

最优估计实验报告

姓名： 秦旗峰

学号： 2023302143029

专业： 导航工程（智能导航实验班）

课程： 最优估计 1

武汉大学

2024.11

目录

1 实验任务	3
2 数据计算和分析方法	3
2.1 实验编程工具	3
2.2 实验流程图	3
2.3 公式与函数	4
1) 读取文件数据	4
2) 时间转换	5
3) 去除 NaN 数据	5
4) 计算数字特征	5
5) 去除粗差	8
6) 时间序列图、直方图和相关系数图代码	9
3 数据分析和结果	12
3.1) 修正前后时间序列图	12
3.2) 修正前后直方图	13
3.3) 相关系数图	14
4 附录	14

1 实验任务

本次实验的任务主要为：熟悉 MATLAB 的使用；用 MATLAB 读取数据文件、计算数据的数字特征同时剔除异常数据，并对数据进行可视化处理。

数据文件名 C30ClockError.txt；数据文件记录的是北斗卫星 C30(MEO)在 2020 年 8 月 1 日-2020 年 10 月 21 日的卫星钟误差，采样间隔为 15min；数据格式说明：

第一行：

YearMonth	Day	Hour	Minute	Second	GPS Week	Second of Week	Delta T[m]
年 月	日	时	分	秒	GPS 周	周内秒	钟误差

