



Modélisation et estimation de la consommation de carburant de moteurs diesels marins

Mémoire présenté

dans le cadre du programme de maîtrise en ingénierie

en vue de l'obtention du grade de Maître ès sciences appliquées (M.Sc.A)

PAR

© VINCENT ST-PIERRE

Novembre 2025

Composition du jury :

Jean-Sébastien Deschênes, président du jury, Université du Québec à Rimouski

Maxime Berger, directeur de recherche, Université du Québec à Rimouski

Antoine Lesage-Landry, codirecteur de recherche, Polytechnique Montréal

Sara Séguin, examinatrice externe, Université du Québec à Chicoutimi

Dépôt initial le 8 septembre 2025

Dépôt final le 5 novembre 2025

UNIVERSITÉ DU QUÉBEC À RIMOUSKI
Service de la bibliothèque

Avertissement

La diffusion de ce mémoire ou de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire « *Autorisation de reproduire et de diffuser un rapport, un mémoire ou une thèse* ». En signant ce formulaire, l'auteur concède à l'Université du Québec à Rimouski une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de son travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, l'auteur autorise l'Université du Québec à Rimouski à reproduire, diffuser, prêter, distribuer ou vendre des copies de son travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de la part de l'auteur à ses droits moraux ni à ses droits de propriété intellectuelle. Sauf entente contraire, l'auteur conserve la liberté de diffuser et de commercialiser ou non ce travail dont il possède un exemplaire.

REMERCIEMENTS

Je tiens à exprimer ma reconnaissance à mon directeur de recherche, M. Maxime Berger, pour son encadrement rigoureux, ses conseils avisés et son soutien constant tout au long de la réalisation de cette recherche. J'adresse également mes remerciements à mon codirecteur, M. Antoine Lesage-Landry, dont l'expertise technique et l'engagement soutenu ont contribué de manière significative à l'avancement de ce travail. Je souhaite également souligner la collaboration du partenaire industriel OpDAQ Systems, dont l'appui technique et financier a joué un rôle déterminant dans la conduite de ce projet et le transfert des résultats. Enfin, cette recherche a été soutenue financièrement par Mitacs dans le cadre du programme Mitacs Accélération, sous la subvention IT37372, ainsi que le Fonds institutionnel de recherche (FIR) de l'UQAR, concours Automne 2024.

RÉSUMÉ

La réduction des émissions de gaz à effet de serre constitue un enjeu majeur pour l'industrie maritime. Parmi les solutions envisagées, les techniques de navigation écoénergétiques, comme l'optimisation de la vitesse et des trajets, permettent de diminuer la consommation de carburant avec un minimum d'investissement. L'efficacité de ces solutions dépend toutefois de la disponibilité de données fiables en temps réel sur la consommation de carburant des moteurs diesels. La méthode traditionnelle pour obtenir la consommation de carburant repose sur l'utilisation d'un débitmètre pour mesurer directement le débit de carburant. Bien que précise, cette solution est relativement coûteuse, invasive et peu attrayante pour les armateurs en raison des modifications nécessaires sur le circuit d'alimentation en carburant pour l'installation d'un débitmètre.

Ce mémoire par articles explore une alternative moins intrusive qui consiste à estimer la consommation de carburant uniquement à partir des mesures de couple et de vitesse à l'arbre du moteur, obtenues via un capteur peu invasif, et d'un modèle mathématique guidé par les données. Pour ce faire, deux approches sont explorées. La première approche (CHAPITRE 1) consiste à identifier et à écarter les données aberrantes autour d'une fonction de régression tridimensionnelle de type *spline*, puis à monter un tableau de correspondance en deux dimensions (2D) à l'aide d'une méthode de moyennage local. Face à certaines limitations de la première approche, notamment concernant la nécessité de l'intervention humaine pour le nettoyage des données, l'hypothèse implicite de linéarité sur un court intervalle dans l'espace des données et la faible garantie de généralisation, une seconde méthode (CHAPITRE 2) est proposée. Cette méthode repose sur un modèle de régression polynomiale du second ordre intégrant cinq (5) termes combinant le couple et la vitesse. Les coefficients de ce modèle sont ajustés par une méthode d'apprentissage robuste en distribution, qui utilise la distance de Wasserstein pour optimiser les performances du modèle dans des conditions hors distribution qui ne seraient pas capturées par les données recueillies, ce qui maximise sa généralisation à d'autres navires.

Des tests sont réalisés sur une flotte de sept (7) remorqueurs similaires. Au total, ces navires ont généré environ 80 millions de points de données représentant un total de plus de 22 000 heures d'opération commerciale sur des remorqueurs en activité dans les eaux nord-américaines durant les cinq (5) dernières années. Le modèle issu de la seconde approche est entraîné sur un (1) navire où les données de couple, de vitesse et de débit réel de carburant sont utilisées puis testé sur les six (6) autres où les données de couple et de vitesse sont utilisées pour estimer le débit. En évaluant l'erreur quadratique moyenne (*Root Mean Square Error*, RMSE) sur l'estimation du débit de carburant sur toute la durée de plusieurs mois des

séries temporelles utilisées pour les tests, on obtient des résultats variants entre 3 % et 10 %, avec une majorité des cas situés autour de 5 %. Ces résultats démontrent la validité et la robustesse de l'approche proposée pour un déploiement en pratique sur une flotte de navires similaires. Un tel déploiement se fera en trois (3) étapes, soit 1) collecter des données opératoires d'un navire type équipé d'un capteur de couple et de vitesse ainsi que d'un débitmètre, 2) entraîner le modèle à l'aide de l'approche proposée, et 3) déployer le modèle sur le reste de la flotte de navires équipés seulement d'un capteur de couple et de vitesse.

L'approche proposée constitue une avenue réaliste, économique et peu invasive pour l'estimation en temps réel de la consommation de carburant des navires. Elle contribue à réduire les risques et les coûts associés à l'installation de débitmètre sur une flotte de navires entière ce qui favorisera la mise en œuvre de pratiques de navigation écoénergétiques par les armateurs.

Mots clés : consommation de carburant des navires, estimation en temps réel, régression robuste en distribution, modèle basé sur les données, couple et vitesse, débitmètre, distance de Wasserstein.

ABSTRACT

The reduction of greenhouse gas emissions is a major challenge for the maritime industry. Among the solutions considered, eco-efficient navigation techniques, such as speed and route planning optimization, make it possible to reduce fuel consumption with minimal investment. The effectiveness of these solutions, however, depends on the availability of reliable real-time data on the fuel consumption of diesel engines.

The traditional method relies on the use of a flowmeter to directly measure fuel flow. Although accurate, this solution is relatively costly, invasive, and unattractive to shipowners due to the modifications required to the fuel supply circuit required to install a flowmeter.

This article-based master's thesis explores a less intrusive alternative that consists in estimating fuel consumption solely from torque and shaft speed measurements obtained via a minimally invasive sensor and a data-driven mathematical model. To this end, two approaches are investigated. The first approach (CHAPITRE 1) consists in identifying and removing outliers around a three-dimensional spline regression function and then constructing a two-dimensional lookup table using a local averaging method. Faced with certain limitations of the first approach, notably the need for human intervention in data cleaning, the implicit assumption of linearity over a short interval, and the weak guarantee of generalization over a larger sample, a second method (CHAPITRE 2) is proposed. This method relies on a second-order polynomial regression model incorporating five (5) terms combining torque and speed. The coefficients of this model are fitted using a distributionally robust learning method, which employs the Wasserstein distance to optimize model performance under out-of-distribution conditions which may not be covered by the available data, thereby maximizing its generalization to other vessels.

Tests were carried out on a fleet of seven (7) similar tugboats. In total, these vessels generated approximately 80 million data points, representing more than 22,000 hours of commercial operation on tugboats active in North American waters over the past five (5) years. The model derived from the second approach is trained on one (1) vessel, where torque, speed, and actual fuel flow data are used, then tested on the other six (6), where torque and speed data are used to estimate fuel flow. By evaluating the Root Mean Square Error (RMSE) of the fuel flow estimation on out of sample data over the entire duration of several months of time series used for testing, results range from 3% to 10%, with most cases around 5%. These results demonstrate the validity and robustness of the proposed approach for practical deployment on a fleet of similar vessels. Such deployment would proceed in three (3) steps: 1) collecting operational data from a reference vessel equipped with a torque and speed sensor as well as a flowmeter, 2) training the model using the proposed approach, and

3) deploying the model across the rest of the fleet of vessels equipped only with a torque and speed sensor.

The proposed approach constitutes a realistic, economical, and minimally invasive avenue for real-time estimation of fuel consumption. It helps reduce the risks and costs associated with installing flowmeters on an entire fleet of vessels, which will foster the implementation of eco-efficient navigation practices by shipowners.

Keywords: ship's fuel consumption, real-time estimation, distributionally robust regression, data-driven model, vessels, torque and speed, flowmeter, Wasserstein distance.

TABLE DES MATIÈRES

REMERCIEMENTS.....	vii
RÉSUMÉ	ix
ABSTRACT.....	xi
TABLE DES MATIÈRES	xiii
LISTE DES FIGURES	xv
LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES.....	xvii
INTRODUCTION GÉNÉRALE	1
MISE EN CONTEXTE	1
PROBLÉMATIQUE.....	1
OBJECTIFS	4
MÉTHODOLOGIE	5
CONTRIBUTIONS	10
STRUCTURE DU MÉMOIRE	11
CHAPITRE 1 Vers une navigation écoénergétique : une approche basée sur les données pour l'estimation en temps réel de la consommation de carburant des moteurs de navires sans débitmètre	13
RÉSUMÉ EN FRANÇAIS DU PREMIER ARTICLE	13
I. INTRODUCTION	16
II. DATASET	17
III. METHODOLOGY	19
IV. RESULTS AND ANALYSIS.....	22
V. CONCLUSION	24
REFERENCES	25
CHAPITRE 2 Estimation de la consommation de carburant sans débitmètre pour l'estimation en temps réel de la consommation des moteurs diesel marins	27

RÉSUMÉ EN FRANÇAIS DU DEUXIÈME ARTICLE	27
I. INTRODUCTION	30
II. FLOWMETERLESS FUEL ESTIMATION MODEL	32
III. CASE STUDY	37
IV. CONCLUSION.....	40
REFERENCES.....	41
CONCLUSION GÉNÉRALE	45
RÉFÉRENCES BIBLIOGRAPHIQUES	47

LISTE DES FIGURES

Figure 1. Représentation simplifiée de l'évolution du système <i>OpHMI</i>	3
Figure 2. Analyse de la consommation en carburant d'un moteur diesel typique de navire.....	4

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

CSV	<i>Comma-Separated Values</i>
OMI	Organisation maritime internationale
<i>OpDAQ</i>	Nom de l'entreprise partenaire (<i>OpDAQ Systèmes Inc.</i>)
<i>OpGhost</i>	Module virtuel d'estimation de carburant d' <i>OpDAQ</i>
<i>OpHMI</i>	Interface opérateur du système d'acquisition d' <i>OpDAQ</i>
<i>OpTS</i>	Capteur de couple d' <i>OpDAQ</i>
Optuna	<i>Framework</i> Python pour l'optimisation d'hyperparamètres
PCHIP	<i>Piecewise Cubic Hermite Interpolating Polynomial</i>
POSTGRESQL	Système de gestion de base de données relationnelle <i>open source</i>
RMSE	<i>Root Mean Square Error</i>
R^2	Carré du coefficient de corrélation de Pearson
scikit-learn	Bibliothèque Python pour l'apprentissage automatique
Sodes	Société de Développement Économique du Saint-Laurent
SQL	<i>Structured Query Language</i>

INTRODUCTION GÉNÉRALE

MISE EN CONTEXTE

Le Canada s'est engagé à réduire ses émissions de gaz à effet de serre de 40 % sous les niveaux de 2005 et à atteindre la carboneutralité d'ici 2050 [1]. Cet engagement nécessitera une transition énergétique majeure qui comprend l'industrie maritime canadienne qui s'engage de plus en plus dans des programmes environnementaux comme celui de l'Alliance verte [2], [3]. Plus particulièrement, au Québec, la Société de développement économique du Saint-Laurent (Sodes) a récemment mandaté une firme externe pour réaliser un plan d'actions pour la réduction des émissions du secteur maritime québécois [3]. Ce plan a pour objectif principal d'harmoniser les efforts provinciaux avec les cibles de décarbonation définies par l'Organisation maritime internationale (OMI) [4].

Les solutions de décarbonation disponibles sont multiples. Cependant, elles diffèrent grandement en termes de complexité d'implémentation, de coûts et de bénéfices sur la réduction des émissions [5]. Parmi ces solutions, l'utilisation de techniques de navigation écoénergétique comme l'optimisation de la vitesse des navires [6], [7] et l'optimisation des voyages en mer [8], permet aux armateurs de réduire de façon significative leur consommation en carburant fossile à court terme tout en minimisant les investissements initiaux et les changements à effectuer sur leurs navires.

