

Window of Deposition Description and Prediction of Deposition Efficiency via Machine Learning Techniques in Cold Spraying

H. Canales, S. Dosta, IG. Cano

Abstract

In this work we describe an energy-based window of deposition, and predict the deposition efficiency for different cold-sprayed powder/substrate systems, using machine learning techniques. We implement several machine learning models to predict whether particles adhere or bounce off during cold spraying. The models are trained using data extracted from several experimental runs taking into account the cumulative particle size distribution and the deposition efficiency of the process. The classification models infer a critical total energy threshold above which deposition occurs. Based on this threshold, we describe an energy-based window of deposition for the powder/substrate systems studied. These models predict the deposition efficiency of different spraying operations for different powder materials with acceptable accuracy. Machine learning techniques provide better understanding of the particle deposition process and enable a more comprehensive exploration of the scope of cold spraying. The use of these techniques opens up new possibilities for the pursuit of links between the spraying process, the structure and different properties for novel cold-sprayed materials.

Keywords: Cold Spray, Machine Learning, Cooper, Aluminum, Deposition Efficiency

1. Introduction

Cold spraying is a solid-state process used for the deposition of particulate materials onto many different substrates. A DeLaval nozzle is used to accelerate particles and then to build dense material layers via their plastic deformation when they impact on the substrate [1]. The deposition mechanisms and the effect of the process parameters on the properties of the deposited material are still not well understood. The main reason for this lack of understanding is the chaotic nature of the process which is the result of thousands of particles with a vast number of different characteristics impacting on the substrate in a very short period of time. In addition, the characteristics of the substrate also influence the deposition of particles, thereby increasing the degree of complexity of the process. Many authors [2, 3, 1, 4, 5] agree that particles achieve deposition only if they surpass a critical velocity; while there is also an upper limit for particle deposition called the erosion velocity. The range between these velocities is known as the *window of deposition* [6]. In [6], the critical velocity is represented by a mathematical model and defined as a function of the particles size, temperature, and the material characteristics. The bonding of the particles is related to adiabatic shear instabilities caused by high strain values during the deformation of the particles. In [7], instead of the critical velocity, the total energy per unit mass (sum of the kinetic and thermal energies) is considered the threshold above which particles adhere. From this threshold, those authors propose an energy-based window of deposition where there is no single critical velocity above which deposition occurs as the amount of energy required depends on the mass. This concept is different from the critical velocity deposition criterion since, with enough thermal energy, deposition can be achieved even at low levels of kinetic energy. The numerical methods used to describe the critical velocity of the particles and eventually a bonding criterion for the cold-spray process usually study the impact of a single or several particles on the substrate. These developments represent an essential contribution to fundamental research. To increase the objectivity and accuracy of the current deterministic models, in [8] the authors propose a method to determine the powder properties. These properties are used in numerical calculations as an alternative to bulk material properties. In spite of the efforts to describe cold spraying bonding criteria, the large number of variables affecting the deposition limit the adequateness of the current numerical approaches in describing the complex process of particle consolidation. Those variables include, among others: substrate preparation [9], powder feed

rate [10], the effect of the gas velocity profile on the particle acceleration [2], and powder oxidation [11]. Data-driven models have been proposed when the current state of the art of our best physics calculations is severely limited [12]. As the cold-spray process is currently described using fundamental causal relations, herein, an alternative approach is proposed based on the developing field of materials informatics [13], also known as materials data science. The main goal of materials data science is to build process–structure–property (PSP) links using methods from data science to help develop new materials and manufacturing processes [14]. Machine learning applies computational algorithms to existing data or past experience in order to optimize performance criteria, thereby deriving new knowledge from existing data [15]. Machine learning algorithms can be divided into three main categories: supervised, unsupervised, and reinforcement learning [16]. Continuous and discrete datasets can be predicted using machine learning algorithms. For example, a regression algorithm can be used to predict a continuous value—as in linear regression—while a classification algorithm can predict a discrete value based on experimental data—as in logistic regression.

Machine learning methods are increasingly being used to predict thermal spray deposition properties. In [17, 18] support vector regression (*SVR*), extreme learning machines (*ELM*), linear regression (*lr*), support vector machines (*SVM*), and gaussian process regression (*GPR*) are used to predict wear loss of coatings deposited by different thermal spray processes. In [19] artificial neural networks (*ANN*) were used to predict the track profile of cold spray depositions aimed at additive manufacturing. Deposition efficiency prediction with machine learning has been studied for plasma spray coatings in [20] using (*ANN*). Another common approach, is to use the surface response method (*SRM*) (a design of experiments methodology based on polynomial regression) to predict coatings properties; i.e. in [21] the authors predicted the porosity of plasma sprayed alumina coatings using (*SRM*). Other authors, have developed models to predict cold spray deposition efficiency using theoretical approaches [22, 2].

In this work, an energy-based window of deposition, such as that proposed in [7], is established using machine learning techniques to define the range in the kinetic–thermal energy field over which a powder material can be deposited. So, instead of considering the concept of critical velocity as the deposition threshold, we infer a critical total-energy boundary from data and define regions within the impact kinetic energy and thermal energy fields of particles

where deposition is achieved. Aluminum and copper powders were sprayed at several combinations of parameters to achieve different particle impact kinetic and thermal energy levels (Section 2). Under the assumption that larger particles usually have higher total energies due to their higher masses—at least under high-pressure cold spraying conditions—particles were designated "deposited" or "not deposited" taking into account the cumulative powder particle size distribution and the deposition efficiency determined in every experimental run. In Section 3, the kinetic and thermal energies of the particles were calculated using computational fluid dynamics and used as features of the labeled particles; the resulting dataset was divided into two parts for training and testing operations. Logistic regression (*LR*), support vector machine (*SVM*), k-nearest neighbor (*kNN*) and random forest (*RF*) statistical classifiers were trained to obtain particle deposition boundaries in the kinetic and thermal energy fields. The prediction accuracy of the models was evaluated against test data to determine which model best defines the window of deposition for the powder material/substrate systems studied in every spraying operation. Finally, the deposition efficiency of the coatings was predicted using classification and regression models, achieving low prediction errors. The implementation of machine learning models for better understating of cold spraying is discussed in Section 4, including further research recommendations.

2. Materials and Methods

As data are the primary input for machine learning models, several cold-spraying operations were planned based on prior experience to achieve high-, medium-, and low-energy states. To classify particles as "deposited" or "not deposited", the portion of the particle size distribution that was preferentially bonded was established using scanning electron microscopy (SEM). Information concerning the bonded particles was obtained from experimentation by many authors. For example, in [23], different methods for determining particle critical velocities are reviewed; some of the methods are: tests with in-flight particle velocity distributions, spraying mono-sized powders, and tests using the deposition efficiency and the cumulative particle size distribution. In [24, 22], the critical velocity of particles is estimated by calculating the impact velocity of the particles through numerical simulation and selecting a specific particle size from the cumulative distribution based on experimentally measured deposition efficiency values, under the assumption that since smaller particles achieve higher

