Maritime Transport Research 2 (2021) 100024

Contents lists available at ScienceDirect

Maritime Transport Research

journal homepage: www.elsevier.com/locate/martra

Predicting vessel speed in the Arctic without knowing ice conditions using AIS data and decision trees

Prithvi S Rao[a], Ekaterina Kim[b,*], Bjørnar Brende Smestad[b], Bjørn Egil Asbjørnslett[b], Anirban Bhattacharyya[a]

[a] Department of Ocean Engineering and Naval Architecture, Indian Institute of Technology Kharagpur, India
[b] Department of Marine Technology, Norwegian University of Science and Technology, Trondheim, Norway

ARTICLE INFO

Keywords:
Arctic
AIS
Decision trees
Speed

ABSTRACT

The vessel speed is one of the important parameters that govern safety, emergency, and transport planning in the Arctic. While previous studies have traditionally relied on physics-based simulations to predict vessel's speed in ice-covered waters, most have not fully explored data-driven approaches and powerful supervised machine learning tools to aid speed prediction. This study offers a perspective of applying supervised machine learning models to predict MV SOG using historical Automatic Identification System (AIS) data and *without* explicit knowledge of local ice conditions. This paper presents a case-study from the region of the Eastern Barents Sea and the Southern Kara Sea. We first analyzed the vessel traffic situation for the years 2017 and 2018, and then used this knowledge to build statistical models to predict vessel speeds. Finally, we evaluated the models' performance on a test dataset from January 2019. Performance of three models (Random Forest, XGBoost, and LightGBM) have been tested with a variety of date-time handling techniques, and data input mode being permuted to arrive at the most optimal model. The results demonstrate the ability of the models to predict the vessel's speed based on its geographical location, time of the year and other engineered features such as daylight information and route. With the proposed approach we were able to achieve mean absolute error 3.5 knots in average on a test dataset without explicit knowledge of local ice conditions around the vessel, with the majority of the errors being in the Kara Strait region and the Sabetta Channel.

1. Introduction

The Arctic experiences an increase in passenger cruise voyages as well as in marine export shipping of non-renewable natural resources from coastal regions and the interior. The prediction of vessel speeds in the Arctic remains a challenge in view of rapidly changing ice conditions (including man-made ice ridges) and specifics of ice navigation (icebreaker assistance, ice channels, etc.). For example, during convoy operations (when several ships are following an icebreaker), the captain of the leading icebreaker nominates the speed, the distance between the vessels, or the engine mode depending on the ice conditions and on the ice-going capability of the vessels in convoy. Thus, the problem of speed prediction on an Arctic route can be related to a system of ships rather than to a single vessel (i.e., its tactical and technical characteristics). As the demand upon shipping in the Arctic and ice infested waters are likely to increase in the future, a more in-depth understanding and modelling of the speed regimes in the ice infested waters becomes increasingly more important.

* Corresponding author.
E-mail address: ekaterina.kim@ntnu.no (E. Kim).

https://doi.org/10.1016/j.martra.2021.100024
Received 15 December 2020; Received in revised form 2 May 2021; Accepted 7 May 2021

Traditionally, the topic of ship's speed has been addressed within the context of ship design, operation, and its implications on transportation and safety. Existing tools for the simulation of ship speed in ice mainly comprise physics-based approaches. The latter, following (Milaković et al., 2020), can be separated into two groups: (1) computer simulations of ship performance in ice (e.g., Valanto 2001; Wang 2001; Lau 2006; Sawamura et al., 2008; Su et al., 2010; Lubbad and Løset 2011; Erceg et al., 2014; Zhou et al., 2016; Li et al., 2019); and (2) semi-empirical methods for ice resistance estimation (e.g., Lindqvist 1989; Riska et al., 1997; Jeong et al., 2017, Frederking 2003; Kotovirta et al., 2009; Valkonen et al., 2013; Bergström et al., 2016; Kuuliala et al., 2017). A detailed summary and comparison of the existing approaches can be found e.g., in Li et al. (2018).

Data-driven methodologies are relatively new to the field of ship transits in ice but gaining momentum fast. Adamovich et al. (1995) derived empirical equations to calculate ship speed in ice based on 10 years of ice navigational experience. Montewka et al. (2013), and Montewka et al. (2015) developed data-driven models that can predict ship speed using Bayesian networks. Similä and Lensu (2018) used a MATLAB-based random forest tree regression model to estimate ship speed from satellite imagery and ship data. Löptien and Axell (2014) applied multilinear regression of forecasted ice concentration, level ice thickness, ridge density, historical ship speed and other parameters to reconstruct ship speed profiles in the Baltic. Lensu (2015) and Afonin et al. (2018) used AIS data and operational ice data to investigate dependency between ship speed and ice thickness in the Baltic and Northern Sea Route, respectively. Recently, Montewka et al. (2019) developed a hybrid model for estimating ship performance in ice combining traditional engineering and data-driven approaches while (Milaković et al., 2020)applied a supervised machine learning technique to predict the expected ship speed profile in an ice field.

All the existing methods for assessment of ship speed *use ice-field data* as an input in one way or another. To achieve accurate results, the underlying local ice conditions around the vessel must be known in advance. This presents a specific challenge for tactical route planning as one needs to input short-term predictions (e.g., days ahead) of ice conditions at local scale (~100–1000 m) for planning the voyages.

Currently, all models that provide a decision support for tactical planning and strategic planning of shipping involve an estimation of short-term or medium-term ice conditions (days ahead, months ahead); see, for example, (Topaj et al., 2019). The ice charts and forecasts, that are often used for route planning and strategic decisions, represent the conditions at a given time and may not be representative of local ice conditions around the vessel. This is because, the ice conditions can change significantly in a short period of time, specifically within straights. Old tracks of vessels and man-made ice channels are often not shown on these charts and ice forecasts, thus missing the state of brash ice (or also called the channel ice). In addition, the vessels will try finding pathways through the open leads and mechanically weak ice, and the latter information is not accessible by remote sensing. All these specifics of operations generate the interest in models that can efficiently simulate a vessel's speed profile along a route without explicit knowledge of the complex ice-field data.

In this view, this paper addresses the following research question: *how accurately can vessel speed be predicted without explicit knowledge of local ice conditions around the vessel*? As a case study we have selected the region of the Eastern Barents Sea and the Southern Kara Sea and a specific vessel type and ice class, but the presented approach can be applied to other ice-exposed areas and vessels.

The hypothesis is that we can predict the vessel speed at a given location and time of the year, provided a representative historical dataset of operational speeds. The latter implies that the data record is large enough to implicitly capture speed changes in response to changing navigational conditions (ice, hydrometeorological conditions), vessel purpose, ice going capabilities (ice class, deadweight, power), and transport system specifics, including the seasonal (and or yearly) variability in ambient conditions as well as human and operational factors.

To address the research question, we analyzed and enhanced AIS data for the years 2017 and 2018 (in total, around 245,000 individual records) and have explored the predictive capabilities of three popular supervised machine learning models (Random Forest, XGBoost, and LightGBM). This includes different data representation techniques such as the date-time handling and input data arrangement. Being able to predict vessels' speed on a route, is important for the planning of marine transportation and search and rescue in ice infested waters.

The paper is arranged as follows: We first briefly report specifics of navigation and underlying conditions in the studied region, present the AIS data along with the methods used to transform data into a state suitable to train and validate the different machine learning models. Next, we analyze and compare predictions from three supervised machine learning models, i.e., the Random Forest model, the XGBoost model, and LightGBM for an unseen dataset followed by a brief discussion about the capabilities and limitations of the models and modelling techniques. One of the best model implementation (i.e., XGBoost model) is provided on GitHub (martra-AISspeed), and the presented approach could be used to support planning and active management of marine transportations as well as search and rescue operations in the Kara Sea region.

2. Region specifics

This study focuses on the region of the Eastern Barents Sea and the Southern Kara Sea for the years 2017 and 2018 – a region of year-round marine navigation along the Northern Sea Route.

2.1. General routes

The western part of the Kara Sea route has two main entrances: the Kara Gate and Cape Zhelaniya (at the northern tip of Novaya Zemlya). The vessels typically move to the Gulf of Ob, Yenisei Bay or on the transit to the eastern part of the Northern Sea Route. For the past two decades, due to effects of the climate change, duration of summertime navigation along the Northern Sea Route, including its segments in the Kara Sea, increased because of the earlier melting and later formation of ice (Shukurov and Semenov, 2018). Furthermore, Arctic shipping fleet expansion is anticipated in the future, and there are difficult navigational conditions in the Kara Sea. The latter is due to a large number of underwater obstacles and shoals, frequent fogs, presence of ice and limited knowledge about the currents (Department of Navigation and Oceanography of the Ministry of Defense of the Russian Federation, 1995). A detailed description of ice regimes along the main transport corridors can be found in e.g., Shumovsky (2012) and Dumanskaya (2013). Hence, only important details, specific to the considered dataset, are presented in the following sub-section.

2.2. Seasonal navigational characteristics

In the beginning of summer navigation (end of June – beginning of July), the ice can be present in the East part of the Barents Sea when approaching the Kara Strait, the Yugorsky Strait, and toward the Cape Zhelaniya. At the end of June, the local ice of the Barents Sea does not create difficulties for navigation, provided there is no ice input from the Kara Sea. Typically, presence of ice complicates navigation (on approach to the Kara Strait and the Yugorsky Strait) until the first decade of July. This is when the local ice in the Barents Sea meets the stronger ice brought from the Kara Sea.

In some years, ice from the Kara Sea can block the entrance to the Kara Strait until end of July. If the ice conditions are severe at the Kara Strait and the southwest part of the Kara Sea, it is recommended then to proceed toward Dikson following the northern route around the Cape Zhelaniya.

In the Kara Sea, winds and wind driven currents have seasonal character. Most difficult navigation conditions are typically in March and April when additional ice from the Ob River collides with the ice in the Kara Sea forming ridges between Bely and Shokalsky islands. In favorable years, the route from the Kara Strait to the Dikson island can be ice free from the beginning of July. It is common to differentiate the following different ice regimes on the route form the Barents Sea to the Kara Sea.

Kara Strait (30-km long) has the most difficult ice conditions. Often ridging and compression are present. The ice is typically present from December to May, in some years – from November until July. In addition, an ice jet phenomenon can be periodically observed. A two-lane traffic separation scheme operates within the strait (Japan Association of Marine Safety, 2016).

Novozemelsky ice massif, NZIM, (a 250-km long accumulation of ice) is in the western part of the Kara Sea between the Novaya Zemlya and the Yamal peninsula. It is typically found in the same region every year and is mostly represented by thick first-year ice with concentration of 8 – 10.

Yamal polynya (70-km long) and *Yamal landfast ice* (20-km long): The Yamal polynya is located between NZIM and the landfast ice of Yamal. In winter, it is frozen and represented by young and relatively thin ice. The landfast ice of Yamal, is up to 1-m thick and present from November until July. The ice channel created by icebreakers or ice strengthened vessels can, in this case, exist up to 4–5 days. A distinction is typically made between three distinct positions of NZIM: A west position, when the core of NZIM is formed near the Novaya Zemlya and its southern parts block the Kara Gate. A central position is when NZIM forms in the central part of the Kara Sea and the leads exist on both sides of it; and an east position – when NZIM forms in the eastern part of the Kara sea, and the lead forms on the west near the coast of Novaya Zemlya. The latter location of NZIM create favorable navigational conditions in the Kara and Yugorsky Straits. However, near the coast of Yamal, the ice conditions can become severe due to ice ridges that form due to interaction between NZIM and landfast ice.

