MODELING MAXIMUM MONTHLY TEMPERATURE IN KATUNAYAKE REGION, SRI LANKA: A SARIMA APPROACH

M.C.Alibuhtto[1]&P.A.H.R.Ariyarathna[2]
[1] Department of Mathematical Sciences, Faculty of Applied Sciences, South Eastern University of Sri Lanka
[2] Faculty of Applied Sciences, South Eastern University of Sri Lanka
mcabuhtto@seu.ac.lk, himashrashmika@gmail.com

ABSTRACT: Time series analysis plays a major role in predicting and analyzing climatological data. Temperature is one of the most vital elements of the climate system and modeling of the temperature helps the interested party those who are depending on it directly or indirectly to prepare in advance. The aim of this study is to develop a time series model which can help in improving the predictions of monthly temperature of Katunayake. This paper describes the Box-Jenkins time series seasonal ARIMA model for prediction of monthly maximum temperature in Katunayake region, Sri Lanka. In this study, 181 monthly temperature data were considered during the period 2001-2016. For the model selection, 169 observations were used while the rest of 12 observations were used to validate the model. The results indicate that the SARIMA(3,0,2)(2,0,2)$_{12}$ model was the best model to predict monthly maximum temperature in Katunayake region.

Keywords: Box-Jenkins, Climatology, Katunayake, SARIMA, Temperature

1. INTRODUCTION

The weather of Sri Lanka can be considered as a tropical. Sri Lanka can be divided into two climatological zones such as dry zone and wet zone. The Northern, North Central and Eastern regions of the Sri Lanka are in Meteorological he dry zone and the South Western region is in the wet zone. The warmest average maximum temperature is 31^{o}C in February, March, April and May. The coolest average minimum temperature is 22^{o}C in January and February.

For the study, maximum temperature data were selected from Katunayake area, where it is located in Gampaha district, Western province, Sri Lanka. The Katunayake area is bounded to the Negombo lagoon from one side.

The Meteorological Department in Sri Lanka is the authorized agency to measure, collect and forecast all climatological factors. The head department is located in BauddhalokaMawatha, Colombo 7. There are many branches located all over the country which are taking the climatological observations such as rainfall, temperature and wind speed.

The main objectives of thisresearch were to explore the behaviour of maximum temperature in Katunayake area, develop a time series model for maximum temperature in Katunayake area and validate the developed model for an independent set of data.

2. MATERIALS AND METHODS

2.1 Data Collection

The data were collected from the Metrological Department, Colombo. The selected data set was a complete record of monthly maximum temperature from 2001 to 2016 (181 observations), which was compiled for Katunayake area. The temperature data were measured in Celsius (^{o}C). The Box-Jenkins modelling was done with 169 data set while keeping the rest 12 observations for the validation of the model.

2.2 ACF and PACF

Autocorrelation function and partial autocorrelation function are a type of graphs which containing correlations of different time lags. ACF and PACF can be used to identify the behaviour of the series whether stationary or not and to identify the number of components in ARMA model. The exponentially decaying spikes in ACF and PACF indicate the stationary series.

The number of significant spikes in ACF indicates the number of MA terms in the model and the number of significant spikes in PACF indicates the number of AR terms in the model.

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{n}(x_t \cdot x_{t-k}) - (n-k-1)\bar{x}_t^2}{\sum_{t=k+1}^{n} x_t^2 - (n-k-1)\bar{x}_t^2} \qquad (1)$$

2.3 Autoregressive Moving Average Model

If current x_t value can be expressed as a linear combination of both past p number of x_t values and past q number of e_t values, then it is an auto regressive moving average time series model of p and q. The autoregressive moving average model can be identified from the number of significant spikes of ACF and PACF graph. The ARMA(p,q) model can be written as;

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \beta_1 e_{t-1} + \cdots + \beta_q e_{t-q} + e_t \qquad (2)$$

2.4 Unit Root Test

The Augmented Dickey Fuller unit root test used to find the stationarity of a series. It uses hypothesis as,

