Problem Chosen	2022 MCM/ICM Summary Sheet	Team Control Number
A		XJ162

test

Summary

Keywords: 123456

Contents

1 Introduction

1.1 Problem Restatement

Finless porpoise is the only freshwater mammal in the Yangtze River at present, which is distributed in the middle and lower reaches of the Yangtze River, Dongting Lake and Poyang Lake, and its population has decreased dramatically in the past 20 years. According to the statistics, the number of finless porpoises in the Yangtze River was more than 2,700 in 1991. However, in the year of 2006, there were fewer than 1,800 finless porpoises surviving in the area. In 2011, there were probably just over 1,000 of them, and in 2018 there were about 1,012.

In fact, since the 1980s, the ecologists along with the government had explored and developed three conservation strategies: in situ conservation, ex situ conservation and artificial breeding. Among them, ex situ protection, that is, selecting some waters with similar ecological environment to the Yangtze River to establish ex situ protection, is the most direct and effective measure to protect the Yangtze finless porpoise.

China has set up five ex-situ protected sites until now, in which more than 150 Yangtze finless porpoises are conserved. On September 18, 2021, CCTV reported that the population of the Yangtze finless porpoise is growing steadily. The population decline of the Yangtze finless porpoise has been curbed, but its critically endangered status remains unchanged.

Based on what has been discussed above, please address the following problems:

1 Establish a mathematical model to predict the population number of finless porpoises in five ex situ protected areas after 20 years, and explain how the sex ratio of 150 finless porpoises in ex situ protected areas affects the population development of finless porpoises.

2 Will the Yangtze finless porpoise become functionally extinct without ex situ conservation strategies?

3 Based on your analysis, please submit no more than 2 pages of recommendations for the protection of finless porpoises to the relevant authorities.

1.2 Overview of Our Work

2 Assumptions and Justifications

These are necessary assumptions for simplifying the model.

1. The carrying capacity per unit area of each ex-situ conservation layout is constant.
2.

3 Notations

Table 3.1: Notation Descriptions

Symbol	Definition
r	Innate rate of increase
λ	Finite rate of increase
R_θ	Intrinsic rate of natural increase
N-extant	Average extant population size
N-all	Average population size
PE	Probability of Extinction
GeneDiv	Genetic Diversity
TE	Time of Extinction(year)
SD	Standard Deviation
K	Carrying capacity
N_t	Size of finless porpoise population in the year of $1991 + t$

4 Intorduction and Results of Models

Note that because of the deficiency of the statistics about the other four ex-situ conservations, the size of finless porpoise population per unit area is considered the same as that in Swan Oxbow of the Yangtse River.

Considering tremendous cost on massive finless porpoise population census, merely six years of data was collected in Swan Oxbow of the Yangtse River during the three decades since 1992. (Zhigang, 2020) Thus, we've applied **Lagrange interpolation** to obtain other years' data in Swan Oxbow.

4.1 Relation between Population size and Time based on Lagrange Interpolation

Given n distinct real values $x_1, x_2, \cdots, x_n$and n real values $y_1, y_2, \cdots, y_n$ (not necessarily distinct), there is a unique polynomial P with real coefficients satisfying $P(x_i) = y_i$ for $i \in \{1, 2, \cdots, n\}$, such that deg($P$)<n.

The polynomial $P(x)$ is defined as follows:

$$P(x) = \sum_{k=1}^{n} y_k p_k(x), \quad p_k(x) = \frac{(x - x_1)\cdots(x - x_{k-1})(x - x_{k+1})\cdots(x - x_n)}{(x_k - x_1)\cdots(x_k - x_{k-1})(x_k - x_{k+1})\cdots(x_k - x_n)}$$

After substituted the number in 1992, 2002, 2005, 2007, 2015 and 2021, the figure of the polynomial is as follows:

