Prediction of chemical plants operating performances: a machine learning approach

Lelio Campanile
Luigi Piero Di Bonito
Mauro Iacono
Dipartimento di Matematica e Fisica
Università degli Studi della Campania "L. Vanvitelli"
viale Lincoln 5
81100, Caserta, Italy

Francesco Di Natale
Dipartimento di Ingegneria Chimica, dei Materiali e della Produzione Industriale
Università degli Studi di Napoli Federico II
Piazzale Vincenzo Tecchio, 80
80125 Napoli, Italy

KEYWORDS

Marine pollution; Sulphur dioxide absorption; Scrubber; Machine Learning; Chemical engineering.

ABSTRACT

Modern environmental regulations require rigorous optimization of operations in process engineering to reduce waste, pollution, and risks while maximizing efficiency. However, the nature of chemical plants, which include components with non-linear behavior, challenges the use of consolidated tuning and control techniques. Instead, ad-hoc, self-adapting, and time-variant controls, with a balanced tuning of parameters at both the subsystem and system level, may be necessary. Needed computing processes may require significant resources and high performance systems, if managed by means of traditional approaches and with exact solution methods. In this regard, domain experts suggest instead the use of integrated techniques based on Artificial Intelligence (AI), which include Explainable AI (XAI) and Trustworthy AI (TAI), which are unique in this industry and still in the early stages of development.

To pave the way for a real-time, cost-effective solution for this problem, this paper proposes an AI-based approach to model the performance of a real chemical plant, i.e. a marine scrubber installed on a Ro-Ro ship. The study aims to investigate Machine Learning (ML) techniques which can be used to model such processes. Notably, this analysis is the first of its kind, at the best of the authors' knowledge. Overall, the study highlights the potential of using ML-based techniques, to optimize environmental compliance in the shipping industry.

I. INTRODUCTION

Modern environmental regulations necessitate rigorous optimization of operations which are involved in process engineering in order to decrease waste, pollution, and risks, as well as maximize the efficiency of each step and sub-system. Managing compliance requires significant computational efforts and non-negligible performances to ensure that systems keep all operational parameters within the boundaries that allow a safe evolution of their dynamics, with real-time verification and adjustment of all internal and external variables. Considering chemical processes, the nature of chemical plants, which include non-linear components and could constitute one-of-a-kind elements of a chemical plant, these requirements challenge the consolidated tuning and control techniques and suggests the use of ad-hoc, self-adapting, and time-variant controls, possibly with a balanced tuning of parameters at both the subsystem and the system level.

As the real-time computing operations have to be performed on-site to guarantee that the control loop is closed and timely, the case of processes which happen on ships, without the constant supervision of a full team of IT personnel and with limited assets in a non-friendly environment, with a need for constant monitoring and intervention, suggests a quest for solutions that may be implemented with reduced devices. Domain experts in the process engineering area suggest the use of integrated techniques based on Artificial Intelligence (AI) or, even more interesting, Explainable (XAI) or Trustworthy AI (TAI), which are unique in this industry and are still in the early stages of development. The use of XAI/TAI techniques is significant for the process safety and the imputation of responsibility in case of failures.

Shipping transports almost 90% of the world's commerce annually and is critical to international trade and the global economy. Shipping produces higher sulphur emissions per tonne-mile of cargo than other modes of transportation, owing to the high sulphur content of the adopted fuels.

Sulphur compound limitations established by the International Maritime Organization (IMO) under Annex MARPOL VI regulation are achieved by using the marine scrubbers. Compliance has been established for two conditions: sulphur emissions must be equivalent to those produced by a fuel containing a sulphur concentration lower than 0.1% w/w for vessels traveling in Sulphur Emission Control Areas (SECA); elsewhere, a