$\mathrm{H_0}$: Model has a unit root

$\mathrm{H_1}$: Model does not have a unit root

2.5Autoregressive Integrated Moving Average Model

If the autoregressive moving average model uses a 1st or 2nd order differenced series due to non-stationarity, then that model is an ARIMA model. The number of differences denoted by d. The ARIMA(p,d,q) model can be written as;

$$\phi_p(B)\nabla^d x_t = \alpha_0 + \theta_q(B)e_t \qquad (3)$$

2.6Seasonal Autoregressive Integrated Moving Average Model

If the data series has differentiated into two parts such as seasonally adjusted component and seasonal component due to the existence of seasonality, then the ARIMA model used in both components can be identified as SARIMA model. The ARIMA analysis of seasonally adjusted component is non-seasonal ARIMA and the ARIMA analysis of the seasonal component is seasonal ARIMA component. The SARIMA model denoted by SARIMA(p,d,q)(P,D,Q)s, where first bracket and second bracket denote the seasonally adjusted and seasonal factor series respectively. The SARIMA model can be written as;

$$\phi_p(B)\cdot\Phi_P\left(B^{(s)}\right)\cdot(1-B)^d\cdot\left(1-B^{(s)}\right)^D x_t=\alpha_0+\theta_q(B)\cdot\Theta_Q\left(B^{(s)}\right)\cdot e_t \qquad (4)$$

Where $d-\text{non}-\text{seasonal difference}$

$D-\text{seasonal difference}$

$s\ -\text{number of seasons}$

2.7 Model Selection

The Box-Jenkins methodology uses the Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC) to select the best model among tentative models.

3. Results and Discussion

3.1 Behaviours of temperature data

The average of maximum temperature for the Katunayakeregion from 2001 January to 2016 January was 31.28^{o}C. The minimum and maximum temperature in this time period were29. 32^{o}C and 33.30^{o}C respectively. To identify the behaviour of the data set, the line chart (Figure 1) and the seasonal graph (Figure 2) of the data were obtained. The identifiable seasonality of the data set can be seen from the seasonal graph.

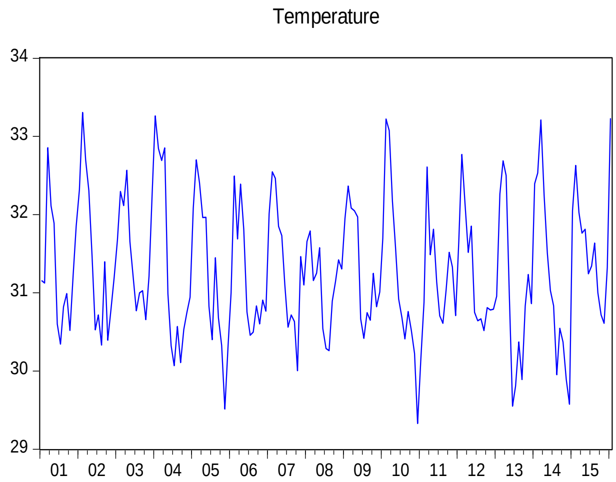

Figure 1: Time series plot of the temperature data

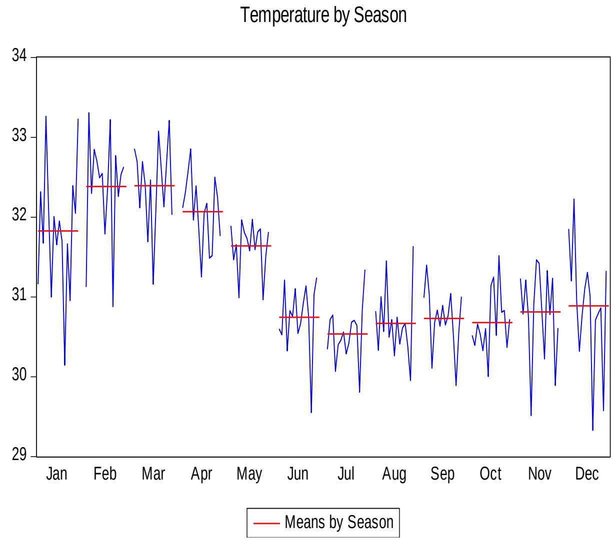

Figure 2: Seasonal plot of the temperature data

The Seasonal ARIMA modelling was done with 169 data which was 2001 January to 2015 January while therest of 12 observations were kept for measuring the accuracy of the forecasted model. According to the figure 2, it can be seen a seasonality of the data exist.

Table 1. Results of Kruskal-Wallis Test.

