



Learning Domain-Invariant Representations via Optimal Transport and Spectral Embeddings

Deep CosmoStat Days

Fred NGOLE, Abdel DJalil SAD SAOUD

February, 13th, 2026



1. Context

2. Background

3. Spectral Embedding of Optimal Transport Plan

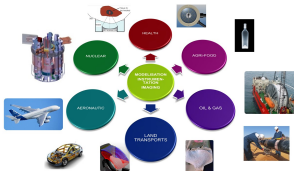
4. Experimental results

5. Conclusion

Context

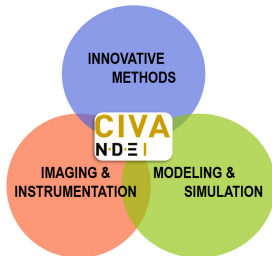
01

- NDE team at CEA List : ~ 110 staff members including 25 PhD & post-doc
- Large industrial partnership in main sectors concerned by NDE
- Strong academic collaborations : through CIVA network, National, European and International projects, ...



02

- Three-fold R&D activity



Addressing the main NDE modalities:
Ultrasound, Electromag. (ET,...), X-Ray & CT, Guided Waves, Thermography (IR), ...

03

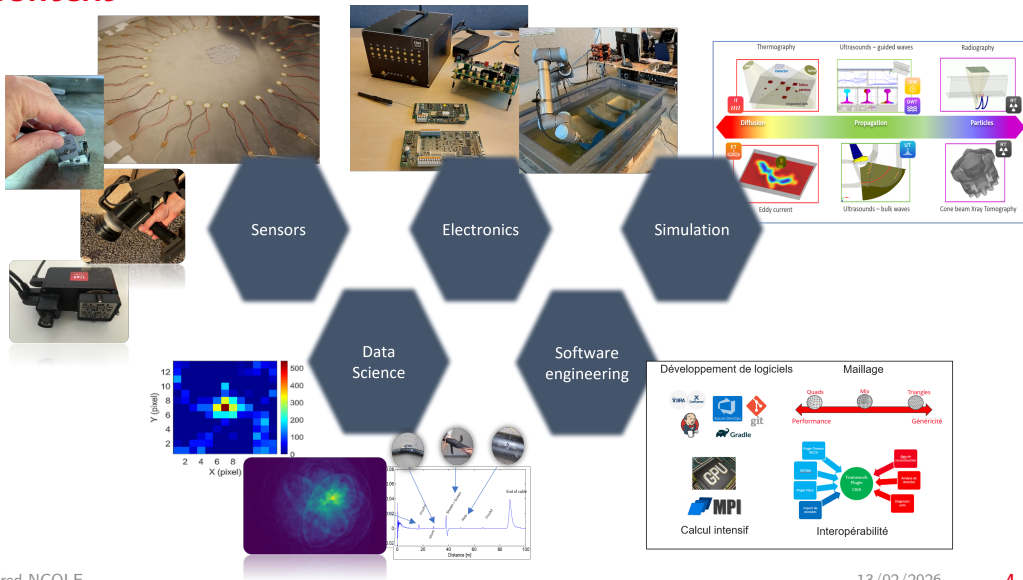
- Experimental & numerical platforms

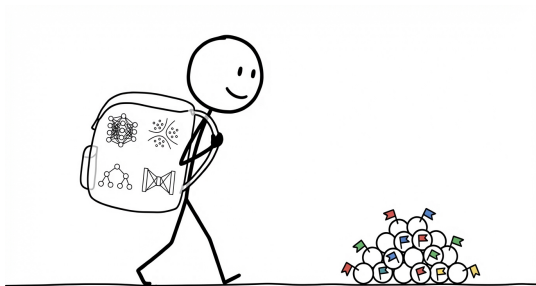


CIVA
NDE I



Context





Empirical risk minimisation

Training and test data are drawn from the same underlying probability distribution :

$$P_{\text{train}}(X, y) = P_{\text{test}}(X, y)$$

This implies that minimizing the empirical risk on the training set leads to low expected risk on the test set :

$$\arg \min_{c \in C} \hat{f}(x), \quad \text{where } \hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \ell(c(x_i^{\text{train}}), y_i^{\text{train}}),$$

$$\Rightarrow f_{\text{test}}(x) = \mathbb{E}_{(X, Y) \sim P_{\text{test}}} [\ell(c(X), Y)] \text{ is small.}$$

Statistical learning tasks : Detection | Classification | Characterization of defects.

