Re-framing Time Series Augmentation Through the Lens of Generative Models

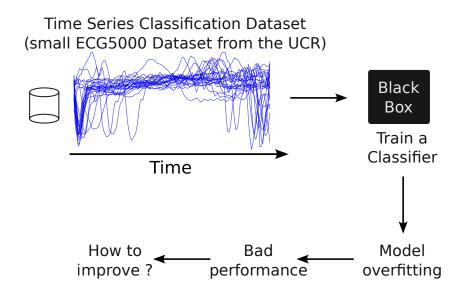
10th Workshop on Advanced Analytics and Learning on Temporal Data - ECML-PKDD 2025

Ali Ismail-Fawaz¹, Maxime Devanne¹, Stefano Berretti², Jonathan Weber¹, Germain Forestier^{1,3}

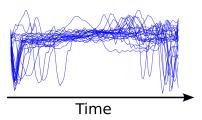
¹IRIMAS, Université Haute-Alsace, Mulhouse, France ²MICC, University of Florence, Florence, Italy ³DSAI, Monash University, Melbourne Australia

September 19, 2025

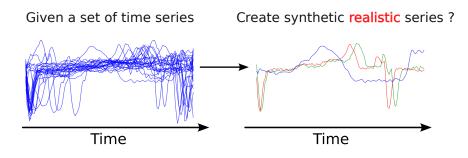
Time Series Classification on Small Datasets



Given a set of time series



Dau, H.A., Bagnall, A., Kamgar, K., Yeh, C.C.M., Zhu, Y., Gharghabi, S., Ratanamahatana, C.A. and Keogh, E., 2019. The UCR time series archive. IEEE/CAA Journal of Automatica Sinica, 6(6), pp.1293-1305.



Dau, H.A., Bagnall, A., Kamgar, K., Yeh, C.C.M., Zhu, Y., Gharghabi, S., Ratanamahatana, C.A. and Keogh, E., 2019. The UCR time series archive. IEEE/CAA Journal of Automatica Sinica, 6(6), pp.1293-1305.

Given a set of time series

Create synthetic realistic series ?

Time

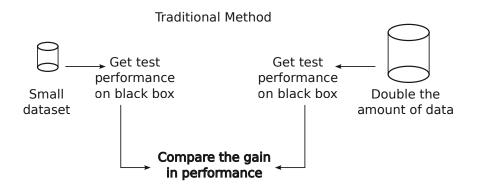
- What makes a synthetic time series realistic ?
- Dau, H.A., Bagnall, A., Kamgar, K., Yeh, C.C.M., Zhu, Y., Gharghabi, S., Ratanamahatana, C.A. and Keogh, E., 2019. The UCR time series archive. IEEE/CAA Journal of Automatica Sinica, 6(6), pp.1293-1305.

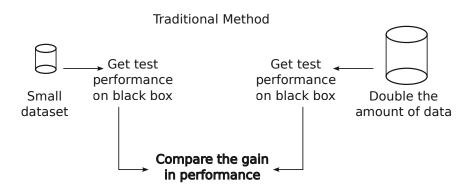
Create synthetic realistic series? Time Time

What makes a synthetic time series realistic ?

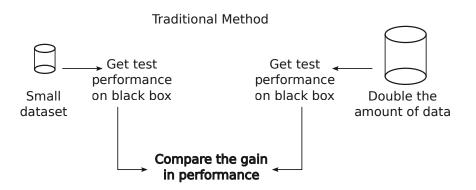
Given a set of time series

- Bigger question: How to evaluate data augmentation?
- Dau, H.A., Bagnall, A., Kamgar, K., Yeh, C.C.M., Zhu, Y., Gharghabi, S., Ratanamahatana, C.A. and Keogh, E., 2019. The UCR time series archive. IEEE/CAA Journal of Automatica Sinica, 6(6), pp.1293-1305.

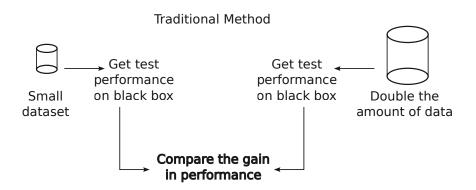




Computationally incomprehensive

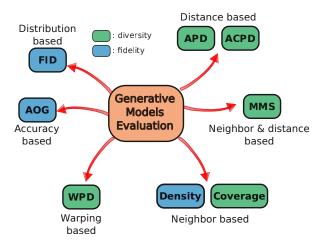


- Computationally incomprehensive
- Why not look at data augmentation as a Generative Model ?



