n-steps non-overlapping, total epochs: 236 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 $t_0 + 1$ 0.06 attn_layer_1 attn_layer_2 0.05 attn_layer_3 0.04 0.03 -0.02 -0.01 0.00 250 300 100 150 200