Context

But labeled defects data are scarce - fortunately.

And distributional shift happens.

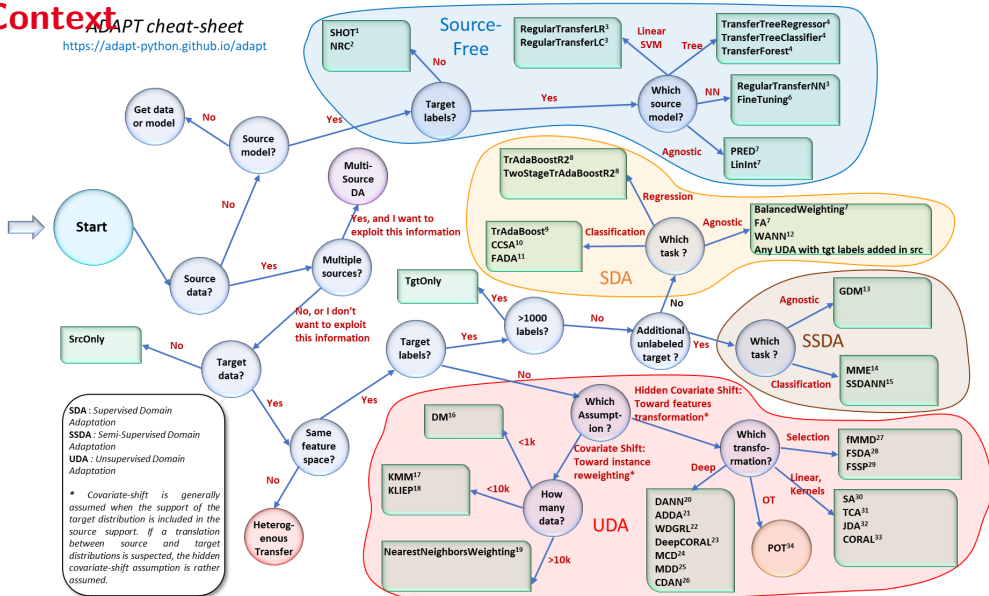
- Hardware differences and control set up (excitation signal, sensors topology, sensors drift...)
- Structure properties (geometry, density, conductivity, etc.)
- Environmental factors (temperature, humidity, (structured) noise...)
- Sim2Real gap



Context

ADAPT cheat-sheet

<https://adapt-python.github.io/adapt>



Context

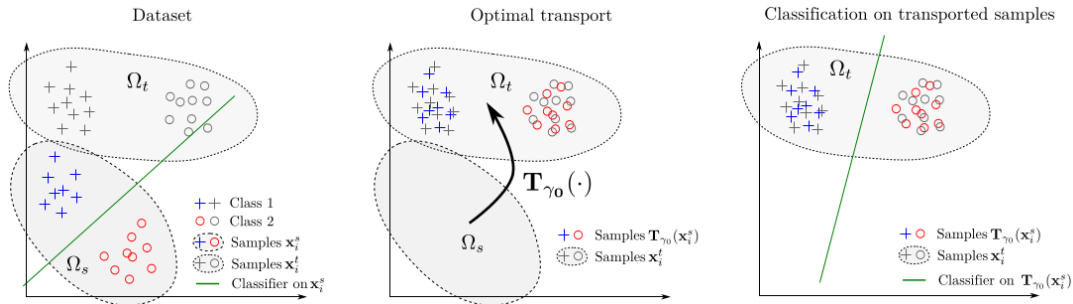


Figure – Optimal Transport for Domain Adaption¹

One leverages distribution level regularity.