- Computationally incomprehensive
- Why not look at data augmentation as a Generative Model ?
- But how to evaluation Generative Models without relying on performance gain ?

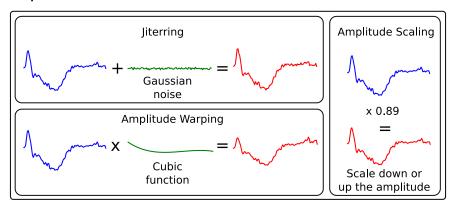
Generative Models Evaluation Measures



Ismail-Fawaz, A., Devanne, M., Berretti, S., Weber, J. and Forestier, G., 2025. Establishing a unified evaluation framework for human motion generation: A comparative analysis of metrics. Computer Vision and Image Understanding, 254, p.104337.

Data Augmentation for Time Series Classification

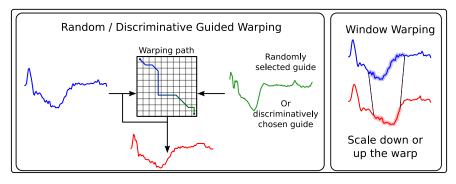
Simple methods



Um, T.T., Pfister, F.M., Pichler, D., Endo, S., Lang, M., Hirche, S., Fietzek, U. and Kulić, D., 2017, November. Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. In Proceedings of the 19th ACM international conference on multimodal interaction (pp. 216-220).

Data Augmentation for Time Series Classification

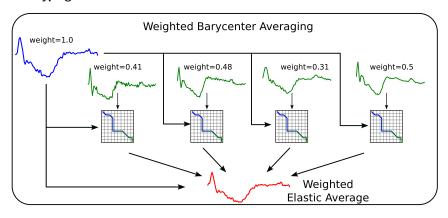
Warping methods



- Iwana, B.K. and Uchida, S., 2021, January. Time series data augmentation for neural networks by time warping with a discriminative teacher. In 2020 25th International Conference on Pattern Recognition (ICPR) (pp. 3558-3565). IEEE.
- Le Guennec, A., Malinowski, S. and Tavenard, R., 2016, September. Data augmentation for time series classification using convolutional neural networks. In ECML/PKDD workshop on advanced analytics and learning on temporal data.

Data Augmentation for Time Series Classification

Prototyping methods



Forestier, G., Petitjean, F., Dau, H.A., Webb, G.I. and Keogh, E., 2017, November. Generating synthetic time series to augment sparse datasets. In 2017 IEEE international conference on data mining (ICDM) (pp. 865-870). IEEE.

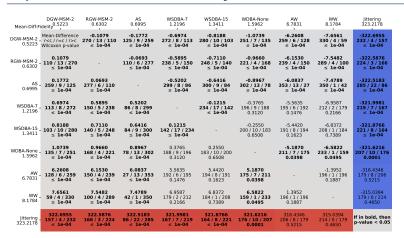
Experimental Setup

- Amplitude Warping (AW), Amplitude Scaling (AS),
 Window Warping (WW) and Jittering fixed parameters
- Random and Discriminative Guided Warping (RGW and DGW)
 - Dynamic Time Warping (DTW) with and without window constraint
 - Move Split Merge (MSM) with different cost value
 - ShapeDTW with different reach value
- Weighted Barycenter Averaging
 - with DTW BA (DBA) [1] with and without windowing
 - with MSM BA (MBA) [2] with different cost value
 - with ShapeDTW BA (ShapeDBA) [3] with different reach value
- 131 univariate Time Series Classification datasets (test sets were not used in this work)
- All series go through z-normalization prior to synthetic generation
- All experiments are repeated multiple times and averaged over multiple seeds to remove any possible bias
- [1] Petitjean, F., Ketterlin, A. and Gançarski, P., 2011. A global averaging method for dynamic time warping, with applications to clustering. Pattern recognition, 44(3), pp.678-693.
- [2] Holder, C., Guijo-Rubio, D. and Bagnall, A.J., 2023. Barycentre Averaging for the Move-Split-Merge Time Series Distance Measure. In KDIR (pp. 51-62).
- [3] Ismail-Fawaz, A., Ismail Fawaz, H., Petitjean, F., Devanne, M., Weber, J., Berretti, S., Webb, G.I. and Forestier, G., 2023, September. Shapedba: Generating effective time series prototypes using shapedtw barycenter averaging. In International Workshop on Advanced Analytics and Learning on Temporal Data (pp. 127-142). Cham: Springer Nature Switzerland.

