WALAILAK JOURNAL

http://wjst.wu.ac.th

Forecasting World Tuna Catches with ARIMA-Spline and ARIMA-Neural Networks Models

Boonmee LEE^{1,*}, Suhartono SUHARTONO², Apiradee LIM¹ and Sung Keuk AHN³

 ¹Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani 94000, Thailand
 ²Department of Statistics, Faculty of Mathematics & Natural Sciences, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
 ³Department of Finance & Management Science, Carson College of Business, Washington State University, WA, USA

(*Corresponding author's e-mail: bm_lee2001@yahoo.com)

Received: 13 April 2020, Revised: 21 May 2021, Accepted: 28 May 2021

Abstract

Tuna is a renewable resource that has been managed regionally, but its worldwide capacity for regeneration is still little known. A time-series dataset of tuna catches was used to develop nonlinear univariate models for monitoring the sustainability of tuna catches. Two approaches were compared: 1) fitting an ARIMA-spline model to the volume of annual tuna catches and 2) combining neural networks with an ARIMA model to fit the annual changes in volume. These models offer competitive forecasting performance with small percentage errors. By averaging results of the best model developed in each of these approaches, our ensemble forecast predicts that world tuna catches will reach the optimal level of 5.09 million tons in 2025, remain stable thereafter until 2033, and start decreasing about 0.78 % annually. These models could be used by regional fishery management groups to discover discrepancies between such projections and other science-based estimations of the maximum sustainable output.

Keywords: ARIMA-spline model, Ensemble forecasts, Neural networks, Nonlinear univariate forecasting, Tuna catches

Introduction

Tuna is a widely traded commodity around the world. It is fished in all oceans, processed in many countries, and marketed around the world. The growing demand for nutrient-rich food, especially in developed countries, drives tuna species' high economic value as does the increase in multinational vessels fishing for tuna. With advanced fishing and storage technologies, tuna fisheries have expanded rapidly since the 1950s into all oceanic areas to which tuna migrates. Purse-seine fishing with drifting fish aggregating devices (FADs) has become the dominant means of catching tuna due to the more than 90 % success rates [1,2]. Consequently, the global capture of the principal commercial tuna species increased dramatically, from 0.4 million tons in 1950 to the 1st hit at 5 million tons in 2014 [3]. Of the total tuna catches in 2015 (4.8 million tons), 80 % came from 18 major countries, which fished, altogether, almost 4 million tons. The top 5 countries were Indonesia (11.8 %), Japan (8.0 %), Taiwan (6.7 %), Ecuador (6.6 %) and Korea (6.5 %). Almost half of the tuna caught were sourced from the Western Pacific Ocean. These tunas were shipped to and processed in several countries, especially Thailand, the world's leading tuna processing country since the 1980s [4].

Economically, the overall volume of tuna catches accounted for more than 10 billion US dollars in landed value [5], importantly contributing to economies of more than 100 countries. Ecologically, however, there have been global concerns about the rapid depletion of tuna stock over the past 2 decades.

The tuna stock assessment in 2014 showed that 39 % of tuna stocks were overfished [6]. By 2015, especially skipjack, the most caught species (58 % of total catches), was no longer plentiful, having been moderately exploited in all oceans, even though it has a 1-year maturity span [7]. Currently, there are 5 regional fishery management organizations (RFMOs) active in governing tuna fishing and controlling the health of tuna stocks through international law [8]. They are in charge of tuna management and conservation in their respective areas of jurisdiction, consisting of 1) Inter-American Tropical Tuna Commission (IATTC); 2) International Commission for the Conservation of Atlantic Tuna (ICCAT); 3) Commission for the Conservation of Southern Bluefin Tuna (CCSBT); 4) Indian Ocean Tuna Commission (IOTC); and 5) Western and Central Pacific Fisheries Commission (WCPFC).

These tuna RFMOs have implemented a series of regulations and measures toward achieving their conservation goals, such as requiring 1) FAD-free fishing to prevent overfishing and the catching of juvenile tunas; 2) catch certification to expose illegal, unreported, and unregulated (IUU) fishing; and 3) extending the duration of FAD closures for purse-seine fisheries [9,10]. However, such marine policy implementation is not simple and has raised concerns of success in conserving tuna stock particularly in the Pacific Ocean, which is the largest source of tuna [11], as it involves monitoring multiple fishing fleets over distant waters under different national jurisdictions.

