Predicting catch of Giant Pacific Octopus *Enteroctopus dofleini* in the Tsugaru Strait, using a machine learning approach

Abstract

An attempt has been made to forecast catches of giant Pacific octopus *Enteroctopus dofleini* using past catch data by using the machine learning approach. Algorithms tested were Seasonal Autoregressive Integrated Moving Average (SARIMA) and Recurrent Neural Network (RNN) latter of which were represented by Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). Firstly, the catch of giant pacific octopus on the southern coast of the Tsugaru Strait was backcasted for each of the 2001–2019 years, using the data for at least 20 years out of 1981 - 2018 for training. Because of 19 trials of backcasting, SARIMA obtained the best agreement with actual catch data (R^2 = 0.849) followed by LSTM and GRU (R^2 = 0.847 and 0.841, respectively) The Root Mean Squared Error (RMSE) was within a narrow range for the three models (27.9, 28.2 and 28.5 metric ton (t) for LSTM, GRU and SARIMA, respectively).The behavior of the SARSIMA model was rather unstable when the performance was assessed by Mean Absolute Percentage Error (MAPE) which is standardized by average catch of each year; it increased substantially after 2010 when the winter catch decreased, whereas LSTM and GRU remained relatively stable. Although the behavior of LSTM and GRU were more stable than the SARIMA model, they were difficult to interpret due to their black box like structures. In contrast, the SARIMA model was advantageous in interpreting the results. Thus, each model has advantages and disadvantages, so it is necessary to choose the best model depending on its purpose.

Keywords: catch forecast, machine learning, SARIMA, LSTM, GRU

Abbreviations: giant pacific octopus (GPO)

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

1. INTRODUCTION

Fishery data often show periodicity. Therefore, it has been expected that future catch can be forecasted using the time series data of past catches. Machine learning methods, such as the Seasonal Autoregressive Integrated Moving Average (SARIMA) model and Recurrent Neural Network (RNN) model, have been used to analyze time series data, and are useful for estimating the future (e.g.,~~-~~Kim~~,~~ et al. 2015; Almanjahie, et al. 2019; Farmer and Froeschke, 2015). Machine learning methods take accounts of non-linear components of time series data as well as linear ones and have the capability of processing complex and huge data. Recently, thanks to the improvements of personal computers, machine learning methods such as deep learning have become easier, and the computer software became available free or at an affordable price One of the characteristics of deep learning is that its accuracy improves as the number of accumulated data increases, which makes it suitable for analyzing fishery catch data. In the Korean waters, the autoregressive recurrent neural network (ARNN) and SARIMA models were used to forecast the abundance (*i.e.* CPUE) of Japanese anchovy based on data from previous 25 (Kim, et al., 2015). Whereas ARNN was effective in the training phase with impressive performance for short to medium term forecasting, SARIMA's overall forecasting ability was also satisfying (Kim, et al., 2015). Thus, machine learning with time series data is expected to be an effective tool for fishing catch forecasts.

In 2017, the maximally sustainably fished stocks and overfished stocks accounted for 93.8% of the total number of assessed stocks globally (FAO, 2020), making resource management increasingly important. Fishery catch data at the regional scale are essential for the purpose. In Japan, some local governments collect monthly fishery data like catches in their area, and Aomori Prefecture is one of them. Catches of more than 50 species of fishery resources including fishes, crustaceans, mollusks, seaweeds have been collected and published by Aomori Prefecture since 1981 (https://opendata.pref.aomori.lg.jp/ , accessed on January 4, 2022.). If time series data like this can be utilized for forecasting catch trends, it will be useful for fishing planning and management.

Giant Pacific octopus *Enteroctopus dofleini* (GPO) are widely distributed in the Subarctic region of the North Pacific Ocean, ranging from Northern California to Alaska and Northern Japan (FAO, 2014; Wülker, 1910). GPO is a commercially and culturally important species in the coastal water of northern Japan. For example, the caught value of octopus was the more than 609 million yen in 2019 in Aomori Prefecture, so it is providing a source of winter income for fishers. Moreover, vinegared octopus as a traditional Japanese New Year's food.

The water temperature of the GPO's habitat was reported to be below 16°C off Canada (Hartwick, 1984) and 15°C off Tsugaru Strait (Robin and Sakurai, 2004), and GPO move to deeper waters offshore as the water temperature rises in summer (Noro and Sakurai, 2012). As a result, the coastal catches of GPO depend on fluctuations in water temperature. Furthermore, in the southern coast of the Tsugaru Strait, fishers have released small octopus (currently 3 kg) and established a voluntary fishery closure season (currently July to October), and these artificial factors also likely affect catches. Considering the background, it is expected that there is a strong periodicity in the catches of GPO along the coast of the Tsugaru Strait and make it possible to forecast catches using past data.

In this study, we forecasted by the model rebuilt year-by-year basis using catches of GPO since 1981 monthly catches from 2001 - 2019, for assessing its performance as forecasting methods. SARIMA as a Box-Jenkins method, and Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997; Cho, et al., 2014) and Gated recurrent unit (GRU) (Chung, et al., 2014), are one type of Recurrent neural network (RNN) model, were used to forecast, and assessed performances of them.

2. Materials and Methods

2.1. Data source

The data used are the monthly catch data of the octopus along the southern coast of the Tsugaru Strait, from

1981 to 2019 (Fig. 1). These were collected and compiled by Aomori Prefecture and published as “the results of a survey on marine fisheries”. In this dataset, GPO is contained in "catch of octopuses" which comprise catch of common octopus *Octopus vulgari*s, chestnut octopus *Octopus conispadiceus* and GPO. Of these, catch in weight of GPO accounted for 97.0 ± 1.9% (mean ± S.D.) of it in 2003 - 2009 (Noro, 2012). Therefore, we approximated "catches of octopuses" as catches of GPO in this study. Besides, in the coastal Tsugaru Strait, almost all GPO was caught mainly by selective fishing methods such as octopus basket, and unselective fishing methods such as trawl nets are prohibited as mentioned earlier. There is a strong seasonality in monthly catches in this area because GPO migrate to the deep from July - October when the water temperature goes greater than 15 °C (Noro and Sakurai, 2012), moreover there is a closed season from July - October (however, no-closed season from 1981 to 1990, and closed season from June-October from 1991 - 2001). When checked by year, the highest-catch year of GPO was 1986, with 1,780 t, and the lowest-catch year was 2015 with 327 t.

In this study, the estimation period was one year (12 months), and the training period used for modeling was at least 20 years (240 months). The dataset was divided into two parts: the training data from 1981 to 1 year before the estimation, and the test data from the estimated year.

2.2. Model

The SARIMA model and RNNs (LSTM and GRU) were used to forecast the catch of GPO. To assess the performance of the model, they were compared with a simple model called Naïve Forecast I (NI), which assumes that the latest value takes the previous value (Carlos Gutiérrez-Estrada, et al., 2007). Such method to compare forecast model with naïve forecast was used to assess forecasting for other fishery resources like Japanese anchovy (Kim, et al., 2015), and if the forecast model was worse than naïve model, it cannot be used.