Kruskal-Wallis statistics	Degrees of freedom	Probability level
130.0749	11	0.0000

Table 1 shows that the existence of a significant seasonality (P=0.000). Therefore the seasonality of the data has been extracted by using Census X-12 seasonal adjustment method. The line chart of the seasonally adjusted series (Figure 3) and seasonal factor series (Figure 4) was obtained to identify the behaviour of the both series.

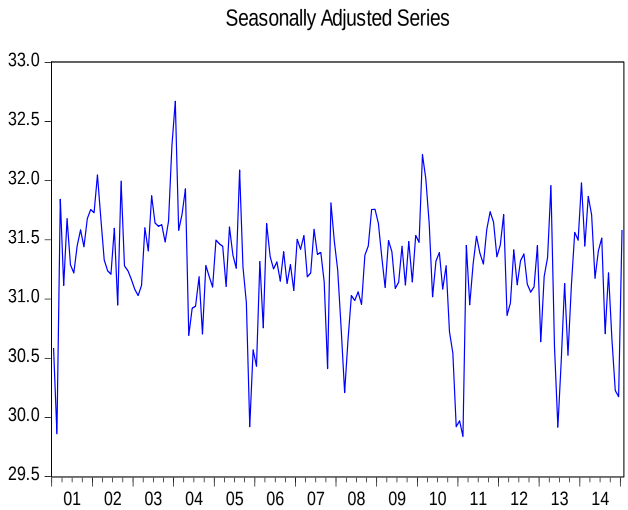

Figure 3: Seasonally adjusted series

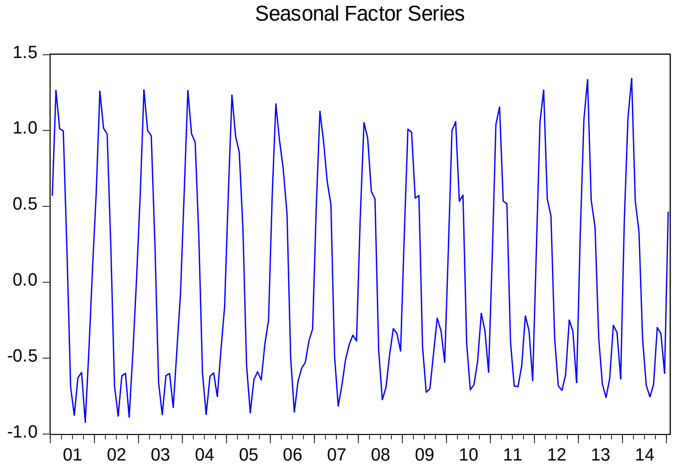

Figure 4: Seasonal factor series

3.2 Model identification and parameter estimates

The stationarity of the both series has checked by using the ADF test. The unit root test for the seasonally adjusted series (P=0.0000) and seasonal factor series (P=0.0000) has confirmed the both series are stationary in 95% confidence level.The ACF and PACF were obtained to identify the number of AR and MA terms in both series.

The model selection criterion was used to identify the best model. Table 2 and Table 3 used to identify the number of components in the seasonally adjusted series and seasonal factor series respectively. AIC, SIC and Durbin Watson (DW) values used for the model selection. The best model can be considered as SARIMA $(3,0,2)(2,0,2)_{12}$because this model has the lowest AIC and SIC values.

Table 2. Parameter estimates of seasonally adjusted series.

Parameters	Significance	Overall sig.	AIC	SIC	DW
AR(1)	0.2477	0.0000	1.1582	1.2693	1.9626
MA(1)	0.8428				
MA(2)	0.5056				
MA(3)	0.9852				
AR(1)	0.0000	0.0000	1.1358	1.2469	1.9528
AR(2)	0.0000				
MA(1)	0.0000				
MA(2)	0.0036				
AR(1)	0.8894	0.0000	1.1657	1.2953	1.9652
AR(2)	0.2498				
MA(1)	0.5657				
MA(2)	0.9785				

MA(3)	0.9964				
AR(1) AR(2) AR(3) MA(1) MA(2)	0.0000 0.0000 0.0001 0.9539 0.9769	0.0000	1.0993	1.2290	1.9760
AR(1) AR(2) AR(3) MA(1) MA(2) MA(3)	0.1577 0.6229 0.2732 0.4446 0.5513 0.5216	0.0000	1.1550	1.3032	1.9433

Table 3. Parameter estimates of seasonal factor series.

