde{t}_1, \tilde{t}_2) - \frac{\mathbb{E}[dN_i(\tilde{t}_1)]}{d\tilde{t}_1}\epsilon_{i,j}\delta(\tilde{t}_1 - \tilde{t}_2),$$

and substituting the second order cumulant density given in (5).

A.4. Discussion on Covariance Density matrix Equations

We show that equations (13) and (15) do not admit unique solutions, but first we need the following Lemma.

Lemma 4. *For a stationary MHP and a random translation of it, both $\rho(\mathcal{L}[\boldsymbol{G}](s))$ and $\rho(\mathcal{L}[\widetilde{\boldsymbol{G}}](s))$ are strictly less than one for all $s \in \mathbb{C}$.*

Proof. From Gelfand's Formula, we know that for any matrix $\mathbf{B}$, $\rho(\mathbf{B}) = lim_{n\to\infty}||\mathbf{B}^n||^{1/n}$, where $||\cdot||$ is any matrix norm.

We will apply Gelfand's formula with $\mathbf{B} = \mathcal{L}[\mathbf{G}](s)$ and $||\cdot||$ chosen as the max norm $||\cdot||_{max}$, but first observe

$$||\mathcal{L}[\mathbf{G}]^n(s)||_{max} \coloneqq max_{i,j}\left|[\mathcal{L}[\mathbf{G}]^n(s)]_{i,j}\right| \leq max_{i,j}\left|[\mathcal{L}[\mathbf{G}]^n(0)]_{i,j}\right| = ||\boldsymbol{G}^n||_{max},\ \forall n \in \mathbb{N}, s \in \mathbb{C}.$$

We used the triangle inequality and the fact that for a positive function f, $|\mathcal{L}[f](s)| \leq |\mathcal{L}[f](0)|$. Now, applying Gelfand's Formula, we obtain

$$\rho\left(\mathcal{L}[\mathbf{G}](s)\right) \leq lim_{n\to\infty}||\boldsymbol{G}^n||_{max}^{1/n} = \rho\left(\boldsymbol{G}\right) < 1.$$

The last inequality is due to the stationarity assumption of the MHP. Following the same steps and the result of Corollary 1, one can show $\rho(\mathcal{L}[\widetilde{\boldsymbol{G}}](s)) < 1$. □

Lemma 5. *Equations (13) and (15) do not admit unique solutions in terms of $\boldsymbol{G}(t)$ and $\widetilde{\boldsymbol{G}}(t)$.*

Proof. We only present the proof for equation (13). The argument for (15) is similar. Let $\boldsymbol{G}(t)$ denote a solution to (13) and $\mathbf{R}(t) \coloneqq \boldsymbol{I}\delta(t) + \boldsymbol{\Psi}(t)$, then

$$\boldsymbol{\Sigma}(t) = \underline{\mathbf{R}} * \boldsymbol{\Lambda}\mathbf{R}^T(t) - \boldsymbol{\Lambda}\delta(t).$$

Define $\mathbf{R}_0(t) \coloneqq \mathbf{R}(t)\boldsymbol{\Lambda}^{1/2}\mathbf{O}\boldsymbol{\Lambda}^{-1/2}$, where $\mathbf{O}$ is any orthogonal matrix, i.e., $\mathbf{O}\mathbf{O}^T = \mathbf{O}^T\mathbf{O} = \mathbf{I}$. It is easy to see that

$$\boldsymbol{\Sigma}(t) = \underline{\mathbf{R}}_0 * \boldsymbol{\Lambda}\mathbf{R}_0^T(t) - \boldsymbol{\Lambda}\delta(t).$$