Despite the fact that tuna is a renewable resource that is managed regionally, the sustainability of global tuna populations has remained dubious, and the future stability of tuna supply is in doubt. Thus, the goal of this research is to create a trustworthy dynamic statistical model for tracking patterns in global tuna catches that have consequences for long-term fishery management. In time-series analysis, both linear and nonlinear statistical methods have been applied to estimate future catches of different marine species. By using no explanatory variable apart from variables which are solely functions of time, the time series model fundamentally is applicable to forecast over the long horizon. If explanatory variables are used as leading indicators, the forecasting would be limited by the availability of the explanatory variables unless these are also forecasted [12]. Thus, to accommodate the uncontrollable conditions in fisheries management, such as environmental changes, variations in fishery capacity, and natural resource governance, different statistical methods were combined to improve forecasting accuracy. These methods involve combining autoregressive integrated moving average (ARIMA) models and neural networks [13,14] and using multiple linear regression models [15,16]. Moreover, spline functions have been wellapplied to smooth time series data in several scenarios [17-21]. Some of those methods were explored and integrated into our approach to develop nonlinear models which provide the most plausibility of forecasts for world tuna catches.

Materials and methods

The world tuna-catch time-series used in this study were retrieved from the Global Fishery and Agriculture Statistics [22]. The dataset contained annual captures in tonnage from 1950 to 2015 of all 7 commercially important species, including skipjack, yellowfin, bigeye, albacore, Pacific bluefin, Atlantic bluefin and Southern bluefin. The data of all 7 tuna species aggregated into total annual tuna catches were used to develop the statistical models.

First, a linear regression model was fit to the volume of annual catches to investigate the trend. The model was also fit to the changes in the volume of annual catches to examine variation patterns in global tuna catches. Then, the need for data transformation was identified by using a Box-Cox transformation [23]. Both the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) were plotted to identify appropriate orders of the ARIMA model [24,25]. Ramsey's RESET test [26] and Tsay's test [27] were used to verify the necessity of nonlinear models. Additionally, White's test [28] was executed to validate the appropriate use of neural networks for handling nonlinearity in the time series.

To take advantage of both perspectives of the time series, 2 nonlinear approaches were conducted and compared as diagramed in **Figure 1**. Approach 1 is modeling the volume of tuna catches with the ARIMA-spline model and approach 2 is modeling the changes in the volume of tuna catches by combining neural networks and the ARIMA model. All data analyses and modeling were carried out by using the R program [29]. In addition, the SAS [30] was used to simultaneously estimate residuals for identifying subset ARIMA models.

Approach 1: Modeling the volume of tuna catches

In this 1st approach, the goal was to model both the signal (linear regressions with spline) and the noise (autoregression) of the volume of tuna catches at the same time. A natural spline function was added as an external regression vector into the *arima ()* function in the R program [31] to smooth change points within the time series without loss of information [32]. However, instead of the commonly used cubic spline [33], a *quintic* spline was proposed to integrate into the ARIMA model. Because the reason was that a *quintic* spline provides quadratic rather than linear extrapolation [19], which more plausibly represents the long-term trend in global tuna catches.

The ARIMA-quintic spline model is constructed as follows:

$$Y_t = S(t) + z_t \tag{1}$$

where t is time (year), Y_t are total tuna catches at time t, z_t are errors identified from stationary processes, and S(t) is a *quintic* spline function for time t expressed below:

$$S(t) = a + bt + ct^{2} + \sum_{k=1}^{p-3} d_{k} \begin{bmatrix} (t - t_{k})_{+}^{5} - \frac{(t_{p-1} - t_{k})(t_{p} - t_{k})}{(t_{p-1} - t_{p-2})(t_{p} - t_{p-2})}(t - t_{p-2})_{+}^{5} \\ + \frac{(t_{p-2} - t_{k})(t_{p} - t_{k})}{(t_{p-1} - t_{p-2})(t_{p} - t_{p-1})}(t - t_{p-1})_{+}^{5} \\ - \frac{(t_{p-2} - t_{k})(t_{p-1} - t_{k})}{(t_{p} - t_{p-2})(t_{p} - t_{p-1})}(t - t_{p})_{+}^{5} \end{bmatrix}$$
(2)

where p is number of knots, $t_1 < t_2 < ... < t_p$ are specified time knots, and $(t - t_k)_+$ is $t - t_k$ for $t > t_k$ and 0 otherwise, and a, b, c, $d_1, d_2, ..., d_{p-3}$ are model coefficients.

Using this *quintic* spline smoothing in the model offers multiple forecasting outcomes because positions of the knots are free parameters. To avoid overfitting, only 4 -- the minimum number of knots -- were implemented in the *quintic* spline, and results of the ARIMA-*quintic*-spline model with different combinations of 4 knots were computed and compared. As equispaced knots or equal distance of knots positions are commonly used [34] for spline interpolation, they are considered as criteria for knots selection of the spline model. The tests for model validity include Akaike Information Criteria (AIC), Root Mean Square Errors (RMSE), Mean Absolute Percentage Errors (MAPE) and adjusted r-squared. Besides such goodness-of-fit validation, the plausibility of forecasts is also essential to finalize the forecasting model.