velocities and therefore are more likely to exceed the critical velocity. The latter assumption implies that particles with a diameter below the "critical diameter"—equivalent to the largest particle that achieves critical velocity—should be part of the deposited material. In [25] the critical velocity is studied by increasing the total pressure of the process; as the total pressure increased, the diameter of particles that adhered onto the substrate increased as well. As opposed to the assumptions made in the references above, [26] concluded that smaller particles are not the most "sprayable", after experimentation with titanium powders with different size distributions. In [26], free surface micrographs were taken, and craters on the surface of large bonded particles led the authors to deduce that smaller particles are more likely to be removed. In [27], small, medium and large particle distributions from the same powder material were sprayed at low, medium and high particle impact temperatures using nitrogen or helium, as well as powder preheating. Higher deposition efficiencies were achieved when spraying the powder with the smallest particles using helium; while higher deposition efficiencies were achieved when spraying the powder with largest particles using nitrogen. When spraying with preheated powders, the highest deposition efficiency was achieved with the powder consisting of medium-sized particles. As determining the portion of the powder that is most likely to be deposited depends on the characteristics of each cold-spray operation, in this section we study the free surface of coatings sprayed from different distances using SEM. For the same spraying operation, if smaller particles dominate at the surface of specimens sprayed at short distances, it can be deduced that the smallest particles within the powder are being deposited first—taking into account that at long spraying distances, larger particles are more likely to achieve deposition due to their higher inertia. Meanwhile, if large particles are found on the surfaces of specimens sprayed at short and large distances, we understand that large particles were more likely to be deposited. In this manner, the "critical diameter" is calculated taking into account the cumulative particle size distribution and the deposition efficiency for each specimen, to finally categorize particles by assigning their corresponding "deposited" and "not deposited" labels.

2.1. Aluminum and copper coating depositions

Aluminum and copper powders were sprayed using the Kinetiks 4000 high-pressure cold-spray equipment (Cold Gas Technology GmbH). A spherical 99% copper powder with smooth surfaces (Figure 2c) was sprayed using a ceramic nozzle with a divergent length of 127 mm and

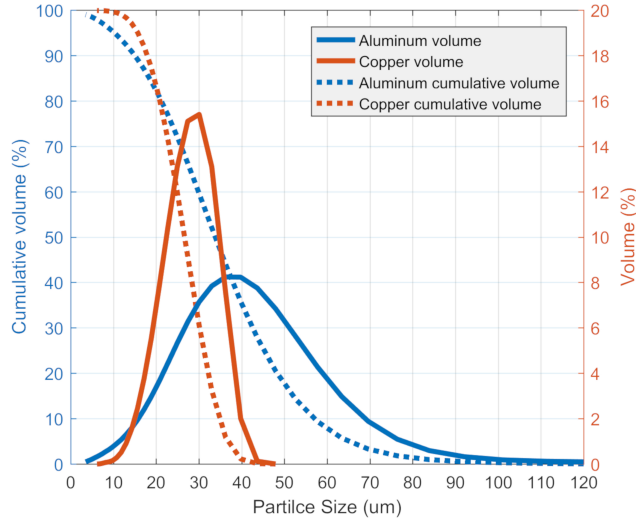


Figure 1: Particle size distribution of copper and aluminum powders

an exit diameter of 6.52 mm. Similarly, a 99.7% aluminum powder with rounded morphology (but not completely spherical) was sprayed using a PBI nozzle with a divergent length of 178 mm and an exit diameter of 10 mm. Figure 1 shows the particle size distribution for the powders, estimated using laser diffraction analysis; it is important to note that in this plot, the cumulative particle size distribution is depicted from the largest to the smallest particles in the distribution, in order to visualize better the contribution of a bonded particle size fraction to the overall deposition efficiency.

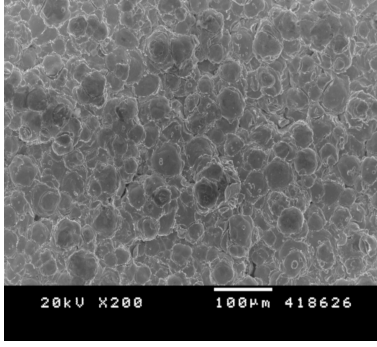
Low-carbon steel substrates were prepared using 220 grid SiC paper, in order to remove impurities from their surfaces. The powder feed rates and robot arm velocities were maintained constant for all the spraying operations using nitrogen as the propelling gas. The deposition efficiency was determined for each spraying operation by comparing the feedstock material mass used in the process and the mass adhered to each specimen. The parameters for each spraying operation and deposition efficiency values are shown in Table 1.

2.2. Analysis of the size of particles bonded within the coating

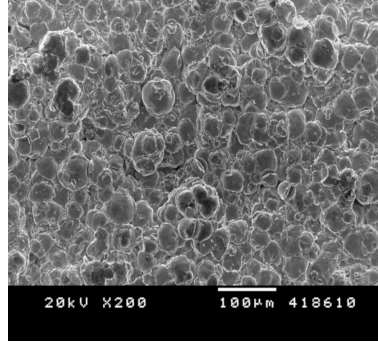
To determine which fraction of the powder material was bonded onto the substrate, free surface micrographs from the sprayed samples and the feedstock powder were obtained using SEM. Figures 2c and 2f show the free surface of the powder feedstock. As can be seen in the figures, both the aluminum and copper powders have a particle size distribution that agrees well with the laser diffraction analysis. Figures 2a-b and 2d-e show the free

Table 1: Cold-spray process parameters for simulation and experimentation

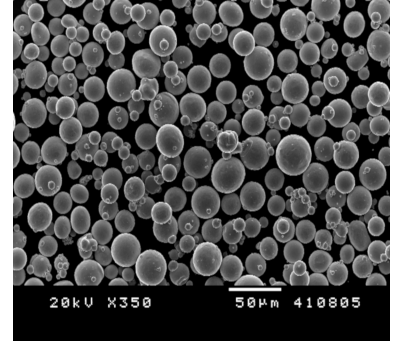
Operation	Powder	Nozzle	Stagnation Temperature, ($^{\circ}\text{C}$)	Total Pressure, (MPa)	Stand-Off Distance, (mm)	Pre-Chamber, (mm)	Percentage DE
(1)A	Aluminum	PBI 33	350	2.5	30	45	43%
(2)A	Aluminum	PBI 33	350	2.5	60	45	35%
(3)A	Aluminum	PBI 33	350	2.5	100	45	22%
(4)A	Aluminum	PBI 33	350	4	30	45	54%
(5)A	Aluminum	PBI 33	350	4	60	45	49%
(6)A	Aluminum	PBI 33	350	4	100	45	44%
(7)A	Aluminum	PBI 33	400	2.5	30	90	57%
(8)A	Aluminum	PBI 33	400	2.5	60	90	48%
(9)A	Aluminum	PBI 33	400	2.5	100	90	36%
(1)C	Copper	MOC 24	300	4	30	45	62%
(2)C	Copper	MOC 24	300	4	60	45	65%
(3)C	Copper	MOC 24	300	4	100	45	69%
(4)C	Copper	MOC 24	450	2.5	30	45	87%
(5)C	Copper	MOC 24	450	2.5	60	45	85%
(6)C	Copper	MOC 24	450	2.5	100	45	85%
(7)C	Copper	MOC 24	450	4	30	45	79%
(8)C	Copper	MOC 24	450	4	60	45	81%
(9)C	Copper	MOC 24	450	4	100	45	81%