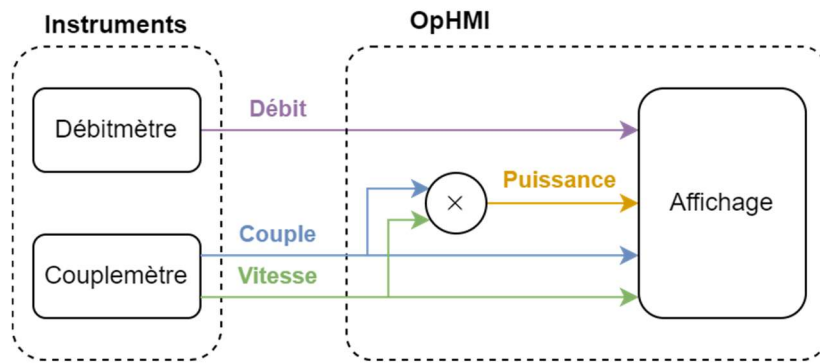
PROBLÉMATIQUE

L'évaluation de l'efficacité des techniques de navigation écoénergétiques est cependant ralentie par le manque de données quant à la consommation réelle de carburant par les moteurs à combustion interne dans différentes conditions d'opération. Les armateurs

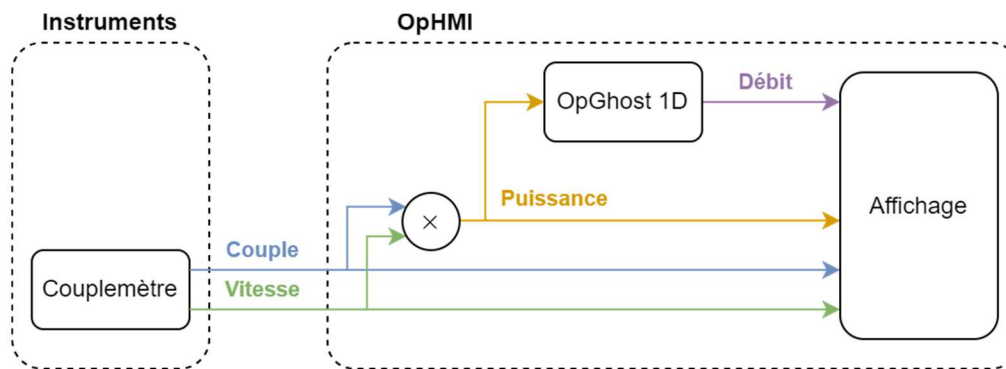
souhaitent donc instrumenter leurs navires afin d'évaluer l'efficacité des mesures opérationnelles mises en place. Cette volonté est cependant freinée par la complexité et les coûts élevés des systèmes actuels de surveillance des émissions qui sont très souvent invasifs et demandent des modifications aux systèmes de carburant et d'échappement des navires.

Dans cette foulée, le partenaire du projet, l'entreprise *OpDAQ Systèmes Inc.*, a mis au point un système d'acquisition de données, appelé *OpHMI* (Figure 1(a)), qui permet aux armateurs de mesurer en temps réel leur consommation de carburant à l'aide d'un débitmètre [9] et de déduire l'efficacité des moteurs à l'aide d'un capteur de couple et de vitesse de rotation installé sur l'arbre de sortie des moteurs [11]. Toutefois, même si la mesure directe du carburant à l'aide d'un débitmètre est une solution très commune dans l'industrie et la littérature, celle-ci nécessite des modifications importantes au niveau de la conduite d'entrée en carburant des moteurs, ce qui tend à réduire sa fiabilité et freine son adoption par les armateurs.

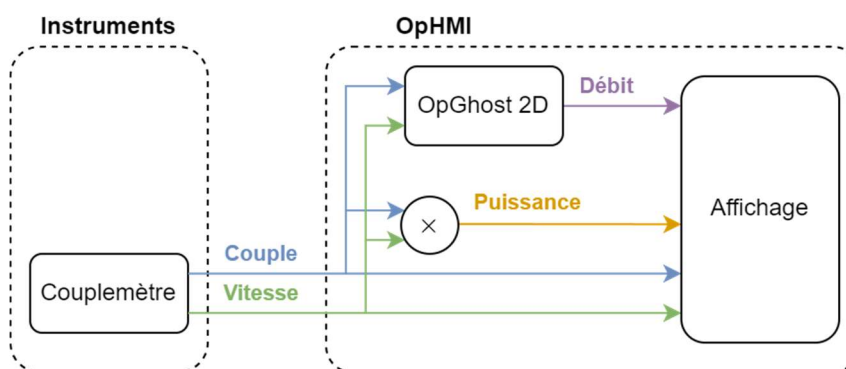
Pour résoudre ce problème, *OpDAQ* a récemment intégré un module de calcul appelé *OpGhost* dans son système d'acquisition *OpHMI*, qui permet d'estimer la consommation de carburant des navires en utilisant uniquement les mesures du couple et de la vitesse à l'arbre des moteurs [13]. La version actuelle du module *OpGhost* consiste en un débitmètre virtuel qui compare la mesure de la puissance mécanique (déduite à l'aide de la mesure du couple et de la vitesse à l'arbre) à un modèle virtuel de moteur à combustion interne pour estimer sa consommation de carburant (Figure 1(b)). Le modèle interne est basé uniquement sur la courbe de consommation fournie par le manufacturier du moteur et est ajusté avec un facteur de correction à chacun des navires à l'aide de différents essais réalisés en mer. Toutefois, malgré son excellente précision pour évaluer la consommation de carburant *annuelle* des navires, les résultats d'essais en mer ont démontré que l'algorithme n'est pas suffisamment précis pour évaluer sa consommation *en temps réel* et dans certaines conditions d'opération. Ce système demeure ainsi limité pour permettre aux armateurs d'estimer avec suffisamment de précision l'efficacité des mesures de navigation écoénergétique mises en place.



(a) Système *OpHMI* avec débitmètre et couplemètre



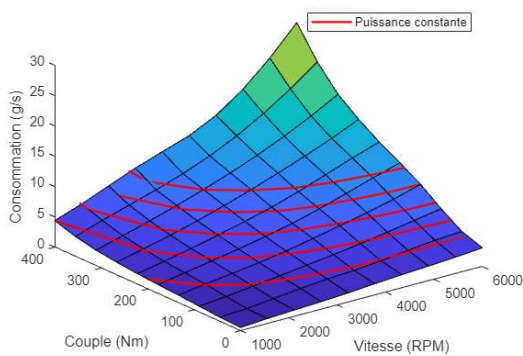
(b) Système actuel *OpHMI* avec *OpGhost* 1D (sans débitmètre)



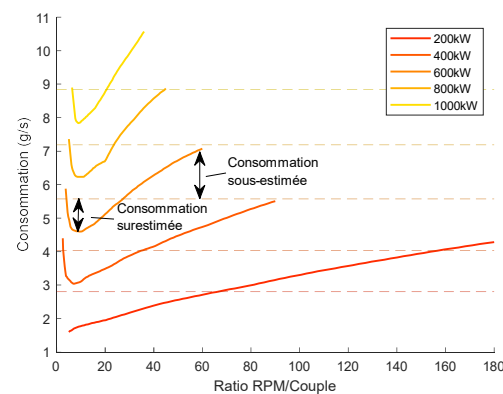
(c) Système *OpHMI* proposé avec la nouvelle génération *OpGhost* 2D (sans débitmètre)

Figure 1. Représentation simplifiée de l'évolution du système *OpHMI*

Une analyse plus fine des performances des moteurs a permis de déterminer que l'imprécision de cette méthode vient du fait qu'elle ne prend pas en compte l'effet individuel du couple et de la vitesse sur l'efficacité du moteur. Ce phénomène est expliqué à l'aide de la Figure 2 qui montre les relations entre la consommation en carburant, le couple, la vitesse et la puissance de sortie d'un moteur diesel typique pour cette étude. Sur la Figure 2(a), on note que la consommation en carburant est une fonction non linéaire du couple et de la vitesse. Pour une puissance de sortie constante (ligne **rouge**), on remarque qu'il existe plus d'une solution pour la consommation en carburant. L'algorithme *OpGhost* actuel renvoie alors la valeur moyenne pour simplifier le problème. Sur la Figure 2(b), en comparant la valeur moyenne (en pointillée [--]) avec la consommation réelle (ligne pleine [--]) on remarque toutefois que l'utilisation de la valeur moyenne peut surestimer ou sous-estimer de manière significative la consommation en carburant pour une même puissance de sortie.



(a) Consommation de carburant en fonction du couple et de la vitesse de rotation



(b) Consommation de carburant en fonction du ratio vitesse sur le couple à puissance constante

Figure 2. Analyse de la consommation en carburant d'un moteur diesel typique de navire

OBJECTIFS

L'objectif principal de cette recherche est de **développer un modèle d'estimation de consommation de carburant** basé sur les données qui utilise uniquement les mesures de

couple et de vitesse de rotation à l'arbre des moteurs de navires pour estimer leur débit de carburant (Figure 1(c)). L'objectif principal est divisé en trois (3) objectifs spécifiques :

O1. Décoder et analyser les données fournies par le partenaire *OpDAQ* ;

O2. Développer un premier modèle d'estimation de carburant pour évaluer la faisabilité d'une estimation basée entièrement sur les données ;

O3. Développer un second modèle offrant une grande généralisabilité à différents navires d'une même flotte et une robustesse accrue face aux données aberrantes et aux variations entre les navires.

MÉTHODOLOGIE

La démarche de réalisation du projet se déroule en trois (3) étapes principales qui sont associées chacune à un objectif spécifique du mémoire :

Étape 1 – Prise en main et décodage des données fournies par *OpDAQ* (O1)

La première étape du projet consiste à recevoir, décoder et s'appropriier les données collectées par *OpDAQ*. Le segment de données utilisées provient de sept (7) remorqueurs similaires qui ont généré environ 80 millions de points de données représentant un total de plus de 22 000 heures d'opération commerciale dans les eaux nord-américaines durant les cinq (5) dernières années. Un script *Python* est développé pour décoder le format binaire propriétaire reçus et le sauvegarder sous la forme d'une base de données SQL pour le premier article (CHAPITRE 1) et en format CSV pour le deuxième article (CHAPITRE 2). À cette étape, un premier nettoyage des données est effectué pour retirer les points qui ne contiennent pas d'information utile tels que les points enregistrés lorsque le navire est inactif ou les valeurs négatives de débit qui sont le résultat d'une erreur de communication avec les capteurs de débit. Ce processus génère un ensemble de données brutes prêt à l'emploi. Cette étape est effectuée à la fois sur les ensembles de données d'entraînement et ceux de validation.

Étape 2 – Élaboration d'un premier modèle et analyse de la faisabilité d'une approche basée sur les données (O2)

À la suite de l'étape 1, le jeu de données est mis à profit pour développer une première version d'un modèle d'estimation de consommation de carburant. Ce modèle est développé en suivant trois (3) sous étapes principales [9].

2.1 – Détection et suppression des données aberrantes : Cette étape vise à supprimer les valeurs aberrantes, aussi dites extrêmes, qui affectent négativement le processus d'apprentissage. Un espace euclidien en trois dimensions (3D) est premièrement construit avec les mesures de débit, de couple et de vitesse issues de l'ensemble de données d'apprentissage. Une sélection de *splines* est ensuite ajustée au nuage de points en utilisant les méthodes *akima* [12] et *Piecewise Cubic Hermite Interpolating Polynomial* (PCHIP) [14] dans l'espace de données. L'erreur quadratique moyenne (*Root Mean Square Error*, RMSE) et le critère de Pearson (R^2) sont ensuite évalués pour chacune des *splines*. Les *splines* les plus prometteuses sont ensuite tracées dans l'espace 3D avec le nuage de points et la *spline* qui représente le mieux la forme du nuage de points des données de l'échantillon d'entraînement est ensuite sélectionnée manuellement. La distance entre chaque point de données et la *spline* précédemment sélectionnée est ensuite calculée et stockée. Les points de données du 5^e percentile le plus éloigné sont considérés comme des données aberrantes et sont supprimés de l'ensemble de données d'apprentissage. D'autres algorithmes non supervisés de détection de données aberrantes comme *Isolation Forest* [15], *Nearest Neighbours* [16] et *Local Outlier Factor* [17] ont été testés à cette étape, mais ils ont tous présenté une capacité moindre à rejeter les données hors de la masse principale de points.

2.2 – Génération d'un tableau de correspondance (*lookup table*) : Les données nettoyées sont ensuite divisées selon une grille de dimensions n par m le long des axes de vitesse et de couple et la valeur moyenne du débit est calculée pour chaque segment rectangulaire de la grille résultante. Ce processus génère un tableau dans lequel une valeur de débit

correspondante peut être interpolée pour chaque binôme de coordonnées de couple et de vitesse de rotation.

2.3 – Validation de l’approche : Le modèle d’estimation est ajusté et validé à l’aide d’un jeu de données, qui comprend des mesures de consommation de carburant, de couple et de vitesse à une fréquence d’échantillonnage d’une donnée par seconde, pour un peu plus de 450 heures de mesures en mer sur deux (2) remorqueurs. Le tableau de correspondance obtenu à la sous-étape 2.2 est utilisé pour estimer le débit de carburant en temps réel avec les mesures de vitesse et de couple d’entrée provenant de l’ensemble de données de validation. Le débit estimé est ensuite comparé au débit réel mesuré dans cet ensemble de données avec deux métriques de calcul d’erreur, soit l’erreur absolue moyenne et le RMSE. La fenêtre de temps pour le calcul de l’erreur a aussi été variée pour visualiser l’impact de la largeur de la fenêtre sur l’erreur.

Afin d’analyser le potentiel de généralisation de la méthode, nous avons effectué deux types de validation : moteur à moteur et navire à navire. La validation moteur à moteur est effectuée pour étudier la pertinence de la méthode proposée pour estimer la consommation de carburant d’un moteur à l’aide de données recueillies à partir d’un autre moteur similaire sur le même navire (note : les remorqueurs ont typiquement deux moteurs à bord). En pratique, cela permettrait à un armateur de n’installer qu’un seul débitmètre sur son navire. La validation navire à navire vise à étudier les performances de la méthode proposée pour estimer la consommation de carburant d’un navire à l’aide de données recueillies à partir d’un autre navire ayant des caractéristiques similaires. En pratique, cela permettrait à un armateur de n’installer qu’un seul débitmètre sur un navire de sa flotte.

