DOI: https://doi.org/10.14311/TPFM.2026.014

DISTANCE BASED SIMILARITY METRICS FOR ARTIFICIAL NEURAL NETWORK ESTIMATES OF SOOT DISTRIBUTION IN CATALYTIC FILTERS

M. Khýr [1, 2], M. Isoz [1, 3]

[1] Czech Academy of Sciences, Institute of Thermomechanics, Dolejškova 5, Prague 182 00, Czech Republic

[2] Department of Chemical Engineering, Faculty of Chemical Engineering, University of Chemistry and Technology, Technická 5, Prague 166 28, Czech Republic

[3] Department of Mathematics, Informatics, and Cybernetics, Faculty of Chemical Engineering, University of Chemistry and Technology, Technická 5, Prague 166 28, Czech Republic

Abstract

Although artificial neural networks have attracted the attention of researchers as potential fast and lightweight substitutes for expensive computational models, their application to real world problems may be hindered by improper choice of training and validation metrics. In this work, we examine the applicability of so called distance based metrics to the problem of soot deposition in porous media and compare their performance with standard pixel-wise metrics such as the root mean square error. Specifically, we showcase the implementation and behavior of the Sliced Wasserstein Distance and the Sinkhorn Distance on a set of benchmarks, and we demonstrate their use with a previously developed artificial neural network for the estimation of soot deposits. Our results suggest that both the Sliced Wasserstein and Sinkhorn distances outperform the root mean squared error; however, their computational complexity may render their use in conjunction with artificial neural networks impractical.

Keywords: CFD, ANN, CNN, CF.

1 Introduction

Catalytic filters (CFs) are devices used in automotive exhaust gas aftertreatment systems to facilitate catalytic abatement of harmful gasses and the capture and subsequent oxidation of particulate matter, the soot. The CF bed is a monolithic porous ceramic support with parallel channels that are plugged alternately at each end. Because the gas is forced to flow through the porous walls between the individual channels, it comes in contact with the catalytic coating, so surface reactions can take place, and solid particles from the gas stream are deposited within the pores. Although using a CF over a multi-device system lowers the overall temperature and pressure losses, the CF performance may be strongly dependent on the exact spatio-temporal distribution of deposited soot, which itself depends on the CF support microstructure. For effective design of the monolithic supports, it is therefore crucial to obtain fast and reliable estimates of the distribution of the deposited soot given the microstructure.

A complete physics-based numerical solver for fluid flow, chemical reaction, and soot deposition in CF walls based on a combined Eulerian–Lagrangian framework has been developed and validated on experimental data [1]. However, its computational cost effectively prohibits its use in fast prototyping. Consequently, we developed an alternative approach to predict the soot distribution in CFs. The proposed approach is based on artificial neural networks (ANNs) and employs convolutional autoencoders and deep neural networks [2]. The ANN model showed promising results with two- and three-dimensional CF structure data in terms of both the location and the total amount of deposited soot. Still, the ANN estimates of the deposited soot structure appear diffuse and lack sufficient detail [2].

Up to this point, standard pixel-wise metrics, such as mean square error (MSE) and binary cross-entropy (BCE), have been used to train and evaluate the ANN model. However, these metrics are generally incapable of capturing fine structural differences. On the other hand, distance based metrics arising from the optimal transport theory are designed to measure the similarity between samples in terms of physical closeness, albeit at higher computational expenses. Note that switching from MSE and BCE to distance based metrics is also motivated by the effort to include probabilistic element in the supervised learning of our ANN model [3]. This effort itself is driven by the probabilistic nature of the underlying physics-based solver that takes into account the random Brownian motion of soot particles.

