Machine learning-based approach for filling gaps in data on plasma-related processes

Prediction of Sputtering Yields

Kateryna Lemishko

k.lemishko@quantemol.com

October 16, 2024

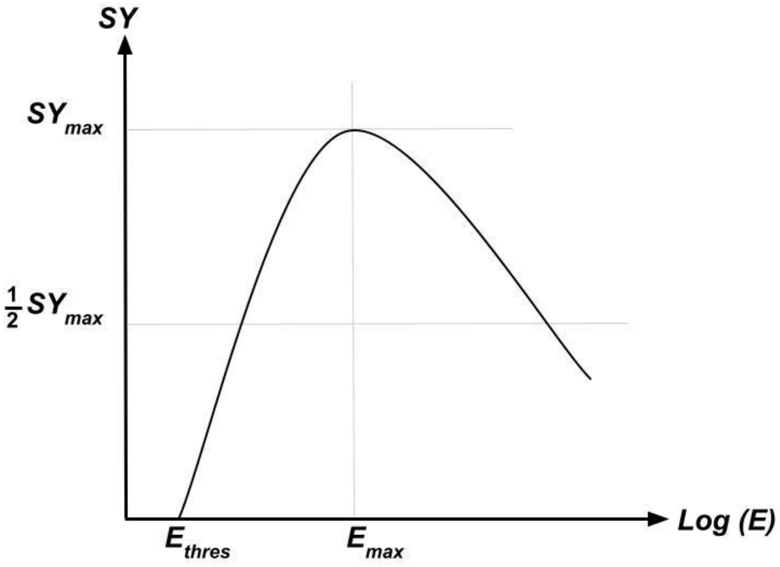
Introduction

Sputtering is the removal of atoms from a lattice by particle (ion) impact

Sputtering yield (SY) is the number of atoms removed from the surface by a single impact

Sputtering Yield depends on impact energy, angle and target-projectile combination

Experiments and MD simulation are used to measure or estimate sputtering yields, but both these approaches are time consuming and difficult to implement



We can use existing sputtering yield data to train a machine learning model capable of predicting sputtering yields for target-projectile combinations that has never been studied before.

Supervised learning

Input features : $X = (x_1, x_2, \dots x_n)$
(independent variables)

→

Model
y=f (x)

→

Target : y
(dependent variable)

A supervised machine learning algorithm learns relationships between input data (input features) and known responses to the data (target) so it can generate reasonable predictions for the response to new data.

Dataset to predict Sputtering Yield for single-element targets

Target dependent variables:

SY_{max} and E_{max} values at normal incidence

267 single-element target-ion combinations

Input descriptors (independent variables):

Easily accessible data describing chemical and physical properties of projectile ions and target materials

Predict SY_{max} and E_{max} with Machine Learning

Dataset to predict Sputtering Yield for single-element targets

quantemol

Target dependent variables:

SY_{max} and E_{max} values at normal incidence

267 single element target-ion combinations

Input descriptors (independent variables):