数据部分示例：

2020 8 10 0 0 2116 518400 0.1040000000000000

钟误差单位为米 (m)；若出现 NaN 表明此处无有效数据

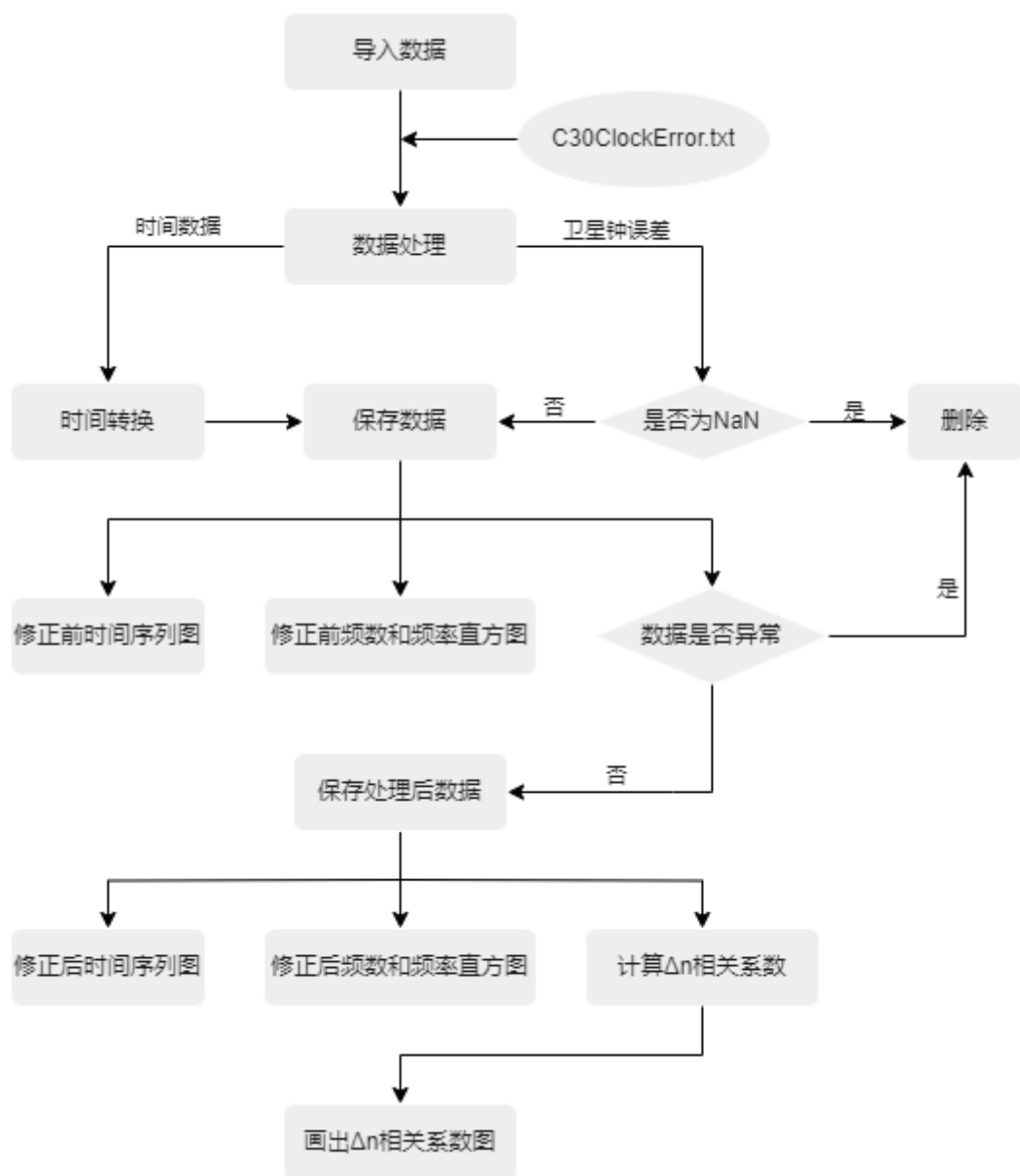
2 数据计算和分析方法

2.1 实验编程工具

本次实验使用的编程工具是 MATLAB R2024a

2.2 实验流程图

本次实验的流程图如下：



2.3 公式与函数

1) 读取文件数据

`importdata()` 函数：对于第一行有字符的数据文档，该函数可以将文本读取为：
`data, textdata, colheaders` 三个部分，`data` 部分为文件数据值，存储形式为矩阵；

```

data1=importdata("C30ClockError.txt");    %导入数据
data2=data1.data;    %获取data1中的数据
clear data1;    %删除data1，释放内存
year=data2(:,1);    %年
month=data2(:,2);    %月
day=data2(:,3);    %日
hour=data2(:,4);    %小时
minute=data2(:,5);    %分钟
second=data2(:,6);    %秒
dt=data2(:,9);    %clock error

```

2) 时间转换

datetime () 函数：可将传入的数据转换为标准的公历时间。

```

%将数据转化为标准时间格式
time=datetime(year,month,day,hour,minute,second);

```

3) 去除 NaN 数据

isnan () 函数：若传入参数为NaN，返回true，否则返回false;

```

%清除NaN异常数据
index=[];    %获取异常数据的索引值
for i=1:length(dt)
    if isnan(dt(i))
        index(end+1)=i;
    end
end
dt(index)=[];    %删除
time(index)=[];

```

4) 计算数字特征

4.1) 数学期望（均值）

对于离散随机变量X，数学期望定义为：

$$E(X) = u_X = \sum_{i=1}^{\infty} x_i p(x_i) \quad (4.1.1)$$

式中， x_i 为第*i*个随机变量样本值， $p(x_i)$ 是 x_i 出现的概率；在本次实验中，卫星钟误差可以看作是独立同分布的随机变量样本，每一个出现的概率相同，可以改写为：

$$E(X) = u_X = \frac{\sum_{i=1}^{\infty} x_i}{n} \quad (4.1.2)$$

mean()函数：传入一个数组或矩阵，返回数组或矩阵的均值；

`,mean(dt)`

4.2) 中误差

中误差反映的是随机变量与其期望的偏离程度，偏离程度越小，说明X的数值越稳定。随机变量X与它的均值 $E(X)$ 的偏离程度定义为：

$$Var(X) = E[(X - E(X))^2] \quad (4.2.1)$$

式中， $Var(X)$ 是随机变量X的二阶中心矩，也称为方差； $E(X)$ 是随机变量X的期望或均值。中误差由下式计算得到：

$$\sigma_X = \sqrt{Var(X)} \quad (4.2.2)$$

对于离散型随机变量X，当其期望 u_X 已知时，样本方差为：

$$S_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_X)^2, \sigma_X = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \mu_X)^2} \quad (4.2.3)$$

当随机变量期望未知时，样本方差为：

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \sigma_X = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \mu_X)^2} \quad (4.2.4)$$