Furthermore, a Monte Carlo simulation was implemented to test the consistency of forecasting performance through Bootstrap [35]. The reason for this rigorous validation is to ensure the model performance for long-term forecasting. There are the following steps: 1) A set of 100 residuals was randomly sampled (with replacement) from the fitted model, and 100 annual tuna catches were generated. 2) Those first 66 tuna catches were modeled, and the last 34 catches were held in reserve as future observations. 3) Using the fitted model, the next 34-step-ahead tuna catches were forecast and compared with those future observations. 4) Steps 1 - 3 were repeated 1,000 times, and the mean and the standard deviation of the relative forecast errors were calculated to measure the model's performance.

Approach 2: Modeling the changes in the volume of tuna catches

In this approach 2, the purpose was to apply neural networks, capable of modeling nonlinear relationships among resource and environmental variables associated with non-controllable changes in fisheries [36,37], especially when predicting complicated phenomena [38]. Thus, a neural network was combined into a model, fitting the changes in the volume of tuna catches, essentially to improve the estimations of errors from the model. For comparison purposes, 2 neural networks were constructed:

$$W_t = \beta_0 + \beta_1 t + \frac{(1 - \theta_1 B^1 - \theta_2 B^2 - \dots - \theta_q B^q)}{(1 - \theta_1 B^1 - \theta_2 B^2 - \dots - \theta_p B^p)} a_t$$
(3)

$$W_t = \beta_0 + \beta_1 t + z_t \tag{4}$$

where t is time (year); W_t is the changes in the volume of tuna catches at time t; B is backshift operator; p and q denote orders of autoregressive and moving average respectively; a_t are residuals series from the ARIMA model (Eq. (3)); z_t are residuals series from the linear regression model (Eq. (4)), and β , \emptyset and θ are model coefficients.



Figure 1 Two nonlinear approaches for forecasting world tuna catches.

In this application, some lagged residuals from the linear regression and ARIMA models were used as input variables, as they explain the nonlinear autocorrelation structure of the data. However, there was no specific theory dictating the specific number of lagged observations we should choose to input in neural network modeling [39]. Thus, we conducted experiments to find an appropriate number of lagged residuals providing minimal forecasting errors. For each neural network structure, the *neural net ()* function in the R program [29] was applied to train up to 10 artificial neurons in a single hidden layer with 30 repetitions. This neural training used the backpropagation algorithm to get linearly-predicted residuals as an output of the simple nonlinear regression model as diagramed in **Figure 1**, step 4 of approach 2. Such repetitions of neural network training were recommended to obtain the best plausible model with 99 % confidence [40]. RMSE was the key measure assessing neural network performance. With the least error estimations of noise, the forecasts of changes in the volume of catches were generated and calculated into the volume of catches. As outlined in **Figure 1**, there are 4 forecasting results from approach 2; two are from linear regression and ARIMA models alone and the other 2 are from those models combined with neural networks. The MAPEs of each model were then compared.



Figure 2 World tuna catches 1950 - 2015 (a) and changes in world tuna catches 1951 - 2015 (b).

Results and discussion

Results from 1^{st} fitting the linear regression model to the 66-year volume of tuna catch demonstrated an upward trend with very high goodness-of-fit, adjusted r-squared 96.88 % as shown in **Figure 2(a)**. When exploring the rates of tuna catches, this procedure demonstrated a downward trend as depicted in **Figure 2(b)**. The opposite directions of these trend analyses showed that the growth in rates of annual tuna catches could soon reach the optimal level of catches, the highest volume of catches before decreasing. Just how soon was one of the key findings from results described in the following 2 approaches:

Results from approach 1: Modeling the volume of tuna catches

As illustrated in **Figure 3(a)**, the Box-Cox transformation plot of tuna catches time-series shows that the optimal λ , an estimate of power transformation was close to 0.5, meaning that a squared-root was needed for data transformation. The normal Q-Q plot in **Figure 3(b)** illustrates that the studentized residuals from modeling the squared-root transformed tuna catches were stabilized. Even though the squared-root transformation improved the data variability of the model up to adjusted r-squared 98.59 %, both Tsay's test and Ramsey's RESET test of the time series resulting in *p*-values < 0.05 indicated the presence of nonlinearity in annual tuna catches. This confirmed tuna catch in **Figures 4(a)** - **4(b)** revealed that there was a linear relationship between past and future volumes of tuna catches at lag-1, lag-5 and lag-10. In the iterative process, the ACF and PACF of residuals from the AR(1) model as illustrated in **Figures 4(c)** - **4(d)** still showed the existence of residual autocorrelations. Finally, this procedure identified that the subset ARIMA([1,9],0,4) model was the appropriate orders of the ARIMA model to remove all significant autocorrelation as shown in **Figures 4(e)** - **4(f)**, the ACF and PACF of residuals from ARIMA([1,9],0,4) model.

