5th Meeting of the Scientific Committee
Shanghai, China, 23 - 28 September 2017

SC5-SQ04
Morphological variation and stock classification of Jumbo flying squid based on the statolith shape using wavelet analysis

Zhou Fang, Xinjun Chen, Hang Su & Yong Chen
Morphological variation and stock classification of jumbo flying squid, *Dosidicus gigas*, based on the statolith shape using wavelet analysis

Zhou Fang, Xinjun Chen, Hang Su, Yong Chen
National Data Center for Distant-water Fisheries, Shanghai Ocean University

**Introduction**

Oceanic squid, especially ommastrephid species, plays an important role in marine ecosystems, both as prey and predators (Navarro *et al*., 2013). They prefer to gather in areas with high primary productivity (Ichii *et al*., 2002; Waluda and Rodhouse, 2006). The dramatic fluctuations of annual landings for these species indicate that they are heavily susceptible to the surrounding environmental changes (Anderson and Rodhouse, 2001; Rodhouse 2001; Zeidberg and Robison, 2007; Keyl *et al*., 2011; Robinson *et al*., 2013; Yu *et al*., 2016). The local population, or stock, also possess distinct biological characteristics, reflecting genetic structures and experienced ambient environment (Ichii *et al*., 2002; Rocha and Vega, 2003; Sandoval-Castellanos *et al*., 2007, 2010; Arkhipkin *et al*., 2014; Liu *et al*., 2015a).

Jumbo flying squid, *Dosidicus gigas*, mainly located in regions with comparatively high productivity that is induced by two main upwelling zones (Ichii *et al*., 2002; Lavín *et al*., 2006; Radenac *et al*., 2012), are the target of commercial fisheries for Asian and Latin American countries in east Pacific Ocean (Chen *et al*., 2008; Ibáñez and Poulin, 2014; Ibáñez *et al*., 2015; Guevara-Carrasco *et al*., in press).
This largest ommastrephid species is highly mobile and often migrate on a large scale (Keyl et al., 2008; Rodhouse, 2008). Historically, the general habitat for *D. gigas* ranged from 30°N to 25°S, with occasional occupation at 40°N in the eastern tropical Pacific (ETP) (Nigmatullin et al., 2001). However, this species has undergone an unusual change in distribution since the severe El Niño and La Niña event during 1997-1999 (Waluda et al., 2006; Arkhipkin et al., 2014), including an expansion of habitat stretched to Alaska in the north and Antarctic in the south (Keyl et al., 2008; Hoving et al., 2013; Stewart et al., 2014). The behavior of these extreme generalists seeking for suitable habitat strengthens the connectivity among different populations, potentially influencing the intrinsic population structure of *D. gigas* (Staaf et al., 2010; Morales-Bojórquez and Pacheco-Bedoya, 2016; Sanchez et al., 2016).

It is critically important to fully understand the population structure of commercial marine organisms for investigating the population dynamics and developing management strategies (Thorrold et al., 2001; Gillanders, 2002). The population structure of *D. gigas* is complicated and usually defined using mantle length size (Nigmatullin et al., 2001) where the small-sized group is thought to be restricted to equatorial waters; the medium-sized group has been found in the entire range of the species; and large-sized group only occurs in the northern and southern peripheries of its ranges. These size based sub-populations are regarded as distinct stocks (Nigmatullin et al., 2001). The seasonal schooling of *D. gigas* in different fishing grounds can be regarded as separated geographical stocks (Liu et al., 2015). However, some researchers have indicated that *D. gigas* should be regarded as two
separate stocks, “northern” and “southern”, based on supposed migration patterns that only occur northern hemisphere or south hemisphere, respectively (Nesis, 1983; Wormuth, 1998; Clarke and Paliza, 2000; Staaf et al., 2010). This perspective is supported by recent RAPD-genetic and mitochondrial DNA analysis, which concluded that samples from the two hemispheres are genetically distinct (Sandoval-Castellanos et al., 2007, 2010). The results of elemental signatures in the statolith nucleus (embryonic stage) and postnucleus (paralarval stage) also showed significant differences among different geographic locations, representing the northern and southern stocks (Liu et al., 2015a).