Communications of the ECMS, Volume 37, Issue 1,
Proceedings, ©ECMS Enrico Vicario, Romeo Bandinelli,
Virginia Fani, Michele Mastroianni (Editors) 2023
ISBN: 978-3-937436-80-7/978-3-937436-79-1 (CD) ISSN 2522-2414

worldwide limit equivalent to sulphur concentration in fuels lower than 0.5% w/w applies. Marine scrubbers work under a range of conditions relating to the route of the ship, weather conditions, and engine running, which is also dependent on ship movement. Marine engines are either four-stroke or two-stroke diesels that run on heavy fuel oils. They are a blend of diesel fuels and mineral oils, and their qualities are governed by ISO 8217:2017 Petroleum products — Fuels (class F) — Maritime fuel specifications also known as Residual Marine Fuels (RMx). These fuels are distinguished by varying sulphur content and a non-negligible amount of ashes, ranging from 0.040 to 0.150% w/w. In most situations, the sulphur concentration of RMx utilized onboard ships varies from 2 to 3.5% w/w of the fuels, resulting in an average $SO_{2_{(g)}}$ concentration in the exhaust gases of 400-800 ppmv. Sulphur is also released in the form of $SO_{3_{(g)}}$, $H_2SO_{4_{(l)}}$, and sulphate particles. Compliance with the ship emission restrictions of the MARPOL Annex VI Regulation 14 [5] indicates that a marine scrubber must be built to ensure $SO_{2_{(g)}}$ removal efficiency above 97% in SECA zones.

In this paper, a modeling approach based on ML techniques is presented for a real scrubber installed on a Ro-Ro ship (cargo ship), considering as target variable the $SO_{2_{(g)}}$ scrubber outlet concentration. The aim of this research is to understand if this kind of processes could be modeled by using explainable machine learning models. The main original contribution is the application of this kind of modeling on a real dataset: at the best of our knowledge, no such analysis is available in literature. After this section, the paper is structured as follows: the next section summarizes related work and provides a brief background on possible AI uses in process engineering. The case study and the used dataset are then described; after that, the methods utilized in this paper to develop the model using machine learning are presented; results and discussion close the paper, as well as future work and advances.

II. RELATED WORKS

There are different examples in literature of AI applications in process engineering and the main problems are related to prediction/modeling, optimization, control and fault diagnosis.

Considering the prediction/modeling challenges, in [2] the authors implement AI techniques to predict NO_X emissions from coal-powder power plants, in [1] the AI was used to evaluate the operation of a wet scrubber system for air pollution management and in [11] the collection efficiency of Venturi scrubbers was evaluated by using different AI techniques; the work in [12] present an artificial intelligence inference system that minimizes the uncertainty of traditional approaches of risk assessment in pipelines by using case study from the Colombian oil transportation network while in [10] an AI technique was implemented to address the numerical solutions of a adsorption fixed-bed column where a monoclonal antibody is purified.

Regarding the optimization problems, in [8] the authors propose a methodology for optimizing the energy efficiency of an atmospheric distillation unit without sacrificing product quality or process throughput, whereas the case studies in [14] present an AI-based real time optimization (RTO) for two chemical process examples: a Continuous Stirred Tank Reactor (CSTR) and a distillation column. Related to control issues, after the RTO analysis, the authors successfully updated the control systems of both processes using AI approaches; always considering the control applications, in [13] an AI based control-logic system was implemented to regulate product compositions of distillation.

As regards fault diagnosis, in [7] the authors monitored and analyzed flows and compositions of the intermediate streams of a wastewater treatment plant, while in [6] they proposed a fault diagnostic system for a distillation process.

Fig. 1: Ship Route

III. THE CASE STUDY

The case study is based on real-time data from an open loop scrubber installed on a cargo ship owned by Grimaldi Group. In the reference year, the maritime route reported in Fig. 1 runs from Bilbao (Spain) to St. Petersburg (Russia).

The ship exhaust gases cleaning technology is based on seawater scrubbing, this cleaning technology depends on chemical-physical properties of the seawater such as salinity, alkalinity and temperature. These parameters of seawater depend on the sea crossed along the trip by the ship.

The equipment used in the process is known as *scrubber*, it is intended to run continuously in wet circumstances. The two combustion units transport exhaust fumes to the scrubber. The seawater is collected and injected into the scrubber, where it is sprayed.