2.3. Current situations

Over the past decade, the ice situation in the Kara Sea has been very unstable, and as reported in Ol'khovik (2019), during September 2018, the speed of ships, on the route between the Kara Strait and Bely Island, varied slightly and was mainly dependent on the density and intensity of the traffic. Analysis of AIS data for March 2018 shows that the speed of ships can significantly change as ships follow the route with a large lateral deviation from it. Furthermore, large Arc7 vessels experience significant speed reduction during summer when transiting through the shallow waters (depths of 10–20 m) located north-northwest of the Bely island (Ol'khovik, 2018).

Navigation is also challenging in the Gulf of Ob. One of the most challenging parts is the approach channel in the waters adjacent to the Sabetta port (SCF Newsletter, 2019). This 42-km long and 295-m wide channel is located in the zone of drift ice. The currents in this area can reach 2.5 knots, and one-way traffic is allowed at a time. For detailed description of navigational conditions, refer SCF Newsletter (2019).

3. Vessel data

The AIS data was retrieved from the Norwegian Coastal Administration (NCA). Along with several AIS base stations, the NCA operates four AIS satellites that circumference the Earth in a polar orbit. These satellites were set in orbit in 2010, 2014 and the two remaining satellites in 2017. Both the increasing number of satellites, and updated technology is reflected in the number of AIS messages retrieved over the years. For this study, we have selected a region of the Eastern Barents Sea and the Southern Kara Sea encompassing the routes that are open all year round (refer to the Fig. 2). This region is outside the reach of the base stations, so all data come from the four satellites. The AIS data collected in the years 2017–2018 was used to train and validate the model. For model testing and comparison, available data during the month of January 2019 was used.

For this study, we have utilized AIS message type 1–3 to retrieve dynamic data such as speed and position, as well as message type 5 to retrieve static data such as IMO-number. The AIS data is transmitted per pre-defined time intervals according to operational state and message content. According to the IMO Resolution A.1106(29) this time interval varies from 3 min when the vessel is anchored, moored or moving at speeds less than 3 knots, and up to an interval of 2 s for speeds higher than 14 knots, when changing the course, or for speeds over 23 knots when moving at a fixed course. The raw satellite data were decrypted using the open source Libais Python library version 0.17 (Schwehr, 2011).

The AIS data was cleaned and enhanced before modelling as described in the following section.

4. Pre-processing of data

The objective of the preprocessing is to produce a clean dataset that preserve the original information but also includes additional information on the vessel characteristics such as size, type, and ice class. The latter should be in accordance with the ice class that is assigned by the Northern Sea Route (NSR) administration.

First, the duplicate rows were deleted, and the records without IMO numbers were dropped as well as the records with abnormal speeds (>21 knots). The Vessel Finder database, combined with published information from the navigational permits on the NSR Administration website were used to extract the information about the vessel size, type, and the corresponding ice class. Records for all special purpose vessels (dredgers, drill ships, tugs, icebreakers, etc.) were removed from the dataset as the speed regime for them is different from that of the transit vessels (cargo ships, tankers). For the interest of clarity, we limited ourselves to the family of vessels with a certain ice class, length, deadweight, and installed power.

Next, we visually inspected the dataset and removed observations on land and sudden jumps in position. The resulting dataset consists of nearly 245,000 individual data samples covering the region as shown in Fig. 1. The dataset contains numerical variables (latitude, longitude) and speed as well as date-time data. The date-time data was handled in multiple ways (for details refer to Section 5.2) to acquire best predicting power of the models. Some of these date-time handling methods included splitting into categorical/cyclic datetime features such as month, day, hour, minute, second as well as splitting into seasonal data.

Regional and temporal data analysis

The considered vessel traffic is driven by marine export shipping of non-renewable natural resources from coastal regions and the interior. We separated the dataset into vessel-specific voyage blocks using *a time-based separation algorithm*, where a data entry was declared as part of a new voyage block if the time-gap between two consecutive points was greater than 8 h. This resulted in a total of 226 voyage-blocks (see Fig. 1 on the right).

In addition, four main navigation regions (Routes) were defined as shown in Fig. 2, namely:

(1) Sabetta – Gulf of Ob;
(2) Barents Sea – Kara Strait – Bely Island;
(3) Barents Sea – Cape Zhelaniya – up to the Gulf of Ob; and
(4) Southeast Kara Sea.

Fig. 2 shows that voyages in the Southeast part of the Kara Sea constitute only 8% of the total data recorded. Furthermore, all the considered vessel traffic in April, May, June, and July mainly occurs via the Cape Zhelaniya (see Fig. 3). There are only 156 AIS messages from the Kara Strait in April, 221 in July, and no messages in May and June. The highest density of AIS messages (in total around 25,000 AIS messages) and most of the records with zero speeds are observed at the entrance to the Ob Bay and in the Gulf of Ob. Similar observations were made by Ol'khovik, 2018 on a different dataset and also for other vessel classes and types. Thus, there seems to be a common 'high message density' trend as two major shipping lines converge (Kara Strait and Cape Zhelaniya).

An interesting observation can be made from the data presented in Figs. 1 and 2. Twenty-six percent of all AIS messages are from vessels transiting through the Kara Strait (Route 2 on Fig. 1), and none of these vessels were registered at 0.0 speed. In contrast to this, despite that the minority of the data are form Route 4 (transits in southeast part of the Kara Sea), the vessels with zero speed

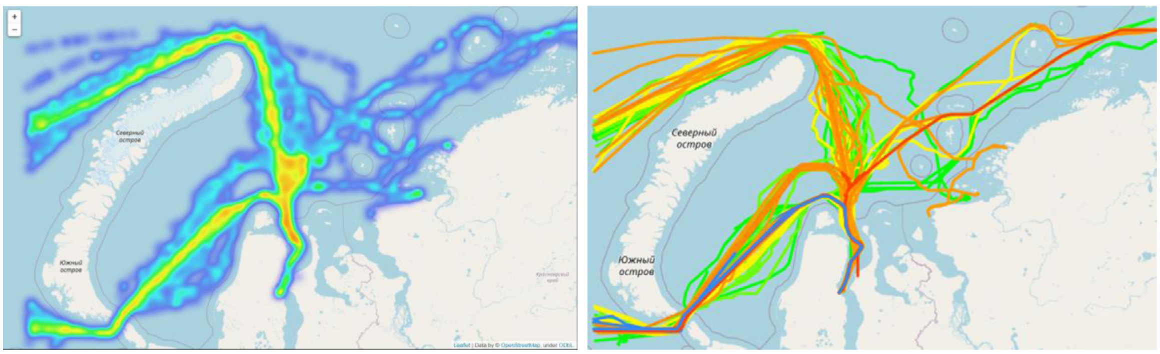

Fig. 1. Spatial visualization of the dataset. Density of AIS messages (on the left) and vessel-specific voyages (on the right). On the left, the color code: orange – high density, blue – low density.

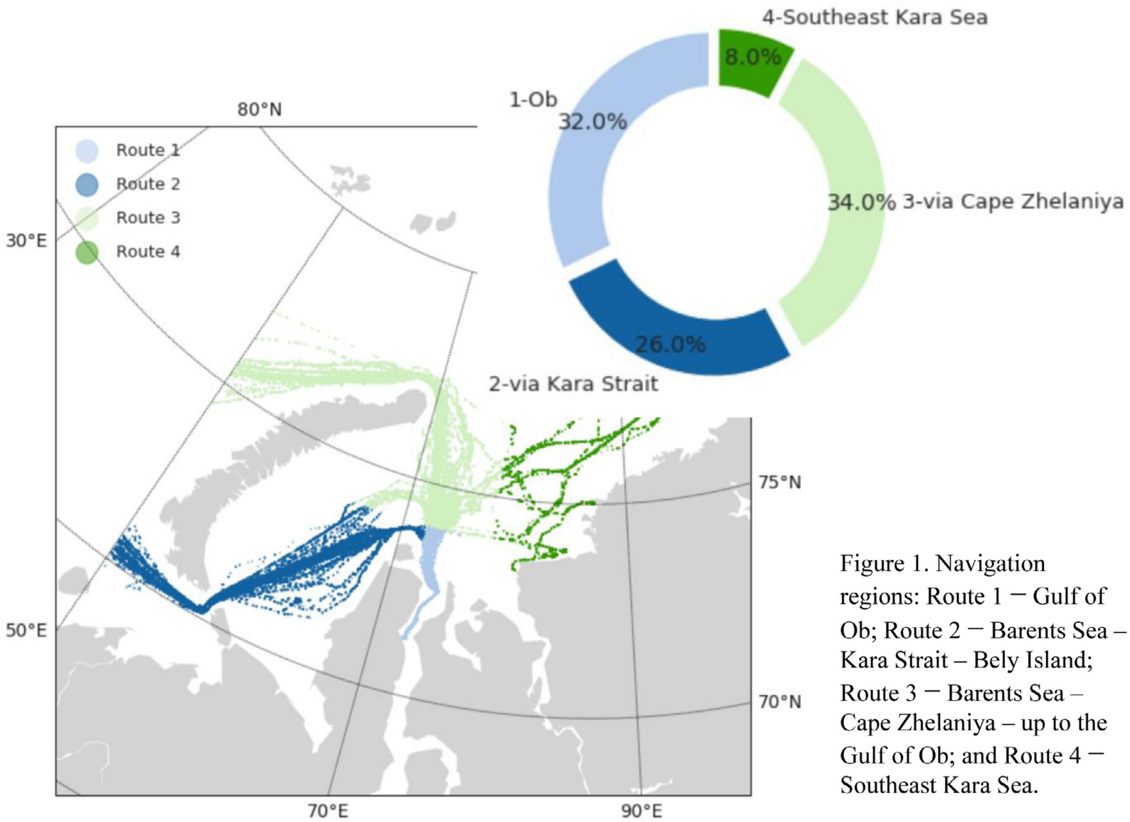

Figure 1. Navigation regions: Route 1 — Gulf of Ob; Route 2 — Barents Sea – Kara Strait – Bely Island; Route 3 — Barents Sea – Cape Zhelaniya – up to the Gulf of Ob; and Route 4 — Southeast Kara Sea.

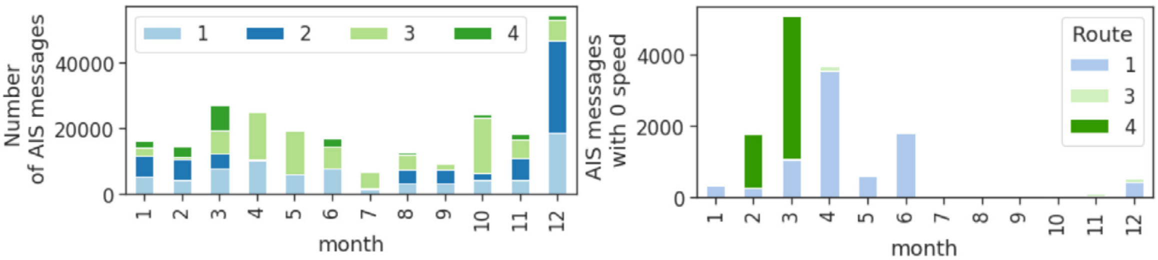

Fig. 2. Total number of AIS messages per month (on the left) and number of AIS messages with zero speed (on the rights) for Route 1: Gulf of Ob; Route 2: Barents Sea – Kara Strait – Bely Island; Route 3: Barents Sea – Cape Zhelaniya – up to the Gulf of Ob; and Route 4: Southeast Kara Sea. Grey color is used for land and white – for sea.