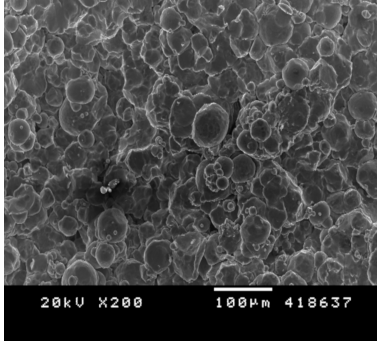
(a) 30 mm SOD Copper Coating



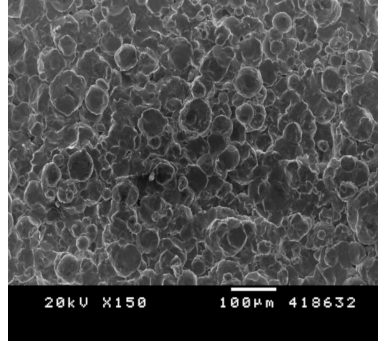
(b) 100 mm SOD Copper Coating



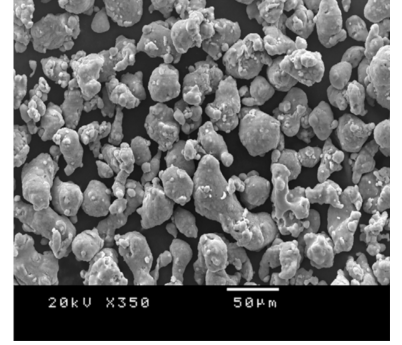
(c) Copper powder



(d) 30 mm SOD Aluminum Coating



(e) 100 mm SOD Aluminum Coating



(f) Aluminum powder

Figure 2: SEM free-surface micrographs for sprayed powders and coatings

surface of the resultant copper and aluminum coatings respectively. As can be seen from these micrographs, the particles in the outer surface of the coating correspond to the largest fraction of the particle size distribution shown in Figure 1 for aluminum and copper powders sprayed at short and large spraying distances. Several smaller craters were seen on deposited particles in all the samples. As the feedstock materials were spherical—with completely smooth surfaces in the case of the copper powder—we conclude that the craters were caused by smaller particles that did not achieve critical deposition energy, but bounced off the coating surface thereby reducing the deposition efficiency. Based on our findings, we consider that larger particles are more likely to be deposited. We use this criterion in later sections to estimate the fraction of the powder that is most likely to be deposited based on deposition efficiency determinations.

3. Machine Learning applied to coatings deposition data

In the previous section, a criterion for selecting the fraction of powder material that is most likely to be deposited onto the substrate was defined. The objective of this section is to implement machine learning models capable of predicting "deposited" and "not deposited" categories for cold-sprayed particles based on prior experience.

3.1. Description of energy-based window of deposition

Machine learning models need high information entropy within the dataset to function properly [28]; for this reason, any extra information on the particles during the cold-spray process can substantially improve the prediction models. In this section, feature engineering, data preprocessing, model training, and validation tasks are presented. As well, classification algorithms are used to describe an energy-based window of deposition based on decision boundaries.

3.1.1. Feature engineering from experimental data

3.2.1 To increase the information entropy in the sprayed-particle dataset, we determined fluid dynamics for the operating gas expanded through the cold-spray nozzle using the computational fluid dynamics software Fluent 16.2. The calculations were performed with a 2-D axisymmetric model and density-based algorithms. For each spraying operation in the experimental plan (Table 1), the combination of cold-spray parameters was used as the boundary conditions. To take into account the effect of turbulence on the nozzle walls, the k- ϵ algorithm was implemented. One hundred particles were uniformly sampled from the particle size distribution, to represent the particle size range for each powder; the impact velocity and temperature of the particles sampled were calculated using the high Mach number option in Fluent 16.2 and they were plotted as shown in Figure 3. To further increase the information in the particle dataset, the impact kinetic and thermal energies were calculated for each particle, as shown in Equations 1 and 2:

$$Ek_p = \frac{m_p V_p^2}{2} \quad (1)$$

$$Et_p = m_p c_p (T_p - T_{ref}) \quad (2)$$

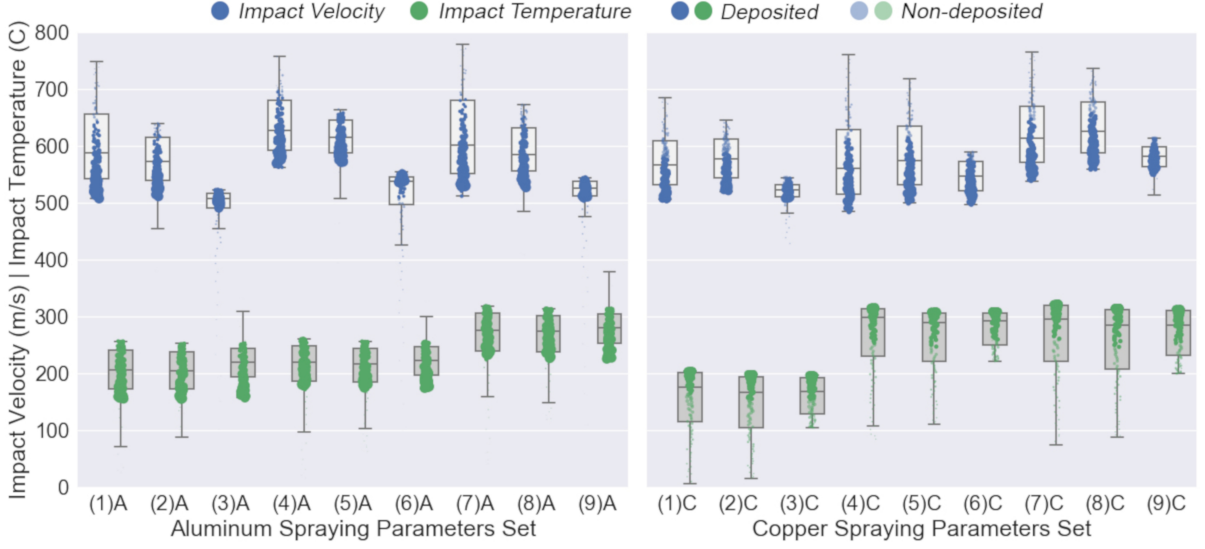


Figure 3: Particle impact velocity and temperature boxplots for each spraying operation. Each point represents a particle size, uniformly sampled from the powder distribution (the point sizes are scaled from the largest to the smallest particle size)

where Ek_p and Et_p are the kinetic and thermal energy, m_p and c_p are the mass and heat capacity of the particle, V_p and T_p are the particle’s impact velocity and temperature, and T_{ref} is the particle’s temperature prior spraying. The impact kinetic and thermal energies obtained after simulation for the d10, d50 and d90 particles are shown in Table 2.

Under the assumption that the largest particles within the distribution are more likely to be deposited, the particles were classified as ”deposited” and ”not deposited” based on the cumulative particle size distribution and using the deposition efficiency value to determine the corresponding critical diameter for each spraying operation. As shown in Figure 3, the particles labeled as ”deposited” have lower impact velocities in both the aluminum and copper coatings. Large copper particles achieved higher impact temperatures than in the aluminum spraying operations, where mid-sized particles reached higher impact temperatures.

