n-steps non-overlapping, total epochs: 50 1.0 -0.5 0.0 Source: $0:t_0$ --- Target: $t_0 + 1: t_0 + 100$ Prediction: $t_0 + 1 : t_0 + 100$ Attention point: $t_0 + 1$?1.0 -Attention pattens for $t_0 + 1$ 0.035 attn_layer_1 attn_layer_2 0.030 attn_layer_3 0.025 0.020 0.015 -0.010 -

200

150

250

300

0.005 -

0.000 -

100