Traffic dataset, linear\_kernel, 24-steps non-overlapping, 29-th time series, total epochs: 300 --- Source:  $0:t_0$ --- 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 0.2 -0.0 ?0.2 ?0.4 ?0.6 Attention layer 2 for  $t_0 + 1$ 0.2 0.0 ?0.2 Attention layer 3 for  $t_0 + 1$ 0.50 0.25 0.00 ?0.25

48000

50000

52000

?0.50

42000

44000

46000