Domain Adaptation of Time Series through Optimal Transport and Temporal Alignment

François Painblanc[1], Laetitia Chapel[3], Nicolas Courty[2], Chloé Friguet[2], Charlotte Pelletier[2], Romain Tavenard[1]

[1] *Université Rennes 2, LETG, IRISA, Rennes, France*
[2] *Université Bretagne Sud, IRISA, UMR CNRS 6074, Vannes, France*
[3] *Institut Agro Rennes-Angers, IRISA, Rennes, France*

Résumé. De grandes quantités de données non étiquetées sont souvent disponibles, mais l'étape d'annotation est généralement une tâche fastidieuse et/ou coûteuse. En apprentissage non supervisé, l'adaptation de domaine peut résoudre ce problème en exploitant les étiquettes d'un domaine source pour classifier des données d'un domaine cible, similaire mais différent. Dans le cas de séries temporelles, des défis supplémentaires surviennent, notamment en raison des décalages temporels potentiels qui s'ajoutent aux décalages de distribution entre les domaines.

Nous présentons une méthode dénommée Match-And-Deform (MAD) qui vise à relever ces défis en identifiant les correspondances entre les séries temporelles des domaines source et cible tout en tenant compte des distorsions temporelles. Le problème d'optimisation associé aligne simultanément (1) les séries en optimisant un coût de transport optimal et (2) les temps à l'aide de *dynamic time warping*. Intégré dans un réseau de neurones profond, MAD permet l'apprentissage de nouvelles représentations des séries temporelles, alignant les domaines et améliorant le pouvoir discriminant du réseau.

La méthode est évaluée empiriquement sur des données de référence et des données réelles de télédétection, démontrant l'efficacité de MAD: les séries sont appariées de façon pertinentes et les décalages temporels sont estimés avec précision. Des performances de classification comparables ou supérieures aux stratégies d'adaptation de domaine de séries temporelles profondes l'état de l'art sont également obtenues.

Cet article est issu de Painblanc et al. [2023]. Le code et le jeu de données sont disponibles publiquement: `https://github.com/rtavenar/MatchAndDeform`

Mots-clés. adaptation de domaines, séries temporelles, transport optimal, dynamic time warping

Abstract. Large amounts of unlabeled data are often available, and the annotation step is usually a tedious and/or costly task. Unsupervised domain adaptation can address this issue by leveraging labels from a source domain to classify data from a related, yet different, target domain. When dealing with time series data, additional challenges arise due to potential temporal shifts alongside the feature distribution shift.

We introduce the Match-And-Deform (MAD) approach to address these challenges. MAD aims at identifying matching between source and target time series while taking into account temporal distortions. To achieve this, the associated optimization problem simultaneously (1)

aligns the series using an optimal transport loss and (2) adjusts the timestamps using dynamic time warping. When integrated into a deep neural network, MAD facilitates the learning of new representations of time series that align the domains and enhance the network's discriminative power.

Empirical evaluations conducted on benchmark datasets and real remote sensing data demonstrate MAD's effectiveness. These numerical experiments show meaningful sample-to-sample matching and accurately estimates time shifts. Comparable or superior classification performance compared to state-of-the-art deep time series domain adaptation strategies are also achieved.

This paper is adapted from Painblanc et al. [2023]. Code and data are publicly available: `https://github.com/rtavenar/MatchAndDeform`

Keywords. domain adaptation, time series, optimal transport, dynamic time warping

1 Introduction

A standard assumption in machine learning is that the training and the test data are drawn from the same distribution. When this assumption is not met, trained models often have degraded performances because of their poor generalization ability. Domain Adaptation (DA) addresses this challenge by considering the generalization problem when there exists a distributional shift, allowing for improving task efficiency on the target domain through the use of comprehensive information from a source domain.

In this work, we consider DA issue for time series data. Our objective is to classify time series from an unlabelled target dataset $(\mathbf{X}') \in \mathbb{R}^{n' \times T' \times q}$ using a labelled source dataset $(\mathbf{X}, \mathbf{Y}) \in \mathbb{R}^{n \times T \times q} \times \mathcal{C}$. This setting can be yield for example by crop-type mapping from remote sensing data (miniTimeMatch dataset), where domains correspond to different geographical areas and class-level temporal shifts are observed, as illustrated on Fig.1.