The catch of GPO on the southern coast of the Tsugaru Strait was backcasted for each of the 2001–2019 years, using data from at least 20 years of 1981 - 2018 for training. If the obtained estimated value was negative, it was assigned 0.

2.3. SARIMA Model

The ARMA model is the Box-Jenkins method, which is a combination of the Auto Regressive model and the Moving Average Model (Whittle , 1951.), and it cannot be used for non-stationary data such as random walks, therefore AutoRegressive Integrated Moving Average (ARIMA) model makes the non-stationary series stationary by taking dth-order differences (Box, et al., 1976). Additionally, SARIMA is the model incorporating periodicity to the ARIMA.

The SARIMA model is represented as (p, d, q) (P, D, Q) $[s]$. Specifically, p and P are the orders of autoregressive regression, d and D are the different hierarchies taken to make the values stationary, q and Q are the orders of moving averages, respectively, and s is the period. (p, d, q) are orders of the time series and $(P, D,$

Q) are that of periodic directions. In this study, we set s=12.

The calculations were performed using the auto.arima function available in the forecast package (Hyndman, 2008) of the statistical analysis software R. The auto.arima is a function that automatically performs the unit root test, applies the ARMA model, check stationarity and decidability, and selects the best model using the AIC.

2.4. Recurrent neural network (RNN)

RNN is a neural network with a recursive structure, moreover it was used to identify and estimate time series such as climate change and dissolved oxygen in aquaculture ponds (Li, et al., 2021). By using a recursive structure for the cells in the Hidden layer, it has a memory of previous time points and can add long-period fluctuations to the model. RNN uses Back Propagation Through Time (BPTT) to propagate the error to the past, but the basic RNN (Fig.2) passes all information to the next term, which causes problems such as gradient loss and divergence, Input Weight Conflict and Output weight conflict, namely long-term data processing problems. Specifically, extreme values of error during the calculation process cause gradient loss or divergence. Meanwhile, Input & Output Weight Conflict occurred to be caused by the contradiction between the weight that requires certain information and the weight that does not require it, then the inability to learn. Firstly, LSTM was invented to solve these two problems (Hochreiter and Schmidhuber, 1997). Secondly, LSTM with forgetting gates (Gers, 2000) and LSTM with peephole coupling (Cho, et al., 2014), which is commonly used today, was devised by improving the basic LSTM. Thus, in this study, we used the peephole LSTM (Cho, et al., 2014). Peephole LSTM prevents gradient loss and divergence by using Constant Error Carrousel (CEC), which preserves the error and prevents Weight Conflict by using Input gate and Output gate to adjust how much to learn. Additionally, CEC is Peephole-coupled to each gate, so that the value at time t-1 acts on the Input gate and Forget gate, and the value at time t acts on the Output gate (Fig. 2).

GRU (Chung, et al., 2014) combines the CEC and output h, and Update gate that is combined the Forget gate and Input gate controls them, removes the Output gate and adds the Reset gate to prevent (Fig. 2), to simplify LSTM. While GRU reduce the computational cost, its performance reported to be equivalent to peephole LSTN (Li, et al., 2021).

The RNN computations were performed in the programming language Python, with the deep learning libraries Tensorflow (Abadi. et. al.,2016) and TFLearn (Tang, 2016). Because LSTM and GRU use random numbers, and the calculation results change a little with each, and so the random numbers were conducted with a fixed seed value (=23) to obtain ensured results of repeatability. Additionally, each parameter was adjusted as follows; the optimization algorithm was Adam, the floating-point number was performed in float64, the number of epochs was 150, the number of Hidden layers was 8, the learning rate was 0.001, the batch size was 1, and the window width was 12 months. Because in LSTM and GRU, data is needed to be normalized or standardized for training, we applied the Min-max Normalization as shown in the following equation (Yuan, 2019).

$$x_t = \frac{Z_t - Z_{MIN}}{Z_{MAX} - Z_{MIN}}$$

Z_t Monthly GPO *catch in the Tsugaru Straits, Aomori*

2.5. Verifying performance

The backcasted results for 19 years were compared with observed catches to assess the performance of the NI, SARIMA, LSTM, and GRU. Specifically, Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Correlation Coefficient (R), Coefficient of Determination (R^2) were used as indicators. Generally, MAE, MSE, RMSE, MAPE are error indicators which are used to assess their accuracy, the lower the value, the better, and R and R^2 indicate goodness of fit, and the higher the value, the better. Each indicator has own merit; MAE is easy to interpret because it is calculated using absolute values; RMSE and MSE can consider the impact of outliers better than MAE because those value calculated by squared value; MAPE can be used to assess different level data, such as case to assess the forecast performance that has seasonality from month to month, because it is percentage standardized by average. The errors assessed by MAE and MAPE show similar trend since their calculation process is almost the same, whereas MSE and RMSE also generate similar results .

$\hat{Z}_t$ *estimated values of Zt* $\hat{Z}_t = \hat{Z}_t$ ($\hat{Z}_t \geqq 0$), $\hat{Z}_t = 0$ ($\hat{Z}_t < 0$)

$\bar{Z}$ *average of Zt* $\bar{\hat{Z}}$ *average of* $\hat{Z}_t$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|Z_t - \hat{Z}_t\right| \quad MSE = \frac{1}{n}\sum_{t=1}^{n}\left(Z_t - \hat{Z}_t\right)^2$$

$$RMSE = \sqrt{MSE}$$

$$MAPE = \frac{100}{n}\cdot\sum_{t=1}^{n}\frac{\left|Z_t - \hat{Z}_t\right|}{Z_t}$$

$$R = \frac{\sum_{t=1}^{n}(Z_t - \bar{Z})\left(\hat{Z}_t - \bar{\hat{Z}}\right)}{\sqrt{\sqrt{\sum_{t=1}^{n}(Z_t - \bar{Z})^2}\sqrt{\sum_{t=1}^{n}\left(\hat{Z}_t - \bar{\hat{Z}}\right)^2}}}$$

*R^2 squared multiple correlation coefficient between the regressand and the regressors.**

** Kvalseth T. O. (1985) Cautionary Note about R^2, The American Statistician 39: 279–285.*

3. Results

3.1. models

The estimation values for each model were calculated. Each model backcasted negative catch value for some months out of the 228 months, therefore the catch of that month was assigned 0 (SARIMA 12 months, LSTM 3months, GRU 11months). For SARIMA model selection, the best model under the condition of s=12 and $0 \leq p$, P, d, D, q, $Q \leq 2$ was SARIMA (2, 0, 0) (0, 1, 2) [12] for 2001- 2018. By contrast, in 2019, the best model was SARIMA (1, 0, 1) (0, 1, 2) [12] , that was different from 2001- 2018 ones. Because there was only a trivial difference between AICs of SARIMA (1, 0, 1) (0, 1, 2) [12] and of SARIMA (2, 0, 0) (0, 1, 2) [12] (4,399.36 and 4,399.37, respectively) in 2019, the latter was adopted for all years in this study. The best model used the AR model with values up to two months ago was used as the time series component, and the MA model of values of first-order residuals up to two years ago as the periodic component with a 12-month cycle.

