25th International Conference on Harbor, Maritime and Multimodal Logistic Modeling & Simulation
20th International Multidisciplinary Modeling & Simulation Multiconference

2724-0339 © 2023 The Authors.
doi: 10.46354/i3m.2023.hms.013

Virtual Gas Analyzer in the Maritime Sector

Agostino Bruzzone [1,2,*], Alberto De Paoli,[1,2,] and Mehrnoosh Mashayekhi[1]

[1]Simulation Team, Via Cadorna 2, Savona, 17100, Italy
[2]University of Genoa, Via Opera Pia 15, Genoa, 16145, Italy

[*]Corresponding author. Email address: agostino.bruzzone@unige.it

Abstract

This research paper focuses on the utilization of Virtual Gas Analyzers (VGAs) within the maritime industry to observe, analyze and predict gas compositions in marine holds. The maritime sector encounters numerous challenges concerning the efficient transportation, particularly considering increasingly stringent regulations. By enabling real-time predicting of water flow, VGAs offer a solution to prevent environmental pollution. The objective was to collect relevant data from IoT sensorts and present it in a comprehensive Machine Learning (ML) model to operators responsible for on board vessels. It works on modeling for exhaust gas cleaning systems (EGCS) and assessing the impact of washwater discharges. The study reveals research on VGAs prediction models which regarding their usage of ML models such as Linear Regression, Random Forest, LightGBM and Multi-Layer Perceptron. The prediction model that VGAs hold great promise in enhancing efficiency and accuracy are capable of effective prediction. Added to the model, the criteria need to consider is how the systems will behave in different situations which, handled by Polynomial chaos expansion (PCE). The study concludes with recommendations for policies and practices, underscoring the significance of ongoing innovation and investment in this field.

Keywords: Machine Learning; pollution; exhaust gas cleaning systems

1. Introduction

Environmental pollution caused by maritime transportation has become a significant concern in recent years due to the release of polluting emissions, including various gases and particulate matter (PM), leading to the depletion of the ozone layer. To address this issue, the International Maritime Organization (IMO) implemented a new amendment on January 1, 2020, imposing restrictions on the sulfur content in the fuel oil used by ships operating outside designated Emissions Control Areas (ECAs). The previous sulfur content limit of 3.5% m/m was reduced to 0.50% m/m to substantially reduce environmental pollution caused by maritime transportation.

In ECAs, where emissions control is crucial, the maximum allowable sulfur content in the fuel was set even lower at 0.10% m/m. These strict measures reflect the maritime industry's commitment to mitigating environmental pollution and minimizing the harmful impact on the ozone layer. The industry is facing the challenge of finding alternative methods to reduce Sulfur Oxide (SOx) emissions to comply with the new regulations. Potential solutions include the use of cleaner fuels, advanced exhaust gas cleaning systems, and improved engine designs.

1.1. The usage of EGCS system

The demand for exhaust gas cleaning systems (EGCS) for ships has increased following the global cap on sulfur emissions in 2020. The EGCS operates based on a scrubbing principle, with components such as an Exhaust Gas Cleaning (EGC) unit, a designated scrubber, an exhaust gas supply and discharge system, and a seawater supply and washwater discharge system. The EGC unit pretreats the exhaust gas by mixing it with seawater to remove SOx and particulate matter emissions. Continuous Emissions Monitoring Systems (CEMS) and scrubber water monitoring sensors are used to monitor CO2 and SO2 levels in the exhaust gas and detect pollutants in the washwater discharge pipe. Two approaches to reduce emissions are discussed in the case study. The first involves using low-sulfur oil or alternative fuels like Liquid Natural Gas (LNG). The second approach is installing an EGCS, which is a cost-effective method for ship owners to reduce SO2 emissions due to the reduction in additional fuel expenses.

1.2. State of the art: VGA

A Virtual Gas Analyzer (VGA) is introduced as a software tool used to analyze and monitor the composition of gases within a particular environment, such as the cargo hold of a ship. VGAs collect data from sensors and use algorithms to calculate gas concentrations, ensuring safe and efficient transportation of cargo and identifying potential hazards during voyages. Overall, by implementing these measures and technologies, the maritime industry can contribute to the preservation of the environment and work towards a more sustainable future for maritime transportation.

While the initial costs of implementing these technologies may be significant, the long-term benefits outweigh the expenses. The reduction in emissions and compliance with regulations can lead to improved air quality, reduced health risks for crew members and individuals living near ports, and a positive image for the maritime industry.

