

---

# Shape and Time Distortion Loss for Training Deep Time Series Forecasting Models

## Supplementary Material

---

Vincent Le Guen<sup>1,2</sup>  
vincent.le-guen@edf.fr

Nicolas Thome<sup>2</sup>  
nicolas.thome@cnam.fr

(1) EDF R&D  
6 quai Watier, 78401 Chatou, France

(2) CEDRIC, Conservatoire National des Arts et Métiers  
292 rue Saint-Martin, 75003 Paris, France

## 1 Experimental setup

### 1.1 Metrics

We describe here the two external metrics used in our experiments to evaluate the shape and timing errors.

**Ramp score** The notion of *ramping event* is a major issue for intermittent renewable energy production that needs to be anticipated for electricity grid management. For assessing the performance of trained forecasting models in presence of ramps, the Ramp Score was proposed in (VCP<sup>+</sup>17). This score is based on a piecewise linear approximation on both input and target time series by the Swinging Door algorithm (Bri90; FHO13). The Ramp Score described in (VCP<sup>+</sup>17) is computed as the integral between the unsigned difference of derivatives of both linear approximated series. For assessing only the shape error component, we apply in our experiments the ramp score on the target and prediction series after alignment by the optimal DTW path.

**Hausdorff distance** Given a set of change points  $\mathcal{T}^*$  in the target signal and change points  $\hat{\mathcal{T}}$  in the predicted signal, the Hausdorff distance is defined as:

$$\text{Hausdorff}(\mathcal{T}^*, \hat{\mathcal{T}}) := \max(\max_{\hat{t} \in \hat{\mathcal{T}}} \min_{t^* \in \mathcal{T}^*} |\hat{t} - t^*|, \max_{t^* \in \mathcal{T}^*} \min_{\hat{t} \in \hat{\mathcal{T}}} |\hat{t} - t^*|) \quad (1)$$

It corresponds to the greatest temporal distance between a change point and its prediction.

We now explain how the change points are computed for each dataset: for Synthetic, we know exactly by construction the positions of the change points in the target signals. For the predictions, we look for a single change point corresponding to the location of the predicted step function. We use the exact segmentation method by dynamic programming described in (TOV18) with the Python toolbox <http://ctruong.perso.math.cnrs.fr/ruptures-docs/build/html/index.html#>.

For ECG5000 and Traffic datasets which present sharp peaks, this change point detection algorithm is not suited (detected change points are often located at the inflexion points of peaks and not at the exact peak location). We thus use a simple peak detection algorithm based on first order finite differences. We tune the threshold parameter for outputting a detection and the min distance between detections parameter experimentally for each dataset.

## 2 Experiments

We provide here additional qualitative predictions for all datasets.