A continuous emission monitoring system (CEMS) is installed on board to demonstrate that the SO_2/CO_2 ratio at discharge is less than or equal to the required SO_2/CO_2 (i.e. 21.7 ppm/%vv out the SECA zones and 4.3 ppm/%vv in the SECA zones) at any loading point and therefore complies with Regulation 14. Characteristics of any wash water discharged into the sea are continuously monitored and data for the following parameters must be recorded with respect to time and

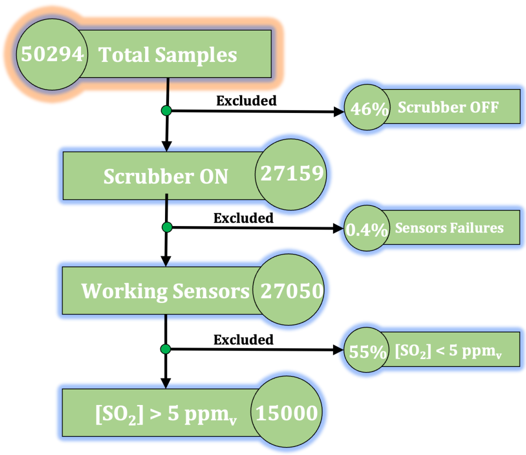

Fig. 2: Data cleaning process

location of the vessel:

- pH (a measure of acidity);
- PAH (a measure of harmful oil components);
- Turbidity (a measure of particulate matter);
- Nitrates.

IV. METHODOLOGY

A. *Dataset Description*

The dataset used in this work was collected directly on board the ship described in the case study.

All the data are recorded every ten minutes and are stored on board in a data base. In addition to the regulated parameters, this database contains several operational variables related to ship operation, Venturi and scrubber operations and several water and environmental parameters. Data retrieval occurred from 00:00 on January 1st, 2017 to 8:00 on December 16th, 2017 for a total of 50294 samples.

Before proceeding with the data analysis, we removed the samples where the scrubber was switched off because the data acquisition was switched on anyway. This operation reduced the dataset by 46%. Then a small number of samples had zero CO_2 concentration with the engines switched on. After this the dataset reduced of 0.4%. Lastly, all samples with a SO_2 concentration of less than 5 ppm were deleted, because they are too close to the detection limit of the instrumentation.

The obtained dataset contains around 15.000 samples and the data cleaning process is summarized in Fig. 2. The features of the dataset could be divided in three main sections:

- ship information
- Venturi/scrubber data
- seawater parameter

The ship information section includes parameters such as latitude, longitude, ship speed, and fuel type. The Venturi/scrubber data section includes parameters such as inlet flow and pressure of seawater, inlet pressure of Venturis, and differential pressure of scrubber vents. The seawater parameter section includes salinity of seawater.

Table I summarizes the features of the dataset.

B. *Data Analysis*

The first step was to gain insight into the attribute properties of the dataset, with descriptive statistics summarized in Table II. Then, a graphical analysis was performed to highlight the attribute characteristics, and Fig. 3 provides a general overview of the distribution of each attribute.

The Python programming language and the Pandas library in the Jupyter Notebook environment were mainly used for the data analysis of the comprehensive dataset.

C. *Evaluation Metrics*

To confirm the ability of the regression model to provide accurate predictions, the dataset was separated into 80% training and 20% test portions. To assess prediction ability, the Mean Square Error (MSE) (Eq.(1)), Mean Absolute Error (MAE) (Eq.(2)) and Coefficient of determination (R^2) (Eq.(3)) were determined.

$$MSE(y, \widehat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \widehat{y}_i)^2 \tag{1}$$

$$MAE(y, \widehat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} |y_i - \widehat{y}_i| \tag{2}$$

$$R^2(y, \widehat{y}) = 1 - \frac{\sum_{i=0}^{n} (y_i - \widehat{y}_i)^2}{\sum_{i=0}^{n} (y_i - \bar{y}_i)^2} \tag{3}$$

where:

- y_i is the prediction;
- $\widehat{y}_i$ is the experimental value;
- $\bar{y}_i = \frac{1}{n} 1 \sum_{k=1}^{n} y_i$ is the mean of the true values.