1. N. COURTY et al., "Optimal transport for domain adaptation", *IEEE transactions on pattern analysis and machine intelligence*, t. 39, n° 9, p. 1853-1865, 2016 .

1. Context

2. Background

3. Spectral Embedding of Optimal Transport Plan

4. Experimental results

5. Conclusion



Background

Kantorovich Relaxation of Optimal transport

Problem Setup : Let $\mu_s \in \Delta^{n_s}$ and $\mu_t \in \Delta^{n_t}$ be discrete probability measures supported on $X_s = \{x_i^s\}_{i=1}^{n_s}$ and $X_t = \{x_j^t\}_{j=1}^{n_t}$. The *Kantorovich formulation* of optimal transport solves :

$$\min_{\gamma \in \Pi(\mu_s, \mu_t)} \sum_{i,j} C_{ij} \gamma_{ij} \quad s.t \quad C_{ij} = \|x_i^s - x_j^t\|^p$$

The solution γ^* defines the optimal transport plan :

$$\gamma^* = \arg \min_{\gamma \in \Pi(\mu_s, \mu_t)} \sum_{i,j} C_{ij} \gamma_{ij},$$

where $\Pi(\mu_s, \mu_t) = \{\gamma \geq 0 \mid \gamma \mathbf{1}_{n_t} = \mu_s, \gamma^\top \mathbf{1}_{n_s} = \mu_t\}$.

Remarks :

- Scales in $O(n^3 \log n)$ with the number of samples, not suitable for large scale adaptation tasks.

Background

To improve computational efficiency and ensure smooth transport plans,^a proposed a regularized optimal transport :

Entropy-Regularized Optimal transport

$$\gamma^* = \arg \min_{\gamma \in \Pi(\mu_S, \mu_T)} \langle C, \gamma \rangle_F - \varepsilon H(\gamma)$$

where

$$H(\gamma) = - \sum_{i,j} \gamma_{ij} (\log \gamma_{ij})$$

is the entropy of γ , and $\varepsilon > 0$ controls the regularization strength.

Advantages :

- Converges exponentially.
- Quadratic complexity in the number of samples.

a. M. CUTURI, "Sinkhorn distances: Lightspeed computation of optimal transport", *Advances in neural information processing systems*, t. 26, 2013 .

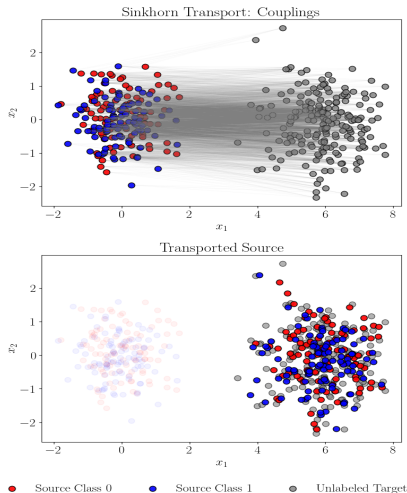
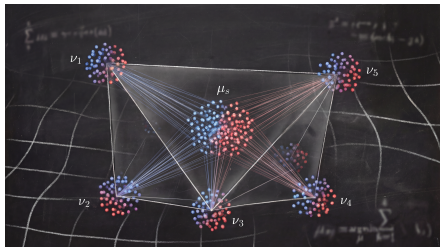


Figure – Visualization of the entropy regularized optimal transport.

Background

Dataset Dictionary Learning for domain adaptation^{a, b}



a. E. F. MONTESUMA et al., "Multi-source domain adaptation through dataset dictionary learning in wasserstein space", in *ECAI 2023-European Conference on Artificial Intelligence*, t. 372, 2023, p. 1739-1746.

b. E. F. MONTESUMA et al., "Optimal Transport for Domain Adaptation through Gaussian Mixture Models", *Transactions on Machine Learning Research*,

