

科研营学习总结报告

2019暑假遥感大数据分析项目总结

姓 名 : Yiran Huang
学 校 : 南京师范大学
专 业 : 自动化转计算机科学与技术
指导老师 : 周增光

目录

PART1 INTRODUCTION	2
1.1 BACKGROUND	2
1.2 DATA	3
1.3 PURPOSE.....	3
1.4 METHOD - Z-SCORES OF SEASON-TREND MODEL RESIDUALS (ZSTR)	3
1.5结果概述	4
1.6项目实施平台——R语言	4
第二章 方法描述	5
2.1总体思路	5
2.2具体过程	5
第三章 实验设计	8
第四章 程序设计	8
4.1区域选择及数据准备	8
4.2函数定义	8
4.3数据的批处理	9
4.4结果处理和分析.....	10
第五章 结果分析	13
5.1实验中间结果	13
5.2实验最终结果	14
5.3应用实际对结果合理性解释.....	15
第六章 主要问题和解决.....	15
第七章 科研心得	16
8.1团队心得	16
8.2个人心得	16

Part1 introduction

1.1 background

自然灾害（森林火灾、洪水和森林开伐）以及人类活动可能造成陆地覆盖的异常或干扰。在研究陆地变化的时空过程中做到连续侦测异常是十分重要且有效的。然而现如今，许多时间序列的分析方法（例如：基于扩展的Kalman Filter方法和Gaussian过程或者基于一些简谐模型）也许适合监测年际间的变化或者能够检测陆地覆盖变化以及监控陆地覆盖干扰和异常的发生。却有很少的方法关注在任何一个卫星图像时间序列之上进行连续监测变化异常去得到陆地覆盖的时空变化结果。

该研究提供了针对连续监测卫星图像时间序列异常的一个简单的方法——基于季节趋势模型残差的Z值——ZSTR。ZSTR是基于一个非常著名的对于陆地覆盖检测的方法——Breaks For Additive Season and Trend (BFAST)。两个方法均可以连续监测有洪水导致的区域时空异常。

本实验的结果主要展示了ZSTR对于卫星时间序列进行连续监测异常的能力和良好表现。

Natural disasters such as forest fires, floods, and deforestation, as well as human activities, can cause anomalies or disturbances in land cover. In studying the spatiotemporal processes of land changes, continuous detection of anomalies is crucial and effective. However, many time series analysis methods (such as extended Kalman filter methods and Gaussian processes or based on some harmonic models) may be suitable for monitoring interannual changes or detecting land cover changes, as well as monitoring land cover disturbances and anomalies. There are few methods that focus on continuously monitoring changes and anomalies in any satellite image time series to obtain spatiotemporal changes in land cover.

This study provides a simple method for continuously monitoring anomalies in satellite image time series based on the Z-value of seasonal trend model residuals, called ZSTR. ZSTR is based on a very well-known method for detecting land cover changes called Breaks For Additive Season and Trend (BFAST). Both methods can continuously monitor spatiotemporal anomalies in areas affected by floods.

The results of this experiment mainly demonstrate the ability and good performance of ZSTR to continuously monitor anomalies in satellite time series.

1.2 data

研究区域：黑龙江同江流域、6000平方千米（120km*50km）

时间区域：2000.2-2015.2

数 据：源于Terra/MODIS的NDVI图像、23张/年

异 常：2013年夏季洪水、2013.8.23河岸破裂导致大面积区域的洪水

1.3 purpose

通过设计ZSTR方法实现连续侦察异常动态

Designing the ZSTR method to achieve continuous surveillance of dynamic anomalies.

1.4 method - Z-scores of Season-Trend model Residuals (ZSTR)

步骤一：Decompose time series by a season-trend model

利用季节趋势模型分解时间序列

步骤二：Detect anomaly based on z-scores of model residuals

基于残差模型的Z值监测异常

1.5 结果概述

		空间上	时间上	整体上
ZSTR	清楚展现意外洪水区域的时空变化	反映意外洪水区域的整体区域	准确揭示洪水周期	利用卫星图像时间序列连续监测异常区域
BFAST	既不能反映意外洪水区域的时空变化也不能揭示实际洪水的时间周期	只能展示一半的实际异常洪水区域并且很多错误检测	监测结果比实际洪水周期早6到8个周期	监测在季节和趋势部分的突然变化或者破坏点

因此，BFAST可能适合检查在时间序列数据中是否有突然变化的存在，而ZSTR适合利用卫星图像时间序列进行持续监测时空异常区域。

Therefore, BFAST may be suitable for checking the presence of sudden changes in time series data, while ZSTR is suitable for continuously monitoring spatiotemporal anomalies in areas using satellite image time series.