3.1.2. Feature selection

An important step during the training data preprocessing is the selection of prediction features for the classification algorithms. The complexity of the classification model depends on the complexity of the data that is fed in: fewer features—without compromising the information content within the dataset—are preferred in order to reduce noise, outliers and dependence on particular samples, thereby making the model robust for most applications

Table 2: Impact energies for the d10, d50 and d90 particles in the cold-spray operations studied

Aluminum ($\rho=2700 \text{ kg/m}^3$, $c_p=902 \text{ J/kg}^\circ\text{C}$ (25°C))						
Operation	d10 (15.46 μm)		d50 (34.58 μm)		d90 (58.47 μm)	
	$E_{k_p} (\mu\text{J})$	$E_{t_p} (\mu\text{J})$	$E_{k_p} (\mu\text{J})$	$E_{t_p} (\mu\text{J})$	$E_{k_p} (\mu\text{J})$	$E_{t_p} (\mu\text{J})$
(1)A	1.39	0.62	12.54	13.11	50.21	62.10
(2)A	1.02	0.62	11.64	12.76	49.06	61.55
(3)A	0.40	0.91	7.23	12.98	38.59	62.04
(4)A	1.29	0.57	13.64	12.94	57.22	65.83
(5)A	1.04	0.55	12.54	12.49	56.28	64.66
(6)A	0.38	0.81	7.26	12.63	42.00	64.45
(7)A	1.55	0.80	13.22	16.19	52.61	84.66
(8)A	1.13	0.82	12.23	15.87	51.51	83.43
(9)A	0.46	1.11	8.10	15.99	41.24	83.36
Copper ($\rho=8960 \text{ kg/m}^3$, $c_p=384 \text{ J/kg}^\circ\text{C}$ (25°C))						
Operation	d10 (17.97 μm)		d50 (26.83 μm)		d90 (34.85 μm)	
	$E_{k_p} (\mu\text{J})$	$E_{t_p} (\mu\text{J})$	$E_{k_p} (\mu\text{J})$	$E_{t_p} (\mu\text{J})$	$E_{k_p} (\mu\text{J})$	$E_{t_p} (\mu\text{J})$
(1)C	4.94	1.09	14.59	5.48	28.90	13.75
(2)C	5.04	0.97	15.10	5.14	30.21	13.20
(3)C	3.83	1.12	12.82	5.18	27.16	13.08
(4)C	5.22	2.43	14.32	10.23	27.29	24.01
(5)C	5.31	2.31	14.98	9.84	29.28	23.24
(6)C	4.57	2.51	13.75	10.01	27.69	23.36
(7)C	6.03	2.29	17.14	10.05	33.56	24.14
(8)C	6.14	2.17	17.82	9.64	35.34	23.39
(9)C	4.93	2.34	16.13	9.68	33.07	23.22

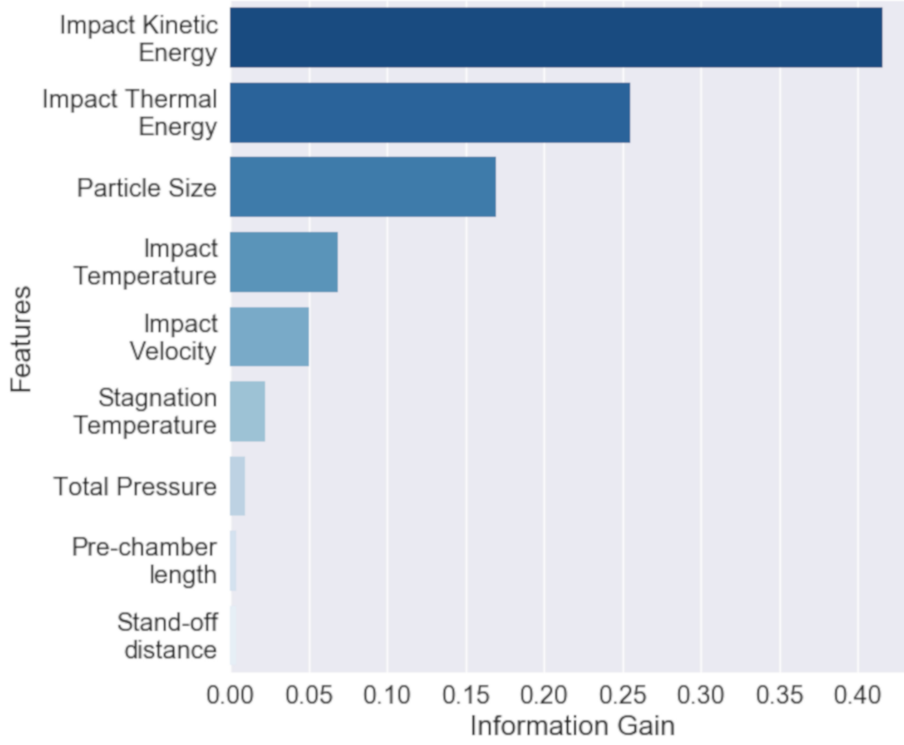


Figure 4: Information gain per feature in the sprayed-particles dataset

[15]. Several feature selection techniques are available in machine learning. In this work, features were selected based on information entropy measurements after splitting a single feature. The reduction in information entropy (impurity decrease) after splitting a feature is known as the information gain—a magnitude used to rate each feature within the dataset, where most significant features achieve higher scores [28]. The information gain for each feature in the sprayed-particles dataset (for both the copper and aluminum operations) was calculated using the *RandomForestClassifier* algorithm from the Scikit-Learn Python library [29]. Figure 4 shows the information gain for the features in the dataset. The kinetic Ek_p and thermal Et_p energies showed greater information gains than the particle size d_p , impact velocity V_p and temperature T_p features; the stagnation temperature T_0 , total pressure P_0 , stand-off distance SOD and pre-chamber length PC presented an information gain below 0.05. As the particle energy includes information from all the other features, the Ek_p and Et_p features were selected as the principal predictors. No data scaling was required since the selected features are of the same magnitude.

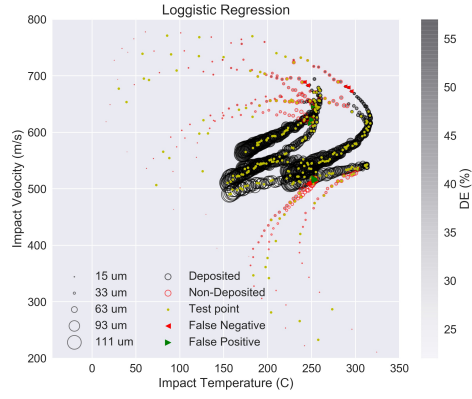
Table 3: Classification model hyperparameters

Classification Model	Hyperparameters			Accuracy	
		Aluminum	Copper	Aluminum	Copper
kNN [29]	Neighbors:	25	5	0.98	0.99
	Leaf Size:	10	10		
	Metric:	Minkowski	Minkowski		
LR [29]	C:	1	50	0.99	0.99
	Penalty:	L2	L2		
RF [29]	Criterion:	Gini	Gini	0.98	0.98
	Estimators:	10	100		
SVM [29]	C:	5000	5000	0.98	0.99
	Degree:	2	2		
	Gamma:	0.85	0.5		

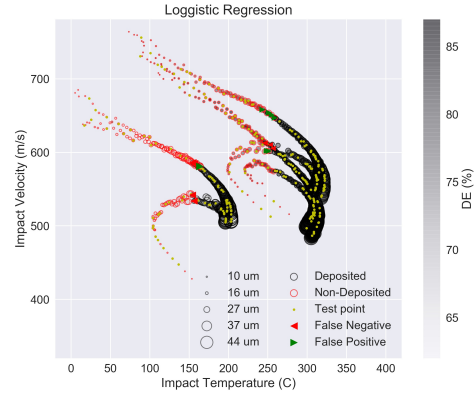
3.1.3. Classification model training and validation

Depending on the problem to be solved, certain machine learning algorithms will perform better than others. The way a machine learning algorithm "learns" is by generalizing assumptions from data [28]. In [7], a linear particle deposition boundary in the Ek_p and Et_p field was proposed under certain theoretical assumptions. Due to the complex nature of the particle consolidation process, we are not sure of the shape of this boundary. For this reason, several classification algorithms from the Scikit-Learn Python library [29] were implemented to test their predictive performance. The "test" and "train" portions were extracted randomly from the original dataset—with the new test dataset representing 30% of the original data—in order to test model performance against unique "never seen" data points. Linear and quadratic particle deposition boundaries were inferred from the training data using the *LR* and support vector machine (*SVM*) models respectively; irregular boundaries were inferred with the random forest (*RF*) and k-nearest neighbor (*kNN*) classifiers. The hyperparameters used for the selected classification models are shown in Table 3.