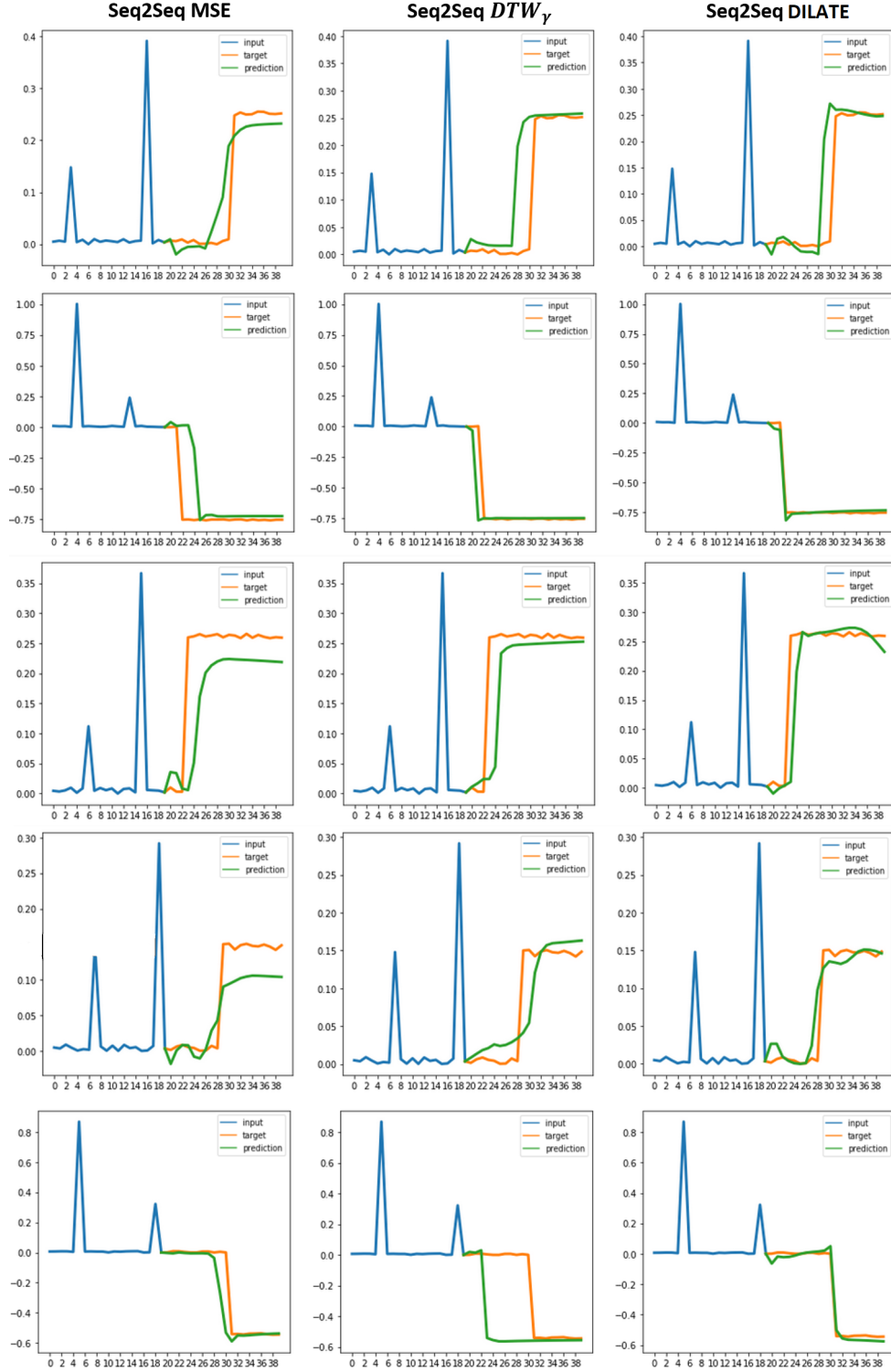


Figure 1: Qualitative predictions for the Synthetic dataset.