In addition, the use of Virtual Gas Analyzers (VGAs) presents an innovative approach to monitoring and ensuring the safety of cargo transportation. By accurately measuring gas concentrations within the cargo hold, ship operators can identify potential hazards and take necessary precautions to prevent accidents or incidents. This technology not only enhances the safety of maritime transportation but also improves the efficiency of cargo handling and delivery, leading to cost savings and increased customer satisfaction.

2. Related work

Flow analysis techniques have been proven to be effective tools for analyzing marine environmental parameters in various samples (Christian, 2003). The continuous discharge of carbon dioxide into the atmosphere by human activities can alter the carbonate equilibrium in oceans, highlighting the importance of accurately characterizing parameters like pH, pCO2, TA, and DIC for understanding ocean chemistry and acidification impacts (Ma, 2016). Flow is also significant in water circulation, carrying materials and impacting organism presence (Dahuri, 2003). Flow patterns in oceans are crucial for determining cruise ship directions (Giribone, 1995).

Marine realm is an extremely complex system and Simulation offers a comprehensive method for researchers, policymakers, and environmentalists to make informed decisions based on accurate predictions and scenario testing (Bruzzone et al., 2021). Advanced simulation tools can predict the spread of pollutants in the atmosphere based on various factors like wind speed, direction, temperature, and the physical and chemical properties of the pollutant itself (Bruzzone et al., 2022).

Marine flow meter monitoring is vital for balanced vessel operations. This project developed an interpretable model to predict SWFlow using sensor data. While related studies use CFD modeling for exhaust gas cleaning systems (Vasilescu, 2021), or study washwater discharges (Faber, 2019), a CDF approach was developed for scrubber washwater dilution simulation in this project (Ristea, 2022).

Collected data, read from installed sensors every three minutes, was used for this project. Exploiting the time order of the data as time-series data improves prediction accuracy (Ciaburro, 2021). ML models like Random Forest (RF) and Gradient Boosting were employed (Breiman, 2001; Friedman, 2001). Random Forests proved to be more accurate with a large number of covariates (Medeiros, 2021). Concerns arise from pollutants in scrubber washwater from ship engine exhausts. Therefore, an onboard prediction system is necessary to prevent exceeding pollution thresholds.

Historical terrestrial AIS data is utilized to analyze ship movements and operational modes in this model. Ship emissions and fuel consumption are calculated based on ship type, size, operating mode, and machinery type (Jalkanen, 2009; Olesen, 2010; Pitana, 2010; Ng, 2013; Johansson, 2013).

3. Methodology

The VGA Application is a comprehensive software solution designed to assist with data analysis and prediction modeling. It includes two main categories: Data preprocessing and prediction Model. Data preprocessing involves collecting and preparing input data for prediction or training. Raw data is gathered from BigQuery data warehouse or through API requests, then undergoes validation and feature calculation. The processed data is stored as historical data for analysis.

The Model category focuses on prediction and training. Pre-trained models can be used for accurate predictions, but in cases where accuracy is lacking due to lack of historical data or different sensor errors, the VGA Application offers a train AI model. This allows users to periodically train the model using new data, adapting it to specific scenarios and improving accuracy.

VGA Application provides Data APIs for data collection and preprocessing, and Model APIs for prediction and training. The training API is especially important for refining accuracy in scenarios where pre-trained models are not sufficient. This empowers users to make informed decisions and enhance predictive capabilities using data-driven insights.

3.1. Data analysis

The VGA Application provides two methods for data collection and analysis:

1. API Call: Users can send real-time or interval-based observation through API call. Each observations follows a specific JSON format for compatibility with the application's processing capabilities.
2. Bulk Upload: Users can upload a folder of DG files during setup, containing historical data in a format supported by the application.

The VGA Application requires input data to follow a specific JSON structure for consistent and efficient processing. JSON is a widely supported format for structured data representation, making it suitable for data transmission.

By providing data in the specified JSON structure, users can seamlessly integrate their data with the VGA Application, allowing for accurate analysis, prediction, and training based on the provided observations. As data is received through API calls, each observation is appended to the corresponding historical data files. To be considered valid and usable, each observation must contain a 'timestamp' or 'ndc_ts' field. If either of these entries is missing, the observation is discarded to maintain data integrity.

By dividing the data into daily CSV files and implementing the

retention policy, the VGA Application strikes a balance between preserving valuable historical data and managing storage resources effectively. This approach enables users to access and leverage a significant amount of historical data for model training, ensuring that the models are trained on a diverse and comprehensive dataset.