1.6 项目实施平台——R语言

国外高校和研究机构广泛使用R语言开展数据分析和论文写作。R语言带来了数据分析领域新的革命——强大高效的数据分析能力和自由美观的可视化绘图效果，在数理统计、互联网、人工智能、金融、生物医学、天文地理等各个领域脱颖而出。R语言因其简洁易懂、免费开源，本着资源共享和共同提高的原则，截至 2018年12月其在线资源库已有远超10000个免费开源的R工具包以及无数的在线资源和使用人员。

R语言有以下五个主要特定——R语言软件、包、程序和文档等全部免费、开源；有强大的数据分析和团队支撑；R包的使用和管理相当便捷；可直接生成报告文档；拓展开发能力强，便于开展重复性实验、创新性研究。

第二章 方法描述

2.1 总体思路

步骤一：Decompose time series by a season-trend model

利用季节趋势模型分解时间序列

步骤二：Detect anomaly based on z-scores of model residuals

基于残差模型的Z值监测异常

2.2 具体过程

步骤一：利用季节趋势模型分解时间序列

季节性时间序列 y_t 可以通过季节趋势模型分解为趋势、季节和残差分量：

$$y_t = T_t + S_t + r_t, \quad t = 1, 2, \dots, n$$

其中 y_t 是时间序列 t 的观测值， T_t ， S_t 和 r_t 分别是时间序列数据的趋势、季节和残差分量。

遥感时间序列常用的加性季节趋势模型是由线性趋势和谐波季节组成的模型。

$$y_t = a_1 + a_2 t + \sum_{k=1}^K \gamma_k \sin\left(\frac{2\pi k t}{T} + \delta_k\right) + r_t, \quad t = 1, 2, \dots, n$$

其中 y_t 是观测值(即时间序列的响应)， α 和 β 分别是线性趋势分量的截距和斜率，

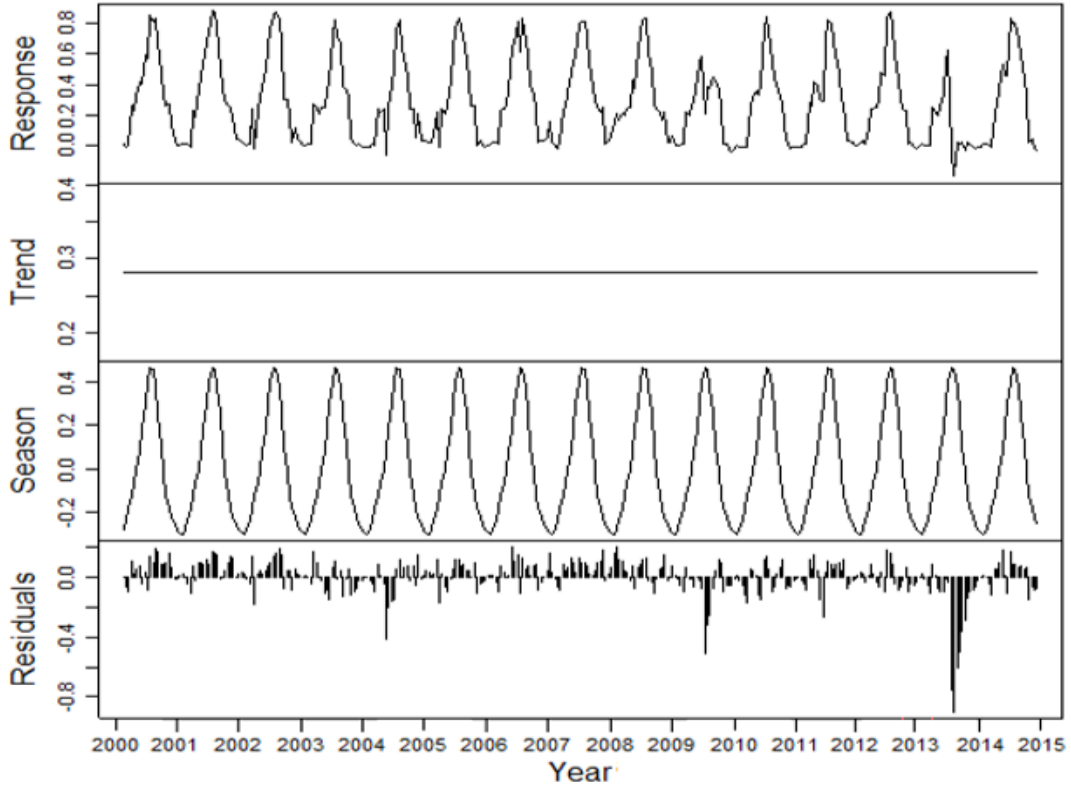
γ_k ， δ_k 和 K 分别是谐波季节分量的振幅、相位和最高阶（一个预置的值），

T 是周期或时间序列的周期（此文献中一年周期中有23个观测值，则 $T=23$ ）

上式中给出的季节趋势模型可转换为标准线性回归模型：

$$\begin{aligned} y_t &= \mathbf{x}_t^T \boldsymbol{\beta} + r_t \\ \mathbf{x}_t &= [1, t, \sin(2\pi 1 t / T), \cos(2\pi 1 t / T), \dots, \sin(2\pi K t / T), \cos(2\pi K t / T)]^T \\ \boldsymbol{\beta} &= [a_1, a_2, \gamma_1 \cos(\delta_1), \gamma_1 \sin(\delta_1), \dots, \gamma_K \cos(\delta_K), \gamma_K \sin(\delta_K)]^T \end{aligned}$$