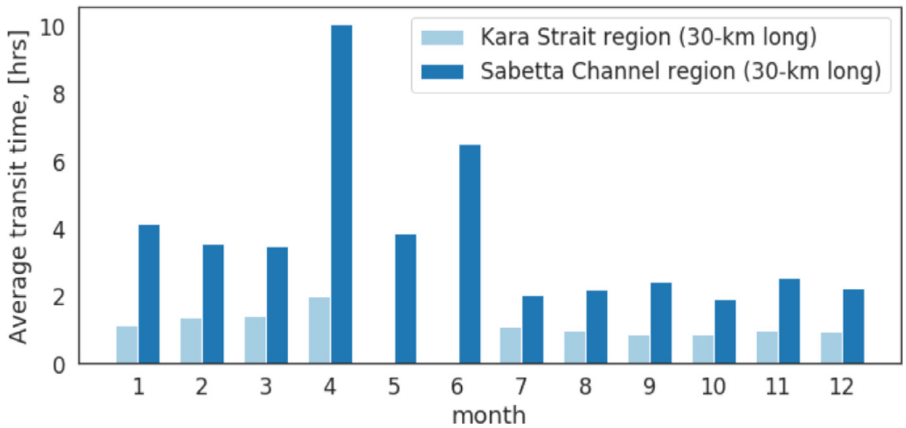

Fig. 3. Average time needed to cross the 30 km long Kara Strait and a 30 km long path within the Sabetta channel.

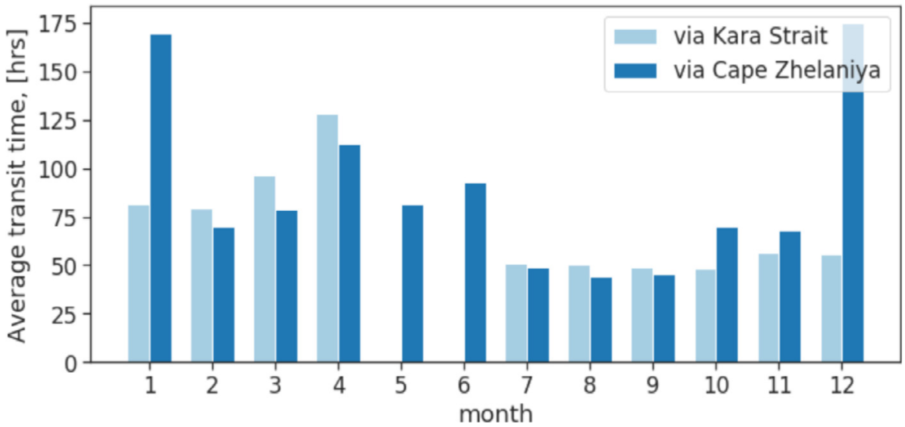

Fig. 4. Comparaison of 'average' voyage durations. The calculated 'average' time for the winter months has a greater uncertainty than that for the summer. This is because in winter the vessels tend to deviate from a predefined Route.

constitute a significant portion of the records, especially during month February and March. Lateral deviations from the general direction of Routes 2 and 3 are typical in the winter months and as noted by Ol'khovik (2019) can be explained by the presence of complex ice situation when the vessels need to search for a route with the least ice thickness and weak ice.

To further investigate specifics of the traffic, we focused on two most navigationally challenging parts (The Kara Strait and the Sabetta Channel) as well as on two most frequently navigated parts (Sabetta – via Cape Zhelaniya: Routes 1 + 3 and Sabetta – the Kara Strait: Route 1 + 2). Studying the 'average' time expenditures at different geographical locations and on specific routes enables us to look for variations in data with respect to time and identify patterns that later can be incorporated into the models. The 'average' time was calculated by dividing the predetermined fixed distance on the route by the average vessels' speeds on this route.

Fig. 3 presents the monthly variation of the normalized average time durations in crossing the Kara Strait (a 30-km long path) and the Sabetta Channel region (a 30-km long path). For the traffic through the Kara Strait, two observations are noteworthy. First, there seems to be no voyages through the strait for the months May and June. Second, the vessels take less time to cross Kara Strait during the latter half of the year (from July to December).

For the 30-km long path within the Sabetta Channel, it can be observed that average transit time is the greatest during the months April, May, June, and that the vessels take less time to in transiting through the Sabetta channel during the latter half of the year (July - December). This latter trend resembles that in the Kara Strait. Note that, both the number of zero speed records in the Gulf of Ob (Fig. 2, on the right) and the average transit time in the Sabetta channel (Fig. 3) are the greatest in April (month no.4).

Fig. 4 reports the monthly variation of the 'average' transit time on the routes 1 + 3 (1252 km) and 1 + 2 (1358 km). On the route via the Cape Zhelaniya, the vessels take significantly more time to complete the voyage during the months December and January. There is a spike in time durations during the month April, May, June but it is insignificant when compared to the spikes in December and January. These voyages' durations in December and January are 2–3 times longer than the voyages' durations via the Kara Strait. The average transit time via the Kara Strait route in winter months (January-March/April) is 1.5 –2.5 times greater than that during second half of the year.

From April till July, the preferred route from Sabetta to Barents Sea is via the Cape Zhelaniya (no vessels were reported in the Kara Strait during May and June. In December and January, the preference is given to transits via the Kara Strait. April is the most challenging for the navigation month as indicated by the number of zero speeds in the AIS data and spikes in the voyage durations in the considered areas.

In summary:

(1) The considered data set is highly unbalanced, both in term of geospatial distribution of data (lower number of messages in the southeastern part of the Kara Sea) and in terms of values of the recorded speeds (the vessels' speed is rarely zero in the Southwest Kara Sea and the considered parts of the Barents Sea region).
(2) There is seasonality in the AIS data, i.e., the vessel routing is affected by the time of the year and this needs to be considered in the models.

5. Modelling approach

In this study we have considered the following factors that can contribute to change in the vessel's speed: the vessel location, daylight information, route, and the time of the year. The task is to investigate how accurate we can predict a vessel's speed without explicit knowledge of local ice conditions around the vessel.

The overall approach is presented in Fig. 5.

We have systematically explored predictive capabilities of three popular supervised machine learning approaches such as Random Forests (Breiman, 2001) XGBoost (E**X**treme **G**radient **Boost**ing) (Chen and Guestrin, 2016), and LightGBM (Ke et al., 2017) which are

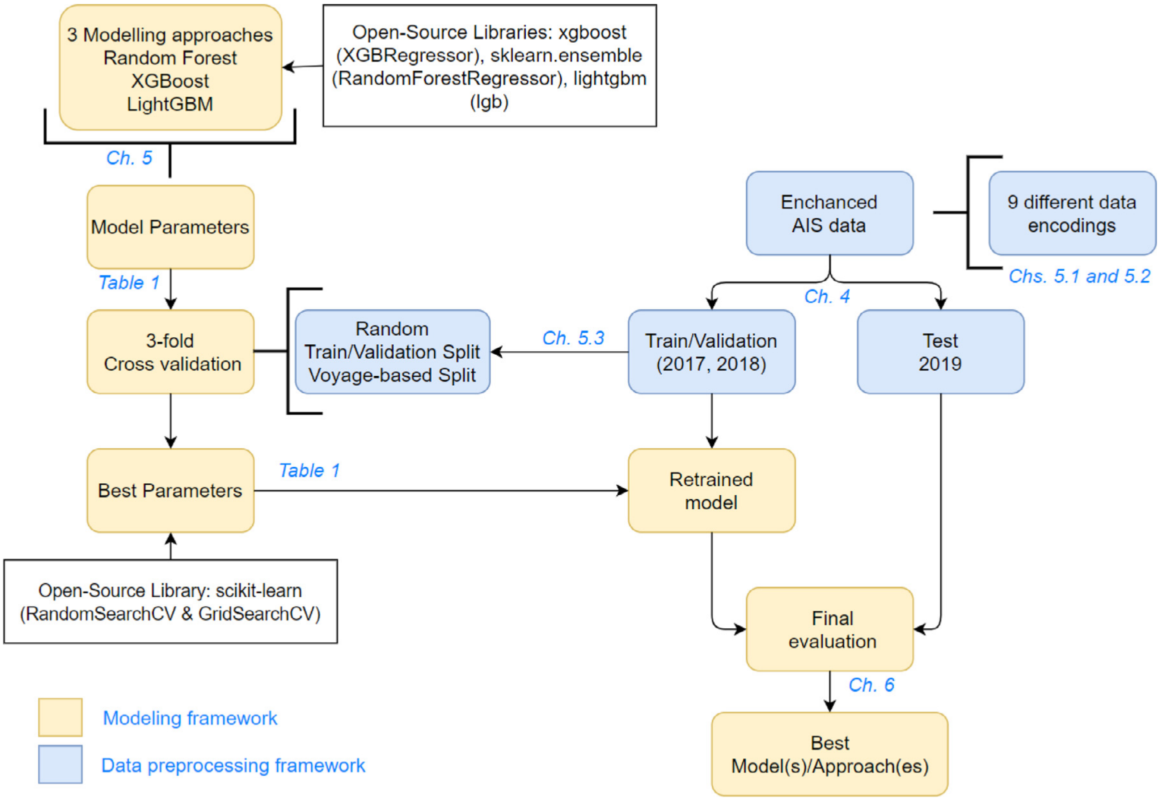

Fig. 5. Flowchart of the overall approach to data-driven speed prediction.

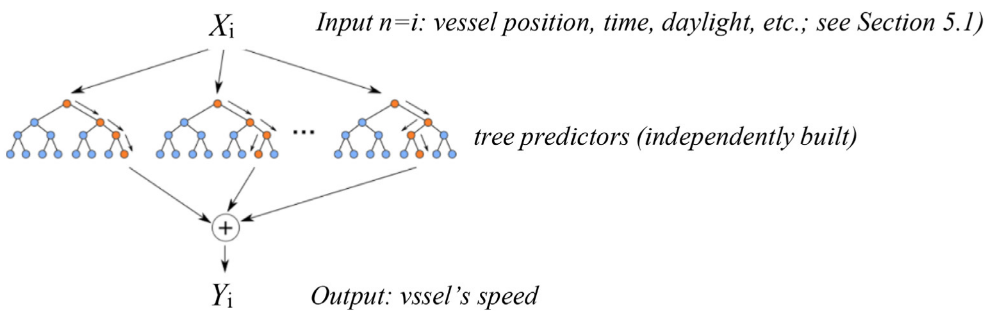

Fig. 6. Random forest predictor, image source (dsc-spidal.github.io/harp/docs/examples/rf/).

recognized for predicting patterns in data and admirably on tabular datasets. These methods are based on a decision threes algorithm (Quinlan, 1986) but are different in how the trees are created and built and in how the predictions of the threes are summed up. The mathematics behind the decision trees can be found in e.g., Breiman et al. (1984) whereas the implementation of the above approaches can be found in Pedregosa et al. (2011), thus only a short summary is provided below. For speed prediction, we have provided the implementation of one of the best-performing model (XGBoost) on GitHub (martra-AISspeed).