Experimental Results - Fidelity

Mean-Diff-F	DGW-MSM-2 delity ^{0.5223}	RGW-MSM-2 0.6302	AS 0.6995	WSDBA-7 1.2196	WSDBA-15 1.3411	WDBA-None 1.5962	AW 6.7831	WW 8.1784	Jittering 323.2178
DGW-MSM-2	Moon Difference	-0.1079	-0.1772 125 / 9 / 259 ≤ 1e-04	-0.6974 272 / 8 / 113 ≤ 1e-04	-0.8188 280 / 10 / 103 ≤ 1e-04	-1.0739 251 / 7 / 135 ≤ 1e-04	-6.2608 259 / 6 / 128 ≤ 1e-04	-7.6561 330 / 4 / 59 ≤ 1e-04	-322.6955 232 / 4 / 157 ≤ 1e-04
RGW-M5M-2 0.6302	0.1079 110 / 13 / 270 ≤ 1e-04	-	-0.0693 110 / 6 / 277 ≤ 1e-04	-0.5895 238 / 5 / 150 ≤ 1e-04	-0.7110 248 / 5 / 140 ≤ 1e-04	-0.9660 221 / 4 / 168 ≤ 1e-04	-6.1530 239 / 4 / 150 ≤ 1e-04	-7.5482 289 / 4 / 100 ≤ 1e-04	-322.5876 224 / 3 / 166 ≤ 1e-04
AS 0.6995	0.1772 259 / 9 / 125 ≤ 1e-04	0.0693 277 / 6 / 110 ≤ 1e-04	-	-0.5202 299 / 8 / 86 ≤ 1e-04	-0.6416 300 / 9 / 84 ≤ 1e-04	-0.8967 302 / 13 / 78 ≤ 1e-04	-6.0837 353 / 13 / 27 ≤ 1e-04	-7.4789 350 / 1 / 42 ≤ 1e-04	-322.5183 285 / 22 / 86 ≤ 1e-04
WSDBA-7 1.2196	0.6974 113 / 8 / 272 ≤ 1e-04	0.5895 150 / 5 / 238 ≤ 1e-04	0.5202 86 / 8 / 299 ≤ 1e-04		-0.1215 234 / 17 / 142 ≤ 1e-04	-0.3765 196 / 9 / 188 0.3120	-5.5635 195 / 6 / 192 0.1476	-6.9587 212 / 2 / 179 0.2166	-321.9981 219 / 7 / 167 ≤ 1e-04
WSDBA-15 1.3411	0.8188 103 / 10 / 280 ≤ 1e-04	0.7110 140 / 5 / 248 ≤ 1e-04	0.6416 84/9/300 ≤ 1e-04	0.1215 142 / 17 / 234 ≤ 1e-04		-0.2550 200 / 10 / 183 0.6508	-5.4420 191 / 8 / 194 0.1623	-6.8372 208 / 1 / 184 0.7389	-321.8766 221 / 8 / 164 ≤ 1e-04
WDBA-None 1.5962	1.0739 135 / 7 / 251 ≤ 1e-04	0.9660 168 / 4 / 221 ≤ 1e-04	0.8967 78 / 13 / 302 ≤ le-04	0.3765 188 / 9 / 196 0.3120	0.2550 183 / 10 / 200 0.6508	-	-5.1870 211 / 7 / 175 0.0398	-6.5822 233 / 1 / 159 0.0495	-321.6216 207 / 10 / 176 0.0001
AW 6.7831	6.2608 128 / 6 / 259 ≤ 1e-04	6.1530 150 / 4 / 239 ≤ 1e-04	6.0837 27 / 13 / 353 ≤ 1e-04	5.5635 192 / 6 / 195 0.1476	5.4420 194 / 8 / 191 0.1623	5.1870 175 / 7 / 211 0.0398		-1.3952 196 / 1 / 196 0.1887	-316.4346 179 / 8 / 206 0.5215
WW 8.1784	7.6561 59 / 4 / 330 ≤ 1e-04	7.5482 100 / 4 / 289 ≤ 1e-04	7.4789 42 / 1 / 350 ≤ 1e-04	6.9587 179 / 2 / 212 0.2166	6.8372 184 / 1 / 208 0.7389	6.5822 159 / 1 / 233 0.0495	1.3952 196 / 1 / 196 0.1887		-315.0394 179 / 0 / 214 0.4650
Jittering . 323.2178	322.6955 157 / 4 / 232 ≤ 1e-04	322.5876 166 / 3 / 224 ≤ 1e-04	322.5183 86 / 22 / 285 ≤ 1e-04	321.9981 167 / 7 / 219 ≤ 1e-04	321.8766 164 / 8 / 221 ≤ 1e-04	321.6216 176 / 10 / 207 0.0001	316.4346 206 / 8 / 179 0.5215	315.0394 214 / 0 / 179 0.4650	If in bold, then p-value < 0.05

