Forecasting the Catch Data of *E.sinensis* by Applying Continuous Time Markov Approach

Guangpeng Feng[1, a], Yanhua Li[2], Bin Wu[2], Benhe Ma[2], Huogen Hu[3] and Haihua Wang[1, 2, *]

[1]East China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Shanghai, China
[2]Jiangxi Institute for Fisheries Sciences, Nanchang, China
[3]Jiangxi Fisheries Technology Extension Station, Nanchang, China

*Corresponding author e-mail: jxswhh@163.com, [a]Fenggp@ecsf.ac.cn

Abstract. The forecast model for catch data of *E.sinensis* was built by using continuous time Markov approach, combining with optimization method (0.618) based on of catch data of *E.sinensis* in Yangtze River estuary from 1970~2012. And a forecast on the catch data of *E.sinensis* in 2013 was made in this paper. The result showed that the forecast was corresponded to the catch data accurately and provides a new method for the prediction of crab resources.

1. Model Introduction

The Continuous Time Markov Approach (CTM) is mainly used to describe the state change and interaction process between the system and the environment. It is particularly applicable to the dynamic analysis of the relationship between biology and environment [1]. The model is based on continuous-time Markov theory and time-sharing analysis techniques in mathematics, and draws on the concept of quantum transitions in atomic physics. It was proposed by P.J.H Sharpe and Wu Xinyi et al. in 1985 [2].

The types of CTM modeling are between the theoretical and empirical models, the modeling principles are clear, the modeling method is simple, the analysis of relevant parameters is flexible, and the data is inclusive (can be theoretical, empirical, or conceptual "precise" data and "inexact" data), the model contains the basic rules of the ecological system such as environmental comprehensive principle and limit factor principle, and can fully display and integrate the system control factors of different sources and different dimensions within the system. Therefore, it is particularly suitable for the study of complex systems controlled by multiple factors, such as applied to biological and ecological systems.

2. Mathematical principles

Definition: Let the parameter set T of the Markov process $\{X_n, n \in T,\}$ be a discrete time set, $T=\{0, 1, 2, ...\}$, the total $\{X_n\}$ of all possible values of X_n is a discrete state space, denoted by $E=\{x_1, x_2, ...\}$, if any positive integer $n \in T$ and any $x_1, x_2, ..., x_n, x_{n+1} \in E$, have $P(X_{n+1}=x_{n+1} \mid X_1=x_1, X_2=x_2, ..., X_n=x_n)=P(X_{n+1}=x_{n+1} \mid X_n =x_n)$, then $\{X_n\}$ is called Markov chain [3-5].

In the Markov chain, $P_{ij}^{(n)}=p\{X_{n+1}=j/X_n=i\}$ is called the one-step transition probability of the Markov chain $\{X_n, n \in T,\}$ at time n, where i, j $\in$ E, the one-step transition probability can be represented by the matrix P.

$$P = \begin{bmatrix} P_{11} & P_{12} & ... & P_{1m} \\ P_{21} & P_{22} & ... & P_{2m} \\ ... & ... & & \\ P_{m1} & P_{m2} & ... & P_{mm} \end{bmatrix}$$

From t to t+1, the ratio of the frequency nij of the state transition from Si to Sj to the total frequency n is the one-step transition probability of the transition of state Si to Sj. Pij $\cong$ 0 and $\sum_{j=1}^{m} P_{ij} = 1$ (i, j=1, 2, ...m). In general, the transition probability Pij(n) is related not only to the state i, j but also to the time n. When Pij(n) does not depend on time n, it means that the Markov chain has a smooth transition probability. If for any i, j $\in$ E, the transition probability Pij(n) of the Markov chain $\{x_n, n \in T\}$ is independent of n, the Markov chain is said to be homogeneous [6].