Parameters	Significance	Overall sig.	AIC	SIC	DW
AR(1) MA(1) MA(2) MA(3)	0.0000 0.9985 0.9989 0.9994	0.0000	0.84459	0.9557	1.8804
AR(1) AR(2) MA(1) MA(2)	0.0000 0.0000 0.2537 0.5676	0.0000	0.0034	0.1145	1.7048
AR(1) AR(2) MA(1) MA(2) MA(3)	0.0000 0.0000 0.2875 0.8199 0.5274	0.0000	-0.0654	0.0642	2.1587
AR(1) AR(2) AR(3)	0.0000 0.0000 0.3645	0.0000	0.4224	0.5335	2.0529

MA(1)	0.9983				

Table 4. Parameter estimates of SARIMA(3,0,2)(2,0,2)$_{12}$ model.

Variable	Coefficient	Standard error	T-statistics	Probability
C	31.2600	0.0703	44.5511	0.0000
AR(1)	2.1367	0.0016	1301.180	0.0000
AR(2)	-1.6982	0.0005	-3109.156	0.0000
AR(3)	0.4027	0.0007	560.8814	0.0000
SAR(1)	1.0022	0.0196	51.0985	0.0000
SAR(2)	-0.9994	0.0052	-189.4518	0.0000
MA(1)	-1.0240	1.0192	-1.0046	0.3166
MA(2)	0.9995	1.9386	0.5155	0.6069
SMA(1)	-1.7204	1.6347	-1.0524	0.2942
SMA(2)	0.9995	1.8974	0.5267	0.5991

Table 4 shows the estimated coefficients of Fitted SARIMA model.

3.3 Residual analysis

The residual plot (Figure 5) was obtained from the developed model. The normal probability plot (Figure 6) of residuals and histogram(Figure 7) of residuals also were obtained. According to the Figure 6 and Figure 7 residuals follow a normal distribution at 95% confidence level (P=0.523). The model suitability can be observed from the plot of observed and fitted data graph (Figure 8).