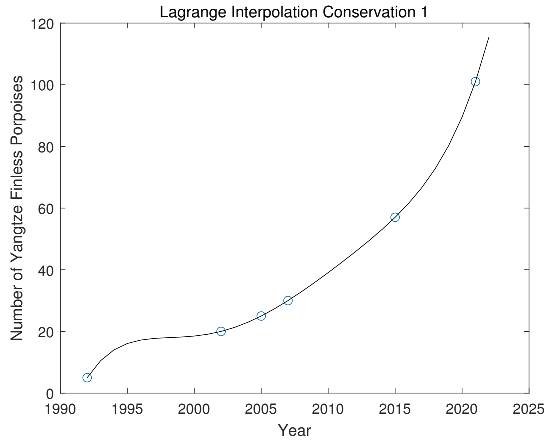


Figure 4.1: Langrange Interpolation Conservation

According to Lagrange interpolation and the definition of N_t :

$$N_t = P(t) \quad t = 1, \cdots, 30$$

and the exact numbers are listed below:

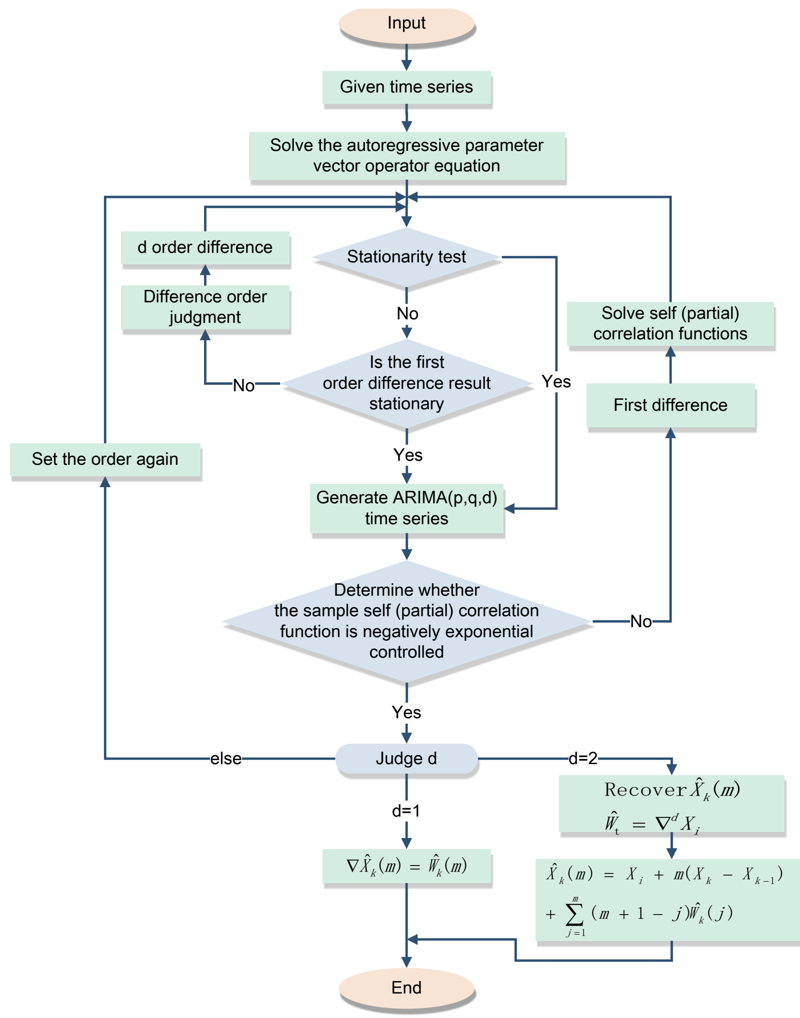
Table 4.1: Estimated size of population on year basis from 1992 to 2022

Year	Number	Year	Number	Year	Number
1992	5	2002	20	2012	45.7364
1993	10.4864	2003	21.2919	2013	49.2555
1994	13.9987	2004	22.9647	2014	52.9737
1995	16.0839	2005	25	2015	57
1996	17.2014	2006	27.3612	2016	61.4941
1997	17.7296	2007	30	2017	66.6734
1998	17.9732	2008	32.8636	2018	72.82
1999	18.1701	2009	35.9015	2019	80.2876
2000	18.4981	2010	39.0724	2020	89.5084
2001	19.0819	2011	42.3514	2021	101

4.2 Model I: Vortex model based finless porpoise analysis

4.3 Model II: Auto Regressive Integrated Moving Average(ARIMA) model

We apply ARIMA model in order to predict the size of finless porpoise population in five ex-situ conservation areas after 20 years.