Thus, the integrated ARIMA([1,9],0,4)-quintic-spline model was the designated nonlinear method to model both the signal and the noise of the annual tuna catches, and the quintic spline function helped smooth the data represented of the slow-down in the catching rates after the year of 2000. Based on the Eq. (1), the ARIMA([1,9],0,4)-quintic-spline model can be re-expressed in Eq. (5), where Y_t are square-root-transformed catches.

$$Y_t = S(t) + \frac{(1 - \theta_4 B^4)}{(1 - \theta_1 B - \theta_9 B^9)} a_t$$
(5)



Figure 3 A Box-Cox transformation plot of tuna catches time-series (a) and a normal Q-Q plot after fitting the linear model to squared-root transformed tuna catches (b).



Figure 4 ACF and PACF plots of data modeling square-root-transformed tuna catches.

Then, this model equation was used to generate forecasting results with statistical measures for each combination of 4 knots. For knots selection of the spline model, AIC was the 1st consideration. As exhibited in Table 1, the smallest AIC was 643.91, resulting from the model M1 with the 4 knots at 1990, 1999, 2000 and 2010. This model M1 predicts that the optimal catches of tuna will reach 5.49 million tons in year 2025 before dropping significantly at the rate of -0.47 %. However, this forecast may not correspond to catchability under sustainable fishing management. Thus, the least change after the optimal catches was the 2^{nd} consideration since it would explain the plausible long-term trend better. The model M2 with 4 knots at 1983, 1992, 2004 and 2012, meets the 2^{nd} criteria with the least change, -0.000004 %, whereas its AIC, 644.28 has no significant difference from M1's. This model M2 predicts higher optimal catches of 5.7 million tons. When considering equally distanced knots as the 3rd criteria, model M3 shown in Figure 5(a) with 7-year equispaced knots at 1990, 1997, 2004 and 2011, has the smallest AIC, 643.93, and its forecast of the optimal catches is 5.45 million tons, 40,000 tons lower than the model M1's forecasts. However, the model M3 has sharply decreasing rates, -0.59 % after the optimal catches, not meeting the 2nd criteria whereas the model M4 exhibited in Figure 5(b) with 9-year equispaced knots at 1984, 1993, 2002 and 2011, satisfies all 3 criteria. It predicts the optimal catches reaching 5.69 million tons, about the same as M2's forecasts. All these 4 spline models (M1 - M4) have high goodness-of-fit. Their adjusted r-squared is between 99.59 to 99.62 % and they have low mean errors, MAPE ranged from 1.85 to 1.90 % and RMSE ranged from 30.45 to 31.61. These validation measures have no statistically significant difference. Thus the model M4 was selected for approach 1 and named in short "A1: ARIMAspline" as expressed below:

Walailak J Sci & Tech 2021; 18(17): 9726

7 of 15

$$Y_t = 584.94 + 27.22t - 0.000009S(t) + \frac{(1 - 0.51919B^4)}{(1 - 0.6961B + 0.2637B^9)}a_t$$

From the Monte Carlo simulation, the relative forecast errors also confirm that the model outperformed. By comparing the next 34 stepped-ahead simulated annual tuna catches with future catches generated from the model, the mean of forecast errors ranged from -0.073 to 0.709 % and the standard deviations were between 0.0130 and 0.0538. Furthermore, the model diagnostics of the residuals do not indicate any heteroscedasticity and autocorrelation of errors.

Based on existing tuna fishery implications, the ARIMA-spline model predicted that the total tuna catches of the 7 major commercial species in 2016 would be 4.81 million tons, slightly decreasing from 4.86 million tons in 2015 before slightly rising 2.7 % annually to the optimal volume of catches at 5.69 million tons in 2025 (**Table 1**).

Table 1 Forecasts and parameter estimations from different knots position in the ARIMA([1,9],0,4)*quintic* spline model.