其中 β 是一个未知参数的向量，可以用普通最小二乘法（OLS）估计。



通过Fig. 1的时间序列分解，参数 K 和 T 分别为3和23。

步骤二：基于残差模型的Z值监测异常

时间序列中的异常可以通过季节趋势模型在时间序列分解后转移到残差分量（Fig. 1）。从上述标准线性回归模型中，残差分量可直接从标准线性回归模型中提取：

$$r_t = y_t - \mathbf{x}_t^T \boldsymbol{\beta}$$

假设遵循 r_t 具有平均值 μ 和标准差 σ 的正态分布，即 $r_t \sim N(\mu, \sigma^2)$ ，则残差时间序列可通过以下方式标准化：

$$z_t = \frac{r_t - \hat{\mu}}{\hat{\sigma}}$$

其中 $\hat{\mu}$ 和 $\hat{\sigma}$ 分别是 μ 和 σ 的估计值。由于 r_t 可能包含异常数据，这些数据将带来 μ 和 σ 的常规估计的误差，这里我们使用它们的一些稳健估计。 $\hat{\mu}$ 等于删除5%最高和5%最低数据后的平均值。 $\hat{\sigma}$ 由下式计算：

$$\hat{\sigma} = \frac{\sum |r_t| \cdot \sqrt{\pi/2}}{n}$$

然后 z_t 为标准化 r_t 之后的结果，将遵循标准正态分布，在本文中，我们称之为模型残差 r_t 的 z -scores。

残差时间序列数据中的异常可以根据其 z -score values的分布来确定，即 z -score values到零平均值越远，相应残差越可能是异常。因此，可以通过以下规则检测异常：

$$z_t \text{ is } \begin{cases} \text{positive anomaly,} & \text{if } z_t > z_{[\alpha/2]} \\ \text{negative anomaly,} & \text{if } z_t < -z_{[\alpha/2]} \end{cases}$$

其中， $z_{[\alpha/2]}$ 和 $-z_{[\alpha/2]}$ 分别是 z -分数的上边界和下边界。 α 为统计检验的显著性水平（如0.05）

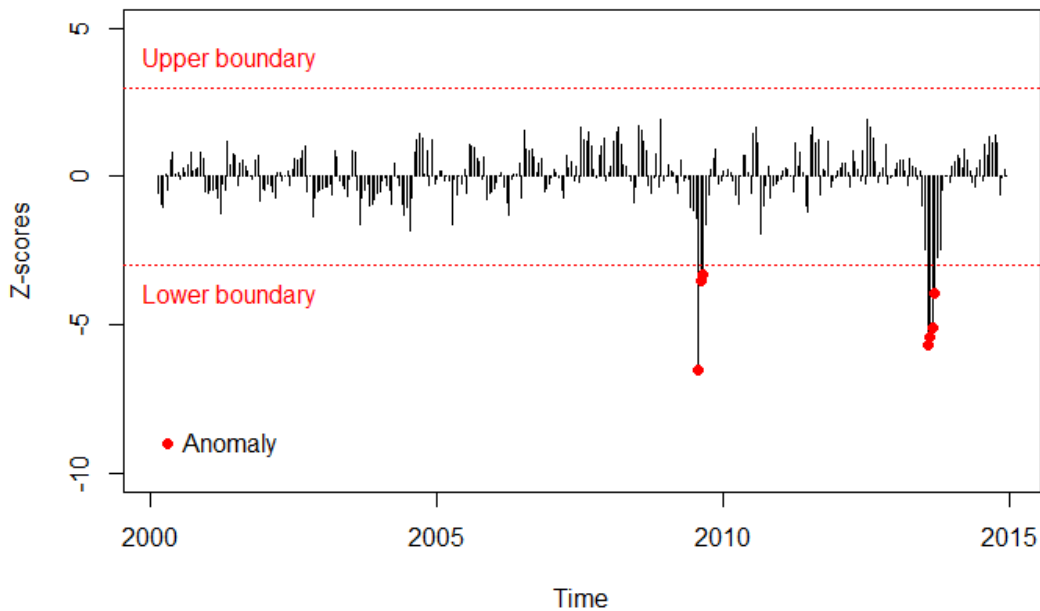


Fig. 2 基于季节趋势模型残差值 Z 值的持续异常监测，边界值为 ± 3 ($\alpha=0.0027$)

第三章 实验设计

实验原材料的选择->

设计ZSTR ->

ZSTR和BFAST对照数据处理->

合成最终的效果图->

对比结果进行分析->

得出结论

第四章 程序设计

4.1 区域选择及数据准备

```
data<-brick("C:/keyanying/Data/layerstack/  
NDVI_layerstack_2000-2014.tif")#导入图片数据  
data1<-crop(data, extent(data, 70, 119, 250, 349))#裁剪选定区域  
data2<-getValues(data1) #转成矩阵
```

