

# 后疫情时代及COVID-19突变株影响与周期性的研究

中山大学 物理学院

郭蒙 20343097

## 1. 前言

### 1.1 背景介绍

自2020年初COVID-19爆发以来，国际形势大变，并且对国际金融造成了重创。尽管这近两年以来对该问题，无论是传染病模型还是国际经济发展的研究都已经比较详细和全面。但是COVID-19的突变株相继出现，对整个国际形势仍然有决定性的影响，我们尝试重复一些较为简单的拟合研究，并且尝试找到引起国际注意的高传染性突变株出现的周期性规律并分析。

### 1.2 研究方法

本文通过利用Mathematica集数据处理分析、符号拟合运算和Wolfram丰富的数据库的支撑，辅以其他渠道得到的.csv文件的数据，在Mathematica中利用可视化和其他功能运算进行分析并得出结论。

## 2. 病毒初期中国湖北省疫情病例曲线及建模

### 2.1 数据导入和分类

利用Wolfram DataBase中自带的Wolfram DataSet格式的数据库直接进行数据抓取

```
In[*]:= DeleteObject[
```

删除对象

```
ResourceObject["Epidemic Data for Novel Coronavirus COVID-19"]]
```

资源对象

(\*删除缓存的对象以便于导入最新的数据\*)

```
In[*]:= datacn =
```

```
ResourceData["Epidemic Data for Novel Coronavirus COVID-19", "ChinaProvinces"]
```

资源数据

(\*导入中国疫情表 数据范围——自2020.1.22到2021.12.12\*)

AdministrativeDivision	Country	GeoPosition	ConfirmedCases
Hubei, China	China	30.98°N 112.3°E	TimeSeries[

Out[ ]=


Guangdong, China	China	23.34°N 113.4°E	TimeSeries [
Shanghai, China	China	31.2°N 121.4°E	TimeSeries [
Heilongjiang, China	China	47.86°N 127.8°E	TimeSeries [
Yunnan, China	China	24.97°N 101.5°E	TimeSeries [
Zhejiang, China	China	29.18°N 120.1°E	TimeSeries [
Henan, China	China	37.9°N 114.9°E	TimeSeries [
Jiangsu, China	China	32.97°N 119.5°E	TimeSeries [
Hebei, China	China	39.55°N 116.1°E	TimeSeries [
Fujian, China	China	26.08°N 118.0°E	TimeSeries [
Sichuan, China	China	30.62°N 102.7°E	TimeSeries [
Hunan, China	China	27.61°N 111.7°E	TimeSeries [
Beijing, China	China	40.18°N 116.4°E	TimeSeries [
Neimenggu, China	China	44.09°N 113.9°E	TimeSeries [
Shandong, China	China	36.34°N 118.1°E	TimeSeries [
Anhui, China	China	31.83°N 117.2°E	TimeSeries [
Xinjiang, China	China	41.11°N 85.24°E	TimeSeries [


Jiangxi, China	China	27.61°N 115.7°E	TimeSeries [
Liaoning, China	China	41.3°N 122.6°E	TimeSeries [
Shaanxi, China	China	35.19°N 108.9°E	TimeSeries [


rows 1–20 of 31

我们只对疫情爆发初期的曲线感兴趣，我们将中国湖北省自2020.1.22至2020.4.30日的确诊、死亡和治愈的病例进行提取

```
templist =
Normal@
|转换为普通表达式
datacn[
SelectFirst[#AdministrativeDivision == Hubei, China ADMINISTRATIVE DIVISION ☒ &],
|选择第一个
{#ConfirmedCases - (#RecoveredCases + #Deaths), #RecoveredCases, #Deaths} &];
(*截断时间至我们将要研究的时间*)
{Hubeiinfected, Hubeirecovered, Hubeideceased} =
TimeSeriesWindow[#, {{2020, 1, 22}, {2020, 4, 30}}] & /@ templist
|时间序列窗
```

Out[ ] = {TimeSeries [  Time: 22 Jan 2020 to 30 Apr 2020 Data points: 100 ],

TimeSeries [  Time: 22 Jan 2020 to 30 Apr 2020 Data points: 100 ],

TimeSeries [  Time: 22 Jan 2020 to 30 Apr 2020 Data points: 100 ] }

## 2.2 数据整体可视化

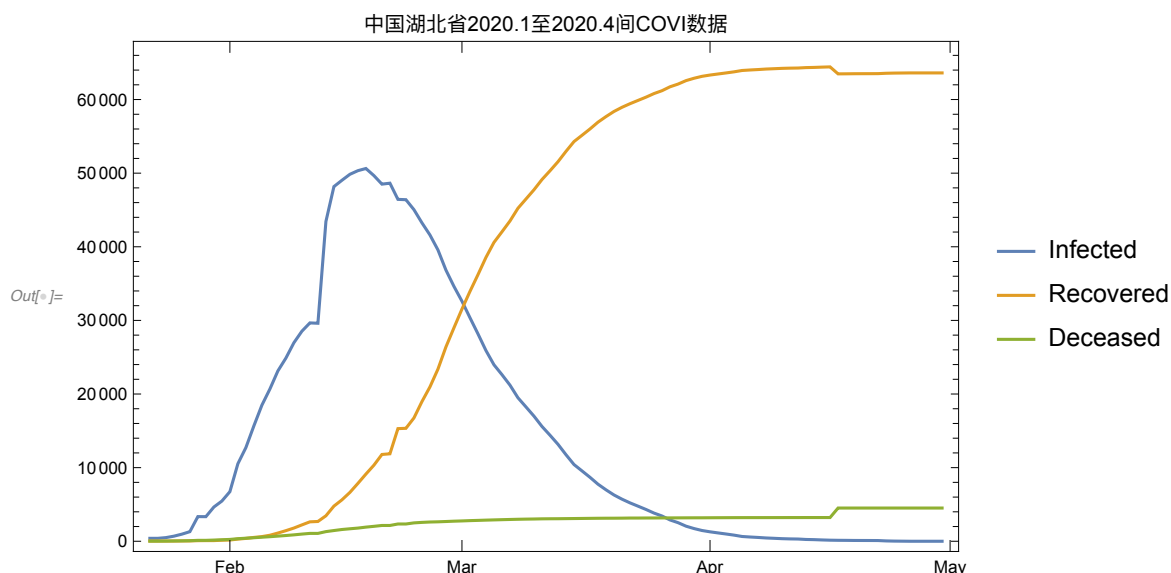
```
DateListPlot[{Hubeiinfected, Hubeirecovered, Hubeideceased},
|日期列表图
```

```
PlotLabel -> "中国湖北省2020.1至2020.4间COVID-19数据",
```

```
|绘图标签
```

```
PlotLegends -> {"Infected", "Recovered", "Deceased"}]
```

```
|绘图的图例
```