Cephalopod statolith has been considered as an archival tissue that records valuable ecological information during the life history within a stable configuration (Lipiński, 2001; Arkhipkin, 2005). Like fish otolith, cephalopod statolith has been widely used for age estimation via the daily deposited increments as observed under microscope (Lipiński, 1981, 1986, 1993; Arkhipkin and Shcherbich, 2012). Statoliths embedded in the statocyst also have unique characteristics of shape (Clarke, 1978). Droplet-shaped statoliths are usually anomalous, compared to the oval-shaped otolith in most fish species (Clarke, 2003), although they have similar function to keep balance (Budelmann, 1975; Lecomtefiniger, 1999). Arkhipkin (2005) reported that the features of statoliths in different sections and modified the terminology based on previous studies (Clarke, 1978; Lipiński, 1993). Genetic structure and environmental factors (e.g. temperature, salinity) during ontogenetic processes are the two main determinants that influence the statolith shape variation in species and ontogenetic
levels (Thomas and Moltschaniwskyj, 1998; Lipiński, 2001; Arkhipkin, 2005).

Several studies of statolith morphometrics have been proposed mostly based on measurements of linear distances (Clarke, 1978; Arkhipkin and Murzov, 1986; Flores and Garland, 2002; Barcellos and Gasalla, 2015). This method, however, has a degree of uncertainty and may increase the risk of potential bias of measurements (Francis and Mattlin, 1986; Bookstein, 1998; Adams et al., 2004). Geometric morphometrics (GM), which can effectively represent and reconstruct the shape of object, has been proposed and developed for more than two decades (Adams et al., 2013). This powerful method, including the landmark (Cadrin, 2013) and outline analyses (Rohlf and Marcus, 1993; Marcus et al., 1996), has been widely used in stock and population identification based on fish otoliths (Stransky et al., 2008; Farias et al., 2009; Crespo et al., 2012). There are limited reports for cephalopod statolith shapes, except for genera and species discrimination for some local species in Europe (Lombarte et al., 1997; 2006; Dommergues et al., 2000).

For outline analysis, Fourier transform (FT) analysis (including Elliptical Fourier Analysis, EFA and Fast Fourier Analysis, FFA) is more commonly used due to its well-known mathematic background that serves to characterize statolith shape with a series of harmonic expansions (Lombarte et al., 1997; Green et al., 2015). It also provides a representation of decomposing the curve in the form of sines and cosines, which are problematic to capture the localized sharp irregularities and different singularities (Libungan et al., 2015). This can therefore weaken the description of morphology for sharp margins (Graps, 1995; Parisi-Baradad et al., 2005). The
Wavelet transform (WT), with the ability to measure sharp transitions and establish locations that effectively and adequately represent morphological characteristics, providing a more powerful alternative to solve the problems inherent with FT analysis (Parisi-Baradad et al., 2005). Previously, WT was found to be an effective tool for species discrimination and stock identification (Lombarte et al., 2006; Libungan et al., 2015, 2016).

In this study, the squid samples commercial caught in four main fishing grounds of *D. gigas* was collected, the effect of wavelet analysis is evaluated to identify different stocks of *D. gigas* using statolith shapes. A large number of wavelet variables were automatically extracted with statistical software and then described positional differences and inter-population variations were derived from the shape of statolith edge. Different partitioning approaches were also compared to improve classification rate and select an appropriate method for stock discrimination. This study aims to enhance the understanding of *D. gigas* statolith morphological features in different stocks and give us an alternative method for stock classification. The statolith shape description and follow-up discriminating method in this study are also applicable to the investigation of other cephalopod species.

**Materials and methods**

**Sampling**

A total of 350 *D. gigas* squid samples were randomly obtained on the deck from the Chinese commercial jigging vessel *F/V Zheyuanyu 807*, which operated off the
Chilean Exclusive Economic Zone (75°-80°W and 20°-24°S) in 2008; the *F/V New Century 52*, which operated off the Peruvian Exclusive Economic Zone (82°-86°W and 10°-13°S) in 2009; the *F/V Puyuan 802*, which operated off the equatorial area (114°-119°W and 2°N-3°S) in 2013; and the *F/V Fenghui 16*, which operated off the Costa Rica Dome (90°-98°W and 6°-10°N) in 2009 (Table 1, Fig. 1). All samples were immediately frozen at -18°C on deck.

