

Predictive modeling of pelagic fish catch in malaysia using seasonal ARIMA models

Hadiza Yakubu Bako¹, Mohd Saifullah Rusiman², Ibrahim Lawal Kane³,
Hazel Monica Matias-Peralta¹

¹Department of Technology and Heritage, Faculty of Science, Technology and Human Development, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia

²Department of Mathematics and Statistics, Faculty of Science, Technology and Human Development, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia

³Department of Mathematics and Computer Science, Umaru Musa Yar'adua University, 2218, Katsina State, Nigeria

Email address:

hadizayakububako@yahoo.com (H. Y. Bako), saifulah@uthm.edu.my (M. S. Rusiman)

To cite this article:

Hadiza Yakubu Bako, Mohd Saifullah Rusiman, Ibrahim Lawal Kane, Hazel Monica Matias-Peralta. Predictive Modeling of Pelagic Fish Catch using Seasonal ARIMA Models. *Agriculture, Forestry and Fisheries*. Vol. 2, No. 3, 2013, pp. 136-140.

doi: 10.11648/j.aff.20130203.13

Abstract: Fish catch prediction is an important problem in the fisheries sector and has a long history of research. The main goal of this paper is to create a model and make predictions using fish catch data of two fish species. Among the most effective and prominent approaches for analyzing time series data is the methods introduced by Box and Jenkins. In this study we applied the Box-Jenkins methodology to build Seasonal Autoregressive Integrated Moving Average (SARIMA) model for monthly catches of two fish species for a period of five years (2007 – 2011). The seasonal ARIMA (1, 1, 0)(0, 0, 1)₁₂ and SARIMA (0, 1, 1) (0, 0, 1)₁₂ models were found fit and confirmed by the Ljung-Box test and these models were used to forecast 5 months upcoming catches of *Trichiurus lepturus* (Ikan Selayor) and *Amblygaster leiogaster* (Tambun Beluru) fish species. The result will help decision makers to establish priorities in terms of fisheries management.

Keywords: Box-Jenkins, SARIMA, Ljung-Box, Fish Catches, *Trichiurus Lepturus*, *Amblygaster Leiogaster*

1. Introduction

The fisheries sector plays a vital role in the Malaysian national economy. It contributes to the national gross domestic product (GDP), as it also a source of employment, foreign exchange and a cheap source of protein supply in the country. Fish constitutes 60-70% of the national animal protein intake, with per capita consumption of 47.8 kg per year. The marine capture fishery in Malaysia covers a total area of 547,200 km² which is categorized into coastal fisheries and deep-sea fisheries [1]. *Trichiurus lepturus* and *Amblygaster leiogaster* are among the most important fish species found in Malaysian waters. In fact, the total catch for these species alone contributes 1.57% to GDP and which also provides employment for more than 79,000 fishermen and 20,000 fish farmers [1].

Modeling and forecasting fish catches have a long history of research, for example, Efthymia [2] modeled and forecasted the monthly pelagic production of fish species in the Mediterranean Sea during 1990–2005 using the

univariate and multivariate autoregressive integrated moving average (ARIMA) models. Stergiou et al. [3] indicated that seasonal ARIMA(1,0,1)(0,1,2)₁₂ models fitted and forecasted the monthly pelagic fish catches in Hellenic waters, the model provided low value of Bayesian information criteria (BIC), medium standard error and mean absolute percentage error. Chi-Lu and Su-Zan [4] evaluated the efficiency of Box and Jenkins ARIMA model in the short-term forecasting of the fishery, using the time series of monthly catch and the monthly catch per unit effort (CPUE) of the South Atlantic albacore harvested by the Taiwanese long-line fishery from January 1968 to December 1993. The model provides a useful measure of the likely ranges of catch and CPUE for the management of albacore stock in the South Atlantic Ocean. In a study by Georgakarakos et al. [5], time series analysis techniques (ARIMA models), artificial neural networks (ANNs) and Bayesian dynamic models were used to forecast annual Loliginid and Ommastrephid landings recorded from the most important fishing ports in the Northern Aegean Sea

