

Stochastic sea ice generator for Monte Carlo modelling of navigation conditions at the Northern Sea Route

Ruslan I. May², Roman B. Guzenko¹, Oleg V. Tarovik², Alex G. Topaj³, Alexander V. Yulin¹ ¹Arctic and Antarctic Research Institute, St. Petersburg, Russia ²Krylov State Research Centre, St. Petersburg, Russia ³LLC Bureau Hyperborea, St. Petersburg, Russia

ABSTRACT

In this study, we applied a stochastic sea ice generator and automatic ice routing to investigate navigation conditions in the Arctic. The stochastic generator provides long (over 100 years) interconnected spatial-temporal fields of sea ice parameters that were used as a source of information to calculate navigation routes using the optimal ice routing tool. A stochastic sea ice generator is a probabilistic model that reproduces synthetic sea ice data. Temporal connectivity of generated data is provided by Markov chains, while spatial connectivity is achieved using empirical random probability fields. To determine the matrices of transition probabilities for Markov chains and empirical random probability fields of sea ice concentration, we used satellite data from the OSI SAF project for the last 20 years. The grid area of the probabilistic model covers the Barents Sea, Kara Sea, and nearby waters with a spatial resolution of 10 km. Time step of the model is one day. We tested the stochastic generator by comparing its results with statistical characteristics of ice concentration obtained from satellite data of the OSI SAF project. The mean absolute error of the statistical characteristics of ice concentration was found to be less than 5%. As an example of the practical use of the stochastic generator, we used ice routing tool to calculate the probability of navigation start at the line from Murmansk to Pobeda field (Kara Sea) on different calendar dates under the assumption that the vessel can operate in open water or very open ice only. The average and extremely early and late dates were estimated.

KEYWORDS: Stochastic ice generator; Monte Carlo method; Ice routing; Arctic navigation

INTRODUCTION

Ice cover determines any human activity in the Arctic and freezing seas. In a number of engineering problems, it is necessary to evaluate the statistical characteristics of not the ice parameters themselves but the processes and phenomena functionally related to the characteristics of ice cover. The examples of the latter ones are the parameters of vessel movement in ice (speed, time, and travelled distance), probability of emergencies, the time of the beginning and end of navigation, parameters of ice load on offshore structures. The use of Monte Carlo methods may be the only way to solve some of these engineering problems. Moreover, the required length of realization of calculations by the Monte Carlo method can

significantly exceed the duration of the available observation series. In this case, the results of probabilistic models can be claimed as the only source of input data on natural parameters.

The Monte Carlo method was used to assess the probability of accidents along the routes of Arctic seas in the studies of Tretyakov&Frolov (2009), Tretyakov et al. (2019), and Bergström et al. (2016). There are also several studies where the Monte Carlo method is applied to model the distribution of ice parameters along the routes of vessel operation (see (Tretyakov & Frolov, 2013) and (Tretyakov et al., 2019)). However, such a linear approach assumes the change of ice parameters along the predefined routes and does not consider that the actual vessel route may vary depending on local ice conditions. E.g., the ship route from the Barents Sea to the Kara Sea can pass through both the north of Cape Zhelaniya, and the Kara Strait depending on ice conditions. At the same time, even the chosen global variant of a path means the possibility of its local tactical variability due to the local peculiarities of ice distribution. Therefore, the use of information on probabilistic characteristics of ice along strictly predefined routes can lead to incorrect results.

We believe that to assess the navigation conditions, it is necessary to consider the space-time fields of ice characteristics. The problem is that such information is limited to several decades. For this reason, we suggest using the results of probabilistic models of spatio-temporal variability of ice cover as an alternative source of information to model Arctic navigation during Monte Carlo experiments. Such probabilistic models that simulate meteorological characteristics are usually called stochastic weather generators (Richardson, 1981). Similarly, we named the probabilistic model of sea ice dynamics as a "stochastic sea ice generator". The stochastic sea ice generator can provide synthetic information for models to simulate Arctic navigation and calculate climatic characteristics of navigation conditions. It also can be used as a source of information for typification of ice conditions, calculating fleet operations, testing methods for automatic ice routing of a ship, and used to simulate synoptic spatial and temporal variability of ice characteristics under various hypothetical climatic scenarios of ice cover dynamics.