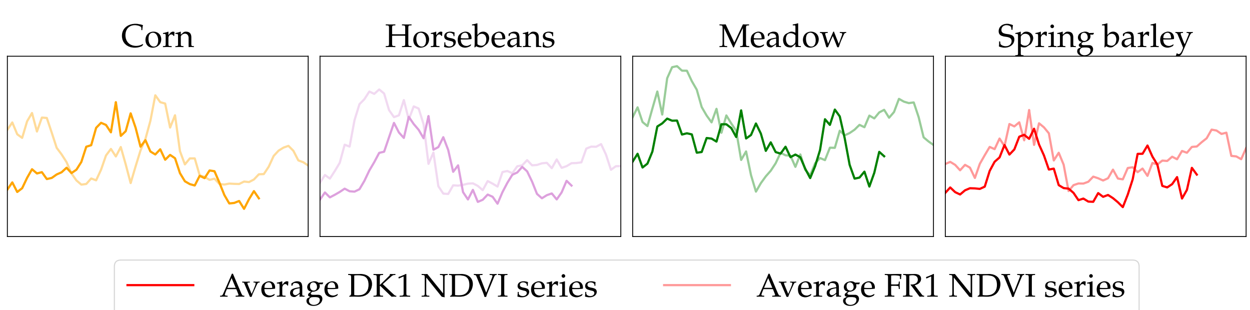

Figure 1: Illustration of temporal shift between averaged NDVI time series for 2 domains, for different types of crops (miniTimeMatch data)

The DA literature primarily focuses on addressing distribution shifts through alignment or common representation space approaches. In unsupervised DA frameworks, training on source domain data leverages this shared representation for improved performance on the target domain. The standard adversarial training aims to induce domain-invariant representations in deep neural networks. Extention to time series can be done with specialized architectures such as convolutional layers (CoDATS, Wilson et al. [2020]). However, the time dimension vanishes with the use of pooled features. Optimal Transport (OT, Peyré and Cuturi [2019]) emerges as a powerful tool in both unsupervised and semi-supervised DA, deriving efficient solutions for assessing distribution shifts and deep neural network losses that capture domain dissimilarities. However, current OT-based DA methods do not encode any temporal coherence for time series data analysis.

In the following, we first introduce basics from OT and time series alignment, and define our optimisation problem to deal with DA for time series. Empirical evaluations are then conducted on benchmark datasets and real remote sensing data to demonstrate the effectiveness of the proposed approach.

2 Domain Adaptation for time series with Optimal Transport and Temporal Alignment

2.1 Background

The optimisation problem for comparing two objects (either time series or distributions) $\mathbf{x}$ and $\mathbf{x}'$ can be stated as:

$$J\Big(\mathbf{C}(\mathbf{x}, \mathbf{x}'), \Pi\Big) = \underset{\boldsymbol{\pi} \in \Pi}{\arg\min} \Big\langle \mathbf{C}(\mathbf{x}, \mathbf{x}'), \boldsymbol{\pi} \Big\rangle \ , \tag{1}$$

in which Π is a set of admissible couplings. A coupling will either be a temporal alignment if $\mathbf{x}$ and $\mathbf{x}'$ are time series or a matching between samples if they are distributions, with appropriate constraint sets (see Fig.2.1). The solution $J(\cdot,\cdot)$ of the optimization problem is called the optimal coupling matrix. The cost matrix $\mathbf{C}(\mathbf{x}, \mathbf{x}') = \{d(x^i, x'^j)\}_{ij}$ stores distances $d(x^i, x'^j)$ between atomic elements x^i and x'^j, respectively from $\mathbf{x}$ and $\mathbf{x}'$. Dynamic time warping (DTW, Sakoe and Chiba [1978]) and Optimal transport (OT) are both instances of the same general optimization problem (1) that (OT) defines a distance between two probability measures or (DTW) matches two (multivariate) time series, as illustrated on Fig.2.1.