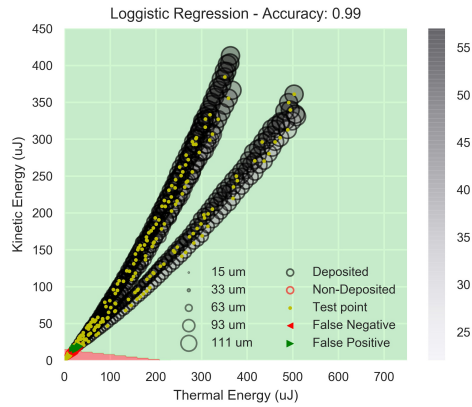
The prediction accuracy of each classification model was calculated against test data representing the fraction of correct predictions (Table 3). We evaluated the precision of the model using a confusion matrix where the true positive and negative, and false positive and



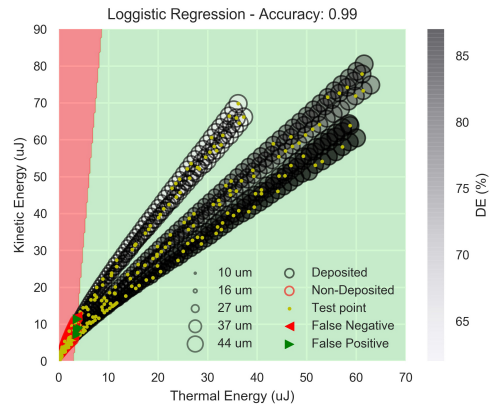
(a) Aluminum impact velocities and temperatures



(b) Copper impact velocities and temperatures



(c) Aluminum decision boundaries



(d) Copper decision boundaries

Figure 5: Decision boundaries and sprayed particle-related data

negative predictions are quantified (Figure 6).

3.1.4. Results

Figure 5c-d shows the decision boundaries inferred from a logistic regression. The "non-deposition" decision region is located in the lower energy zone of the plot - and highlighted in red - for the chosen classification algorithm; smaller particles are located in this region due to their lower energy values—in spite of their higher velocities. As expected, the inferred decision boundaries were different for each classification model. The particles sampled were grouped according to the deposition efficiency value of each spraying operation, as shown in Figure 5. Higher deposition efficiencies were achieved for copper particles, as they carried more thermal energy. In the case of aluminum, higher deposition efficiencies were obtained either under the spraying operations with the most kinetic energy or with the most ther-

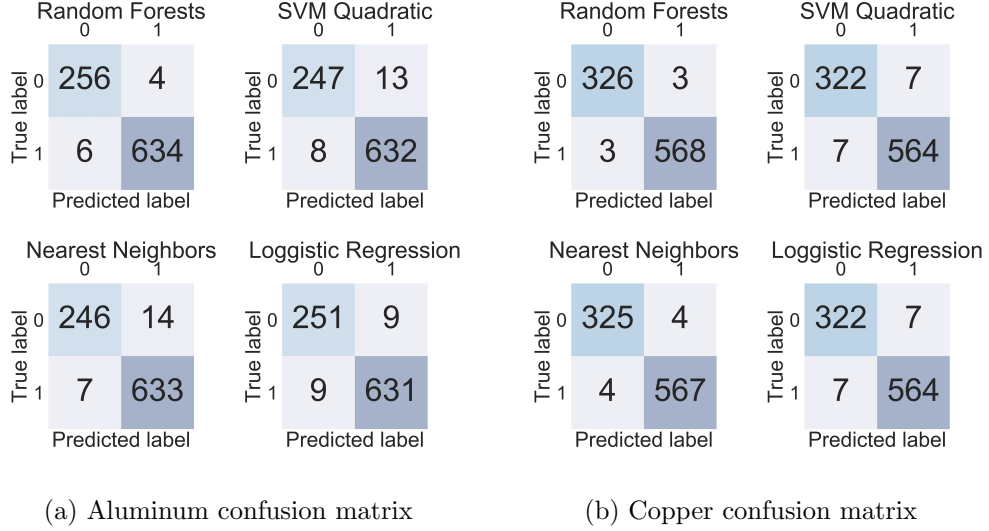


Figure 6: Confusion matrix for the classification models

mal energy. The decision boundaries inferred by the machine learning algorithms describe the energy-based window of deposition proposed in [7], with alternative decision boundary shapes. To understand better the relation between the classification of the particle bonding status and the impact velocities and temperatures, in Figure 5a-b the corresponding data-points are plotted. The *k-nearest neighbor* classification model was used in the visualization to distinguish “deposited” from “not deposited” particles, due to its performance, as assessed in Sections 3.1.3 and 3.2. As shown in Figure 5a, for the copper spraying operations, particles with higher impact temperatures were more likely to be classified as “deposited”; their deposition velocities are located in the 500-600 m/s range, which agrees well with deposition velocities reported by other authors [5]. The false positive and negative markers can be used to identify the temperature threshold at which deposition occurs for a certain particle impact velocity range. Figure 5b shows two ways to increase the deposition efficiency for aluminum powder operations: the first, by increasing the particle impact velocity; and the second, by increasing the impact temperature. Under aluminum spraying operations, the false positive and negative markers can be used to understand the dependence of the deposition threshold on particle size. In spite of the small aluminum particles reaching impact velocities and temperatures that are higher than those of their counterparts, their intrinsic total impact energy does not reach the deposition criterion of the classification models.

3.2. Deposition Efficiency Prediction

In the previous section, an energy-based window of deposition was described, taking advantage of different machine learning techniques. Other than describing deposition thresholds from experimental data, the models were not implemented for practical uses in spite of providing better understanding of the cold-spray process. In this section, we present the implementation of machine learning models for the prediction of deposition efficiency in the cold-spray process. Two different approaches are reviewed:

1. Deposition efficiency using classification models (from Section 3.1)
2. Deposition efficiency from regression models (implemented in this section)

Each approach is implemented and its advantages and limitations are discussed in Section 4.

3.2.1. Deposition efficiency using classification models

Deposition efficiency in the cold-spray process refers to the ratio between the masses of the material bonded on the substrate and the feedstock material fed into the system during the spraying operation. Different machine learning models were trained using datasets obtained from simulation and experiment, as presented in Section . We now use these models in this section to determine the size of the particles that are most likely to bond, in order to estimate the deposition efficiency for each spraying operation based on the cumulative particle size distribution shown in Figure 1.

In [30], an approach for predicting the deposition efficiency using a size-dependent function is presented, where the critical and erosion velocities of the particles were used as deposition thresholds. In the present study, we modified that approach by replacing the size-dependent function with the decision function from the machine learning classifiers presented in Section 3.1, as shown in Equation 3:

$$\hat{DE} = \int_0^\infty g([Ek_p, Et_p]) \times f(d_p) dd_p \quad (3)$$

where dd_p is the particle diameter differential, d_p is the particle diameter, $f(d_p)$ is the particle size distribution function and $g([Ek_p, Et_p])$ is the decision function from a classification model which processes the kinetic and thermal energies of particles as shown in Equation 4:

$$g([Ek_p, Et_p]) = \begin{cases} 0 & \text{if } [Ek_p, Et_p] \in \mathcal{R}_{Non-deposition} \\ 1 & \text{if } [Ek_p, Et_p] \in \mathcal{R}_{Deposition} \end{cases} \quad (4)$$

where $\mathcal{R}_{Non-deposition}$ and $\mathcal{R}_{Deposition}$ are the decision regions from the classification model.