Random Forest Approach

The random forest predictor $H(\boldsymbol{\Theta})$ (Fig. 6) is formed by taking the average over K tree predictors $\{h(\boldsymbol{X}, \Theta_k)\}$.

$$H(\boldsymbol{\Theta}) = \left\{h\left(\boldsymbol{X}, \boldsymbol{\Theta}_k\right)\right\},\ k = 1,\ \ldots,\ K \tag{1}$$

where $\boldsymbol{X}$ is the input data matrix with n rows and p features such as vessel's position, route, etc.; $\boldsymbol{\Theta} = [\Theta_1, \ldots, \Theta_K]$, represents the parameters in H and includes splitting variables and their splitting values; K is the total number of trees in the model. These parameters are obtained by training data $\boldsymbol{X}$ and $\boldsymbol{Y}$, where $\boldsymbol{Y}$ is the outcome vector containing vessels speeds. Through the fitted forest predictor, for any set of features X_i, $i = 1, \ldots, n$, we obtain the speed prediction from each tree in $\boldsymbol{H}$.

Gradient boosting approaches

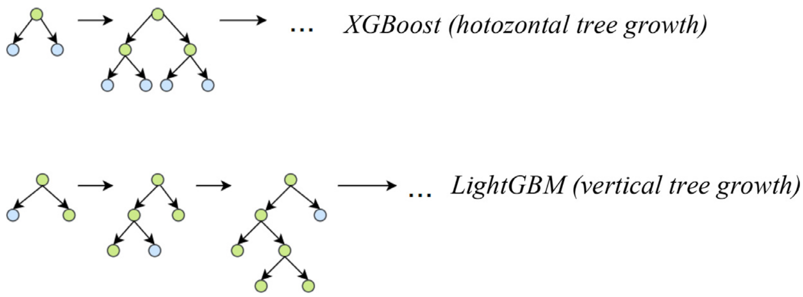

Fig. 7. XGBoost vs. LightGBM; green nodes cannot expand, blue nodes can expand.

Like the random forest approach, gradient boosting approach is a set of decision trees. There are two main differences: (1) how trees are built and (2) how the trees' predictions are combined. The random forest builds each tree independently while gradient boosting builds one tree at a time. The random forest combines the predictions at the end of the process (by averaging) while gradient boosting combines the tree's predictions along the way. XGboost method applies horizontal three growth, whereas LightGBM applies vertical growth (see Fig. 7).

The model (or the exact mathematical structure by which the prediction of speed is made from the input variables), including the model parameters were determined by learning from the two years of enhanced AIS data as described in the following paragraphs.

5.1. Feature engineering

We have introduced new features into the preprocessed AIS data (described in Sections 3 and 4). These features are described in the following paragraphs and include daylight information, route information, and a distance feature.

Daylight feature

To capture the effect of visibility caused by the amount of sunlight at various times of the day at a given location, we have introduced an additional daylight feature to the dataset. The daylight feature varies between 0 and 1, has its maximum at the solar noon of the day and varies linearly. It assumes its minimum value during the nighttime, i.e., the time after sunset and before the sunrise of the next day. The daylight information was collected from the *dateandtime.info* (2020) website using the Sabetta (Yamalo-Nenets Autonomous Okrug, Russia) as a reference.

Route feature

Since the vessels in the dataset are generally transit via one of the four major regions (as shown in Fig. 1, Route 1 – 4), we provide a sense of the 'region' to the predictive model by using a 'route' feature.

Distance feature

As discussed in Sections 2 and 4, there are distinct regional characteristics (the Kara Strait, Sabetta Channel, etc.) which are of great importance to shipping in the Kara Sea. To relay the importance of these geographical places we have introduced four 'distance' features. These are: A – The distance from Sabetta (71.2733, 72.0725); B - The distance from the Kara Strait (70.4807, 58.2225); C – The distance from the point located 3 km west from the Cape Zhelaniya (76.9550, 68.4511); and D – the distance from the Vilkitsky Strait (77.9044, 102.7267).

5.2. Encoding datetime

The AIS data contains information about which year, month, day, hour, minute, and second an AIS message was transmitted. This time information is essential for the model to be realistic; however, it is uncertain of what is the most efficient way of incorporating time information. We have explored the following four options:

(1) no specific date or time encoding – for the purpose of computations only.

(2a) categorical encoding – the time information is represented by month, day, hour, minute, and second.

(2b) categorical encoding – the time information is represented solely by month and day.

(3) seasonal encoding – the time information is coded as two seasons: season one being April, May, and June, season two is from July to March. This is based on the findings in Section 4.

(4) cyclic encoding – cyclic nature of time is encoded using sine and cosine transformations of the datetime by employing *add_cyclic_datepart* function from the fastai library.

In this study, we have used the above daytime encodings with and without additional distance features (refer Results and Discussion Section 6).

5.3. Model training and validation

The data set was split into a training set (66%) and a validation (34%) set. To select data into these two categories, two options were considered. They were:

Table 1
List of tunable and default hyperparameters (optimized model parameters are highlighted in bold green).

Random Forest	XGBoost	LightGBM
Tunable parameters [values considered for optimization]		
bootstrap	n_estimators	n_estimators
[**True**, False]	[50, **100**, 500]	[50,**100**, 500]
max_depth	max_depth	max_depth
[5, 6, **7**, 8, 9]	[3, **5**, 7, 9]	[3, **4**, 5, 6, 7, 8, 9]
max_features	min_child_weight	subsample
[**auto**, sqrt]	[1, 3, **5**]	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, **0.7**, 0.8, 0.9]
min_samples_leaf	gamma	colsample_bytree
[1, 2, **4**]	[0, 0.1, 0.2, 0.3, **0.4**]	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, **0.7**, 0.8, 0.9]
min_samples_split	subsample	learning_rate
[2, 5, **10**]	[0.6, 0.7, **0.8**, 0.9]	[0.01, **0.1**, 0.2, 0.3]
n_estimators	colsample_bytree	
[**50**,100,500]	[0.6, **0.7**, 0.8, 0.9]	
Default parameters		
criterion	eta	num_leaves
min_weight_fraction_leaf	max_delta_step	tree_learner
max_leaf_nodes	sampling_method	min_data_in_leaf
min_impurity_decrease	lambda	min_sum_hessian_in_leaf
min_impurity_split	alpha	bagging_fraction
oob_score	tree_method	pos_bagging_fraction
max_samples	scale_pos_weight	feature_fraction
ccp_alpha	process_type	early_stopping_round
class_weight	max_leaves	min_gain_to_split
warm_start		drop_rate
		boosting

(1) Split based on the voyage blocks – First, split the data into individual voyages by using the time-based algorithmic separation explained in Section 4 (ref. *Regional and temporal data analysis*) Then, we randomly split these voyages into training and validation sets.
(2) Random split – Split each datapoint randomly into training and validation sets. Data from the same voyage can be present in both the training and validation set.

After training each of the three models: Random Forest, XGBoost, and LightGBM, we exposed it to the validation dataset for the estimation and optimization of its performance. The optimization is done by tuning the hyperparameters of the model using 3-fold cross-validation. The hyperparameters for each model are listed in Table 1. The hyperparameters of all the models in different input data settings were tuned using *RandomSearchCV* and *GridSearchCV* method from the scikit-learn library (Pedregosa et al., 2011).

After optimizing the hyperparameters of the models, each model was trained again from scratch on a full dataset (training + validation datasets). For each of the models we tested its predictive capability on the available vessels' data from January 2019 in the same geographical region.

5.4. Performance metrics

Error measures (performance metrics) are vital component of the model evaluation. In this study we employed *mean absolute error* (MAE) (see Eq. (1)) and *standard deviation of the absolute error* (STD) (see Eq. (2)) to evaluate the modelling approaches.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}, \tag{2}$$

$$\mathrm{STD} = \sqrt{\frac{\sum_{i=1}^{n} \left(|y_i - x_i| - \mu\right)^2}{n}} \tag{3}$$

where n is the number of data samples, y_i – single model prediction of speed (knots), x_i – corresponding AIS record of speed (knots), and μ – average of absolute error.

5.5. Models limitations

The models' parameters are based on the historical AIS data (2017, 2018) from four Norwegian satellites. Any limitations of the underlying training data have been transferred to the models. The models will not be able to produce predictions for data points beyond the scope of AIS training data (i.e., to capture any trends in data outside the observed training set). Additional details are provided in the following section.

Table 2
Mean absolute error (MAE, knots) and standard deviation (STD, knots) in speed predictions on the validation dataset (seasonal time encoding). The operational speed range (0–20 knots).

Model	Random Forest		XGBoost		LightGBM	
	MAE	STD	MAE	STD	MAE	STD
Voyage-based data split	3.4	2.6	3.3	2.7	3.2	2.5
Random data split	3.4	2.7	3.3	2.6	3.2	2.5

6. Results and discussion

This paper has addressed the following research question: *how accurately can one predict vessel speed without explicit knowledge of local ice conditions around the vessel*? As a case study we have selected the region of the Eastern Barents Sea and the Southern Kara Sea and a specific vessel type and ice class. It was hypothesized that it is possible to predict the vessel speed at a given location and time of the year, provided a representative historical dataset of operational speeds. We have analyzed and enhanced AIS data for the period of 2017–2018 (two full years) and have explored the ability of three distinct supervised machine learning models (Random forest, XGBoost, and LightGBM) to capture complex relationship in the data and then tested predictive capabilities of several models using different modeling techniques such as time encoding and input data arrangement.

The following paragraphs present and discuss results of this evaluation.

After 3-fold cross-validation of the models (with parameters in Table 1; see values highlighted in green), it was observed that the *random-split* between training and validation datasets gives similar model performance than the *voyage-based split* for seasonal time encoding (Table 2). Since the voyage-based data split requires considerably more preprocessing, random data split is further adopted in the study.

Mean absolute error (knots) and standard deviation (knots) in speed predictions on the test dataset can be found in Table 3. Accuracy scores for each of the data handling method (Random Forest, XGBoost, and LightGBM) and approaches (time encoding, distance features) are plotted in Fig. 8.

As a general trend, it was observed that as the cardinality of the datetime features increased, the performance of the machine-learning models decreased, which is in accordance with the fact that the feature importance score of components such as seconds, minutes, hours, and days was very low.

Based on the accuracy scores presented in Table 3 and Fig. 8 it can be concluded that *XGBoost,* and *LightGBM* models *with seasonal time encoding and with additional four distance features* were better than other approaches with MAE up to 3.5 knots and the standard deviation of 2 – 3 knots on a test dataset from January 2019 (7 vessels in total, speeds ranging between 0 and 20 knots). Similar model performance was achieved without time information, but this have limited applicability in a real-life setting as time is a crucial factor for vessels in transit. Furthermore, this can be an artefact of our test data as only transits in January were considered.

Other models based on feature variations such as cyclic time encoding and categorical time encoding performed significantly worse (MAE up to approximately 6 knots) than those based on seasonal time encoding.

The feature importance of the best performing model is presented in Fig. 9. The vertical axis is the normalized mean decrease in impurity score or Gini importance (Louppe, 2014). Each feature importance was calculated as the sum over the number of splits (across all trees) that include the feature, proportionally to the number of samples it splits (ref. scikit-learn library implementation, Pedregosa et al., 2011). The higher the value of the importance score, the more important the feature is.

Longitude was the most important feature to explain the speed variations in our dataset, followed by season, distance to the Vilkitsky Strait and latitude. Other features (e.g., route and daylight) were found to be less important but still contributed to the predictive capabilities of the models. Note that the distances are derived from the vessel position, thus, features such as 'longitude', 'latitude', and 'distances' (A, B, C, and D) are interrelated. Nevertheless, referring to Fig. 8 and the MAE score in Table 3, the explicit encoding of distances helped the model to better 'understand' the relationships in the data. This improvement is however less pronounced than that with seasonal and cyclic time encodings.