For a set of datasets $\mathcal{Q} = \{\hat{Q}_\ell\}_{\ell=1}^N$, atoms $\mathcal{P} = \{\hat{P}_k\}$ and $\mathcal{A}\{\alpha_\ell\}_{\ell=1}^N$, $\sum_k \alpha_{\ell,k} = 1$, we propose the Dataset Dictionary Learning (DaDiL) framework through,

$$(\mathcal{P}^*, \mathcal{A}^*) = \mathcal{P}, \mathcal{A} \frac{1}{N} \sum_{\ell=1}^N \mathcal{L}(\hat{Q}_\ell, \mathcal{B}_{\mathcal{P}}(\alpha_\ell)),$$

$$\underbrace{\text{Reconstruction}} = \mathcal{B} \left(\underbrace{[\alpha_{\ell 1}, \alpha_{\ell 2}, \alpha_{\ell 3}]}_{\text{Representation 1}}, \underbrace{\{\text{Atom 1}, \text{Atom 2}, \text{Atom 3}\}}_{\text{Atom 1}} \right)$$

$$\mathcal{L}(\hat{Q}_\ell, \mathcal{B}_{\mathcal{P}}(\alpha_\ell)) = W_2 \left(\underbrace{\text{Dataset}}_{\text{Dataset}}, \underbrace{\text{Reconstruction}}_{\text{Reconstruction}} \right)$$

1. Context

2. Background

3. Spectral Embedding of Optimal Transport Plan

4. Experimental results

5. Conclusion



Motivation

Entropy-Regularized Optimal transport

$$\gamma^* = \arg \min_{\gamma \in \Pi(\mu_s, \mu_t)} \langle C, \gamma \rangle_F - \varepsilon H(\gamma)$$

where

$$H(\gamma) = - \sum_{i,j} \gamma_{ij} (\log \gamma_{ij})$$

is the entropy of γ , and $\varepsilon > 0$ controls the regularization strength.

Entropic bias

- The entropic regularization biases the domain alignment.
- The weaker the regularization, to lower its numerical benefits.
- There is no simple rule to set ε .

Can we mitigate the bias and keep the computational benefit of the entropic regularization ?

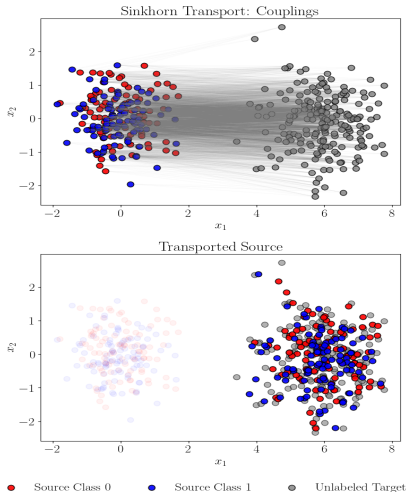


Figure – Visualization of the entropy regularized optimal transport.

Spectral Embedding of Graphs

From Graph to Spectral Embedding

- Graph $G = (V, E, W)$ with adjacency matrix :

$$A_{ij} = \begin{cases} w_{ij}, & (i, j) \in E, \\ 0, & \text{otherwise.} \end{cases}$$

- Normalized Laplacian :

$$\mathcal{L}_{\text{sym}} = I - D^{-1/2} A D^{-1/2}, \quad \text{where } D : \text{degree matrix}$$

- Spectral embedding $F \in \mathbb{R}^{n \times k}$ obtained by :

$$F^* = \arg \min_{F^T F = I} \text{Tr}(F^T \mathcal{L}_{\text{sym}} F)$$

using eigenvectors corresponding to the k smallest eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_k$.

- Embedding $\rightarrow$ well-separated clusters (nodes of the same class group together).