Optimal Transport

$$\underset{\boldsymbol{\gamma}\in\Gamma(\mathbf{w},\mathbf{w}')}{\arg\min}\left\langle \mathbf{C}(\mathbf{X},\mathbf{X}'),\boldsymbol{\gamma}\right\rangle$$

$\mathbf{X}$, $\mathbf{X}'$: Datasets with weights $\mathbf{w},\mathbf{w}'$
$\Gamma(\mathbf{w},\mathbf{w}')$: set of all couplings

Dynamic Time Warping

$$\underset{\boldsymbol{\pi}\in\mathcal{A}(T,T')}{\arg\min}\left\langle \mathbf{C}(\mathbf{x},\mathbf{x}'),\boldsymbol{\pi}\right\rangle$$

$\mathbf{x}$, $\mathbf{x}'$: Time series of lengths T, T'
$\mathcal{A}(T,T')$: set of all temporal alignments

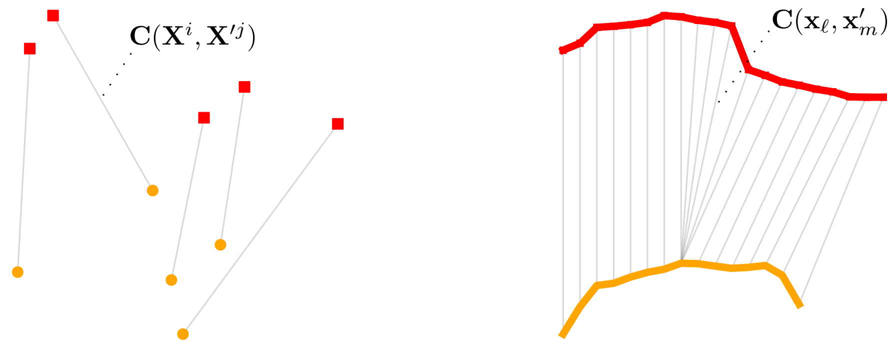

Figure 2: Illustration of the 2 optimisation problems: (left) OT defines a distance between two probability measures and (right) DTW matches two (multivariate) time series

2.2 Match-And-Deform

We introduce Match-And-Deform (MAD) that combines OT with DTW to achieve time series matching and timestamp alignment. In other words, MAD evaluates the feature distribution shift between domains up to a global temporal alignment. MAD jointly optimizes a global DTW alignment and an OT coupling to match two sets of time series. Let us therefore define MAD as:

$$\begin{aligned}\mathrm{MAD}(\mathbf{X},\mathbf{X}') &= \underset{\substack{\boldsymbol{\gamma}\in\Gamma(\mathbf{w},\mathbf{w}')\\ \boldsymbol{\pi}\in\mathcal{A}(T,T')}}{\arg\min}\langle \mathbf{L}(\mathbf{X},\mathbf{X}')\otimes\boldsymbol{\pi},\boldsymbol{\gamma}\rangle \\ &= \underset{\substack{\boldsymbol{\gamma}\in\Gamma(\mathbf{w},\mathbf{w}')\\ \boldsymbol{\pi}\in\mathcal{A}(T,T')}}{\arg\min}\sum_{i,j}\sum_{\ell,m} d(x_\ell^i, x_m'^j)\pi_{\ell m}\gamma_{ij}. \end{aligned} \tag{2}$$

with, $\mathbf{L}(\mathbf{X},\mathbf{X}')$ is a 4-dimensional tensor whose elements are $L_{\ell,m}^{i,j} = d(x_\ell^i, x_m'^j)$, with $d : \mathbb{R}^q\times\mathbb{R}^q\to\mathbb{R}^+$ being a distance. $\otimes$ is the tensor-matrix multiplication. $\boldsymbol{\pi}$ is a global DTW alignment between timestamps and $\boldsymbol{\gamma}$ is a transport plan between samples from $\mathbf{X}$ and $\mathbf{X}'$.

This optimization problem can be further extended to the case of distinct DTW mappings for each class c in the source data. This results in the following optimization problem, coined $|\mathcal{C}|$-MAD:

$$|\mathcal{C}|\text{-MAD}(\mathbf{X},\mathbf{X}',\mathbf{Y}) = \underset{\substack{\boldsymbol{\gamma}\in\Gamma(\mathbf{w},\mathbf{w}')\\ \forall c,\boldsymbol{\pi}^{(c)}\in\mathcal{A}(T,T')}}{\arg\min}\sum_{i,j}\sum_{\ell,m} L_{\ell,m}^{i,j}\pi_{\ell m}^{(y^i)}\gamma_{ij}\,. \tag{3}$$

In that case, $|\mathcal{C}|$ DTW alignments are involved, one for each class c. $\boldsymbol{\pi}^{(y^i)}$ denotes the DTW matrix associated to the class y^i of x^i. This more flexible formulation allows adapting to different temporal distortions that might occur across classes.

