

Supplementary Information

To make our experiments more complete, we validate some rhythm-based and stability-based metrics using real-world dataset. Three metrics mentioned in related work are calculated, which are Regularity Index (RI) [6], Inter-daily Stability (IS) [38], and Intra-daily Variability (IV) [38]. RI assesses the difference between the same time points across two different days. Activity data of each individual are rescaled to $[-0.5, 0.5]$ firstly. The similarity of two different days is the product of two rescaled values. Positive production means they are close and negative means they are not similar. Then RI is calculated as the average similarity of all combinations of two different days. IS reflects the stability of rest-activity rhythms over multiple days. It is calculated as the ratio of the variance within the same time interval each day and the overall variance. Higher IS values reflect more regular patterns of activity. IV detects fragmentation of rest-activity rhythms. It is calculated as the ratio of the mean squared the first deviation of the data and the population variance of the data. Higher IV values indicate more rest-activity alternations. We calculate these metrics of samples in SJTU dataset and associate them with prediction errors from different forecasting models. The correlation coefficients are presented in table S1. Compared to entropy models, these rhythm-based and stability-based metrics have much lower correlation coefficients, and all of them do not exceed 0.2. Such results indicate that rhythm-based and stability-based metrics can not serve as an appropriate measurement for human physical activity.

Table S1: The correlation coefficients between different metrics and prediction errors from different forecasting models.

Models	ES	Prophet	RNN	N-BEATS
RI	-0.0190	0.0834	-0.1022	-0.1085
IS	-0.0533	0.0962	0.1475	0.0750
IV	-0.1886	-0.0944	-0.1149	-0.1425