MAE and MSE are metrics that measure the expected value of the error and the quadratic error, respectively, while R^2 represents the proportion of variance of y and provides a general measure of the adequacy of fit of the model.

D. *Regression Models*

In order to perform the regression task, we implemented six different regression models, four are related to Linear Regressions, these are Ordinary Least Square Regression (OLS), Ridge Regression, Lasso Regression and the Stochastic Gradient Discent (SGD). The last two are the k-Nearest Neighbors Regressor (kNN-R) and Support Vector Machine Regressor (SVM-R).

Regarding the set of models for the linear regression, in these models the target value is expected to be a linear combination of the features. We use $x \in R^n$ to describe the input data, with n input features, y for the target variable SO_2 concentration, $\widehat{y}$ for the the predicted value and w for the coefficients as reported in Eq.(4).

$$\widehat{y}(w, x) = w_0 + w_i x_i + \cdots + w_n x_n \tag{4}$$

Variable	Unit	Description
Ship Information		
Lat	degrees	Latitude of the ship
Lon	degrees	Longitude of the ship
Ship Speed	kNot	Speed of the ship
IFO	% w/w	Composition of intermediate fuel oil in terms of sulphur
MGO	% w/w	Composition of marine gas oil in terms of sulphur
SFOC.ME.1	g/kWh	Specific fuel oil consumption of main engine 1
SFOC.ME.2	g/kWh	Specific fuel oil consumption of main engine 2
Venturi/Scrubber Data		
SO_2	ppm	Concentration of sulphur dioxide in exhaust gases
SC.200.SW.Inlet.Flow	m^3/h	Flow rate of seawater entering the scrubber
SC.200.SW.Inlet.Press.	bar	Pressure of seawater entering the scrubber
Venturi.1.Inlet.Pressure	mmWC	Pressure of gas entering the first Venturi
Venturi.2.Inlet.Pressure	mmWC	Pressure of gas entering the second Venturi
Scrubber.Vent.1.Diff. Press.	mmWC	Pressure drops across the first Venturi and the scrubber
Scrubber.Vent.2.Diff. Press.	mmWC	Pressure drops across the second Venturi and the scrubber
Venturi.1.Inlet.Temperature	°C	Temperature of gas entering the first Venturi
Venturi.2.Inlet.Temperature	°C	Temperature of gas entering the second Venturi
Scrubber.Outlet.Temperature	°C	Temperature of gas exiting the scrubber
Seawater parameter		
PAH Scrubbing Water	ppb	Concentration of PAH in the scrubbing water
pH Scrubbing Water	pH	pH value of the scrubbing water
Turbidity Scrubbing Water	FNU	Turbidity of the scrubbing water
Temperature Scrubbing Water	°C	Temperature of the scrubbing water
Salinity	g/L	Salinity of the seawater