Based on probability theory, the CTM model predicts the possible state of the system at a future moment (ti>t) according to the state of the system at the time t at the transition probability matrix P of each order. When applying this model for prediction, it is necessary to satisfy the following two assumptions on a certain space-time scale. First, the system must be stable during the study period, and its transition probability matrix remains unchanged from period to period; second, "no post-effect", that is, the transition of the system state is only related to the previous state, but not to any other state [6].

3. Modeling Methods

3.1. Historical data processing
Historical data is processed using the method of optimization (0.616 method), and the state level is divided.

The partitioning principle is: (1) All elements of the transition probability matrix of each order of the Markov chain are non-negative; (2) The sum of all the rows of the transition probability matrix of the Markov chain is about 1; (3) The probabilities of various states of the Markov chain are roughly similar. According to this principle, the final grading standard is determined.

3.2. Markov property test
Before analyzing the actual problem with the Markov chain model, it must first test whether the random sequence has Markov property. The x^2 statistic is usually used to test the Markov property of discrete sequences.

$$x^2 = 2\sum_{i=1}^{m}\sum_{j=1}^{m} f_{ij}\left|\log\frac{P_{ij}}{P*j}\right|, \quad P_{*j} \text{ is called marginal probability, } P_{*j} = \frac{\sum_{i=1}^{m} f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{m} f_{ij}} \text{ indicates that the}$$

historical data sequence has m possible states. The f_{ij} is used to represent the frequency of the historical data sequence from state i to the state j in only one step. i, j $\in$ E, where P_{ij} is the transition probability of one step. If n is large enough, the statistic obeys the x^2 distribution with a degree of freedom of $(m-1)^2$, given the significance level a, the table can be scored at the point $x_a(m-1)^2$ if $x^2 > x_a(m-1)^2$. It can be determined that the sequence $\{X_n\}$ is Markovian, but the sequence cannot be treated with Markov chain.

3.3. Establish transition probability matrix

According to the state of the historical data sequence, the frequency matrix and each order transition probability matrix are listed.

f_{ij} represents the frequency of transition from state i to step j, step 2, step... until step m to state j, i, j $\in$ E. In the transfer frequency matrix, divide the element f_{ij} of the jth column of its ith row by the sum of the rows and obtain the value, which is called transition probability P_{ij}; $P_{ij} = \dfrac{f_{ij}}{\sum\limits_{j=1}^{m} f_{ij}}$ is the transition probability matrix of each order, indicating that the historical data sequence contains m states.

3.4. Computation of Markov Chain Weights

r_k is an autocorrelation coefficient of each order, k$\in$E, normalized to it, ie $w_k = \left| r_k \right| / \left| \sum\limits_{k=1}^{m} r_k \right|$, w_k is the Markov chain weight of each order delay, and m is the maximum order of calculation.

4. Construction of the CTM model of the crab in the Yangtze Estuary

The CTM model provides an effective tool for quantitatively studying the resource dynamics of the E.sinensis, but its application requires long-term accumulation of resource monitoring data. In general, the total amount of fishing can indirectly reflect the resource density [7]. Due to data limitations, this paper uses the method of Zhan Bingyi [8] to replace the population with the amount of E.sinensis in the Yangtze River estuary to establish CTM. The model predicts the resource changes of E.sinensis in the Yangtze River estuary.

4.1. Sources of data

Based on the monitoring data collected from the Yangtze River Estuary Crab Eel Crab Harvesting data from 1970 to 2012 (Table 1), the 2013 E.sinensis population catch data was used as an example of model test and prediction.

Table 1. Catch data of *E.sinensis* in Yangtze River estuary from 1970~2012.