It should be noted that numbers in Table 3 and in Fig. 8, correspond to the assumption that in a voyage, the datapoints are independent of each other. This is not true, since each data point is dependent on the previous predictions in a voyage (speed variation in a voyage must be continuous). This problem is caused by the fact that AIS data is not recorded with the same intervals, but the frequency of recordings depends on the speed of the vessel and satellite coverage. This causes the time series to have readings at unequal intervals. Models such as rolling window regression (see e.g., Zivot and Wang, 2006) which account for inter-dependencies cannot be applied in such cases. We anticipate that the presented models may exhibit shortcomings when are used to predict a continuous speed variation.

To gain deeper insight into the results, we generated speed-distance plots from our predictions and compared those to actual speed records (AIS data). Figs. 10 and 11 present original worst and best predictions on the test set, whereas Figs. 12 and 13 present averaged values. Comparisons for the remaining test cases are provided in Appendix.

Note that in Figs. 10 and 11, referring the constant speed regions, the predictions with Random Forest are less smooth and look like a combination of various step functions. Furthermore, these predictions do not exceed the maximum values of the AIS speed or the minimum of the AIS speed. The observed performance is typical for the random forest models. It can be explained by the inability

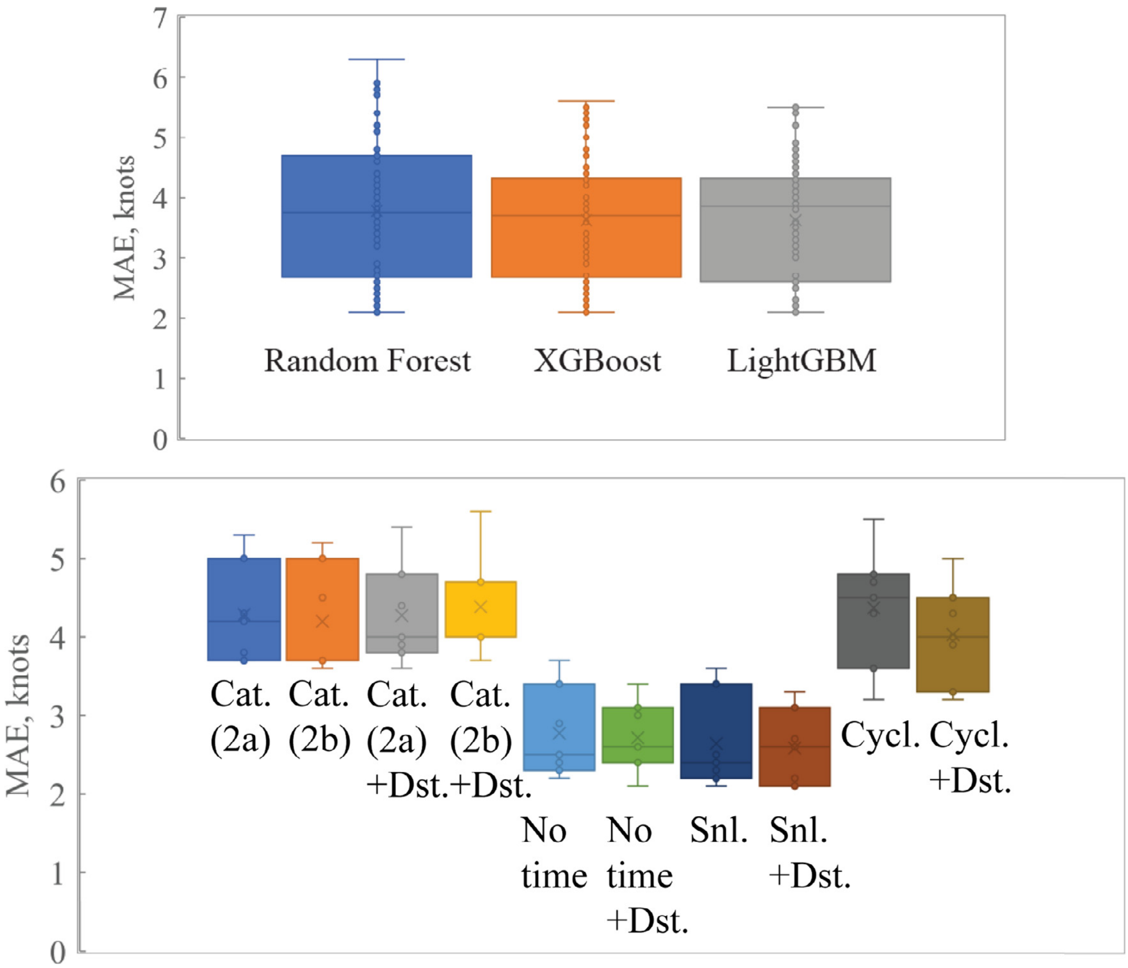

Fig. 8. Accuracy scores (MAE) for different data handling methods (top) and for data handling techniques with the XGBoost model (bottom). Notations: Cat. – Categorical time encoding, Dst. – Distance features (4 in total), No time – No time encoding, Snl. – Seasonal time encoding, Cycl. – Cycic time encoding.

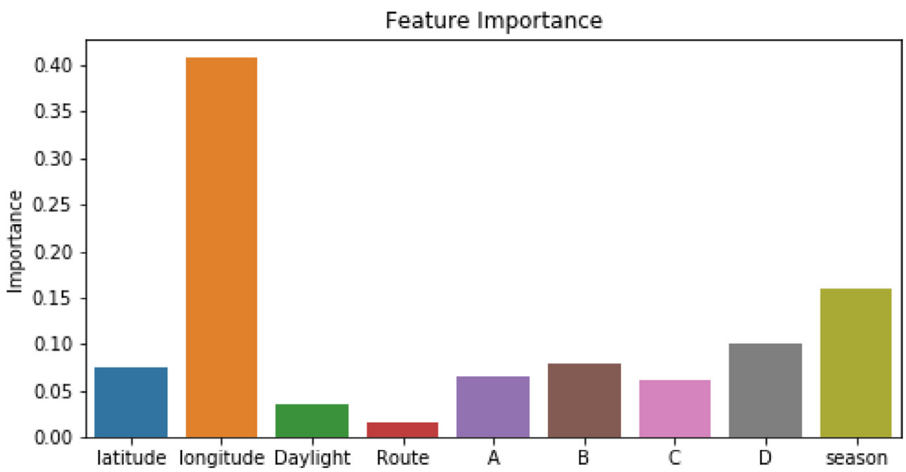

Fig. 9. Feature importance of the XGBoost built on seasonal time encoding and distance features. A - The distance from Sabetta; B - The distance from the Kara Strait; C – the distance from the point located 3 km west from the Cape Zhelaniya, and D - The distance from the Vilkitsky Strait.

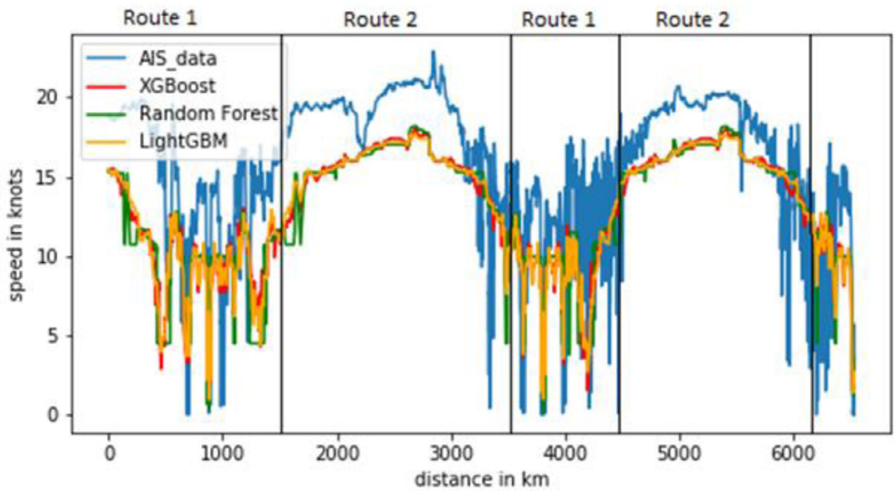

Fig. 10. Worst result (Test no. 6) of the models. The x-axis is the computed distance covered by the vessel from its starting position, and the y-axis is the speed of the vessel predicted by the model and the recorded speed value (AIS data).

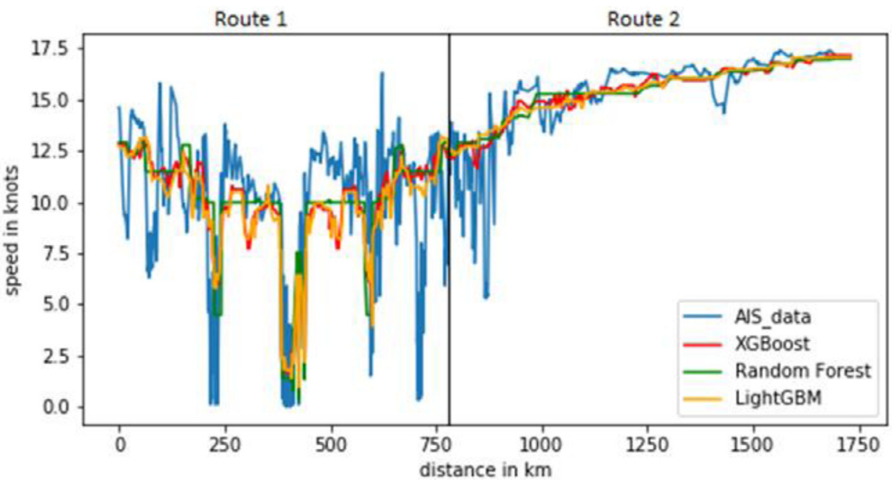

Fig. 11. Best result (from Test no. 7) of the models.

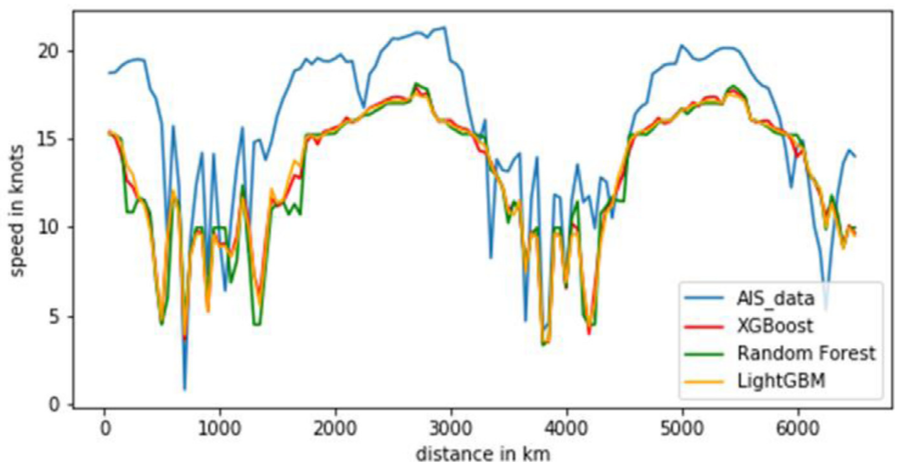

Fig. 12. 50-km averaged worst predictions of the models and recorded AIS speed.