**Table 1 Sample information of *D. gigas* in this study**

<table>
<thead>
<tr>
<th>Location</th>
<th>Code</th>
<th>Time</th>
<th>Fishing vessel</th>
<th>Coordinates</th>
<th>N</th>
<th>ML (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[Mean (range)]</td>
</tr>
<tr>
<td>Chile</td>
<td>CH</td>
<td>May-Jul. 2008</td>
<td><em>Zheyuanyu 807</em></td>
<td>75°-80°W, 20°-24°S</td>
<td>103</td>
<td>395.6(312.4-517.1)</td>
</tr>
<tr>
<td>Equator</td>
<td>EQ</td>
<td>Apr.-Jun. 2013</td>
<td><em>Puyuan 802</em></td>
<td>115°-119°W, 2°-1°S</td>
<td>103</td>
<td>296.0(226.7-397.5)</td>
</tr>
<tr>
<td>Peru</td>
<td>PE</td>
<td>Aug.-Sep. 2009</td>
<td><em>New Century 52</em></td>
<td>82°-85°W, 10°-13°S</td>
<td>102</td>
<td>342.4(220.7-494.5)</td>
</tr>
</tbody>
</table>
Fig 1 Sampling locations for *Dosidicus gigas* and related surface currents in eastern Pacific Ocean. NEC: North Equatorial Current; NECC: North Equatorial Countercurrent; SEC: South Equatorial Current; PC: Peru Current; CCR: Countercurrent Ridge; CCT: Countercurrent Trough; ER: Equatorial Ridge. Revised from Ichii et al (2002)

Dorsal mantle length (ML) was measured to the nearest 0.1 cm in the laboratory after samples were thawed. Sex maturity stage was evaluated by naked eyes following Lipinski and Underhill (1995). Maturing and matured squid (stage III to stage V) was selected for the future analysis to avoid the effect of allometric growth. Finally, 260 complete statoliths were extracted from the statoecyst and preserved in a 0.5 ml centrifuge tube. As the left and right statoliths were tested with no significant difference in morphometrics (t-test, P>0.05), only the right statolith was used in the
latter process.

**Image capture and shape analysis**

Images of the 260 statoliths were taken with Charge-coupled Device (CCD) with stereomicroscope (Olympus Corp. Tokyo). The contour of statolith in every picture should be clear to ensure the margin of the statolith visibly. The wavelet coefficients were obtained using the wavethresh package (Nason, 2012) and scaled by adjusting for allometric relationships with ML as in Lleonart *et al.* (2000) to avoid the influence of squid size. The image analysis was conducted with R 3.1.3 (R Core Team, 2015) using the package shapeR (Libungan and Pálsson, 2015) to reconstruct the outline of statolith with wavelet analysis.

A series of size variables concerning the shape of the statolith were automatically calculated by shapeR (Libungan and Pálsson, 2015) based on the digitized images and MLs and used to calculate the shape indices according to Tuset *et al.* (2003) (Table 2). The relationship between ML and statolith length were assess the effect of body size on statolith shape using linear regression. And the stock differences of statolith growth were evaluated using ANCOVA. Statolith shape indices were analyzed using an ANOVA test to assess the stock variation of *D. gigas*. In order to detect the section variation in statolith shape, the wavelet coefficients were plotted against the angle according to its mean and standard deviation. The proportion of variation within stocks along the outline was concluded with intraclass correlation (ICC).
Table 2 Statolith shape indices for different stocks of *D. gigas*

<table>
<thead>
<tr>
<th>Size variables</th>
<th>Shape indices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Area (A) Formfactor = ((4\pi A)^{\frac{1}{2}})</td>
</tr>
<tr>
<td></td>
<td>Perimeter (P) Roundness = ((4A)(\pi SL)^{\frac{1}{2}})</td>
</tr>
<tr>
<td>Statolith Length (SL)</td>
<td>Circularty = (P^2A^{-1})</td>
</tr>
<tr>
<td>Statolith Width (SW)</td>
<td>Rectangularity = (A(SL \times SW)^{\frac{1}{1}})</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shape indices</th>
<th>Mean±standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CH</td>
</tr>
<tr>
<td>Formfactor</td>
<td>3.83±0.24</td>
</tr>
<tr>
<td>Roundness</td>
<td>11.05±2.07</td>
</tr>
<tr>
<td>Circularty</td>
<td>19.67±0.94</td>
</tr>
<tr>
<td>Rectangularity</td>
<td>0.65±0.05</td>
</tr>
</tbody>
</table>

