n-steps non-overlapping, total epochs: 38 3.75 3.50 -3.25 3.00 -2.75 Source: $0:t_0$ 2.50 ---- Target: $t_0 + 1: t_0 + 100$ Prediction: $t_0 + 1 : t_0 + 100$ 2.25 • Attention point: $t_0 + 1$ Attention pattens for $t_0 + 1$ 0.10 attn_layer_1 attn_layer_2 attn_layer_3 0.05 0.00 -?0.05 ?0.10 -

200

150

100

250

300