Experimental Results - Fidelity



- DGW / RGW with MSM high cost value is the most fidelious approach
- WSDBA outperforms WDBA and WMBA
- AS could be promising
- · Jittering completely fails in terms of fidelity

300

200

100

-100

-200

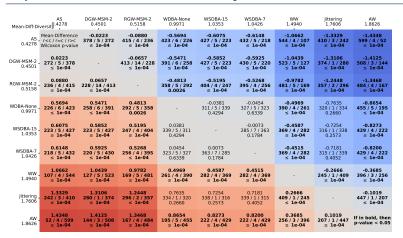
-300

0

Experimental Results - Diversity

Mean-Diff-Di	AS versity 4278	DGW-MSM-2 0.4501	RGW-MSM-2 0.5158	WDBA-None 0.9971	WSDBA-15 1.0353	WSDBA-7 1.0426	WW 1.4940	Jittering 1.7606	AW 1.8626
AS - 0.4278	Mean-Difference r <c r="">c/r=c/r>c Wilcoxon p-value</c>	-0.0223 378 / 5 / 272 ≤ 1e-04	-0.0880 415 / 4 / 236 ≤ 1e-04	-0.5694 423 / 6 / 226 ≤ 1e-04	-0.6075 427 / 5 / 223 ≤ 1e-04	-0.6148 432 / 5 / 218 ≤ 1e-04	-1.0662 544 / 4 / 107 ≤ 1e-04	-1.3329 410 / 3 / 242 ≤ 1e-04	-1.4348 599 / 4 / 52 ≤ 1e-04
DGW-MSM-2 0.4501	0.0223 272 / 5 / 378 ≤ 1e-04	-	-0.0657 413 / 14 / 228 ≤ 1e-04	-0.5471 391 / 6 / 258 ≤ 1e-04	-0.5852 427 / 5 / 223 ≤ 1e-04	-0.5925 430 / 5 / 220 ≤ 1e-04	-1.0439 523 / 5 / 127 ≤ 1e-04	-1.3106 374 / 1 / 280 ≤ 1e-04	-1.4125 508 / 3 / 144 ≤ 1e-04
RGW-MSM-2 0.5158	0.0880 - 236 / 4 / 415 ≤ 1e-04	0.0657 228 / 14 / 413 ≤ 1e-04	-	-0.4813 358 / 5 / 292 0.0026	-0.5195 404 / 4 / 247 ≤ 1e-04	-0.5268 395 / 4 / 256 ≤ 1e-04	-0.9782 481 / 5 / 169 ≤ 1e-04	-1.2448 357 / 2 / 296 ≤ 1e-04	-1.3468 484 / 4 / 167 ≤ 1e-04
WDBA-None 0.9971	0.5694 - 226 / 6 / 423 ≤ 1e-04	0.5471 258 / 6 / 391 ≤ 1e-04	0.4813 292 / 5 / 358 0.0026	-	-0.0381 311 / 5 / 339 0.4294	-0.0454 327 / 5 / 323 0.6339	-0.4969 390 / 4 / 261 ≤ 1e-04	-0.7635 320 / 1 / 334 0.2660	-0.8654 455 / 5 / 195 ≤ 1e-04
WSDBA-15 1.0353	0.6075 223 / 5 / 427 ≤ 1e-04	0.5852 223 / 5 / 427 ≤ 1e-04	0.5195 247 / 4 / 404 ≤ 1e-04	0.0381 339 / 5 / 311 0.4294	-	-0.0073 285 / 7 / 363 0.1784	-0.4587 369 / 4 / 282 ≤ 1e-04	-0.7254 316 / 1 / 338 0.2573	-0.8273 429 / 4 / 222 ≤ 1e-04
WSDBA-7 1.0426	0.6148 218/5/432 ≤ 1e-04	0.5925 220 / 5 / 430 ≤ 1e-04	0.5268 256 / 4 / 395 ≤ 1e-04	0.0454 323 / 5 / 327 0.6339	0.0073 363 / 7 / 285 0.1784	-	-0.4515 369 / 4 / 282 ≤ 1e-04	-0.7181 315 / 1 / 339 0.4052	-0.8200 429 / 4 / 222 ≤ 1e-04
WW - 1.4940	1.0662 107 / 4 / 544 ≤ 1e-04	1.0439 127 / 5 / 523 ≤ 1e-04	0.9782 169 / 5 / 481 ≤ 1e-04	0.4969 261 / 4 / 390 ≤ 1e-04	0.4587 282 / 4 / 369 ≤ 1e-04	0.4515 282 / 4 / 369 ≤ 1e-04	-	-0.2666 245 / 1 / 409 ≤ 1e-04	-0.3685 396 / 3 / 256 ≤ 1e-04
Jittering _ 1.7606	1.3329 242/3/410 ≤ 1e-04	1.3106 280 / 1 / 374 ≤ 1e-04	1.2448 296 / 2 / 357 ≤ 1e-04	0.7635 334 / 1 / 320 0.2660	0.7254 338 / 1 / 316 0.2573	0.7181 339 / 1 / 315 0.4052	0.2666 409 / 1 / 245 ≤ 1e-04	-	-0.1019 447 / 1 / 207 ≤ 1e-04
AW 1.8626	1.4348 52 / 4 / 599 ≤ 1e-04	1.4125 144 / 3 / 508 ≤ 1e-04	1.3468 167 / 4 / 484 ≤ 1e-04	0.8654 195 / 5 / 455 ≤ 1e-04	0.8273 222 / 4 / 429 ≤ 1e-04	0.8200 222 / 4 / 429 ≤ 1e-04	0.3685 256 / 3 / 396 ≤ 1e-04	0.1019 207 / 1 / 447 ≤ 1e-04	If in bold, then p-value < 0.05