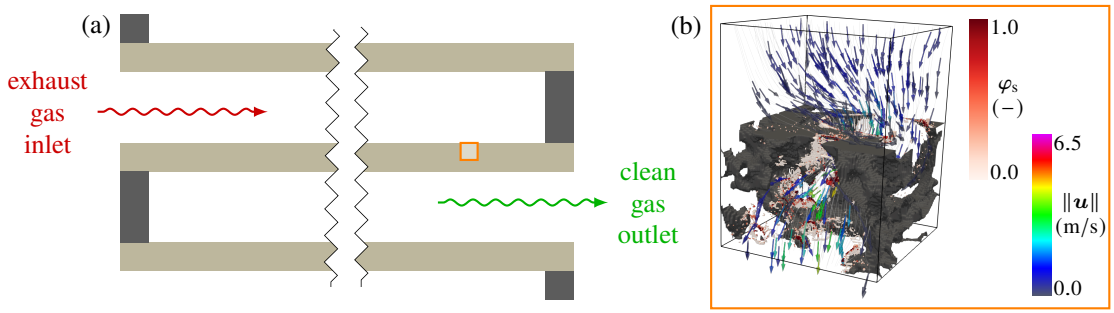

Figure 1: Visualization of the problem of interest. (a) Two adjacent channels of a catalytic filter. (b) Flow field and soot distribution in a small section of the catalytic filter wall, highlighted in (a).

Hence, in this work, we examine the use of selected distance based metrics with the existing ANN model for soot deposit estimation. First, we briefly discuss the ANN model and introduce the necessary background for distance based metrics used in this work, namely the Sliced Wasserstein Distance (SWD) and the Sinkhorn Distance (ShD). Subsequently, we test the properties of these metrics implemented in the Keras/TensorFlow [4, 5] machine learning framework on a benchmark data set comprising simple geometric transformations. And finally, we show how metrics' behavior translates to the real dataset.

2 Methods

Applied problem of interest. First, let us recall the problem at hand. As stated above, we are interested in estimating the spatial distribution of soot within the porous wall of a catalytic filter. In Figure 1a, a schematics of two CF channels is depicted, together with the intended function of cleaning the engine exhaust gas. In Figure 1b, a small portion of the porous CF wall is shown, together with a typical flow field and soot distribution as estimated from the physics-based solver from [1]. Note that φ_s and $\boldsymbol{u}$ mark the soot volume fraction and the exhaust gas velocity, respectively.

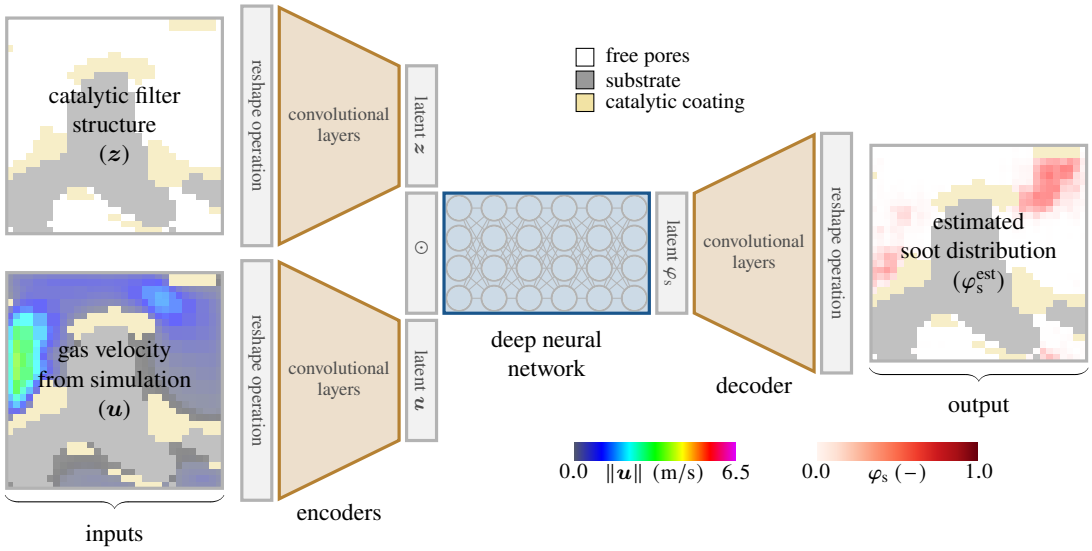

Figure 2: Schematics of the ANN model used to estimate soot distribution inside a catalytic filter.