Lemma 4 implies that $\mathcal{L}[\mathbf{R}](s)$ is bounded and equals $(\mathbf{I} - \mathcal{L}[\mathbf{G}](s))^{-1}$, therefore

$$\begin{aligned}\mathcal{L}[\mathbf{R}_0](s) &= \mathcal{L}[\mathbf{R}](s)\boldsymbol{\Lambda}^{1/2}\mathbf{O}\boldsymbol{\Lambda}^{-1/2} = (\mathbf{I} - \mathcal{L}[\mathbf{G}](s))^{-1}\boldsymbol{\Lambda}^{1/2}\mathbf{O}\boldsymbol{\Lambda}^{-1/2} \\ &= \left(\boldsymbol{\Lambda}^{1/2}\mathbf{O}^T\boldsymbol{\Lambda}^{-1/2}(\mathbf{I} - \mathcal{L}[\mathbf{G}](s))\right)^{-1} = \left(\mathbf{I} - \mathbf{I} + \Lambda^{1/2}\mathbf{O}^T\boldsymbol{\Lambda}^{-1/2} - \boldsymbol{\Lambda}^{1/2}\mathbf{O}^T\boldsymbol{\Lambda}^{-1/2}\mathcal{L}[\mathbf{G}](s)\right)^{-1} = \left(\mathbf{I} - \mathbf{A}(s)\right)^{-1},\end{aligned}$$

where

$$\mathbf{A}(s) \coloneqq \mathbf{I} - \boldsymbol{\Lambda}^{1/2}\mathbf{O}^T\boldsymbol{\Lambda}^{-1/2} + \boldsymbol{\Lambda}^{1/2}\mathbf{O}^T\boldsymbol{\Lambda}^{-1/2}\mathcal{L}[\mathbf{G}](s).$$

This means $\boldsymbol{G}_0(t) = \mathcal{L}^{-1}[\mathbf{A}](t)$ is also a solution of (13).

□

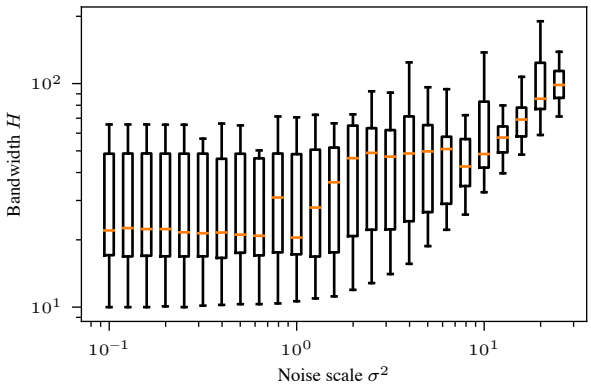


Figure 9: Tuning of the bandwidth H of the cumulants estimator used in the experiments of NPHC. We observe that the for noise power in range $\sigma^2 < 1/\beta = 1$, the best bandwidth H remains stable. For $\sigma^2 < 1/\beta = 1$ the best bandwidth H was increases linearly with σ^2.

B. Estimators for the Integrated Cumulants

We used the same empirical estimates for the cumulants as in (Achab et al., 2017) as follows.

$$
\begin{aligned}
\widehat{K}_i =& \frac{1}{T} \sum_{\tau \in \tilde{\mathbf{t}}_i} 1 = \frac{N_i(T)}{T}, \\
\widehat{\overline{K}}_{i,j} =& \frac{1}{T} \sum_{\tau \in \tilde{\mathbf{t}}_i} \big(N_j(\tau + H) - N_j(\tau - H) - 2H\widehat{K}_j \big), \\
\widehat{\overline{K}}_{i,j,k} =& \frac{1}{T} \sum_{\tau \in \tilde{\mathbf{t}}_i} \big(N_j(\tau + H) - N_j(\tau - H) - 2H\widehat{K}_j \big) \big(N_k(\tau + H) - N_k(\tau - H) - 2H\widehat{K}_k \big) \\
& - \frac{\widehat{\overline{K}}_i}{T} \sum_{\tau \in \tilde{\mathbf{t}}_j} \sum_{\tau' \in \tilde{\mathbf{t}}_k} \max\{(2H - |\tau' - \tau|), 0\} + 4H^2 \widehat{K}_i \widehat{K}_j \widehat{K}_k,
\end{aligned}
$$

where $\tilde{\mathbf{t}}_i$ denotes the set of all events up to time T in dimension i and H is a hyper-parameter used to truncate the interval $(-\infty, \infty)$ to $[-H, H]$. See (Achab et al., 2017) for a proof of the consistency of these estimators.

C. Reproducibility

In this section, we provide details of the experimental setup used to produce the figures reported in the main text. The open-source code and all the scripts used to generate the figures are released publicly.