Les limites de l’approche proposée : L’analyse des résultats a permis de constater le potentiel d’une approche basée entièrement sur les données. Toutefois, nous avons identifié trois (3) limites importantes à surmonter pour permettre son déploiement à plus grande échelle.

- **Limite 1 :** La méthodologie nécessite l'intervention humaine pour le nettoyage des données (2.1), ce qui empêche son automatisation et introduit un biais de sélection dans le processus.
- **Limite 2 :** La méthode de moyennage utilisée pour la définition du tableau de correspondance (2.2) est limitée par la simplification des informations, la dépendance à la définition des dimensions de la grille et l'hypothèse implicite sur la linéarité due à l'utilisation d'une méthode d'interpolation linéaire. Cette méthode semble aussi sensible aux données aberrantes.
- **Limite 3 :** La méthode a été validée sur seulement deux (2) remorqueurs très similaires (2.3) et des validations supplémentaires non présentées dans le mémoire ont confirmés que la méthode est peu généralisable pour l'estimation navire à navire.

Étape 3 – Élaboration d'un second modèle robuste et généralisable (O3)

L'étape 3 vise à améliorer les sous étapes 2.1 à 2.3 de la chaîne de traitement de données pour surmonter les trois limites identifiées. Une version améliorée a été mise au point. Cette nouvelle version permet de systématiser et d'automatiser le plus possible le processus selon certains points importants. Deux améliorations principales ont été proposées.

Amélioration 1 : Le tableau de correspondance (2.2) est remplacé par une relation polynomiale du second degré. Ceci règle les limites associées aux discontinuités aux points de la table, rend le modèle moins sensible à la densité de distribution des points de mesure et se rapproche plus de la nature continue du système modélisé.

Amélioration 2 : L'étape de détection et de suppression des données aberrantes (2.1) est complètement supprimée et l'intégration d'un algorithme de régression linéaire robuste en distribution gère maintenant leur influence. Cet algorithme utilise une forme minimax afin de minimiser l'erreur dans le cas le moins avantageux de l'ensemble de données pour obtenir un modèle final qui est robuste en distribution et donc plus résistant au bruit de mesure et aux

légères variations du système réel. La distance de Wasserstein est utilisée pour définir l'ensemble d'ambiguïté du problème minimax.

Le nouveau modèle est développé en suivant les trois (3) sous étapes suivantes :

3.1 – Réglage de l'hyperparamètre ϵ : L'ajout du composant robuste de la régression introduit également l'hyperparamètre ϵ qui règle l'intensité de la régularisation du modèle. Celui-ci doit être ajusté à l'aide d'au moins deux (2) navires de validation pour réduire au maximum l'erreur sur l'ensemble des scénarios possibles. Pour ce faire, le modèle suit les sous étapes suivantes :

3.1.1 – Régression des coefficients : L'algorithme débute avec une supposition de valeur ϵ et procède à la régression d'un modèle pour chaque navire disponible à l'aide du solveur mathématique intégré à `scikit-learn` [18]. Les valeurs de tous les moteurs de propulsion d'un même navire sont combinées pour cette étape.

3.1.2 – Validation croisée avec les autres navires : Une fois les coefficients obtenus par régression, le modèle est utilisé pour estimer la consommation de carburant de chacun des navires individuellement. Le RMSE est par la suite évalué pour chacune des combinaisons de modèles et navires.

3.1.3 – Suggestion d'un nouveau ϵ : Les résultats de la sous-étape 3.1.2 sont utilisés pour évaluer la performance de l'hyperparamètre ϵ suggéré et une nouvelle valeur est suggérée avant de relancer une nouvelle itération à partir de l'étape 3.1.1. La suggestion de valeurs de ϵ et le processus itératif pour sa sélection est géré avec la librairie `Optuna` pour l'optimisation automatique des hyperparamètres [19].

3.2 – Entraînement du modèle sur un navire spécifique : Avec la valeur optimale de ϵ , il est maintenant possible d'obtenir les coefficients par régression pour un seul navire et ainsi établir un modèle d'estimation de carburant robuste en distribution et spécifique pour un navire particulier.

3.3 – Évaluation de la performance finale du modèle : Une fois le modèle final entraîné, sa performance est évaluée à l'aide du calcul du RMSE sur l'ensemble du jeu de données distincts de l'ensemble d'entraînement. Ceci donne une attente quant à sa performance future quand il sera déployé sur d'autres navires dépourvus d'un débitmètre.

CONTRIBUTIONS

Ce mémoire propose une nouvelle approche d'estimation en temps réel de la consommation de carburant peu intrusive qui présente plusieurs avantages.

Intégration non intrusive : le capteur de couple et de vitesse à fixation externe peut être installé aisément, sans nécessiter de modifications des composants existants du moteur ;

Estimation en temps réel : l'approche permet une évaluation précise instantanée de la consommation de carburant durant l'opération grâce aux mesures continues fournies par les capteurs étant donné que le modèle est optimisé sur une base de temps de l'ordre de la seconde ;

Indépendance de l'environnement : l'utilisation d'un capteur mesurant directement l'impact des conditions extérieures sur le fonctionnement du moteur rend superflu le recours à des données environnementales externes comme l'état de la mer ou des courants ;

Simplification de la modélisation : le caractère fondé sur les données du modèle évite le recours à des modèles thermodynamiques, hydrodynamiques et aérodynamiques complexes ;

Généralisable : l'intégration d'une régression robuste en distribution améliore la transférabilité du modèle malgré des conditions d'exploitation variées et incertaines.

STRUCTURE DU MÉMOIRE

La structure de ce mémoire est orientée de manière à s'harmoniser avec les deux articles (un publié et un en soumission) qui représente le cœur du mémoire. Le CHAPITRE 1 présente le premier article de conférence (publié) qui porte sur le modèle basé sur un tableau de correspondance. Le CHAPITRE 2 présente le second article (en soumission) qui porte sur le modèle polynomial robuste en distribution. La CONCLUSION GÉNÉRALE résume le mémoire et propose finalement des pistes d'améliorations pour la suite du projet.

CHAPITRE 1

VERS UNE NAVIGATION ÉCOÉNERGÉTIQUE : UNE APPROCHE BASÉE SUR LES DONNÉES POUR L'ESTIMATION EN TEMPS RÉEL DE LA CONSOMMATION DE CARBURANT DES MOTEURS DE NAVIRES SANS DÉBITMÈTRE

RÉSUMÉ EN FRANÇAIS DU PREMIER ARTICLE

Cet article propose une nouvelle méthode peu invasive pour estimer en temps réel la consommation de carburant des moteurs de navires sans recourir à des débitmètres. Traditionnellement, les débitmètres fournissent une mesure directe et précise, mais leur installation exige des modifications coûteuses et contraignantes des conduites de carburant, ce qui limite leur adoption par les armateurs.

La méthode présentée repose uniquement sur les mesures du couple et de la vitesse de rotation des arbres des moteurs. Un modèle de régression est développé et validé à partir d'un vaste ensemble de données réelles (environ 464 heures de navigation sur des remorqueurs en activités dans les eaux nord-américaines), comprenant les valeurs mesurées à la seconde du débit de carburant, du couple et de la vitesse de rotation à l'arbre des moteurs. Le traitement comporte plusieurs étapes : décodage et stockage des données, ajustement de *splines* 3D, élimination des valeurs aberrantes, puis construction d'un tableau de correspondance (*lookup table*) reliant couple, vitesse et consommation (débit).

La validation est effectuée selon deux approches : moteur à moteur (d'un moteur à l'autre sur un même navire) et navire à navire (d'un navire à un autre d'une même flotte). Les résultats montrent une précision moyenne de l'ordre de 5 à 10 % pour une estimation temps réel à l'échelle de quelques secondes, et une précision améliorée à moins de 1–2 % lorsqu'on considère des fenêtres de moyennage de plusieurs heures.

Malgré certaines limitations observées au niveau de la robustesse et de la généralisation, l'article démontre le potentiel d'un modèle basé entièrement sur les données et que l'usage d'une table de correspondance permet de capturer les relations non-linéaires non connues ou trop complexes pour être modélisées. Ce modèle constitue une première alternative économique et fiable au débitmètre, permettant aux armateurs d'évaluer l'efficacité énergétique de leurs pratiques opérationnelles et de mieux gérer la consommation de carburant de leur flotte.

En tant que premier auteur, j'ai développé l'intégralité du modèle, monté les scripts *Python* pour le tester et rédigé la vaste majorité de l'article. Le professeur Maxime Berger (2^e auteur) a supervisé la démarche scientifique lors du développement et révisé la rédaction de l'article. En tant que superviseurs des travaux chez le partenaire (*OpDAQ Systèmes Inc.*), Tomy Pineau (3^e auteur) et Charles Massicotte (4^e auteur) ont fourni le jeu de données et révisé l'article pour approuver sa diffusion en tout respect de la propriété intellectuelle de *OpDAQ*.

L'article a été présenté le 23 septembre 2024 à la conférence *OCEANS 2024 – Halifax*, et publié plus tard sur *IEEE Xplore* le 25 novembre 2024. La référence complète est donnée ci-dessous :

V. St-Pierre, M. Berger, T. Pineau, and C. Massicotte, "Toward Energy-Efficient Navigation: A Data-Driven Approach for Flowmeterless Estimation of Ship's Engine Fuel Consumption in Real-Time", *OCEANS 2024*, Halifax, NS, Canada, 2024, pp. 1–5.

Toward energy-efficient navigation: a data-driven approach for flowmeterless estimation of ship's engine fuel consumption in real-time

Vincent St-Pierre
*Department of mathematics,
informatics, and engineering*
*Université du Québec à
Rimouski*
Rimouski, Canada
vincent.st-pierre@uqar.ca

Maxime Berger
*Department of mathematics,
informatics, and engineering*
*Université du Québec à
Rimouski*
Rimouski, Canada
maxime_berger@uqar.ca

Tomy Pineau
Research and development
OpDAQ Systems
Rimouski, Canada
tpineau@opdaq.com

Charles Massicotte
Research and development
OpDAQ Systems
Rimouski, Canada
cmassicotte@opdaq.com

Abstract—Reducing greenhouse gas emissions from ships is a significant challenge facing the maritime industry. In particular, energy-efficient navigation techniques allow shipowners to adapt their itineraries to minimize fuel consumption. However, the evaluation of the effectiveness of these techniques is slowed down by lack of data on the actual consumption of the engines. Direct measurement of fuel using a flowmeter seems an ideal solution to this problem. However, many shipowners do not adopt it since it requires significant modifications to the fuel inlet piping, reducing ships reliability and requiring significant initial investments. This paper proposes a new, less invasive, data-driven approach to estimating the real-time fuel consumption of engines without using a flowmeter on each ship or each engine of a fleet. The strategy is based on a lookup table fitted in a three-dimensional space composed of torque, rotational speed, and flow measurements gathered on real ships. The model is trained and validated with approximately 464 hours of actual measurements on tugboats operating in North American waters over the last five years. The developed model enables shipowners to evaluate at a low cost the effectiveness of operational measures to reduce their fuel consumption.

Keywords— *Energy-efficient navigation techniques, tugboats, fuel estimation, regression*

I. INTRODUCTION

Reducing greenhouse gas emissions from ships is an important issue the maritime industry faces. Available solutions differ greatly regarding their implementation complexity, costs, and emissions reduction benefits [1]. Among these solutions, the use of energy-efficient navigation techniques, such as optimizing ship speed [2], [3] and optimizing sea voyages [4], are known to permit shipowners to reduce their fuel consumption while minimizing initial investments and vessel modifications. However, implementing energy-efficient navigation techniques and evaluating their effectiveness require accurate real-time data on the fuel consumption of ship engines under different operating conditions.

The *OpDAQ OpHMI* system [5] allows shipowners to measure their fuel consumption in real time using a flowmeter and to deduce the efficiency of the engine using a torque and speed sensor installed on the output shaft of the engine [6] (Fig. 1 (a)). Even if the direct measurement of fuel using a flowmeter is the most accurate solution to measure fuel consumption, significant modifications to the engine's fuel inlet pipe are required, which increase installation time and cost and reduce the overall system's reliability, thus slowing its adoption by shipowners. A less invasive solution consists of estimating the fuel consumption using only the engine's output torque and shaft speed measurements [7]. However, its practical implementation requires an accurate mathematical model to estimate the engine's efficiency in real time under different operating conditions [8], [9]. The analysis of consumption data on tugboats obtained using the *OpHMI* system showed that the model must consider the highly nonlinear effect of both torque and speed on the engine's efficiency (Fig. 2).

A physics-based approach to estimating an engine's efficiency requires complex and computationally expensive mathematical modelling and the knowledge of the engine's exact physical and environmental parameters, which are rarely available to shipowners. To overcome the limitations of a physics-based approach, this paper proposes a data-driven approach permitting the real-time estimation of ship fuel consumption using only shaft torque and speed measurements, thus providing a flowmeterless *OpHMI* solution to

shipowners (Fig. 1(b)). The core of the proposed estimation model is a multi-step process. The first step comprises a 3D spline regression used for data cleaning with Euclidean distance rejection. The second step relies on a 2D data partitioning followed by a local average computation. The estimation model is fitted and validated using an extensive dataset, which includes fuel consumption, torque, and speed measurements at a sampling rate of 1 sample per second for hundreds of hours of sea operation on tugboat ships operating in North American waters over the last five years.

