n-steps non-overlapping, total epochs: 49 4.5 4.0 3.5 3.0 -2.5 -**---** Source: $0:t_0$ --- Target: $t_0 + 1: t_0 + 100$ 2.0 Prediction: $t_0 + 1 : t_0 + 100$ Attention point: $t_0 + 1$ 1.5 -Attention pattens for $t_0 + 1$ 0.1 attn_layer_1 attn_layer_2 0.0 attn_layer_3 ?0.1 ?0.2 -?0.3 -?0.4 ?0.5 -?0.6 -?0.7 -

200

150

100

250

300