4.2 函数定义

#ZSTR异常检测函数

```
fun_ZSTR_detection<-function(x)  
{  
  if(anyNA(x)) return(c(1:342)*NA)  
  else{  
    x<-(x-mean(x))/sd(x)  
    data6<-ts(x, start=c(2000, 4), frequency = 23)  
    model1<-lm(data6~time(data6))  
    har.<-harmonic(data6, 3)  
    model2<-lm(ts(residuals(model1), frequency = 23, start =  
c(2000, 4))~har.)
```

```

    data7<-residuals(model2)
    index<-(data7>quantile(data7, 0.05)&data7<quantile(data7, 0.95))
    data71<-data7[index]
    sd<-1/length(data71)*sum(abs(data71))*sqrt(pi/2)
    z<-(data7-mean(data71))/sd
    zt<-ts(z, frequency = 23, start = c(2000, 4))
    anomalies<-(zt < -3)
    anomalies[anomalies == FALSE]<-NA
    return(anomalies)
  }
}

#ZSTR异常开始时刻函数
fun_ZSTR_breakpoints<-function(x) {
  return(which(x == TRUE)[1])
}

#bfast异常检测函数&异常开始时刻记录
fun_bfast_detection<-function(x) {
  if(anyNA(x)) return(NA)
  else{
    ts_data<-ts(x, frequency = 23, start = c(2000, 4))
    bfast_md<-bfast01(ts_data, formula = response ~ harmon, bandwidth =
0.05)
    ifelse(bfast_md$breaks, return(bfast_md$breakpoints), return(NA))
  }
}

```

4.3 数据的批处理

```

#ZSTR异常检测批处理
data2_ZSTR_rslt<-apply(data2, 1, fun_ZSTR_detection)
#bfast异常检测批处理
data2_bfast_rslt<-apply(data2, 1, fun_bfast_detection)

```

4. 4结果处理和分析

(1) ZSTR方法小样本检验

```

data1<-as.array(data)
data2<-ts(data1[79, 275, ], start= (c(2000, 4)), frequency=23)
data2<-(data2-mean(data2))/sd(data2)
data6<-ts(data2, start=c(2000, 4), frequency=23)
model1<-lm(data6~time(data6))
har.<-harmonic(data6, 3)
model2<-lm(ts(residuals(model1), frequency=23, start= c(2000, 4))~har.)
data7<-residuals(model2)
index<-(data7>quantile(data7, 0.05)&data7<quantile(data7, 0.95))
data71<-data7[index]
sd<-1/length(data71)*sum(abs(data71))*sqrt(pi/2)
z<-(data71-mean(data71))/sd
zt<-ts(z, frequency=23, start=c(2000, 4))
data2<-ts(data1[79, 275, ], start=(c(2000, 4)), frequency=23)
plot(data2, type= 'h' ) #Fig. 1. a
plot(ts(fitted(model1), start=c(2000, 4), frequency=23)) #Fig. 1. b
plot(ts(fitted(model2), frequency=23, start=
c(2000, 4), ylab= "Season" )) #Fig. 1. c
plot(ts(residuals(model2), start=c(2000, 4), frequency=23),
type = "h", ylab= "Residuals" ) #Fig. 1. d

#Fig. 2
plot(ts(zt, start=c(2000, 4), frequency = 23), type= "h", ylim =
range(-10, 5), ylab = "Z-scores" ) #ylim y轴坐标范围
a<-zt<-3 #a为异常点
b<-ts(zt, start=c(2000, 4), frequency=23)
abline(h = 3, lty = 3, col = "red" ) #lty=3 虚线
abline(h = -3, lty = 3, col = "red")
text(time(b)[30], 4, labels = "Upper boundary", col = "red" )
#time取b的第三十个横坐标
text(time(b)[30], -4, labels = "Lower boundary", col = "red")
points(time(b)[a], b[a], col = "red", pch = 16)

```

#pch=16 实心圆点

```
points(time(b)[5], -9, col = "red", pch = 16)
text(time(b)[30], -9, labels = "Anomaly")
```

(2) 结果处理和分析绘图---bfast

#bfast检测结果处理

```
data2_bfast_rslt1<-t(data2_bfast_rslt)  #? ?
```

```
#data2_bfast_rslt1<-as.matrix(data2_bfast_rslt)
```

#转回raster文件，赋予地理信息，投影信息

```
b f a s t _ b r e a k p o i n t s < -
raster(matrix(data2_bfast_rslt1, nrow=50, ncol=100, byrow= TRUE))
extent(bfast_breakpoints)<-extent(data1)
projection(bfast_breakpoints)<-projection(data1)
```

#显示bfast处理后302~308波段异常检测结果

“Fig. 3. bfast. q-t”

```
for (i in c(302:308))
{
  plot(bfast_breakpoints, col = gray.colors(256))
  bbp_temp<-bfast_breakpoints
  bbp_temp[bfast_breakpoints!=i]<-NA          #绘制302图
  plot(bbp_temp, col='red', add=TRUE, legend = FALSE)
  mtext('E', at=134.24, line=0.95, side=1)
  mtext('N', at=48.31, line=0.95, side=2)
}
```