(1984–1999). The techniques were evaluated based on their efficiency to forecast and their ability to utilize auxiliary environmental information. Applying a “stepwise modeling” technique, namely by adding stepwise predictors and comparing the quality of fit, certain inferences concerning the importance of the predictors were made. The ARIMA models predicted the test data remarkably well, after they were first differenced to obtain stationarity. The time series approach is generally preferred for its fitting and forecasting accuracy for fisheries production [6-7, 2]. In Malaysia, fish catch statistics has only recently been given attention by the research bodies. As such, in 2008, Shitan, et al. [8] has used the time series analysis to forecast demersal and pelagic marine fish production in Malaysia. Using ARFIMA and ARIMA they build a model to predict demersal and pelagic marine fish catch, respectively. Unfortunately, no other predictive model has been reported afterwards. Therefore, the objective of this paper is to build a model to predict monthly catch of five years fish data following a different methodology.

2. Materials and Methods

2.1. Data Collection and Analysis

The time series data analyzed in this paper corresponds to the monthly fish catch of *Trichiurus lepturus* and *Amblygaster leiogaster* for the period of five years 2007-2011 (<http://www.dof.gov.my/en/fishery-statistics>) [9]. *Trichiurus lepturus* Linnaeus, 1758 is a member of the cutlassfish family, Trichiuridae, it is a long, slender fish found throughout the tropical and temperate waters of the world. They grow to over 2m in length and can reach 5.0 kg in 15 years [10]. *Amblygaster leiogaster* (Valenciennes, 1847) is from clupeidae family and it currently contains three species, they grow from 18cm to 23cm [11]. Total fish catch data are catches using various methods including purse-seine, beach seine and long line methods, and they are measured in tons [1].

In this paper we tend to develop a suitable forecasting models Box-Jenkins methodology, because it is flexible enough to allow choosing the best model out of complete model using a designated and acceptable criteria, and it has the ability for error measurement. All statistical analysis were run using R software.

2.2. Time Series Decomposition

A time series is a stochastic procedure that describes the evolution of the random variable. It consists of four different components as follows [12];

- Trend (T): A trend is a long-term component that represents a growth or a decline of a time series over an extended period of time.
- Seasonal component (S): This term of seasonality is used for time series defined at time intervals which are fractions of a year. It is a pattern of change that repeats itself from year to year.

- Cyclical component (C): Changes in time series sometimes show a wave-like fluctuation around a trend, which shows the possible existence of periodicity with longer intervals.
- Irregular component (\mathcal{E}): This is a part of a time series represented by residuals, after the above-mentioned components have been removed.

Any time series data that contained the above components is usually described using the following in (1).

$$y_t = T_t + S_t + C_t + \varepsilon_t. \quad (1)$$

2.3. Box-Jenkins Methodology

Box-Jenkins [13] proposed a methodology for modeling and forecasting time series data, wherein the models are the family of autoregressive integrated moving average models, or ARIMA models. These are linear, stochastic models that can describe fairly complex dynamics in a time series. Basically, there are three phases included in Box-Jenkins Methodology: model identification, parameter estimation, and diagnostic checking.

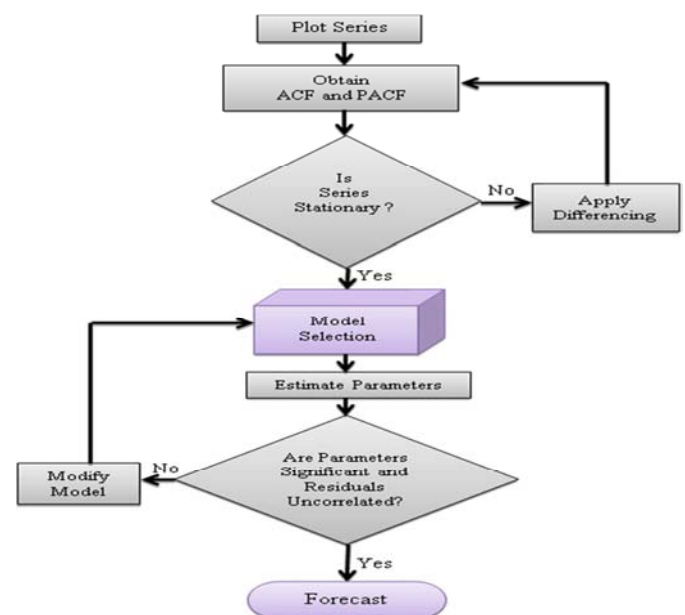


Figure 1. Box-Jenkins model building flow chart.