Neural network model. The artificial neural network (ANN) model implemented in the Keras/TensorFlow framework is schematically outlined in Figure 2, which has been previously described in detail in [2]. The model takes binary representations of small segments of the CF microstructure $\boldsymbol{z}$ together with precalculated discretised velocity fields $\boldsymbol{u}$ as input and estimates the values of soot volume fractions φ_s in the corresponding cells. The inputs and outputs are represented by data arrays corresponding to cell-center values of a structured orthogonal computational mesh. The model is modular and its overall architecture consists of several building

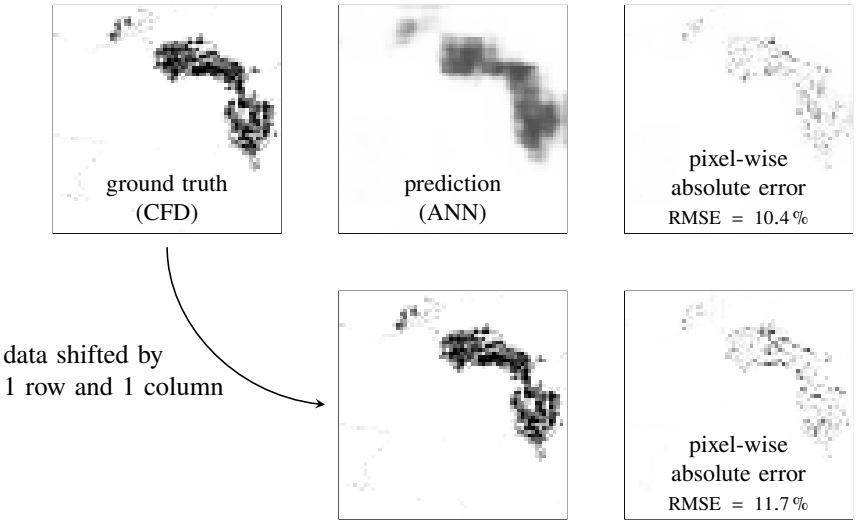

Figure 3: Illustration of MSE sensitivity to slight shifts of the same structures within the compared datasets.

blocks: (i) two convolutional encoders are used to convert the input samples into their low-dimensional latent representations, (ii) a deep neural network is used for estimation in the latent space, and finally (iii) a convolutional decoder is used to reconstruct the latent representation of the estimate.

Summary of ANN training principles and importance of similarity metric. To train the ANN model described above, we employ supervised learning methods. This learning is based on (a) knowing the network input $\mathcal{I}$, expected output $\mathcal{O}$, and estimate $\mathcal{E}$ and (b) adjusting the network weights W to minimize some metric of the error between $\mathcal{O}$ and $\mathcal{E}$, i.e., $\min_W \mathcal{M}^*(\mathcal{O}, \mathcal{E})$, where by using the superscript "$*$", we emphasize the fact that the metric $\mathcal{M}$ can be selected as problem specific. Common choices for $\mathcal{M}$ are the mean squared error (MSE) or binary cross-entropy (BCE).

In the case of the ANN model presented above, the training is performed in two stages. First, the auto-encoders, i.e., only the convolutional encoders and decoders, without the deep neural network (DNN), for $\boldsymbol{z}$, $\boldsymbol{u}$, and φ_s are trained to minimize the error between the inputs and the corresponding estimates $\boldsymbol{z}^\mathrm{est}$, $\boldsymbol{u}^\mathrm{est}$, and $\varphi_\mathrm{s}^\mathrm{est}$. Second, the trained encoders for $\boldsymbol{z}$ and $\boldsymbol{u}$ are connected to the trained decoder for φ_s via the DNN and the DNN weights in the combined ANN are trained to minimize the error between the pre-computed and estimated φ_s.

The specificity of our case lies in the fact that we are predominantly interested in the general topology of the soot distribution. In particular, if the soot mass distribution and the shape of the soot deposits are estimated correctly in the overall computational domain, slight shifts of individual structures are not paramount for the quality of the estimate. This is further promoted by the fact that even our physical CFD model that represents the ground truth for the ANN training contains a stochastic component due to the consideration of the Brownian motion of the sub-micron soot particles. Consequently, even a repeated CFD simulation will not yield exactly the same results. Unfortunately, the standard error metrics are sensitive to any shifts in the data, which is for root MSE (RMSE) illustrated in Figure 3.