#Fig4. b bfast图像绘制

```
bfast_breakpoints<-bfast_breakpoints - 296
```

```
bfast_breakpoints[bfast_breakpoints < 6 | bfast_breakpoints > 20]<-NA
```

```
bfast_breakpoints[c(1, 2)]<-c(6, 20)          #为什么是6~20
```

#显示NDVI背景图

```
plot(data1[[312]], col = colorRampPalette(c("black", "white"))
(256), legend = FALSE)
```

#叠加

```
plot(bfast_breakpoints, col = rainbow(15), add = TRUE)
mtext('E', at=134.28, line=0.95, side=1)
```

```

mtext('N', at=48.31, line=0.95, side=2)
(3) 结果处理和分析绘图---ZSTR
#ZSTR检测结果处理(2013全年)
data2_ZSTR_rslt1<-t(data2_ZSTR_rslt)
data2_ZSTR_rslt1_2013<-data2_ZSTR_rslt1[, 297:319]
#ZSTR异常开始时间批处理计算
data2_ZSTR_rslt1_2013_breakpoints<-
  apply(data2_ZSTR_rslt1_2013, 1, fun_ZSTR_breakpoints)
#转回raster文件, 赋予地理信息, 投影信息
ZSTR_breakpoints<-
  raster(matrix(data2_ZSTR_rslt1_2013_breakpoints, nrow=50, ncol=100, byrow = TRUE))
extent(ZSTR_breakpoints)<-extent(data1)
projection(ZSTR_breakpoints)<-projection(data1)

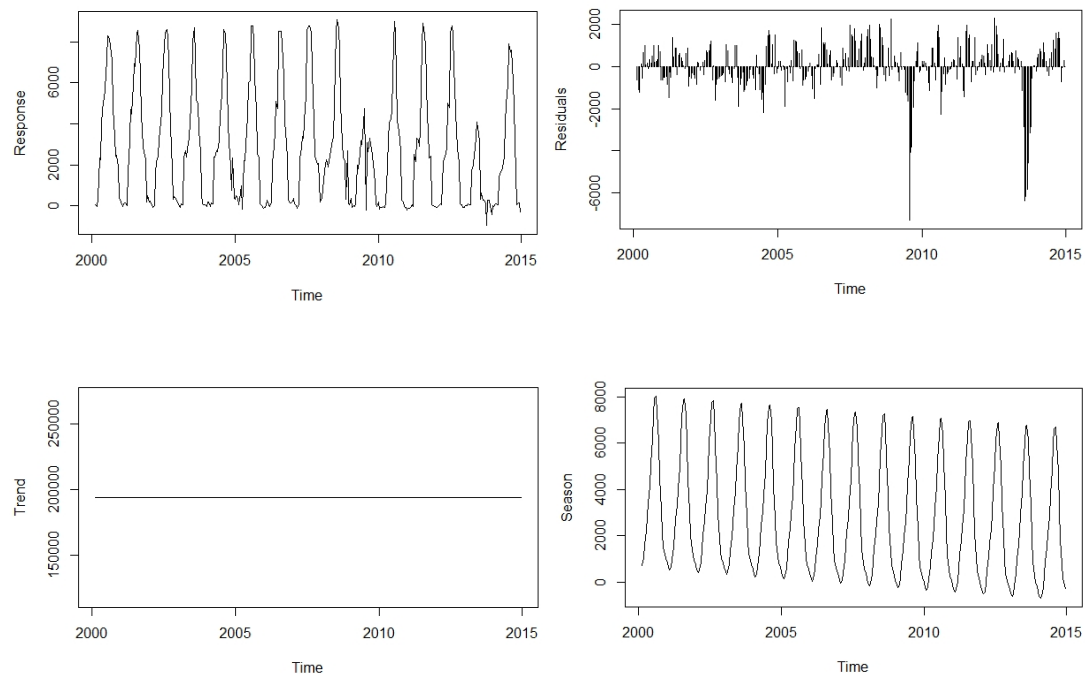
#Fig4. a ZSTR图像绘制
ZSTR_breakpoints[ZSTR_breakpoints < 6 | ZSTR_breakpoints > 20]<-NA
ZSTR_breakpoints[c(1, 2)]<-c(6, 20) #为什么是6~20
#显示NDVI背景图
plot(data1[[312]], col = colorRampPalette(c("black", "white"))
      (256), legend = FALSE)
#叠加
plot(ZSTR_breakpoints, col = rainbow(15), add = TRUE)
mtext('E', at=134.28, line=0.95, side=1)
mtext('N', at=48.31, line=0.95, side=2)