Note: ACF: Autocorrelation Function; PACF: Partial Autocorrelation Function.

2.4. Stationarity

In order to use any time series data for analysis using Box-Jenkins methodology, the time series should meet the stationarity conditions. A time series X_t is said to be weakly stationary or wide-sense stationary or covariance stationary if it fulfills the following three properties [14].

1. Mean is constant over time: $E[X_t] = \mu$, for all t .
2. Variance is constant over time:

$$Var[X_t] = E[(X_t - \mu)^2] = \sigma_x^2 \text{ for all } t. \quad (2)$$

3. Covariance between any two values of the series depends only on their distance apart in time (k) not on their absolute location in time $\{t\}$;

$$\text{Cov}[X_t, X_{t-k}] = E[(X_t - \mu)(X_{t-k} - \mu)] = \gamma(k) \quad (3)$$

2.5. Seasonal ARIMA Model

A time series is said to be seasonal if there exists a tendency for the series to exhibit a periodic behavior after certain time interval. The usual ARIMA models cannot really cope with seasonal behavior; it only model time series with trends. Seasonal ARIMA models are formed by including an additional seasonal terms in the ARIMA models and are defined by seven parameters. Following Fadhilah & Ibrahim [12], the seasonal ARIMA denoted by ARIMA (p, d, q)(P, D, Q)s is given as in (4).

$$\begin{aligned} & (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - \beta_1 B^s - \beta_2 B^{2s} - \dots - \beta_P B^{Ps}) \\ & (1 - B)^d (1 - B^s)^D y_t \\ & = c + (1 - \psi_1 B - \psi_2 B^2 - \dots - \psi_q B^q)(1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_Q B^{Qs}) \varepsilon_t \end{aligned} \quad (4)$$

where

$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ and $(1 - \beta_1 B^s - \beta_2 B^{2s} - \dots - \beta_P B^{Ps})$ are the autoregressive part of order p and seasonal autoregressive part of order P respectively. $(1 - \psi_1 B - \psi_2 B^2 - \dots - \psi_q B^q)$ and $(1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_Q B^{Qs})$ represents the moving average part of order q and seasonal moving average part of order Q respectively. $(1 - B)^d$ is the differencing polynomial of order d and $(1 - B^s)^D$ the seasonal differencing polynomial of order D . Finally, s is the period of the seasonal pattern appearing. The seasonal ARIMA is to look at what are the best explanatory variables to model a seasonal pattern.

2.6. Ljung-Box Test

Ljung and Box [15] proposed a Q-Test called Ljung-Box test which is commonly used in linear models following Box-Jenkins methodology. This test is applied to the residuals of a fitted model, not the original series, and in such applications the hypothesis to be tested is that the residuals from the model have no autocorrelation. Perhaps it performs a lack-of-fit hypothesis test for model misspecification based on the Q -statistic given as:

$$Q = N(N+2) \sum_{j=1}^L \frac{\hat{\rho}_j^2}{(N-j)}$$

Where N = sample size, L = number of autocorrelation lags included in the statistic, and $\hat{\rho}_j^2$ is the squared sample autocorrelation at lag j . Under the null hypothesis of no serial correlation, the Q -test statistic is asymptotically Chi-Square distributed. The p-values above 0.05 indicate the acceptance of the null hypothesis of model accuracy under 95% significant levels [17].

3. Results and Discussion

The time series plots of the catches of Selayor (*Trichiurus lepturus*) and Tamban Beluru (*Amblygaster leiogaster*) show no stability over time (Figure 2). In modeling Seasonal ARIMA processes, the first step is to determine whether the time series is stationary or non-stationary. There exist several tests for stationarity among others are the ADF, PP and KPSS. In this paper we relaxed such test, and alternatively we observed the autocorrelation function (ACF) depicted (Figure 3). The autocorrelation that provide initial information relevant to the internal organization of time series data gives clear evidence of non stationarity in the time series considered by displaying a hyperbolic decay pattern rather than exponential pattern.