3.2. Data preprocess

In this research, regression functions are used to estimate the pH set-point in the second rack of an unspecified system. These functions are evaluated within a variable space that adheres to specific validity conditions outlined in Table 2. These conditions relate to fundamental system parameters and aim to accurately represent the typical operation of AAQS systems with a specific SO2 to CO2 ratio. By analyzing data that meets these conditions, the researchers aim to make precise conclusions about the relationship between pH and other variables in the system, leading to insights for optimizing its performance.

It's important to note that all subsequent analysis considerations refer to the dataset that fulfills the mentioned conditions. An observation is only considered valid if all its variables are valid as well, ensuring the analysis focuses on high-quality and reliable data.

Table 1. The parameters of valid threshold.

Signal	Valid range
pH rack 2	2.8-5.5
Dg Load (%)	40-90
Sea water flow	200-1200
So_2 / CO_2 (ppm/%)	0.2-4.3

All the considerations developed in Table 1 will refer to the dataset defined by applying the conditions described above. The entire observation is considered valid if all variables are valid as well. Table 1 is showing the thresholds of data validity. The stability analysis conducted in the study is crucial to ensure that the data used for analysis is reliable and free from noise or transitional effects. By excluding moments affected by noise or transitions between different speeds, the researchers aim to focus on data points that reflect the stable and representative operating conditions of the system. This approach helps to mitigate any potential confounding factors that could distort the relationship between pH and the influencing variables. It is important to note that the pH in the second rack is influenced by four main factors: Engine Power, Seawater Flow, Alkalinity of water, and the percentage of sulfur in the fuel. However, due to the availability of data, only Engine Power and Seawater Flow can be directly used for estimating the pH.

To overcome this limitation, the researchers introduce the concept of Theoretical pH, which is derived by utilizing the measured values of Engine Power and Seawater Flow, assuming average values for alkalinity and sulfur content. Through the analysis of system stability, researchers detect and eliminate noise or transitional effects, ensuring that the data used for regression analysis accurately reflects the system's stable operation. This approach enhances the accuracy of estimating the relationship between pH and influencing variables within the specified validity conditions.

3.3. Geographic coordinate sampling

By considering the geographical location of the ship, the model captures complex relationships and makes accurate predictions, optimizing ship operations and ensuring efficiency in varying marine environments. Geolocation data also offers contextual information for analysis and decision-making, providing insights into the specific marine environments where data samples were collected. The integration of geolocation data enables the model to capture the potential impact of environmental factors on Sea Water Flow, such as varying temperatures, salinity levels, or tidal patterns. It facilitates the identification of geographical patterns or trends, uncovering spatial variations and correlations that would otherwise go unnoticed. This information is valuable for maritime planning, resource allocation, and decision-making processes. Moreover, geolocation data allows for the development of location-specific models, tailored to the unique characteristics and challenges of specific geographical areas. This enhances prediction accuracy and provides targeted insights for localized decision-making. Incorporating geolocation data into the prediction model for Sea Water Flow improves prediction accuracy, enables the consideration of environmental factors, reveals spatial patterns, and trends, and supports location-specific analysis and decision-making. It enhances our understanding of Sea Water Flow dynamics and enables more effective management of ship operations in diverse marine environments.

3.4. Training the model

After evaluating a sufficient number of observations for each DG, the methodology proceeds to calculate polynomial features by generating combinations of principal variables at a degree of 3. This captures complex relationships among variables. The Lasso model is then fitted with 10-fold cross-validation to identify important features and manage multicollinearity. Cross-validation ensures model evaluation and selection.

To enhance model performance, the stable K value is incorporated, updated during data processing to reflect recent observations. This accounts for system behavior changes and enhances prediction accuracy. The obtained model is saved, and the previous version is archived for reference, enabling tracking and comparison of different versions.

The methodology involves calculating polynomial features, fitting a Lasso model with cross-validation, enriching the model with the stable K value, and saving the obtained model while archiving the previous version. These steps develop an effective model for estimating pH set-point, incorporating variable complexity and recent system behavior. Enriching with the stable K value accommodates ongoing changes. Additionally, for more reliable and complex development, tree-based models like Decision Trees, LightGBM, and Random Forest can be explored, with a focus on LightGBM.

Utilizing LightGBM, we achieved highly accurate SW Flow predictions. This model excels in capturing complex, non-linear data feature relationships, yielding valuable insights. The underlying decision tree algorithm in LightGBM assesses feature importance, aiding result interpretation and understanding SW Flow drivers.