The comparison between observed catches and estimated catches for the 19-year (2001–2019) as backcasted by NI, SARIMA, LSTM, and GRU is shown in Table1. All of the four models had better value than NI. Additionally, RMSE was within a narrow range for the three models (27.9, 28.2, 28.5, and 48.3 t for LSTM, GRU, SARIMA and NI, respectively), and the best RMSE model was LSTM. Although the SARIMA had the best fit according to R and R^2, both the LSTM and GRU models had R values more than 0.90, indicating that these methods were sufficiently good fit to the actual catch. The relationship between the observed and estimated monthly catches for 2001–2019 are shown in scatter plots (Fig. 3). These models well matched with the actual catch, although the errors of all models tended to increase at moderate to greater catches (>10 t). Furthermore, all of these models well backcasted the overall fluctuation of the catch (Fig.4).

3.2. Yearly comparison

The RMSE, a measure of error, was calculated using the estimated and observed values for each year and compared (Table 2, Fig.5). Among the years 2001 - 2019, there were the lowest RMSE of value backcasted by SARIMA for 6 years (2004, 2005, 2009, 2013, 2016, and 2018), by LSTM for 8 years (2001, 2002, 2003, 2007, 2008, 2010, 2014, and 2015), by GRU for 5 years (2006, 2011, 2012, 2017, and 2019), and no year by NI. Simultaneously, according to the average RMSE for each year from 2001 - 2019, the best model was LSTM (25.8, 26.1, 26.8, and 44.7 t for LSTM, GRU, SARIMA and NI, respectively). SARIMA had larger error than NI for two years, while LSTM and GRU were consistently more accurate than NI throughout the years.

When the models were assessed by using RMSE, NI showed the comparable performance to other models after 2009, reflecting the reduced catches of GPO. To assess the trend of errors without being affected by the reduction of catch, we calculated MAPE, which is standardized by average catch of each year, to assess the trend of the estimation accuracy of each model (Fig. 5). To assess the performance of the models during the fishing season, data from the closed season (July – October) were excluded. As a result, the MAPE of the estimation backcasted by SARIMA and LSTM was higher than NI in 4 years (2010, 2011, 2015, and 2017) and 1 year (2011), respectively. In contrast, GRU showed smaller MAPE than NI throughout the years. The mean

MAPA was similar for LSTM and GRU (28.2% and 28.5%, respectively) while SARIMA showed an intermediate value (36.6%) between that for NI (49.9%).

3.3. Comparison of Monthly Catch

Similarly, we assessed each model by the monthly RMSE and fishing-season MAPE (Fig. 6). RMSE of SARIMA, LSTM and GRU had the equal seasonality as the level of octopus catch; fishing-season values were larger than closed-season ones. Additionally, the RMSE of SARIMA was lower values than ones of LSTM and GRU in closed season (July – October).

In fishing season, the lowest month of MAPE for LSTM, GRU, SARIMA and NI was in January (20.2%), June (18.9%), January (21.8%) and May (13.7%), respectively. The highest month of them were in November (36.1%), February (30.9%), November (67.4%) and June (101.1%), respectively. The model showing the lowest average of monthly fishing-season MAPE was LSTM (28.2%), followed by GRU (28.5%), SARIMA (36.6%) and NI (49.9%). Thus, MAPE and LSTM showed almost equal ability in predicting monthly catch. For April and May, the result was somehow exceptional with NI being most accurate. However, according to a comparison with MAPE of the other model's estimate, these models were not more inaccurate specifically than other months in these two months. Thus, there was no problem for forecasting in April and May because they were backcasted equally accurately to the other months.

4. Discussion

In this study, LSTM, GRU and SARIMA were used to backcast the catch of GPO for 19 years (2001 - 2019) and the estimation performance of each model was assessed. It shown all of the machine learning models takes much more accurate and stable forecasting than NI, have a good-fit to observed catches, according to MAE, MSE, RMSE, R and R^2 (Table1). Moreover, MAPE of NI scores significantly worse when catch fluctuated greatly due to fishery closure in July and reopening in November. Therefore, MAPE of NI for fishing season only should be better than that included all months because it was not included July value. Nevertheless, according to the mean MAPA by year from 2001 - 2019, all models backcasted more accurately than NI Besides, the accurate model was LSTM, LSTM and GRU were more effective than SARIMA model.

According to MAE and RMSE, the most accrete method was LSTM. However, SARIMA was lower error than GRU assessed by MAE, on the other hand GRU has lower RMSE than SARIMA (Table1). This means that SARIMA estimate is closer value to observation catches than GRU, meanwhile it means there are more outlier in comparison to GRU. Thus, SARIMA can forecast values closer to actual catches overall than GRU, but also can miss predictions by a wide margin. SARIMA is difficult to forecast structural changes due to switching regimes because SARIMA model's forecasting and time series are stationary and linear(Koutroumanidis, et al., 2006). Certainly, one of SARIMA's weakness is often discussed that the future parameter is affected by past time series excessively (Georgakarakos et al.,2002; Prista, et al.,2011). Thus, SARIMA's forecasting does not improve their accuracy after regime switching, thus it's possible that the number of outliers has increased.

In 2010, RMSE and MAPE of SARIMA were higher than those of the NI model; therefore, its forecasting had more outliers and less accuracy than NI. The reasons for this, octopus catch in 2010 year-end and new-year months (2009 December and 2010 January) was lowest since 1981, at 183.2 t ,compared to the average of 444.1 t from 1981 to 2009.Thus, SARIMA model was unable to have learn the first movement of extremely low catches in these months. However, for 2010, LSTM and GRU could backcasted more accurately. The catch in 2011 December and January was 149.3 t, which was less than 2010, SARIMA, and LSTM had higher MAPE than NI. According to the LSTM backcasting for 2011, it was assumed to be due to the catch data was not reproduced well in the training phase, 1981 to 2010. However, backcasting of 2011 could have been improved by tuning them individually because all backcasting used the same values of parameters in this study. Whereas the performance of LSTM and GRU was stable after 2012, the SARIMA model was unstable, with higher MAPE value than NI model in 2015 and 2017 fishing seasons. Generally, SARIMA cannot forget old information; SARIMA's forecasting is affected by past data excessively, moreover it has no means to adapt to the structural changes such as switching regimes (Koutroumanidis, et al., 2006; Prista, et al.,2011). Moreover, SARIMA models often had good fit either short or long time series in other study, SARIMA modeling does not necessarily need larger sample sizes (e.g., Lloret et al., 2000; Prista, et al.,2011). For this reason, the high-level-catch data of the past had strongly influence for SARIMA estimate. Therefore, this SARIMA model performance should improve by making dataset that has the training period encompassing the switching regimes. Apart from SARIMA, LSTM and GRU have mechanisms to prevent gradient loss/divergence and weight conflict, and to forget past information to maintain long-term forecast performance. Furthermore, their training mechanism improve their accuracy quickly after regime switching making them more capability to adapt to switching regimes.