In this article, we describe the developed stochastic sea ice generator and its application to estimate the time of navigation start by means of Monte Carlo experiments. The latter task was solved with the use of a previously developed ice routing tool that we applied to find optimal ice routes in generated spatial ice fields. As the criterion of navigation start, we adopted the case when a ship moves through open water or very open ice, which concentration is less than 30%.

STOCHASTIC SEA ICE GENERATOR

Concentration is one of the most important parameters of ice cover that characterize the degree of ice coverage of a particular part of the water area. Obviously, it is necessary to start developing the stochastic generator from this parameter, supplementing it with probabilistic models of other ice characteristics (ice thickness/age, horizontal size of floes, ridging, compression, etc.).

To develop a probabilistic model we chose a free database OSI-409 v.1.2 from the study (EUMETSAT, 2019). This data was obtained from satellite information and covers the period from 10/25/1978 to 12/31/2019. However, daily time series in this database are available only since July 1987. The values of ice concentration are placed in a rectangular grid in a polar stereographic projection, with a spatial step of 10 km and a time step of one day. Data is available at ftp://osisaf.met.no/reprocessed/ice/.

Traditionally, ice concentration is rated from 0 to 100% with a step of 10%. Therefore, the time series of ice concentration changes can be represented as a discrete Markov chain, which matrix of transition probabilities has a size of 11×11 :

$$p_{(C_{t+1}|C_t)} = \begin{pmatrix} p_{1,1} & \cdots & p_{1,11} \\ \vdots & \ddots & \vdots \\ p_{11,1} & \cdots & p_{11,11} \end{pmatrix}$$
(1)

Each row of the matrix (1) contains the probabilities of the transition of a certain gradation (equal to the row number) of initial concentration C_t to different gradations of concentration at the next time step C_{t+1} . Thus, the conditional probabilities contained in the rows of this matrix can be interpreted as the values of the corresponding discrete probability density function $f_{(C_{t+1}|C_t)}$. By means of integrating (series summing), it is easy to obtain the values of the conditional cumulative probability distribution function $F_{(C_{t+1}|C_t)}$:

$$F_{(C_{t+1}|C_t)} = \sum_{i=0}^{C_{t+1}} p_{(i|C_t)}$$
(2)

To ensure the coherence of simulated values of ice concentration, both in time and space, we implemented in the stochastic sea ice generator an algorithm, which can be described by the equation:

$$C_{t+1,x,y} = F_{(C_{t+1,x,y}|C_{t,x,y})}^{-1} (P_{t,x,y}),$$
(3)

where $C_{t+1,x,y}$ is the concentration field in grid coordinates (x, y) at the next time step. In equation (3), in each grid cell, a spatially connected probability field is specified $(P_{t,x,y})$, which is defined for each grid node.

According to the OSI-409 database, empirical distribution functions of conditional probabilities $F_{(C_{t+1}|C_t)}$ for 11 gradations of concentration were calculated in each cell of the grid area. OSI-409 initial data accuracy is the 0.1%. To estimate the initial conditions C_t , we round the data to integers and obtained 11 gradations. We use the concentration C_{t+1} values following the initial conditions in their original form (without rounding to integer). In this case, for the convenience of further calculations, the empirical distribution function is written in a form of quantiles for a given predefined discrete set of probabilities $P \in [0, \Delta, 2\Delta, 3\Delta, ... 1]$:

$$C_{t+1}(C_t = 0, P = 0) \quad \cdots \quad C_{t+1}(C_t = 0, P = 1) , \ C_t = 0 \\ \vdots \\ C_{t+1}(C_t = 11, P = 0) \quad \cdots \quad C_{t+1}(C_t = 11, P = 1) , \ C_t = 11 \} = F_{(C_{t+1}|C_t)}^{-1}(P)$$
(4)

The results presented in this paper were calculated by quantiles calculated from probabilities with a step of $\Delta = 0.05$. Thus, in each cell of the spatial computational grid, information of size 11 × 21 is stored (11 gradations of the initial concentration C_t , for each of which 21 quantiles of the subsequent concentration C_{t+1} are given). This representation allows to concisely store information about the empirical cumulative probability distribution function. At the same time, the excessive accuracy of the concentration of the OSI-405 project (0.1%) makes it possible to reduce to a minimum the number of identical quantile values falling within the range of probability values.