Optimal transport framework. Metrics based on the optimal transport theory, such as the Wasserstein distance or its alternatives [6], provide an alternative to standard error metrics [3]. To better understand the upcoming paragraphs, we first need to review some basic notions of optimal transport theory. For a more general and mathematically sound introduction to the field, the reader is referred to the monograph [7]. Let us now consider two one-dimensional probability measures α and β defined on a common equidistant discrete support $\boldsymbol{x} \in \mathbb{R}^n, x_i = i/n, n \in \mathbb{N}$. Thus, the probability measures α and β can be represented by histogram vectors $\boldsymbol{a} \in \mathbb{R}^n$ and $\boldsymbol{b} \in \mathbb{R}^n$, such that $a_i = \alpha(x_i)$ and $b_i = \beta(x_i)$. The set $\boldsymbol{U}$ of all possible transport plan matrices $\boldsymbol{P}$ is given as

$$\boldsymbol{U}(\boldsymbol{a}, \boldsymbol{b}) = \left\{ \boldsymbol{P} \in \mathbb{R}_{0,+}^{n \times n}; \sum_{j=1}^{n} P_{ij} = a_i \, \forall i \in \{1, 2, \ldots, n\}, \sum_{i=1}^{n} P_{ij} = b_j \, \forall j \in \{1, 2, \ldots, n\} \right\}, \quad (1)$$

which means that $\boldsymbol{a}$ and $\boldsymbol{b}$ are the row-wise and the column-wise sums (marginals) of $\boldsymbol{P}$. Additionally, we define the cost matrix $\boldsymbol{C} \in \mathbb{R}_{0,+}^{n\times n}$ with the elements $C_{ij} = |x_i - x_j|$ being the Euclidean distances between the elements of the discrete support. The *Wasserstein distance* between $\boldsymbol{a}$ and $\boldsymbol{b}$ is then defined as

$$\mathrm{WD}(\boldsymbol{a}, \boldsymbol{b}) = \min_{\boldsymbol{P}\in\boldsymbol{U}} \langle \boldsymbol{C}, \boldsymbol{P}\rangle_{\mathrm{F}}\,, \tag{2}$$

where $\langle\cdot,\cdot\rangle_{\mathrm{F}}$ is the Frobenius inner product (i.e. the sum of elements of a matrix given as an element-wise product of two matrices). If we interpret the histogram values as mass and the support values as spatial coordinates, the Wasserstein distance is proportional to the minimum amount of energy needed to transport the mass from arrangement $\boldsymbol{a}$ to $\boldsymbol{b}$. Together, relations (1) and (2) define a special case of the Kantorovich (optimal transport) problem. For the n^2 elements of the transport plan matrix $\boldsymbol{P}$, formula (1) gives $2n$ linear equations (constraints), which makes finding the Wasserstein distance a linear programming (LP) problem. Note that by proper modification of the structure of the cost matrix $\boldsymbol{C}$, the WD can be easily generalized to distinct non-equidistant supports and multidimensional probability measures.

For one-dimensional problems, the Wasserstein distance has the closed form

$$\mathrm{WD}(\boldsymbol{a}, \boldsymbol{b}) = \frac{1}{n}\sum_{i=1}^{n} |A_i - B_i|\,, \quad A_i = \sum_{j=1}^{i} a_j\,, \quad B_i = \sum_{j=1}^{i} b_j\,, \tag{3}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are the cumulative sums (cumulative distributions) of the elements of $\boldsymbol{a}$ and $\boldsymbol{b}$. However, for multidimensional problems, the exact evaluation of the Wasserstein distance requires the solution of the optimal transport problem (2) and is therefore unfeasible in most applications. Consequently, WD may be replaced by a metric with similar properties and approximated numerically, as will be discussed in the following paragraphs.

Sliced Wasserstein distance. One metric derived from the Wasserstein distance is the Sliced Wasserstein distance (SWD), which is defined as an average of Wasserstein distances between one-dimensional projections $\mathrm{proj}_\theta\, \boldsymbol{a}$ and $\mathrm{proj}_\theta\, \boldsymbol{b}$ of (generally multi-dimensional) measures α and β over all directions θ within a unit sphere [8]. In applications, the Monte Carlo approximation [9] of SWD may be calculated from a finite number n_θ of projections in randomly selected directions θ as

$$\mathrm{SWD}(\boldsymbol{a}, \boldsymbol{b}) \simeq \frac{1}{n_\theta}\sum_{k=1}^{n_\theta} \mathrm{WD}(\mathrm{proj}_{\theta_k}\, \boldsymbol{a}, \mathrm{proj}_{\theta_k}\, \boldsymbol{b})\,. \tag{4}$$