```

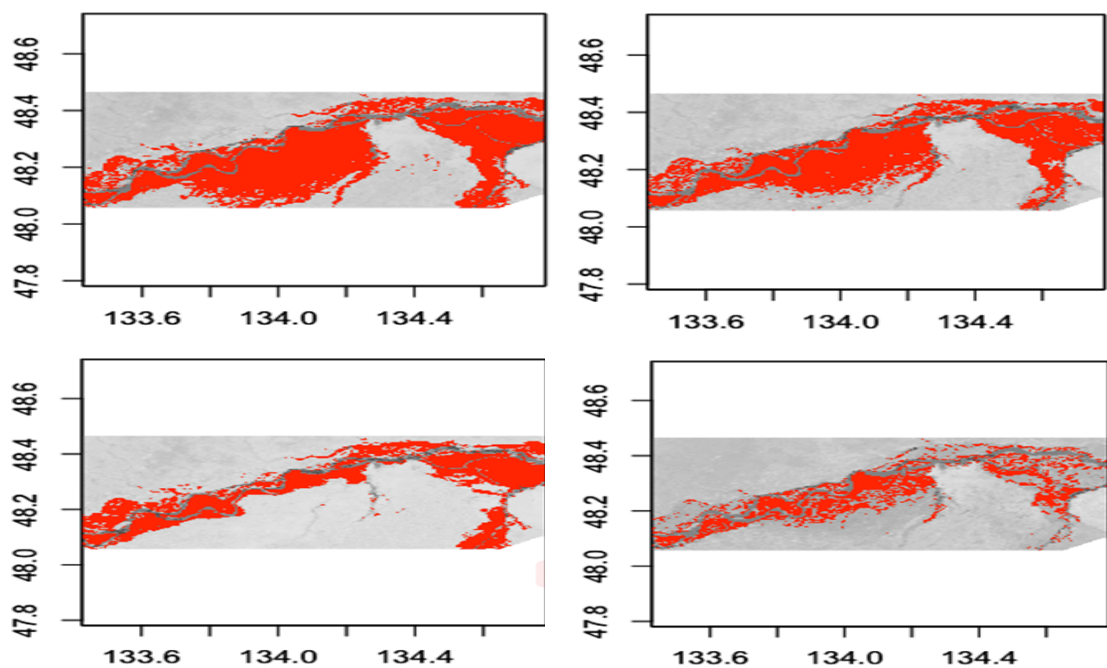
第五章 结果分析

5.1 实验中间结果

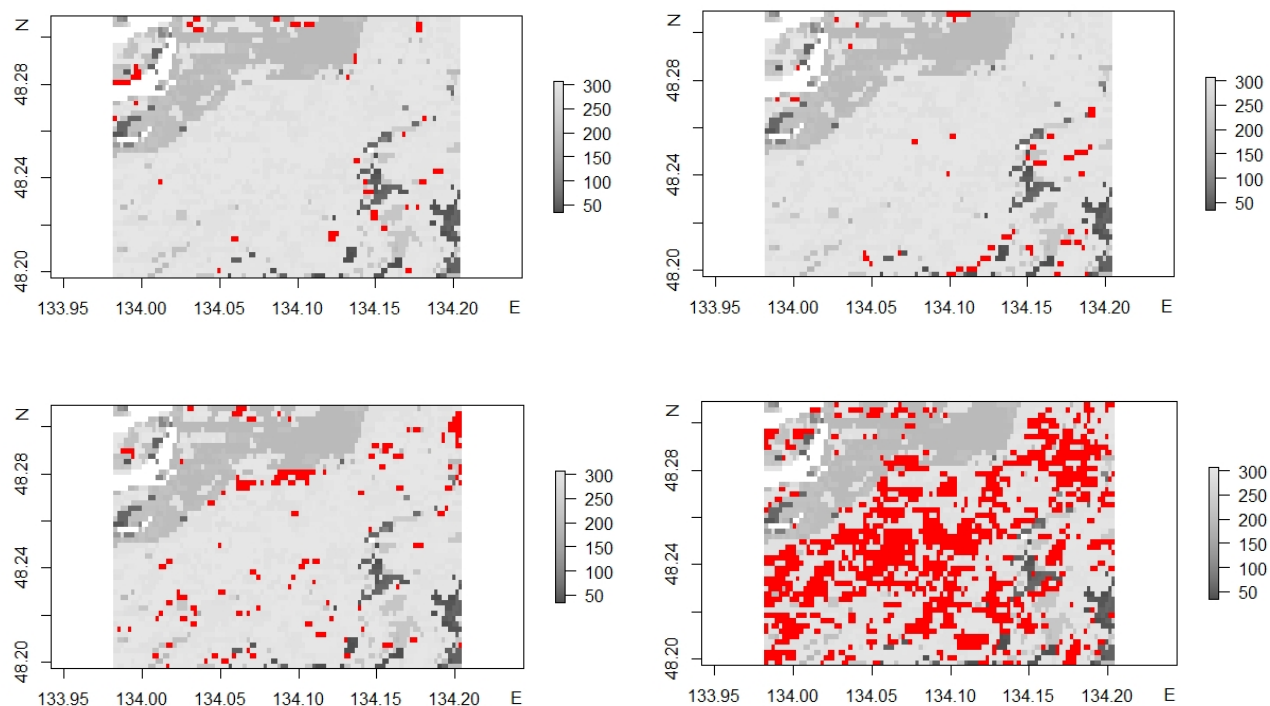
季节时间序列yt可以分解成趋势、季节、季节趋势、和残差分量的模型



生成Zstr方法检测到的异常区域:

