n-steps non-overlapping, total epochs: 185 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 $t_0 + 1$ attn_layer_1 0.035 attn_layer_2 0.030 attn_layer_3 0.025 -0.020 -0.015 -0.010 -0.005 -0.000 250 300 100 150 200