The joint optimization problem introduced in Eq. (3) involves $|\mathcal{C}|$ finite sets of admissible DTW paths and a continuous space with linear constraints for the OT plan. We perform a Block Coordinate Descent (BCD) to optimize the corresponding loss.

Fig. 3 illustrates the general workflow of MAD.

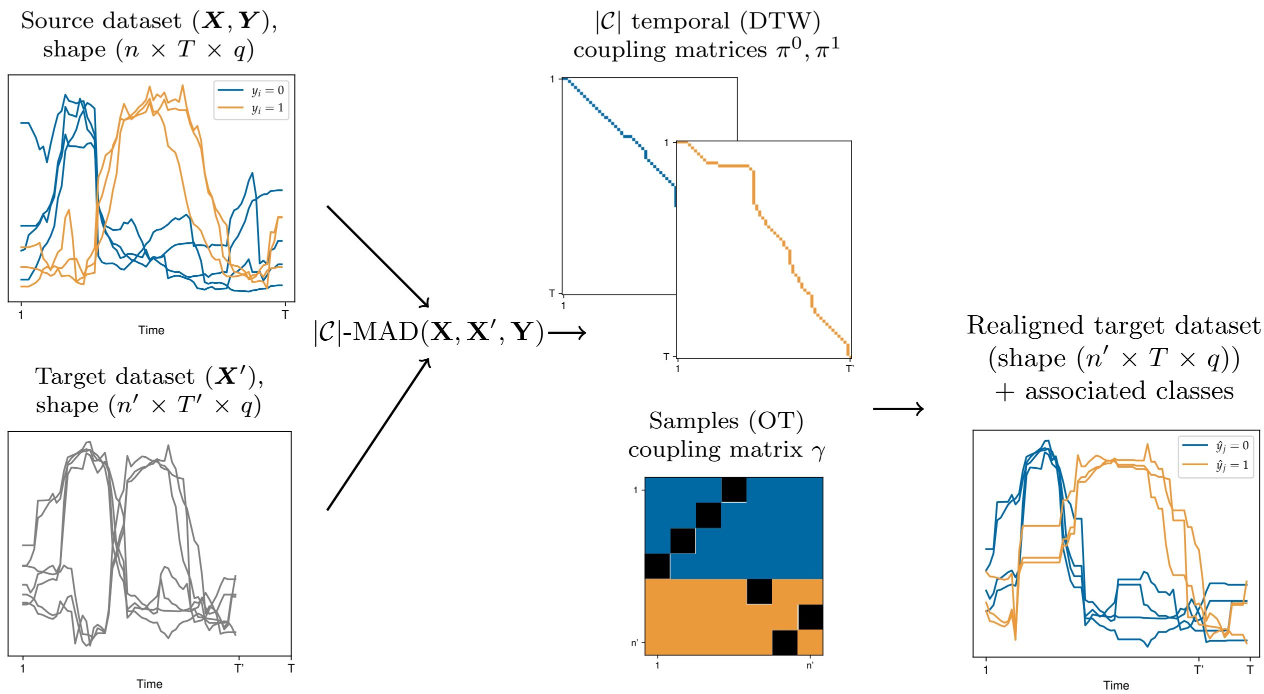

Figure 3: Match-And-Deform ($|\mathcal{C}|$-MAD) takes two time series datasets as inputs: a source (labelled) dataset and a target (unlabelled) dataset. It jointly computes an optimal transport (OT) coupling matrix $\boldsymbol{\gamma}$ and $|\mathcal{C}|$ class-wise dynamic time warping (DTW) paths $\{\boldsymbol{\pi}^{(c)}\}_{c\in\mathcal{C}}$. The OT cost is derived from the pairwise distances yielded by the DTW paths while the DTW cost is weighted by the OT plan. These outputs are then used to improve classification in the target dataset (figure from Painblanc et al. [2023])

2.3 Deep Domain Adaptation Loss

MAD can be used as a loss function in a neural network to learn a domain-invariant latent representation. Indeed, OT has been successfully used as a loss to measure the discrepancy between source and target domain samples embedded into a latent space. Similarly to Deep-JDOT (Damodaran et al. [2018]), our proposal considers a deep unsupervised temporal DA model that relies on MAD or $|\mathcal{C}|$-MAD as a regularization loss function, as illustrated on Fig.4.