3.3. Deposition efficiency from regression models

In the last section, the deposition efficiency was calculated based on discrete particle data and the continuous cumulative mass value from the particle size distribution. As the accuracy of calculations derived from discrete data models can be compromised by the Δd_p used to uniformly sample the particle population, in this section polynomial regression models are used to predict the deposition efficiency value directly from continuous experimental data.

First- and second-order polynomial models (Equations 5 and 6 respectively, using 2 prediction features) were implemented to predict and describe deposition efficiency, $\hat{DE}$, values for the cold-spray process. The kinetic and thermal energies of the particles (Ek_p and Et_p respectively) were selected as predictors due to their higher information gain within the particle dataset, as concluded in Section 3.1.2. The polynomial models were trained by implementing a least-squares regression, where the models weights, ω_n , were calculated using experimental data. The regression was performed using 2, 4 and 6 features [31] (corresponding to the $d90$, $d50$ and $d10$ particle kinetic and thermal energies) in order to examine the dependence of the model accuracy on the number of features.

$$\hat{DE} = \omega_0 + \omega_1 Ek_p + \omega_2 Et_p \quad (5)$$

$$\hat{DE} = \omega_0 + \omega_1 Ek_p + \omega_2 Et_p + \omega_3 Ek_p Et_p + \omega_4 Ek_p^2 + \omega_5 Et_p^2 \quad (6)$$

3.3.1. Results

The regression and classification models were trained with all the available data to evaluate how well they describe the cold-spray process in terms of the deposition efficiency. In order to evaluate their prediction performance, the models were trained using data from 6 spraying operations (1-6 in Table 1 for copper and aluminum) and tested against the rest of the experimental results. The *mean absolute percentage error* (MAPE) and the regression residuals were calculated in order to describe the performance of the models using Equations 7 and 8, where i is the operation number and n is the total number of operations.

Table 4: MAPE of machine learning models for description and prediction of deposition efficiency

Model Reference	Model Name	Prediction MAPE (%)		Description MAPE (%)	
		Copper	Aluminum	Copper	Aluminum
(C) kNN [29]	k-Nearest Neighbor Classifier	4.04	25.77	1.33	3.65
(C) LR [29]	Logistic Regression	4.66	18.47	3.75	16.59
(C) RF [29]	Random Forest Classifier	4.66	24.10	2.51	12.79
(C) SVM [29]	Support Vector Machine Classifier	2.15	14.81	4.16	19.13
Classification models average MAPE:		3.88	20.79	2.94	13.04
(R) L(2)	First-order Polynomial (2 features)	5.40	45.22	7.69	22.58
(R) L(4)	First-order Polynomial (4 features)	2.80	45.44	2.02	7.20
(R) L(6)	First-order Polynomial (6 features)	29.51	308.40	1.94	2.85
(R) Q(4)	Second-order Polynomial (4 features)	13.94	60.83	0.00	0.00
(R) Q(6)	Second-order Polynomial (6 features)	7.21	65.56	0.00	0.00
Regression models average MAPE:		11.77	105.09	2.33	6.53

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{DE_i - \hat{DE}_i}{DE_i} \right| \quad (7)$$

$$Residual_i = DE_i - \hat{DE}_i \quad (8)$$

Smaller MAPE values indicate better model performance. As shown in Table 4, aluminum deposition efficiency predictions resulted in higher MAPE values than those corresponding to copper operations. Regression models were better at describing the deposition efficiency response, while the classification-based models presented higher prediction accuracy for both aluminum and copper spraying operations.

Figure 7 shows a boxplot of the deposition efficiency calculation residuals for the models implemented. As depicted in this figure, the classification-based models presented a bias towards underestimation of deposition efficiency when trained with all the available data. Meanwhile, regression models presented a better distribution of residuals that decreased with the number of features and the degree of the polynomial model. The residuals of the regression models when performing the prediction tasks show overestimation of the deposition efficiency, with the worst results when predicting the aluminum deposition efficiency.

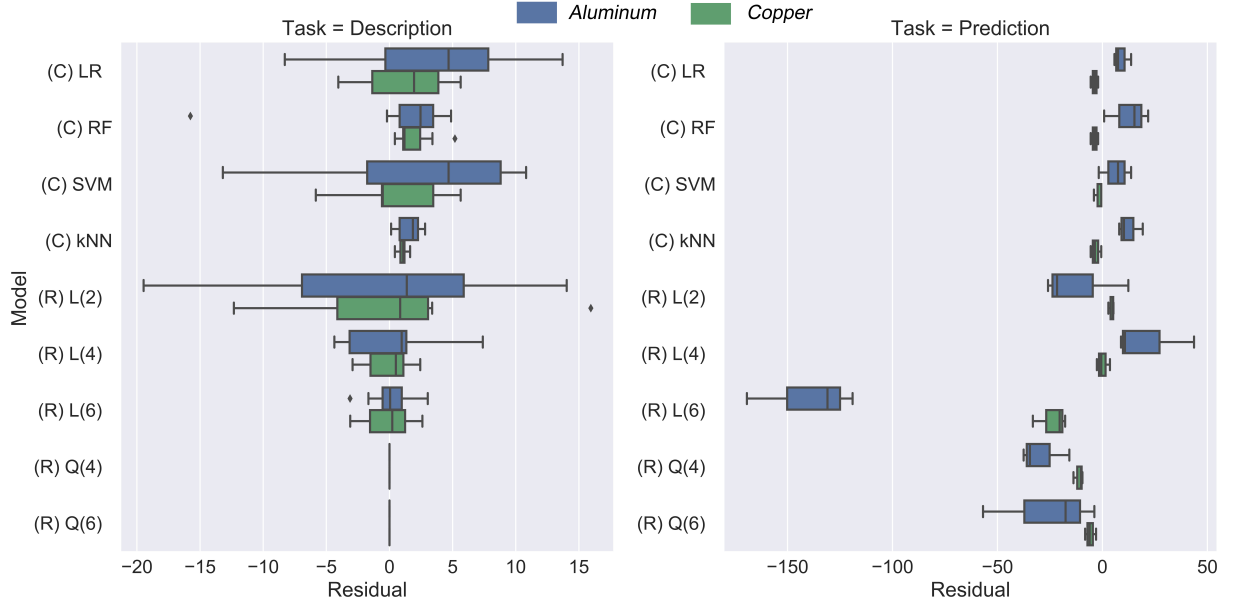


Figure 7: Model residuals

4. Discussion and further work

The main objective of this work is to propose a data driven methodology for studying cold spray depositions. The reason for setting attention on the methodology and not on the obtained models, is that a general model that would work for any powder material, would need a large dataset including powder material properties that are either hard or expensive to measure, leading to an impractical application of our research. We believe that less prior-spraying characterization and less experimental runs can lead to great results while taking advantage of machine learning methodologies - with a substantial impact on the research costs. Other powder materials may be better represented with models not covered in this work. The reason for this, is that other powder characteristics such as oxidation, plasticity, morphology, etc., may be unique from powder to powder with the same chemistry due to their specific processing, handling and storage.