Model ·	Coefficients							MAPE	RMSF	% Adi	Forecasts of tuna catches (Million MT)						%change after
	intercept	year	spline	ar1	ar9	ma4	АС	MALE	RMSE	r-squared	2016	2017	2018	2019	2020	2025	optimal catches
M1	585.74	27.14	-0.000028	0.682	-0.269	0.5148	643.91	1.86	30.63	99.62	4.79	4.99	5.09	5.24	5.36	5.49	-0.472829
M2	585.02	27.22	-0.000007	0.697	-0.263	0.5196	644.28	1.90	31.61	99.59	4.81	5.02	5.15	5.31	5.44	5.70	-0.000004
M3	586.12	27.12	-0.000026	0.682	-0.270	0.5148	643.93	1.85	30.45	99.62	4.79	4.97	5.09	5.23	5.34	5.45	-0.591688
M4	584.94	27.22	-0.000009	0.696	-0.263	0.5191	644.23	1.90	31.60	99.59	4.81	5.02	5.14	5.30	5.44	5.69	-0.000262



Figure 5 Graphical forecasts of tuna catches from ARIMA([1,9],0,4)-*quintic* spline models with different equispaced knots.

Walailak J Sci & Tech 2021; 18(17): 9726

Results from approach 2: Modeling the changes in the volume of tuna catches

As illustrated in **Figure 2(b)**, the changes in the volume of tuna catches during 1951 - 2015 fluctuated between -0.0864 and 0.2235, and demonstrated a downward trend. Based on Eq. (4), this trend linear regression model labeled as "A2: TLR" has the following formula:

$W_t = 1.9019234 - 0.000939t + Z_t$

From autocorrelation diagnoses, the ACF and PACF plots in **Figures 6(a)** - **6(b)** show results from the initial linear model while the plots in **Figures 6(c)** - **6(f)** exhibit significant correlations that remained after fitting ARIMA(1,0,0) and ARIMA(1,0,8) models respectively. Last, only lag-23 remains significantly correlated as identified in **Figures 6(e)** - **6(f)**. This iterative process demonstrated that ARIMA(1,0,[8,23]) was the most appropriate subset ARIMA model to eliminate all serially correlated noise. The parameters for this subset ARIMA model were simultaneously estimated as expressed below in Eq. (3). This forecasting model of changes in the volume of tuna catches was labeled "A2: ARIMA".

$$W_t = 1.917963 - 0.00095t + \frac{(1 - 0.32938B^8 - 0.60693B^{23})}{(1 + 0.41642B)}a_t$$

The results from White's test with *p*-value < 0.001 confirmed that there was nonlinearity in the annual changes of tuna catches and this validation indicated the suitable use of neural networks to handle such nonlinear data. From experiments using neural network implementation with different autoregressive lags in the initial 2 models ("A2: TLR" and "A2: ARIMA"), it was found that 1-year lagged residuals were preferred for both neural network structures. It is obvious that both combined neural network models, either "A2: TLR-NN" or "A2: ARIMA-NN" greatly improve the forecasting performance of annual changes in the volume of tuna catches. The results from model testing also show outperformance, with a small MAPE of 0.37 % for "A2: TLR-NN" and 0.41 % for "A2: ARIMA-NN".

As shown in **Figure 7(a)**, the "A2: TLR-NN" model predicts that annual changes in the volume of tuna catches will linearly decrease to close to zero in 2025, while in **Figure 7(b)**, the "A2: ARIMA-NN" model forecasts fluctuating changes with a very slightly decreasing trend, which will be close to zero in 2033. Based on these predicted changes, we calculated the forecasts for the volume of tuna catches exhibited in **Table 2**. The projections of the optimal volume of tuna catch derived from "A2: TLR-NN" and "A2: ARIMA-NN" models will be 4.77 and 4.51 million tons respectively. Among the 4 models from this approach 2, "A2: ARIMA-NN" model obtains the lowest MAPE 4.62 %, a little bit higher than the "A1: ARIMA-spline" (MAPE 3.59 %).



Figure 6 ACF and PACF plots of data modeling changes in tuna catches.



Figure 7 Graphical results from data modeling changes in tuna catches.

Models	MAPE	2016	2017	2018	2019	2020	2025	2026	2027	2033	2034	2035
A1: ARIMA-spline	3.59	4.81	5.02	5.14	5.30	5.44	5.70	5.70	5.68	5.49	5.46	5.45
A2: TLR	12.04	5.50	5.52	5.57	5.60	5.63	5.71	5.70	5.70	5.55	5.50	5.46
A2: TLR-NN	5.17	4.62	4.64	4.68	4.70	4.73	4.77	4.77	4.76	4.62	4.58	4.54
A2: ARIMA	7.67	5.25	5.21	5.37	5.33	5.20	5.22	5.11	5.11	5.30	5.26	4.98
A2: ARIMA-NN	4.62	4.57	4.52	4.66	4.62	4.50	4.49	4.39	4.38	4.51	4.47	4.22
Average†		4.69	4.77	4.90	4.96	4.97	5.09	5.04	5.03	5.00	4.97	4.84

 Table 2 Forecasts of world tuna catches from alternative models of the 2 approaches.

