n-steps non-overlapping, total epochs: 291 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  $t_0 + 1$ 0.012 attn\_layer\_1 attn\_layer\_2 attn\_layer\_3 0.010 0.008 -0.006 0.004 -

200

150

100

250

300