Using the described representation of matrices of transition probabilities, it is easy to calculate the concentration for the next time step C_{t+1} as a result of two-dimensional interpolation of values of the matrix (3) from the known initial concentration (C_t) and the current probability (P). At the same time, due to interpolation, the new value can take an

intermediate value between the gradations (10%) of concentration, i.e. the concentration is calculated up to several digits after the dot.

Ice cover is characterized by seasonal variability. To consider it in the stochastic generator, conditional cumulative probability distribution functions were calculated over a time series consisting of segments with the duration of w = 31 days separated from each other by one year (the total length of the series is $w \times N_Y$, where N_Y is the number of years in initial data). Since the ice regime of Arctic seas in recent decades differs significantly from past periods, only the data for 2000 - 2019 was used to model the seasonal ice dynamics, i.e. $N_Y = 20$ years. This conditionally corresponds to the current state of the climate. To eliminate the uncertainties associated with time interpolation of the matrices of transition probabilities, we used moving analysis with a step of 1 day. Thus, the conditional cumulative probability distribution functions $F_{(C_{t+1}|C_t)}$ are calculated for each day of the year and reflect the changes in concentration during the season, which falls on that day of the year. The index *t* in equation (3) now also denotes the seasonal modulation of the parameters of the equation.

On the one hand, the probability fields $P_{t,x,y}$ in equation (3) must be uniform distributed in time to provide a realistic reproduction of a random variable. On the other hand, the spatial correlation characteristics of the probability fields $P_{t,x,y}$ should provide realistic sizes, shape, and orientation of the elements of the fields of the modeled parameter $C_{t+1,x,y}$. The assumption that the spatial characteristics of the probability fields $P_{t,x,y}$ correspond to the spatial characteristics of concentration fields themselves $C_{t,x,y}$ is not substantiated. Moreover, as it was shown in (Wilks, 1998), under such an approach the spatial correlations of simulated quantities will differ from the spatial correlations of measured parameters.

To solve this problem, we propose to estimate the empirical probability fields $P_{t,x,y}^*$ that determine the observed variability of the concentration over time for a known conditional cumulative probability distribution function $F_{(C_{t+1}|C_t)}$. Equation (3) can be transformed in such a way to perform a reverse calculation, i.e to estimate the only possible value of the probability $P_{t,x,y}^*$ at which the concentration $C_{t,x,y}$ transits to $C_{t+1,x,y}$. It can be done by substituting the values of the concentration at two adjacent time steps ($C_{t,x,y}$ and $C_{t+1,x,y}$) to the conditional distribution function $F_{(C_{t+1}x,y|C_{t,x,y})}$:

$$P_{t,x,y}^* = F_{(C_{t+1,x,y}|C_{t,x,y})} \Big(C_{t,x,y}, C_{t+1,x,y} \Big)$$
(5)

Based on the obtained array of empirical probability fields, $P_{t,x,y}^*$ it is possible to find the probability fields $P_{t,x,y} = f(P_{t,x,y}^*)$, which can be used in equation (3) for stochastic modeling of concentration fields. There is a large number of algorithms in applied geostatistics to model random fields with given characteristics of the spatial correlation of values in closely spaced points or areas. To construct them there used the approaches of moving average, separation in local subdomains by average value, Cholesky's decomposition of the full covariance matrix, discrete Fourier transformation, calculation of spatial correlation radii of parameter changes in time, etc. Many of these approaches are included in standard statistical packages and software tools (Shlachter et al., 2015).

When modeling hypothetical situations, e.g, ice dynamics under possible scenarios of climate change, it is necessary to specify both a model of changes in conditional cumulative probability distribution functions $F_{(C_{t+1,x,y}|C_{t,x,y})}$, and a model of probability fields $P_{t,x,y}$. Models of fields of random probabilities will always contain errors caused by one or another approximation method. When calculating the dynamics of ice cover in the current climate

state, it is possible to replace the model of probability fields $P_{t,x,y}$ with an array of empirical probability fields $P_{t,x,y}^*$. In this case, it is necessary to determine the limitations on the duration of stochastic generator realization (run); they should be determined depending on the number of empirical probability fields.