TABLE I: Dataset features

Feature	Mean	StDev	Min	P25%	P50%	P75%	Max
SC.200.SW.Inlet.Flow	853	71.6	385	802	819	934	967
SC.200.SW.Inlet.Press.	3.02	0.45	1.20	2.70	2.70	3.50	4.10
So2	13.12	5.82	5.00	8.00	12.0	18.0	161
Venturi.1.Inlet.Pressure	44.1	17.6	-29.00	31.00	51.0	58.0	83.0
Venturi.2.Inlet.Pressure	31.2	16.8	-22.0	19.0	34.0	46.0	68.0
Scrubber.Vent..1.Diff..Press.	59.0	13.4	-1.00	50.0	63.0	69.0	89.0
Scrubber.Vent..2.Diff..Press.	46.1	13.5	-7.0	36.00	49.0	57.0	75.0
Venturi.1.Inlet.Temperature	258.6	19.2	32.0	255.0	260.0	267.0	293.0
Venturi.2.Inlet.Temperature	263.1	18.0	33.0	258.0	264.0	269.0	293.0
Scrubber.Outlet.Temperature	13.0	4.67	1.00	10.0	14.0	17.0	27.0
PAH.Scrubbing.Water	4.25	2.96	-3.00	1.00	5.00	6.00	16.0
pH.Scrubbing.Water	8.31	0.25	6.70	8.20	8.30	8.50	9.40
Turbidity.Scrubbing.Water	0.61	4.42	0.00	0.00	0.00	0.10	176
Temperature.Scrubbing.Water	16.7	4.29	4.00	14.0	18.0	20.0	27.0
Lat	56.2	3.04	48.0	54.8	56.5	58.8	60.5
Lon	14.3	8.59	-5.77	8.84	15.4	21.1	29.7
Salinity	17.4	11.8	4.00	8.00	8.00	35.0	35.0
Ship.Speed	8.94	6.71	0.00	6.17	9.13	12.01	164.03
IFO	2.39	0.24	1.62	2.33	2.42	2.51	2.64
MGO	0.07	0.01	0.04	0.07	0.07	0.07	0.08
SFOC.ME.1	182	8.13	175	177	179	186	225
SFOC.ME.2	180	6.74	175	176	177	183	225

TABLE II: Statistical outlook of attributes of the dataset

We also use the notation X to describe the matrix of input features and $w = (w_1, \ldots, w_n)$ for the vector of coefficients. The solution of the following problem provides us with the values of the coefficients w of the linear model, using the aforementioned methods.