Under regular circumstances, the time series we obtain in the real world has tendency, seasonality and non-stationarity. Thus, it's vital for us to transfer the non-stationary time series to stationary time series and make an assumption that the time series is an Auto Regressive Moving Average (ARMA) series to predict the future data. ARMA series is defined as follows.

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q}$$

ϵ_1 is a stationary white noise whose average is zero and deviation is σ_ϵ^2;$X_t$is an ARMA series with p and q degree, recorded briefly as $\mathbf{ARMA}(p,q)$ series. Akaike Information Criterion(AIC) is one of the most commonly used criterion to determine the degree of $\mathbf{ARMA}(p,q)$: choose p,q such that

$$\min \mathbf{AIC} = n \ln \hat{\sigma_\epsilon^2} + 2(p+q+1) \tag{4.1}$$

n is the capacity of sample;$\hat{\sigma_\epsilon^2}$ is the estimation of σ_ϵ^2 relating to p and q. Suppose $p = \hat{p}, q = \hat{q}$, such that equation(4.1) reaches the minimum, than we deem the series is $\mathbf{ARMA}(\hat{p},\hat{q})$.

Suppose $\mathbf{ARMA}(p,q)$ series has an unknown average parameter μ, the model becomes

$$\phi(B)(X_y - \mu) = \theta(B)\epsilon_t,$$

meanwhile, the number of unknown parameters is $k = p + q + 2$, the AIC is: choose p,q such that

$$\min \mathbf{AIC} = n \ln \hat{\sigma_\epsilon^2} + 2(p+q+2). \tag{4.2}$$

In fact, equations(4.1)and(4.2)have the same minimum point $\hat{p},\hat{q}$. After that, we usually choose $p = 1, q = 1$ to make parameter estimation over ARMA model.

It's demonstrated that the differential operation can stabilize certain class of non-stationary series. And It's emphasized that stationary test must be conducted previously. Stationary test can be applied by calculating sample autocorrelation function and sample coefficient of partial function.

If the functions are truncated or trending to 0(meaning being controlled by negative index), than the series belongs to ARMA model.

If at least one of the functions above is not truncated or trending to 0, than it's not stationary.

Suppose the series is non-stationary, which can be transformed to a stationary series by d -degree differential operation, denoted as $\textbf{ARIMA}(p, q, d)$ series,than differentiate the sample by d -degree:

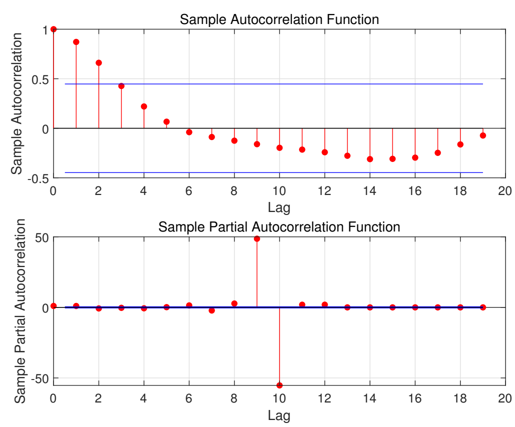
$$W_t = \nabla^d X_t, \quad t = d + 1, \dots, n$$

After that, apply stationary test on W_t and repeat steps above until it becomes a stationary series, Than W_t(which is denoted as X_t) complies ARMA model.

