n-steps non-overlapping, total epochs: 44 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  $t_0 + 1$ ?0.03 attn\_layer\_1 attn\_layer\_2 ?0.04 attn\_layer\_3 ?0.05 -?0.06 -?0.07 -?0.08 ?0.09 ?0.10 -

200

150

100

250

300