$$\mathcal{L}\Big(\mathbf{X}, \mathbf{Y}, \mathbf{X}', f_\theta, g_\Omega, \gamma, \{\pi^{(c)}\}_c\Big) = \frac{1}{n}\sum_i \overbrace{\mathcal{L}_s\Big(y^i, f_\theta(g_\Omega(\mathbf{x}^i))\Big)}^{\text{Cross-Entropy (CE) on source domain}} \tag{4}$$
$$+ \alpha \underbrace{\sum_{i,j}\sum_{\ell,m} d\left(g_\Omega(\mathbf{x}^i)_\ell, g_\Omega(\mathbf{x}'^j)_m\right) \pi^{(y^i)}_{\ell m}\gamma_{ij}}_{\substack{\text{Matching series under temporal alignment} \\ \text{(MAD cost on intermediate features)}}} + \beta \underbrace{\sum_{i,j} \mathcal{L}_t\Big(y^i, f_\theta(g_\Omega(\mathbf{x}'^j))\Big)\gamma_{ij}}_{\substack{\text{CE on target domain} \\ \text{transporting source labels}}}$$

Optimisation is computed over two groups of parameters: (i) the neural network parameters θ and Ω and (ii) MAD transport plan $\boldsymbol{\gamma}$ and DTW paths $\{\boldsymbol{\pi}^{(c)}\}_c$. Similar to what is done in Damodaran et al. [2018], we use an approximate optimization procedure that relies on stochastic gradients.

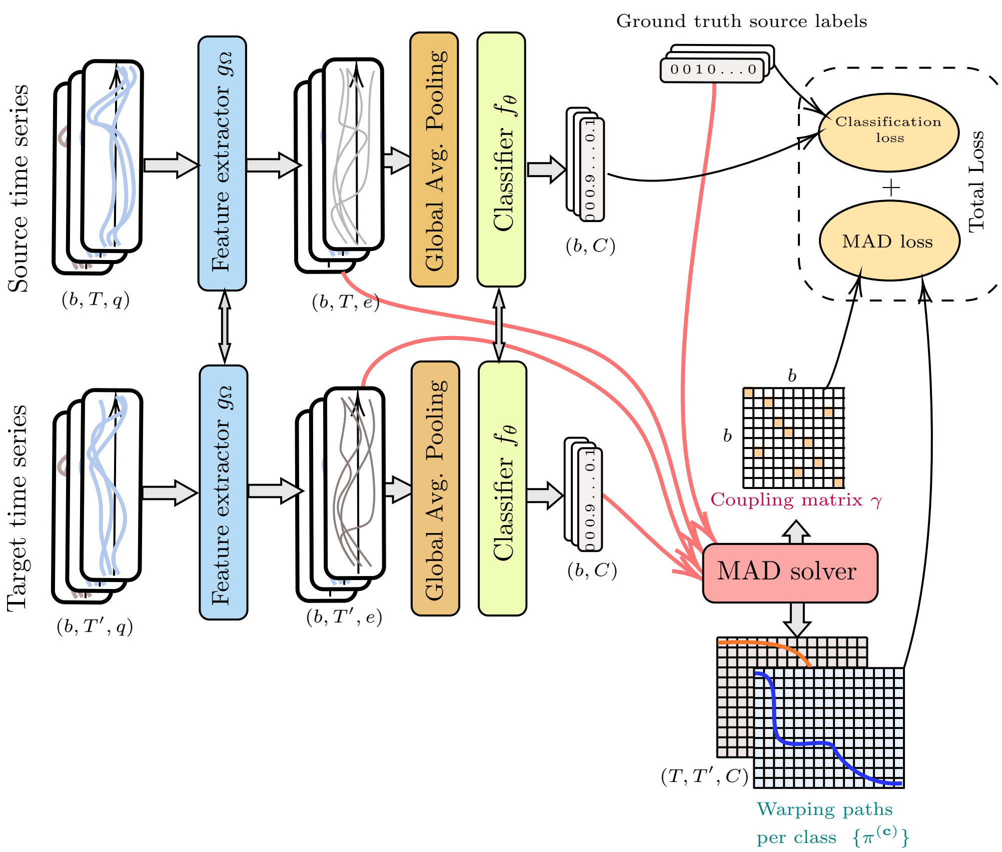

Figure 4: MAD backbone architecture and schematic view of the loss computation (figure from Painblanc et al. [2023])