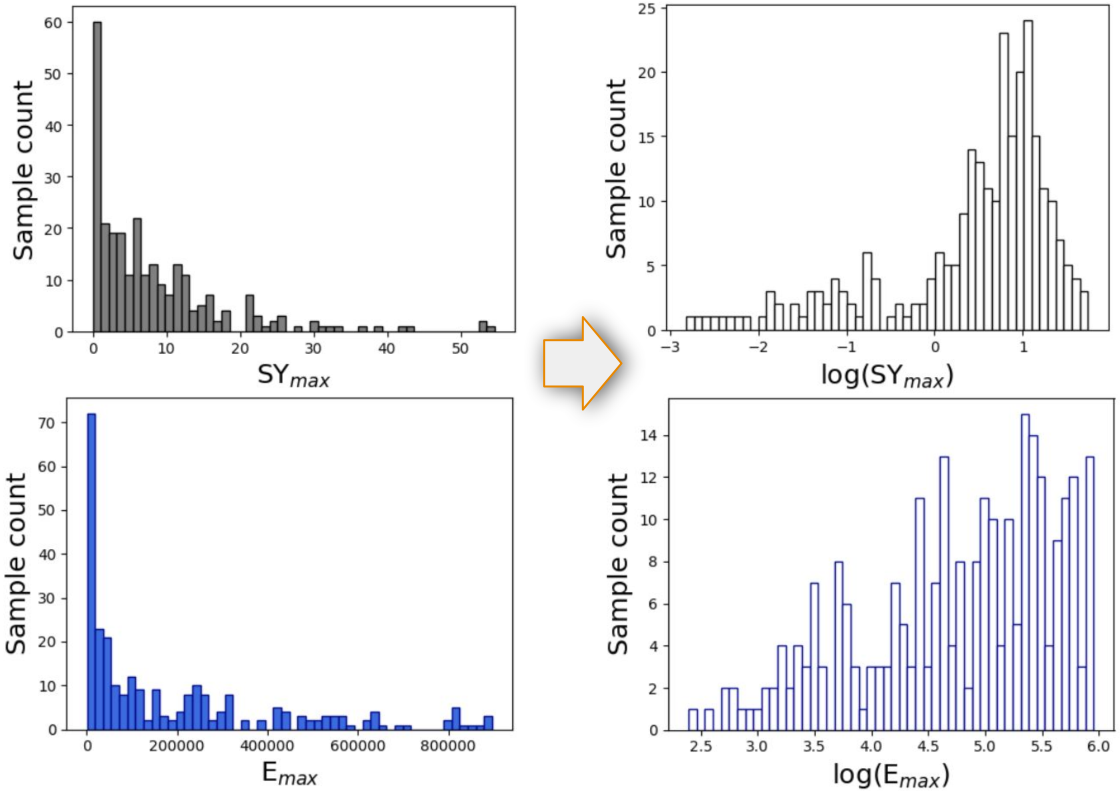
Easily accessible data describing chemical and physical properties of projectile ions and target materials

19 descriptors

Target	Projectile
Atomic number, Atomic mass	Atomic number, Atomic mass
Melting point, Boiling point	Melting point, Boiling point
Energy of evaporation, Enthalpy of formation (gas)	Energy of evaporation (neutral), Energy of formation (neutral, gas phase)
Density	—
First ionization potential	First ionization potential (neutral particle)
Atomic, Covalent and VdW radii	Atomic, Covalent and VdW radii

Data pre-processing for Machine Learning

1. Log-transformation of the target variables



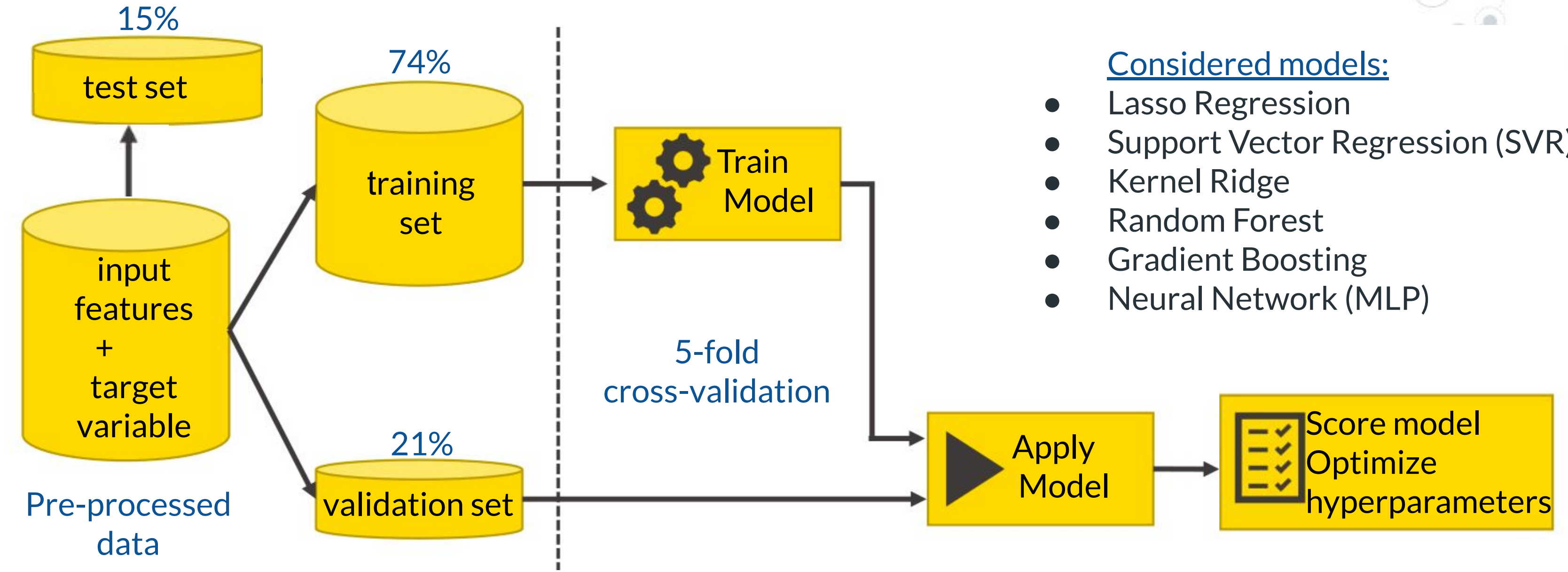
2. Missing value imputation with a simple Machine Learning algorithm