Obviously, the direct sequence of empirical probability fields $(P_{t,x,y} = P_{t,x,y}^*)$, calculated from observational data, makes it possible to reproduce the results of the stochastic generator for a period of N_Y years. The time sweep of probability fields $P_{t,x,y}^*$ is uniformly distributed. Consequently, to increase the length of stochastic generator realization, the sequence of fields $P_{t,x,y}^*$ can be swapped within one season. With a random selection of the empirical probability field from the set $P_{t,x,y} = P_{t+[R_1 \times N_Y \times N_D]x,y}^*$, the number of non-repeating combinations of annual circles will be equal to $N_Y^{N_D}$, where N_Y is the number of years, N_D is the number of days in a year. At $N_Y = 20$, the maximum generator realization length will be limited to 10^{472} years. However, the matrices of transition probabilities at each point were calculated over a series of lengths $w \times N_Y$. Therefore, the number of combinations of a sequence of empirical probability fields for each day of the year will be equal to $(w \times N_Y)^{N_D}$. And accordingly, the maximum length of generator realization will be equal to 10^{780} years at w = 7 days, or 10^{901} years at w = 15 days, or 10^{1016} years at w = 31 days. Such a length is quite enough to estimate navigation parameters by the Monte Carlo method.

The choice of one probability field for the moment of time *t* from the set $P_{t,x,y}^*$ is performed according to the formula:

$$P_{t,x,y} = P_{[R_1 \times N_Y \times N_D] + [U^{-1}(R_2, \mu, \sigma)], x, y}^{*},$$
(6)

where $U^{-1}(R_2, \mu, \sigma)$ is the normal inverse cumulative distribution function with mean $\mu = t$ and standard deviation, $\sigma = w/3$; R_1 and R_2 are the random variables generated by the random number generator; [] is the rounding to integer operation; N_Y is the number of years in the series; N_D is the number of days in a year; t is the number of a day in a year. In formula (6), the normal distribution law is selected to preserve seasonal variability, which is present in the series of empirical probability fields $P_{t,x,y}^*$.

The realization of a generator begins with a random field of initial concentration. The first year of model time is given to reach a stationary solution. Starting from the second year, three-dimensional arrays of concentration are recorded for subsequent analysis and use in the simulation of navigation conditions in the Arctic during Monte Carlo experiments. Fig. 1 shows an example of concentration fields simulated by the stochastic generator for June 5 of various model years.

The main requirement for the results of stochastic modeling is the coincidence of statistical parameters of modelled values and observed natural ones. We verified the stochastic generator by comparing the statistical characteristics of generated concentration fields with similar parameters of satellite observation data. Figure 2 shows the mean field of concentration and standard deviation (STD) of concentration for March. To quantify the similarity of the mean and standard deviation, we use the mean absolute error (MAE):

$$MAE = \frac{\sum |c_D - c_G|}{N} \tag{7}$$

where C_D is the field of characteristics calculated from satellite data; C_G is the field of characteristics calculated from generated data; N is the number of compared pairs of cells in the fields C_D and C_G .



Figure 1. Examples of concentration fields on June 5th generated by a stochastic generator.

Table 1 shows the values of the mean absolute error for the mean concentration value and standard deviation of concentration. As it can be seen, the average absolute deviation between the fields of the mean value and STD compared with the original satellite data is less than 5%. The maximum MAE values for the fields of average concentration are observed in September and October (5.0% for STD and 4.3% for mean values).

Table 1.	Comparison	of the	statistical	characteristics	of	concentration	fields	obtained	from
satellite	data (OSI-409) projec	et) and stoc	chastic generato	or.				

MAE 0/	Month											
MAE, %	Ι	II	III	IV	V	VI	VII	VIII	IX	Х	XI	XII
Mean values	1.7%	1.6%	1.4%	1.5%	1.6%	2.1%	1.9%	2.2%	2.6%	4.3%	3.3%	2.3%
STD	2.5%	2.4%	1.7%	1.9%	2.3%	2.6%	2.1%	3.9%	5.0%	3.6%	3.3%	3.1%



Figure 2. Mean value (top) and standard deviation (bottom) of the changes in the concentration for March based on the results of calculating the stochastic generator (left) and data from the OSI-409 project (right) for 2000-2019.