Sinkhorn Distance. By adding an entropy regularization term $h(\boldsymbol{P})$ to the objective function in (2), we obtain the Sinkhorn distance [6]

$$\mathrm{ShD}_\lambda(\boldsymbol{a}, \boldsymbol{b}) = \min_{\boldsymbol{P}\in\boldsymbol{U}(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{C}, \boldsymbol{P}\rangle_{\mathrm{F}} - \lambda\, h(\boldsymbol{P})\,, \quad h(\boldsymbol{P}) = -\langle \log \boldsymbol{P} - \boldsymbol{E}, \boldsymbol{P}\rangle_{\mathrm{F}}\,, \tag{5}$$

where λ is the regularization parameter. Crucially, regularization makes the computation of the ShD a convex problem and one that can be solved efficiently using the iterative Sinkhorn–Knop algorithm.

3 Results

Benchmark cases. The goal of the numerical experiments is to evaluate whether the sliced Wasserstein distance (SWD) and/or the Sinkhorn distance (ShD) are suitable metrics for application to our specific problem of training our ANN model to predict distribution of soot deposits inside catalytic filters with acceptable accuracy. In particular, we need to examine whether the tested metrics (i) are not overly sensitive to slight shifts in the data, (ii) do not suffer from saturation when two identical structures are moved away one from another, and (iii) have a distinct minimum with respect to an "almost identity" transformation of a tested object Consequently, we defined two types of benchmark cases, the first comprising well-defined transformations (rotation, translation, uniform scaling) of simple two-dimensional shapes (ellipse, rectangle, triangle). The second benchmark follows the same transformations, but is focused on real-life data. The two datasets and the transformations tested are shown in Figure 4. A Python class has been prepared to define the shapes, manipulate them and project them onto a pixel mesh (rasterization). For all the cases, evolution of MSE, SWD, and ShD with respect to the prescribed transformations was studied.