C.1. Synthetic data

The experimental setup for the experiments on synthetic data is as follows. We simulated 20 datasets, each comprised of 5 realizations of 10^5 events. We then randomly translated each dataset with distributions $F_i \sim \mathcal{N}(0, \sigma^2),\ 1 \leq i \leq d$, for 20 noise powers σ^2 ranging from 0.1 to 25, sampled in log-space.

The hyper-parameters used to produce the figures in Section 5 can be found in notebook and scripts. Details are as follows.

- ADM4. The exponential decay was set to its ground-truth value, $\beta = 1$.
- Desync-MLE. Similar to ADM4, the exponential decay was set to its ground-truth value, $\beta = 1$.
- NPHC. The bandwidth H of the estimator of the cumulants was set using binary search to minimize the $L_{2,2}$ distance to the ground-truth cumulants. The resulting bandwidth used to run the algorithm are discussed in Figure 9. In short,

we observed that the for noise powers $\sigma^2 < 1/\beta = 1$, the best bandwidth H remained stable. For $\sigma^2 < 1/\beta = 1$ the best bandwidth H increased linearly with σ^2.

- WH. The maximum support of the excitation matrix was set to 10.0 to roughly match the same scale as the ground-truth excitation functions. The number of quadrature points was set to 20. This value, which has a quadratic cost in the computational complexity of the algorithm, was found to be large enough to provide prefect PR-AUC and Precision@n on noiseless observations.

C.2. Real data

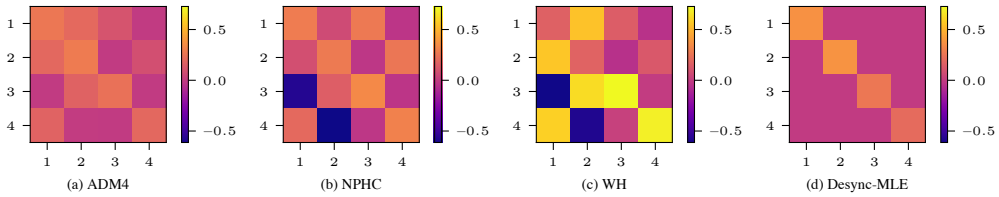
The dataset used in the experiments is that of Bund Futures traded at Eurex over 20 days in April 2014 [8]. This dataset has already been modeled using MHPs in (Bacry et al., 2016). Each day is considered an independent realization of the process. The timestamps are recorded at the microsecond timestamp resolution. As explained on the download website, the data was preprocessed as follows. Market opens at 8AM which corresponds to a timestamp of $28\,800$. This timestamp has been subtracted to all timestamps to have a realizations that starts at time 0. As markets closes at 10PM, the end time of the realizations is $50\,400$. No additional preprocessing was performed on the dataset.

The first 5 days were used to tune the hyper-parameters of the learning algorithms, and the remaining 15 days were used to measure the performance reported in Figure 7.

The hyper-parameters used to produce the figures in Section 5 can be found in notebook and scripts. Details are as follows.

- ADM4. To set the exponential decay β of the excitation functions, we ran a grid search for values between 10^2 and 10^4, and we found that $\beta = 1291.0$ maximized the log-likelihood. We used grid search between 1 and 10^5 to tune the regularization weight and found that 10^3 maximized the log-likelihood.
- NPHC. Following the observation made in the experiments on synthetic data, the bandwidth H of the estimator of the cumulants, was set to $H = 1/\beta + \sigma^2$. This value provided stable results.
- WH. The hyper-parameters of this algorithm were set as in (Bacry et al., 2016).

We were unable to reproduce the results of (Bacry et al., 2016) on the noiseless dataset (*i.e.*, without added random translation) using Desync-MLE. Since the dataset consists of timestamps in a high-frequency trading application, it is not expected to hold any synchronization noise. However, as shown in Figure 10, Desync-MLE converged to a non-zero synchronization noise with a diagonal excitation matrix.



(a) ADM4 (b) NPHC (c) WH (d) Desync-MLE

Figure 10: Excitation matrices $\hat{\boldsymbol{G}}_0$ learned by the different learning methods on the noiseless dataset of Bund Futures traded at Eurex.

[8] Dataset available at: https://github.com/X-DataInitiative/tick-datasets/