

# 2020MCM 养鱼选手的 Summary

Kevin

2020 年 2 月 18 日

# 目录

|          |   |          |
|----------|---|----------|
| <b>1</b> | <b>知识性收获</b>                              | <b>1</b> |
| 1.1      | L <sup>A</sup> T <sub>E</sub> X . . . . . | 1        |
| 1.1.1    | word 和 latex 的比较 . . . . .                | 1        |
| 1.2      | 时间序列分析 . . . . .                          | 1        |
| 1.2.1    | 时间序列基本知识 . . . . .                        | 1        |
| 1.2.2    | SPSS 批斗 . . . . .                         | 3        |
| 1.3      | 小波分析 . . . . .                            | 3        |
| 1.4      | matlab 工具箱以及文档 . . . . .                  | 3        |
| <b>2</b> | <b>体验性收获</b>                              | <b>3</b> |
| 2.1      | 赛题感受 . . . . .                            | 3        |
| 2.2      | 核心过程 . . . . .                            | 3        |
| 2.2.1    | 第一部分 . . . . .                            | 3        |
| 2.2.2    | 第二部分 . . . . .                            | 4        |
| 2.2.3    | 第三部分 . . . . .                            | 4        |
| 2.3      | 其他 . . . . .                              | 4        |
| <b>3</b> | <b>附录</b>                                 | <b>4</b> |

# 1 知识性收获

在美赛前期，以及比赛的四天当中，我被迫学习了各种各样，五花八门的神奇技能 (简称不学无术)。接下来介绍一下我近期的不学无术。

## 1.1 L<sup>A</sup>T<sub>E</sub>X

latex 是专门为了打美赛去学的，有一说一，真的不提倡这种速成的行为 (好像也不是第一次了)。原因有二：一是在短期学习过程中会遇到很多难以解释的 bug，如果没有写过代码是几乎没有办法解决的。另外我在学 latex 的过程中，发现用的人很多，资源也很多，但是大同小异，关键性的实际问题往往得不到解决；二是 latex 真的是个好东西，值得多花时间去学习并且练习。

### 1.1.1 word 和 latex 的比较

我在这里比较一下 word 和 latex 的使用体验。

**word 体验：**

- word 图形界面，使用方便，而且生态比较好，遇到问题能很快找到解决的方法。
- 图形接口众多，功能强大，但是不够专业的话，长篇的排版任务会带来很大压力。

**latex 体验：**

- 学术论文排版利器！szd！
- 因为是编程实现，结构非常清晰，甚至对行文思路有帮助。
- 对文献，图表有统一管理，很符合项目开发的习惯。
- 虽然直接生成 pdf，但是编译速度真的有点慢 (也可能是机子的问题)。
- 可能是刚刚上手，总感觉 latex 生态不大行。

至于这俩玩意儿哪个好，我也说不出来，但是 latex 我真的锁了。

## 1.2 时间序列分析

学这个东西真的一头雾水，所幸回家之前从图书馆 (@ 某财图书馆) 借了堆书，正好有本时间序列，上面正好有 SPSS 操作过程。既然提到 SPSS，后面必对这玩意开批斗会。

### 1.2.1 时间序列基本知识

- 面板数据，截面数据

这里不得不引用一下我 PPT 建模的两张图。

图 1 就是面板数据，多个数据单元的全部时序，像面板一样；图 2 就是截面数据，时间序列的截点数据。

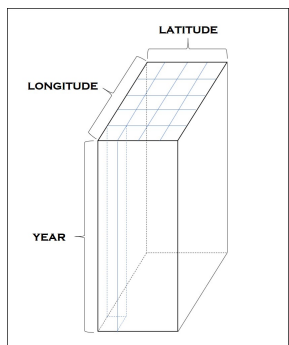


图 1: 面板数据

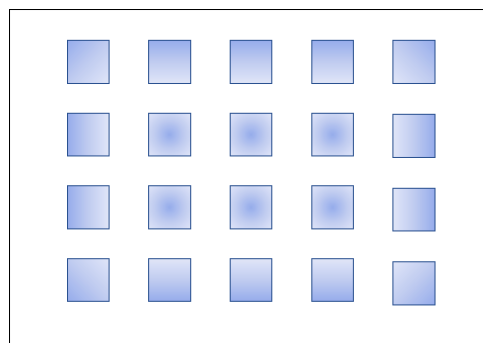


图 2: 截面数据

- 平稳性

判断时间序列是否平稳可以做 ADF 单位根检验，但是我觉得人眼观察法也挺不错的 (bushi。实际上就是看时间序列是否一直在同一水平上振动，发生倾斜就是不平稳。这里贴一张不平稳的，差分那里看平稳的。(这里贴的矢量图，随便放大，前面的 book)