## 2.3 使用SIR模型进行拟合

下面利用传染病动力学中最为经典的SIR模型对该组数据进行拟合，下面简要介绍一下SIR模型——SIR模型的建立基于以下三个假设：

- 不考虑人口的出生、死亡、流动等种群动力因素。人口始终保持一个常数，即 $N(t) \equiv K$
- 一个病人一旦与易感者接触就必然具有一定的传染性。假设 $t$ 时刻单位时间内，一个病人能传染的易感者数目与此环境内易感者总数 $s(t)$ 成正比，比例系数为 $\beta$ ，从而在 $t$ 时刻单位时间内被所有病人传染的人数为 $\beta s(t)$ 。
- $t$ 时刻，单位时间内从染病者中移出的人数与病人数量成正比，比例系数为 $\gamma$ ，单位时间内移出者的数量为 $\gamma i(t)$

基于以上假设，我们有感染机制的模型描述

$$\begin{cases} S(i) + I(j) \xrightarrow{\beta} I(i) + I(j) \\ I(i) \xrightarrow{\gamma} R(i) \end{cases}$$

基于以上理论，我们给出微分方程形式的SIR模型并用于拟合

```
In[ ]:= odeSIR = {S'[t] == -β S[t]  $\frac{I[t]}{N}$ , I'[t] == β S[t]  $\frac{I[t]}{N}$  - γ I[t], R'[t] == γ I[t]};
```

(\*在给定时=0时刚出现感染情况的初值情况下, 获得参数数值解\*)

```
pmNDSolve = ParametricNDSolve[{odeSIR, S[0] == N - i0, I[0] == i0, R[0] == 0},  
[参数数值解]
```

```
{S, I, R}, {t, 0, 100}, {N, i0, β, γ}];
```

(\*回代至ODE中\*)

```
{solI, solR} = {I, R} /. pmNDSolve
```

```
Out[ ]:= {ParametricFunction[ Expression: I  
Parameters: {N, i0, β, γ}],  
  
ParametricFunction[ Expression: R  
Parameters: {N, i0, β, γ}]}
```

```
In[ ]:= (*定义拟合模型*)
```

```
fitModelSIR[N_, β_, γ_][tag_, t_] :=
```

```
Which[tag == 1, solI[N, 1, β, γ][t], tag == 2, solR[N, 1, β, γ][t]]
```

[Which循环]

由于前文的数据仍是含时间序列的列表, 需要改写为顺序序列的列表才能参与拟合

```
In[ ]:= Length[Normal@Hubeiinfected]
```

[长度] [转换为普通表达式]

```
Out[ ]:= 100
```

```
HubeiinfectedData = MapIndexed[{#2[[1]], #1} &][Hubeiinfected["Values"]][[;; 100];  
[映射索引] [获取值]
```

```
HubeideceasedData = MapIndexed[{#2[[1]], #1} &][Hubeideceased["Values"]][[;; 100];  
[映射索引] [获取值]
```

整合在统一同一数据集中

```
In[ ]:= dataSIR = Join[Prepend[#, 1] & /@ HubeiinfectedData,  
[连接] [加在前面]
```

```
Prepend[#, 2] & /@ HubeirecoveredData];  
[加在前面]
```

```
In[ ]:= fitSIR = NonlinearModelFit[dataSIR, fitModelSIR[N, β, γ][tag, t],  
[非线性拟合模型]
```

```
{{N, 12000}, {β, 0.27}, {γ, 0.03}}, {tag, t}];
```

(\*拟合后给出可能的参数条件\*)

```
In[ ]:= fitSIR["ParameterTable"]
```

	Estimate	Standard Error	t-Statistic	P-Value
N	67613.3	726.955	93.0089	$1.00305 \times 10^{-164}$
β	0.563189	0.00669564	84.1127	$2.46842 \times 10^{-156}$
γ	0.0458924	0.0012546	36.5794	$9.02522 \times 10^{-90}$

```
In[ ]:= fitSIR["BestFitParameters"]
```

```
Out[ ]:= {N → 67613.3, β → 0.563189, γ → 0.0458924}
```