Annual	Catch	Annual	Catch	Annual	Catch	Annual	Catch	Annual	Catch
1970	1.9	1979	54.5	1988	6.5	1997	0.8	2006	6.0
1971	19.2	1980	60.0	1989	6.5	1998	0.8	2007	16.0
1972	32.6	1981	56.1	1990	8.0	1999	1.2	2008	14.0
1973	21.1	1982	90.0	1991	25.5	2000	0.9	2009	14.0
1974	20.0	1983	95.8	1992	10.0	2001	0.9	2010	25.0
1975	45.0	1984	35.0	1993	15.0	2002	0.8	2011	26.0
1976	114.1	1985	12.5	1994	12.3	2003	0.5	2012	26.0
1977	47.3	1986	12.5	1995	11.0	2004	1.8	2013	25.7
1978	29.0	1987	10.0	1996	5.5	2005	10.7	-	-

4.2. Status Level Division

According to Table 3-15, it can be concluded that the maximum catch of the migrating *E.sinensis* in the Yangtze River Estuary is 114.1t and the minimum catch is 0.5t. First make the first point according to the optimization method.

The first point: 0+(114.1-0.5)×0.618=70.2

The second point: 0.5+114.1-70.2=44.4

The third point: 0.5+70.2-44.4=26.3

The fourth point: 0.5+44.4-26.3=18.6

The fifth point: 0.5+26.3-18.6=8.2

Then fold the first remaining line left:
The first point: 70.2+(114.1-70.2)×0.618=97.3
The second point: 70.2+(114.1-97.3)=87.0

According to the above data, according to the Markov chain transition probability matrix data partition principle, the catch of *E.sinensis* is divided into six levels, which are: 8.2 is the first level, 8.2 to 18.6 is the second level, 18.6 to 26.3 For grade 3, 26.3 to 44.4 are grade 4, 44.4 to 70.2 are grade 5, and 70.2 and above are grade 6. From 1970 to 2012, the ranking of *E.sinensis* in the Yangtze River estuary is shown in Table 2.

Table 2. Statistics of state hierarchical division of catch data of *E.sinensis* in Yangtze River estuary from 1970~2012.

Annual	Grade	Annual	Grade	Annual	Grade	Annual	Grade	Annual	Grade
1970	1	1979	5	1988	1	1997	1	2006	1
1971	3	1980	5	1989	1	1998	1	2007	2
1972	4	1981	5	1990	1	1999	1	2008	2
1973	3	1982	6	1991	3	2000	1	2009	2
1974	3	1983	6	1992	2	2001	1	2010	3
1975	5	1984	4	1993	2	2002	1	2011	3
1976	6	1985	2	1994	2	2003	1	2012	3
1977	5	1986	2	1995	2	2004	1		
1978	4	1987	2	1996	1	2005	2		

4.3. Markov Chain Inspections

From the measured data of the catch amount of *E.sinensis* in Table 3-18, the frequency transition matrix f_{ij} and the one-step transition probability matrix $P(1)$ are calculated to obtain the frequency transition matrix of the *E.sinensis*:

$$F_{ij} = \begin{bmatrix} 10 & 2 & 0 & 0 & 0 \\ 3 & 7 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

Combining one step transition probability matrix P(1)

$$P_{ij} = \begin{bmatrix} 0.714 & 0.143 & 0.143 & 0 & 0 & 0 \\ 0.273 & 0.636 & 0.091 & 0 & 0 & 0 \\ 0 & 0.167 & 0.5 & 0.167 & 0.167 & 0 \\ 0 & 0.333 & 0.333 & 0 & 0.333 & 0 \\ 0 & 0 & 0 & 0.2 & 0.4 & 0.2 \\ 0 & 0 & 0 & 0.333 & 0.333 & 0.333 \end{bmatrix}$$

The statistic X^2 of the catch of the *E.sinensis* was calculated to be 69.937, with a given significance level a=0.05. Table lookup available $x_a^2(25)=37.65$, due to $X^2 > x_a^2[(m-1)^2]$, Therefore, the catch data of *E.sinensis* meet the Markov nature and can be treated as a Markov chain.

4.4. Weights of markov chains of various orders

The measured data from Table 3-18 were used to calculate the catch of the *E.sinensis* in the Yangtze River Estuary.