MONTE CARLO SIMULATION OF ARCTIC NAVIGATION PARAMETERS

Under the navigation parameters, we understand all the parameters that characterize vessel voyage, such as ice parameters on a route, vessel speed, voyage time, and others. To estimate navigation parameters based on a set of generated concentration fields, we used the method of optimal ice routing described in (May, et al, 2020). This is an isochrone method that is based on polygon operations and does not require a grid area or graph system. Isochrones are modelled as the closed polygonal objects that best convey the physical meaning of the isochrone. The vertices of isochrone polygon can be interpreted as the coordinates of possible ship location that are farthest from the starting route point. The coordinates of the polygon vertices are used to determine ice zones and vessel speeds. Depending on the speed, new polygons of possible movement of ships are calculated from each vertex of the polygon. The combination of these polygons gives a new isochronous polygon for the next model time

step. The described procedure is repeated until the end point or several end points of the path get inside the polygon of maximum displacement. After that, according to the saved isochrones, the optimal route of the vessel movement is calculated in reverse order, as a line that connects the closest vertices of isochronous polygons neighboring in time (Fig. 3). The algorithm considers the change in ice parameters over time, small-scale spatial discontinuities in ice cover, as well as calculates vessel movement along narrow channels, polynyas, fractures, and ice leads. The grid-free concept of the algorithm is best suited for the use of standard electronic ice charts in the SIGRID-3 format.



Figure 3. Schematic diagram of the isochrone method for optimal route calculation.

The stochastic sea ice generator described in this paper simulates only the fields of ice concentration, while other ice parameters are out of consideration. Therefore, we will try to illustrate the prospects of the proposed approach for navigation problems that depend only on a presence or absence of ice. Examples of such a problem are the passage of a non-ice class vessel through ice-covered waters, towing a drilling rig, building routes for seismic ships, etc. Let us assume that the speed of a vessel depends only on total ice concentration. The speed in open water is maximal and exponentially decreases to zero when the concentration reaches 40% or more, i.e., a vessel can operate in very open ice only. Such speed function will lead to bypassing ice floes encountered on the way. If we know the spatial distribution of vessel speed, we can use the route optimization methods to calculate the optimal path between two points. After that, we will be able to estimate the statistical characteristics of navigation parameters using the calculated optimal route. Multiple runs of such an algorithm make it possible to obtain estimates of navigation parameters during Monte Carlo experiments where the stochastic generator provides the input information for ice routing tools.

The fields of total concentration simulated by the stochastic generator were converted into vector polygons as the isolines of the same concentration with the discreteness of 5%. We chose the start point of a route in a non-freezing part of the Barents Sea (Murmansk), while

the end point is located in the area of Pobeda field in a southwestern part of the Kara Sea (73.996524°N, 66.751593°E) (see Fig. 4). The optimal route was calculated for each calendar day, i.e. in this task, we neglect the duration of the vessel voyage. Also, we did not consider the ice performance of the particular vessel and assess only the possibility of reaching the final point of a route through the open water or very open ice (concentration less than 30% inclusively). The date when the ship becomes able to pass the given route during a year circle is called the date of navigation start. The probability of navigation start on a certain calendar day is the ratio of several successful passes on a given day to the total number of simulated years.



Figure 4. Examples of the optimal routes between Murmansk and the Pobeda field, built using synthetic information generated by the stochastic ice generator.

Fig. 5 shows the probabilities of navigation start at a given sailing line. A broken red line in this figure is the cumulative distribution function of the probability of navigation start calculated over 20 years. With an increase in the length of stochastic realization up to 100 years, the probability distribution function takes a smoother form (blue line).

Calculation of navigation start dates between the Barents Sea and the Pobeda field showed the following results. On average, the way to the field opens on 3 July. The earliest date of passage along the specified route is June 6, the latest date of the beginning of navigation in the specified conditions is July 20 (the indicated dates have a probability of 1 time in a hundred years). The interquartile range of navigation start is limited to the dates of June 25 and July 9, i.e. in 50% of cases, the beginning of navigation should be expected in this interval. By setting the length of stochastic realization, one can directly obtain the probabilities of extremely rare events (once every 100 years, once every 1000 years, etc.), without the unreasonable use of theoretical distribution functions.