Table 3
Mean average error (knots) and standard deviation (knots) in speed predictions on the test dataset for seven different vessels (Test nos. 1–7), R – Random forest, X – XGBoost, and L – LightGBM.

Method Score	R Mean (knots)	X Mean (knots)	L Mean (knots)	R Stdev. (knots)	X Stdev. (knots)	L Stdev. (knots)	Method Score	R Mean (knots)	X Mean (knots)	L Mean (knots)	R Stdev. (knots)	X Stdev. (knots)	L Stdev. (knots)
Categorical time encoding (2a)	3.3	4.2	3.8	1.9	2.1	2.1	Categorical time encoding (2b) (day, month)	3.3	3.7	4	1.9	2.4	2
	4.4	3.7	3.9	2.7	2.3	2.3		4.4	3.6	3.9	2.7	2.2	2.4
	4.7	4.3	4.5	2.7	2.3	2.3		4.7	4.5	4.4	2.6	2.5	2.3
	5.8	5.3	5.5	2.4	2.2	2.3		5.7	5.2	5.4	2.5	2.3	2.2
	3.8	3.7	4	1.9	1.7	1.8		3.8	3.7	3.9	1.9	1.7	1.8
	5.2	5	4.8	3.1	2.8	3.2		5.1	5	4.9	3.1	2.8	3.2
	4	3.8	3.9	2.3	2	2.2		4	3.7	3.9	2.4	1.9	2.1
Categorical time encoding (2a) with four additional distance features	4.8	3.9	4.1	2.6	2	2.1	Categorical time encoding (2b) with four additional distance features	4.8	4	4.3	2.6	2.2	2.1
	4.4	3.6	4	2.6	2.3	2.4		4.4	3.7	3.9	2.6	2.4	2.3
	4.8	4.4	4.6	2.6	2.6	2.5		4.8	4.7	4.6	2.6	2.6	2.4
	5.9	5.4	5.5	2.6	2.2	2.3		5.9	5.6	5.4	2.6	2.2	2.3
	4.3	4	4	2	1.8	1.8		4.2	4	3.9	1.9	1.8	1.7
	5.2	4.8	4.8	3	3.1	3.2		5.2	4.7	4.7	3	2.9	3.2
	4.3	3.8	4	2.5	2.2	2.2		4.2	4	3.9	2.5	2.3	2.3
No time information	2.8	2.9	2.7	1.9	1.9	1.9	No time information and with four additional distance features	3.2	3	3	2	1.9	1.9
	2.3	2.4	2.3	1.9	1.9	1.9		2.3	2.4	2.3	1.9	1.9	1.9
	2.4	2.5	2.5	2.3	2.4	2.4		2.6	2.6	2.6	2.5	2.6	2.6
	3.7	3.4	3.1	2.4	1.9	1.6		3.6	3.1	3.2	2.2	1.6	1.6
	2.3	2.3	2.2	1.8	1.9	1.8		2.4	2.4	2.3	1.9	2	1.9
	3.2	3.7	3.6	2.6	2.4	2.4		3.3	3.4	3.4	2.6	2.5	2.4
	2.1	2.2	2.2	2.1	1.9	1.9		2.2	2.1	2.1	2.1	2	2
Seasonal time encoding with four additional distance features	2.6	2.6	2.5	2.2	2	2	Cyclic time encoding	3.3	4.7	3.8	2.2	2.3	2.2
	2.2	2.2	2.3	1.9	1.8	1.8		3.3	3.6	3.6	2.6	2.2	2.2
	2.7	2.7	2.6	2.6	2.5	2.4		4.6	4.3	4.4	2.2	2.3	2.2
	2.9	3.1	3.3	1.5	1.6	1.8		6.3	5.5	5.2	2.6	2.5	2.3
	2.2	2.1	2.1	1.9	1.9	1.9		3.9	4.5	3.9	1.8	2	1.8
	3.4	3.3	3.3	2.7	2.5	2.5		4.7	4.8	4.6	2.9	2.8	2.7
	2.1	2.1	2.1	2.2	2.1	2		3.2	3.2	3.5	2.2	1.6	1.6
Cyclic time encoding with four additional distance features	3.3	3.9	3.9	2.2	2.4	2.1							
	3.5	3.3	3.5	2.2	2	2							
	4.7	4.3	4.2	2.7	2.4	2.2							
	5.9	5	5.2	2.6	2.4	2.2							
	4.1	4	3.8	1.9	1.8	1.7							
	5.4	4.5	4.5	3	2.8	3							
	3.5	3.2	3.3	2.2	1.5	1.6							

of random forest models to produce expected predictions for data points beyond the scope of AIS training data (i.e., to capture any trends in data outside the observed training set).

A common feature among all the models is the inferior performance for one vessel in the test set. Although the models are somewhat able to capture the overall speed trend, the predictions are still off. There are two areas where the errors in prediction are high. These areas correspond to the areas within and in proximity to the Sabetta Channel and the Kara Strait.

In several other test cases (Test no. 1 and Test no. 3 in Appendix), the predictions seem to be off initially, which correspond to the Cape Zhelaniya route (Route 3) and the Kara Strait route (Route 2), both regions where the vessels deviate laterally from the route. Excluding these regions, our models seem to be more accurate in predicting 'averaged' speed variations. Nevertheless, even in the previously discussed seemingly erroneous areas, the models have performed well in several other test cases. Despite these limitations, the developed methodology has a potential as a data-driven approach in support of physics-based simulations of vessels speeds in ice covered waters.

Using neural networks to predict vessel speed without ice information (in the same geographical region) yields to similar performance in terms of the overall speed trends on an unseen dataset (ref. Fig. 7 in Kim et al., 2020). Thus, both the decision trees and

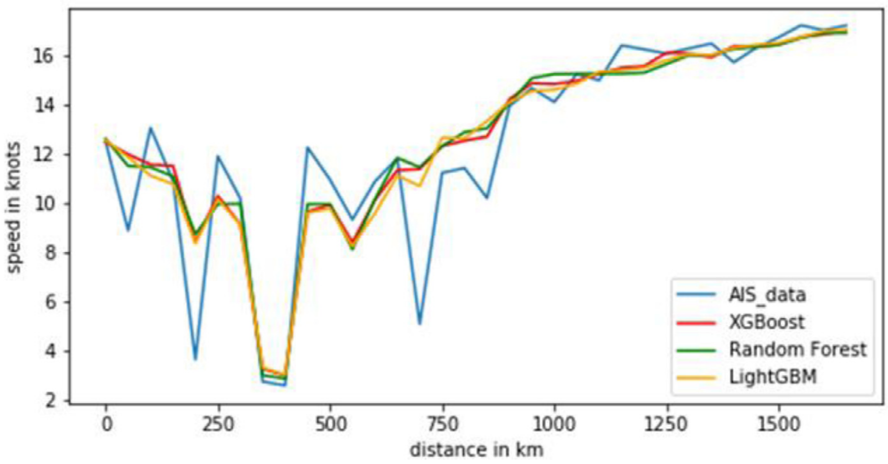

Fig. 13. Averaged best prediction of the models and recorded AIS speed. Averaging every 50-km.

neural networks can be used for predictions, however putting a neural network in production requires a dedicated interpretation algorithm, whereas a decision tree that can be converted to if-then-else statements and implemented in most computer languages.

In winter months, vessels can significantly deviate (in lateral direction) from the route in search of weak ice (see, e.g., Fig. 1, Route 2, on the path between the Kara Strait and Bely Island). Thus, the future data-based modelling should also focus on predicting these deviations in addition to speed predictions.

The presented discussion and the modeling approach are relevant to route planning (see e.g., Andersen et al., 2021), speed optimization and management (e.g., Fagerholt et al., 2010; Ng, 2019; von Westarp and Brabänder, 2021), and fleet deployment (see, e.g., Wetzel and Tierney, 2020) in ocean shipping.

7. Conclusions

The vessel speed is one of the important parameters that govern safety, emergency, and transport planning in the Arctic. In this study we have investigated a possibility of predicting transit speeds in ice using historical AIS data, without explicit knowledge of ice conditions around the vessel. It was assumed that the speed of the vessel will indirectly reflect the ice state and its dynamics as well as the ship's level of strengthening (i.e., ice class) and the operational mode (independent navigation or navigation in convoy).

First, we have analyzed AIS data from the Eastern Barents Sea and the Southern Kara Sea for the period of 2017–2018 (two full years), and then we trained three machine learning models with specific date-time encoding on the enhanced AIS data (extra information about daylight, the route, etc.). Then we evaluated the models' performance on the unseen data by comparing the model predictions with the ground-truth records. Next, we have exposed the models to new transit data from January 2019 to evaluate their predictive capability.

The main results of this study show that

(1) Supervised machine learning methods can be used for predicting vessel speeds (0 – 20 knots) with mean absolute error up to approximately 3.5 knots in average, without using complex ice field data. Majority of the errors are in the Kara Strait region and the Sabetta Channel.
(2) The LightGBM, and XGBoost models, built on seasonal time encoding with additional distance features, are a better modeling approach in a view of the considered dataset and modelling techniques.
(3) Seasonal time encoding (April–June and July–March) is shown to yield better predictions than categorical- or cyclic encoding.
(4) Visibility metrics such as daylight data and geographical information such as a distance from the predefined places can be engineered to enhance the predictive capabilities of the models.

We provide implementation of one the best models (XGBoost model for speed prediction) on GitHub. The presented machine learning approach to speed modelling could be considered as a source of additional and complimentary information for tactical planning (months ahead) of transits in the Arctic as well as to support development of new strategies that can dynamically change the speed of ships, the geography of the route, etc., in response to changes in ambient conditions. Future work should focus on supplementing existing AIS data with the data from the PAME's Arctic Ship Traffic Database, development and training of the models based on a continuously growing dataset, predicting ship's speeds together with deviations from the route as well as on coupling the data-driven approach with physics-based modelling and risks.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The computations were performed on resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway. We thank the Norwegian Coastal Administration for providing access to the AIS data.

Appendix

Figs. A1–A10.

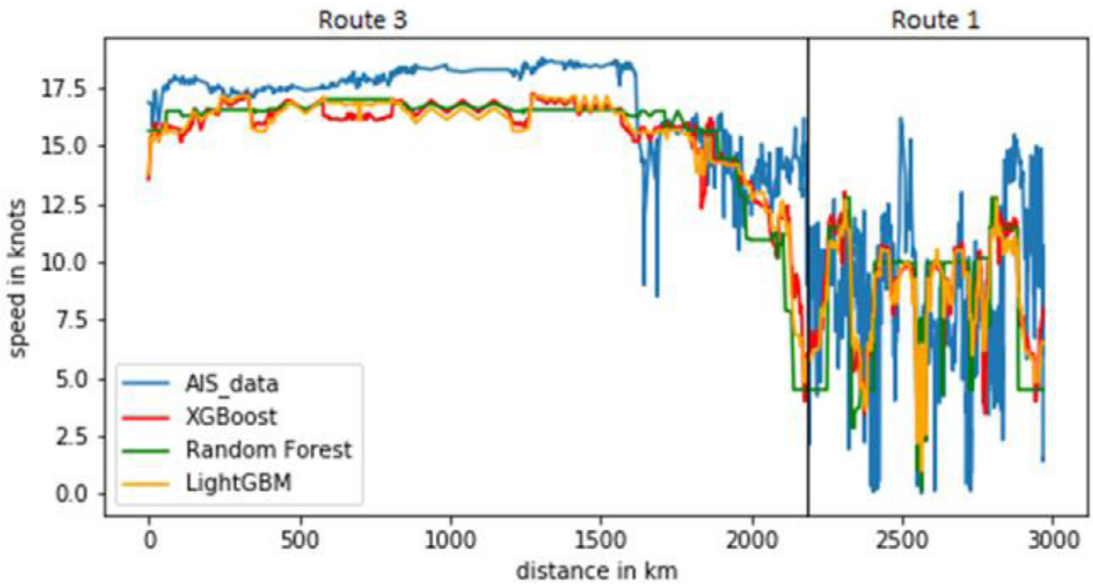

Fig. A1. Models' performance on a test dataset (from Test no. 1).