The paper is organized as follows. Section II presents an overview of the dataset for training and validating the estimation model. Section III presents the proposed methodology for automated data cleaning, developing and validating the flowmeterless model. In section IV, the accuracy of the proposed method is validated and analyzed over various averaging time windows. Engine-to-engine and ship-to-ship validation results are provided to demonstrate the flexibility of the proposed method.

II. DATASET

OpDAQ possesses an extensive dataset comprising over 100,000 hours of commercial operation across 19 different vessels. In this paper, the analysis is limited to two comparably sized tugboats, referred to as ship 1 and ship 2. The dataset for ship 1 encompasses 646,530 data points over approximately 179 hours, while the dataset for ship 2 includes 1,026,757 data points spanning around 285 hours. Nevertheless, results presented in this paper showed that the dataset is large enough to validate the methodology and enabling its use on a larger scale. Each vessel under investigation has two main engines (engine 1 and engine 2), and around 50 measurements are logged each second across all sensors on the ship. For training and validation, only the input flowrate, output torque, and shaft rotational speed are used from those measurements.

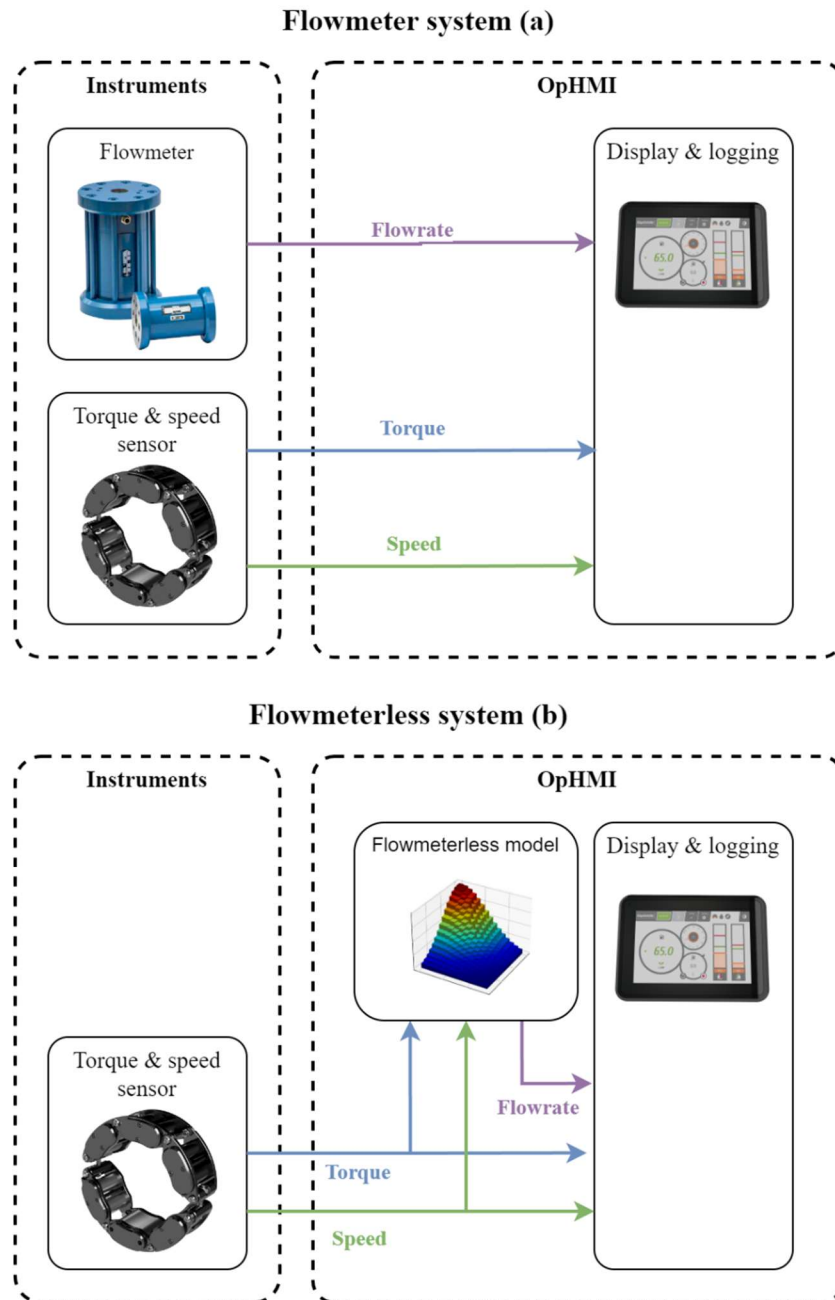


Fig. 1: Evolution of the *OpHMI* acquisition system
(a) flowmeter enabled. (b) flowmeterless.

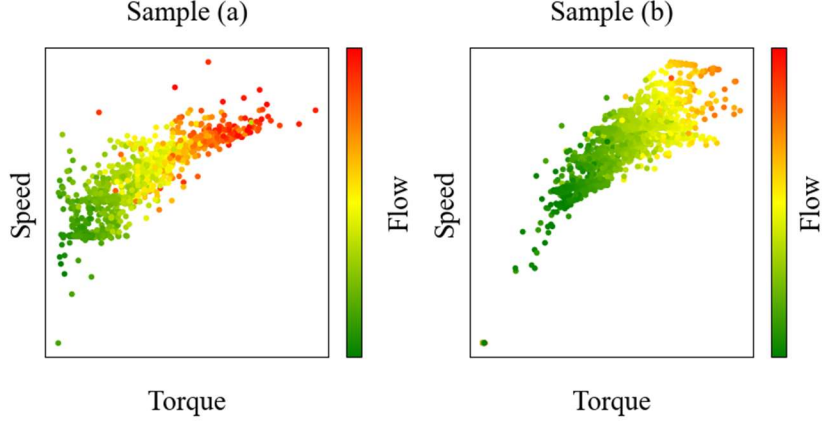


Fig. 2: Variability of flow over similar speed and torque range showcasing the highly nonlinear character of the engine/propeller system.

Engine-to-engine validation is performed to investigate the suitability of the proposed method to estimate the fuel consumption of one engine with data gathered from another similar engine on the same ship. For such validation, measured data from engine 1 are used to train the model, and measured data from engine 2 are used for validation. Ship-to-ship validation aims to investigate the performance of the proposed method to estimate the fuel consumption of one ship with data gathered from another ship with similar characteristics. In this case, the fitted model from engine 1 on ship 1 is first used to estimate the fuel consumption of engine 2 on ship 2. The other way around is also validated, i.e. the fitted model from engine 1 on ship 2 is also used to estimate the fuel consumption of engine 2 on ship 1.

III. METHODOLOGY

A. Data decoding

The first step in processing the measured data is to convert it from the raw binaries produced by the logging infrastructure aboard to a format that is readily usable. An importation tool is developed to this effect which, after decoding the data, saves it to a *PostgreSQL* database for further processing. This process generates a raw dataset ready for use, as shown in Fig. 3(a). This step is performed on both the training and the validation datasets.

B. Spline regression and selection

Next, a 3D Euclidean space is constructed with the flowrate, the torque and the rotational speed measures from the training dataset. A selection of splines are then fitted to the resulting scatter plot using both *akima* [10] and *Piecewise Cubic Hermite Interpolating Polynomial* (PCHIP) [11] methods over 2 to 10 points in the data space. These two regression methods are selected for their ability to generate smooth continuous curves. The *Root Mean Square Error* (RMSE) and *Pearson's Criterion* (R^2) are then evaluated for each of the 18 splines (9 *akima* and 9 PCHIP). The most promising splines are then plotted in the 3D space with the scatter plot (Fig. 3(b)). The spline that best represents the sample training data scatter plot shape is then selected for the next processing step.

C. Outliers removal

The outlier removal step removes extreme data that negatively affects the averaging process performed in the next processing step. To do so, the distance between each data point and the previously selected spline is then computed and stored. The data points in the 5th farthest percentile are considered as outlier data and are removed from the training dataset. The result of this step is a cleaned dataset as represented in Fig. 3(c).

D. Flow lookup table generation

The cleaned data is then split following a n by m grid along the rotational speed and torque axes and the mean flowrate value is calculated for every rectangular segment of the resulting grid. This process generates a table where a corresponding flowrate value can be interpolated for every torque and rotational speed couple coordinates. This process is shown in Fig. 3(d). In this paper, $n=m=30$ is selected as it showed the best overall results.

E. Validation

The generated lookup table is then validated using two different approaches. First, the original scatter plot and the resulting lookup table are plotted in the same space to validate the quality of the fit as shown in Fig. 3(e). Second, the lookup table is used to estimate the

real-time fuel flowrate with the speed and torque measurements input from the validation dataset. The estimated flowrate is then compared with the real measured flowrate in this dataset, and the absolute error is calculated sample by sample as shown in Fig. 3(f). The mean error and RMSE are then computed from this sample-wise absolute error. A rolling window average is also applied to both the measured and estimated flow data for different window size (Fig. 4). The error between the averaged measured and estimated flow data is then calculated sample by sample and the mean error and RMSE are calculated. In this paper, the window size has been varied from 2 seconds to 16 hours to show the impact of averaging window width on the mean error and the RMSE. This analysis is performed to evaluate the impact of temporal averaging on flow estimation error and temporal fidelity.

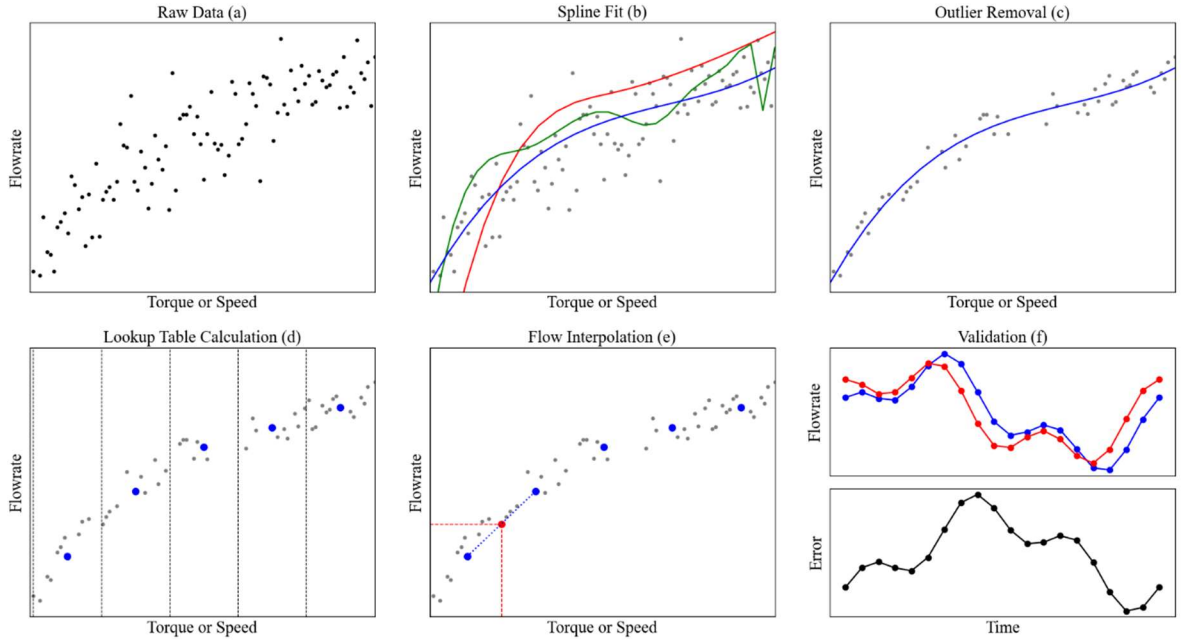


Fig. 3: 2D simplification of the data treatment and model fitting process. (a) The data is first decoded from the raw binaries. (b) Then, a multitude of splines with varying parameters is fitted to the scatter plot. (c) The best-fitting spline is kept, and all the 5th percentile farthest away data from this spline is removed as outlier data. (d) The sample is split into segments, and the mean flowrate value is calculated for each segment. (e) The lookup table generated is used to estimate flowrate from torque and speed data. (f) The lookup table generated from another dataset is used and validated against measured flowrate.

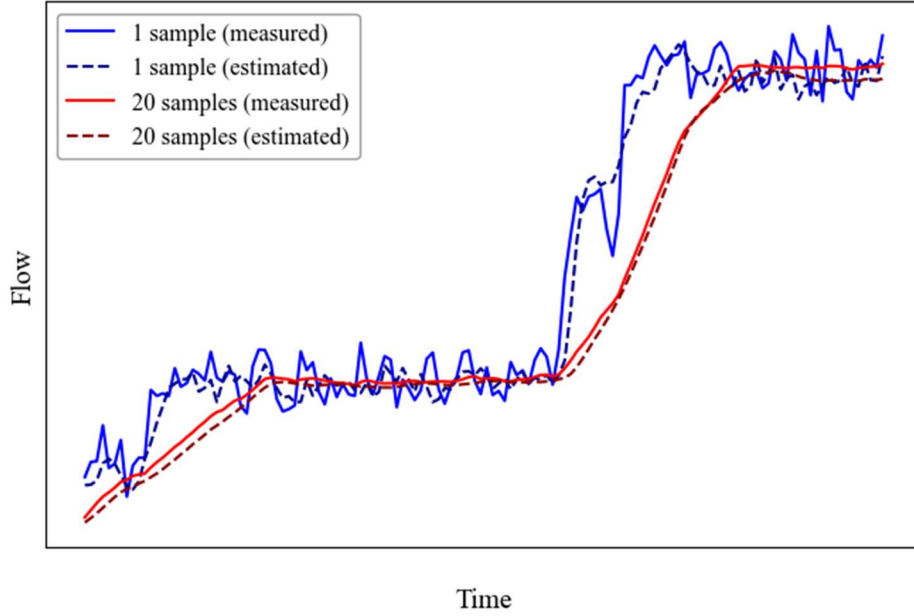


Fig. 4: Effect of increasing the rolling average window sampling width (size) on both the measured and estimated flow samples. Comparing the original measured values (solid blue line) with the averaged measured sample (solid red line) permits noticing the loss of temporal fidelity and the time delay induced by a larger averaging window. However, for each window size (1 and 20), the comparison of the averaged measured values (solid lines) with their respective averaged estimated values (dashed lines) also allows to visually notice that the error with respect to its respective averaged measured data is reduced as the window size increases.