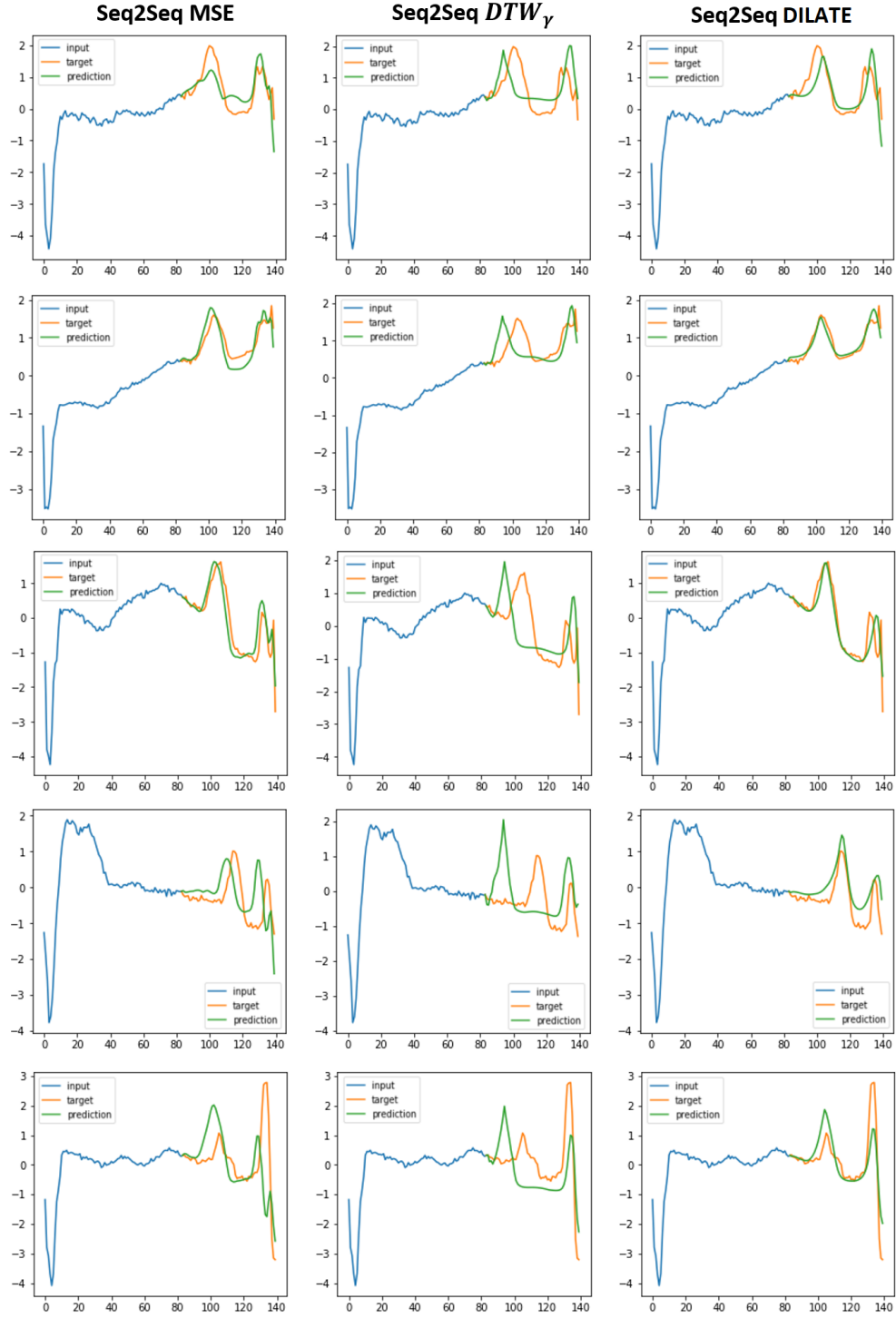


Figure 2: Qualitative predictions for the ECG5000 dataset.