在本实验中，随机变量的期望未知，采用(4.2.4)式计算中误差。

std () 函数：传入一个数组或矩阵，返回数组或矩阵的标准差；

```
std(dt)
```

4.3) (RMS) 均方根值

均方根值，也称方均根。对于离散型随机变量 X ，其 RMS 由如下公式求得：

$$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \quad (4.3.1)$$

```
%计算均方根值
sum=0;
for i=1:length(dt)
    sum=sum+dt(i)*dt(i);
end
rms=sqrt(sum/length(dt));
```

4.4) 计算 Δn 相关系数

Δn 相关系数计算公式如下：

$$\rho(\Delta n) = \frac{\frac{1}{N - \Delta n} \sum_{i=1}^{N - \Delta n} (x_i - \bar{X})(x_{i+\Delta n} - \bar{X})}{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})^2} \quad (4.4.1)$$

式中 N 表示样本容量， \bar{X} 表示样本均值。计算相关系数的代码如下：

```

%计算时间序列的相关系数
sum=0;
for i=1:length(dt)
    sum=sum+(dt(i)-mean(dt))*(dt(i)-mean(dt));
end
V=sum/length(dt); %计算分母
P=[]; %存放Δn对应的相关系数
index=[]; %存放Δn
for i=4:192
    index(end+1)=i;
    sum=0;
    for j=1:length(dt)-i
        sum=sum+(dt(j)-mean(dt))*(dt(j+i)-mean(dt));
    end
    sum=sum/(length(dt)-i); %计算分子
    P(end+1)=sum/V; %放入
end

```

5) 去除粗差

对于大量的偶然误差, 其分布服从高斯分布 $\Delta \sim N(0, \sigma_{\Delta}^2)$, 无论其方差为何值, 它在一定区间内出现的概率是不变的, 即

$$P(-m\sigma_{\Delta} < \Delta < m\sigma_{\Delta}) = \frac{1}{\sqrt{2\pi}\sigma_{\Delta}} \int_{-m\sigma_{\Delta}}^{m\sigma_{\Delta}} \exp\left\{-\frac{1}{2\sigma_{\Delta}^2}\Delta^2\right\} d\Delta \quad (5.1)$$

式中 m 为非负数。当 m 分别为 1, 2 和 3 是, 观测值在 $(-m\sigma_{\Delta}, m\sigma_{\Delta})$ 出现的概率为:

$$\begin{cases} P(-\sigma_{\Delta} < \Delta < 1\sigma_{\Delta}) \approx 68.3\% \\ P(-2\sigma_{\Delta} < \Delta < 2\sigma_{\Delta}) \approx 95.5\% \\ P(-3\sigma_{\Delta} < \Delta < 3\sigma_{\Delta}) \approx 99.7\% \end{cases} \quad (5.2)$$

(5.2)式表明, 随机误差在 3 倍标准差范围内出现的概率为99.7%, 超出此范围的概率仅为0.3%, 如果某个观测值在其均值和 3 倍标准差决定的置信区间外, 可将此观测值视为异常观测值 (粗差观测值), 检验并去除粗差的代码如下:


```

%清除大于3被方差的异常数据
index=[]; %获取异常数据的索引值
for i=1:length(dt)
    if abs(dt(i)-mean(dt))>3*std(dt)
        index(end+1)=i;
    end
end
dt(index)=[]; %删除
time(index)=[];

```

6) 时间序列图、直方图和相关系数图代码

6.1) 创建图窗

figure 命令，用于创建一个空白图窗，可以让一个脚本文件同时画出多幅图像。

```
figure; %创建一个图窗
```

6.2) 时间序列图

plot (x, y) 函数，在图窗中画出 $x - y$ 的图像，可以加入不同参数用于改变形式。

plot(time,dt,'DisplayName','C30'); 例如 'DisplayName','C30' 是创建名

'C30'的图例。 

图例常与 legend () 函数并用; legend('show'); %显示图例

annotation () 函数，在图窗上添加注解标签

```
annotation('textbox',dim,'String',str,'FitBoxToText','on'); %添加注解标签
```

xlabel ()、ylabel ()、title () 函数，分别添加 x 轴标签、y 轴标签和图片标题。

```

xlabel("时间");
ylabel("clock error");
title("BDS3 clock error 时间序列图");