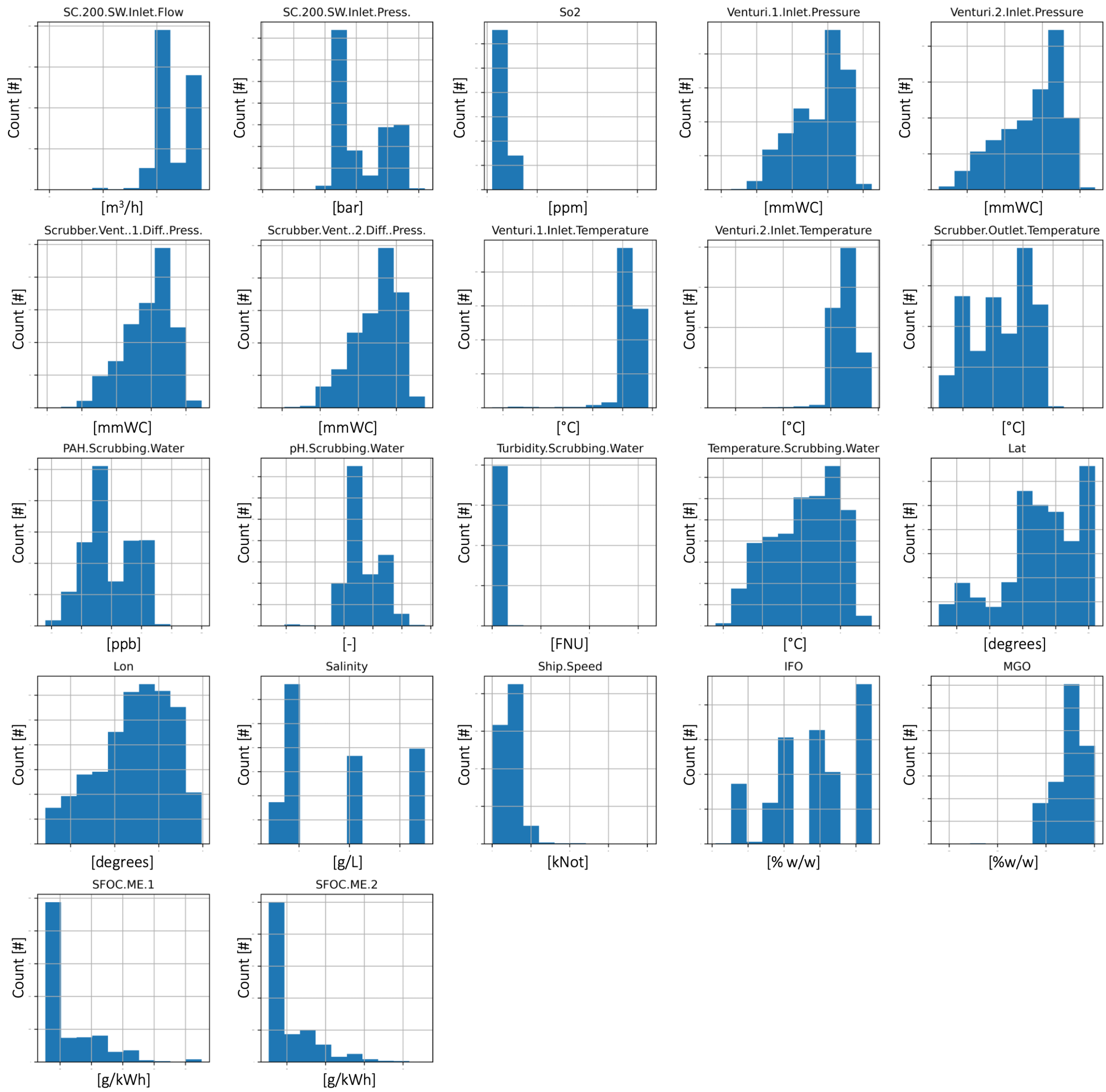

Fig. 3: Outlook of distribution attributes

$$OLS : \min_{w} ||Xw - y||_2^2 \tag{5}$$

$$Ridge : \min_{w} ||Xw - y||_2^2 + \alpha||w||_2^2 \tag{6}$$

$$Lasso : \min_{w} \frac{1}{2n_{samples}} ||Xw - y||_2^2 + \alpha||w||_1 \tag{7}$$

Ridge (Eq. (6)) and Lasso regressions (Eq. (7)) address some of the problems of OLS by imposing a penalty on the size of the coefficients. The $\alpha \geq 0$ parameter controls the entity of penalty.

$$SGD : \min_{w} \frac{1}{n_{samples}} \sum_{i=1}^{n} L(y_i, f(x_i)) + \alpha R(w) \tag{8}$$

Indeed, considering the SGD Regressor in Eq. (8), where L is a loss function that measures model (mis)fit and R is a penalty; also in this case $\alpha \geq 0$ is a non-negative hyperparameter that controls the regularization strength. Interested readers may find more details on the most suitable L and R and about the overmentioned ML algorithms in [3], [9], [4].

Concluding, the kNN-R and the SVM-R were selected because they are non linear algorithms which use a different approach on a different basis with respect to the other linear ones: consequently, it is not possible to define an analogous, yet consistent, formal expression.

V. RESULTS AND DISCUSSION

We have used Python 3.10 and the Sci-Kit Learn library to implement all previous described regression models. Before showing regression results per each

Model	R^2	MSE	MAE
OLS	0.5753	14.56	2.817
RIDGE	0.5753	14.56	2.815
LASSO	0.5502	15.42	2.857
SGD	0.5744	14.59	2.824
kNN-R	0.8360	5.620	1.368
SVM-R	0.8346	5.669	1.374