Notes: Forecasts are in million tons and bold values are smallest MAPE of each approach. The Average[†] is an ensemble forecasts by averaging results from A1: ARIMA-spline and A2: ARIMA-NN models.

Graphing the forecasts from all 5 alternative models in **Figure 8**, A1: ARIMA-spline is the ARIMA([1,9],0,4)-*quintic* spline model and A2: ARIMA-NN is the ARIMA(1,0,[8,23])-neural network model. It clearly shows that the models with neural networks combined in the statistical calculations offer much more conservative results even though they improve model performance statistically. To take advantage of the unique strengths of the 2 approaches, we, therefore, propose the ensemble forecasts for world tuna catches by averaging results from the first 2 outstanding models (ARIMA-spline and ARIMA-NN models). As illustrated in **Figure 8**, the black bold line, named "Avg+" is averaged results of those 2 outstanding models and the 2 vertical-cyan-dashed lines are optimal years, the highest level of predicted catches before declining trends. 2025 is the year of optimal catches for A1: ARIMA-spline, A2: TLR, and A2: TLR-NN models and 2033 is the year of optimal catches for A2: ARIMA and A2: ARIMA-NN models. These ensemble forecasts predict that world tuna catches would fall 3.6 % in 2016 before beginning the rise of an average 0.9 % to the level of optimal catches at 5.09 million tons in 2025 and remaining relatively stable at 4.9 - 5 million tons thereafter till 2033. Then, total tuna catches are expected to slowly decrease by about 0.78 % annually during 2034 - 2040.

In 2019, the Food and Agriculture Organization (FAO) reported that global tuna captures of 2016 - 2017 were 4.91 and 5.07 million tons respectively [41]. The actual catches were 5 % higher than the ensemble forecast whereas the actual catches were 2 % higher than the forecasts from the ARIMA-spline model. With this small difference, it is convinced that there might be an effort to sustainably manage the global tuna capture at 5 million tons per year, maximum. Knowing this optimal expected global supply of tuna would encourage stakeholders to employ tuna fishery management practices that could ecologically secure global tuna stock and consequently sustain world tuna industries. Moreover, in the future, if a statistically based projection of maximum sustainable yield (MSY) of all tuna species in all oceans was publicly available, we could identify gaps between that projected MSY and our forecasts. Such gaps would be useful for reviewing the adequacy of existing tuna-related conservation measures and would help support decisions made by the tuna RFMOs.

At the time of revising this paper in 2021, the global tuna captures of 2018 - 2019 were reported 5.52 and 5.72 million tons respectively [42]. These catch rates dramatically increased 9 and 3.6 % subsequently. The volume of these 2-year catches was 14 % greater than the ensemble forecasts and 8 % higher than the forecasts from the ARIMA-spline model. This could be further investigated whether such catch abilities were under controllable fishing quota. Industry groups expressed the need for transparency in management practices from the WCPFC, the fifth RFMO governing tuna fisheries in the Western and Central Pacific Ocean (WCPO), the area that provides the largest tuna resources in the world. Industry groups have requested that the WCPFC share information on decisions made regarding compliance with conservation measures since these groups believe tuna conservation progress has been undermined [43]. Considering this communication gap in information on tuna governance, our nonlinear approaches could

also be a tool dynamically refined further to fit data quantifying catches of specific tuna species or catches in specific fishing areas like the WCPO to help explain a situation by fact and statistics.



Figure 8 Graphical forecasts for world tuna catches 2016 - 2040.

Conclusions

The tuna ecosystem remains imbalanced and the nonlinear relations in global tuna catch result from multiple interacting factors, including exploitation intensity, environmental changes, economic dynamics and/or interventions to facilitate sustainable fishery management. Researchers face daunting challenges when trying to capture and interpret existing information within an available dataset. Both scientific and statistical approaches are needed to improve technical procedures for dealing with uncertainty in fisheries management [44].