Note: CH = Chile; CR = Costa Rica; EQ = Equator; PE = Peru

Intra-species statolith shape variation was also analyzed by applying a canonical analysis of principal coordinates (CAP) (Anderson and Willis, 2003) with the “capscale” function in package vegan (Oksanen *et al*., 2013). Ordination of the population averages was also examined graphically with the shape descriptors (wavelet coefficients).

**Statistical analyses with different classification methods**

There were 52 wavelet coefficients retained for latter data processing after the wavelet data standardization. Combined with the four shape indices, a total of 56 shape variables were selected for stock discriminant analysis. The assumption of normality and homogeneity of variance for each variable was examined using Kolmogorov–Smirnov (KS) and Levene's tests (Zar, 2007). The shape index “Roundness” and 14 wavelet coefficients were excluded due to their failure to pass these normality and homogeneity tests. In order to unify the data analysis, the
remained 41 variables were prepared for future analysis.

Linear Discriminant Analysis (LDA), Classification Tree (CT), Naive Bayes (NB), Random Forest (RF) and Support Vector Machine (SVM) were employed to evaluate the classification effect. The first method is a traditional classification method commonly used for stock identification based on linear regression analysis (Martínez et al., 2002; Chen et al., 2012; Fang et al., 2014; Liu et al., 2015b). The latter four methods, which belong to supervised learning method, were used in varied fields due to their powerful classification ability (Lantz, 2013) for fish stock discrimination (Mercier et al., 2011).

The NB is a family of simple probabilistic classifiers based on applying Bayesian rules with strong independence assumptions between the features. This method requires a number of linear parameters in the number of predictors in a learning problem, and has also used in fish recruitment forecasting (Fernandes et al., 2014).

The response variable of decision trees (DTs), which have a branching structure that distinguishes values in a hierarchical form (Breiman et al., 1984), is either categorical (classification trees, CT) or numeric (regression trees, RT), and has also been used in the establishment of cephalopod maturity stages (Crespi-Abril et al., 2015).

The RF (Breiman, 2001) utilizes an ensemble of classification or regression trees to predict the dependent variable as a result of majority vote or average assignment across trees. The advantage of RF is that it allows correlated predictors without any data transformation or exclusion to estimate the importance of variables (Cutler et al., 2017).
The SVM, also referred to Maximum Margin Classifiers (MMC), is a method that minimizes empirical errors and maximizes geometric margins. It constructs an N-dimensional hyper-plane based on an algorithm that finds the maximum-margin hyper plane with the greatest division between classes (Hu et al., 2012). This method has been considered as one of the most influential data mining algorithms (Wu et al., 2008).

All the above analyses were carried out with R 3.1.3 (R Core Team, 2015) using the packages “MASS” for LDA, “C50” for CT, “e1071” for NB and SVM, “randomForest” for RF. The classification results from different methods were compared with the K-fold cross validation, which is a popular method to compare different models (Arlot and Celisse, 2010). The 10-fold cross-validation procedure has shown good performance in model selection (Kohavi, 1995) and was used to evaluate the performance of each model. Considering the sample size in this study, a stratified sampling was performed in the cross-validation (Zhao et al., 2014), where 70% of the total samples were chosen as training data and the remaining 30% as validation data to test the model performance. Some relevant measurements (e.g. sensitivity, specificity, Kappa) were listed to assess the effect among different methods (Kuhn, 2008).

**Results**

*Statolith shape indices variation among stocks*
Statolith length was not changed by the squid body growth in each stock (Table.3). Statolith shape also differed among four geographic stocks. The result showed that the Chilean (CH) stock had the highest value of the shape indices, except for circularity, among the four geographic stocks (Fig. 2). Chilean, Peruvian (PE) and equatorial (EQ) stocks had the similar circularity, except for Costa Rica (CR), which had the highest value (Fig. 2). The shape indices had significant difference between stocks (ANOVA, F=90.66, P<0.01). Pairwise analysis (Post-hoc, Tukey HSD) also showed significant differences between stocks (Fig.3, P<0.01). Individuals in PE and CH had the most similar shape indices, while CR and CH had the most different shape indices among stocks (Fig.3). The statolith in different stocks also presented distinguish growth pattern with their body size (ANCOVA, F=54.171, P<0.01).