TABLE III: Evaluation Metrics

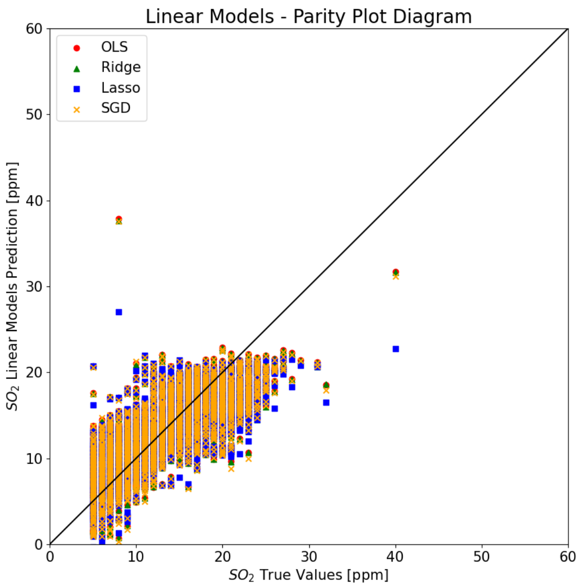

Fig. 4: Linear Models Results

model, some considerations are needed about the tuning of models parameters. The optimal values of α for all linear models and also the optimal functions for the loss function $L(y_i, f(x_i)$, as well as the penalty function $R(w)$ for the SGD model, have been selected by using a k-Fold Cross Validation approach. Following the same approach, for the kNN-R model 7 has been found to be the optimal value for k and the Radial Basis Fucntion (RBF) has been found to be the best kernel for the SVM-R.

All the regression models were applied using the variable $[SO_2]_{OUT} = y$ as target variable and the other variables described in the Subsection IV-A as features matrix X. The regressions result are reported in Fig. 4 and 5, respectively for the linear and no-linear models. The evaluation metrics for each model are shown in Table III.

VI. CONCLUSIONS AND FUTURE WORKS

In this work, we explored the possibility of modeling the behavior of a naval scrubber for reducing SO_2 emissions using machine learning tools.

We utilized a real dataset that gathered information on a ship during an entire route, including data on the scrubber and other relevant characteristics. The

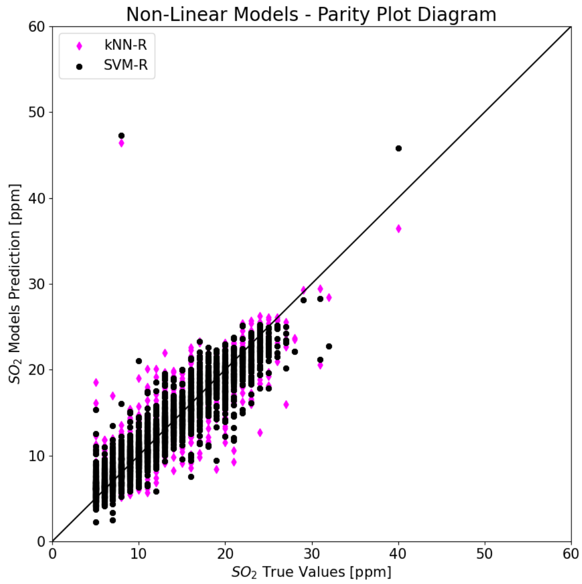

Fig. 5: kNN-R and SVM-R Results

results were encouraging, particularly for support vector regression (SVR) and k-nearest neighbors (KNN) algorithms, while not optimal for linear regression algorithms, as expected due to the non-linearity of the problem itself.

The importance of testing the use of machine learning algorithms, not necessarily deep learning, stems from the possibility of future implementation of such a system on board a ship, to set the scrubber parameters in real-time for optimal performance. Under such operational conditions, the use of algorithms with low computational impact could make a significant difference.

Finally, as the next step in optimizing the proposed algorithms, we will undertake a careful feature engineering phase, involving a detailed analysis of the physical parameters involved, and the possible need to add sensors to the onboard instruments to obtain additional data that can enhance the model.

VII. ACKNOWLEDGMENTS

This work has been funded by the research activity realized within the project PON Ricerca e Innovazione 2014-2020, FSE REACT-EU Fund, Azione IV.4 “Dottorati e contratti di ricerca su tematiche dell’innovazione”. The authors would like to thank Grimaldi Group for their support in the study.

REFERENCES

[1] M.-j. S. Bashir Ahmed Danzomo and R. M. Khan. Artificial neural network model for wet scrubber performance. *International Journal of scientific and Engineering Research*, 3:8–10, 2012.

[2] I. Bukovský and M. Kolovratník. A neural network model for predicting nox at the mělník 1. *Acta Polytechnica*, 52, 2012.

[3] A. Burkov. The hundred-page machine learning book-andriy burkov. *Expert Systems*, 5:132–150, 2019.

[4] A. Geron. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems.* O'Reilly Media, Inc., 1st edition, 2017.

