n-steps non-overlapping, total epochs: 68 3.75 3.50 -3.25 3.00 -2.75 Source: 0 : *t*₀ 2.50 ---- Target: $t_0 + 1: t_0 + 100$ Prediction: $t_0 + 1 : t_0 + 100$ 2.25 • Attention point: $t_0 + 1$ Attention pattens for $t_0 + 1$ 0.00735 attn_layer_1 attn_layer_2 0.00730 attn_layer_3 0.00725 -0.00720 -0.00715 0.00710 -0.00705 0.00700 200 250 300 150 100