Table 3 Regression analysis of mantle length and statolith length of *D. gigas*

<table>
<thead>
<tr>
<th>Location</th>
<th>Code</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>R²</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chile</td>
<td>CH</td>
<td>1</td>
<td>0.1497</td>
<td>0.1497</td>
<td>20.9732</td>
<td>0.172</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>102</td>
<td>0.7209</td>
<td>0.0071</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Costa Rica</td>
<td>CR</td>
<td>1</td>
<td>0.0591</td>
<td>0.0591</td>
<td>9.4013</td>
<td>0.150</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>53</td>
<td>0.3337</td>
<td>0.0063</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equator</td>
<td>EQ</td>
<td>1</td>
<td>0.1940</td>
<td>0.1940</td>
<td>32.1856</td>
<td>0.241</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>102</td>
<td>0.6089</td>
<td>0.0060</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peru</td>
<td>PE</td>
<td>1</td>
<td>0.3643</td>
<td>0.3643</td>
<td>47.8026</td>
<td>0.323</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.7620</td>
<td>0.0076</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig 2 Shape indices of statoliths for the four stocks of *Dosidicus gigas*. CH: Chile; CR: Costa Rica; EQ: equatorial; PE: Peru.

Fig 3 Pair-wise comparisons of shape indices among the four geographic stocks of *Dosidicus gigas*
**Shape variation of different geographic stocks**

The reconstructed mean statolith shapes were shown in Fig. 4 and differed among four stocks. The most obvious modification sections of statolith shape were located in rostrum (0 to 30°), rostral angle (30 to 60°) and wing (300 to 360°) (Figs. 4 and 5). The rostrum section of CR was the most outward from the centroid, CH was moved inward toward the centroid, PE and CR were between the above two stocks. On the contrary, CR presented a sharp rostrum with an innermost rostral angle, CH presented a blunt rostrum with an outermost rostral angle (Fig. 4). For the wing, CH had a plump wing than any other stocks, PE and EQ were similar and more outward than CR (Fig. 4). Dorsal section (150 to 210°) showed little variation among stocks, which was confirmed by examining the proportion of variation within groups summarized with the ICC, although the mean and standard deviation of wavelet coefficients were relatively higher than other sections (Fig. 5).

![Fig 4 Mean shape of statoliths for the four stocks of *Dosidicus gigas*. CH: Chile; CR: Costa Rica; EQ: equatorial; PE: Peru. The four numbers (0, 90, 180 and 270) represent angle in degrees (°) along the outline](image)
Fig 5 Mean and standard deviation (SD) of the wavelet coefficients (dot with error bar) for all combined statoliths and the proportion of variance within stocks, or the intraclass correlation (ICC, black solid line).

A total of 52 wavelet coefficients were selected to analyze the stock variation by canonical analysis of principal coordinates (CAP). The canonical scores revealed large differences between stocks (Fig. 6). The first two canonical axes explained a large proportion (96.0%) of the variation between stocks (CAP 1: 87.1%; CAP 2: 8.9%). The first scores (CAP 1) showed that southern hemisphere stocks (CH and PE) had negative values while northern hemisphere stock (CR) had positive values (Fig. 6). CAP 2 also showed difference between neritic stocks (CH, PE and CR) with negative values and oceanic stocks (EQ) with positive value (Fig. 6).
Discrimination rate comparison among difference methods

Among these four stocks, PE had a relatively higher misclassification rate than the other two stocks (Table 4). The classification rate in CH changed little among these five methods (Table 4). CR had a higher classification rate using CT and RF. EQ had the highest rate using SVM (Table 4).