Thus, in the purpose of octopus catch forecasting only, LSTM and GRU were better method than SARIMA. However, there was the advantage of SARIMA-based forecasting; it is easy to interpret the model and results (Kim, et al., 2015). Moreover, several studies found that SARSIMA model better explained the fish life history and fleet dynamics (e.g., Lloret, et al. 2000; Prista, et al.,2011). In this study, SARIMA (2, 0, 0) (0, 1, 2) [12] was selected as the best GPO catch model except for 2019. This fact means that the model has in addition to the AR (2) model at time series component, the MR (2) model with 12 periods and first-order differences at the periodic component. Altogether, this indicates that the catch of GPO is affected by the autoregressive of the last two months as the time series component, and the moving average of the same month of the last two years as the periodic component. Moreover, the cyclic component is negatively correlated with both years, indicating that if the catch of octopus increase excessively in the past two years, the model calculate the catch as the decreasing direction this year, and *vice versa*. GPO's lifespan, growth rate, and releasing policy of small-octopus probably caused this behavior. In the Tsugaru Strait, GPO's lifespan ranging from 2.5 to 5 years, though their growth rate varies greatly among individuals, they weigh less than 3 kg during the first year after hatching and that most of them will exceed 3 kg after the age of 2 (Noro and Sakurai, 2012). Additionally, fishers have released octopus weighting <3 kg. Thus, MA (2) of periodic component probably reflect the recruitment at the age of 2 and the intensive catch pressure thereafter , consistent with the interpretation of SARIMA (2, 0, 0) (0, 1, 2) [12]. The forecasting performance of SARIMA was better than that of LSTM and GRU in July - October, which is the closed season, it was due to the periodic component of the SARIMA worked effectively in the closed season. In contrast, the performance of SARIMA forecasting was poor in November. It was perhaps due to the fact that the previous two months (September and October) were closed season. Therefore, the catches for November were forecast only by the cyclic component. As above, although the SARIMA can forecast cyclical fluctuations well, it is probable that there is a variation of the performance of forecasting before and after seasons when catch has been reduced for prolonged periods due to closed season or migration.

According to the fishing-season MAPA, the performance of LSTM and GRU decreased in 2010–2011 equivalent to NI, but it became rather accurate since then (Fig.5). It is often reported that the larger the training dataset is, the more the model improves (e.g., Linjordet and Balog, 2019). Thus, the more training data an RNN has, the more accuracy its forecasting will be, because there are more opportunities to train a variety of data. Therefore, it will improve as the amount of the number of catch data increases. The dataset used in this study did not experience and train the situation where the level of catch recovered. Therefore, the LSTM and GRU estimation will be inaccurate temporarily when the GPO catch level switch increasing phase at the first time. However, it is expected that LSTM and GRU accuracy will improve quickly due to training the case. Although LSTM was slightly better than GRU in overall performance, the difference was not significant, and this result was consistent with the previous study (Li, et al., 2021). Moreover, GRU has clear advantages, in that the computational cost is less than that of LSTM Incidentally, the disadvantage of RNN forecasting is that it is a black box, making it difficult to interpret and discuss models in contrast to SARIMA. However, RNN has a great advantage in that it is easy to make precise forecasting.

As described above, LSTM, GRU and SARIMA were all accurate enough for practical use in estimating the future catch of the GPO, although there are areas for improvement. Because each model has advantages and disadvantages, so it is necessary to select the best model depending on its purpose. By applying these methods, it will be possible to make forecast of catches and resources easily for not only the GPO but also other species. In the future, more necessity of fishery and resource management is expected due to globally increasing demands for fishing products. The forecasting method like SARIMA and RNN of being able to make estimation with data such as fish catches without the need for additional data and research, and the convenience and immediateness are great advantages. Therefore, these methods based on machine learning of time series data will become even more essential approaches for the planning and management of fishery in the future.

Acknowledgments

References

Table 1 Comparison of model performance in forecasting annual catch data of octopus in the Tsugaru Strait, based on backcasting of 2001 – 2019 data.

Table 2 Root Mean Square Error (RMSE) for each model in backcasting annual catch of octopus in Tsugaru Strait. Shaded areas indicate the best estimate.

Figure caption

Fig1. Annual Catch of "octopus" in Aomori Prefecture, shown by months.

Fig 2. Chart showing the algorithm of A) RNN, B) LSTM and C) GRU; x_t : input, y_t : output, h_t : hidden state, c_t: Constant Error Carrousel, Dashed arrows: peep-hole.

Fig.3 Plots showing relationships between observed and model-estimated catches; A) NI model, B) SARIMA model, C) LSTM model, D) GRU model. Dashed line is a diagonal showing points of precise prediction.

Fig.4 Monthly change in actual catch (dashed line) and estimated catch of octopus in the Tsugaru Strait by different models; A) SARIMA, B) LSTM, C) GRU models.

Fig.5 Root mean squared error (RMSE) for annual estimate (Upper) and mean absolute percentage error (MAPE) for fishing season (Bottom) of octopus catch, generated by different models.

Fig.6 Root mean squared error (RMSE) for monthly estimate (Upper) and mean absolute percentage error (MAPE) for fishing season (Bottom) of octopus catch, generated by different models.

Table 3
Summary of machine learning methods studied; pros, cons and suitable use in catch prediction are summarized.

Methods	Pros	Cons	Suitable use
SARIMA	1) Clear and simple algorithm. 2) Easy model interpretation.	1) Overfitting to past data in prediction. 2) Unable to adapt to regime switch.	1) Model-based analysis of variance structure. 2) Forecasting within the same regime.
LightGBM	1) Able to forecast non-time-series data. 2) Speedy computation. 3) Effective even with missing or incomplete data. 4) Simple to add and interpret explanatory variables.	1) Overfitting to past data in prediction. 2) Unable to adapt to regime switch.	1) Feature/variable importance analysis. 2) Prediction with incomplete time-series (i.e. frequent missing data). 3) Forecasting within the same regime.
LSTM	1) Stable forecasting performance. 2) Automatically able to ignore past data not needed for forecasting (e.g. data before regime switch).	1) Prolonged computation time. 2) Difficulty in interpretation.	1) Forecasting across different regimes (i.e. Forecasting based on long-term time series) 2) Estimating abundance and catch trends
GRU	1) Stable forecasting performance. 2) Automatically able to ignore past data not needed for forecasting (e.g. data before regime switch). 3) Faster computation than LSTM.	1) Prolonged computation time. 2) Difficulty in interpretation.	1) Forecasting across different regimes (i.e. Forecasting based on long-term time series) 2) Estimating abundance and catch trends