The time series decomposition plots (Figure 4) separate the series into its constituent components i.e, the estimated trend component and the estimated seasonal component. The plots show that the estimated trend component follows a declining trend since 2011 and some noticeable seasonal component also displayed. Since all the time series are non-stationary, they have to be transformed into a stationary time series by applying the appropriate order of differencing d or D . The time series plot of the first difference series of monthly fish catch data of *Trichiurus lepturus* and *Amblygaster leiogaster* species are made stationary by observing pattern with zero mean and constant variance in the plot (Figure 5). The ACF plot (Figure 6) decays rapidly which confirmed the stationarity assumption (Figure 5).

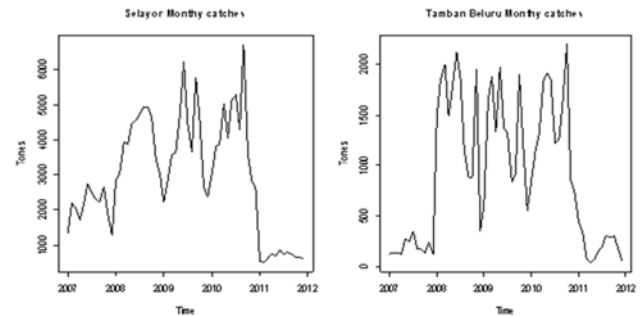


Figure 2. Time series plots of monthly catches of ikan Selayor and ikan Tamban Beluru

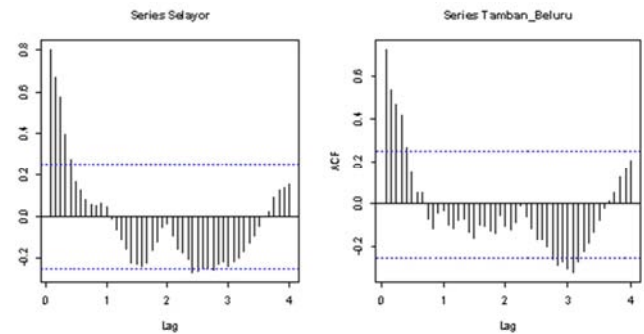


Figure 3. Autocorrelation function of monthly fish catch data of ikan Selayor and ikan Tamban Beluru

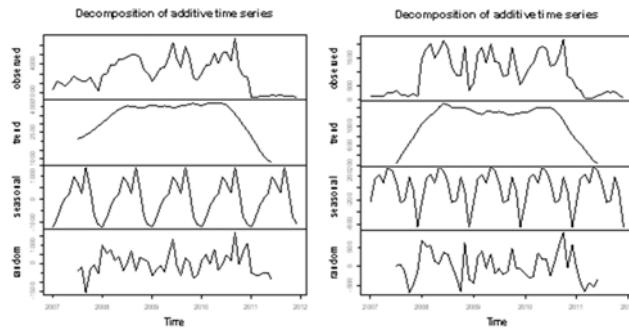


Figure 4. Time series decomposition plots separating trend, seasonal and random components from the monthly fish catch data of ikan Selayor and ikan Tamban Beluru

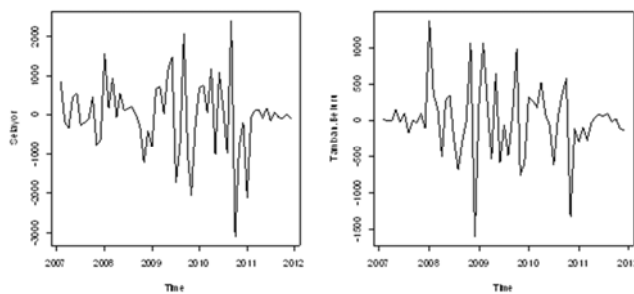


Figure 5. Time series plot of the first difference series of monthly fish catch data of ikan Selayor and ikan Tamban Beluru