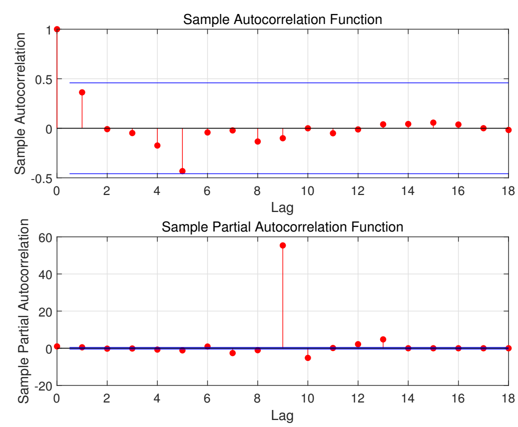
The figures below describe the result of ARIMA model on the time series of the size of finless porpoise $N_t, \quad t = 1 \cdots 30$, in which the Figure(4.2(c)) clearly shows that the number will decline and settle around 62 in the future 20 years.

4.4 Model III: Cellular Automata based population size prediction

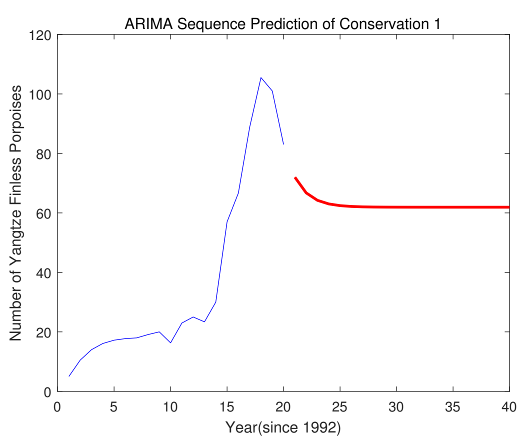
4.5 Model IV: Gray Forecast model



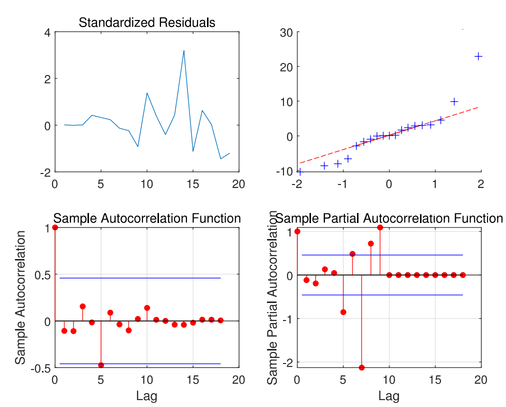
(a) Sample Autocorrelation Function and Sample Partial Autocorrelation Function



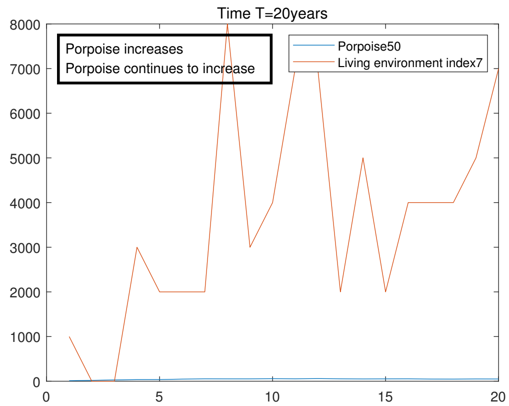
(b) Sample Autocorrelation Function and Sample Partial Autocorrelation Function



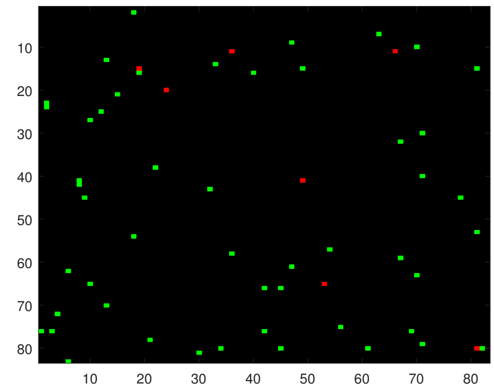
(c) ARIMA Sequence Prediction of Conservation



(d) Standardaized Residuals and QQ figure



(e) Living Environment Index and Population



(f) Cell Distribution

5

6 Sensitivity Test

7 Evaluation of Model

8 Conclusions

REPORT

To: 123

From: 123

Date: January 14, 2022

References

Zhigang, L. (2020). *The changes of micro-ecological of diseased yangtze finless and research on its protection under ex-situ* (Unpublished doctoral dissertation). Huazhong Agricultural University.