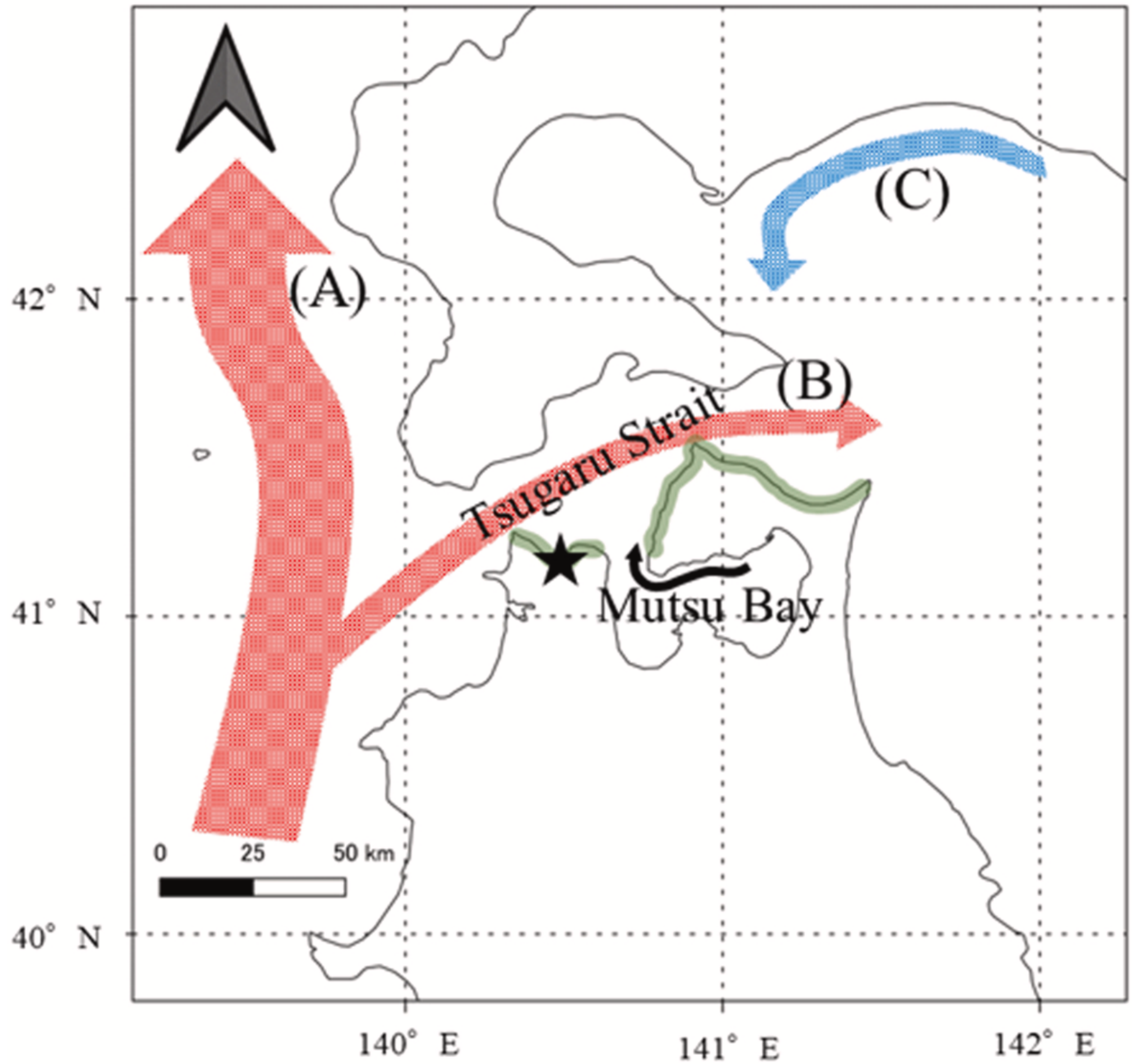
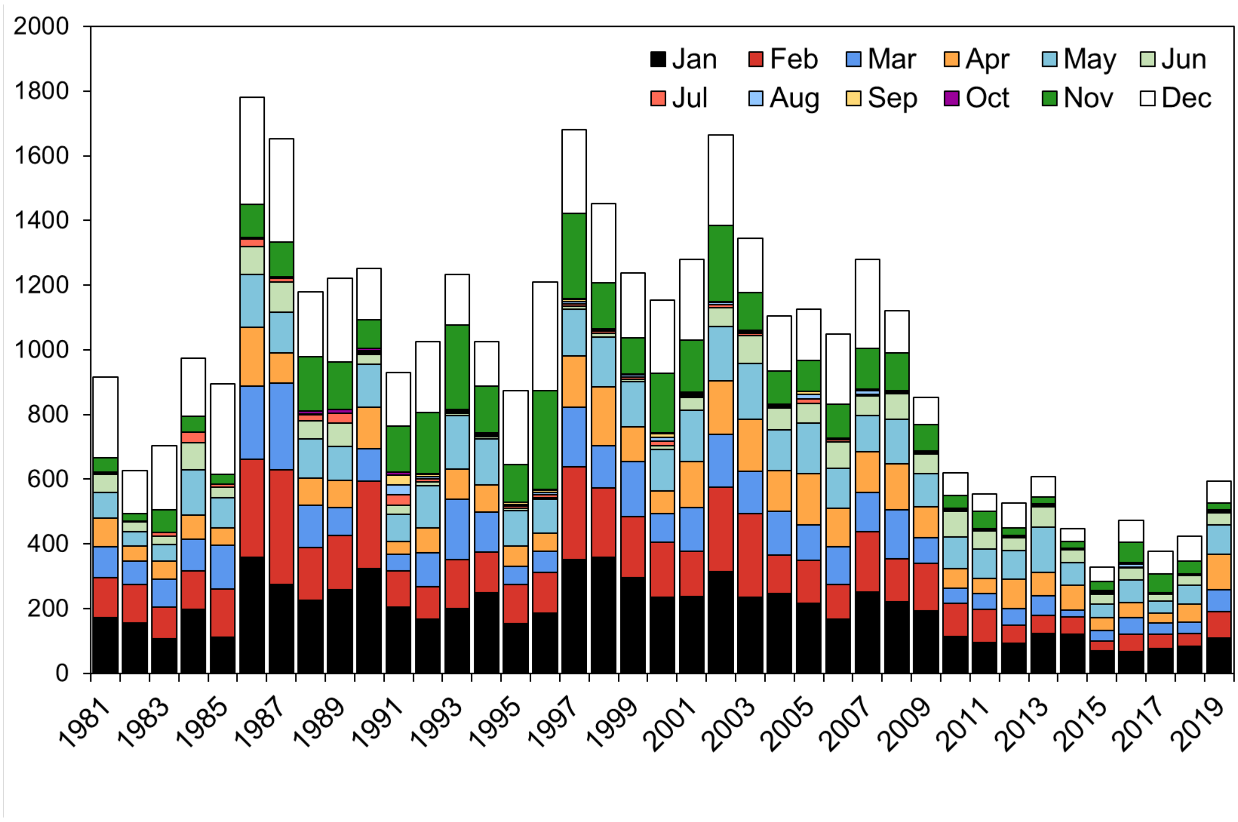


Fig. 1 Map of the area around Tsugaru Straits. Red/blue arrows indicate warm/cold current. A) Tsushima Warm Current, B) Tsugaru Warm Current, C) Coastal Oyashio Current. Green indicates the southern Tsugaru Straits area, where catch was counted. Black star indicates the Automated Meteorological Data Acquisition System (AMeDAS) station at Imabetsu.



Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
0
200
400
600
800
1000
1200
1400
1600
1800
2000
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
2015
2017
2019

Fig. 2

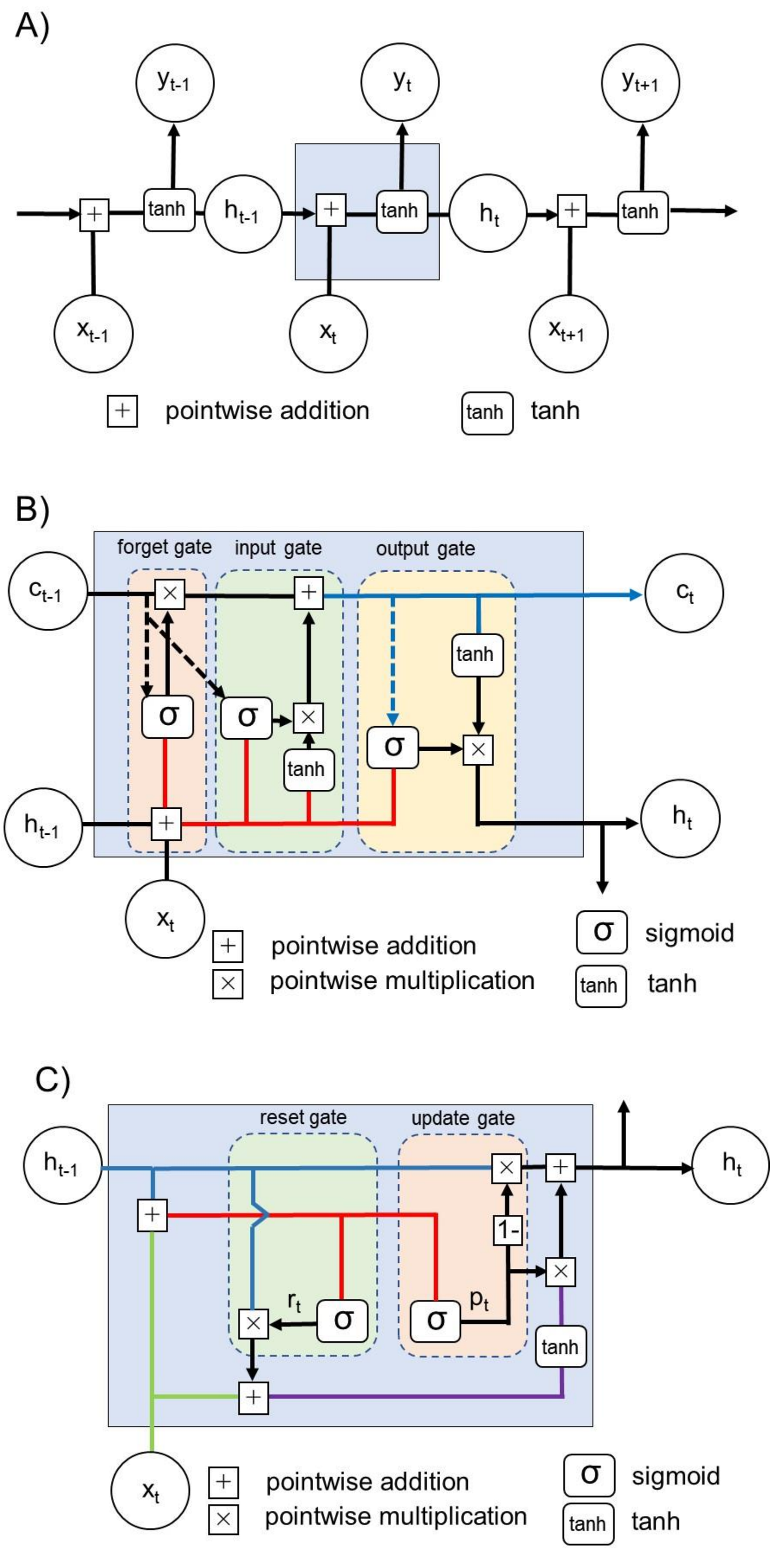
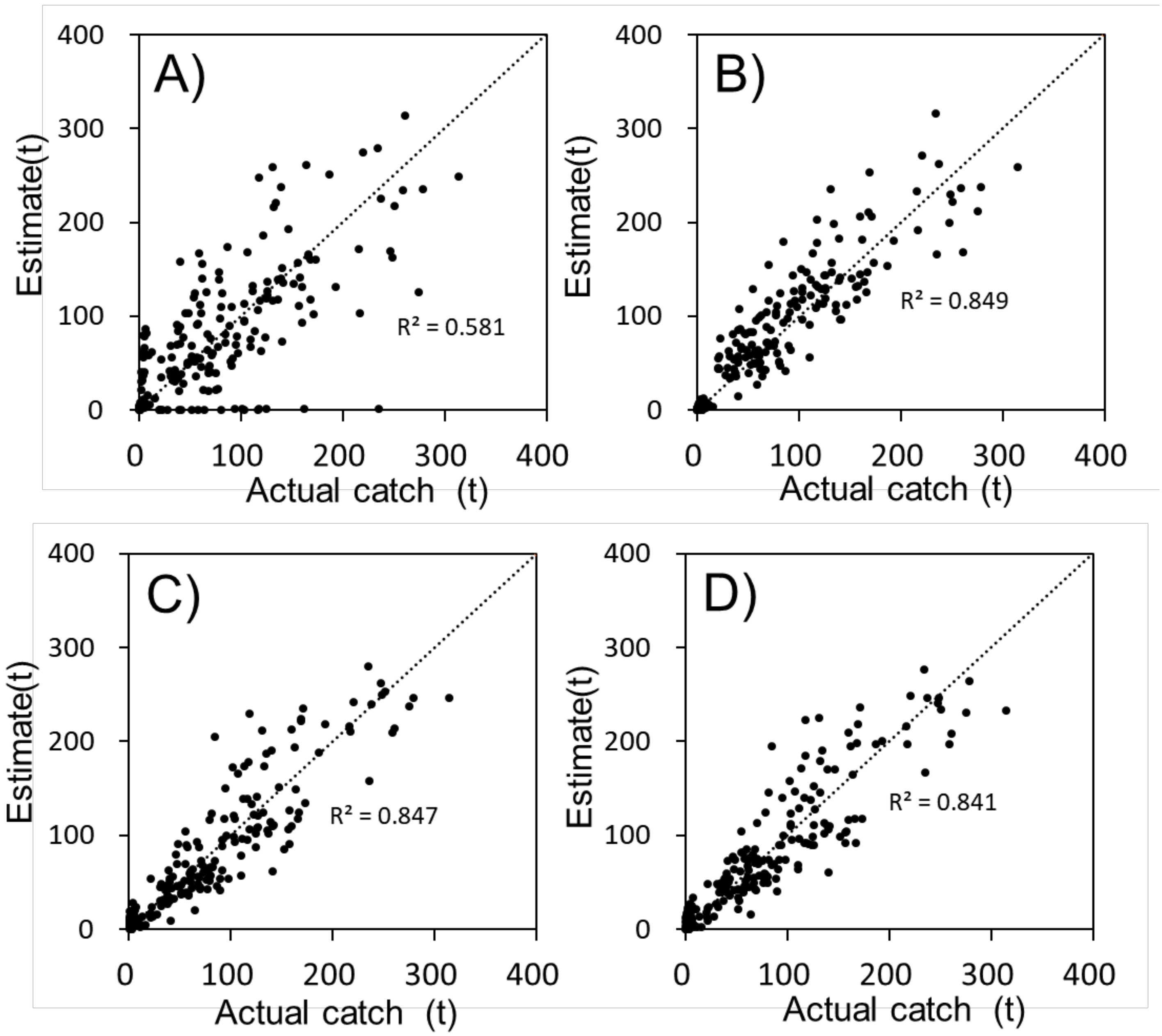