The average value from 1970 to 2012 was x=23.309. The autocorrelation coefficients of *E.sinensis* were calculated based on the measured data from 1970 to 2012 and the average value. The weights were calculated from the autocorrelation coefficients. The results are shown in Table 3.

Table 3. Each orders auto correlation coefficient and the weigh of Markov chain of each steps of *E.sinensis*.

| Project	Time delay (year)					
	1	2	3	4	5	6
Autocorrelation coefficient rk	0.6931	0.4404	0.4386	0.3890	0.2416	0.2372
Weight wk	0.2841	0.1805	0.1798	0.1594	0.0990	0.0972

5. Model Calculation and Prediction

According to the catch data of Chinese mitten crab from 1970 to 2012 (excluding the 2013 data, the model was left as a test), the transition probability matrix of *E.sinensis* was calculated as:

$$P_1 = \begin{bmatrix} 0.714 & 0.143 & 0.143 & 0 & 0 & 0 \\ 0.273 & 0.636 & 0.091 & 0 & 0 & 0 \\ 0 & 0.167 & 0.5 & 0.167 & 0.167 & 0 \\ 0 & 0.333 & 0.333 & 0 & 0.333 & 0 \\ 0 & 0 & 0 & 0.2 & 0.4 & 0.2 \\ 0 & 0 & 0 & 0.333 & 0.333 & 0.333 \end{bmatrix}$$

$$P_2 = \begin{bmatrix} 0.549 & 0.217 & 0.187 & 0.024 & 0.024 & 0 \\ 0.369 & 0.459 & 0.142 & 0015 & 0.015 & 0 \\ 0.046 & 0.245 & 0.321 & 0.117 & 0.206 & 0.067 \\ 0.091 & 0.267 & 0.197 & 0.122 & 0.189 & 0.133 \\ 0 & 0.067 & 0.067 & 0.213 & 0.360 & 0.293 \\ 0 & 0.111 & 0.111 & 0.177 & 0.355 & 0.244 \end{bmatrix}$$

$$P_3 = \begin{bmatrix} 0.451 & 0.256 & 0.199 & 0.036 & 0.049 & 0.010 \\ 0.388 & 0.373 & 0.171 & 0.027 & 0.035 & 0.006 \\ 0.100 & 0.255 & 0.228 & 0.117 & 0.197 & 0.105 \\ 0.138 & 0.257 & 0.176 & 0.115 & 0.193 & 0.120 \\ 0.018 & 0.124 & 0.110 & 0.181 & 0.324 & 0.242 \\ 0.030 & 0.148 & 0.125 & 0.171 & 0.301 & 0.223 \end{bmatrix}$$

$$P_4 = \begin{bmatrix} 0.392 & 0.272 & 0.199 & 0.046 & 0.068 & 0.023 \\ 0.379 & 0.330 & 0.184 & 0.038 & 0.053 & 0.016 \\ 0.141 & 0.254 & 0.190 & 0.112 & 0.191 & 0.114 \\ 0.169 & 0.251 & 0.170 & 0.108 & 0.185 & 0.117 \\ 0.047 & 0.160 & 0.129 & 0.164 & 0.289 & 0.210 \\ 0.062 & 0.176 & 0.137 & 0.155 & 0.272 & 0.195 \end{bmatrix}$$

$$P_5 = \begin{bmatrix} 0.354 & 0.278 & 0.196 & 0.054 & 0.083 & 0.035 \\ 0.361 & 0.308 & 0.189 & 0.047 & 0.070 & 0.027 \\ 0.170 & 0.251 & 0.176 & 0.108 & 0.183 & 0.114 \\ 0.189 & 0.248 & 0.168 & 0.104 & 0.177 & 0.113 \\ 0.077 & 0.185 & 0.140 & 0.149 & 0.261 & 0.185 \\ 0.092 & 0.196 & 0.145 & 0.142 & 0.248 & 0.174 \end{bmatrix}$$

