Traffic dataset, CP_kernel, 24-steps non-overlapping, 29-th time series, total epochs: 300 --- Source: 0 : t₀ Target: $t_0 + 1 : t_0 + 24$ Prediction: $t_0 + 1 : t_0 + 24$ Attention point: $t_0 + 1$?1 Attention layer 1 for $t_0 + 1$?0.4 ?0.5 ?0.6 ?0.7 Attention layer 2 for $t_0 + 1$?0.4 ?0.5 ?0.6 ?0.7 Attention layer 3 for $t_0 + 1$?0.50 ?0.55 ?0.60 ?0.65 ?0.70

48000

50000

52000

?0.75

42000

44000

46000