In Section 3.1, we describe an energy-based window of deposition via machine learning techniques. Machine learning algorithms "learn" by generalizing assumptions about the data; for this reason, the results obtained from the developments presented here should be considered as heuristics: a generalization of the truth from which we hope to derive conclusions that are approximately correct [32]. The deposition boundaries generated by the proposed machine learning models may substantially change as more data is fed into them

from exploring other regions within the Ek_p and Et_p particle fields; these boundaries may present a different shape depending on the hyperparameter values used during the model training step. As the trade-off between bias and variance will define the correctness of the model [33], in Figure 5a-b the LR classifier was chosen to analyze the effect of impact velocity and temperature on the particle deposition, since the shape of the corresponding decision boundaries is in better agreement with the theoretical developments presented in [7].

The implementation of machine learning algorithms for the description of the window of deposition presented herein only considers a lower bound for the particle deposition threshold. A probable upper bound needs further research. Based on the findings of Sections 2.2 and 3.1.4, we interpret the phenomena of coating erosion as accelerated particles having sufficient energy to deform the substrate or coating but not enough to reach deposition conditions. This situation is typical in the smallest fraction of the particle size distribution; in spite of the higher impact velocities, the smallest particles are not deposited but cause craters on the surface of the coating, as shown in Figure 2 and reported in [26]. This interpretation of the erosion phenomenon agrees well with the total impact energy criterion introduced in Section 1, where the total energy of particles is considered as the deposition threshold. The proposed lack of small particles being deposited due to their low total impact energy—despite their higher impact velocities—coincides with the critical velocity size dependence concept presented in [6], where it is explained that smaller particles need higher impact velocities in order to be deposited.

In Section 2, we concluded that the fraction of the powder being bonded within the deposited material depends on the specific combination of spraying parameters. If the combination of parameters results in a high kinetic process with low thermal energy transferred to the particles—e.g., when using helium as the process gas—smaller particles are more likely to be deposited. In cases where higher thermal energies are transferred to the particles, larger particles will store more thermal energy—due to their lower velocity and larger heat capacity [7]—promoting their deformation and resulting in it being more likely that they are deposited. Since the work required to cause plastic deformation decreases at higher temperatures, and larger particles can store more thermal energy during a cold-spray operation, we can infer a correlation between the work performed during plastic deformation and the higher adhesion of smaller particles, as reported in [34]—further research on this subject is recommended.

The findings we present in Section 3.1.4 imply that the material critical velocity, as a function of the particle diameter and impact temperature, is not a concept that is easy to visualize—especially because of its dependence on particle size—whose complexity may increase considerable if other process characteristics (such a substrate preparation, powder oxidation, powder feed rate and particle interaction during deposition) are taken into account. For this reason, we consider that the generalizations made using machine learning methods are a good alternative to describe the scope of the cold-spray process, and demonstrate that their development should be further promoted. It is important to implement and develop new numerical methods for the study of materials and their transformation processes, since much of the equipment that is used today for different research activities is derived from numerical methods proposed in the past [35]. We consider that new knowledge can be derived from characterization data relating to existing materials using data science and visualization techniques.

From Figure 5c-d, it can be seen that particle impact kinetic and thermal energies show a quadratic relation that reveals a trend for process deposition efficiency across every spraying operation. This quadratic relation is the result of particle momentum and heat transfer during residence in the nozzle and represents a specific characteristic of the combination of cold-spray parameters. Further research into this characteristic of the cold-spray process is recommended. False positive and negative points shown in Figure 5 may represent the physical state of particles where deposition happen. As they are located within the decision boundaries of classification models, their physical state may be used as a boundary condition in explicit finite element simulations, in order to learn more about the deposition threshold in the cold-spray process.

Better methods to acquire training data are required to improve the implementation of machine learning techniques for the cold-spray process. In the present work, the criterion for particle classification was explained in Section 2.2, under the assumption that larger particles are more likely to be deposited. This assumption could be incorrect for some magnitudes since there could be large particles that are unable to adhere (supposing a hidden "non-deposition" region beyond the scope of the experimental procedures presented in Section 2) or very small particles may be deposited due to their large impact velocities and temperatures; as the mass contribution of these particles is very small compared with the contribution of

a particle of average size (at least for the powders studied), we consider that the sampling error in the present implementation is negligible. The particle kinetic and thermal energies were calculated ignoring the gas velocity profile across the cold-spray nozzle. It is assumed that machine learning models can generalize the results despite the fact that the gas velocity profile directly affects the overall particle acceleration.

In Section 3.2, we presented a prediction of the cold-spray process deposition efficiency using both classification and regression models. From the results for model accuracy, it was concluded that regression models will not predict the deposition efficiency accurately from "never seen" data—this was especially noticeable in the aluminum spray operations. As shown in Figure 7, these models overestimated the process deposition efficiency as a result of the bias after the polynomial regression. Regression models predicted higher deposition efficiencies due to the higher thermal energies resulting from operations 7-9 in Table 1 for aluminum powders. It is important to note that despite their low accuracy, the prediction gradient pointing towards an increase in efficiency which agrees well with experimental results. This property is important if we wish to take into account whether optimization tasks are planned in future work. Second-order polynomial models described the deposition efficiency data excellently using 4 and 6 prediction features; in spite of their probable over-fitted state, we should note that the quadratic nature of the models successfully describes the dependence between features (which in this case are the impact kinetic and thermal energies). Classification models predicted deposition efficiency from "never seen" data better, implying that their generalization of the "deposition" and "non-deposition" regions is probably approximately correct, in spite of the different characteristics of each material. It is important to add that the accuracy of predicting deposition efficiency from classification-based models is dependent on the correctness of the cumulative particle size distribution.

Uncertainty may be injected to the models either from numerical errors in the simulation or from experimental measurements. To assess the models prediction capabilities they were tested against a test dataset containing data "never-seen" by the model [36] (Figure 7). We acknowledge that the prediction capabilities of the models may be biased due to explored experimental domain, and the models may perform different as more data is collected.

Here, we have implemented machine learning models in order to find a numerical tool that will describe the effects of cold-spray parameters on the prediction of the deposition

efficiency of the process—which is highly correlated with many properties of the materials deposited [30]. The right selection of numerical models and predictive features will lead to a proper numerical representation of powder material/substrate systems for a wide range of cold-spray parameters. This representation can be called “the application fingerprint” and besides its potential application in optimization tasks, its predictive power—which increases as more experimental data is accumulated—may represent a core element in the pursuit of process–structure–property links for cold-sprayed materials.

5. Conclusions

In this work we have presented a description of the window of deposition and predictions of deposition efficiency for cold spraying via machine learning techniques. “Deposition” and “non-deposition” regions were inferred from experimental data—under the assumption that larger particles within the distribution are more likely to be deposited—using different classification models and taking into account the cumulative particle size distributions of the powders sprayed and the deposition efficiency values, as determined for each spraying operation. Sprayed particles were plotted using their impact kinetic and thermal energies, showing the simplicity of the total impact energy concept in comparison with the critical velocity criterion. Regression and classification models (originally implemented to describe the window of deposition for particles) were used to predict the deposition efficiency of the process. In copper spraying operations, the deposition efficiency increased when the thermal energy of the particles was higher; while for aluminum operations, this efficiency increased as either the kinetic or thermal energy of the particles augmented. The classification models showed better performance in predicting the deposition efficiency from “never seen” data; while the regression models described the deposition efficiency better when trained with all the available data points. Numerical representations of the powder material/substrate systems (referred to as “application fingerprints”) derived from machine learning models may play a key role in the pursuit of process–structure–property links and in the optimization of the properties of cold-sprayed materials. Despite the results presented here being heuristics, the methodologies implemented may help to uncover the main trends in the cold-spray process, thereby leading to its standardization in today’s industries.