3 Experimental Results on Remote Sensing Data

The use of MAD and $|\mathcal{C}|$-MAD as losses for neural domain adaptation is now assessed considering a real remote sensing dataset, for which there exists a known global temporal shift between the (classes of the) domains due to different weather conditions. The proposed approach has also been further studied on benchmarked datasets, results are not reported here.

miniTimeMatch dataset is a subsample of TimeMatch Nyborg et al. [2022], a crop-type mapping dataset of different geographical areas, with assumed temporal shifts. Pre-processing performed on the raw data is described in Painblanc et al. [2023]. It finally leads to a dataset of $28,858$ time series and 8 classes per domain, each being described by 10 features per timestamp. Both variant of MAD (considering a single DTW path *vs* one DTW path per class) are evaluated, in comparison with state-of-the-art method CoDATS-WS (Wilson et al. [2020]). Main results are reported in Tab.1: MAD and $|\mathcal{C}|$-MAD outperform CoDATS-WS in 4 out of the 5 DA problems, sometimes with an important improvement (see DK1 $\rightarrow$ FR1 for example). This illustrates the fact that MAD and $|\mathcal{C}|$-MAD are of prime interest when global or class-specific temporal deformations occur between domains (see Fig. 1).

Problem	No adaptation	CoDATS-WS	MAD	$\lvert\mathcal{C}\rvert$-MAD	Target only
DK1 $\rightarrow$ FR1	69.2 ± 1.3	74.8 ± 1.5	$\mathbf{88.4 \pm 0.4}$	88.3 ± 0.9	95.8 ± 0.9
DK1 $\rightarrow$ FR2	62.2 ± 3.5	$\mathbf{87.0 \pm 3.4}$	82.5 ± 1.1	81.0 ± 1.1	94.2 ± 1.7
DK1 $\rightarrow$ AT1	73.9 ± 0.2	71.6 ± 15.4	$\mathbf{93.1 \pm 1.2}$	92.3 ± 2.2	96.7 ± 0.7
FR1 $\rightarrow$ DK1	61.9 ± 5.2	78.0 ± 10.7	$\mathbf{88.2 \pm 0.3}$	$\mathbf{88.2 \pm 0.5}$	96.2 ± 0.3
FR1 $\rightarrow$ FR2	78.8 ± 0.9	82.1 ± 8.2	$\mathbf{90.5 \pm 0.2}$	89.6 ± 0.4	94.2 ± 1.7
Average	69.2 ± 2.2	78.7 ± 7.8	$\mathbf{88.5 \pm 0.6}$	87.9 ± 1.0	95.4 ± 1.1

Table 1: Mean and std classification accuracy over 3 repetitions (DK: Denmark, FR: France, AT: Austria)

4 Conclusion and perspectives

In this paper, we introduce Match-And-Deform (MAD) that combines optimal transport and dynamic time warping for time series domain adaptation in the presence of global time shifts. We furthermore embed MAD as a regularization loss in a neural domain adaptation setting and evaluate its performance in different settings: MAD reaches better performance than state-of-the-art strategies thanks to its ability to capture temporal shifts.

Nevertheless, inter-domain class balance is an implicit OT hypothesis. Extension of MAD could alleviate this OT assumption by using unbalanced optimal transport. An application of MAD could be to consider an estimate the quality of missing values imputation with MAD score.

References

Bharath Bhushan Damodaran, Benjamin Kellenberger, Rémi Flamary, Devis Tuia, and Nicolas Courty. DeepJDOT: Deep joint distribution optimal transport for unsupervised domain adaptation. In *European Conference on Computer Vision*, pages 447–463, 2018.

Joachim Nyborg, Charlotte Pelletier, Sébastien Lefèvre, and Ira Assent. Timematch: Unsupervised cross-region adaptation by temporal shift estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 188:301–313, 2022.

François Painblanc, Laetitia Chapel, Nicolas Courty, Chloé Friguet, Charlotte Pelletier, and Romain Tavenard. Match-And-Deform: Time Series Domain Adaptation through Optimal Transport and Temporal Alignment. In *ECML PKDD 2023*, Torino, Italy, September 2023. URL `https://hal.science/hal-04189149`.

Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1): 43–49, 1978.

Garrett Wilson, Janardhan Rao Doppa, and Diane J Cook. Multi-source deep domain adaptation with weak supervision for time-series sensor data. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1768–1778, 2020.