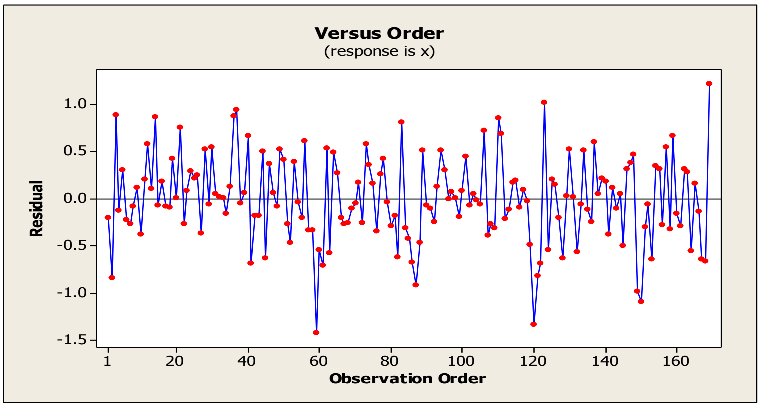

Figure 5: Residual plot

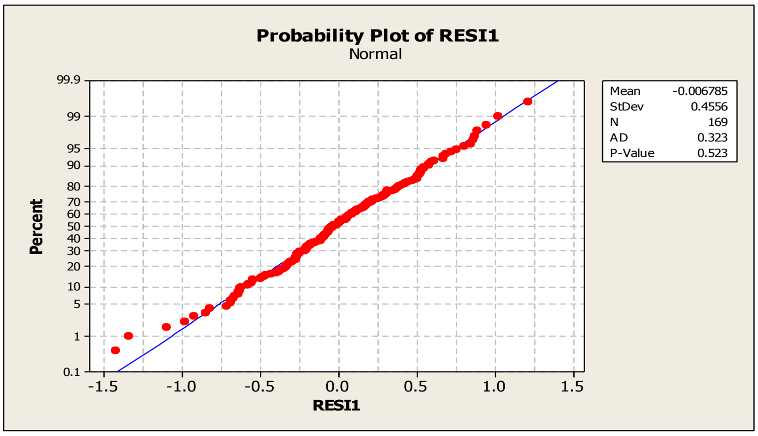

Figure 6: Normal probability plot of residuals

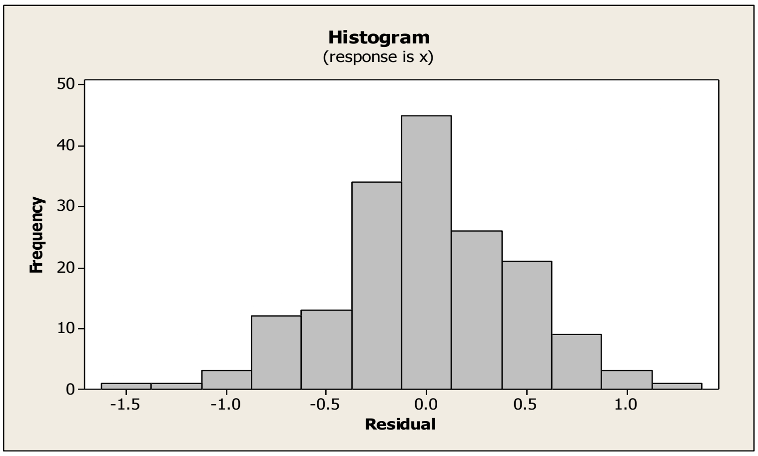

Figure 7: Histogram of residuals

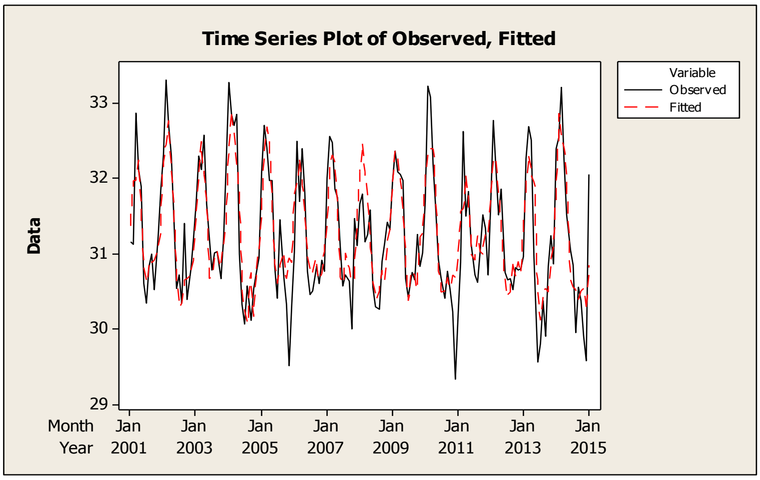

Figure 8: Plot of observed & fitted data

The forecasted values were obtained from the periods of 2015 February to 2016 January using the fitted model and check the prediction accuracy of the fitted model. The results are given in table 5.

Table 5: The observed and forecasted maximum temperature.

Observed temperature	Predicted temperature	Prediction error
32.6286	32.1597	0.4689
32.0290	32.4736	-0.4446
31.7633	32.1770	-0.4137
31.8129	31.8244	-0.0115
31.2433	30.9343	0.309
31.3419	30.7551	0.5868
31.6355	30.9338	0.7017
31.0033	30.8277	0.1756
30.7194	30.9144	-0.195
30.6100	30.9490	-0.339
31.3258	30.8012	0.5246
33.2323	31.7577	1.4746

The error of the prediction was plotted to observe the behaviour (Figure 9). The residuals were distributed normally (P=0.540) which can be observed from the normal probability plot of residuals (Figure 10).