Given the multiple uncertainties and complex interaction of variables, forecasting tuna catches in this study was also no simple task. Thus, a diverse set of techniques had to be adopted and combined to obtain the most plausible forecasts. Model validation and rigorous testing are crucial, and the 2 criteria always considered for model selection are 1) *how well the model fits past data* and 2) *how plausible the model forecasts future data.* It is obvious that integrating a spline function into an ARIMA model and combining neural networks with an ARIMA model have similar accuracy and offer competitive forecasting performance for modeling tuna catches data in our study. As the primary interest of the research is the plausibility of long-term forecasts, rather than making both approaches competitive, the results of both models were averaged. These ensemble forecasts ultimately give a convincing picture of the future of world tuna catches that has implications for fishery sustainability management. The forecasts these models enable will be useful not only for monitoring long-term sustainable fisheries of tuna regionally or globally but also for reformulating future allocation programs for internationally shared fisheries.

Acknowledgements

The authors wish to express their gratitude to the Prince of Songkla University, Thailand, for granting a Ph.D. the overseas Thesis Research scholarship to B. Lee to do this research at Washington State University (WSU), WA, U.S.A. We are grateful to Emeritus Professor Don McNeil from Macquarie University, Sydney, Australia, for rewriting mathematical expressions of *quintic* spline functions and coaching on statistical modeling. We highly appreciate Dr. Kathleen J. Warren for her assistance with our articulation in English. Moreover, we wish to acknowledge the Department of Finance and Management Science, College of Business, WSU for sponsoring data sources used in this tuna research. We sadly note that one of our co-authors, Dr. Suhartono succumbed to the COVID-19 after the 1st draft of this paper was submitted. His contribution to this paper, especially, the statistical method is invaluable as well as to the profession.

References

- [1] J Joseph. *Managing fishing capacity of the world tuna flee*. Fisheries Circular No.982. FAO, Rome, Italy, 2003, p. 23-34.
- [2] TK Davies, CC Mees and EJ Milner-Gulland. The past, present and future use of drifting fish aggregating devices (FADs) in the Indian Ocean. *Mar. Policy*. 2014; **45**, 163-70.
- [3] Food and Agriculture Organization (FAO). Global production by production source, Available at: http://www.fao.org/fishery/statistics/software/fishstatj/en, accessed March 2017.
- [4] A Hamilton, A Lewis, MM McCoy, E Havice and L Campling. *Market and industry dynamics in the global tuna supply chain*. FFA, Honiara, 2011, p. 18-52.
- [5] Food and Agriculture Organization (FAO). Tuna: A global perspective, Available at: http://www.fao.org/docrep/017/ap939e/ap939e.pdf, accessed March 2017.
- [6] International Seafood Sustainability Foundation. *ISSF tuna stock status update-2016*. Washington, D.C., USA, 2016, p. 3-5.
- [7] A Fonteneau, E Chassot and N Bodin. Global spatio-temporal patterns in tropical tuna purse seine fisheries on drifting fish aggregating devices (DFADs): Taking a historical perspective to inform current challenges. *Aquat. Living Resour.* 2013; **26**, 37-48.
- [8] International Seafood Sustainability Foundation. Management RFMOs, Available at: https://iss-foundation.org/about-tuna/management-rfmos/, accessed March 2017.
- [9] J Lenoci. *Sustainable tuna fisheries for blue economy*. Partnerships in Environmental Management for the Seas of East Asia, Quezon City, Philippines, 2018, p. 4-15.
- [10] Western and Central Pacific Fisheries Commission. Summary report: The commission for the conservation and management of highly migratory fish stocks in the Western and Central Pacific Ocean. Thirteenth Regular Session of the Scientific Committee, August 9 - 17. Rarotonga, Cook Islands, 2017, p. 137-9.
- [11] E Havice. The structure of tuna access agreements in the Western and Central Pacific Ocean: Lessons for Vessel Day Scheme planning. *Mar. Policy* 2010; **34**, 979-87.
- [12] A Harvey. Forecasting, structural time series models and Kalman filter. Cambridge University Press, Cambridge, UK, 1989, p. 14.
- [13] JC Gutierrez-Estrada, C Silva, E Yanez, N Rodriguez and I Pulido-Calvo. Monthly catch forecasting of anchovy *Engraulis ringens* in the north area of Chile: Non-linear univariate approach. *Fish. Res.* 2007; 86, 188-200.
- [14] L Naranjo, F Plaza, E Yanez, MA Barbieri and F Sanchez. Forecasting of jack mackerel landings (Trachurus murphyi) in central-southern Chile through neural networks. *Fish. Oceanogr.* 2015; 24, 219-28.