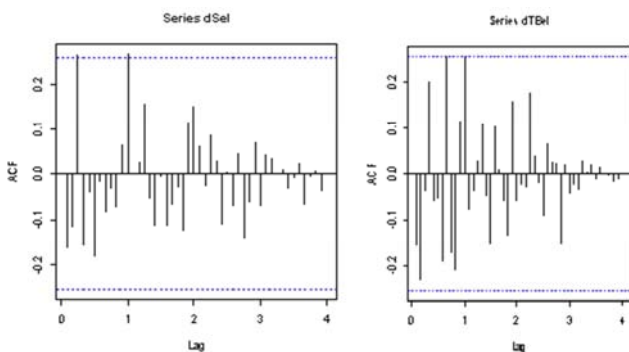


Figure 6. Autocorrelation function (ACF) of the first difference series of monthly fish catch data of ikan Selayor and ikan Tamban Beluru

After first differencing, the time series considered appears to be quite stable over time (Figure 5). The orders p , q , P and Q of the seasonal ARIMA models were identified and estimated for both the stationary series following Box and Jenkins methodology. Model with low Akaike information criteria (AIC) is selected based on Akaike information criteria [17], which is a common procedure in Box-Jenkins modeling (Table 1). Only the best models were displayed. The best fitted models for *Trichiurus lepturus* and *Amblygaster leiogaster* are $SARIMA(1,1,0)(1,0,0)_{12}$ and $SARIMA(0,1,1)(0,0,1)_{12}$ respectively. Before the interpretation and use of the model, we are to look and check whether the models are specified correctly. Tests for residual autocorrelation (AC) are prominent tools for this task. A well-known example is the Ljung-Box test for residual

autocorrelation. If the residuals are correlated, then the model should be refined. Otherwise, the residuals are white noise and the model is adequate to represent the time series [18]. The Ljung-Box test is applied to the residuals of the fitted models (Table 2). The results showed that all P-values at lags 12, 24, 36 and 48 for both models exceeds 0.05 which indicates acceptance of models accuracy at 95% significant levels (Table 2). Since forecasting future catch is one of the main reasons for developing time series model, a 5 months forecast using the developed models was presented (Figure 7).

Table 1. Summary results for the fitted seasonal ARIMA models

Fish type	Selayor		Tamban Beluru	
Model	$SARIMA(1,1,0)(1,0,0)_{12}$		$SARIMA(0,1,1)(0,0,1)_{12}$	
Parameter	Estimate	S.E	Estimate	S.E
Ar1	-0.2035	0.1269	-	-
Ma1	-	-	-0.3219	0.1388
Sar1	0.2912	0.1225	-	-
Sma1	-	-	0.3608	0.1502

Table 2. Ljung-Box test results for residuals from fitted Seasonal ARIMA model

Lags	$SARIMA(1,1,0)(1,0,0)_{12}$		$SARIMA(0,1,1)(0,0,1)_{12}$	
	Chi-Square	P-value	Chi-Square	P-value
12	9.2355	0.5099	13.2267	0.2113
24	18.5434	0.6733	18.1341	0.6980
36	27.0698	0.7948	31.1556	0.6078
48	31.3229	0.9515	33.1964	0.9211

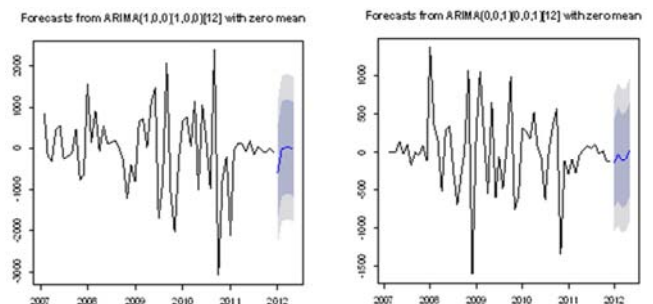


Figure 7. Five month out-sample forecast from the fitted seasonal ARIMA models

4. Conclusion

In this paper the Box-Jenkins methodology was used to build Seasonal Autoregressive Integrated Moving Average (SARIMA) model for monthly catches of two fish species found in Malaysian waters for the period from 2007 – 2011, and it has proven successful in describing and forecasting the monthly fishery dynamics of *Trichiurus lepturus* and *Amblygaster leiogaster*. The seasonal ARIMA (1, 1, 0) (0, 0, 1)₁₂ and SARIMA (0, 1, 1)(0, 0, 1)₁₂ models were used to forecast a five-month upcoming catches of Selayor (*Trichiurus lepturus*) and Tamban Beluru (*Amblygaster leiogaster*). The information provided by the model provides estimated amount of the monthly fish catches, enough to provide sufficient information that can help the decision makers to establish priorities and

strategies for proper fish management in the Malaysian fishery sector. Further research can be conducted to compare the forecast ability of the model with other time series models like ARFIMA, exponential smoothing, ANN etc.