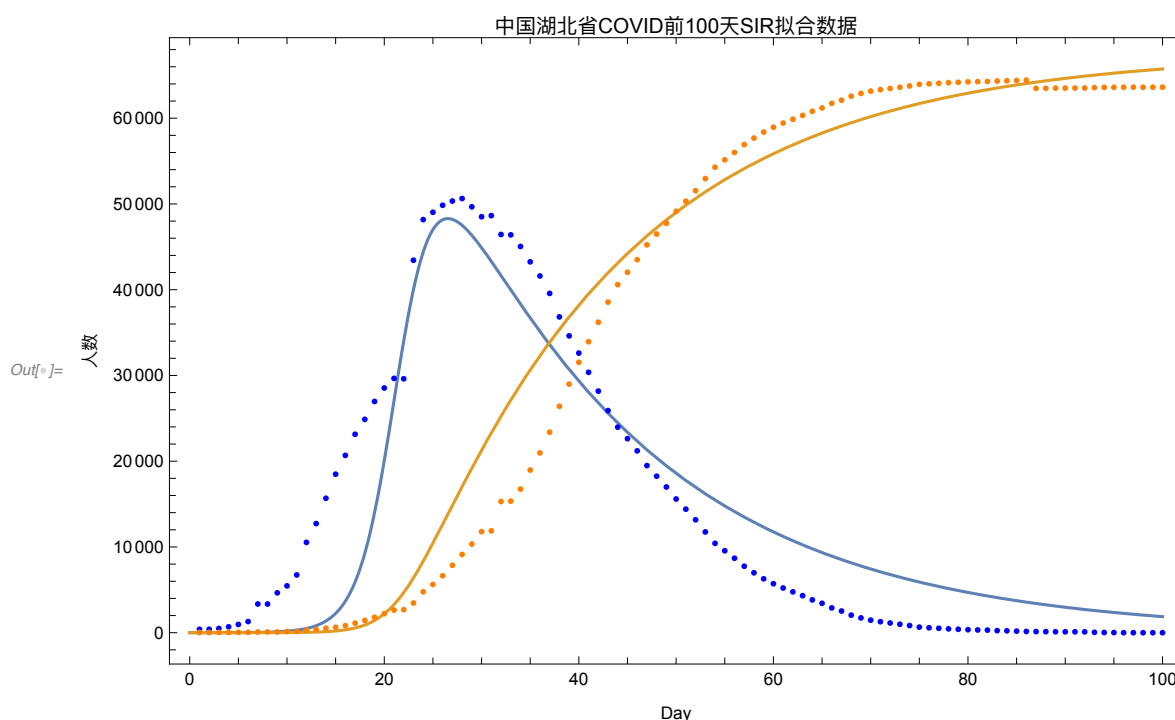
在最佳参数下绘出拟合图像

```

Plot[{solI[N, 1,  $\beta$ ,  $\gamma$ ][t] /. fitSIR["BestFitParameters"],
  solR[N, 1,  $\beta$ ,  $\gamma$ ][t] /. fitSIR["BestFitParameters"]}, {t, 0, 100},

Epilog -> {Blue, Point@HubeiinfectedData, Orange, Point@HubeirecoveredData},
Frame -> True,
FrameLabel -> {"Day", "人数"},
PlotLabel -> "中国湖北省COVID-19前100天SIR拟合数据",
ImageSize -> Large]

```



由上图可知，在拟合模型下，对于峰值和拐点有着较为合理的预测，但是在过渡部分仍然不太准确，是真实情况过于复杂和拟合模型的参数数量较少所致的。我们可以看见，在Day21左右，确诊病例因为重新统计更新有着较大的改变，在其余位置，对模型的拟合还是较为准确的。

## 2.4 拟合模型的分析与展望

在上文我们采用了SIR中最为简单的拟合形式，我们不得不承认尽管对于理想模型来讲其微分方程的形式已经较为复杂，但是对与真实事件的模拟还是略显无力，可以看见，上图在基本趋势和峰值位置上已经和真实情况略为接近。对于SIR模型，还有增加单位时间治愈人数的参数变量，但是本文的重心并不在单一优化传染病模型上，对于一年前的疫情数据，在互联网上已经能够轻易的找到模型更为精确的拟合，我们在此不在重复。只是通过一个较为简单的例子展现Mathematica集数据收集、处理、分析和可视化为一体的能力，接下来我们会分析在各国防疫政策已基本稳定和COVID在Delta变异株和OmiCron变异株影响下其他值得研究的特征。

### 3. Delta和Omicron变异株对各国COVID-19曲线的影响

#### 3.1 基本介绍和数据处理

新冠疫情在2021年的国际防控形式已经逐渐常态化，但是隔一段时间就会有的突变株对国际防控形式会起到决定性的作用，尽管其重病率和死亡率在医疗条件稳定的情况下已经较为乐观，但是其高传染性给各国带来的压力仍然不少。

下面以Delta和Omicron变异株为例，尝试寻找出COVID-19突变的规律和其他特征。

#### 3.2 近一年来的世界确诊病例曲线

获取数据集中世界国家的确诊曲线并处理

```
In[ ]:= wdData = ResourceData[  
    资源数据
```

```
"Epidemic Data for Novel Coronavirus COVID-19", "WorldCountries"];
```

我们以美国和英国为例，做出2020.9.1-2021.12.12的数据曲线

```
In[ ]:= UsAndUkConfirmedCases = wdData[  
    Select[#Country == United States COUNTRY ☒ ||  
    选择  
    #Country == United Kingdom COUNTRY ☒ &],  
    {#Country, #ConfirmedCases} &]
```

