Table 1: Performance comparison of Residual TPPs with RED method using different base models on example benchmark datasets: MIMIC-II and Earthquake. For each baseline neural TPP, the first row reports its original performance, the second row (e.g. Res RMTPP) shows Residual TPP with RED using Hawkes as base model and the third row (e.g. NHP+RMTPP) presents the new Residual TPP with RED using NHP as base model. 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	MIMIC-II		Earthquake	
	L-L	Time/Type	L-L	Time/Type
MHP	-2.839	0.925/27.9%	-4.155	1.475/60.6%
MHP+MHP	-3.883	0.919/43.0%	-3.746	1.452/71.8%
RMTPP	-2.626	0.998/37.8%	-4.643	1.956/52.9%
Res RMTPP	**-2.045**	**0.915/26.7%**	**-3.689**	**1.420/52.7%**
NHP+RMTPP	-2.059	0.929/37.8%	-4.098	1.797/52.8%
NHP	-2.031	1.010/26.7%	-2.389	1.910/53.9%
Res NHP	-1.825	**0.913/20.9%**	**-1.930**	**1.416/52.8%**
NHP+NHP	**-1.792**	1.061/22.1%	-2.089	1.711/52.8%
SAHP	-4.672	0.971/23.8%	-3.338	1.463/54.1%
Res SAHP	-4.488	**0.935/18.0%**	-3.335	1.455/53.3%
NHP+SAHP	**-4.126**	0.935/20.9%	**-3.152**	**1.452/52.8%**
THP	-2.048	1.129/35.5%	-3.498	1.857/54.7%
Res THP	**-2.040**	**0.930/27.9%**	**-3.415**	**1.403/52.8%**
NHP+THP	-2.743	1.067/35.5%	-4.112	1.755/53.0%
AttNHP	-2.500	1.030/35.4%	-2.896	1.822/54.5%
Res AttNHP	**-2.197**	**0.932**/36.0%	-3.147	**1.413/52.8%**
NHP+AttNHP	-2.649	1.094/**23.8%**	**-2.133**	1.757/52.8%
ODETPP	-1.855	1.416/22.1%	-2.203	2.396/56.0%
Res ODETPP	**-1.371**	**0.934/19.2%**	-2.340	**1.412/53.0%**
NHP+ODETPP	-2.480	1.263/22.7%	**-2.089**	2.099/54.8%