3. Normalization on the input data to bring all the descriptors to the same scale

Machine learning workflow

Data partitioning

Model training and validation



Considered models:

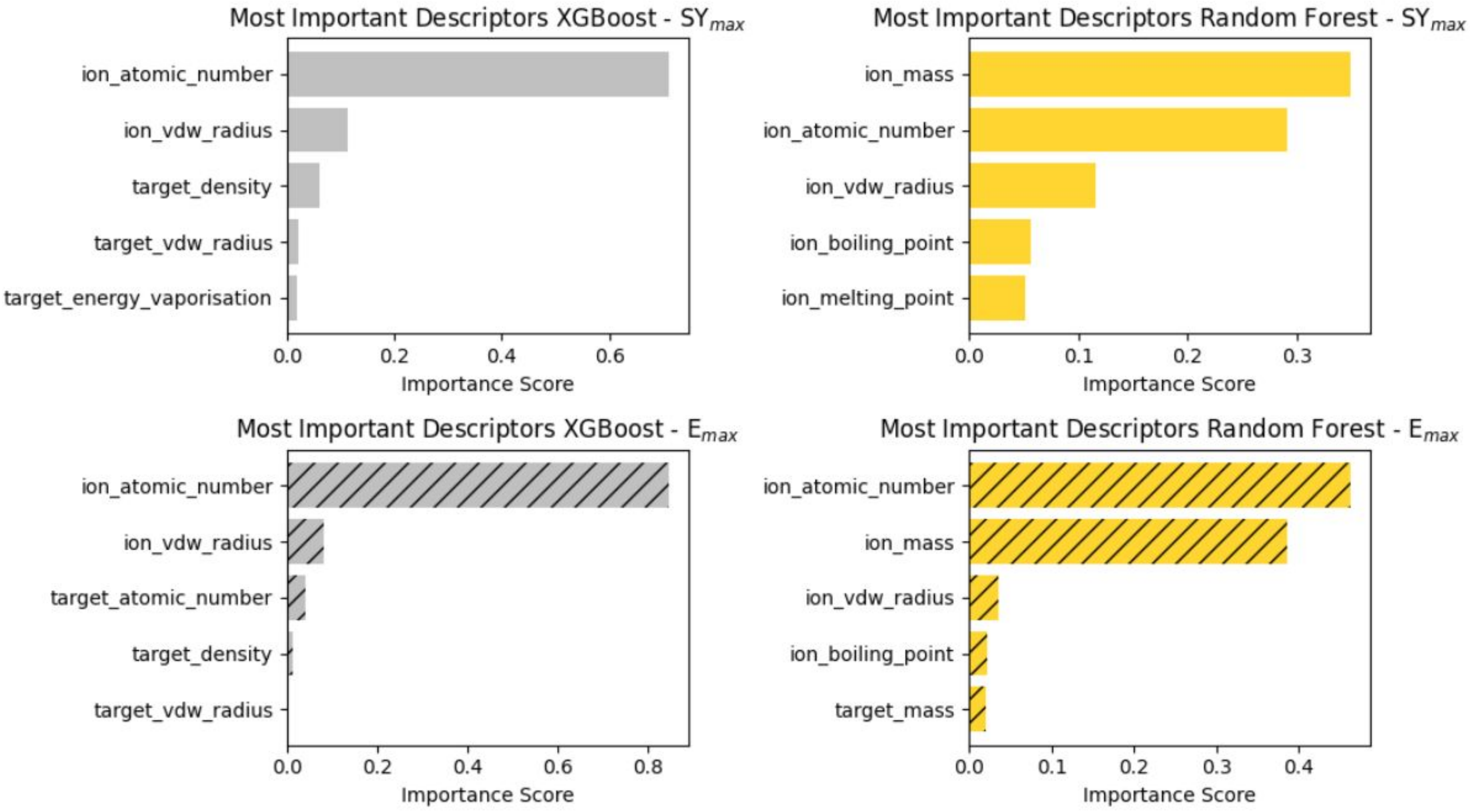
- Lasso Regression
- Support Vector Regression (SVR)
- Kernel Ridge
- Random Forest
- Gradient Boosting
- Neural Network (MLP)