Comparing traditional method (LDA) with machine learning methods (NB, DT, SVM and RF), the discrimination rates of machine learning methods were higher (Table 4). All the results showed that these methods could effectively partition different stocks of *D. gigas* with high classification rates (more than 60%, Table 4). For the result of classification in every stock, the performance of RF was better than SVM due to the precisely partitioning CR stock (92.86%) rather than the result in SVM (73.68%, Table 4).
Table 4 Discriminant analysis results and accuracy measures of different methods considered in this study

<table>
<thead>
<tr>
<th>Method</th>
<th>Original</th>
<th>Predict (%)</th>
<th>Total (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CH 68.93</td>
<td>1.94 2.91 26.21</td>
<td>64.46</td>
<td>54.7</td>
<td>85.6</td>
<td>0.396</td>
</tr>
<tr>
<td></td>
<td>CR 0.00</td>
<td>63.64 23.64 12.73</td>
<td></td>
<td>58.3</td>
<td>93.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EQ 4.85</td>
<td>9.71 78.64 6.80</td>
<td></td>
<td>63.9</td>
<td>88.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PE 27.45</td>
<td>7.84 18.63 46.08</td>
<td></td>
<td>36.2</td>
<td>73.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CH 67.57</td>
<td>0.00 8.11 24.32</td>
<td></td>
<td>76.7</td>
<td>81.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CR 0.00</td>
<td>58.33 37.50 4.17</td>
<td>65.09</td>
<td>68.7</td>
<td>90.0</td>
<td>0.440</td>
</tr>
<tr>
<td></td>
<td>EQ 5.88</td>
<td>0.00 76.47 17.65</td>
<td></td>
<td>53.3</td>
<td>84.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PE 14.29</td>
<td>7.14 17.86 60.71</td>
<td></td>
<td>40.0</td>
<td>88.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CH 69.70</td>
<td>0.00 3.03 27.27</td>
<td></td>
<td>66.7</td>
<td>86.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CR 0.00</td>
<td>92.86 7.14 0.00</td>
<td>68.87</td>
<td>56.2</td>
<td>93.3</td>
<td>0.333</td>
</tr>
<tr>
<td></td>
<td>EQ 3.33</td>
<td>10.00 70.00 16.67</td>
<td></td>
<td>56.7</td>
<td>75.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PE 20.69</td>
<td>0.00 24.14 55.17</td>
<td></td>
<td>26.7</td>
<td>77.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CH 74.07</td>
<td>0.00 3.70 22.22</td>
<td></td>
<td>80.0</td>
<td>89.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CR 0.00</td>
<td>73.68 21.05 5.26</td>
<td>70.75</td>
<td>62.5</td>
<td>94.4</td>
<td>0.487</td>
</tr>
<tr>
<td></td>
<td>EQ 4.17</td>
<td>4.17 83.33 8.33</td>
<td></td>
<td>56.7</td>
<td>78.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PE 25.00</td>
<td>2.78 13.89 58.33</td>
<td></td>
<td>50.0</td>
<td>85.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CH 71.88</td>
<td>0.00 3.12 25.00</td>
<td></td>
<td>80.0</td>
<td>89.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CR 0.00</td>
<td>92.86 7.14 0.00</td>
<td>70.74</td>
<td>62.5</td>
<td>96.7</td>
<td>0.472</td>
</tr>
<tr>
<td></td>
<td>EQ 6.90</td>
<td>6.90 72.41 13.79</td>
<td></td>
<td>63.3</td>
<td>76.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PE 16.13</td>
<td>3.23 22.58 58.06</td>
<td></td>
<td>40.0</td>
<td>84.2</td>
<td></td>
</tr>
</tbody>
</table>

Note: LDA = Linear Discriminant Analysis; NB = Naive Bayes; CT = Classification Tree; SVM = Support Vector Machine; RF = Random Forest

Discussion

Since the statolith was discovered and studied in the 1960s (Clarke, 1962), its morphology has attracted numerous researchers’ interests and is considered as an appropriate material to explore cephalopod biology and evolution (Clarke, 2003). The statolith analyses of linear measurements based on the terminology of previous studies revealed that some statolith morphometrics (e.g. statolith length, SL) had a strong relationship with ML (Arkhipkin and Murzov, 1986; Flores and Garland 2002;
Barcellos and Gasalla, 2015) and that some applications were also utilized to identify the stock variation (Fang et al., 2012; 2014; Vyver et al., 2016). However, it is a problematic method to describe the whole statolith shape with linear measurements. In this study, the otolith-derived shape indices (Tuset et al., 2003) and wavelet coefficients quantitatively described the irregular-shaped statolith via images and computer software, producing credible results, which were statistically in accordance with the statolith shape reconstruction (Fig. 4). These shape indices are supplemental variables for describing statolith morphometrics which help distinguish different stocks, as the significant variation presented in the result of this study was able to demonstrate.