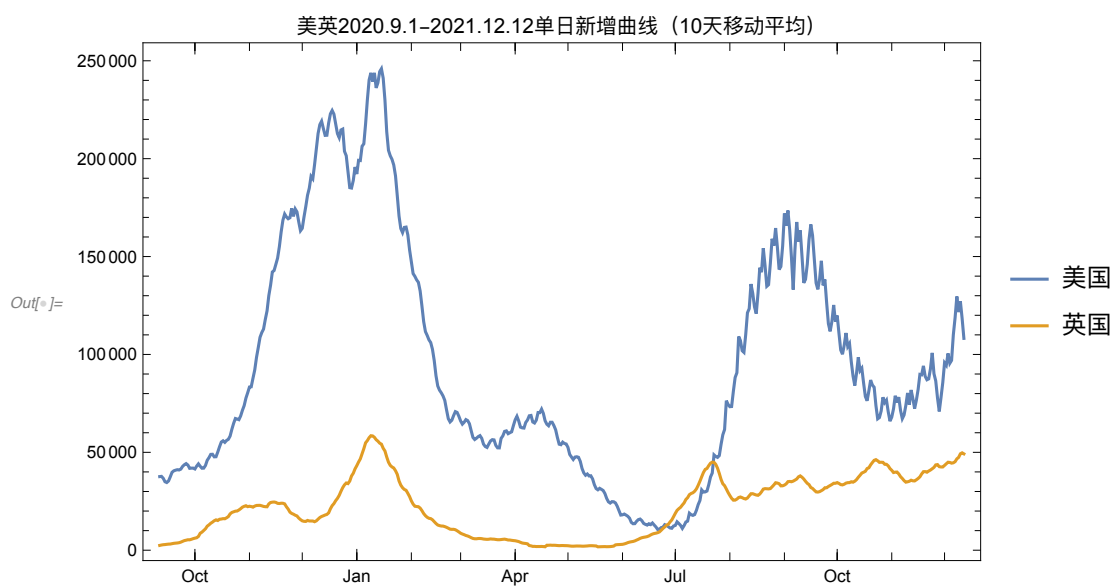
United States	TimeSeries[ Time: 22 Jan 2020 to 12 Dec 2021 Data points: 691 ]
United Kingdom	TimeSeries[ Time: 22 Jan 2020 to 12 Dec 2021 Data points: 691 ]

Out[ ]:=

```

In[ ]:= DateListPlot[
  [日期列表图]
  {MovingAverage[
    [移动平均]
    Differences[
    [差分]
    TimeSeriesWindow[UsAndUkConfirmedCases[1, 2],
    [时间序列窗]
    {{2020, 9, 1}, {2021, 12, 12}}]]
    , 10],
  MovingAverage[
  [移动平均]
  Differences[
  [差分]
  TimeSeriesWindow[UsAndUkConfirmedCases[2, 2],
  [时间序列窗]
  {{2020, 9, 1}, {2021, 12, 12}}]]
    , 10]],
  PlotLabel -> "美英2020.9.1-2021.12.12单日新增曲线（10天移动平均）",
  [绘图标签]
  PlotLegends -> {"美国", "英国"}]
  [绘图的图例]

```



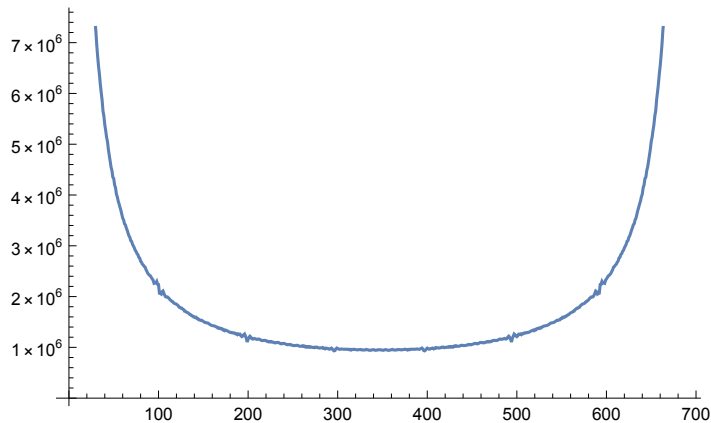
(\*第一次尝试寻找是否存在周期性规律\*)

```

Fourier[
  [傅立叶]
  UsAndUkConfirmedCases[1, 2][["Values"]] // Abs // ListLinePlot
  [获取值] [绝对值] [绘制点集的线条]

```





(\*未能获得明显的周期性结果\*)

通过查询资料得到，Delta变异株于2020.10在印度被发现，以高传染性著称；

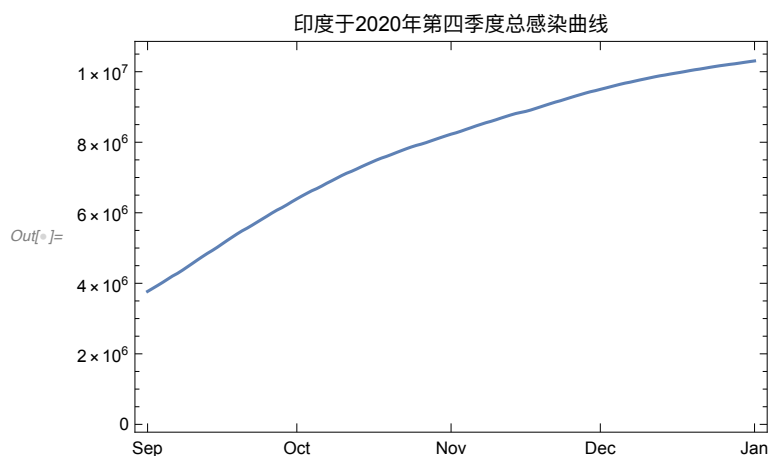
Omicron于2021.12月在南非发现，后面我们简单比较一下Delta和Omicron在爆发地的确诊曲线。