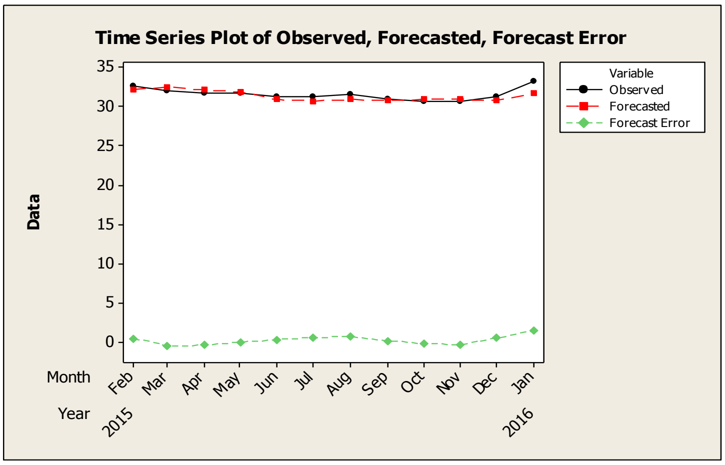

Figure 9: Prediction error plot

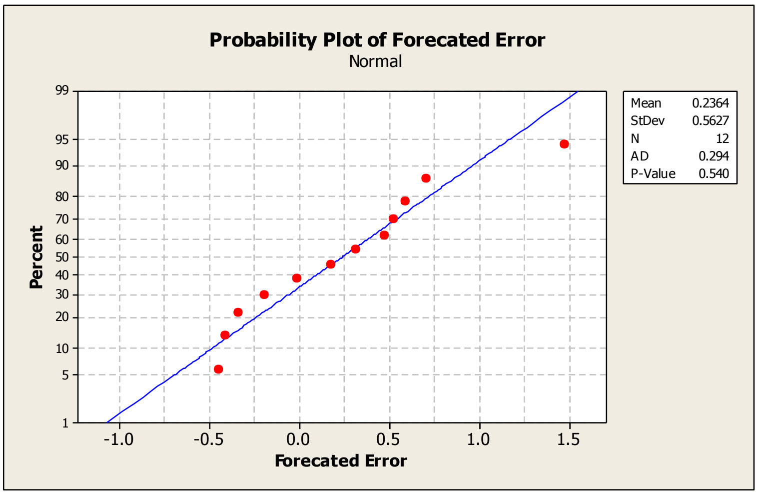

Figure 10: Histogram of prediction error

Table 6: The results of t-test for mean differences

Test of mean difference = 0 (vs. not = 0)	
t-Value = 1.46	p-Value = 0.173

According to the table 6, there is no significant difference between predicted temperature and observed temperatureat 95% confidence level (P=0.173). Thus, the developed model is adequate to predict temperature.

4. CONCLUSION

The aim of this paper is to develop a time series model using recorded monthly temperature data from 2001-2016, of the Katunayake region.This paper has considered the seasonal ARIMAmodelling of Katunayake monthly maximum temperature.Numerous seasonal models were developed using Box-Jenkins methodology and theSARIMA $(3,0,2)(2,0,2)_{12}$ model was selected as the best model by using model selection criteria. Furthermore, the fitted model is confirmed and validated by the residual tests. The fitted SARIMA model is very useful to forecast future values of temperature in the Katunayakeregion.

REFERENCES

Kaushik, I. and Singh, S.M., (2008). Seasonal ARIMA model for forecasting of monthly rainfall and temperature. *Journal of Environmental Research and Development Vol*, *3*(2).

Khajavi, E., Behzadi, J., Nezami, M.T., Ghodrati, A. and Dadashi, M.A., (2012). Modeling ARIMA of air temperature of the southern Caspian Sea coasts. *Intl Res J Appl Basic Sci*, *3*(6), pp.1279-1287.

Nury, A.H., Koch, M., Alam, M.J.B. (2013). Time Series Analysis and Forecasting of Temperatures in the Sylhet Division of Bangladesh. In *4th international conference on environmental aspects of Bangladesh (ICEAB),* pp. 24-26.

Sarraf, A., Vahdat, S.F. and Behbahaninia, A., (2011). Relative Humidity and Mean Monthly Temperature Forecasts in Ahwaz Station with ARIMA Model in Time Series Analysis. In *Proceedings of International Conference on Environment and In-dustrial Innovation* (Vol. 6).

Shamsnia, S.A., Shahidi, N., Liaghat, A., Sarraf, A. and Vahdat, S.F., (2011). Modeling of weather parameters using stochastic methods (ARIMA model)(case study: Abadeh Region, Iran). In *International Conference on Environment and Industrial Innovation. IPCBEE* (Vol. 12).

Stein, M. and Lloret, J., (2001). Forecasting of air and water temperatures for fishery purposes with selected examples from Northwest Atlantic. *Journal of Northwest Atlantic Fishery Science*, *29*, pp.23-30.