Statoliths and statocysts were described as receptors that control and regulate squid position and movement in the water (Budelmann, 1977). Different feeding and swimming behaviors are the main reason for morphological variations in statoliths (Arkhipkin and Bizikov, 2000). The rostrum and wing of statoliths have the most significant difference between the reconstruction image and statistical result (Figs 3 and 4). These two sections (rostrum and wing) are the main parts that regulate accelerations during linear movement for pelagic squids (Arkhipkin and Bizikov, 2000). It has been reported that the statolith is only attached with macula statica princeps (MSP, located in the anterior chamber of statocysts) in the region of the wing (Budelmann, 1990). The main role of rostrum is inducing the endolymph flows across the crista transversalis anterior (CTA) in the inner posterior corner of statocysts and serves mainly to support the lower part of the wing (Budelmann, 1990; Arkhipkin and
Bizikov, 2000). As the above structures described, the gravity of center of statolith lies outside the wing. When the squid move or change direction, the statolith body deviates by a small acute angle to maintain a certain endolymph stream flows from under the statolith body along the anterior statocyst wall outwards (Arkhipkin and Bizikov, 2000). The size of wing and the volume of endolymph stream are closely related to the size of statolith and statocyst, which indicate the intensity of movement, including swimming speed. The CR squid have a narrower and longer rostrum and wing than those of CH (Fig. 4). Considering the explanation of the statolith morphological difference between demersal and pelagic decapods (Arkhipkin and Bizikov, 2000), an appropriate hypothesis of stock statolith variation in this study is that the northern stock (CR) has a greater sensitivity to the low angular accelerations compared with those of other geographic stocks in this study.

The movement of squid is highly related to its ambient oceanographic environment and the mobility of its prey and predators. *D. gigas* are mainly distributed in a high productivity region in the eastern tropical Pacific Ocean (Ichii *et al.*, 2002). The mechanism of squid aggregation in each geographic area depends on the local oceanographic characteristics. The Peruvian and Chilean waters are dominated by Peru Current (PC) in the Humboldt Current System (HCS), characterized by a low surface temperature and higher primary production levels (Pennington *et al.*, 2006). The North Equatorial Countercurrent (NECC) veers to the north in a counter-clockwise direction to merge with the North Equatorial Current (NEC) in the area of the countercurrent ridge (Costa Rica Dome) concentrating
surface nutrients to attract the aggregation of squid (Wyrtki, 1966; Lavín et al., 2006). A divergence of westward component of South Equatorial Current (SEC), from 3°N to 3°S and across the eastern tropical Pacific, is called “Equatorial cold tongue” and also known for macronutrients in euphotic zone (Chavez and Barber, 1987; Pennington et al., 2006). These features are the important factors that attract squid in the equatorial area in eastern tropical Pacific (Sanchez et al., 2016). Characterized by a higher surface temperature and a more complicated, weaker current system in the Costa Rica Dome, the squids living in this area tend to move more to prey and swim in such a wind-driven eddy area (Kessler, 2006) in comparison to their counterparts in south hemispheres of eastern Pacific with low temperature and strong current (Kessler et al., 2006). Thus, the statolith shapes of CH and PE individuals become “rounded”, the result of living in relatively cold temperatures and stronger currents, while the northern stock (CR) statolith shape (Fig. 3) is “sharp”. *D. gigas* prey on zooplankton over all life stages (Liu, 2012). Some studies reported that the zooplankton feature in Costa Rica Dome, including the species and biomass, was distinctively different from other area in the ETP (Fernández-Álamo and Färber-Lorda, 2006). Food composition and availability also lead to the variability of squid prey behavior, which indirectly influences the statolith shape diversity (Nigmatullin et al., 2001; Ulloa et al., 2012).