It should be emphasized once again that the obtained results consider the ice parameters not at some particular point, but in a whole computational domain, i.e., the entire Kara and Barents Sea with the adjacent waters of the Arctic Ocean. E.g., if most of the water area is free of ice, but the Cape Zhelaniya and the Kara Strait are blocked by narrow tongues of ice, then the algorithm will show that open-water navigation is impossible in such conditions.



Figure 5. The probability of navigation start at the line from Murmansk to Pobeda field on different calendar dates, calculated using OSI-409 data and realizations of a stochastic generator for 20 and 100 years.

CONCLUSIONS

The results obtained in this study indicate that the stochastic probabilistic models of spatiotemporal variability of ice cover can be used to assess the parameters of navigation in the Arctic and to plan marine transport operations.

Our experience shows that to simulate the temporal connectivity of ice cover parameters, it is possible to use a probabilistic model based on Markov chains. At the same time, to provide spatial correlations, we propose to use empirical probability fields, which can be estimated from the satellite data and calculated matrices of transition probabilities. Moreover, we showed that it is not necessary to create a model of random fields to obtain feasible synthetic information, but it is possible to use random sequences of calculated empirical probability fields. Comparison of the results of stochastic modeling showed that the field-average absolute error of statistical characteristics (mean value and STD) of concentration is less than 5%.

The probabilistic model of ice cover described in the paper can be used for the short-term predictive calculation of an ensemble of all possible concentration states with an estimate of the probability of each state onset. However, the main task of the stochastic generator is to simulate time-unlimited synthetic fields of ice cover characteristics, which can be used for Monte Carlo estimates of ice-dependent parameters. The combination of a stochastic ice cover generator and ice routing algorithm makes it possible to estimate navigation parameters that depend on the ice cover.

The dates of navigation start at the line between Murmansk and Pobeda field calculated by the Monte Carlo method showed that navigation along this route in very open ice begins in late June - early July on average. Besides, the developed method may allow us to evaluate the statistical characteristics (including extreme values) of ship speed, route length, probability of passage, navigation costs, etc. However, in this case, the stochastic generator should take into account not only the total concentration but also other parameters that affect the movement of a vessel. They are the age gradation of ice, partial concentrations, size of ice floes, ridging, melting stage, snow thickness, and others. We treat these factors as a subject for further research and development of a comprehensive stochastic sea ice generator.

ACKNOWLEDGEMENTS

This study was supported by the Russian Science Foundation (grant for Project No. 17-79-20162- Π).

REFERENCES

Bergström, M., Erikstad, S.O. & Ehlers, S., 2016. A simulation-based probabilistic design method for arctic sea transport systems. *Journal of Marine Science and Application*, 15, pp. 349–369.

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre, 2019. Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration Climate Data Record. OSI-450, OSI-430-b. Ver. 1.2.: 33 p.

May, R.I., Tarovik, O.V., Topaj, A.G., Fedyakov, V.E. & Frolov, S.V., 2020. Method for finding the optimal ship route in ice based on vector geo-algorithms. *International Journal of Offshore and Polar Engineering.*, 30(1), pp 78-85.

Richardson, C.W., 1981. Stochastic simulation of daily precipitation, temperature and solar radiation. *Water Resources Research*, 17, pp.182-190.

Schlather, M., Malinowski, A., Menck, P.J., Oesting, M. & Strokorb, K., 2015. Analysis, simulation and prediction of multivariate random fields with package Random Fields. *Journal of Statistical Software.*, 63 (8), pp. 1-25.

Tretyakov, V.Yu., Frolov, S.V. & Sarafanov M.I., 2019. The results of computer simulation of the probability of accidents due to ship compressions by drifting ice along the Northern Sea Route. *Russian Arctic*, 5, pp.4-10.

Tretyakov, V.Yu., Frolov, S.V. 2009. Model of an estimation of risk of navigations of vessels in ices. *The Meteorological Bulletin*, 4(5), pp.1-36. (in Russian).

Wilks, D., 1998. Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, 210(1), pp. 178-191.