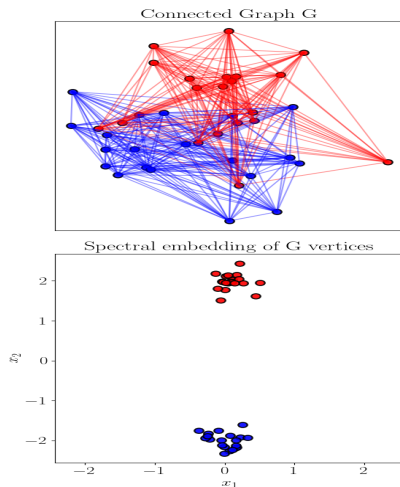


Figure – Spectral embedding intuition

13/02/2026

15 / 26

Spectral Embedding of Optimal Transport Plan - single source

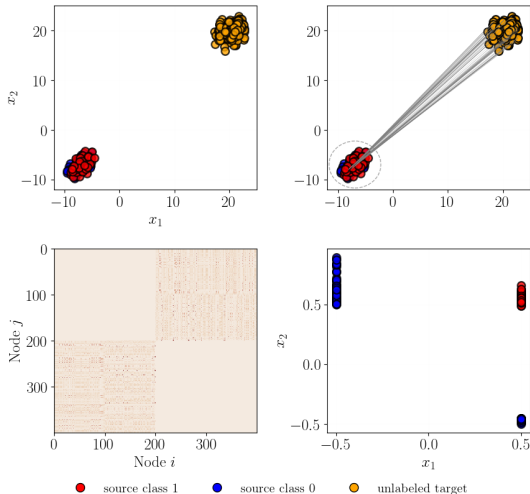
- Consider a labeled source domain $\mathcal{D}_s = (\mathcal{X}_s, \mu_s)$, and an unlabeled target domain $\mathcal{D}_t = (\mathcal{X}_t, \mu_t)$, we solve the entropy regularized OT problem to obtain γ^* :

$$\gamma_{s \rightarrow t}^* = \begin{bmatrix} \gamma_{11}^{s_1} & \gamma_{12}^{s_1} & \dots & \gamma_{1n_t}^{s_1} \\ \gamma_{21}^{s_1} & \gamma_{22}^{s_1} & \dots & \gamma_{2n_t}^{s_1} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n_s 1}^{s_1} & \gamma_{n_s 2}^{s_1} & \dots & \gamma_{n_s n_t}^{s_1} \end{bmatrix}$$

- The corresponding adjacency matrix of the bipartite graph is :

$$\mathcal{A}^* = \begin{bmatrix} 0 & \gamma_{s \rightarrow t}^* \\ (\gamma_{s \rightarrow t}^*)^\top & 0 \end{bmatrix}.$$

- The Resulting spectral embedding *i.e.*, by computing $\mathcal{L}_{sym}$ and solving for F^* , produces a **domain-invariant** and **discriminative** representation for classifier training.

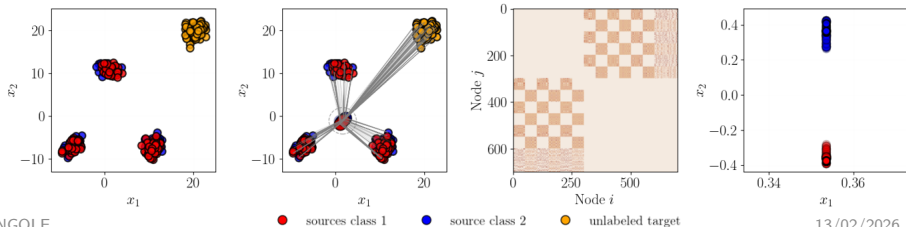


Spectral Embedding of Optimal Transport Plan - Multisource

- Represent multiple domains $\{\mathcal{D}_s^i\}$ and $\mathcal{D}_t$ as a star graph by connecting them via a **Wasserstein barycenter** $\mathcal{D}_b$, thus $\mathcal{A}^*$ will be :

$$\mathcal{A}^* = \begin{bmatrix} 0 & \gamma_{b \rightarrow s_1}^* & \dots & \gamma_{b \rightarrow s_{N_s}}^* & \gamma_{b \rightarrow t}^* \\ (\gamma_{b \rightarrow s_1}^*)^\top & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (\gamma_{b \rightarrow s_{N_s}}^*)^\top & 0 & \dots & 0 & 0 \\ (\gamma_{b \rightarrow t}^*)^\top & 0 & \dots & 0 & 0 \end{bmatrix}$$

- Spectral embedding F^* represents the nodes of the graph onto an euclidean space suitable for classification.



