n-steps non-overlapping, total epochs: 42 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 $t_0 + 1$?0.076 attn_layer_1 attn_layer_2 ?0.078 attn_layer_3 ?0.080 ?0.082 ?0.084 ?0.086 -

200

100

150

250

300