

# 海冰面积与全球温度的建模与研究

Modeling and Reasearch of Sea Ice Extend and Global Temprature

涂博海PB16150212

李传赞PB16080964

朱永鑫PB16150773

2019 年 7 月 8 日

# 目录

<b>1 课题介绍</b>	<b>3</b>
1.1 问题介绍 . . . . .	3
1.2 海冰介绍及相关研究介绍 . . . . .	3
1.3 数据简介 . . . . .	6
1.4 章节安排与任务分工 . . . . .	6
1.5 源代码 . . . . .	6
<b>2 数据预处理</b>	<b>6</b>
<b>3 北极海冰面积的建模与预测</b>	<b>8</b>
3.1 季节性 ARIMA 模型 . . . . .	8
3.1.1 模型构建 . . . . .	8
3.1.2 模型的预测 . . . . .	10
3.2 时间序列基本分解模型 . . . . .	10
3.2.1 趋势项 $T$ 的估计 . . . . .	10
3.2.2 周期项的估计 . . . . .	12
3.2.3 随机项的估计 . . . . .	12
3.2.4 模型的预测 . . . . .	18
3.3 LOESS 分解模型 . . . . .	18
3.3.1 趋势项 $T$ 的估计 . . . . .	20
3.3.2 周期项的估计 . . . . .	20
3.3.3 随机项的估计 . . . . .	20
3.3.4 模型的预测 . . . . .	25
3.4 预测结果对比与模型评论 . . . . .	25
<b>4 南极海冰面积的建模与预测</b>	<b>26</b>
4.1 时间序列的基本分解模型 . . . . .	26
4.1.1 趋势项 $T$ 的估计 . . . . .	26
4.1.2 周期项的估计 . . . . .	28

4.1.3 随机项的拟合 . . . . .	28
4.1.4 模型的预测 . . . . .	29
4.2 LOESS 分解模型 . . . . .	29
4.2.1 趋势项 $T$ 的估计 . . . . .	31
4.2.2 周期项的估计 . . . . .	32
4.2.3 随机项的拟合 . . . . .	32
4.2.4 模型的预测 . . . . .	34
4.3 预测结果对比与模型评论 . . . . .	34
<b>5 全球平均温度的建模与预测</b>	<b>35</b>
5.1 时间序列基本分解模型 . . . . .	35
5.1.1 趋势项 $T$ 的估计 . . . . .	35
5.1.2 周期项的估计 . . . . .	35
5.1.3 随机项的拟合 . . . . .	36
5.1.4 模型的预测 . . . . .	38
5.2 LOESS 分解模型 . . . . .	38
5.2.1 趋势项 $T$ 的估计 . . . . .	39
5.2.2 周期项的估计 . . . . .	40
5.2.3 随机项的拟合 . . . . .	40
5.2.4 模型的预测 . . . . .	42
5.3 预测结果对比与模型评论 . . . . .	42
<b>6 多元数据的 VAR 模型构建与预测</b>	<b>44</b>
6.1 模型滞后阶数的确定 . . . . .	44
6.2 拟合模型 . . . . .	44
6.3 模型的诊断性检验 . . . . .	45
6.4 脉冲响应分析 . . . . .	47
6.4.1 来自全球平均温度的脉冲反应 . . . . .	47
6.4.2 来自南极海冰的脉冲反应 . . . . .	47
6.4.3 来自北极海冰的脉冲反应 . . . . .	47

6.4.4 脉冲响应结果分析 . . . . .	49
6.5 预测与评估 . . . . .	49
<b>7 总结</b>	<b>50</b>

**摘要:** 针对全球变暖与南北极海冰面积的相互影响, 本文提出多种模型来对全球温度和海冰面积进行预测并尝试找出其因果关系。首先, 基于 Berkley Earth 的全球温度数据和南北极海冰数据, 本文通过对原始变量做回归和分解对缺失数据进行填补。其次, 对每组数据进行独立建模, 使用了 ARIMA 模型和 GARCH 模型, 得到较好的预测效果。最后, 考虑到温度与海冰面积的交互关系, 使用 VAR 模型建模, 找出了温度与南北极海冰的因果关系, 验证了南北极海冰与温度的不平等关系, 并得到较好的预测效果。