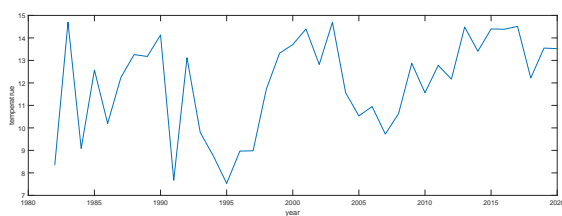


图 3: 非平稳时间序列

- 差分

差分其实就是前后差值的时间序列，如果原始序列有线性变化趋势，那么一阶差分就会平稳 (等差数列)。

对上面非平稳的序列做一阶差分，肉眼可见的平稳。

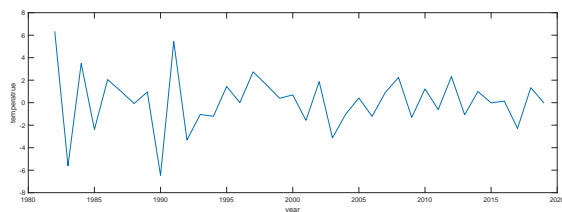


图 4: 一阶差分

- 自相关性，偏自相关性

这玩意我说不清楚，不能误人子弟。自相关大概就是跟自己相关，是数据的内部关联。

- AR MA ARIMA

这个更说不清楚了，大概就是回归模型，根据自相关和偏自相关的结果可以判定时间序列  $p$  和  $q$  阶数，然后选择模型  $AR(p)$   $MA(q)$   $ARIMA(p,d,q)$ ，这里的  $d$  就是差分阶数。

### 1.2.2 SPSS 批斗

这个鬼玩意儿，真的  $tu$  了。想想那天晚上，我和另一个队友一晚上各自做了 60 组  $ARIMA(0,1,0)$  预测。做 1 个：操作简便，输出美观；做 3 个：怎么这么多窗口；...；做  $n$  个：操... 操作真简便。这个 SPSS 的窗口真的有 bug，但是我用盗版，没处说理 (嚶.jpg)。

## 1.3 小波分析

小波分析是我这次学的最  $nb$  的东西，我真的能吹一年。简单说，小波是傅里叶的进阶版，对于信号  $S$  有  $S = s(A) + s(D)$ ，意思就是把一个复杂波拆开成若干波，这里  $A$  是小波近似系数， $D$  是小波细节系数。

然后我干了什么呢？为什么能吹一年呢？我们要做小波分析检测时间序列的异常点，然后我扒了一篇硕士论文，用  $matlab$  复现出来了... 就真的啥也不懂，硬做出来了... 后面会附上代码 (嚶.jpg，这个代码我能吹一年，其实也没什么，只是看起来很  $nb$ 。

## 1.4 matlab 工具箱以及文档

我觉得  $matlab$  不用多说，锁死来。工具箱真的很全面，在我用  $S**S$  做完 60 个预测之后，发现了  $matlab$  的金融分析的  $toolbox$  和  $app$ 。  $matlab$  只要有  $app$  就会有对应的  $toolbox$ ，相当于库。那意味着  $S**S$  手工预测白瞎了。

# 2 体验性收获

## 2.1 赛题感受

说实话，我看到题目之后真的没思路，一头雾水。后来队友说  $b$  站上有人直播讲解美赛思路，然后就去了，不得不说，人家的思路挺有逻辑的。最后参考他的大框架，用了他提到的小波异常检测，没想到做出来了。另外他提了一个克里金插值，是用来空间插值的，我把这个想法用在了构建监测点上，没想到能有后面数据结构化的思路。(有趣.jpg)

## 2.2 核心过程

没想到这次建模能做核心建模，不得不说我捏造模型的能力又提高了。简单说说我的整个思路。

### 2.2.1 第一部分

要求预测鱼群未来 50 年的生存位置。实质上是对鱼类最适生存水温的追踪，于是在苏格兰海域设置 20 个监测点，通过图像  $rgb$  值到温度逐步回归的映射得到温度变化的时间序列数据。用时间序列分析方法进行预测，并在监测点矩阵中追踪最适温度，最后得到迁移轨迹。

### 2.2.2 第二部分

要求求解最好，最好情况的持续时间。用小波异常分析得出全年时间序列的异常点，异常点间隔为最好情况持续时间，异常点密集连续区域为最坏情况持续时间。

### 2.2.3 第三部分

判断渔业公司是否改变现有经营模式。本来想构建渔业公司出海距离和持续时间的广义线性概率模型，但是没做出来，也不知道是否可行。然后怎么办呢？PPT 建模，over。

## 2.3 其他

- 全球变暖真的很严重！我算过了！一直变暖的话海水表皮温度可以到 20 摄氏度！（纯粹的线性增长预测）只想说好好保护环境。
- 写代码好起来了，所以要多写！
- 复现一篇论文让我有种可以直接毕业的错觉（？
- 我想看 SCI

## 3 附录

别问我为什么没有注释，latex 没整好，中文乱码（嚶.jpg

---

```
clear;
clc;

tempSeries = diff(load('D:\desk\matrixSequence\ (3,4).txt'));
outliers = [];
flag = 1;
count = 0;

Md = garch(1,1);
estMd = estimate(Md,tempSeries);
variance = infer(estMd,tempSeries);
Res = (tempSeries-estMd.Offset)./sqrt(variance);
res = Res;

while flag == 1
    Mean = mean(res);
    Std = std(res);
```

```

t = 0;
for i = 1:10000
    resDistribution = random( 'Normal',Mean,Std,length(res)*3,1);
    t = [t,max(abs(resDistribution))];
end
threshold = mean(t);

[A,D] = dwt(res, 'haar');

if find(abs(D)>threshold^0.05)
    out = max(find(abs(D)>threshold^0.05));
    D(out) = 0;
    outliers = [outliers, out];
else
    flag = 0;
end

res = idwt(A,D, 'haar');

count = count + 1;
disp(count);
end

outliers = 2*outliers;
OUT = []
for i = 1:length(outliers)
    s = outliers(i);
    r = Res;

    r(s) = 0; r(s+1) = 0;
    if abs(Res(s) - mean(r))>abs(Res(s-1)-mean(r))
        OUT = [OUT, s];
    else
        OUT = [OUT, s-1];
    end
end

```