LightGBM's noteworthy speed in training and prediction

enhances its suitability for real-time applications where timely predictions matter. Its capability to handle missing data and outliers adds to its robustness, ensuring reliable performance even with imperfect data. Need to mention for training we kept the default to XGboost.

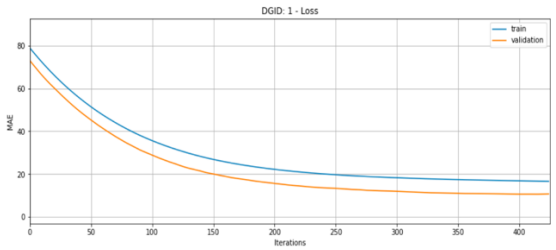

Figure 1. A loss function of train and validation

3.5. Metrices and evaluation

In statistics, the coefficient of determination (Steel and Torrie, 1960), denoted R2 or R and pronounced "R squared", is the proportion of the variation in the dependent variable that is predictable from the independent variable(s).

If $\bar{y}$ is the mean of the observed data:

$$\bar{y} = \frac{1}{n}\sum_{i=0}^{n} y_i \tag{1}$$

The most general definition of the coefficient of determination is, R-Squared (R^2) is a statistical metric in regression models that indicates the portion of variance in the dependent variable explained by the independent variable. Essentially, it gauges how closely the data align with the regression model, indicating the model's goodness of fit:

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \tag{2}$$

In the best case, the modeled values exactly match the observed values, as formula 2 demonstrated SS_{res} is sum of squares of residuals and SS_{tot} is total sum of squares. which results in SS_{res}=0, and R^2=1. A baseline model, which always predicts $\hat{y}$, will have R^2=0. Models that have worse predictions than this baseline will have a negative R^2.

4. Results and Discussion

The exhaust gas cleaning system pollution prevention model was utilized to predict wash water by altering the discharge water's pH upon ocean release. Predicted dilution rates in the model determine discharge water pH. Analysis shows pH increases to 6.5 at the discharge point 4.0 meters away. This approach monitors pH as fluid exits the nozzle, using convection, turbulence, and diffusion to predict downstream pH based on boundary conditions. It's passive and doesn't simulate chemical reactions.

Table 2. Model score

Score	Value
MAE	20.919
Max error	369.406
R^2	0.87

The analysis projects pH by comparing titration curve and seawater with discharge wash water concentration from CFD analysis. Table 2 illustrates typical pH estimations based on seawater dilution levels. For better accuracy, valid data merged with positions is used. A new time-series is created with given intervals, omitting invalid data, and applying LightGBM model to valid data.

The usage of sliding window is to take into account such that the need for reusing the loops gets reduced and hence the program gets optimized. In this technique, we use the data as an aggregation in a specific function for example average, in this case we use each average of 10 record of the data one record to compute the result of the next step.

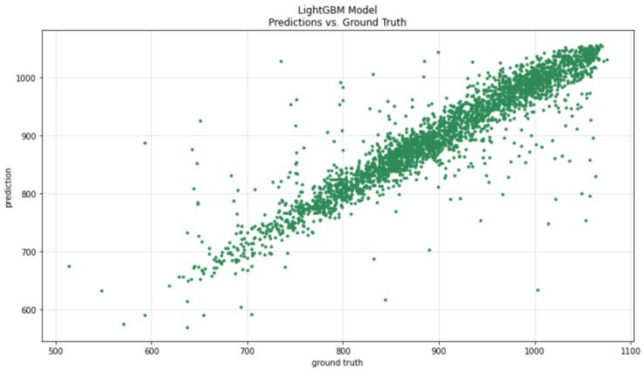

Figure 2. Prediction versus ground truth of the model

In Fig.2 We compute the loss/error between the prediction (of the model) and the ground truth. Here, the ground-truth refers to the "officially correct" label (categorical or numerical) for a given input with which you compute the prediction. So, in this case, ground-truth would be a synonym for a ground-truth label. A better way to identify bias in the regression model predictions is plotting the fitted values against the residuals $r_i = y_i - \hat{y}_i$. Also note that if you include an intercept in regression, the residuals will always sum up to zero.

Table 3. Model score comparison

Models	MAE	R^2
Linear regression	0.241	0.84
Random forest	0.226	0.846
LightGBM	0.13	0.924
MLP	0.01	0.859
PCE	0.232	0.84

In this study, we compare the performance of three different state of the art machine learning models as presented in Table 3 in prediction sector, to the matter of efficiency and accuracy LightGBM offer valuable insights into accurate predicting. Further research and experimentation are necessary to fully leverage the potential of these models and contribute to.