```

总体代码如下：

```

%画未修正的时间序列图
figure; %创建一个图窗
plot(time,dt);
%添加注解标签
dim=[0.2 0.6 0.3 0.3];
str=sprintf('均值: %f;中误差: %f;RMS: %f',mean(dt),std(dt),rms); %注解字符串
annotation('textbox',dim,'String',str,'FitBoxToText','on'); %添加注解标签
clear str; %删除str, 释放内存
[maxValue, maxIndex] = max(dt); %获得最大值坐标
[minValue, minIndex] = min(dt); %获得最小值坐标
%为图像增加元素
xlabel("时间");
ylabel("clock error");
title("BDS3 clock error 时间序列图");
hold on;
plot(time(maxIndex), maxValue, 'r*', 'MarkerSize', 10); % 用红色星号标记最大值
text(time(maxIndex), maxValue, sprintf('最大值: %.4f', maxValue), ...
    'VerticalAlignment', 'bottom', 'HorizontalAlignment', 'right'); %最大值处标记
hold on;
plot(time(minIndex), minValue, 'b*', 'MarkerSize', 10); % 用蓝色星号标记最小值
text(time(minIndex), minValue, sprintf('最小值: %.4f', minValue), ...
    'VerticalAlignment', 'bottom', 'HorizontalAlignment', 'left'); %最小值处标记
legend('C30', '最大值点', '最小值点'); %显示图例

```

6.3) 直方图

histogram (dt) 函数, 画出 dt 数据的直方图, 默认为频数直方图。

```
histogram(dt); histogram(dt, 'Normalization', 'probability');
```

'Normalization','probability'表示画出频率直方图。

set(gca, 'YScale', 'log')函数, 可以改变坐标轴形式, 'YScale','log'表示 Y 轴用对数形式。

subplot () 函数, 可以将一个图窗分为多个画布, 实现多幅图像同时展现, 提高对比度。

总体代码如下:

```

%画频数直方图
figure('Position', [300, 300, 1000, 400]);
subplot(1,2,1);
histogram(dt);
set(gca, 'YScale', 'log');
xlabel("clock error");
ylabel("number");
title("BDS3 clock error 频数直方图");
legend('C30');

subplot(1,2,2);
histogram(dt, 'Normalization', 'probability');
xlabel("clock error");
ylabel("rate");
title("BDS3 clock error 频率直方图");
legend('C30');

```

6.4) 相关系数图

max () 函数，可以获取一个序列的最大值和最大值对应的索引

```
[maxValue, maxIndex] = max(P);    %存储最大值的坐标信息
```

相关系数图与时间序列图画法类似，总体代码如下：

```

%画出相关系数图像
figure;
plot(index,P,'-o');
xlabel("Δn");
ylabel("ρ (Δn)");
title("相关系数");
[maxValue, maxIndex] = max(P);    %存储最大值的坐标信息
[minValue, minIndex] = min(P);    %存储最小值的坐标信息
hold on;
plot(index(maxIndex), maxValue, 'r*', 'MarkerSize', 10); % 用红色星号标记最大值
text(index(maxIndex), maxValue, sprintf('Δn=%d 最大值: %.4f', index(maxIndex), maxValue), ...
    'VerticalAlignment', 'bottom', 'HorizontalAlignment', 'right'); %最大值处标记
hold on;
plot(index(minIndex), minValue, 'b*', 'MarkerSize', 10); % 用蓝色星号标记最小值
text(index(minIndex), minValue, sprintf('Δn=%d 最小值: %.4f', index(minIndex), minValue), ...
    'VerticalAlignment', 'bottom', 'HorizontalAlignment', 'left'); %最小值处标记
hold on;
yline(mean(P), 'r--', 'LabelHorizontalAlignment', 'left', 'LabelVerticalAlignment', 'middle');
%plot(index,mean(P),'g');
% 显示图例
legend('相关系数', '最大值', '最小值', '平均值线');