IV. RESULTS AND ANALYSIS

The results obtained by applying the second validation approach for engine-to-engine and ship-to-ship comparisons are summarized in Table I.

The results in Fig. 5 show engine-to-engine validation on ship 1, and the results in Fig. 6 present engine-to-engine validation on ship 2. The results in Fig. 5 show that the procedure yields commendable outcomes, as the mean error is approximately 8% and the RMSE is roughly 20% for a 2 s averaging window. Those values steadily decrease as the averaging window width increases, down to a mean error of 4% and a RMSE of about 5% for a 16 h averaging window. A comparable pattern is observed for ship 2 as depicted in Fig. 6. Specifically, the mean error initiates at approximately 5% and diminishes to roughly under 1%, while the RMSE begins at about 25% and reduces to approximately 1%.

The validation procedure also demonstrates the robustness of the proposed approach when one lookup table that is fitted on one ship is used to estimate the fuel consumption on another ship. In the first case, when ship 2 lookup table is used to estimate fuel consumption of ship 1, the mean error stays relatively constant between 8% and 11% across varying averaging window size while the RMSE decreases from 21% to 11% as demonstrated in Fig. 7. Doing the opposite, i.e. using ship 1 lookup table to estimate fuel consumption of ship 2, yields even better results as demonstrated in Fig. 8 where the RMSE starts at 31% but quickly decreases to 2%, and the mean error starts at 7% and ends as low as 2%.

An analysis of these results along with the explanation in Fig. 4 also permits concluding that the final choice of averaging window width is a trade-off between flow estimation error and temporal fidelity. It is possible to strike a compromise by using a relatively small window (up to 10 samples) for real-time flow assessment where the gain in RMSE accuracy is maximal and a relatively large window when only assessing total fuel quantity burned. The averaging window can be as large as the desired assessment period to minimize error between the averaged measured and estimated values. In all cases, the mean error value is also always lower than the RMSE because an overestimation error can be compensated by an underestimation error in the same averaging window. The RMSE keeps information on the absolute error value and is, therefore, intrinsically higher.

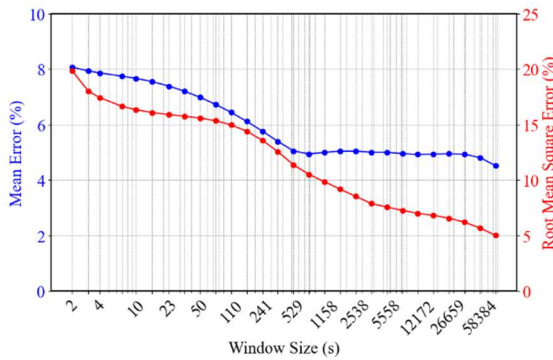


Fig. 5: Results for engine-to-engine validation on ship 1.

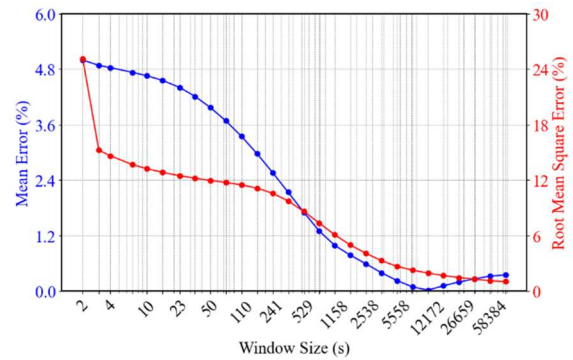


Fig. 6: Results for engine-to-engine validation on ship 2.

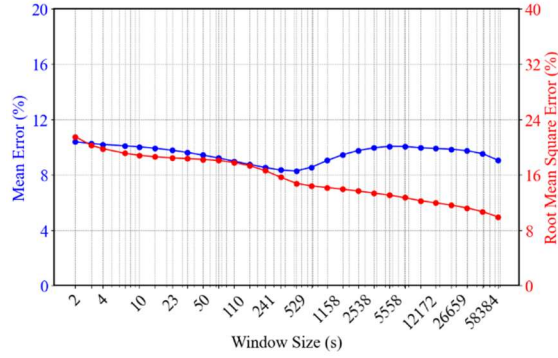


Fig. 7: Results for ship-to-ship validation on ship 1 (fuel estimation of ship 1 with ship 2 lookup table).

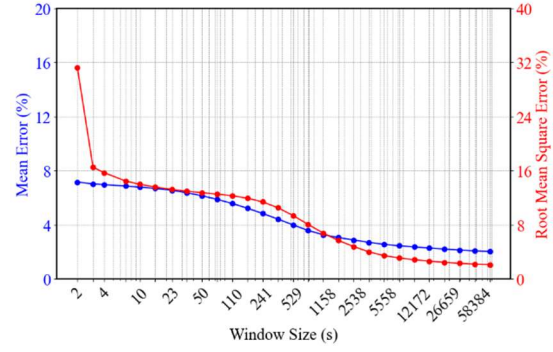


Fig. 8: Results for ship-to-ship validation on ship 2 (fuel estimation of ship 2 with ship 1 lookup table).

TABLE I: VALIDATION RESULTS SUMMARY

	Validation dataset	Lookup table	Mean error		RMSE	
			2s	16h	2s	16h
Fig. 5	ship 1; eng. 1	ship 1; eng. 2	8%	4%	20%	5%
Fig. 6	ship 2; eng. 1	ship 2; eng. 2	5%	<1%	25%	1%
Fig. 7	ship 1; eng. 1	ship 1; eng. 2	10%	10%	21%	11%
Fig. 8	ship 2; eng. 1	ship 2; eng. 2	7%	2%	31%	2%

V. CONCLUSION

This paper presented a data-driven approach to evaluating the fuel-saving capability of energy-efficient navigation techniques. The method is minimally invasive as it only requires the installation of a flowmeter on a single ship or engine. The data gathered on this ship or engine can be used to predict the fuel consumption of another engine or ship having similar characteristics. This method has been tested on 2 similarly sized ships for a total of 4 engines with mean fuel estimation accuracy from 5% error rate for a real-time 2 seconds averaging window and down to less than 1% over a 16 hours window for total fuel quantity estimation. The cross-ship capability is also tested by using the flow lookup table from one ship to evaluate the fuel consumption of the other ship. This approach yields a similar result with mean accuracy ranging between 7% for a real-time 2 second assessment period down to 2% for a 16 hours averaging window.

REFERENCES

- [1] E. A. Bouman, E. Lindstad, A. I. Rialland, and A. H. Strømman, “State-of-the-art technologies, measures, and potential for reducing GHG emissions from shipping— A review,” *Transportation Research Part D: Transport and Environment*, vol. 52, pp. 408–421, 2017.
- [2] G. Finger, A. U. Schubert, T. Riebe, S. Fischer, M. Gluch, and M. Baldauf, “Consumption and Emission Minimised Ship Manoeuvring,” *Global Oceans 2020*, Biloxi, MS, USA, 2020, pp. 1–6.
- [3] H. Lindstad, and G.S. Eskeland, “Low carbon maritime transport: How speed, size and slenderness amounts to substantial capital energy substitution,” *Transportation Research Part D: Transport and Environment*, vol. 41, pp. 244–256, 2013.
- [4] H. Lindstad, B. E. Asbjørnslett, and E. Jullumstrø, “Assessment of profit, cost and emissions by varying speed as a function of sea conditions and freight market,” *Transportation Research Part D: Transport and Environment*, vol. 19, pp. 5–12, 2013.
- [5] “*OpHMI*” *OpDAQ*, <https://OpDAQ.com/en/products/OpHMI/>.
- [6] “*OpTS*” *OpDAQ*, <https://OpDAQ.com/produits/opts/>.
- [7] K. Recep, “An Experimental Approach for Engine Mapping,” *Modern Applied Sci.*, vol. 3, no. 3, 2009, pp. 3–9.
- [8] L. Mak, M. Sullivan, A. Kuczora, and J. Millan, “Ship performance monitoring and analysis to improve fuel efficiency,” *2014 Oceans*, St. John’s, NL, Canada, 2014, pp. 1–10.
- [9] A. Coraddu, L. Oneto, F. Baldi and D. Anguita, “Ship efficiency forecast based on sensors data collection: Improving numerical models through data analytics,” *OCEANS 2015*, Genova, Italy, 2015, pp. 1–10.
- [10] H. Akima, “A new method of interpolation and smooth curve fitting based on local procedures,” *J. ACM*, vol. 17, no. 4, pp. 589–602, 1970.
- [11] F. N. Fritsch and J. Butland, “A method for constructing local monotone piecewise cubic interpolants,” *SIAM J. Sci. Comput.*, vol. 5, no. 2, pp. 300–304, 1984.

CHAPITRE 2

ESTIMATION DE LA CONSOMMATION DE CARBURANT SANS DÉBITMÈTRE POUR L'ESTIMATION EN TEMPS RÉEL DE LA CONSOMMATION DES MOTEURS DIESEL MARINS

RÉSUMÉ EN FRANÇAIS DU DEUXIÈME ARTICLE

Un suivi précis de la consommation de carburant est essentiel pour optimiser l'efficacité énergétique des opérations maritimes. Toutefois, les méthodes traditionnelles telles que les débitmètres installés en ligne avec les conduites de carburant sont invasives, coûteuses et présentent des risques en matière de fiabilité.

Cet article propose une nouvelle approche non invasive pour l'estimation en temps réel du débit de carburant, reposant uniquement sur les mesures de couple et de vitesse de rotation de l'arbre de sortie du moteur. La méthode remplace les débitmètres physiques par un modèle fondé sur les données. Le modèle est entraîné à l'aide d'une régression linéaire robuste en distribution, garantissant ainsi une résilience face au bruit des capteurs et à l'incertitude du modèle. L'aspect robuste en distribution repose sur la minimisation de l'erreur entre l'ensemble de probabilité des valeurs estimées par le modèle par rapport aux mesures réelles. Comme il s'agit d'ensembles de probabilité et non de points précis, la distance euclidienne ne peut pas être utilisée pour calculer l'erreur, elle est donc remplacée par la distance de Wasserstein de 1^{er} ordre. Une procédure de déploiement étape par étape est également proposée afin de permettre le transfert du modèle entre des navires similaires d'une même flotte, à partir d'un seul navire de référence équipé d'un débitmètre.

Le modèle est validé sur un jeu de données réelles comprenant 80 millions de points collectés à chaque seconde sur sept (7) remorqueurs, avec une erreur quadratique moyenne (*Root Mean Square Error*, RMSE) de 5,2 % en moyenne et une erreur maximale

inférieure à 10 % lors de la validation croisée. Ces résultats démontrent la pertinence de la méthode en tant qu'alternative économique, précise et évolutive pour l'estimation de la consommation de carburant dans les applications maritimes.

En tant que premier auteur, j'ai développé la grande majorité de l'algorithme, monté l'intégralité de la programmation *Python* et rédigé la vaste majorité de l'article. Le professeur Maxime Berger (2^e auteur) a dirigé le projet de recherche depuis le début et révisé l'article. Le professeur Antoine Lesage-Landry (3^e auteur) a contribué à l'enrichissement du fondement mathématique du modèle notamment quant à la définition de la régression linéaire robuste en distribution et a révisé l'article.

L'article sera soumis simultanément à l'évaluation de ce mémoire au journal « *IEEE Journal of Oceanic Engineering* ».

Flowmeterless Fuel Estimation For Real-Time Fuel Consumption Estimation in Marine Diesel Engines

Vincent St-Pierre
*Department of mathematics,
informatics, and engineering*
Université du Québec à Rimouski
Rimouski, Canada
vincent.st-pierre@uqar.ca

Maxime Berger
*Department of mathematics,
informatics, and engineering*
Université du Québec à Rimouski
Rimouski, Canada
maxime_berger@uqar.ca

Antoine Lesage-Landry
Department of Electrical Engineering
Polytechnique Montréal
Montréal, Canada
antoine.lesage-landry@polymtl.ca

Abstract— Accurate fuel consumption monitoring is critical for optimizing energy efficiency in maritime operations. Yet traditional methods—such as inline flowmeters—are invasive, costly, and yield potential reliability risks. This paper presents a novel non-invasive approach to estimating real-time fuel consumption using only torque and rotational speed measurements from the engine’s output shaft. The method replaces physical flowmeters with a data-driven model trained via distributionally robust linear regression, ensuring resilience to sensor noise and model uncertainty. A deployment workflow is proposed to enable model transferability across similar vessels within a fleet using a single flowmeter-equipped reference ship. The model is validated using a real-world dataset comprising 80 million data points collected at a sampling frequency of 1 Hz from 7 tugboats, yielding an average root mean square error (RMSE) of 0.052 and a maximum error under 0.100 in cross-validation. These results demonstrate the method’s practicality as a low-cost, accurate, and scalable alternative for fuel consumption estimation in maritime applications.