1. Context
2. Background
3. Spectral Embedding of Optimal Transport Plan
- 4. Experimental results**
5. Conclusion



Experiments



Datasets

■ MSD (Music-Speech Discrimination)

- Domain shifts : noise conditions
- Samples : 128 excerpts (64 music + 64 speech)
- Features : 56
- Task : Classification (2 classes)

■ MGR (Music Genre Recognition)

- Domain shifts : varying noise conditions
- Samples : 1,000 recordings
- Features : 56
- Task : multi-class classification (10 genres)

■ CS-RT (Cable fault diagnosis)

- Domain shifts : 3 compression factors + different physical characteristics
- Samples : 200 signals per domain
- Features : 512
- Task : 4-class classification (no fault, soft fault, short, open circuit)

Results

Benchmark	MSD						MGR				
Domains	Noiseless	Buccaneer2	Destroyerengine	F16	Factory2	Average	Buccaneer2	Destroyerengine	F16	Factory2	Average
Source-only *	67.99 $\pm$ 8.62	82.43 $\pm$ 1.75	51.57 $\pm$ 2.56	88.89 $\pm$ 2.72	50.02 $\pm$ 2.21	68.18 $\pm$ 3.47	22.90 $\pm$ 0.84	38.25 $\pm$ 0.91	51.57 $\pm$ 1.11	47.80 $\pm$ 0.34	40.13 $\pm$ 11.07
KMM *	74.64 $\pm$ 6.70	87.12 $\pm$ 2.79	52.35 $\pm$ 2.94	74.86 $\pm$ 5.58	50.41 $\pm$ 2.17	67.88 $\pm$ 4.04	21.75 $\pm$ 0.99	39.25 $\pm$ 0.66	49.81 $\pm$ 1.69	47.37 $\pm$ 0.71	39.54 $\pm$ 10.99
TCA *	50.01 $\pm$ 2.53	90.43 $\pm$ 1.40	87.14 $\pm$ 4.99	95.12 $\pm$ 2.02	84.76 $\pm$ 3.30	81.49 $\pm$ 2.75	<u>58.95$\pm$1.27</u>	60.67 $\pm$ 2.07	<u>68.75$\pm$2.11</u>	59.82 $\pm$ 0.50	62.04 $\pm$ 3.91
OT-IT *	89.46 $\pm$ 1.22	89.26 $\pm$ 1.56	82.84 $\pm$ 2.78	84.97 $\pm$ 3.09	91.21 $\pm$ 2.04	89.76 $\pm$ 2.34	56.35 $\pm$ 0.84	61.92 $\pm$ 1.64	66.72 $\pm$ 1.86	61.77 $\pm$ 1.65	61.69 $\pm$ 3.67
OT-Laplace *	90.44 $\pm$ 1.37	87.28 $\pm$ 2.97	84.38 $\pm$ 1.76	86.14 $\pm$ 2.79	90.61 $\pm$ 1.68	87.27 $\pm$ 2.11	58.02 $\pm$ 1.45	60.47 $\pm$ 1.75	66.55 $\pm$ 1.60	63.87 $\pm$ 1.51	<u>62.23$\pm$3.24</u>
JCPOT *	65.66 $\pm$ 5.71	92.55 $\pm$ 2.11	87.89 $\pm$ 1.39	88.67 $\pm$ 1.67	82.41 $\pm$ 2.22	83.44 $\pm$ 2.62	35.87 $\pm$ 0.41	48.47 $\pm$ 2.97	51.92 $\pm$ 3.25	51.95 $\pm$ 1.75	47.05 $\pm$ 6.60
JCPOT-LP *	12.89 $\pm$ 1.67	89.06 $\pm$ 1.38	84.97 $\pm$ 3.23	90.24 $\pm$ 1.71	86.13 $\pm$ 1.88	72.66 $\pm$ 1.97	36.40 $\pm$ 0.39	52.92 $\pm$ 1.32	56.30 $\pm$ 0.37	51.52 $\pm$ 2.28	49.28 $\pm$ 7.62
WBT *	52.74 $\pm$ 3.82	56.88 $\pm$ 9.54	56.63 $\pm$ 6.88	56.63 $\pm$ 6.56	59.38 $\pm$ 2.61	58.56 $\pm$ 4.80	21.37 $\pm$ 2.25	24.30 $\pm$ 2.71	25.30 $\pm$ 6.02	22.70 $\pm$ 2.25	23.41 $\pm$ 1.50
WBT _{reg} *	94.34 $\pm$ 2.55	96.27 $\pm$ 1.60	92.98 $\pm$ 1.38	94.92 $\pm$ 0.68	96.87 $\pm$ 0.94	95.08 $\pm$ 1.43	70.60$\pm$1.27	83.05$\pm$0.97	84.40$\pm$1.71	90.17$\pm$0.46	82.05$\pm$7.13
Target-only *	96.88 $\pm$ 2.97	90.51 $\pm$ 3.98	93.07 $\pm$ 3.81	89.23 $\pm$ 4.25	92.30 $\pm$ 3.62	92.40 $\pm$ 3.73	67.43 $\pm$ 1.43	67.96 $\pm$ 2.91	66.86 $\pm$ 2.00	68.37 $\pm$ 1.87	67.41 $\pm$ 0.56
SPOT ²	99.22$\pm$0.00	96.61$\pm$0.97	97.40$\pm$0.37	95.31$\pm$0.00	98.70$\pm$0.37	97.45$\pm$0.34	45.53 $\pm$ 0.12	61.63 $\pm$ 0.31	58.17 $\pm$ 0.12	<u>70.77$\pm$0.19</u>	59.03 $\pm$ 0.19