- [15] S Chesoh and A Lim. Forecasting fish catches in the Songkhla Lake basin. ScienceAsia 2008; 34, 335-40.
- [16] P Komontree, P Tongkumchum and W Karntanut. Trends in marine fish catches at Pattani Fishery Port (1999 - 2003). Songklanakarin J. Sci. Technol. 2006; 28, 887-95.
- [17] N McNeil, P Odton and A Ueranantasun. Spline interpolation of demographic data revisited. *Songklanakarin J. Sci. Technol.* 2011; **33**, 117-20.
- [18] K Watanabe. In-season forecast of chum salmon return using smoothing spline. Fish. Aquac. J. 2016; 7, 173.
- [19] B Lee, D McNeil and A Lim. Spline interpolation for forecasting world tuna catches. *In*: Proceedings of the International Statistical Institute Regional Statistics Conference: Enhancing Statistics, Prospering Human Life, Bali, Indonesia. 2017, p. 801-7.
- [20] N Wongsai, S Wongsai and AR Huete. Annual seasonality extraction using the cubic spline function and decadal trend in temporal daytime MODIS LST data. *Remote Sens.* 2017; 9, 1254.
- [21] B Lee, P Tongkumchum and D McNeil. Forecasting monthly world tuna prices with a plausible approach. *Songklanakarin J. Sci. Technol.* 2020; **42**, 398-405.
- [22] Food and Agriculture Organization (FAO). FishstatJ application version 3.02.0: Global capture production 1950 2015, Available at: http://www.fao.org/fishery/statistics/software/fishstatj/en, accessed March 2017.
- [23] GEP Box and DR Cox. An analysis of transformations. J. Roy. Stat. Soc. Ser. B Methodol. 1964; 26, 211-52.
- [24] GEP Box and GM Jenkins. *Time series analysis: Forecasting and control.* Holden-Day, San Francisco, 1970, p. 47-85.
- [25] WWS Wei. Time series analysis: Univariate and multivariate methods. Pearson Education, Boston, 2006, p. 78-103.
- [26] JB Ramsey. Tests for specification errors in classical linear least squares regression analysis. J. Roy. Stat. Soc. Ser. B Methodol. 1969; **31**, 50-71.
- [27] RS Tsay. Nonlinearity tests for times series. Biometrika 1986; 73, 461-6.
- [28] TH Lee, H White and CWJ Granger. Testing for neglected nonlinearity in time series models: A comparison of neutral network methods and alternative tests. J. Econom. 1993; 56, 269-90.
- [29] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Available at: http://www.R-project.org, accessed March 2017.
- [30] SAS Institute Inc. SAS University Edition, Available at: https://www.sas.com/enus/software/ university-edition.html, accessed June 2017.
- [31] WN Venables and BD Ripley. Modern applied statistics with S. Springer, New York, 2002, p. 397-403.
- [32] S Wold. Spline functions in data analysis. *Technometrics* 1974; 16, 1-11.
- [33] MA Lukas, FR Hoog and RS Anderssen. Efficient algorithms for robust generalized crossvalidation spline smoothing. J. Comput. Appl. Math. 2010; 235, 102-7.
- [34] N Molinari, JF Durand and R Sabatier. Bounded optimal knots for regression splines. *Comput. Stat. Data Anal.* 2004; 45, 159-78.
- [35] B Efron and RJ Tibshirani. An Introduction to the Bootstrap. CRC Press LLC., New York, 1993, p. 45-57.
- [36] S Lek and JF Guegan. Artificial neural networks as a tool in ecological modeling, an introduction. *Ecol. Model.* 1999; **120**, 65-73.

14 of 15

- [37] I Suryanarayana, A Braibanti, RS Rao, VA Ramam, D Sudarsan and GN Rao. Neural networks in fisheries research. *Fish. Res.* 2008; 92, 115-39.
- [38] SS Warren. Neural networks and statistical models. *In*: Proceedings of the 19th Annual SAS Users Group International Conference, SAS Institute Inc., USA. 1994, p. 1538-50.
- [39] GP Zhang. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 2003; **50**, 159-75.
- [40] MS Iyer and RR Rhinehart. A method to determine the required number of neural-network training repetitions. *IEEE Trans. Neural Netw.* 1999; **10**, 427-32.
- [41] Food and Agriculture Organization. Global capture production (online query), Available at: http://www.fao.org/fishery/statistics/global-capture-production/query/en, accessed September 2019.
- [42] Food and Agriculture Organization. Global capture production (online query), Available at: http://www.fao.org/fishery/statistics/global-capture-production/query/en, accessed May 2021.
- [43] Atuna. Industry presses for needed transparency in WCPFC, Available at: http://www.atuna.com/ index.php/en/global-news-registerd/all-atuna-news-registerd, accessed September 2017.
- [44] MN Maunder and AE Punt. A review of integrated analysis in fisheries stock assessment. *Fish. Res.* 2013; **142**, 61-74.