### 3.2 Delta突变株爆发地曲线

```
In[ ]:= IndiaConfirmCases = wdData[
  Select[#Country == India COUNTRY &],
  {#Country, #ConfirmedCases} &];
IndiaConfirmCasestimeWindow = TimeSeriesWindow[
  IndiaConfirmCases[1, 2],
  {{2020, 9, 1}, {2021, 1, 1}}]
```

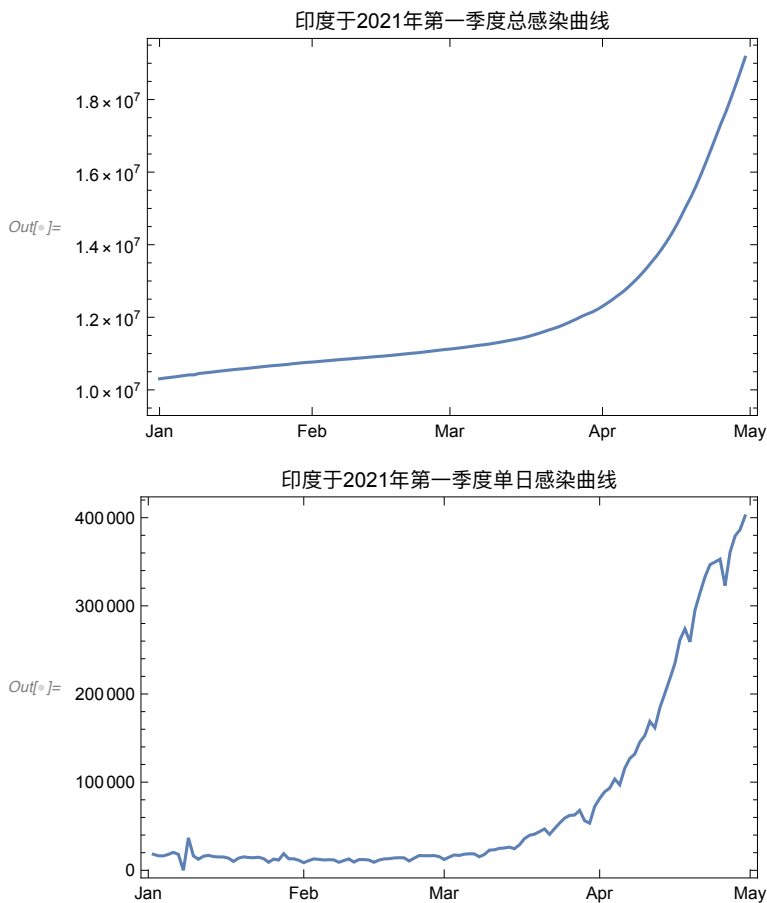
```
Out[ ]:= TimeSeries[
  Time: 01 Sep 2020 to 01 Jan 2021
  Data points: 123]
```

```
In[ ]:= DateListPlot[IndiaConfirmCasestimeWindow,
  PlotLabel -> "印度于2020年第四季度总感染曲线"]
```



我们发现在爆发当时并没有达到感染率高发期，我们进一步拉宽时间进行分析

```
In[ ]:= IndiaConfirmCasestimeWindow2 = TimeSeriesWindow[
    IndiaConfirmCases[1, 2],
    {{2021, 1, 1}, {2021, 4, 30}}];
DateListPlot[%,
    PlotLabel → "印度于2021年第一季度总感染曲线"]
DateListPlot[Differences[IndiaConfirmCasestimeWindow2],
    PlotLabel → "印度于2021年第一季度单日感染曲线"]
```



考虑到病毒从发现到爆发的延迟期，由于各国的医疗水平和条件不一样，我们无法预知不同国家发现新突变病毒的能力。但是值得注意的是，上图分别为总增长曲线和每日新增曲线，但是两者走势近乎一样，说明单日新增的量级已经和总体量相当！

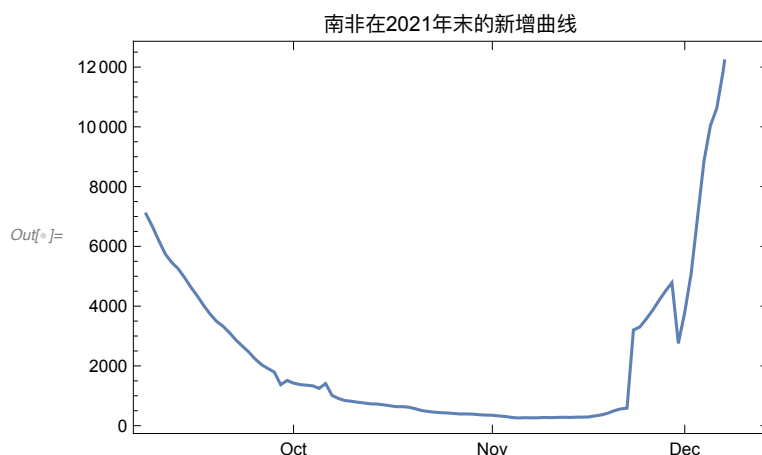
### 3.3 Omicron突变株爆发地曲线

```
In[ ]:= SouthAfConfirmedCases = wdData[
  Select[#Country == South Africa COUNTRY ☒] &]
  , {#Country, #ConfirmedCases} &]
```