Best predictions: Voting regressors combining different machine learning models

$log(SY_{max})$ = 50% **XGBoost +** 20% **Random_Forest** + 20% **SVR +** 10% **Kernel_Ridge**

$log(E_{max})$ = 40% **XGBoost +** 40% **Random_Forest** + 20% **SVR**

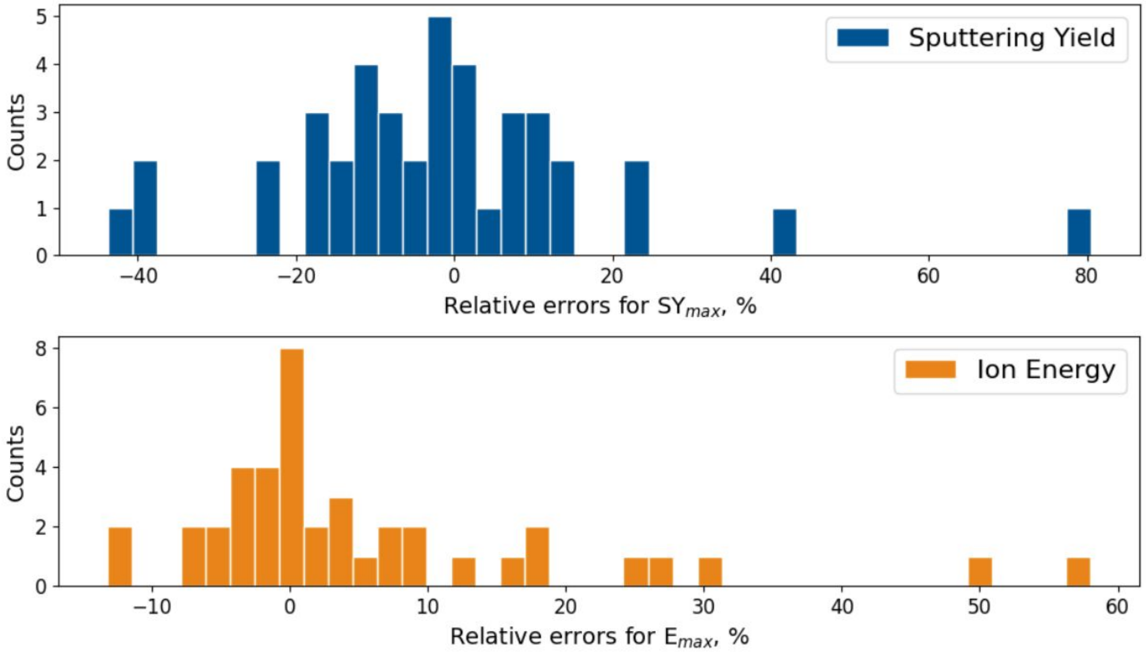
Most important descriptors



Machine Learning predictions for test target-ion pairs

Test set: 41 target-ion pairs

Target	Mean Absolute Error **Validation sets**	Mean Absolute Error **Test set**
Sputtering Yield log(**SY**$_{max}$)	0.0712	0.0653
Ion energy log(**E**$_{max}$)	0.0382	0.0366