Table – Classification results on the MSD and MGR benchmarks. * denotes results reported from⁴, except for the Noiseless domain and the overall average across all domains for MSD. Each column shows results with that domain as target, with values reported as mean accuracy $\pm$ standard deviation, averaged over three independent runs.

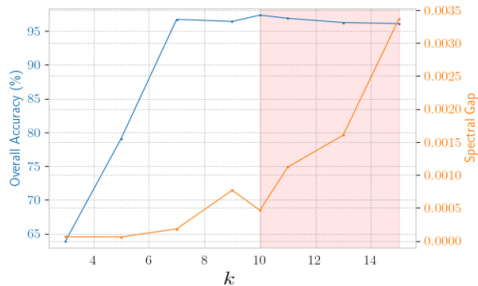
2. A. D. S. SAOUD et al., "Beyond Mapping: Domain-Invariant Representations via Spectral Embedding of Optimal Transport Plans", *arXiv preprint arXiv:2601.13350*, 2026 .
3. E. F. MONTESUMA et F. M. N. MBOULA, "Wasserstein barycenter for multi-source domain adaptation", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, p. 16 785-16 793.
4. E. F. MONTESUMA et F. M. N. MBOULA, "Wasserstein barycenter for multi-source domain adaptation", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, p. 16 785-16 793.

Results

Benchmark	CS-RT				
Algorithm	CF ₁₆	CF ₄	CF ₂	Phys.	Average
Source-only	23.00 $\pm$ 0.00	28.00 $\pm$ 0.00	51.00 $\pm$ 0.00	47.00 $\pm$ 0.00	37.25 $\pm$ 0.00
KMM	23.00 $\pm$ 0.00	25.00 $\pm$ 0.00	45.00 $\pm$ 0.00	44.50 $\pm$ 0.00	34.38 $\pm$ 0.00
OT-Laplace	18.00 $\pm$ 0.00	25.50 $\pm$ 0.00	46.50 $\pm$ 0.00	<u>54.50</u> $\pm$ 0.00	36.12 $\pm$ 0.00
JCPOT	<u>29.50</u> $\pm$ 0.00	<u>30.00</u> $\pm$ 0.00	30.50 $\pm$ 0.00	41.00 $\pm$ 0.00	32.75 $\pm$ 0.00
JCPOT-LP	24.50 $\pm$ 0.00	25.00 $\pm$ 0.00	26.00 $\pm$ 0.00	4.50 $\pm$ 0.00	20.00 $\pm$ 0.00
WBT	20.00 $\pm$ 0.71	19.67 $\pm$ 1.55	31.00 $\pm$ 2.86	30.00 $\pm$ 6.48	23.75 $\pm$ 2.32
WBT _{reg}	28.83 $\pm$ 0.24	16.17 $\pm$ 0.24	<u>52.17</u> $\pm$ 3.88	38.50 $\pm$ 2.55	33.92 $\pm$ 1.73
SPOT	49.90 $\pm$ 3.72	62.88 $\pm$ 1.17	65.90 $\pm$ 0.64	69.59 $\pm$ 5.70	62.07 $\pm$ 2.25