Out[ ]:=

South Africa	TimeSeries[ Time: 22 Jan 2020 to 12 Dec 2021 Data points: 691
--------------	---------------------------------------------------------------------

```
In[ ]:= DateListPlot[MovingAverage[
  Differences[
    TimeSeriesWindow[SouthAfConfirmedCases[1, 2],
      {{2021, 9, 1}, {2021, 12, 12}}], 7],
  PlotLabel -> "南非在2021年末的新增曲线"]
```

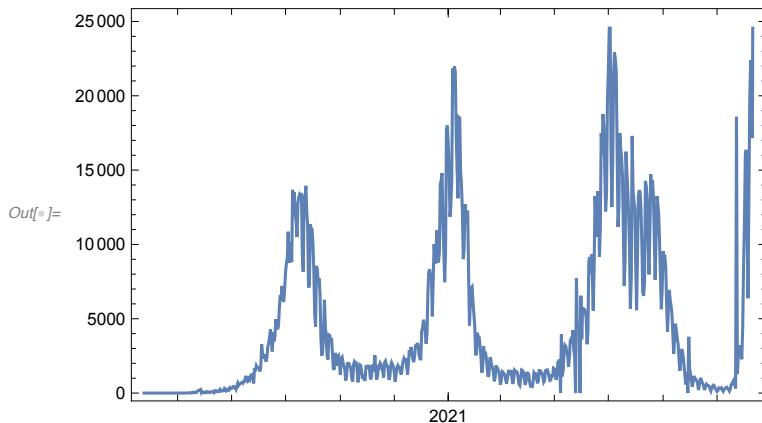


Omicron变异株从发现到爆发的时间远远低于印度的Delta变异株，这一方面可能是两个国家人口基数和医疗条件的差别，当然也有其他（包括气候等因素）的影响，但是由于变量太过复杂，最重要的原因是我们暂时还无法预知Omicron的传染峰值是否已经来到（因为2021.12才于南非发现）我们暂无法进行进一步讨论。

### 3.4 世界疫情曲线的周期性（突变周期性）的分析与预测

我们在研究上节的时候发现南非的确诊曲线有着极为良好的周期性

```
In[*]:= SouthAfConfirmedCases[1, 2] // Differences // DateListPlot
```



(图中有一个很有意思的现象，在具有较为良好的周期性的同时，每次大规模的感染时的峰值总是逐渐拔高，变相暗示了每个突变株的感染能力逐渐升高，这是很不乐观的一件事)

我们猜测，在南非这样的医疗水平较为低下的国家，政府医疗系统对疫情的干涉能力非常小，且疫苗接种率仍然较低，一定程度上能反映较为简单的卫生条件的情况下病毒传染的性质。

我们使用傅立叶变换对单日新增的数据进行周期性分析，这里需要注意的是，由于每日的波动不能客观的反映一段时间的趋势，我们使用合理移动平均对其先进行处理。

```
In[*]:= (*将时间序列转化为一般值序列*)
```

```
SouthAfConfirmedCasesValue = SouthAfConfirmedCases[1, 2][["Values"]];
```

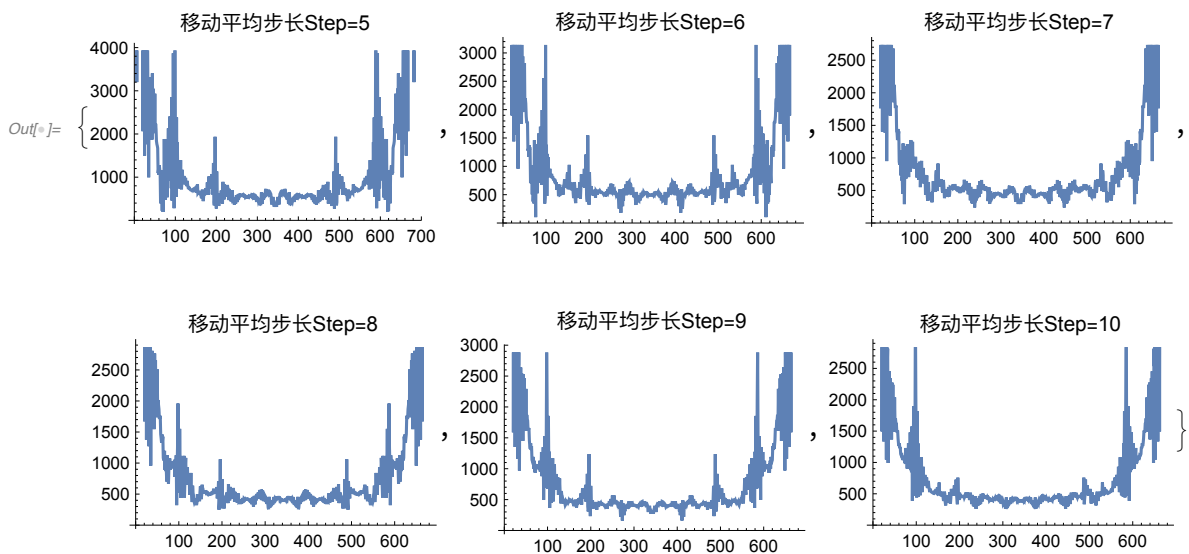
```
% // Short
```

```
Out[*]//Short= {0, 0, 0, 0, 0, 0, <<679>>, 3 051 222,
3 071 064, 3 093 452, 3 112 469, 3 129 622, 3 167 497}
```

```

In[ ]:= Table[
  |表格
  ListLinePlot[Fourier[
    |绘制点集的线条 |傅立叶
    MovingAverage[
      |移动平均
      Differences[SouthAfConfirmedCasesValue], step]] // Abs,
    |差分 |绝对值
    PlotLabel -> "移动平均步长Step=" <> ToString[step]],
    |绘图标签 |转换为字符串
    {step, 5, 10, 1}]