$$P_6 = \begin{bmatrix} 0.329 & 0.278 & 0.192 & 0.061 & 0.096 & 1.052 \\ 0.342 & 0.294 & 0.190 & 0.054 & 0.084 & 0.084 \\ 0.190 & 0.249 & 0.171 & 0.104 & 0.177 & 0.177 \\ 0.202 & 0.247 & 0.168 & 0.101 & 0.171 & 0.171 \\ 0.106 & 0.202 & 0.148 & 0.137 & 0.239 & 0.239 \\ 0.119 & 0.209 & 0.151 & 0.132 & 0.229 & 0.229 \end{bmatrix}$$

The advantage of the CTM model is that it can be considered as a black box model without considering the various changes in the population state, simplifying the complicated and difficult to describe population changes into states, thereby facilitating the simplification of complex ecological problems. In order to intuitively reflect the resource change trends of the *E.sinensis* in the Yangtze River Estuary, a map of the state transition probability of catches of the *E.sinensis* in the Yangtze River Estuary was drawn (Figure 1).
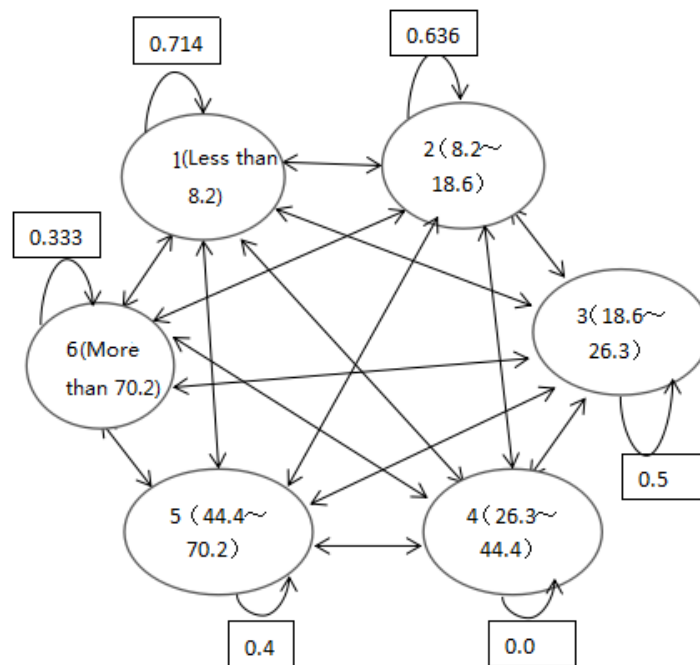
Figure 1. Statistics figure of state hierarchical division of catch data of *E.sinensis* in Yangtze River estuary.

According to the rank of transition probability matrix, the weight values of each state, and the state of the initial year, the catch status of the *E.sinensis* was predicted in 2013. The results are shown in Table 3. It can be seen from the table that in the column of *E.sinensis* catch status, the probability of the state "3" is maximally 0.307 after the weighted sum of the forecasted probabilities of the same state, therefore, in 2013, the catch level of *E.sinensis* was "3", and the estimated catch range was 18.6t to 26.3t, which was in line with the measured data of 25.7t. The measured value is within the predicted value range, which proves that the established CTM model of the Yangtze River Estuary has higher sensitivity.

Table 4. Predicted statistics of catch data of *E.sinensis* in Yangtze River estuary in 2013.