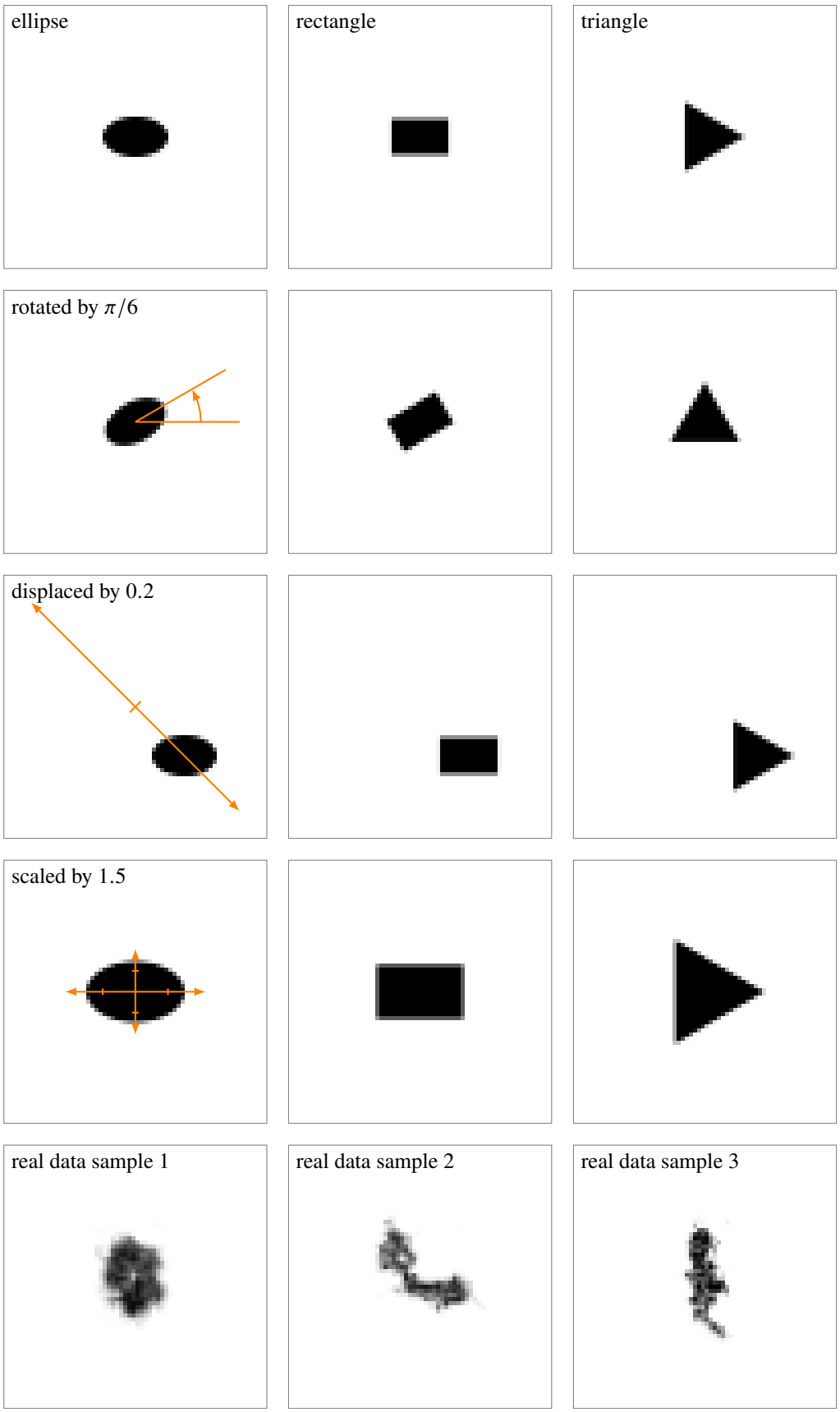

Figure 4: Rows one and five: visualization of tested datasets. Rows two to four: visualization of operations performed on the tested datasets.

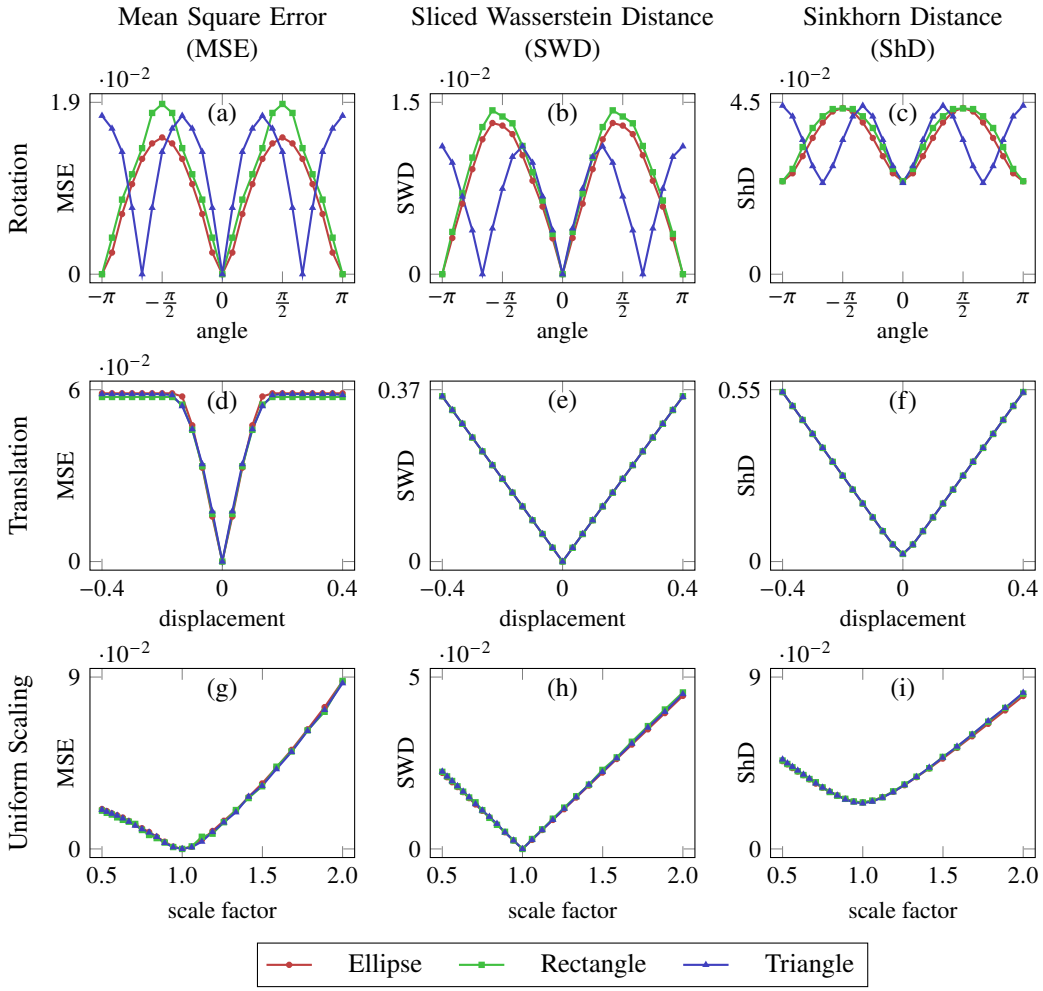

Figure 5: Comparison of metrics on the benchmark cases for simple geometrical transformations.