A)
y_{t-1}
y_t
y_{t+1}
tanh
h_{t-1}
h_t
x_{t-1}
x_t
x_{t+1}
pointwise addition
tanh
B)
forget gate
input gate
output gate
c_{t-1}
c_t
σ
tanh
h_{t-1}
h_t
x_t
pointwise addition
pointwise multiplication
sigmoid
tanh
C)
reset gate
update gate
h_{t-1}
h_t
1-
r_t
p_t
σ
tanh
x_t
pointwise addition
pointwise multiplication
sigmoid
tanh

Fig. 3



A)
B)
C)
D)
Estimate(t)
Actual catch (t)
$R^2 = 0.581$
$R^2 = 0.849$
$R^2 = 0.847$
$R^2 = 0.841$
0
100
200
300
400

Fig. 4

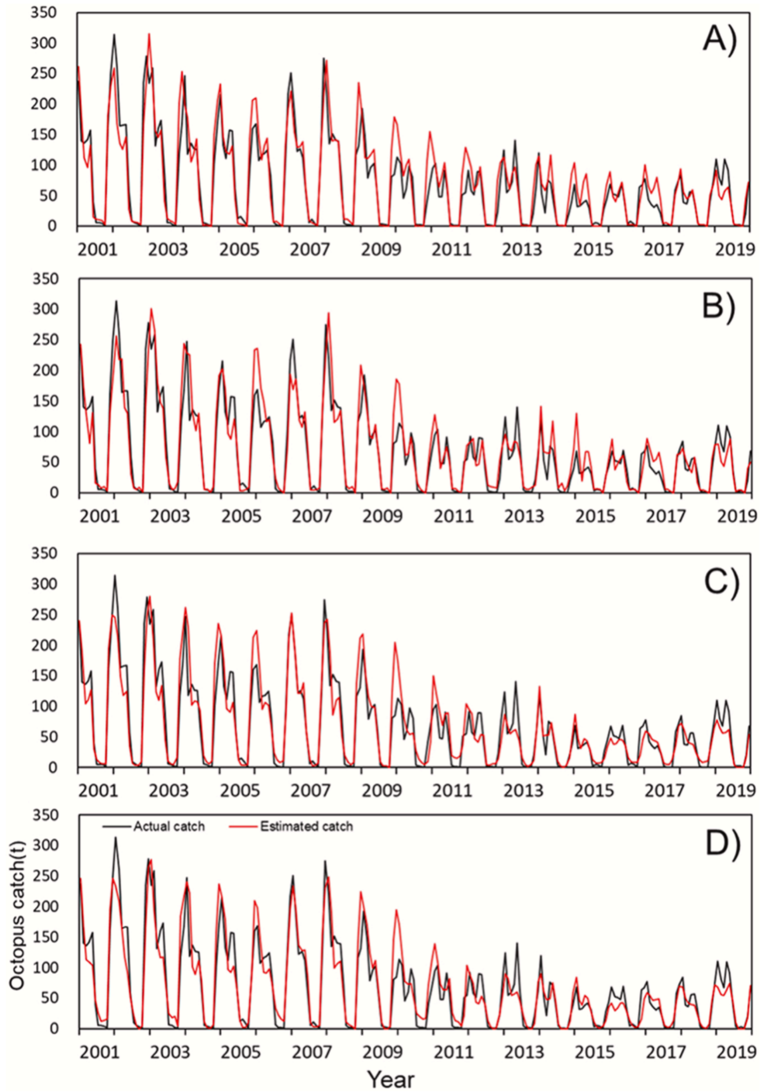


Fig. 5. Monthly changes in the actual catch and estimated catch of octopus in the Tsugaru Strait using the different models; A) SARIMA, B) L-GBM, C) LSTM, and D) GRU.

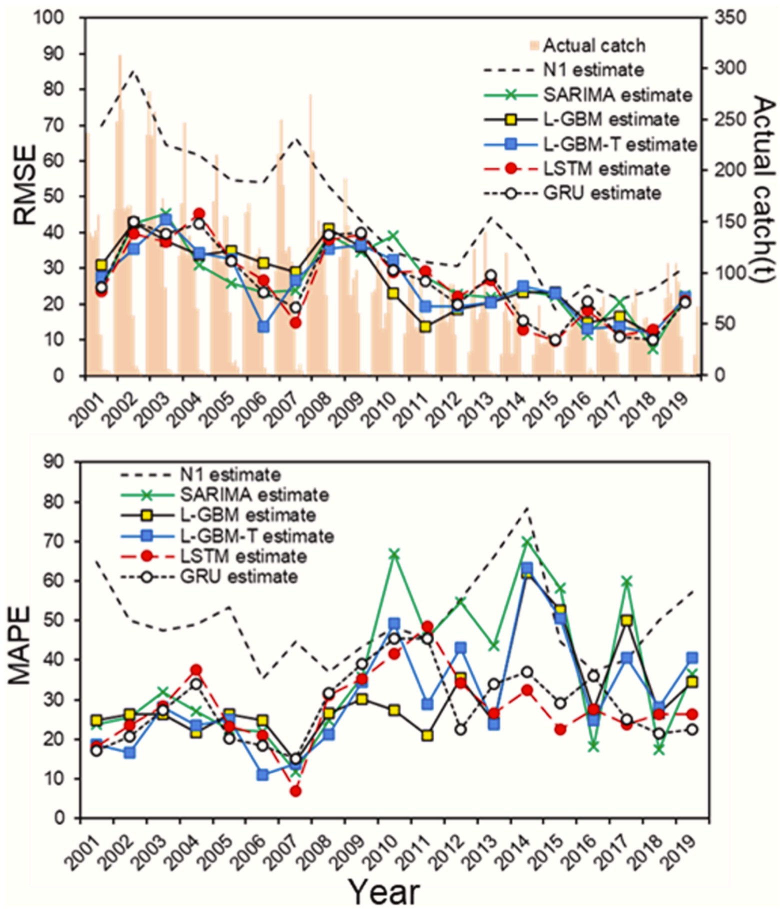


Fig. 6. Monthly actual catch and root mean square error (RMSE) for the annual estimate (Upper) and mean absolute percentage error (MAPE) for the fishing season (Bottom) of octopus catch, generated with different models.