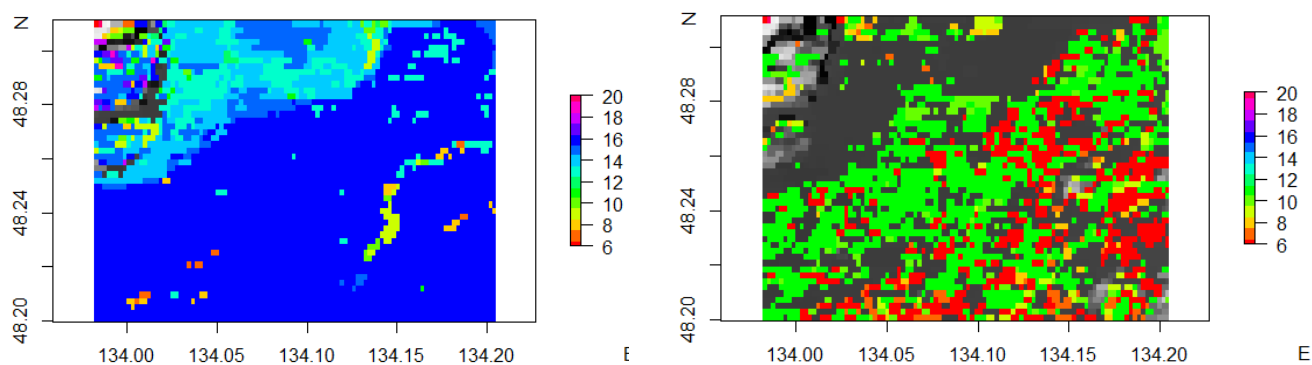


生成Bfast方法检测到的异常区域:



5.2实验最终结果

两种方法检测到异常区域的时空综合信息图（局部） 左图ZSTR 右图BFAST



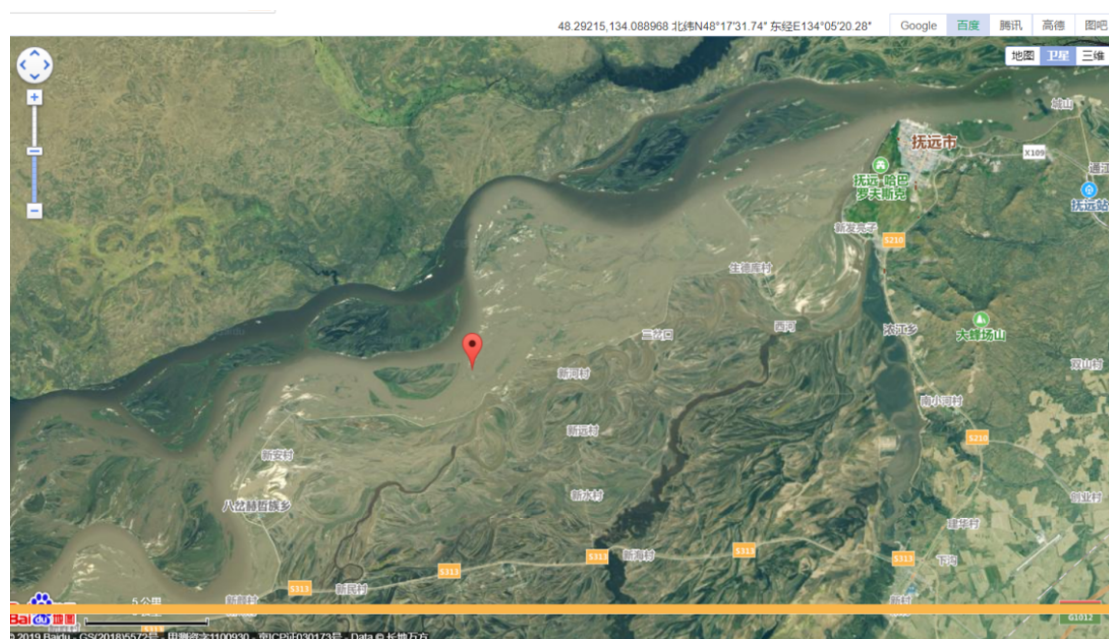
对比两种方法检测异常点以及异常开始时间所生成的模拟图后
ZSTR不仅模拟的更加准确并且错误的信息更加少