```

3 数据分析和结果

3.1) 修正前后时间序列图

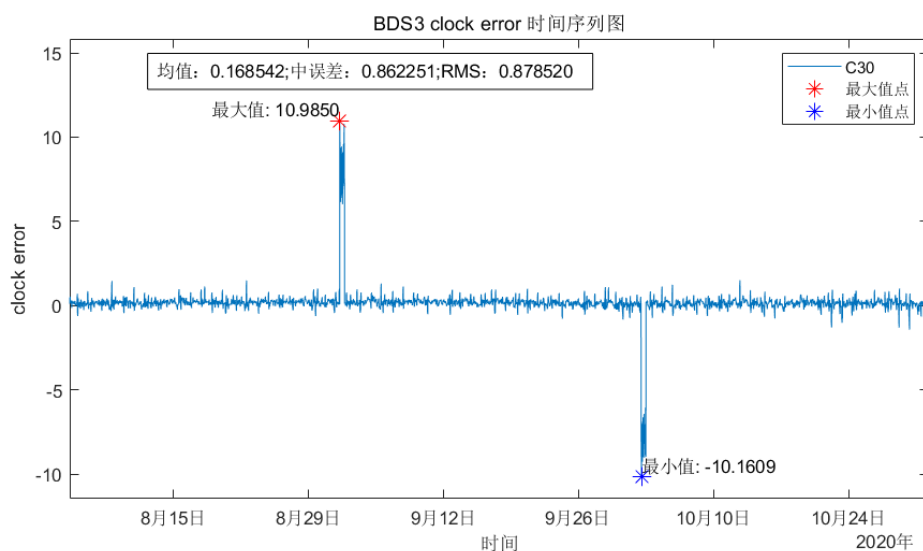


图 3.1.1: 修正前时间序列图

修正前, 数据范围为 $(-10.1609, 10.9850)$, 差异较大; 数据均值为 0.168542; 中误差为 0.862251; RMS 为 0.878520. 可以清晰地从时间序列图中观察到数据存在着粗差, 需要剔除。

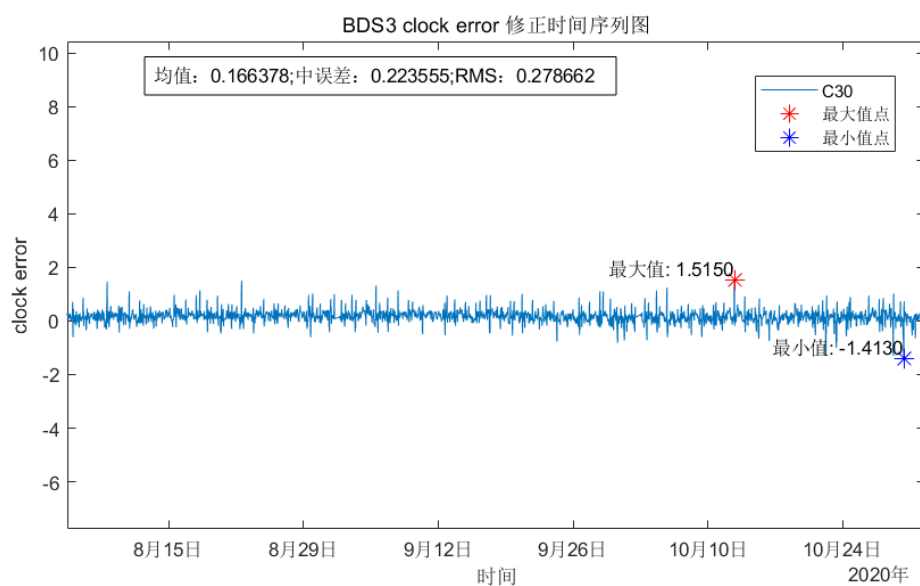


图 3.1.2: 修正后时间序列图

修正后，数据范围位于 $(-1.4130, 1.5150)$ ，差异显著下降；数据均值为 0.166378；中误差为 0.223555；RMS 为 0.278662。中误差和 RMS 显著降低，说明粗差得到了剔除。