Keywords — *Distributionally robust optimization, fuel consumption estimation, marine propulsion, non-invasive sensing, linear regression, real-time monitoring, ship energy efficiency.*

I. INTRODUCTION

Spanning trade, energy, and environmental stewardship, the maritime industry serves as a lifeline for modern civilization. The shipping industry, in particular, is the backbone of the contemporary global economy, handling more than 80 percent of the volume of goods transported worldwide [1]. While contributing to only 3% of global greenhouse gas emissions, the sector remains a heavily targeted area for the International Maritime Organization’s (IMO) net-zero emissions by 2050 initiative [2], [3]. While mid-to-long-term solutions include the transition to alternative fuels, the complete electrification of propulsion systems [4] or the integration of advanced shore power connections [5], short-term solutions mainly focus on energy efficiency and eco-navigation initiatives like path or speed optimization to reduce fuel consumption [6], [7]. These methods require high resolution and reliable operational data to fine-tune and evaluate their efficiency, especially for maneuvering when engine regime and load vary significantly on the second timescale. Ideally, a real-time fuel consumption measurement would provide direct information on the ship’s fuel efficiency with varying operating regimes and conditions. Currently, two approaches are widely used to obtain fuel consumption: via CAN bus in the case of an electronically controlled engine [8] or via a dedicated inline flowmeter in the case of a purely mechanical engine [9]. However, the deployment of an inline flowmeter introduces new issues; the installation of additional hardware inherently adds to the financial burden and complexifies the measurement system, and flowmeter obstruction can lead to fuel starvation of the engine, thereby diminishing the reliability of the vessel’s propulsion system. In this paper, we propose an alternative, data-driven fuel assessment method based on flow estimation from engine output torque and speed, which are readily available measurements.

A. Related work

Many studies on operational ship fuel consumption reduction focus on path or speed planning [6], [10], [11]. In the case of [12], weather data is fed into a hydrodynamic model to evaluate the power needed to accomplish a certain amount of work, and the manufacturer’s specific fuel consumption (SFC) is then used to estimate the fuel consumption from the

power measurement. The model is then validated by comparing the fuel estimation against the measured fuel consumption data recorded in the daily ship reports. Although the study demonstrates promising results, its implementation requires an accurate SFC curve for the engine, which can be challenging to acquire, especially on older or smaller vessels [13]. Additionally, as the model is only validated daily, its temporal precision over shorter time intervals remains uncertain.

Reference [14] highlights similar concerns regarding these limitations and addresses them by implementing an automated direct fuel flow measurement method at 15-minute intervals. The study focuses on a single 349-meter container carrier operating at a speed between 10 and 30 knots. While this range includes most cargo transit operations, it does not account for other operational types, such as towing or trolling. For example, tugboats consume most of their fuel at speeds below 10 knots, which falls outside the scope of this study. Specifically, [14] uses an artificial neural network to model and then predict fuel consumption for a specific vessel. However, model generalization from one ship to another and robustness to noisy measurements are not discussed [15]–[17].

Our initial work [18] uses 1 Hz torque and speed measurements of the engine’s output shaft as input features for the model to estimate fuel consumption. The model relies on the use of an outlier detection method and on the development of a two-dimensional lookup table resulting from locally averaged fuel flow for torque-speed couples coordinates sector. This approach yields limited performance, with an RMSE as high as 25% of full scale, when transferring the model from one ship to another.

Contribution

This paper proposes a new real-time fuel estimation approach based on a non-invasive sensor system. It leverages a combined torque and speed sensor designed to be a clamp-on addition to the engine’s output shaft. The proposed fuel estimation method uses the torque and speed measurements from this sensor to estimate the fuel consumption in real-time via distributionally robust linear regression (DRLR). This method offers the following benefits:

- **Non-invasive integration:** The clamp-on torque and speed sensor can be easily installed without modifying existing engine components;
- **Real-time estimation:** Enables accurate instantaneous assessment of fuel consumption during operation thanks to continuous measurements provided by the sensors, given that the model is optimized on a second-wise time scale;
- **Environment-independent:** Eliminates the need for external environmental data such as wave or current conditions, given that the torque and speed sensor directly measures the impact on engine operation;
- **Simplified modelling:** The data-driven nature of the model removes the requirement for complex thermodynamic, hydrodynamic and aerodynamic models;
- **Model generalizability:** The use of distributionally robust linear regression improves the model's ability to generalize across varying and uncertain operating conditions;

II. FLOWMETERLESS FUEL ESTIMATION MODEL

This section presents the proposed approach for estimating fuel consumption without relying on a physical flowmeter.

A. *Flowmeterless system*

The data acquisition system used for this study comprises sensors distributed across a vessel, which communicate with a centralized data logging interface using the Modbus protocol. Currently, the most common configuration for evaluating fuel efficiency involves the installation of at least one sensor capable of measuring both torque and rotational speed on each engine output shaft, in addition to a flowmeter for each engine. This is illustrated in Fig. 1(a).

To minimize the invasiveness of the current system, in this study, we replace the physical flowmeter from Fig. 1(a) with a virtual one. The virtual flowmeter estimates the fuel consumption using the torque and rotational speed measured from the torque sensors, as shown in Fig. 1(b).

Given the recurrent case of shipowners having multiple very similar ships in their fleet, a new workflow is proposed where only one of those ships is equipped with the invasive flowmeter. Data is then collected from this single ship before being used to train a model that estimates real-time fuel flow from the torque and speed measurements from the engine output shaft. This model is finally validated with the data from the flowmeter-equipped ship to be later deployed to the rest of the fleet, as illustrated in Fig. 2.

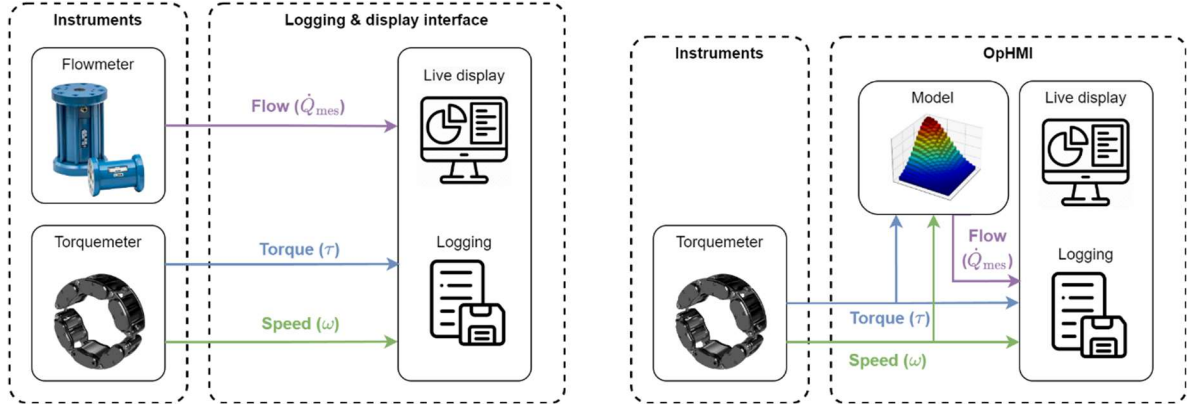


Fig. 1. Comparison between (a) the current acquisition with a physical flowmeter and (b) the proposed new sensor architecture with virtual model-based flowmeter.

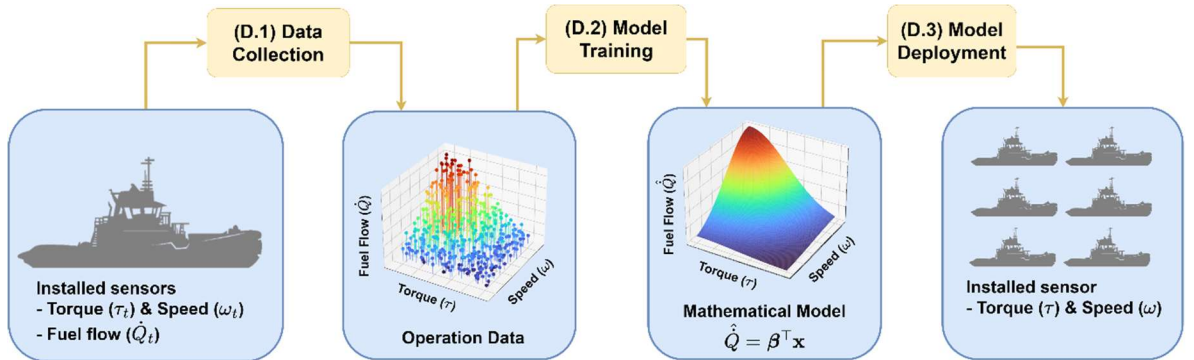


Fig. 2. Step-by-step deployment process on a fleet using a single sample ship and a known ϵ .

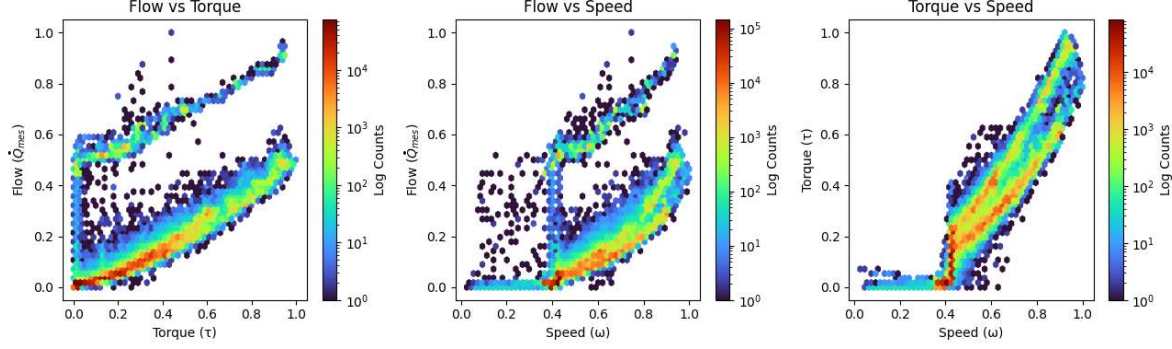


Fig. 3. Visualization of the dataset where the points in 3D space. Axes $\dot{Q}$, τ and ω are projected on 3 orthogonal planes.

B. Fuel estimation model

Let $\dot{Q}$, τ , and ω denote the fuel flow, torque, and rotational speed of the engine, respectively. In this work, we seek to estimate the fuel flow based on measurements of the torque and rotational speed. To guide the model design, the dataset $\mathcal{D} = \{\dot{Q}_t, \tau_t, \omega_t\}_{t=1}^T$, where the subscript t denotes the t^{th} measurement and $T \in \mathbb{N}$ is the number of observations, is projected against its three possible planes, producing the plots in Fig. 3. From this, we discern a general low-order polynomial pattern. Combined with empirical validation, this motivates the choice of a second-order polynomial linear regression defined as:

$$\hat{\dot{Q}} = \beta_1 \tau + \beta_2 \tau^2 + \beta_3 \omega + \beta_4 \omega^2 + \beta_5 \tau \omega, \quad (1)$$

where $\hat{\dot{Q}}$ is the estimated fuel flow. Defining $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_5)^\top$ and $\mathbf{x} = (\tau, \omega, \tau^2, \omega^2, \tau\omega)^\top$, we re-express (1) in vector form as

$$\hat{\dot{Q}} = \boldsymbol{\beta}^\top \mathbf{x}. \quad (2)$$

We define the optimal coefficients of (2), $\boldsymbol{\beta}^*$, in the expected least absolute deviation (LAD) sense, yielding:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^5} \mathbb{E}[|\dot{Q} - \hat{\dot{Q}}|] = \mathbb{E}[|\dot{Q} - \boldsymbol{\beta}^\top \mathbf{x}|], \quad (3)$$

where the expectation is taken with respect to $\mathbb{P}$, the probability distribution of $\mathbf{x}$. Here, the LAD objective is chosen over alternative models due to its robustness against large residuals generated from outlier data points [19]. Motivated by our goal to find a generalizable model that is robust to real-world measurement errors of the specific dataset, (3) is reformulated as a distributionally robust optimization (DRO) problem. To this end, we first need to introduce further notation. Let $\hat{\mathbb{P}}_T \in \mathcal{P}(\mathbb{R}^5)$ be the empirical probability distribution of $\mathbf{x}$ consisting of T samples:

$$\hat{\mathbb{P}}_T \triangleq \frac{1}{T} \sum_{t=1}^T \delta_{\mathbf{x}_t}(\mathbf{x}), \quad (4)$$

where $\delta_{\mathbf{x}_t}$ is the Dirac function used to set a unit probability mass to $\mathbf{x}_t$. We denote the ρ -order Wasserstein distance with respect to some norm $\|\cdot\|$ between distributions $\mathbb{U}, \mathbb{V} \in \mathcal{P}(\mathbb{R}^5)$

$$W_{\|\cdot\|, \rho}(\mathbb{U}, \mathbb{V}) = \left(\inf_{\pi \in \mathcal{J}(\mathbb{U}, \mathbb{V})} \int_{\mathbb{R}^5 \times \mathbb{R}^5} \|\mathbf{z}_1 - \mathbf{z}_2\|^\rho d\pi(\mathbf{z}_1, \mathbf{z}_2) \right)^{\frac{1}{\rho}}, \quad (5)$$

where $\mathcal{J}(\mathbb{U}, \mathbb{V})$ denotes the set of all joint probability distributions over $\mathbb{R}^5 \times \mathbb{R}^5$ whose marginals are $\mathbb{U}$ and $\mathbb{V}$. The distributionally robust optimization problem for flow estimation is:

$$\inf_{\boldsymbol{\beta}} \sup_{\mathbb{D} \in \Omega} \mathbb{E}^{\mathbb{D}}[|\dot{Q} - \boldsymbol{\beta}^\top \mathbf{x}|], \quad (6)$$

where letting $\epsilon > 0$, the ambiguity set Ω is defined as:

$$\Omega = \Omega_{\epsilon}^{s, \mathcal{P}}(\hat{\mathbb{P}}_T) \triangleq \{ \mathbb{D} \in \mathcal{P}(\mathbb{R}^5) : W_{s, \mathcal{P}}(\mathbb{D}, \hat{\mathbb{P}}_T) \leq \epsilon \}$$

using the Wasserstein distance metric. The loss function in (6) is minimized with respect to the worst-case distribution within an ϵ -Wasserstein centred at the empirical distribution of

our observation. Because (6) is intractable, we reformulate it as a convex optimization problem using [19]. Using the 1-order Wasserstein distance with respect to the ℓ_1 -norm leads to:

$$\min_{\boldsymbol{\beta}} \sum_{t=1}^T |\dot{Q}_t - \boldsymbol{\beta}^\top \mathbf{x}_t| + N\epsilon \left\| \begin{pmatrix} \boldsymbol{\beta} \\ -1 \end{pmatrix} \right\|_{\infty}. \quad (7)$$