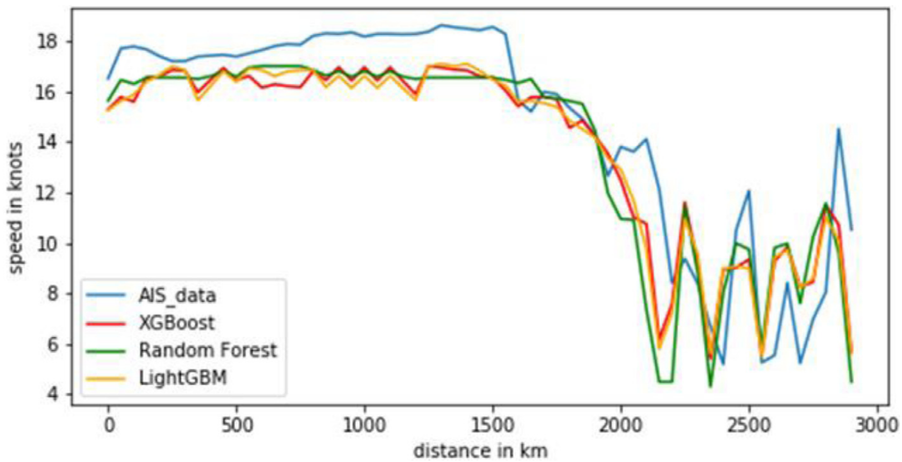

Fig. A2. Models' performance (50-km averaged) on a test dataset (Test no. 1).

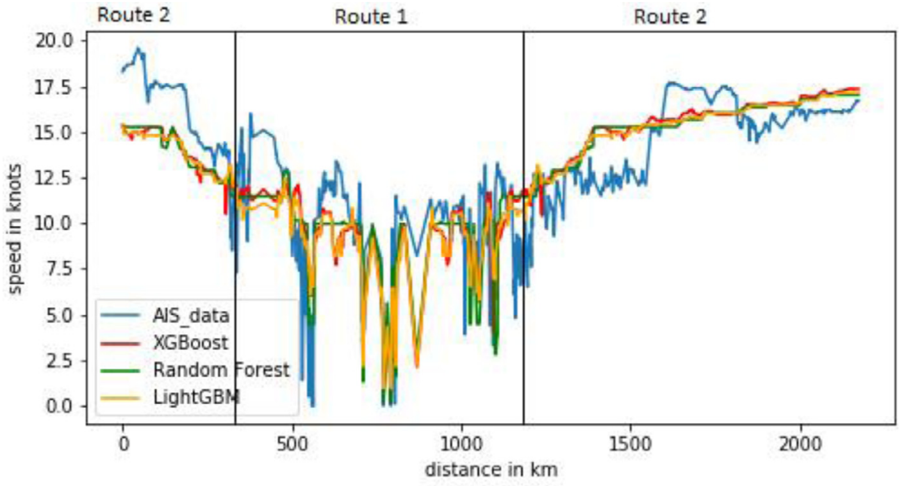

Fig. A3. Models' performance on a test dataset (Test no. 2).

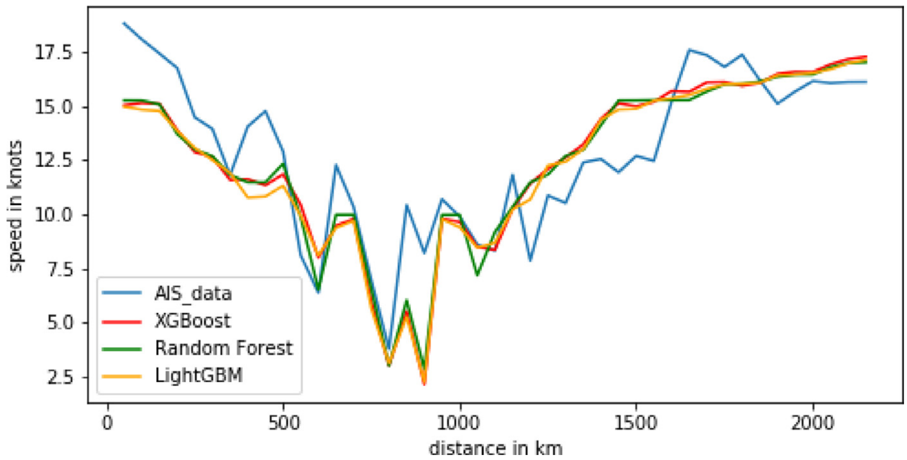

Fig. A4. Models' performance (50-km averaged) on a test dataset (Test no. 2).

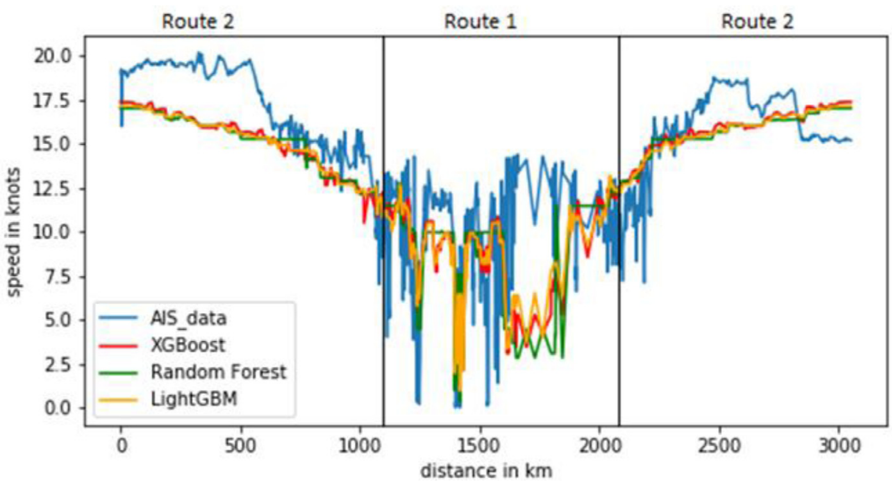

Fig. A5. Models' performance on a test dataset (from Test no. 3).

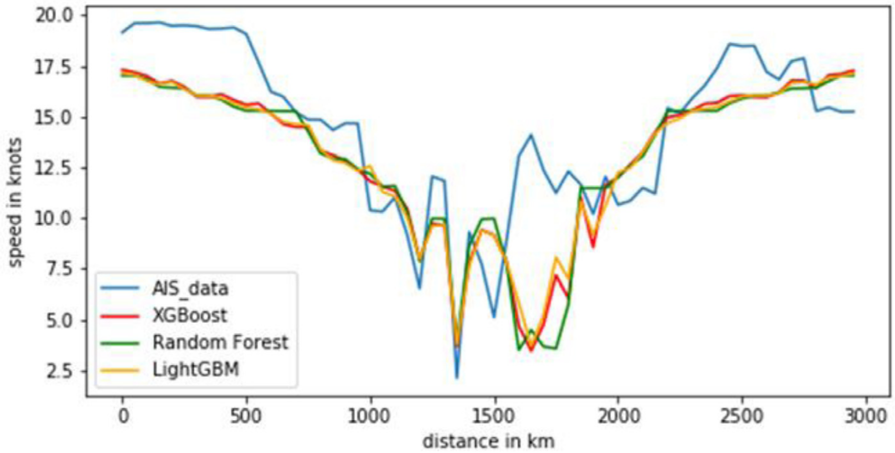

Fig. A6. Models' performance (50-km averaged) on a test dataset (from Test no. 3).

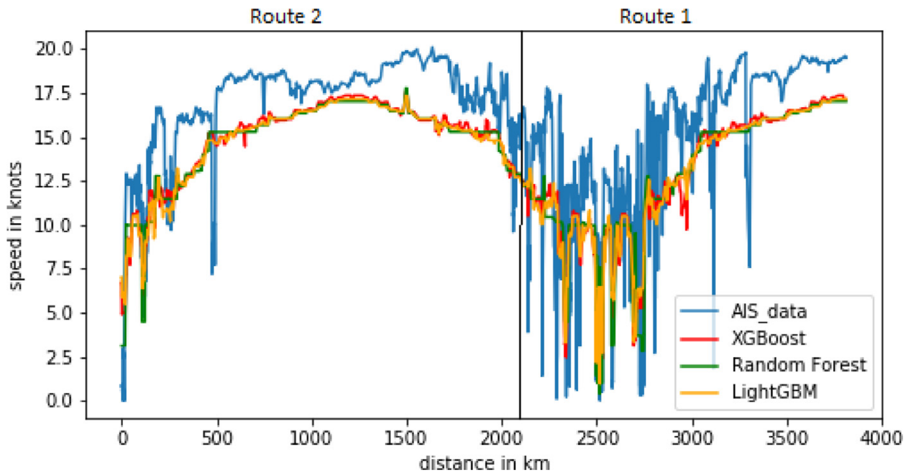

Fig. A7. Models' performance on a test dataset (Test no. 4).

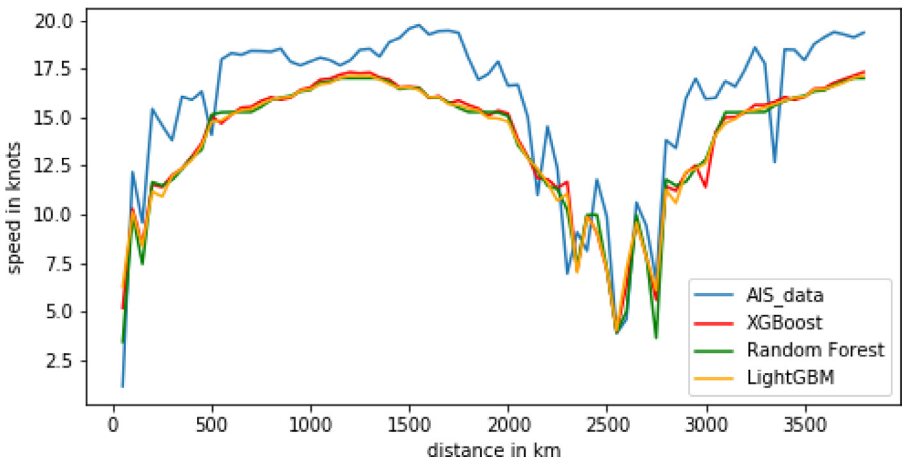

Fig. A8. Models' performance (50-km averaged) on a test dataset (Test no. 4).