```

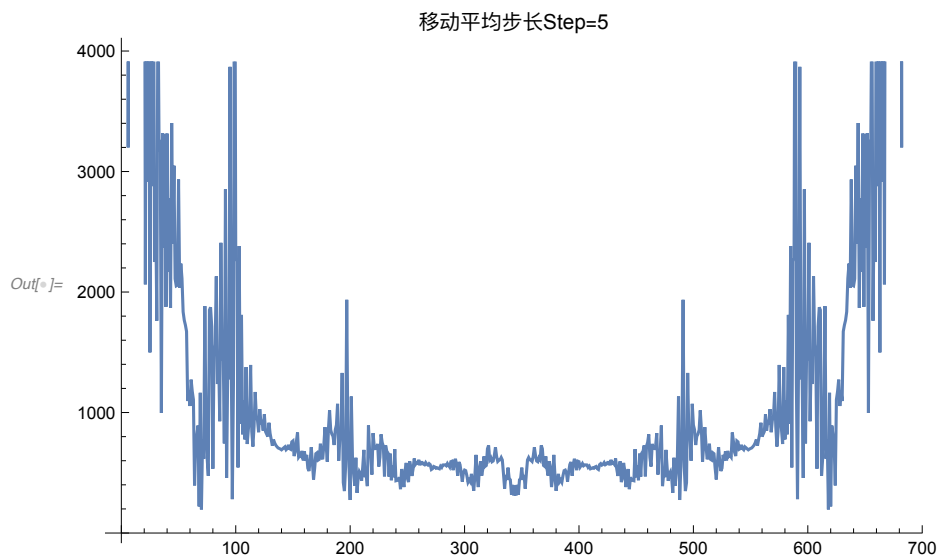


我们发现四个突出的峰经常出现，大约在100、200、500、600天之间，由于COVID-19出现时间仅仅在2年左右，对于我们现在的数据而言，我们舍掉大于一年的可能周期，在100、200天之间，我们选择周期特征更大的100天，并且取移动平均的平均步长为5时，绘出更细节的图像。

```

In[ ]:= ListLinePlot[Fourier[
  |绘制点集的线条 |傅立叶
    MovingAverage[
      |移动平均
        Differences[SouthAfConfirmedCasesValue], 5]] // Abs,
      |差分 |绝对值
    PlotLabel -> "移动平均步长Step=" <> ToString[5]]
  |绘图标签 |转换为字符串

```



在所选择的南非样本中，我们根据已知数据认为COVID-19在以非洲为代表的低医疗水平地区，每一次新增的峰值的周期在100天左右。

## 4\*. COVID-19对航班的影响

我们在国际民用航空组织上下载了2020年（缺少11、12月，最新的数据还未更新）美国的各月的航班人数。

In[ ]:= (\*导入文件为Dataset格式\*)

```
UsaAirlineData = Import["/Users/menakuniay/Desktop/tempairdata.csv",
  导入
```

```
"Dataset",
```

```
数据集
```

```
HeaderLines -> 1]
```

```
标头行
```

Out[ ]:=

Sum_PASSE	AIRLINE_ID	CARRIER_NAME	ORIGIN	ORIGIN_CITY_NAME
197			BIM	Bimini, The Bahamas
50			BIM	Bimini, The Bahamas
1501			CLD	Carlsbad, CA
18			EWB	New Bedford/Fall River, MA
165			EWR	Newark, NJ
35			EWR	Newark, NJ
545			EWR	Newark, NJ
0			FLL	Fort Lauderdale, FL
0			GCM	Grand Cayman, Cayman Islanc
231			GCM	Grand Cayman, Cayman Islanc
223			HPN	White Plains, NY
1600			LAS	Las Vegas, NV
80			MLB	Melbourne, FL
126			MLB	Melbourne, FL
181			PWM	Portland, ME
189			PWM	Portland, ME
179			SRQ	Sarasota/Bradenton, FL
94			SRQ	Sarasota/Bradenton, FL
177			TPA	Tampa, FL
21			VRB	Vero Beach, FL

(\*下面使用Table函数分别提取自2018年以来各月的航班总人数\*)

(\*使用Select函数对数据进行提取和合并\*)

```
SumPassengersUsa = Table[UsaAirlineData[
```

```
表格
```

```
Select[#[[16]] == year && #[[17]] == month &],
```

```
选择
```

```
#[[1]] &] // Total, {year, 2018, 2020}, {month, 1, 12}]
```