The resulting problem now comprises both the LAD loss and an ℓ_{∞} -regularization term and represents a distributionally robust model for a given ϵ . From this, we obtain the new hyperparameter ϵ that must be tuned to achieve the best estimation performance across a fleet. To this end, we propose a workflow comprising the subsequent steps for a given ϵ :

- W.1** Collect multiple datasets $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n$ from n different ships;
- W.2** Train separate models $\boldsymbol{\beta}_j^*, j = 1, 2, \dots, n$, on each unique dataset $\mathcal{D}_j$;
- W.3** Evaluate the error $E_{ij}(\epsilon)$ for each model $\boldsymbol{\beta}_j^*$ on every dataset $\mathcal{D}_i, i = 1, 2, \dots, n$;
- W.4** Evaluate the overall loss value across all trials (i, j) for a given ϵ defined as:

$$\mathcal{L}(\epsilon) = \frac{1}{n \cdot m} \sum_{i=1}^n \sum_{j=1}^m [E_{ij}(\epsilon)]^2, \quad (8)$$

where $E_{ij}(\epsilon)$ denotes the RMSE for a specific test with model $\boldsymbol{\beta}^*$ over a dataset $\mathcal{D}_i$ and a given ϵ . This workflow is employed to obtain adequate performance when a model is to be deployed across a complete fleet. The workflow W.1-W.4 can be used to tune ϵ using Optuna [20]. Finally, a trained model can be deployed with the following process (Fig. 2):

- D.1** Collect a single dataset $\mathcal{D}$ from a fleet sample ship;
- D.2** Evaluate $\boldsymbol{\beta}^*$ using the tuned ϵ ;
- D.3** Deploy model to the rest of the fleet using $\boldsymbol{\beta}^*$;

III. CASE STUDY

A. Dataset

The dataset used to train and validate the model and deployment workflow is provided by OpDAQ Systems [21]. Their dataset is developed over a five-year period of operations of their OpHMI acquisition system by anonymously collecting operational data from their clients with their explicit prior consent. To date, it includes over 100,000 hours of continuous real-world operation across 19 different vessels of varying sizes and functions, ranging from 30 meters tugboats to 80 meters icebreakers.

For this study, the scope is limited from the 19 ships down to 7 tugboats around 30 meters in length. The dataset from these tugboats comprises 80 million unique data points sampled at 1 Hz.

B. Results

The procedure described in Section 2 is run over this dataset using scikit-learn [22] to solve (7) during workflow step W.2 and deployment step D.2. Optuna [20] is used to tune ϵ with (8) as loss function when iterating through steps W.2 to W.4. The hyperparameter optimization is run 50 rounds using a dedicated validation data subset in step W.3, resulting in the value $\epsilon^* = 0.0997$.

We set ϵ to the tuned value and retrain the model. Then, the models are cross-tested using an out-of-sample test subset. The results are presented in Fig. 4 where each vertical line represents a scenario in which a different ship is used for training and then the remaining ships are used to assess its performance. Fig. 5 shows the test error distribution for all model and dataset pairing. The numerical results for the 49 cross-tests are summarized in the Table I.

In this case study, the ship owner would be subjected to a RMSE error of 0.0519 from the model when deployed using the proposed non-invasive flowmeterless fuel consumption

estimation method. Fig. 6(a) showcases a sample comparison between real flow and estimated flow for a 33-minute span (2000 s) for a model resulting from Ship C's data, and shows the quality of the fit. Fig. 6(b) zooms on a finer timescale to further showcase the estimation fit over a small timescale. Fig. 7 illustrates a case where the fit is comparatively less accurate, with Fig. 7(b) providing a zoomed-in view to better visualize the local discrepancies between the two curves. We can observe that the oscillations in the estimated flow are of much greater amplitude than the real flow and that there is a constant offset in the estimated value. The predominant hypothesis suggests that vibrations originating from the engine are the primary source of error in this case, potentially inducing deviations in the sensor readings and, subsequently, the estimated flow values. Fig. 8 illustrates the noise in the sensor measurements. It is expected that applying appropriate signal filtering techniques could mitigate the effects of engine-induced vibrations and further improve the quality of the estimated flow. In sum, Ship C's model leads to a RMSE of 0.057 when deployed aboard the same ship with local RMSEs indicated in Figs 6 and 7 captions.

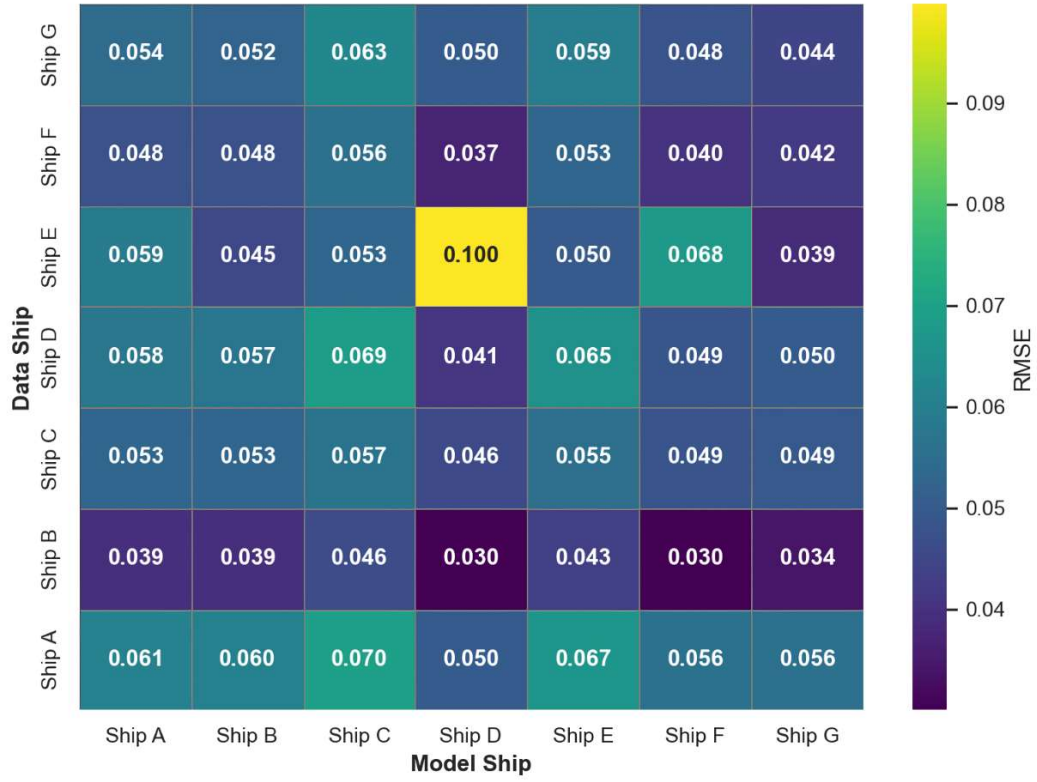


Fig. 4. Results of the cross-validation process

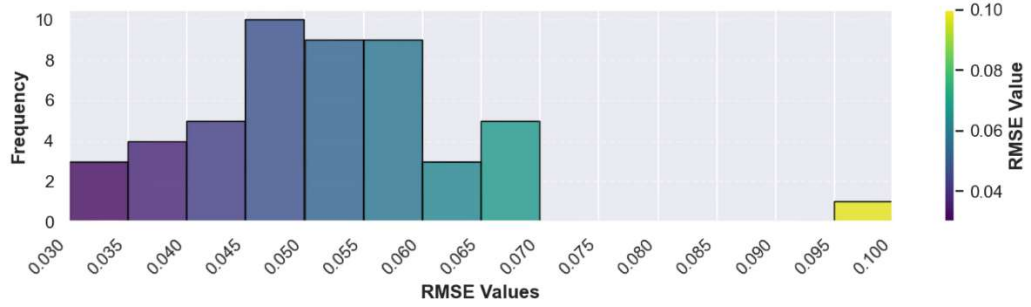
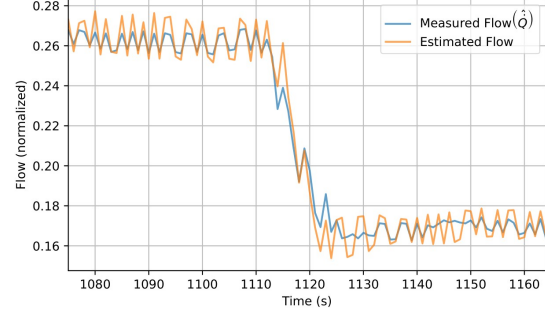
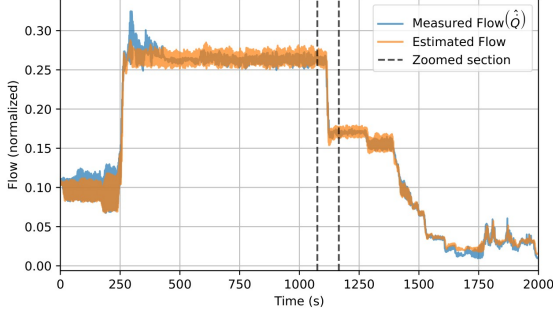


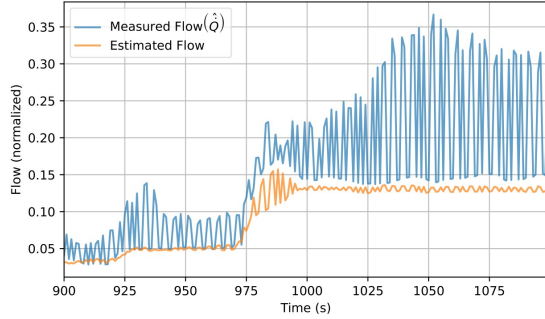
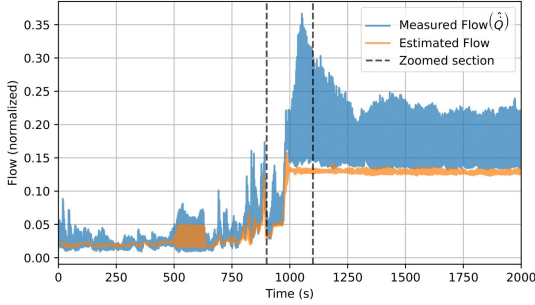
Fig. 5. RMSE distribution during the cross-validation process

TABLE I: CROSS-TESTING ERRORS RELATIVE TO FLOWMETER MESUREMENTS

Mean	Median	Std Deviation	Min	Max	Range
0.0519	0.0500	0.0117	0.0301	0.0998	0.0697



(a) Timespan = 2000 (s) : Local RMSE = 0.0082 (b) Timespan = 90 (s) : Local RMSE = 0.0065
Fig. 6. Sample comparison between real flow and estimated flow from the model

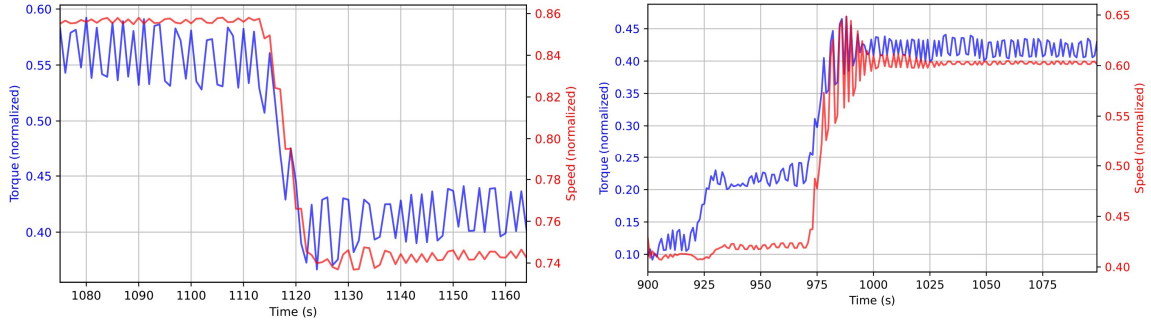


(a) Timespan = 2000 (s) : Local RMSE = 0.0525 (b) Timespan = 200 (s) : Local RMSE = 0.0861
Fig. 7. Validation time series affected by engine-induced vibrations: (a) less accurate fit and (b) detailed view showing localized deviations between estimated and measured flow.