3.2) 修正前后直方图

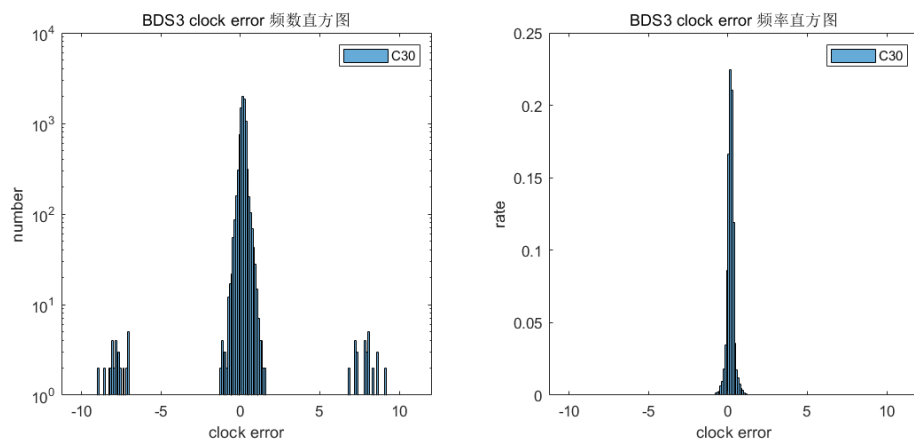


图 3.2.1：修正前直方图

修正前，在两端出现了异常分布的情况。

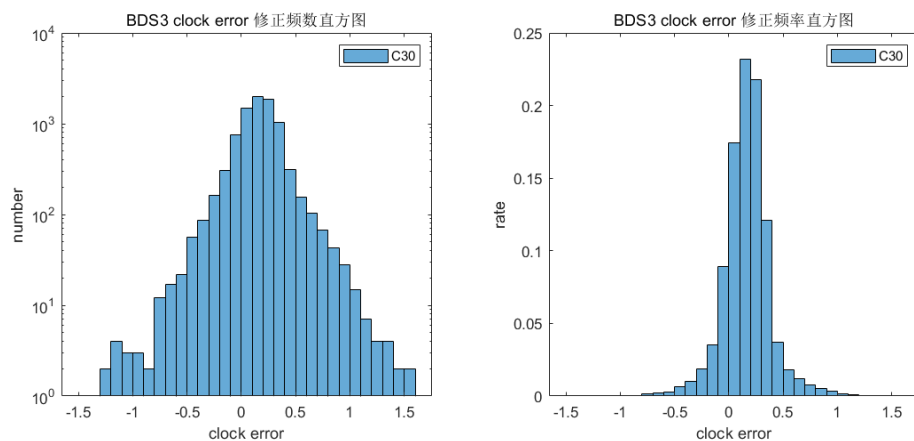


图 3.2.2：修正后直方图

修正后，图形分布得到明显改善

3.3) 相关系数图

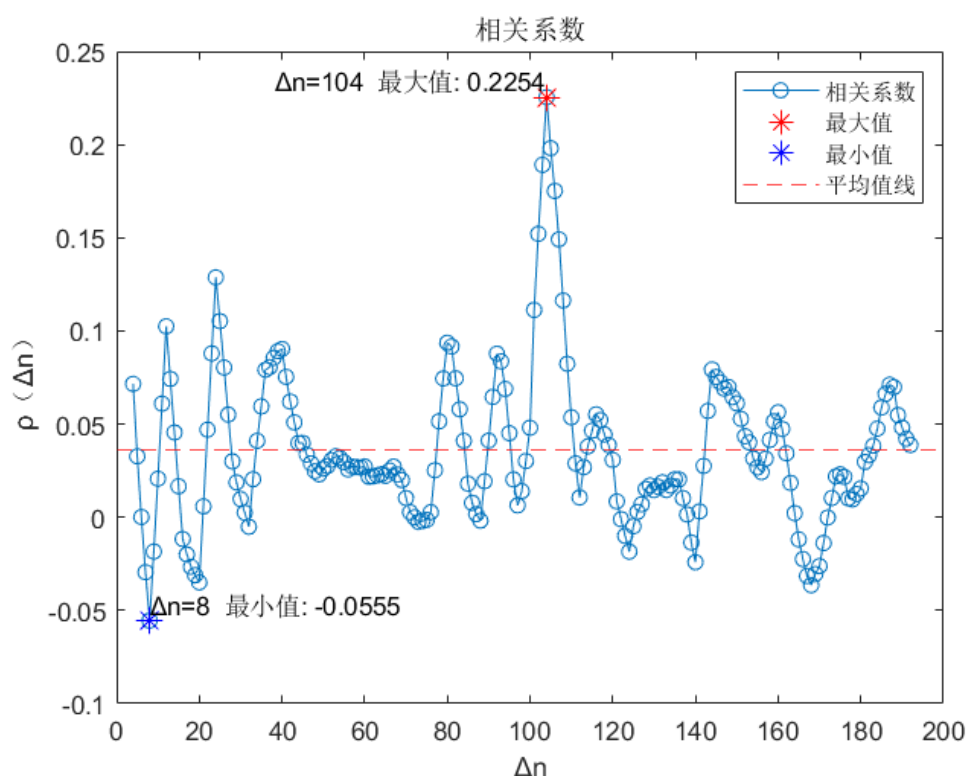


图 3.3: Δn 相关系数图

由图示可以观察得到当 $\Delta n = 104$ 时，相关系数有最大值为 0.2254.正相关性最大；当 $\Delta n = 8$ 时，相关系数有最小值为-0.055，负相关性最大。当 Δn 位于 101~108时相关系数相比之下更大，但总体来说，数据的相关系不够强。

4 附录

代码附录：

```
clc;clear;
data1=importdata("C30ClockError.txt");    %导入数据
data2=data1.data;    %获取 data1 中的数据
clear data1;    %删除 data1，释放内存
year=data2(:,1);    %年
month=data2(:,2);    %月
day=data2(:,3);    %日
```

```

hour=data2(:,4);    %小时
minute=data2(:,5); %分钟
second=data2(:,6); %秒
dt=data2(:,9);     %clock error

%将数据转化为标准时间格式
time=datetime(year,month,day,hour,minute,second);

%清除 NaN 异常数据
index=[];          %获取异常数据的索引值
for i=1:length(dt)
    if isnan(dt(i))
        index(end+1)=i;
    end
end
dt(index)=[];      %删除
time(index)=[];

%计算均方根值
sum=0;
for i=1:length(dt)
    sum=sum+dt(i)*dt(i);
end
rms=sqrt(sum/length(dt));

%画未修正的时间序列图
figure;           %创建一个图窗
plot(time,dt);
%添加注解标签
dim=[0.2 0.6 0.3 0.3];
str=sprintf('均值: %f;中误差: %f;RMS: %f',mean(dt),std(dt),rms); %注解字符串
annotation('textbox',dim,'String',str,'FitBoxToText','on'); %添加注解标签