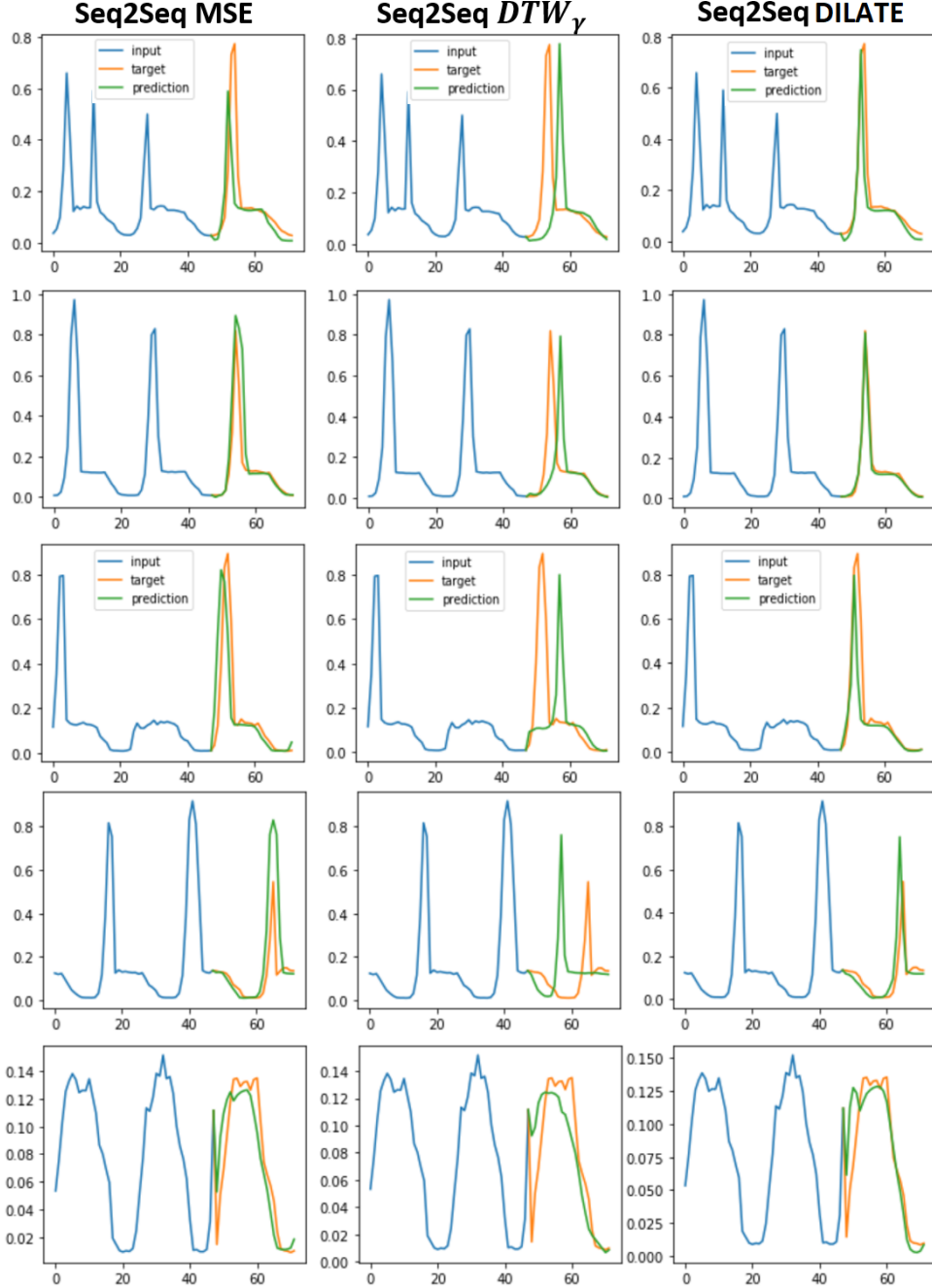


Figure 3: Qualitative predictions for the Traffic dataset.

### 3 Comparison to the temporally constrained variants of DTW

To understand the difference of behaviour between DILATE and its variant  $\text{DILATE}^t\text{-W}$ , we build a synthetic example by starting with a target step function as in Fig 4 (b) and a perfect step prediction. Then we translate progressively the step prediction and we compute three error metrics (Fig 4 (a)). As expected, the DTW error remains zero since we only have temporal shift.  $\text{DILATE}$  and  $\text{DILATE}^t\text{-W}$  increasingly penalize the temporal shift but the  $\text{DILATE}^t\text{-W}$  curve is flatter around 0 (lower gradient for the small temporal shifts). This suggests that DILATE is better suited to recover finer temporal

alignment, as the quantitative results in Table 1. We provide as illustration two experimental results on Fig 4 (b) and (c) on the Synthetic dataset: DILATE leads to better temporal localization.

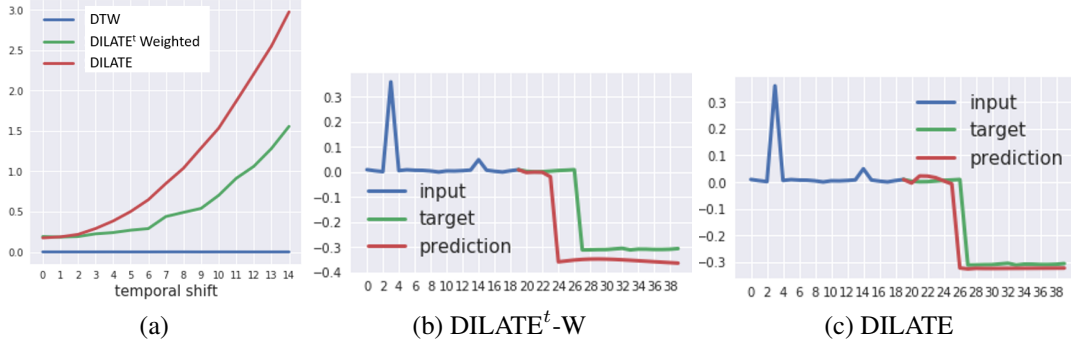


Figure 4: Qualitative comparison between DILATE and DILATE<sup>t</sup>-W.