SY predictions for targets composed of multiple elements

quantemol

New data set

Sputtering yield data for

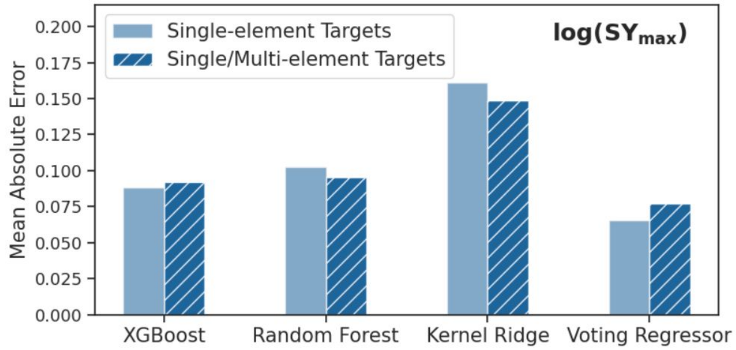
- 51 target - ion pairs with target materials composed of multiple atoms (N_{atoms} =1...7)
- 284 single - element target - ion pairs

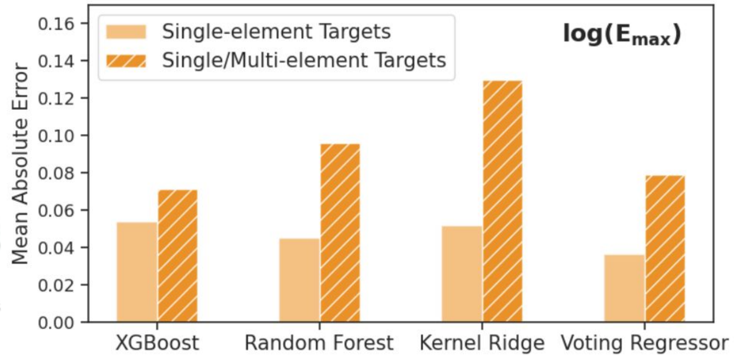
Updated descriptors

Target	Projectile
Total atomic number, Average atomic number, Molecular mass	Atomic number, Atomic mass
Melting point, Boiling point	Melting point, Boiling point
Enthalpy of formation (gas)	Energy of formation (neutral, gas phase)
Density	—
—	First ionization potential (neutral particle)
Bulk Modulus, Lattice Volume,Lattice Type	—
VdW radius	Atomic, Covalent and VdW radii
N atoms from different groups and blocks in the periodic table	Group and block in the periodic table

Machine Learning predictions for test target-ion pairs

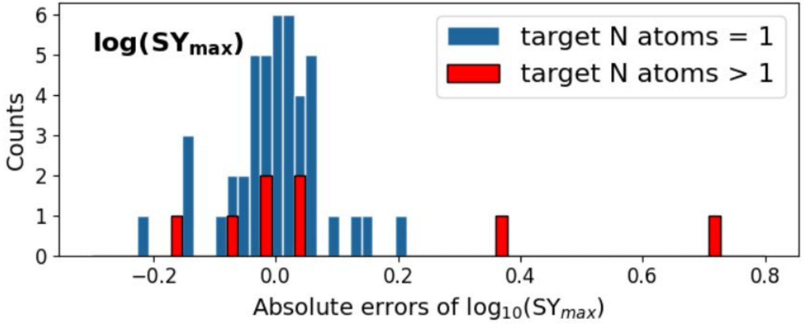
Test data error comparison: single-element vs. mixed targets