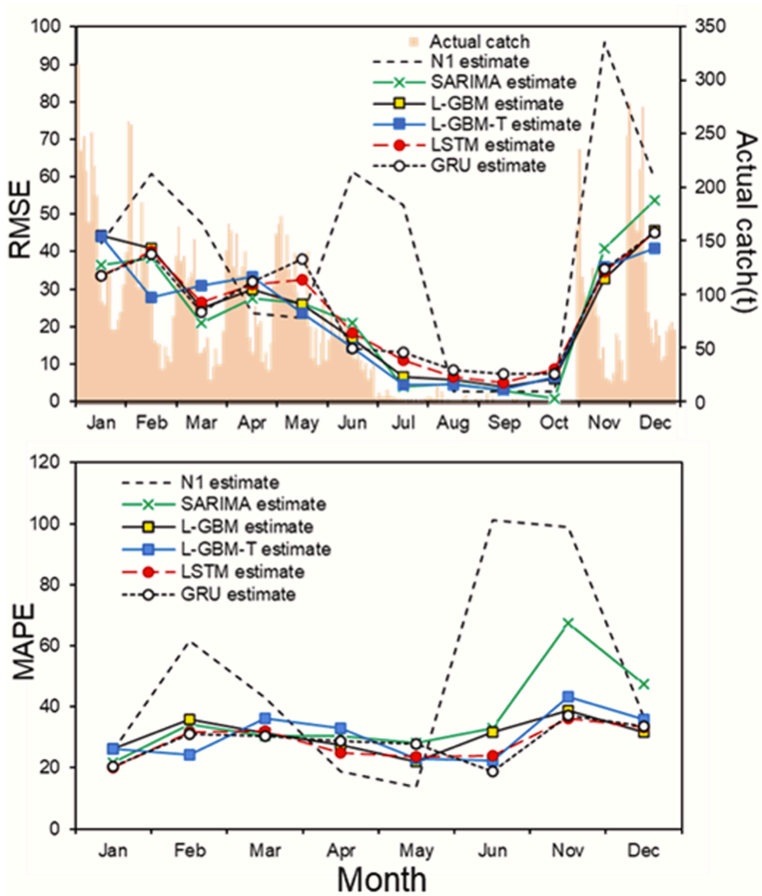


Fig. 7. Monthly actual catch and root mean square error (RMSE) for the monthly estimate (Upper) and mean absolute percentage error (MAPE) for the fishing season (Bottom) of octopus catch, generated with different models.

Table 1
Comparison of model forecasting performance using the annual catch data of octopus in the Tsugaru Strait, based on backcasting of the 2001–2019 data. The bolded entries are the first or second most accurate models.

	MAE	MSE	RMSE	MAPE	R	R^2
Naïve Forecast I	32.081	532,521.593	48.328	49.9	0.762	0.581
SARIMA	19.175	184,925.460	28.479	36.6	0.921	0.848
SARIMA-20	20.092	201508.334	29.729	35.8	0.903	0.815
SARIMA-5	22.978	275,011.402	34.730	34.5	0.869	0.756
L-GBM	**18.403**	178,115.657	27.950	30.7	**0.921**	**0.847**
L-GBM-T	**17.922**	**163,320.427**	**26.764**	30.8	**0.926**	**0.858**
LSTM	19.002	**176,997.364**	**27.862**	**28.2**	0.920	**0.847**
GRU	19.582	181,355.338	28.203	**28.5**	0.916	0.839

Table 2
Root mean square error (RMSE) for each model in the backcasted annual catch of octopus in the Tsugaru Strait. The bolded entries represent the models with the lowest RMSE values. The bottom row indicates the number of years in which each model achieved the lowest RMSE values from the 2001–2019 period.

YEAR	RMSE SARIMA	L-GBM	L-GBM-T	LSTM	GRU	N1
2001	24.331	30.807	27.643	**23.468**	24.694	69.753
2002	42.592	42.763	**35.402**	39.544	43.063	85.358
2003	45.400	37.667	43.580	**37.554**	39.699	64.479
2004	**30.906**	33.874	34.361	45.310	42.373	61.562
2005	**25.830**	34.851	32.544	32.144	32.068	54.723
2006	23.400	31.487	**13.792**	26.726	23.351	53.980
2007	23.795	29.024	26.548	**14.745**	19.091	66.579
2008	39.414	41.054	**35.582**	38.036	39.498	52.896
2009	**34.539**	35.911	36.437	39.284	40.071	43.219
2010	39.224	**22.982**	32.302	28.944	29.575	34.656
2011	27.608	**13.642**	19.401	29.328	26.478	31.818
2012	22.826	**18.636**	19.307	22.311	19.781	30.571
2013	21.807	20.582	**20.392**	26.551	28.224	44.213
2014	23.877	23.313	25.028	**12.755**	15.353	35.019
2015	22.185	23.356	22.999	**9.797**	10.148	18.556
2016	**11.424**	14.966	13.112	18.192	20.683	25.165
2017	20.361	16.430	13.622	11.046	**10.952**	21.333
2018	**7.479**	11.354	11.788	12.847	10.095	24.130
2019	22.361	20.853	21.952	20.973	**20.529**	30.439
	5 years	3 years	4 years	5 years	2 years	0 years

Table 3
Summary of machine learning methods studied; pros, cons and suitable use in catch prediction are summarized.

Methods	Pros	Cons	Suitable use
SARIMA	1) Clear and simple algorithm. 2) Easy model interpretation.	1) Overfitting to past data in prediction. 2) Unable to adapt to regime switch.	1) Model-based analysis of variance structure. 2) Forecasting within the same regime.
LightGBM	1) Able to forecast non-time-series data. 2) Speedy computation. 3) Effective even with missing or incomplete data. 4) Simple to add and interpret explanatory variables.	1) Overfitting to past data in prediction. 2) Unable to adapt to regime switch.	1) Feature/variable importance analysis. 2) Prediction with incomplete time-series (i.e. frequent missing data). 3) Forecasting within the same regime.
LSTM	1) Stable forecasting performance. 2) Automatically able to ignore past data not needed for forecasting (e.g. data before regime switch).	1) Prolonged computation time. 2) Difficulty in interpretation.	1) Forecasting across different regimes (i.e. Forecasting based on long-term time series) 2) Estimating abundance and catch trends
GRU	1) Stable forecasting performance. 2) Automatically able to ignore past data not needed for forecasting (e.g. data before regime switch). 3) Faster computation than LSTM.	1) Prolonged computation time. 2) Difficulty in interpretation.	1) Forecasting across different regimes (i.e. Forecasting based on long-term time series) 2) Estimating abundance and catch trends