5. Conclusions

In conclusion, the choice between low sulphur fuel and scrubbers offers potential environmental benefits, although the increased use of scrubbers has raised concerns about the impact of

washwater discharge on marine ecosystems. This dilemma poses a trade-off between air and water quality, necessitating a comprehensive evaluation that goes beyond examining air emissions alone. Flow analysis techniques have proven valuable for marine environmental parameter analysis due to their high sample throughput, versatility, and robustness. Accurate characterization of key parameters like pH, pCO2, TA, and DIC is vital for understanding carbon dioxide emissions' impact on marine ecosystems. Flow analysis also influences water circulation, carrying material that impacts marine organism presence.

Continuous monitoring with marine flow meters ensures proper vessel operation efficiency. This project developed an interpretable model predicting future flow patterns using machine learning techniques and sequential patterns. The use of exhaust gas cleaning systems effectively reduces SOx emissions, but concerns arise from washwater pollutants. On-board prediction systems are essential to prevent excessive pollutant levels. Various application models, including AIS data, analyze ship movements, emissions, and operational modes. These models calculate emissions and fuel consumption based on ship characteristics.

To summarize, combining flow analysis techniques, machine learning models, and emission monitoring systems enhances understanding and management of environmental aspects in maritime transportation. This pursuit promotes sustainable practices and reduces environmental impact. Further research and innovation are needed to improve prediction accuracy, optimize monitoring, and devise effective strategies for emissions reduction and pollution control in the maritime sector.

References

Bruzzone, A. G., Vairo, T., Cepolina, E. M., Massei, M., De Paoli, A., Ferrari, R., ... & Pedemonte, M. (2022, October). Cooperative Use of Autonomous Systems to Monitor Toxic Industrial Materials and Face Accidents & Contamination Crises. In International Conference on Modelling and Simulation for Autonomous Systems (pp. 231-242). Cham: Springer International Publishing.

Bruzzone, A. G., Massei, M., Sinelshchikov, K., Giovannetti, A., & Gadupuri, B. K. (2021, July). Strategic engineering applied to complex systems within marine environment. In 2021 Annual Modeling and Simulation Conference (ANNSIM) (pp. 1-10). IEEE.

Ciaburro, Giuseppe, and Gino Iannace. "Machine learning-based algorithms to knowledge extraction from time series data: A review." Data 6.6 (2021): 55.

David A. Freedman (2009). Statistical Models: Theory and Practice. Cambridge University Press. p. 26. A simple regression equation has on the right-hand side an intercept and an explanatory variable with a slope coefficient. A multiple regression e right hand side, each with its own slope coefficient

Dietterich, Thomas G. "Machine learning for sequential data: A review." Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshops SSPR 2002 and SPR 2002 Windsor, Ontario, Canada, August 6–9, 2002 Proceedings. Springer Berlin Heidelberg, 2002.

Faber, J. "The impacts of EGCS washwater discharges on port water and sediment." (2019).

Hornik, K., Stinchcombe, M. and White, H. (1989) 'Multilayer feedforward networks are universal approximators',Neural Networks, 2(5), pp. 359–366. Available at: https://doi.org/10.1016/0893-6080(89)90020-8.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.

Luostarinen, M. (2019). Exhaust Gas Cleaning Scrubbers, Operation Monitoring and Maintenance.

Medeiros, M.C., Vasconcelos, G.F., Veiga, Á. and Zilberman, E., 2021. Forecasting inflation in a data-rich environment: the benefits of machine learning methods. Journal of Business & Economic Statistics, 39(1), pp.98-119.

Steel, R. G. D.; Torrie, J. H. (1960). Principles and Procedures of Statistics with Special Reference to the Biological Sciences. McGraw Hill.

Rencher, A. C., & Christensen, W. F. (2012). Chapter 10, Multivariate regression–Section 10.1, Introduction. Methods of multivariate analysis, Wiley Series in Probability and Statistics, 709(3).

Ristea, M., Popa, A., & Scurtu, I. C. (2022). Computational Fluid Dynamics Simulation Approach for Scrubber Wash Water pH Modelling. *Energies*, *15*(14), 5140.

Vasilescu, Mihail-Vlad, et al. "Research on Exhaust Gas Cleaning System (EGCS) used in shipping industry for reducing SOx emissions." E3S Web of Conferences. Vol. 286. EDP Sciences, 2021.