Table – Results for Compressed Sensing Reflectometry Dataset. Each column shows results with that domain as target.

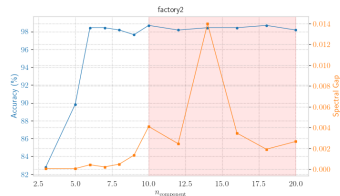
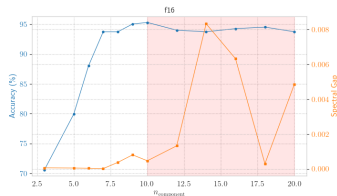
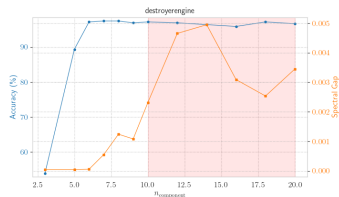
Sensitivity to Hyperparameters



Observations : How to choose k and ϵ ?

- As $k \uparrow$, accuracy and spectral gap $\uparrow$.
- Maximum Spectral gap at N_c indicates well-separated connected component.
- More discriminative embedding.
- Provides a way to select ϵ .

SEWB: Domain-wise Accuracy & Spectral Gap vs $n_{\text{component}}$



1. Context
2. Background
3. Spectral Embedding of Optimal Transport Plan
4. Experimental results
- 5. Conclusion**

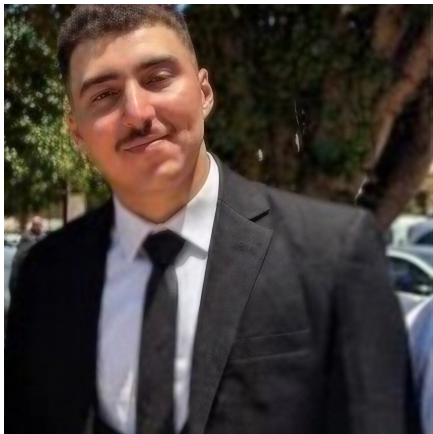


Conclusion

- We proposed *SPOT* a domain adaptation method for addressing distributional shifts, our contributions are as follow :
 - A novel perspective on the optimal transport plan, using it to construct a graph that connects domains, later embedded into a discriminant and domain invariant representation rather than explicit mapping.
 - We propose a multisource domain adaptation algorithm in this framework.
 - We evaluate and compare *SPOT* with the SOTA methods on benchmarks datasets + a cable defects diagnosis use case based on time domain reflectometry, Achieving the best performance on most benchmarks.
- Ongoing work : estimating eigenfunctions of a space continuous transfer operator, which will provide out-of-sample embeddings + end-to-end optimization.

Acknowledgement

Abdel Djalil SAD SAOUD ^a



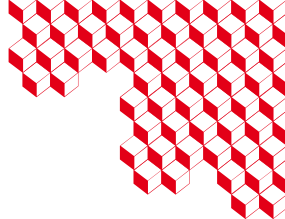
Eduardo Fernandes MONTESUMA



a. <https://github.com/CEA-LIST/SpectralDomainAdaptation>



list



Thank you for your attention !

CEA SACLAY

91 191 Gif-sur-Yvette Cedex

Plateau de Saclay, France

fred-maurice.ngole-mboula@cea.fr