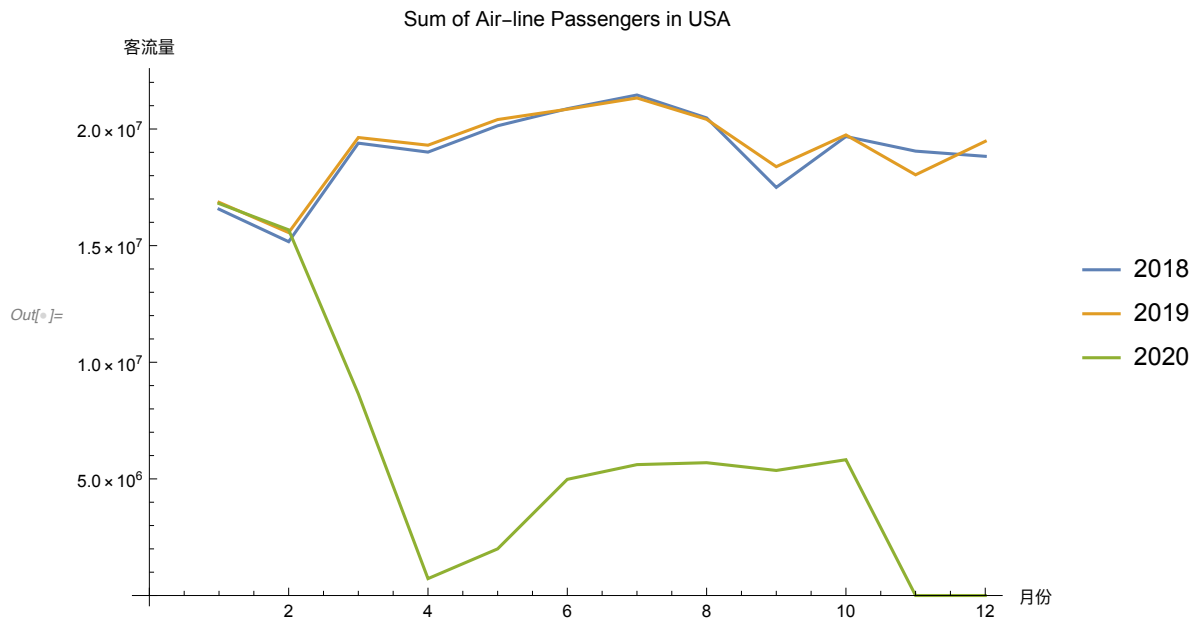
```
总计
```

```

Out[ ]:= {{16 563 412, 15 169 649, 19 393 935, 19 009 785, 20 138 742, 20 868 853,
           21 457 736, 20 476 253, 17 497 945, 19 678 863, 19 051 637, 18 829 148},
          {16 857 305, 15 560 395, 19 634 813, 19 305 834, 20 405 211, 20 847 922,
           21 330 406, 20 414 740, 18 385 639, 19 745 311, 18 036 903, 19 471 563},
          {16 804 782, 15 677 289, 8 645 905, 724 287, 2 000 367, 4 980 362,
           5 612 023, 5 694 449, 5 361 418, 5 822 046, 0, 0}}

In[ ]:= {SumPassengersUsa2018, SumPassengersUsa2019, SumPassengersUsa2020} =
SumPassengersUsa;
ListLinePlot[
绘制点集的线条
{SumPassengersUsa2018, SumPassengersUsa2019, SumPassengersUsa2020},
PlotLegends -> {"2018", "2019", "2020"},
绘图的图例
PlotLabel -> "Sum of Air-line Passengers in USA",
绘图标签 求和
AxesLabel -> {"月份", "客流量"}]
坐标轴标签

```



可以直观的看到，2018年和2019年的人流量基本持平，2020年1月和2月也和往年持平，但从3月开始一直下降，下降到4月达到最低值，仅为2019年同期的

```

In[ ]:= ToString[N[SumPassengersUsa2020[[4]] / SumPassengersUsa2019[[4]]] 10^3] <> "%"
转换为... 数值

```

Out[ ]:= 37.5165%

直到6月客流量开始逐渐回升，但是仍不及同期的一半，但是我们缺少2021年最新的数据，缺少对后疫情时代稳定的航班客流量的分析，留有一部分遗憾。

## 5. 小结

我们通过SIR传染病模型简单的复现了COVID-19爆发初期在爆发地的各项曲线，并且得到了较为准确的拟合图形，说明了疫情的发展仍可以用客观的数学模型所描述。然后通过研究以美



国、英国为代表的发达国家和国际人流量流通较大的国家近一年的疫情曲线，无法得到较为合理的周期性关系，我们认为是发达国家的医疗资源充足对一段时间的高爆发传染期有着抑制作用。然后Delta毒株爆发地印度的曲线过于极端，在2021年第一季度有着令人咂舌的爆发式增长、然而这一数据也不能客观的表示周期性规律。

医疗资源尚不发达、作为非洲国家的南非，是2021年末OmiCron毒株的爆发地，自2021年12月以来，其增长曲线也是一骑绝尘，侧面反映了本次Omicron的高传染性引起强烈国际注意的原因。并且南非曲线有着非常明显周期性，通过FFT分析认为每次大规模高传染性COVID-19变异株的爆发周期在100天左右，且传染时日增峰值逐渐增高。按此规律，那么2022年4月左右将会迎来新一个传染性较高的变异株，我们希望疫苗的发展可以抑制这种事件的发生，但我们也必须要有所准备。

最后本文简单的以美国航空总人流量的变化在体现COVID-19对人类社会经济打击之大，希望引起人们的重视。由于部分数据敏感和数据更新的不及时，本文有部分分析工作无法开展，在以后将会有更多数据供人类以分析，可以得出更科学更有信服力的规律助以COVID-19的研究之中。

---

## 参考文献

- Wolfram language reference <https://reference.wolfram.com/language/>
- COVID-19 Data Analysis and Visualization, Wolfram U <https://www.wolfram.com/wolfram-u/covid-data-analysis-and-visualization/>
- FFT in Mathematica Q&A <https://mathematica.stackexchange.com/questions/192193/which-fft-algorithm-does-fourier-use>
- COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University <https://github.com/CSSEGISandData/COVID-19>
- 2019冠状病毒病. (2021, December 12). Retrieved from 维基百科, 自由的百科全书: <https://zh.wikipedia.org/w/index.php?title=2019%E5%86%A0%E7%8A%B6%E7%97%85%E6%AF%92%E7%97%85&oldid=69066021>
- IMF World Economic Outlook Database <https://www.imf.org/en/home>