Appendices

Input matlab source:

```
clc;
clear;
x = [1992, 2002, 2005, 2007, 2015, 2021]    % Year
y = [5, 20, 25, 30, 57, 101]    % Number of Yangtze Finless Porpoises
xi = 1992:1:2022    % Prediction
yi = lagrange(x, y, xi)     % Lagrange Interpolation
plot(x, y, 'o', xi, yi, 'k')
title('Lagrange Interpolation Conservation 1')
xlabel('Year')
ylabel('Number of Yangtze Finless Porpoises')
```

```
function yy=lagrange(x,y,xx)    % Lagrange Function
m = length(x);
n = length(y);
if m~= n, error('Length of vector x and y should be the same');
end
s = 0;
for i = 1:n
    t = ones(1,length(xx));
    for j = 1:n
        if j~=i,
            t = t.*(xx - x(j))/(x(i) - x(j))    %Data (x, y) at interpolation point xx
        end
    end
    s = s + t * y(i)
end
yy = s
```

```
clc;
clear;
data = textread('conservation_1.txt');
data=nonzeros(data');     % Remove the zero elements in the order of the original data
r11=autocorr(data);        % Calculate the self correlation coefficient
r12=parcorr(data);     % Calculate partial correlation coefficient
figure
subplot(211),autocorr(data);
subplot(212),parcorr(data);    % The autocorrelation and partial autocorrelation of the ori
diff=diff(data);     % R11 is positive, not controlled by a negative exponential, so calcul
r21=autocorr(diff);        % Calculate the self correlation coefficient
r22=parcorr(diff);         % Calculate the partial correlation coefficient
adf=adftest(diff);     % If adf == 1, stable time sequence
figure
subplot(211),autocorr(diff);
subplot(212),parcorr(diff);       % Plot on the same figure, stable time sequence
n=length(diff);        % Caculate the differencial data
k=0;
  for i = 0:3
    for j = 0:3
```

```
        if i == 0 & j == 0
            continue
        elseif i == 0
            ToEstMd = arima('MALags', 1:j, 'Constant', 0);
        elseif j == 0
            ToEstMd = arima('ARLags', 1:i, 'Constant', 0);
        else
            ToEstMd = arima('ARLags', 1:i, 'MALags', 1:j, 'Constant', 0);     % Model struct
        end
        k = k + 1;
        R(k) = i;
        M(k) = j;
        [EstMd,EstParamCov,LogL,info] = estimate(ToEstMd,diff);   % Fitness, and estimate r
        numParams = sum(any(EstParamCov));
        [aic(k), bic(k)] = aicbic(LogL,numParams,n);
    end
end
fprintf('R, M, AIC, BIC value: \n%f');
check  = [R',M',aic',bic'];
res=infer(EstMd,diff);       % Verification
figure
subplot(2,2,1)
plot(res./sqrt(EstMd.Variance))      % Standardlized residual
title('Standardized Residuals')
subplot(2,2,2),qqplot(res)      % Fit the hypothesis of normality
subplot(2,2,3),autocorr(res)
subplot(2,2,4),parcorr(res)


% The autocorrelation coefficient rapidly decreases to 0 after 1 order lag,
% and the partial correlation coefficient is the same as the self correlation coefficient,
% so p = 1, q = 1
p=input('p = ');
q=input('q = ');
ToEstMd=arima('ARLags', 1:p, 'MALags', 1:q, 'Constant', 0);
 [EstMd, EstParamCov, LogL, info] = estimate(ToEstMd, diff);
 dx_forest = forecast(EstMd, 20, 'Y0', diff);       % 20 years prediction
 x_forest = data(end)+cumsum(dx_forest)
figure
h4 = plot(data, 'b');
hold on
h5 = plot(length(data)+1: length(data)+20, x_forest, 'r', 'LineWidth', 2);
title('ARIMA Sequence Prediction of Conservation 1')
xlabel('Year(since 1992)')
ylabel('Number of Yangtze Finless Porpoises')
hold off
```