n-steps non-overlapping, total epochs: 275 4.00 3.75 -3.50 -3.25 -3.00 -2.75 Source:  $0:t_0$ 2.50 ---- Target:  $t_0 + 1: t_0 + 100$ 2.25 -Prediction:  $t_0 + 1 : t_0 + 100$ Attention point:  $t_0 + 1$ 2.00 Attention pattens for  $t_0 + 1$ 0.08 attn\_layer\_1 attn\_layer\_2 0.07 attn\_layer\_3 0.06 0.05 0.04 0.03 -0.02 -0.01 0.00

200

100

150

250

300