IV. CONCLUSION

This paper introduces a flowmeterless, real-time fuel estimation method based on torque and speed measurements from a non-invasive sensor installed on a ship's engine output shaft. By employing a distributionally robust linear regression framework, the approach provides reliable fuel flow estimates without requiring environmental inputs or intrusive hardware. The model's design explicitly accounts for sensor variability and dataset uncertainty, ensuring robustness and transferability across similar vessels within a fleet. A case study involving 7 tugboats and over 80 million real-world data points confirmed the method's effectiveness, yielding a cross-tested RMSE between 0.030 and 0.100, with most predictions clustered around 0.05. This illustrates the potential for wide deployment in operational environments where reducing sensor footprint is essential. Future work will

explore generalization to other vessel classes and reduction of time-varying error dynamics with signal filtering to further refine and extend the method’s applicability.



(a) Measurements corresponding to Fig. 6(b) (b) Measurements corresponding to Fig. 7(b)
Fig. 8. Torque and speed measurements used for flow estimation

REFERENCES

- [1] United Nations Conference on Trade and Development (UNCTAD), “World investment report 2024: Investment facilitation and digital government,” Geneva, June 2024, sales No. E.24.II.D.6. [Online]. Available: <https://unctad.org/publication/world-investment-report-2024>
- [2] International Energy Agency, “International shipping,” IEA, 2024, accessed: June 12, 2024. [Online]. Available: <https://www.iea.org/energy-system/transport/international-shipping>
- [3] International Maritime Organization, “2023 IMO strategy on reduction of GHG emissions from ships,” IMO, London, Resolution MEPC.377(80), July 2023, annex 15 to MEPC 80/INF.7. [Online]. Available: <https://www.imo.org/en/OurWork/Environment/Pages/2023-IMO-Strategy-on-Reduction-of-GHG-Emissions-from-Ships.aspx>
- [4] S. Qazi, P. Venugopal, G. Rietveld, T. B. Soeiro, U. Shipurkar, A. Grasman, A. J. Watson, and P. Wheeler, “Powering maritime: Challenges and prospects in ship electrification,” *IEEE Electrification Magazine*, vol. 11, no. 2, pp. 74–87, 2023.
- [5] MÉRLIN and Innovation Maritime, “Électrification des quais au Québec,” Innovation Maritime, Tech. Rep., July 2024, accessed: 2024-07-04. [Online]. Available: <https://www.innovationmaritime.ca/wp-content/uploads/sites/288/2024/07/Electrification-des-quais-au-Quebec-MeRLIN-IMAR.pdf>

- [6] L. Zhang, X. Peng, Z. Liu, N. Wei, and F. Wang, "An application of augmented Lagrangian differential evolution algorithm for optimizing the speed of inland ships sailing on the Yangtze River," *International Journal of Naval Architecture and Ocean Engineering*, vol. 14, p. 100488, 2022.
- [7] E. A. Bouman, E. Lindstad, A. I. Riialand, and A. H. Strømman, "State-of-the-art technologies, measures, and potential for reducing ghg emissions from shipping – a review," *Transportation Research Part D: Transport and Environment*, vol. 52, pp. 408–421, 2017.
- [8] CSS Electronics, "Marine telematics & boat data: The complete guide," CSS Electronics, 2024, accessed: 2024-05-15. [Online]. Available: <https://www.csselectronics.com/pages/marine-telematics-boat-data>
- [9] KROHNE, "Fuel consumption & bunker monitoring for supply ships," KROHNE, 2024, accessed: 2024-05-15. [Online]. Available: <https://www.krohne.com/en/industries/marine-industry/supply-ships/fuel-consumption-bunker-monitoring-supply-ship>
- [10] H. N. Psaraftis, "Speed optimization vs speed reduction: The choice between speed limits and a bunker levy," *Sustainability*, vol. 11, no. 8, 2019.
- [11] E. Lindstad, B. Asbjørnslett, and A. Strømman, "Reductions in greenhouse gas emissions and cost by shipping at lower speeds," *Energy Policy*, vol. 39, pp. 3456–3464, 06 2011.
- [12] R. Lu, O. Turan, E. Boulougouris, C. Banks, and A. Incecik, "A semi-empirical ship operational performance prediction model for voyage optimization towards energy efficient shipping," vol. 110, pp. 18–28.
- [13] N. Rehmatulla and T. Smith, "Barriers to energy efficient and low carbon shipping," *Ocean Engineering*, vol. 110, pp. 102–112, 2015.
- [14] Z. Hu, Y. Jin, Q. Hu, S. Sen, T. Zhou, and M. T. Osman, "Prediction of fuel consumption for enroute ship based on machine learning," *IEEE Access*, vol. 7, pp. 119 497–119 505, 2019.
- [15] A. Coraddu, L. Oneto, A. Ghio, S. Savio, D. Anguita, and M. Figari, "Machine learning approaches for improving condition-based maintenance of naval propulsion plants," *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, vol. 231, no. 2, pp. 649–666, 2017.

- [16] X. Han, S. Zhou, X. Yan, and F. Li, “Fuel consumption prediction of container ships based on random forest regression,” *Journal of Marine Engineering & Technology*, vol. 18, no. 1, pp. 30–39, 2019.
- [17] S. Zhou, S. Wang, X. Yan, Y. Yuan, and F. Li, “A hybrid approach for ship fuel consumption prediction using data and knowledge-based models,” *Ocean Engineering*, vol. 220, p. 108398, 2021.
- [18] V. St-Pierre, M. Berger, T. Pineau, and C. Massicotte, “Toward energy-efficient navigation: A data-driven approach for flowmeterless estimation of ship’s engine fuel consumption in real-time,” in *OCEANS 2024 - Halifax*, 2024, pp. 1–5.
- [19] R. Chen, I. C. Paschalidis et al., “Distributionally robust learning,” *Foundations and Trends® in Optimization*, vol. 4, no. 1-2, pp. 1–243, 2020.
- [20] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.
- [21] OpDAQ Systems Inc., “OpDAQ: Home,” <https://opdaq.com/en/home/>, 2021.
- [22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

CONCLUSION GÉNÉRALE

Les travaux présentés dans ce mémoire démontrent la faisabilité et la pertinence d'une approche basée entièrement sur les données pour l'estimation en temps réel de la consommation de carburant des navires, sans recourir à des débitmètres intrusifs. L'ensemble des résultats met en lumière une progression méthodologique claire : d'une première version exploratoire fondée sur des tables de correspondance vers un modèle polynomial robuste en distribution, mieux adapté aux conditions variées de l'opération maritime. Cette trajectoire illustre non seulement l'évolution technique de la recherche mais aussi le renforcement de sa validité opérationnelle et de son niveau de maturité technologique en vue d'un déploiement sur le terrain.

Le premier modèle développé a permis de confirmer la faisabilité de l'approche. En s'appuyant sur des données réelles recueillies auprès de navires comparables, un modèle basé sur un tableau de correspondance reliant couple, vitesse et consommation a été mis au point. Cette méthode, bien que simple, a démontré qu'une approche entièrement basée sur les données est viable avec des erreurs de l'ordre de 5 à 31 % sur des fenêtres de temps courtes de deux (2) secondes et une amélioration notable de l'ordre de 1 à 11 % lors d'agrégations sur des horizons plus longs allant jusqu'à 16 heures. Elle a ainsi constitué une preuve de concept fonctionnelle, tout en révélant ses limites : dépendance à une étape manuelle de nettoyage des données, sensibilité aux valeurs aberrantes et manque de généralisabilité lorsque le modèle était transféré d'un navire à un autre.

Ces constats ont conduit au développement d'un modèle polynomial de second ordre intégré dans un cadre de régression robuste en distribution. Ce raffinement méthodologique a permis d'automatiser le traitement des données, de réduire la dépendance à l'intervention humaine et de mieux prendre en compte la non-linéarité intrinsèque entre le couple, la vitesse et la consommation. Surtout, l'introduction de la robustesse en distribution a doté le modèle d'une capacité de généralisation accrue, validée sur une flotte de sept (7) navires distincts. Les résultats obtenus, avec des erreurs centrées autour de 5 %, montrent que l'approche n'est

pas seulement applicable dans un contexte expérimental restreint, mais peut être envisagée comme une solution viable et transférable sur le terrain.

Certaines limites demeurent néanmoins. Le modèle a été principalement validé sur des remorqueurs, ce qui restreint encore son extension à d'autres types de navires. De plus, bien que les marges d'erreur se situent dans un intervalle acceptable (3 à 10 %), elles peuvent varier selon les conditions particulières d'opération, ce qui incite à poursuivre l'analyse de la dynamique temporelle de l'erreur. Enfin, l'approche repose sur la disponibilité de jeux de données représentatifs, condition indispensable à l'efficacité du modèle.

La prochaine étape du projet serait d'étendre la validation du modèle sur une plus grande variété de types de navires et d'opérations tels que le transit océanique ou le chalutage. De plus, il pourrait être avantageux de tester d'autres méthodes d'apprentissage qui mettent à profit les relations temporelles tels que les réseaux de neurones récurrents.

Pour conclure, ce travail aura contribué à l'effort collectif visant à réduire l'empreinte environnementale du transport maritime et à soutenir la transition vers une navigation plus écoénergétique en facilitant l'évaluation de la consommation de carburant des navires.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] OMI, An opportunity for zero-emission shipping in Canada,” [wwf.ca](https://wwf.ca/wp-content/uploads/2021/04/Alternative-Fuels-for-Canada-Pointing-the-direction-towards-low-zero-emission-shipping.pdf), <https://wwf.ca/wp-content/uploads/2021/04/Alternative-Fuels-for-Canada-Pointing-the-direction-towards-low-zero-emission-shipping.pdf> (consulté 16 juin 2023).
- [2] Alliance verte, “Accueil: Alliance verte,” Alliance verte - Green Marine, <https://allianceverte.org/> (consulté 16 juin 2023).
- [3] Pesca Environnement, KPMG et Dunsky, (2024) “Plan d’action pour décarboner l’industrie maritime du Québec — Version finale, mandaté par la Société de développement économique du Saint-Laurent (Sodes) ”, Carleton-sur-Mer, Québec
- [4] OMI, “Stratégie initiale de l’OMI en matière de GES”, Ministère de l’Environnement, “Reducing greenhouse gas emissions from ships”, [imo.org](https://www.imo.org) (consulté 29 août 2025).
- [5] E. A. Bouman, E. Lindstad, A. I. Rialland, and A. H. Strømman, (2017) “State-of-the-art technologies, measures, and potential for reducing GHG emissions from shipping — A review”, Transportation Research Part D: Transport and Environment, vol. 52, pp. 408–421.
- [6] V. N. Armstrong, (2013) “Vessel optimisation for low carbon shipping”, Ocean Engineering, vol. 73, pp. 195–207
- [7] H. Lindstad, & G. S. Eskeland, (2015). “Low carbon maritime transport: How speed, size and slenderness amounts to substantial capital energy substitution. Transportation Research Part D: Transport and Environment”, vol. 41, pp. 244–256.
- [8] H. Lindstad, B. E. Asbjørnslett, & Jullumstrø, E. (2013). “Assessment of profit, cost and emissions by varying speed as a function of sea conditions and freight market”. Transportation Research Part D: Transport and Environment, vol 19, pp. 5–12.
- [9] V. St-Pierre, M. Berger, T. Pineau, and C. Massicotte, “ Toward Energy-Efficient Navigation: A Data-Driven Approach for Flowmeterless Estimation of Ship’s Engine Fuel Consumption in Real-Time, “ OCEANS 2024, Halifax, NS, Canada, 2024, pp. 1–5.
- [10] “Kral,” *OpDAQ*, <https://OpDAQ.com/produits/kral/> (consulté 16 juin 2023).
- [11] “Opts,” *OpDAQ*, <https://OpDAQ.com/produits/opts/> (consulté 16 juin 2023).
- [12] H. Akima, “A new method of interpolation and smooth curve fitting based on local procedures,” J. ACM, vol. 17, no. 4, pp. 589–602, 1970.

- [13] Massicotte, C. (2022) ‘Less GHGs on the Seas: Practical Solutions to Measure and Record Energy Consumption’. rapport interne *OpDAQ* Systems
- [14] F. N. Fritsch and J. Butland, “A method for constructing local monotone piecewise cubic interpolants,” *SIAM J. Sci. Comput.*, vol. 5, no. 2, pp. 300–304, 1984.
- [15] F. T. Liu, K. M. Ting, and Z. H. Zhou, “Isolation-based anomaly detection,” *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 6, no. 1, pp. 3, 2012.
- [16] T. Hastie, R. Tibshirani, and J. Friedman, “Prototype Methods and Nearest-Neighbors,” in *The Elements of Statistical Learning*, Springer Series in Statistics, Springer, New York, NY, 2009.
- [17] M. M. Breunig, H. P. Kriegel, R. T. Ng, and J. Sander, “LOF : identifying density-based local outliers,” in *ACM SIGMOD Record*, May 2000.
- [18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “*Scikit-learn*: Machine learning in *Python*,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [19] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.
- [20] Marimo, “marimo | a next-generation *Python* notebook,” *marimo.io*, <https://marimo.io/> (consulté 27 août 2025).