- [1] A. Papyrin, V. Kosarev, S. Klinkov, A. Alkhimov, V. M. Fomin, Cold spray technology, Elsevier, 2006.
- [2] D. Gilmore, R. Dykhuizen, R. Neiser, M. Smith, T. Roemer, Particle velocity and deposition efficiency in the cold spray process, *Journal of Thermal Spray Technology* 8 (1999) 576–582.
- [3] H. Assadi, F. Gärtner, T. Stoltenhoff, H. Kreye, Bonding mechanism in cold gas spraying, *Acta Materialia* 51 (2003) 4379–4394.
- [4] C.-J. Li, W.-Y. Li, H. Liao, Examination of the critical velocity for deposition of particles in cold spraying, *Journal of Thermal Spray Technology* 15 (2006) 212–222.
- [5] M. Grujicic, C. Zhao, W. DeRosset, D. Helfrich, Adiabatic shear instability based mechanism for particles/substrate bonding in the cold-gas dynamic-spray process, *Materials & design* 25 (2004) 681–688.
- [6] T. Schmidt, F. Gärtner, H. Assadi, H. Kreye, Development of a generalized parameter window for cold spray deposition, *Acta materialia* 54 (2006) 729–742.
- [7] H. Assadi, H. Kreye, F. Gärtner, T. Klassen, Cold spraying—a materials perspective, *Acta Materialia* 116 (2016) 382–407.
- [8] H. Assadi, I. Irkhin, H. Gutzmann, F. Gärtner, M. Schulze, M. V. Vidaller, T. Klassen, Determination of plastic constitutive properties of microparticles through single particle compression, *Advanced Powder Technology* 26 (2015) 1544–1554.
- [9] Z. Arabgol, M. V. Vidaller, H. Assadi, F. Gärtner, T. Klassen, Influence of thermal properties and temperature of substrate on the quality of cold-sprayed deposits, *Acta Materialia* 127 (2017) 287–301.
- [10] T. Van Steenkiste, J. Smith, R. Teets, J. Moleski, D. Gorkiewicz, R. Tison, D. Marantz, K. Kowalsky, W. Riggs, P. Zajchowski, et al., Kinetic spray coatings, *Surface and Coatings Technology* 111 (1999) 62–71.
- [11] K. Kang, S. Yoon, Y. Ji, C. Lee, Oxidation dependency of critical velocity for aluminum feedstock deposition in kinetic spraying process, *Materials Science and Engineering: A* 486 (2008) 300–307.

- [12] A. Agrawal, P. D. Deshpande, A. Cecen, G. P. Basavarsu, A. N. Choudhary, S. R. Kalidindi, Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters, *Integrating materials and manufacturing innovation* 3 (2014) 8.
- [13] K. Rajan, Materials informatics, *Materials Today* 8 (2005) 38–45.
- [14] S. R. Kalidindi, M. De Graef, Materials data science: current status and future outlook, *Annual Review of Materials Research* 45 (2015) 171–193.
- [15] E. Alpaydin, Introduction to machine learning, 2nd edn. adaptive computation and machine learning, 2010.
- [16] T. Wuest, D. Weimer, C. Irgens, K.-D. Thoben, Machine learning in manufacturing: advantages, challenges, and applications, *Production & Manufacturing Research* 4 (2016) 23–45.
- [17] T. Gurgenc, O. Altay, M. Ulas, C. Ozel, Extreme learning machine and support vector regression wear loss predictions for magnesium alloys coated using various spray coating methods, *Journal of Applied Physics* 127 (2020) 185103.
- [18] O. Altay, T. Gurgenc, M. Ulas, C. Özel, Prediction of wear loss quantities of ferro-alloy coating using different machine learning algorithms, *Friction* 8 (2020) 107–114.
- [19] D. Ikeuchi, A. Vargas-Uscategui, X. Wu, P. C. King, Neural network modelling of track profile in cold spray additive manufacturing, *Materials* 12 (2019) 2827.
- [20] A. Behera, S. Mishra, Prediction and analysis of deposition efficiency of plasma spray coating using artificial intelligence method, *Open Journal of Composite Materials* 2 (2012) 54–60.
- [21] D. Thirumalaikumarasamy, K. Shanmugam, V. Balasubramanian, Establishing empirical relationships to predict porosity level and corrosion rate of atmospheric plasma-sprayed alumina coatings on az31b magnesium alloy, *Journal of Magnesium and Alloys* 2 (2014) 140–153.

- [22] C.-J. Li, W.-Y. Li, Y.-Y. Wang, G.-J. Yang, H. Fukanuma, A theoretical model for prediction of deposition efficiency in cold spraying, *Thin Solid Films* 489 (2005) 79–85.
- [23] C.-J. Li, G. Yang, C. Li, H. Bang, W. Li, S. d. S. LERMPS-UTBM, F. Belfort Cedex, Examination of the estimating approaches for the critical velocity in cold spraying, in: *Thermal Spray 2007: Global Coating Solutions: Proceedings of the 2007 International Thermal Spray Conference*, ASM International, 2007, p. 128.
- [24] T. Stoltenhoff, H. Kreye, H. Richter, An analysis of the cold spray process and its coatings, *Journal of Thermal Spray Technology* 11 (2002) 542–550.
- [25] F. Raletz, M. Vardelle, G. Ezo'o, Critical particle velocity under cold spray conditions, *Surface and Coatings Technology* 201 (2006) 1942–1947.
- [26] W. Wong, E. Irissou, A. N. Ryabinin, J.-G. Legoux, S. Yue, Influence of helium and nitrogen gases on the properties of cold gas dynamic sprayed pure titanium coatings, *Journal of Thermal Spray Technology* 20 (2011) 213–226.
- [27] G. Bae, K. Kang, H. Na, J.-J. Kim, C. Lee, Effect of particle size on the microstructure and properties of kinetic sprayed nickel coatings, *Surface and Coatings Technology* 204 (2010) 3326–3335.
- [28] J. R. Quinlan, Induction of decision trees, *Machine learning* 1 (1986) 81–106.
- [29] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: Machine learning in python, *Journal of Machine Learning Research* 12 (2011) 2825–2830.
- [30] H. Assadi, T. Schmidt, H. Richter, J.-O. Kliemann, K. Binder, F. Gärtner, T. Klassen, H. Kreye, On parameter selection in cold spraying, *Journal of thermal spray technology* 20 (2011) 1161–1176.
- [31] J. VanderPlas, *Python data science handbook: Essential tools for working with data*, ”O'Reilly Media, Inc.”, 2016.
- [32] L. Valiant, *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World*, Basic Books (AZ), 2013.

- [33] Y. S. Abu-Mostafa, M. Magdon-Ismail, H.-T. Lin, Learning from data, volume 4, AML-Book New York, NY, USA:, 2012.
- [34] D. Goldbaum, J. M. Shockley, R. R. Chromik, A. Rezaeian, S. Yue, J.-G. Legoux, E. Irissou, The effect of deposition conditions on adhesion strength of ti and ti6al4v cold spray splats, Journal of thermal spray technology 21 (2012) 288–303.
- [35] J.-C. Zhao, High-throughput experimental tools for the materials genome initiative, Chinese science bulletin 59 (2014) 1652–1661.
- [36] H. S. Atamturktur, B. Moaveni, C. Papadimitriou, T. Schoenherr, Model Validation and Uncertainty Quantification, Volume 3, Springer, 2015.