Artificial data. The results of the first batch of benchmark cases presented in Fig. 5 provide several key insights. First, in Fig. 5a–c, we may notice the difference in rotational symmetry of the equilateral triangle compared to both the ellipse and rectangle. For all metrics, the triangle rotation exhibits a periodicity of $2\pi/3$, while the ellipse and rectangle show a periodicity of π.

With respect to translation, see Fig. 5d–f, all metrics show a linear increase at first; however, MSE (Fig. 5d) saturates quickly, while both SWD and ShD (Fig. 5e–f) continue to increase linearly throughout the range of displacements. The point of saturation of MSE corresponds to the point where the shapes no longer overlap – this explains the small variations between the individual shapes. In the case of pure translation, the Wasserstein distance simplifies to the Euclidean distance between the centroids of the shapes, hence there are no variations between shapes and the points overlap perfectly. In this case, we might also observe that ShD, Fig. 5f, overestimates the Wasserstein distance due to preference for higher entropy transport plans. Furthermore, in Fig. 5c,f and i, we may clearly see that the ShD remains convex but non-zero even for zero transformation (it is not a true metric), which is another consequence of the entropy regularization.

Real data. Moving onto selected samples from the real soot deposit data set, as shown in Fig. 6, we may observe similar general trends as in the benchmark cases. For rotation, Fig. 6a–c, the real deposit shapes no longer exhibit symmetry, although a local minimum appears at $\pi/2$. Notably for translation, in Fig. 6d–f, MSE again saturates quickly, while both SWD and ShD continue to increase linearly over the whole range of displacements – independently of the individual shapes. And once again for uniform scaling, Fig. 6g–i, the distance based metrics outperform MSE by showing a linear increase with the scale factor, with a smoothed minimum for ShD for the untransformed shape.