Stock identification is a critically important priority for stock assessment, as biased results can lead to mismanagement (Ying et al., 2011). Some materials and methods have been proposed and used for stock discrimination with successful results (Chen et al., 2012; Fang et al., 2012, 2014; Liu et al., 2015b). The results of this study
have also successfully classified the different stocks from north and south hemispheres, supporting the results of previous studies (Sandoval-Castellanos et al., 2007, 2010; Staaf et al., 2010; Liu et al., 2015b). Conventional statistical methods should rigorously keep the normality and homogeneity for all the data (Zuur et al., 2010). This rule limited the data availability in some extents. The machine learning approaches, which create of algorithm from data and build models for decisions, can process all kinds of data with related statistical analysis, regardless of the data format (Lantz, 2013). Supervised learning methods have been widely used in fisheries science because of the merits of automatically building a model to make prediction from data without any human assistance, (Dub et al., 2013; Fernandes et al., 2014; Crespi-Abril et al., 2015) and these methods have also been tested as an effective way to solve classification issues (Vyver et al., 2016). This is also verified within this study, with a higher classification rate than traditional methods (Table 4). Comprehensively considering the sensitivity, specificity and Kappa values, SVM and RF were good models for the D. gigas statolith shape classification in this study (Table 4). The SVM and RF techniques produced a better result than other methods, which has been observed in other studies (Li et al., 2015). RF tended to be more successful than SVM in our study because of the diminution associated the averaging results in SVM (Table 4).

The stock classification of D. gigas was evaluated based on the statolith shape in this study. The statolith shape of CH stock had the largest size and CR stock had the smallest size within the shape indices and all the shape indices had significant
differences among stocks. Statolith shapes from different geographic stocks were reconstructed by wavelet coefficients derived from the statolith images via computer software, and stock classification was performed. The results showed that reconstructed statolith shapes could realistically reflect the stock morphological difference, which was mainly observed in the rostrum and wing sections of statoliths. Machine learning methods were also used for stock classification in this study and produced a superior classification to traditional methods. This study supports the conclusions of previous studies that two separate stocks of *D. gigas* exist in the ETP. Physical oceanography and feeding behaviour are the two main causes of statolith shape variation among stocks. This study further reinforces the idea that analysis of statolith shape, using techniques such as RF and SVM, is a useful tool for stock recognition.
References


Arkhipkin, A. I., & Z. N. Shcherbich, 2012. Thirty years' progress in age
determination of squid using statoliths. Journal of the Marine Biological
Association of the United Kingdom, 92(06), 1389-1398.

temperature influences adult size and life span in jumbo squid. Canadian
Journal of Fisheries & Aquatic Sciences, 72(3), 1-10.

Barcellos, D. D., & M. A. Gasalla, 2015. Morphology and morphometry of
_Doryteuthis plei_ (cephalopoda: loliginidae) statoliths from the northern shelf
off São Paulo, southeastern Brazil. Journal of Natural History, 49, 1305-1317.

Scientarum Hungaricae, 44: 7–59.


Budelmann, B. U. 1975. Gravity receptor function in cephalopods with particular

Budelmann, B. U. 1977. Structure and function of the angular acceleration receptor
systems in the statocysts of cephalopods. Symp. zool. Soc. Lond. No. 38:
309-324.

Budelmann, B. U. 1990. The statocysts of squids. In Squid as experimental animals:


Crespi-Abril, A. C., N. Ortiz, & D. E.Galván, 2015. Decision tree analysis for the
determination of relevant variables and quantifiable reference points to establish maturity stages in *Enteroctopus megalocyathus* and *Illex argentinus*. ICES Journal of Marine Science, 72(5). DOI: 10.1093/icesjms/fsu202


Lantz, B. 2013. Machine learning with R: learn how to use R to apply powerful
machine learning methods and gain an insight into real-world applications.

Packt Publishing.


United Kingdom, 86(04): 767-771.


Staaf, D. J., Ruiz-Cooley, R. I., Elliger, C., Lebaric, Z., Campos, B., Markaida, U., &


stomach contents of jumbo flying squid Dosidicus gigas (cephalopoda: teuthoidea): how important is it?. Revista De Biologia Marina Y Oceanografia, 47(3), 413-428.


on habitat suitability of jumbo flying squid, *Dosidicus gigas*, in the southeast Pacific Ocean off Peru. ICES Journal of Marine Science, 9(8), 863-868.


