Table 1: End-to-end training time for neural TPPs and their corresponding Residual TPPs. The first row, "MHP + RED", reports the combined runtime for Steps 1 and 2, representing the mean computational time across 10 independent trials. For each Residual TPP, the training parameters are kept the same as in the baseline model, and the result reflects the total time cost of the 3-step procedure (Hawkes fitting + RED filtering + neural TPP training). All training was conducted on a CPU.

Model	End-to-end Runtime (seconds)					
	MIMIC-II	Retweet	Earthquake	StackOverflow	Amazon	Volcano
MHP+RED	1.42	0.27	0.11	0.84	1.49	0.02
RMTPP	**9.70**	195.3	26.00	71.20	95.75	34.20
Res RMTPP	9.72	**188.7**	**20.01**	**70.44**	**95.19**	**26.42**
NHP	34.70	492.6	25.32	206.3	256.6	77.05
Res NHP	**29.52**	**441.9**	**20.57**	**190.8**	**245.1**	**54.47**
SAHP	128.5	498.4	24.90	498.0	505.8	37.30
Res SAHP	**113.4**	**434.9**	**16.96**	**461.9**	**450.3**	**24.42**
THP	**10.45**	1183	19.20	81.20	257.7	39.90
Res THP	10.67	**1029**	**17.77**	**76.44**	**244.8**	**29.82**
AttNHP	68.60	9475	162.2	1924	6375	1093
Res AttNHP	**52.92**	**7195**	**143.1**	**1863**	**5646**	**657.5**
ODETPP	9.96	106.4	42.00	243.6	196.4	51.96
Res ODETPP	**9.64**	**102.7**	**37.43**	**224.2**	**192.1**	**40.88**

Table 2: Statistics of the simulated datasets. (1) *Poisson-based* dataset represents sequences following a periodic non-homogeneous Poisson process with homo Poisson anomalies. (2) *AttNHP-based* dataset represents TPP sequences generated by the AttNHP model, with homo Poisson anomalies added randomly. (3) *Poisson+AttNHP* dataset represents sequences following a periodic non-homo Poisson process, with commission events generated by AttNHP. The table columns, from left to right, show the dataset, number of event types, totoal number of events, number of residual events generated by $\lambda^{(2)}$, sequence length, and number of sequences.

Dataset	# Types	# Events	# Residuals	Sequence Length			# Sequences		
				Min	Mean	Max	Train	Valid	Test
Possion-based	5	33,243	4,253	20	29	100	600	200	200
AttNHP-based	5	100,052	12,807	9	44	79	1,000	500	500
Possion+AttNHP	5	54,737	16,625	24	55	126	600	200	200

Table 3: Performance comparison of Residual TPPs on simulated datasets. We evaluate the model's *goodness-of-fit* (Log-Likelihood, higher is better) and *prediction performance* (Time RMSE / Type Error Rate, lower is better).

Model	Possion-based		AttNHP-based		Possion+AttNHP	
	L-L	Time/Type	L-L	Time/Type	L-L	Time/Type
MHP	-0.132	0.298/62.2%	-1.435	0.329/79.4%	-0.203	0.257/68.1%
RMTPP	-1.030	0.337/63.2%	-1.277	0.373/78.5%	-0.788	0.271/68.6%
Res RMTPP	**-0.531**	**0.305/62.3%**	**-0.664**	**0.334/77.9%**	**-0.027**	**0.237/67.8%**
NHP	-1.023	0.343/62.1%	-1.267	0.374/77.4%	-0.776	0.277/67.8%
Res NHP	**-0.527**	**0.301/62.0%**	**-0.661**	**0.334/77.1%**	**-0.009**	**0.237**/67.8%
SAHP	-1.033	0.330/62.1%	-1.267	0.376/77.5%	-0.789	0.282/67.8%
Res SAHP	**-0.556**	**0.303**/62.1%	**-0.660**	**0.335/77.3%**	**-0.002**	**0.238**/67.8%
THP	-1.026	0.336/62.6%	-1.270	0.374/77.6%	-0.785	0.266/67.8%
Res THP	**-0.566**	**0.300/62.0%**	**-0.655**	**0.331**/77.6%	**0.004**	**0.234**/67.8%
AttNHP	-1.026	0.337/62.1%	-1.266	0.374/77.3%	-0.789	0.269/67.8%
Res AttNHP	**-0.534**	**0.302**/62.1%	**-0.658**	**0.336/77.2%**	**0.001**	**0.237**/67.8%
ODETPP	-1.028	0.339/62.1%	-1.267	0.382/77.6%	-0.789	0.270/67.8%
Res ODETPP	**-0.565**	**0.302**/62.1%	**-0.663**	**0.332**/77.6%	**-0.001**	**0.238**/67.8%