Initial year	Status	Time delay (year)	State weight	1	3	4	5	6	Probability source
2012	3	1	0.2841	0	0.5	0.167	0.167	0	P_1
2011	3	2	0.1805	0.046	0.321	0.117	0.206	0.067	P_2
2010	3	3	0.1798	0.100	0.228	0.117	0.197	0.105	P_3
2009	2	4	0.1594	0.379	0.184	0.038	0.053	0.016	P_4
2008	2	5	0.0990	0.361	0.189	0.047	0.070	0.027	P_5
2007	2	6	0.0972	0.342	0.190	0.054	0.084	0.084	P_6
Weighted value				0.156	0.307	0.106	0.144	0.044	

6. Discussion and Outlook

6.1. Application of CTM Model

The CTM prediction method can be divided into three prediction methods based on the absolute distribution of Markov chain, superimposed Markov chain and weighted Markov chain. It is generally believed that the application of weighted Markov chain prediction method can fully and reasonably use information [9], so this paper selects the weighted Markov chain prediction method for modeling and analysis, predicted and verified the status of crabs in 2013, the results are more accurate. This shows that the CTM model does not require high data, as long as there are many years of continuous resource monitoring data, and the data is in line with the requirements of the Markov chain prediction model, it can be modeled and forecasted, without considering other factors that affect the population. The CTM model has simple calculation, accurate prediction and strong reliability, and provides a new method for predicting the quantity of *E.sinensis*.

6.2. Improvements in the CTM Model

Considering the characteristics of the river-sea migratory resources of *E.sinensis*, their resources are greatly affected by natural and human factors. Over 40 years of actual measurement of the resources of the *E.sinensis* in the Yangtze River Estuary have also shown that the population changes drastically. Therefore, although the CTM model is used to analyze the population trend for the recent trend and the credibility of the forecast results is high, according to the modeling principle of the model, once the natural or human factors change, the steady state of the system will change accordingly. For specific improvement, we can consider the introduction of control factors that lead to changes in the resources of the *E.sinensis*, within the range of values (0-1), by investigating or studying the corresponding values to construct a new state transition probability matrix, so that the model parameters and the structure is more in line with resource conditions to improve the accuracy of the CTM model [9-11].

References

[1]   Jr R L O, Sharpe P J H, Wu H I. Whole-plant modelling: A continuous-time Markov (CTM) approach [J]. Ecological Modelling, 1985, 29(1):171~187.

[2]   Sharpe P J H, Rykiel E J J, Olson R L, et al. A continuous time markov (CTM) model of plant growth [J]. 1987.

[3]   Stenseth N C. Evolutionary Aspects of Demographic Cycles: The Relevance of Some Models of Cycles for Microtine Fluctuations [J]. Oikos, 1977, 29(3):525~538.

[4]   Stenseth N C, Hansson L, Myllymäki A, et al. General Models for the Population Dynamics of the Field Vole Microtus agrestis in Central Scandinavia [J]. Oikos, 1977, 29(3):616~642.

[5]   Stenseth N C. Modelling the Population Dynamics of Voles: Models as Research Tools [J]. Oikos, 1977, 29(3):449~456.

[6]   Stenseth N C, Hansson L, Myllymaki A. Population Dynamics of the Field Vole (Microtus agrestis [L]): a Model [J]. Eppo Bulletin, 1977, 7(2):371~384.

[7]   W G SHI, M Y ZHANG, K LIU, et al. Stress of hydraulic engineering on fisheries in the lower reaches of the Yangtze River and compensation [J].2009, 21(1):10-20 (in Chinese).

[8]   B Y Zhan, Y M Chen, X J Dai, et al. The alteration and rational utilization of megalopa resources of Eriocheir sinenses population in Changjiang River Estuary [J]. Journal of Shanchai Fisheries University, 1999, (4):322~328.

[9]   Yin G G, Zhang Q. Continuous-Time Markov Chains and Applications [J]. Technometrics, 2013, 43(2):240.

[10]   Mangel M. Continuous-Time Markov Chains: An Applications-Oriented Approach. by William J. Anderson [J]. Journal of the American Statistical Association, 1991, 88(422):901~904.

[11]   Gaitsgory V. Continuous-time Markov chains and applications: a two-time-scale approach. Second edition [book review of MR2985157]. [J]. Siam Review, 2013, (4):799~801.