Acknowledgements

The research work is supported by ERGS (Exploratory Research Grant Scheme) grant (vot E020), Ministry of Higher Education, Malaysia.

References

- [1] Jabatan Perikanan Malaysia, "Annual Fisheries Statistics 2003 Jilid 1," 2003, Jabatan Perikanan Malaysia.
- [2] Efthymia, V. T., Christos, D. M. and John, H. "Modeling and forecasting pelagic fish production using univariate and multivariate ARIMA models". *Fisheries Science*; 73: 5; 979–988, 2007
- [3] Stergiou, K.I., Christou, E.D. and Petrakis, G. "Modelling and forecasting monthly fisheries catches: comparison of regression, univariate and multivariate time series methods". *Fish Res.*; 29, 55–95, 1997.
- [4] Chi-Lu, S. and Su-Zan, Y. "ARIMA Modeling and Forecasting the Albacore Catch and CPUE of the Taiwanese Tuna Longline Fishery in the South Atlantic Ocean." <http://sol.oc.ntu.edu.tw/aot/1997/362/E362D1e.html>, 1998.
- [5] Georgakarakos, S., Koutsoubas, D. and Valavanis, V. "Time series analysis and forecasting techniques applied on loliginid and ommastrephid landings in Greek waters". *Fisheries Research* 78, 55–71, 2006.
- [6] Lloret, J., Lleonart, J. and Sole, I. "Time series modeling of landings in Northwest Mediterranean Sea". *Journal of Marine Science*; 57, 171–184, 2000.
- [7] Defrancesco, E. and Samuele, T. "An activity-based decision support system to evaluate the economic viability of fisheries management tools for the small pelagic species in the northern Adriatic Sea". http://www.iamb.it/share/img_new_medit_articoli/426_12defrancesco.pdf, 2012.
- [8] Shitan, M., Jin Wee P.M., Ying Chin, L. and Ying Siew, L. "Arima and Integrated Arfima Models for Forecasting Annual Demersal and Pelagic Marine Fish Production in Malaysia". *Malaysian Journal of Mathematical Sciences*, 2:2, 41–54, 2008. (<http://www.dof.gov.my/en/fishery-statistics>).
- [9] Trichiurus. <https://en.wikipedia.org/wiki/largeheadhairtail>.
- [10] Reece. kahaku.go.jp/zoology/fishes-ofAndaman-sea/content/clu/pedae/01.html
- [11] Fadhilah, Y. & Ibrahim, L. K. "Modeling Monthly Rainfall Time Series Using ETS State Space and SARIMA Models. *International Journal of Current Research*. Vol. 4, Issue, 09, 195–200, 2012a.
- [12] Box, G.E.P., Jenkins G.M. "Time series analysis: forecasting and control. Holden-Day, Boca Raton, 1976.
- [13] Giraitis, L., Leipus, R. and Philippe, A. "The Test for Stationarity versus Trends and Unit Roots Errors". Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom, 2–3, 2002.
- [14] Ljung, G.M., Box, G.E.P "On a measure of a lack of fit in time series models". *Biometrika* 65(2): 297–303. doi: 10.1093/biomet/65.2.297, 1978
- [15] Wang, W., Van Gelder, P. H. A. J. M., Vrijling, J. K., and Ma, J. "Testing and modelling autoregressive conditional heteroskedasticity of streamflow processes", *Nonlin. Processes Geophys.*, 12, 55–66, doi:10.5194/npg-12-55-2005, 2005.
- [16] Laurini M. P. and Portugal, S. P. (2003). Long memory in the R\$/US\$ exchange rate: A robust Analysis, Finance lab working paper 3.
- [17] Fadhilah, Y. & Ibrahim, L. K. "Volatility modeling of rainfall time series". *Theor Appl Climatol.* DOI 10.1007/s00704-012-0778-8, 2012b.