5.3 应用实际对结果合理性解释

通过生成模拟图中的经纬度信息检索到经纬度对应的卫星地图

如图中检索（48.27北纬, 134.02东经）周边为没有受到洪水侵扰区域

实际应用中可以通过ZSTR方法，将更加准确定位洪水侵扰区域，对受灾区域进行提前预警



第六章 主要问题和解决

（1）函数问题——对数据进行线性拟合的理解。

我们组最初整体理解都有所偏差，导致在第五天做作业的时候继续延续第四天犯下的错误以至于将近三点才完成。但是在作业的讨论、其他小组的帮助以及老师的讲授下，我们对原有的函数进行修改，最终获得正确结果（分开拟合、标准化）

（2）矩阵问题——对矩阵转置的疏忽

在编程过程中反复修改，反复试错，得出错误点并改出，然后组员共同进行理解分析，克服难关。可以说是先改正错误再研究原因的过程。

（3）细节问题

许许多多的小错误，一次次改正——对R语言从了解到熟练的过程。

第七章 科研心得

8.1 团队心得

进入科研营之前听说是小组合作完成任务，说实话我有点茫然，但更多是害怕与慌张。茫然因为从未接触过真正的分工式小组合作，可以说平时即使小组合作也是做一些简单的事情而非像这次科研要完成一个不小的项目，慌张始于自己的不自信，不自信自己是否能够在这个团队负责好自己的责任、尽好整个小组的责任。

第一天的任务相对比较简单，在没有经过小组成员的讨论下我成功地将代码实现，这也是我第一次对自己燃起了信心和自信。我开始敢于自己去主动写、主动思考。然而，第一天的我们并没有很深入地一起讨论，可以说讨论的时候人都没有来齐，而且仅简简单单进行了五分钟就匆匆结束。第一天的第一组没有分工，没有更多的交流。

随着后面的任务及作业越来越难、任务量越来越大，每个小组成员在毫无交流沟通的前提下都把作业完成一遍已然是不太可能的了。我们开始有了对每一题甚至是每一个步骤的分工，以及后期的文献阅读分工、ppt制作分工、图像收集分工，逐渐地细化了起来，也逐渐成熟了起来，同时在合作当中不断提出问题对于整个工作的进行也起着很大的作用和帮助。小组同学之间思想的碰撞就能在不经意间解决很多意象不到的问题。我自己有很多想不通的东西在与我的队友激烈的讨论中得到了解决，在解决组员疑问的时候自己对知识有了更深的理解，同时也增强了自己的自信心。这样下来，我对于代码的理解和编写也在这个过程中得到极大的提升。所以，团队合作在任务完成中是必不可少的，在这里我要感谢我的队友们，感谢大家的帮助。

在整个科研营的过程中，我们组员间对彼此能力的认知是随时间不断增进的，所以最初因不熟悉、配合度不高而造成的低效问题，也随着时间的发展，因彼此之间默契的上升而得到了解决。

8.2 个人心得

其一在我印象中最深刻是要敢说敢做，从自己脑子里面迸发出来的点子不要怕说，更不要怕说错，说对了会让整个陷入瓶颈的团队柳暗花明、明确道路，即使说错了但也会给团队提供另一种不同的想法，同时也帮助自己纠正的错误。

其二在自己的深入思考之后切不要闭门造车，要多寻求外界帮助。一是体

现在编程的过程中：R语言拥有庞大的函数集，对于我来说并不可能掌握所有函数运用，即使是老师有时候也会遇见不认识或者不熟悉用法的函数。所以在完成一项任务编程的时候要多用R语言为我们提供的帮助模块，同时积极上网寻求更多资源。二是体现在小组之间：有时候一个小组想了很久都解决不了某个问题的时候也不要闭门造车，组与组之间也要保持沟通交流，每个人甚至每个组的切入点、想法都不一样，所以要多进行思维碰撞，多寻求外界帮助。

其三是分析数据首先要想好整体分析的框架，有一定的思路再开始写程序，而不是边写边改。掌握多种数据分析的方法，遇到问题多思考。

其四是阅读论文的需求和能力：多读论文、多研究论文不仅会帮助我们提升英语能力，更多的是传授我们分析问题解决问题以及对数据处理的一个整体的方法与框架，以及开阔我们的视野，让我们了解到更多的关于数据分析处理的新方法、新思路。

其五是学习的深刻度：老师曾问我们一个问题`anomalies[anomalies == FALSE]<-NA`是什么意思。很多同学都答不上来。其实这个就是R语言其中矩阵用法的一个基础。由于是老师给我们的代码，我们大多数同学都没有深入了解，而只是从表面上大致知道什么意思。其实想要理解其中的意义我们可以简单举一个矩阵的例子就可以清楚明白了，只不过我们大多数同学还没有具备深入思考发现问题的能力。

最后，感谢老师的指导、组员的合作以及其他小组对我们的帮助。老师教会我在短时间内进行小组分工完成项目报告上台展示，其精髓为先追求完成，再追求完美；老师也教会我无论经过小组合作完成什么项目都要记得明确小组分工以及感谢那些在项目进行过程中那些帮助过我们的人；以及积极参与课堂讨论，积极和学生进行互动，永远记住课堂是学生的课堂。