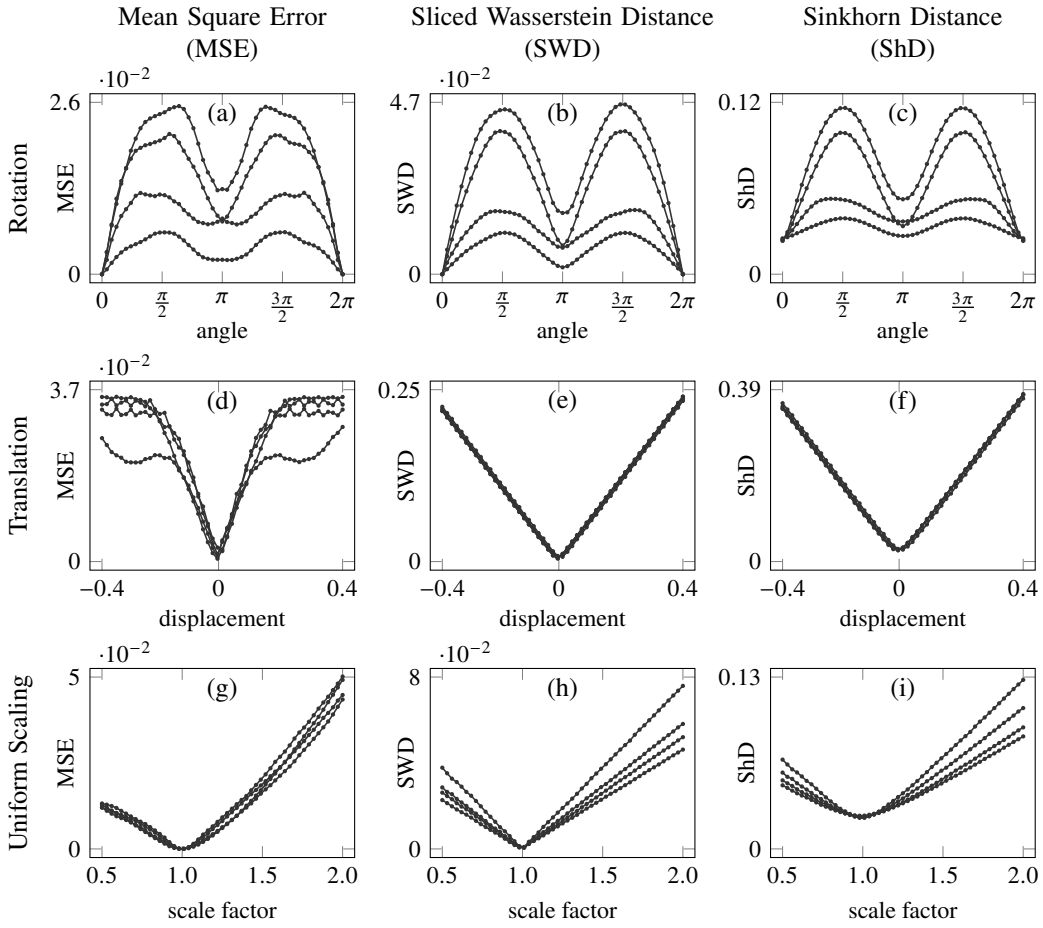

Figure 6: Comparison of metrics on real soot deposit data for simple geometrical transformations.

4 Conclusions

We have investigated the applicability of distance based metrics for the training and validation of artificial neural networks for the estimation of soot distribution in catalytic filters. Specifically, we have focused on the Sliced Wasserstein Distance and the Sinkhorn Distance, showcasing their implementation on a set of benchmarks and comparing their performance with the standard Mean Squared Error metric. Our results suggest that specific properties of the distance based metrics, namely the linear response to translation and scaling of the distributions, generalize well to the soot deposits, which promotes application of these metrics in training of our ANN model. However, an open question remains regarding how distance-based metrics should be incorporated into the training process. In particular, it is unclear whether these metrics can be directly employed in the training of the deep neural network operating on latent representations of inputs and outputs, or whether their inclusion necessitates a more substantial modification of the current two-stage training procedure.

Acknowledgment

The authors acknowledge the financial support provided by the Ministry of Education, Youth, and Sports of the Czech Republic via the project No. CZ.02.01.01/00/23_020/0008501 (METEX), co-funded by the European Union. The work was financially supported by the institutional support RVO:61388998. Special thanks belongs to Strategy AV21 of the Czech Academy of Sciences (VP34 – AI: Artificial Intelligence for Science and Society).

References

[1] M. Plachá, M. Isoz, P. Kočí, M. P. Jones, M. Svoboda, D. S. Eastwood, and A. P. York, "Particle accumulation model in 3D reconstructed wall of a catalytic filter validated with time-resolved X-ray tomography," *Fuel*, vol. 356, p. 129603, Jan. 2024.

[2] M. Khýr, M. Isoz, and M. Plachá, "Artificial intelligence for simulation of soot deposits: from 2D to 3D estimates," in *Topical Problems of Fluid Mechanics 2025*, TPFM, pp. 135–142, Institute of Thermomechanics of the Czech Academy of Sciences; CTU in Prague Faculty of Mech. Engineering Dept. Tech. Mathematics, 2025.

[3] E. F. Montesuma, F. N. Mboula, and A. Souloumiac, "Recent advances in optimal transport for machine learning," 2024.

[4] F. Chollet *et al.*, "Keras." `https://keras.io`, 2015.

[5] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from `tensorflow.org`.

[6] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," in *Advances in Neural Information Processing Systems* (C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, eds.), vol. 26, pp. 2292–2300, Curran Associates, Inc., 2013.

[7] C. Villani, *Topics in Optimal Transportation*. Graduate studies in mathematics, American Mathematical Society, 2003.

[8] S. Kolouri, K. Nadjahi, U. Simsekli, R. Badeau, and G. Rohde, "Generalized sliced Wasserstein distances," in *Advances in Neural Information Processing Systems* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds.), vol. 32, pp. 261–272, Curran Associates, Inc., 2019.

[9] M. Kalos and P. Whitlock, *Monte Carlo Methods*. John Wiley & Sons, Ltd, 2008.