Loss Convergence Curve - Transformer_Seq2Seq trained with Time Window Length 1 back and 20 forward