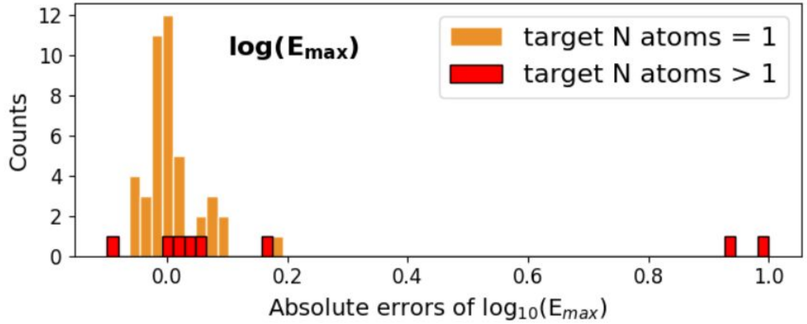




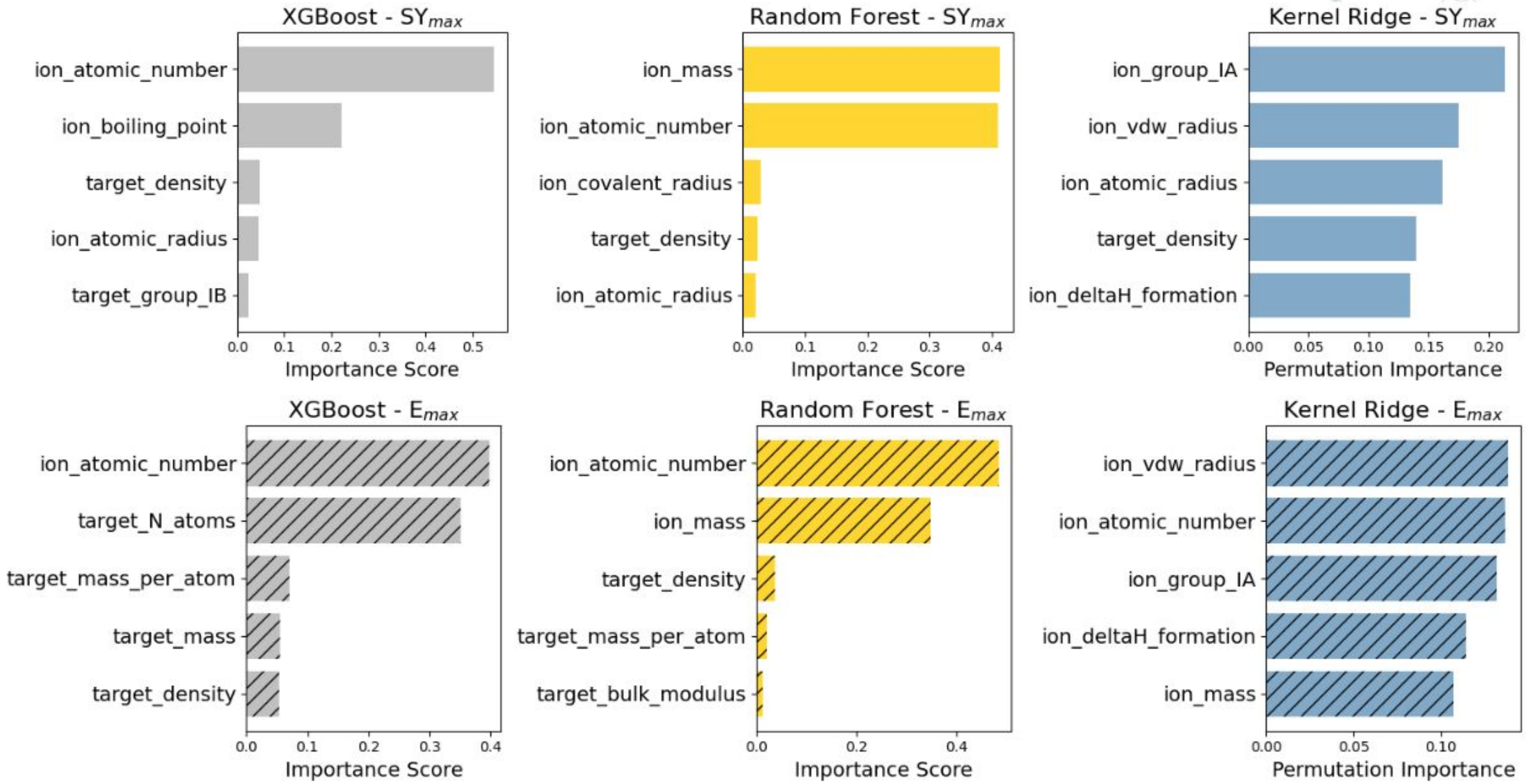
Mixed targets test set: 51 target-ion pairs

Mean Absolute Error distribution for Voting Regressor





Most important descriptors - targets composed of multiple elements



Conclusions

- Machine learning provides a cost-effective and efficient alternative to traditional experiments and simulations for predicting complex outcomes.

- We have developed a machine learning-based algorithm capable of accurately predicting sputtering yields using readily available data on the chemical and physical properties of target materials and projectile ions.