clear str;        %删除 str, 释放内存
[maxValue, maxIndex] = max(dt); %获得最大值坐标
[minValue, minIndex] = min(dt); %获得最小值坐标
%为图像增加元素
xlabel("时间");
ylabel("clock error");
title("BDS3 clock error 时间序列图");
hold on;
plot(time(maxIndex), maxValue, 'r*', 'MarkerSize', 10); % 用红色星号标记最大值
text(time(maxIndex), maxValue, sprintf('最大值: %.4f', maxValue), ...
    'VerticalAlignment', 'bottom', 'HorizontalAlignment', 'right'); %最大值处标记
hold on;

```

```

plot(time(minIndex), minValue, 'b*', 'MarkerSize', 10); % 用蓝色星号标记最小值
text(time(minIndex), minValue, sprintf('最小值: %.4f', minValue), ...
    'VerticalAlignment', 'bottom', 'HorizontalAlignment', 'left'); %最小值处标记
legend('C30','最大值点','最小值点'); %显示图例

```

%画频数直方图

```

figure('Position', [300, 300, 1000, 400]);
subplot(1,2,1);
histogram(dt);
set(gca, 'YScale', 'log');
xlabel("clock error");
ylabel("number");
title("BDS3 clock error 频数直方图");
legend('C30');

```

```

subplot(1,2,2);
histogram(dt, 'Normalization', 'probability');
xlabel("clock error");
ylabel("rate");
title("BDS3 clock error 频率直方图");
legend('C30');

```

%清除大于 3 被方差的异常数据

```

index=[]; %获取异常数据的索引值
for i=1:length(dt)
    if abs(dt(i)-mean(dt))>3*std(dt)
        index(end+1)=i;
    end
end
dt(index)=[]; %删除
time(index)=[];