[5] I. M. O. (IMO). Sulphur oxides (sox) and particulate matter (pm) – regulation 14. *https://www.imo.org/en/OurWork/Environment.*

[6] C. Li, D. Zhao, S. Mu, W. Zhang, N. Shi, and L. Li. Fault diagnosis for distillation process based on cnn–dae. *Chinese Journal of Chemical Engineering*, 27:598–604, 3 2019.

[7] B. Mamandipoor, M. Majd, S. Sheikhalishahi, C. Modena, and V. Osmani. Monitoring and detecting faults in wastewater treatment plants using deep learning. *Environmental Monitoring and Assessment*, 192, 2 2020.

[8] F. N. Osuolale and J. Zhang. Thermodynamic optimization of atmospheric distillation unit. *Computers and Chemical Engineering*, 103:201–209, 2017.

[9] S. Russell and P. Norvig. *Artificial Intelligence A Modern Approach (4th Edition)*. 2021.

[10] V. V. Santana, M. S. Gama, J. M. Loureiro, A. E. Rodrigues, A. M. Ribeiro, F. W. Tavares, A. G. Barreto, and I. B. Nogueira. A first approach towards adsorption-oriented physics-informed neural networks: Monoclonal antibody adsorption performance on an ion-exchange column as a case study. *ChemEngineering*, 6, 4 2022.

[11] M. Taheri and A. Mohebbi. Design of artificial neural networks using a genetic algorithm to predict collection efficiency in venturi scrubbers. *Journal of Hazardous Materials*, 157:122–129, 2008.

[12] A. G. Urbina and A. Aoyama. Pipeline risk assessment using artificial intelligence: A case from the colombian oil network. *Process Safety Progress*, 37:110–116, 3 2018.

[13] M. F. A. Yousif Al-Dunainawi. Pso-pd fuzzy control of distillation column. pages 554–558. Institute of Electrical and Electronics Engineers Inc., 12 2015.

[14] Z. Zhang, Z. Wu, D. Rincon, and P. D. Christofides. Real-time optimization and control of nonlinear processes using machine learning. *Mathematics*, 7, 10 2019.

AUTHOR BIOGRAPHIES

LELIO CAMPANILE is a tenured teacher of Computer Science and a research associate at Dipartimento di Matematica e Fisica, Università degli Studi della Campania "L. Vanvitelli", Caserta, Italy, where he has been a technician, a network administrator and an expert for many local and regional projects, and is a member of the Data and Computer Science group. He holds a M. Sc. Degree in Computer Science and a Ph.D.. His email is lelio.campanile@unicampania.it.

LUIGI PIERO DI BONITO is a Chemical Engineer and PhD Candidate in Mathematics, Physics and Application for the Engineering at Dipartimento di Matematica e Fisica, Università degli Studi della Campania "L. Vanvitelli", Caserta, Italy. His email is luigipiero.dibonito@unicampania.it.

FRANCESCO DI NATALE is a Chemical Engineer and a Professor of Unit Operations and Sustainable Process Design at the University of Naples Federico II. He has more than 20 years of experience in the development of sustainable technologies for pollution control and recovery of added-value chemicals (precious metals, combustion nanoparticles...) from waste streams. He is specialized in the design of multiphase-flows (absorbers, adsorbers, fabric filters, fluidized bed units) and electrical-driven (chemi-electrohydrodynamics and cold plasma technologies) processes. His typical field of application is in the treatment of gas and water streams produced by industrial and civil processes and by internal combustion engines, as well in the filtration of bioaerosols. His email is francesco.dinatale@unina.it.

MAURO IACONO is an Associate Professor in Computing Systems at Dipartimento di Matematica e Fisica, Università degli Studi della Campania "L. Vanvitelli", Caserta, Italy, where he leads the Computer Science section of the Data and Computer Science research group. He received the Ph.D. in Electrical Engineering from Seconda Università degli Studi di Napoli. His research activity is mainly centered on the field of performance modeling of complex computer-based systems, with special attention for multiformalism modeling techniques. His email is mauro.iacono@unicampania.it.