Experimental Results - Diversity



- AS is more promissing compared to fidelity
- DGW / RGW MSM high cost still winning
- WSDBA less diverse than WDBA
- Jittering and AW fail by a lot

1.5

1.0

0.5

0.0

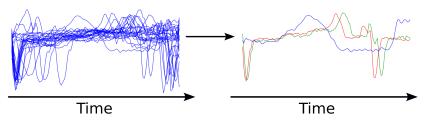
-0.5 4

-1.5

Takeaway

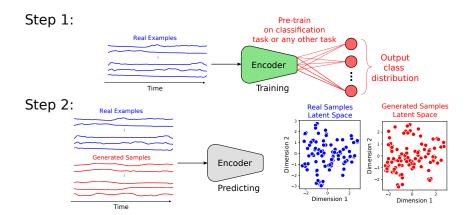
Given a set of time series

Create synthetic realistic series?

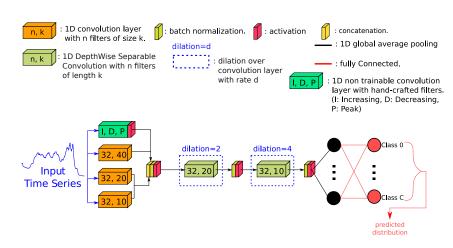


- Title: Re-framing Time Series Augmentation Through the Lens of Generative Models
- Why not evaluate synthetic generation methods prior to using them for data augmentation?
- Is there a relation with performance gain ?
- Used offline or online data augmentation?
- Are all metrics important for augmentation ?
- Contact: ali-el-hadi.ismail-fawaz@uha.fr
- Website: https://hadifawaz1999.github.io/
- GitHub: https://github.com/MSD-IRIMAS/Data-Augmentation-4-TSC

Feature Extraction Metrics



Feature Extractor



Ismail-Fawaz, A., Devanne, M., Berretti, S., Weber, J. and Forestier, G., 2023, October. Lite: Light inception with boosting techniques for time series classification. In 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA) (pp. 1-10). IEEE.