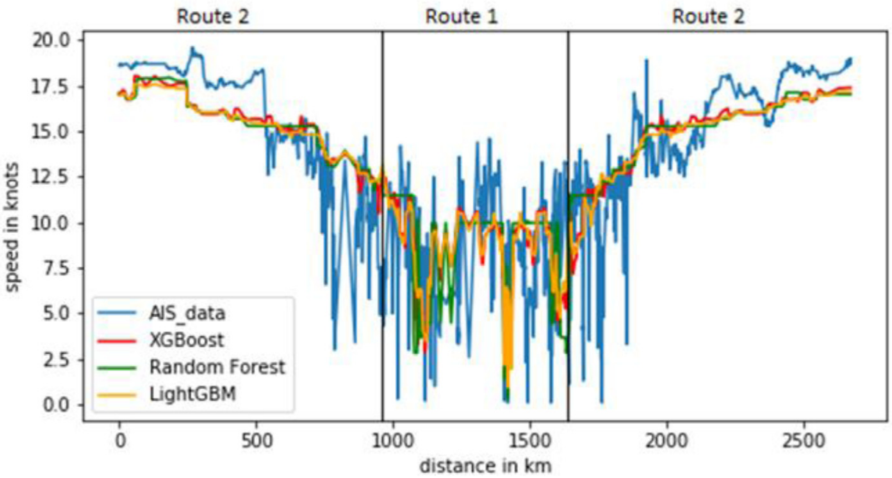

Fig. A9. Models' performance on a test dataset (from Test no. 5).

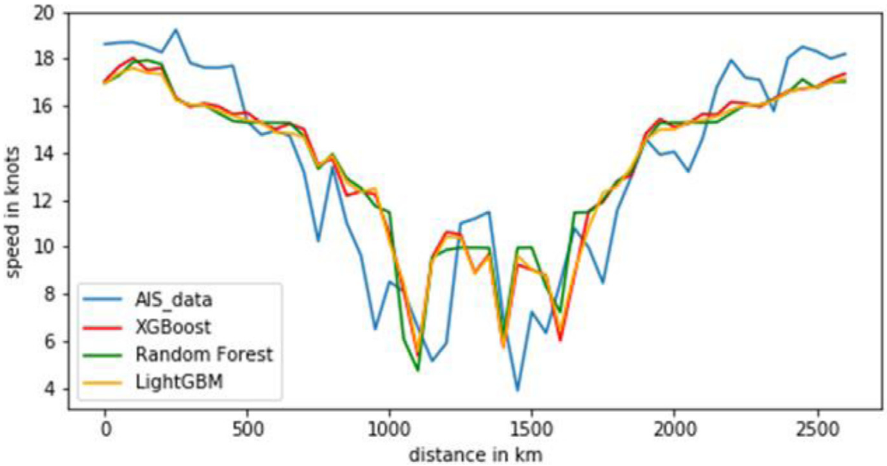

Fig. A10. Models' performance (50-km averaged) on a test dataset (from Test no. 5).

References

Adamovich, N.M., Buzuyev, A.Ya., Fediakov, V.E., 1995. The empiric model of vessel movement in ice and generalization of the experience of the model usage in hydrometeorological support of shipping in the Arctic. In: Proceedings of the POAC-1995, 2, pp. 30–41.

Afonin, A., Ol'Khovik, E., Tezikov, A., 2018. 2018. Study of ship speed regimes in the Arctic sea ice conditions. IOP Conf. Ser. Earth Environ. Sci. 194, 072012.

Andersen, T., Hove, J.H., Fagerholt, K., Meisel, F., 2021. Scheduling ships with uncertain arrival times through the Kiel Canal. Marit.Transp. Res. 2, 100008.

Bergström, M., Erikstad, S.O., Ehlers, S., 2016. A simulation-based probabilistic design method for Arctic sea transport systems. J. Mar. Sci. Appl. 15, 349–369.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth, Belmont.

Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

dateandtime.info, 2020. Retrieved June 3 from https://dateandtime.info/city.php?id=1538111.

Department of Navigation and Oceanography of the Ministry of Defense of the Russian Federation, 1995. Руководство для сквозного плавания судов по Северному морскому пути. – С.П-б: Изд. ГУНИО МО РФ, 415 с. (In Russian)

Dumanskaya, I.O., 2013. Типовые условия на основных судоходных трассах морей европейской части России для зим различной суровости. Тр. Гидрометцентра РФ. 2013350, 142–166 (In Russian).

Erceg, S., Ehlers, S., von Bock und Polach, R.U.F., Leira, B., 2014. A numerical model to initiate the icebreaking pattern in level ice. In: Proceedings of the ASME 2014 33rd International Conference on Ocean, Offshore and Arctic Engineering. OMAE2014-23409.

Fagerholt, K., Laporte, G., Norstad, I., 2010. Reducing fuel emissions by optimizing speed on shipping routes. J. Oper. Res. Soc. 61 (3), 523–529.

Frederking, R., 2003. A model for ship routing in ice. In: Proceedings of the 17th International Conference on Port and Ocean Engineering under Arctic Conditions, POAC '03, vol. 2, pp. 467–476.

Jeong, S.Y., Choi, K., Kang, K.J., Ha, J.S., 2017. Prediction of ship resistance in level ice based on empirical approach. Int. J. Nav. Arch. Ocean 9 (6), 613–623.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-.Y., 2017. LightGBM: a highly efficient gradient boosting decision tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Red Hook, NY, USA. Curran Associates Inc. 3149–31 57.

Kim, E., Smestad, B.B., Asbjørnslett, B.E., 2020. Predicting ship speeds in the Arctic using deep learning on historical AIS data. In: Proceedings of the 30th International Ocean and Polar Engineering Conference. Shanghai, China.

Kotovirta, V., Jalonen, R., Axell, L., Riska, K., Berglund, R., 2009. A system for route optimization in ice-covered waters. Cold Reg. Sci. Technol. 55 (1), 52–62.
Kuuliala, L., Kujala, P., Suominen, M., Montewka, J., 2017. Estimating operability of ships in ridged ice fields. Cold Reg. Sci. Technol. 135, 51–61.
Lau, M., 2006. Discrete element modeling of ship manoeuvring in ice. In: Proceedings of the 18th International Symposium on Ice (IAHR Ice Symposium 2006). Sapporo, Japan.
Lensu, M., 2015. Assessing the Ice Performance of Ships in Terms of AIS Data. Proceedings of the International Conference on Port and Ocean Engineering Under Arctic Conditions.
Li, F., Goerlandt, F., Kujala, P., Lehtiranta, J., Lensu, M., 2018. Evaluation of selected state-of-the-art methods for ship transit simulation in various ice conditions based on full-scale measurement. Cold Regions Science and Technology 151, 94–108.
Li, F., Kotilainen, M., Goerlandt, F., Kujala, P., 2019. An extended ice failure model to improve the fidelity of icebreaking pattern in numerical simulation of ship performance in level ice. Ocean Eng. 176, 169–183.
Lindqvist, G., 1989. A straightforward method for calculation of ice resistance of ships. In: Proceedings of the 10th International Conference on Port and Ocean Engineering under Arctic Conditions, POAC '89. Lulea, Sweden.
Louppe, G., 2014. Understanding Random Forests: From Theory to Practice. University of Liege PhD Thesis.
Löptien, U., Axell, L., 2014. Ice and AIS: ship speed data and sea ice forecasts in the Baltic Sea. Cryosphere 8, 2409–2418.
Lubbad, R., Løset, S., 2011. A numerical model for real-time simulation of ship–ice interaction. Cold Reg. Sci. Technol. 65 (2), 111–127.
Milaković, A.-S., Li, F., Marouf, M., Ehlers, S., 2020. A machine learning-based method for simulation of ship speed profile in a complex ice field. Ships and Offshore Structures 15 (9), 974–980.
Montewka, J., Goerlandt, F., Kujala, P., Lensu, M., 2015. Towards probabilistic models for the prediction of a ship performance in dynamic ice. Cold Reg. Sci. Technol. 112, 14–28.
Montewka, J., Goerlandt, F., Lensu, M., Kuuliala, L., Guinness, R., 2019. Toward a hybrid model of ship performance in ice suitable for route planning purpose. Proc. Inst. Mech. Eng. O J. Risk Reliab. 233 (1), 18–34.
Montewka, J., Sinclair, H., Kujala, P., Haapala, J., Lensu, M., 2013. Modelling ship performance in ice using Bayesian networks. In: Proceedings of the22nd International Conference on Port and Ocean Engineering under Arctic Conditions, POAC '13. Espoo, Finland.
Ng, M.W., 2019. Vessel speed optimization in container shipping: a new look. J. Oper. Res. Soc. 70 (4), 541–547.
Ol'khovik, E.O., 2018. Исследование влияния мелководья на изменение скоростных режимов судов в акватории Северного морского пути. Вестник Государственного университета морского и речного флота имени адмирала С. О. Макарова. 10 (3), 486–496. (In Russian)
Ol'khovik, E.O., 2019. Study of changes of vessel's speed in ice conditions on the Northern Sea route. In: Proceedings of the IOP Conference Series: Earth and Environmental Science.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. JMLR 12, 2825–2830.
Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1, 81–106.
Riska, K., Wilhelmson, M., Englund, K., Leiviskä, T., 1997. Performance of Merchant Vessels in Ice in the Baltic. Winter Navigation Research Board, Helsinki Research report no 52.
Sawamura, J., Riska, K., Moan, T., 2008. Finite element analysis of fluid–ice interaction during ice bending. In: Proceedings of the IAHR 2008, pp. 191–202.
SCF Newsletter, 2019. Корпоративное издание группы СКФ December 05 (65). (In Russian)
Shukurov, K.A., Semenov, V.A., 2018. A link between sea ice concentration in Kara Sea in November and large-scale atmospheric circulation. In: Proceedings of the XXIV International Symposium, Atmospheric and Ocean Optics, Atmospheric Physics, Tomsk, Russia, p. v10833.
Shumovsky, S.A. 2012. О диких проблемах с которыми сталкнется ЯМАЛ-СПГ при вывозе продукции. Профессионально об энергетике. (In Russian)
Similä, M., Lensu, M., 2018. Estimating the speed of ice-going ships by integrating SAR imagery and ship data from an Automatic identification system. Remote Sens. 10, 1132.
Su, B., Riska, K., Moan, T., 2010. A numerical method for the prediction of ship performance in level ice. Cold Reg. Sci. Technol. 60 (3), 177–188.
Topaj, A.G., Tarovik, O.V., Bakharev, A.A., Kondratenko, A.A., 2019. Optimal ice routing of a ship with icebreaker assistance. Applied Ocean Research 86, 177–187.
Valanto, P., 2001. The resistance of ships in level ice. SNAME Trans. 109, 53–83.
Valkonen, J., Løvoll, G., Eide, M.S., Walter, E.L., 2013. COSSARC – concept selection for shipping in the arctic. In: Proceedings of the ASME 2013 32nd International Conference on Ocean, Offshore and Arctic Engineering OMAE2013-10063.
von Westarp, A.G., Brabänder, C., 2021. Support of the speed decision in liner operation by evaluating the trade-off between bunker fuel consumption and reliability. Marit. Transp. Res. 2, 100009.
Wang, S., 2001. A Dynamic Model For Breaking Pattern of Level Ice By Conical Structures. Helsinki University of Technology [doctoral thesis].
Wetzel, D., Tierney, K., 2020. Integrating fleet deployment into liner shipping vessel repositioning. Transp. Res. Part E Logist. Transp. Rev. 143, 102101.
Zhou, Q., Peng, H., Qiu, W., 2016. Numerical investigations of ship–ice interaction and maneuvering performance in level ice. Cold Reg. Sci. Technol. 122, 36–49.
Zivot, E., Wang, J., 2006. Vector Autoregressive Models for Multivariate Time Series. In: Modelling Financial Time Series with S-PLUS. Springer Science, pp. 385–429.