```

%修正数据后

%计算均方根值

```

sum=0;
for i=1:length(dt)
    sum=sum+dt(i)*dt(i);
end
rms=sqrt(sum/length(dt));

```

%画修正后的时间序列图


```

figure;
plot(time,dt);
%添加注解
dim=[0.2 0.6 0.3 0.3];
str=sprintf('均值: %f;中误差: %f;RMS: %f',mean(dt),std(dt),rms);
annotation('textbox',dim,'String',str,'FitBoxToText','on');
clear str; %释放内存
[maxValue, maxIndex] = max(dt); %获取最大值坐标
[minValue, minIndex] = min(dt); %获取最小值坐标
%增加元素
xlabel("时间");
ylabel("clock error");
title("BDS3 clock error 修正时间序列图");
ylim([-10,10]); %改变 y 轴显示范围, 使得对比更明显
hold on;
plot(time(maxIndex), maxValue, 'r*', 'MarkerSize', 10); % 用红色星号标记最大值
text(time(maxIndex), maxValue, sprintf('最大值: %.4f', maxValue), ...
      'VerticalAlignment', 'bottom', 'HorizontalAlignment', 'right'); %最大值处标记
hold on;
plot(time(minIndex), minValue, 'b*', 'MarkerSize', 10); % 用蓝色星号标记最小值
text(time(minIndex), minValue, sprintf('最小值: %.4f', minValue), ...
      'VerticalAlignment', 'bottom', 'HorizontalAlignment', 'right'); %最小值处标记
legend('C30','最大值点','最小值点'); %显示图例

%画出修正后的频数直方图
figure('Position', [100, 100, 1000, 400]);
subplot(1,2,1);
histogram(dt,'BinWidth',0.1);
set(gca, 'YScale', 'log');
xlabel("clock error");
ylabel("number");
title("BDS3 clock error 修正频数直方图");
legend('C30');

subplot(1,2,2);
histogram(dt, 'Normalization', 'probability', 'BinWidth', 0.1);
xlabel("clock error");
ylabel("rate");
title("BDS3 clock error 修正频率直方图");
legend('C30');

%计算时间序列的相关系数
sum=0;
for i=1:length(dt)

```

```

        sum=sum+(dt(i)-mean(dt))*(dt(i)-mean(dt));
    end
    V=sum/(length(dt)-1);      %计算分母
    P=[];      %存放  $\Delta n$  对应的相关系数

    index=[];  %存放  $\Delta n$ 
    for i=4:192
        index(end+1)=i;
        sum=0;
        for j=1:length(dt)-i
            sum=sum+(dt(j)-mean(dt))*(dt(j+i)-mean(dt));
        end
        sum=sum/(length(dt)-i);      %计算分子
        P(end+1)=sum/V;      %放入
    end

    %画出相关系数图像
    figure;
    plot(index,P,'-o');
    xlabel(" $\Delta n$ ");
    ylabel("p ( $\Delta n$ ) ");
    title("相关系数");
    [maxValue, maxIndex] = max(P);      %存储最大值的坐标信息
    [minValue,minIndex]=min(P);      %存储最小值的坐标信息
    hold on;
    plot(index(maxIndex), maxValue, 'r*', 'MarkerSize', 10); % 用红色星号标记最大值
    text(index(maxIndex), maxValue, sprintf('Δn=%d 最大值: %.4f',index(maxIndex), maxValue), ...
        'VerticalAlignment', 'bottom', 'HorizontalAlignment', 'right'); %最大值处标记
    hold on;
    plot(index(minIndex), minValue, 'b*', 'MarkerSize', 10); % 用蓝色星号标记最小值
    text(index(minIndex), minValue, sprintf('Δn=%d 最小值: %.4f',index(minIndex), minValue), ...
        'VerticalAlignment', 'bottom', 'HorizontalAlignment', 'left'); %最小值处标记
    hold on;
    yline(mean(P), 'r--', 'LabelHorizontalAlignment', 'left', 'LabelVerticalAlignment', 'middle');
    %plot(index,mean(P),'g');
    % 显示图例
    legend('相关系数','最大值','最小值','平均值线');
    clear;

```