We report in Table 1 the full quantitative results on the 3 datasets:

Dataset	Eval loss		DILATE (ours)	DILATE <sup>t</sup> -W (JJO11)	DILATE <sup>t</sup> BC (SC90)
Synth	Euclidian	MSE ( $\times 100$ )	<b>1.21 <math>\pm</math> 0.130</b>	1.36 $\pm$ 0.107	1.83 $\pm$ 0.163
	Shape	DTW ( $\times 100$ )	23.1 $\pm$ 2.44	<b>20.5 <math>\pm</math> 2.49</b>	<b>21.6 <math>\pm</math> 1.74</b>
		Ramp	<b>4.99 <math>\pm</math> 0.460</b>	<b>5.56 <math>\pm</math> 0.87</b>	<b>5.23 <math>\pm</math> 0.439</b>
	Time	TDI ( $\times 10$ )	<b>14.8 <math>\pm</math> 1.29</b>	17.8 $\pm$ 1.72	19.6 $\pm$ 1.72
		Hausdorff	<b>2.70 <math>\pm</math> 0.16</b>	<b>2.85 <math>\pm</math> 0.210</b>	3.30 $\pm$ 0.273
ECG	Euclidian	MSE ( $\times 100$ )	<b>30.3 <math>\pm</math> 4.10</b>	<b>31.1 <math>\pm</math> 4.51</b>	66.2 $\pm$ 11.4
	Shape	DTW ( $\times 10$ )	<b>16.1 <math>\pm</math> 0.156</b>	17.9 $\pm$ 2.09	17.6 $\pm$ 0.20
		Ramp	<b>4.80 <math>\pm</math> 0.249</b>	<b>4.63 <math>\pm</math> 0.307</b>	<b>4.49 <math>\pm</math> 0.255</b>
	Time	TDI ( $\times 10$ )	<b>6.59 <math>\pm</math> 0.786</b>	8.08 $\pm$ 0.845	16.8 $\pm$ 3.09
		Hausdorff	4.23 $\pm$ 0.414	<b>3.56 <math>\pm</math> 0.305</b>	6.81 $\pm$ 1.41
Traffic	Euclidian	MSE ( $\times 100$ )	<b>1.00 <math>\pm</math> 0.260</b>	<b>1.06 <math>\pm</math> 0.12</b>	1.87 $\pm$ 0.25
	Shape	DTW ( $\times 100$ )	<b>23.0 <math>\pm</math> 1.62</b>	<b>22.2 <math>\pm</math> 1.07</b>	<b>22.6 <math>\pm</math> 2.01</b>
		Ramp	5.93 $\pm$ 0.235	5.92 $\pm$ 0.177	<b>5.63 <math>\pm</math> 0.46</b>
	Time	TDI ( $\times 10$ )	<b>14.4 <math>\pm</math> 1.58</b>	<b>15.5 <math>\pm</math> 1.50</b>	17.6 $\pm$ 0.23
		Hausdorff	<b>2.13 <math>\pm</math> 0.514</b>	<b>1.79 <math>\pm</math> 0.437</b>	<b>2.08 <math>\pm</math> 0.331</b>

Table 1: Comparison to the tangled variants of DILATE. Results averaged over 10 runs (mean  $\pm$  standard deviation). For each experiment, best method(s) (Student t-test) in bold.

## References

- [Bri90] EH Bristol, *Swinging door trending: Adaptive trend recording?*, ISA National Conf. Proc., 1990, 1990, pp. 749–754.
- [FHO13] Anthony Florita, Bri-Mathias Hodge, and Kirsten Orwig, *Identifying wind and solar ramping events*, 2013 IEEE Green Technologies Conference (GreenTech), IEEE, 2013, pp. 147–152.
- [JJO11] Young-Seon Jeong, Myong K Jeong, and Olufemi A Omitaomu, *Weighted dynamic time warping for time series classification*, Pattern Recognition **44** (2011), no. 9, 2231–2240.
- [SC90] Hiroaki Sakoe and Seibi Chiba, *Dynamic programming algorithm optimization for spoken word recognition*, Readings in speech recognition **159** (1990), 224.
- [TOV18] Charles Truong, Laurent Oudre, and Nicolas Vayatis, *A review of change point detection methods*, arXiv preprint arXiv:1801.00718 (2018).
- [VCP<sup>+</sup>17] Loïc Vallance, Bruno Charbonnier, Nicolas Paul, Stéphanie Dubost, and Philippe Blanc, *Towards a standardized procedure to assess solar forecast accuracy: A new ramp and time alignment metric*, Solar Energy **150** (2017), 408–422.