**关键词:** 全球温度; 海冰面积; LOESS; ARIMA; GARCH; VAR; 预测

**Abstract:** Considering the interaction between global warming and sea ice area in the Antarctic and the Arctic, some models are proposed in this paper to predict the global temperature and sea ice area and try to find out the causal relationship. First, based on global temperature data and sea ice data of the Antarctic and Arctic from Berkley Earth, we fills in the missing data by linear regression and decomposition of the original variables. Secondly, each group of data is independently modeled, using ARMA model and GARCH model, and the prediction is good. Finally, considering the interaction between temperature and sea ice area, we use VAR model to find the causal relationship between temperature and sea ice area in the Antarctic and Arctic, verifying the unequal relationship between sea ice and temperature in the Antarctic and Arctic, and get better prediction results.

**Key words:** Global temperature; Sea ice; LOESS; ARIMA; GARCH; VAR; Forecast

# 1 课题介绍

## 1.1 问题介绍

全球变暖一直以来都是倍受关注的问题。它是由于温室效应不断累积，导致地气系统吸收与发射的能量不平衡，能量不断累积，从而导致温度上升，造成全球变暖。它会产生极端天气，海平面上升等恶劣影响。其中，全球变暖导致的冰盖萎缩，冰川融化，海冰面积减小是导致海平面上升和一些极端天气的主要因素。

我们将以海冰面积和全球平均温度作为指标，研究全球变暖的情况。我们将分别建立南北极海冰面积和全球平均温度的模型，以预测未来一年的情况，再联合二者建立二维时间序列模型，进行因果推断和综合预测，给出适当建议。

## 1.2 海冰介绍及相关研究介绍

海冰是冰冻的海水，它在海洋中形成，生长和融化。海冰仅在北极和南极存在。在北半球，它目前可以存在于中国渤海湾以南<sup>1</sup>，在南半球，海冰仅在南极洲附近<sup>2</sup>形成。海冰在冬季增长，在夏季部分融化，在一年中部分时间，世界上约 15% 的海洋被海冰覆盖。

虽然海冰主要发生在极地地区，但是它会对全球气候产生很大的影响<sup>3</sup>。这个影响主要来自四个方面：

- **海冰的融化。**太阳光在赤道近乎垂直照射，而在极地地区则格外倾斜。太阳的角度是造成极地寒冷和赤道温暖的主要原因。海冰是白色的，所以它有很高的反照率，能反射约 80% 的太阳光。高反照率有助于保持极地的寒冷因为，海冰覆盖的区域不会吸收太多太阳能。但当气候变暖使得极地变暖并融化海冰时，极地区域的放射面减小，就会吸收更多热量，导致更多冰的融化，从而加速升温。这是一个正反馈回路，会改变大气的循环。
- **大气与海洋环流。**大气和海洋是“热力发动机”，总是试图通过向极地输送热量来恢复温度平衡。我们的天气正是这种现象的表现。一些低压系统，比如风暴，在冬季特别强烈，是大自然通过大气环流向极地输送热量的最佳方式之一。相比之下，海洋倾向于以较慢且不太强烈的方式传递热量。海冰量的变化改变了极地的寒冷程度，这会影响大气和海洋的循环。另一方面，洋流通过温盐环流将热量从赤道传递到极地。温水从赤道沿着海面向北移动并最终冷却。随着它的冷却，它变得密集而沉重并且下沉。然后，这种冷水沿着海洋的下部向南移动，并在赤道附近上升以完成循环。与大气热传输一样，这是一个有助于在地球上实现适当温度平衡的自然过程。这也解释了为什么欧洲相对温暖，因为大西洋北部流动的地表水变冷，热量释放到大气中。
- **海冰的移动。**海水通常是咸的，但凝结成冰时，会将盐分排出到冰下，这就使海冰中的水含盐量低，可以饮用。受北极风的影响，北极海冰会向南穿过弗拉姆海峡进入北大西洋时，它会融化，在海面上形成一层淡水。这种淡水的密度低于咸水，所以它往往会留在海洋的顶部。这种较低的密度阻碍了在支持温盐环流的高纬度（极地）下沉的正常过程，这使得温水从赤道向北移动变得更加困难。有力的证

---

<sup>1</sup>大约北纬 38 度。

<sup>2</sup>在南纬 55 度以南。

<sup>3</sup>这个影响主要来自北极海冰，下面也主要以北极海冰为例，后面会具体解释原因。

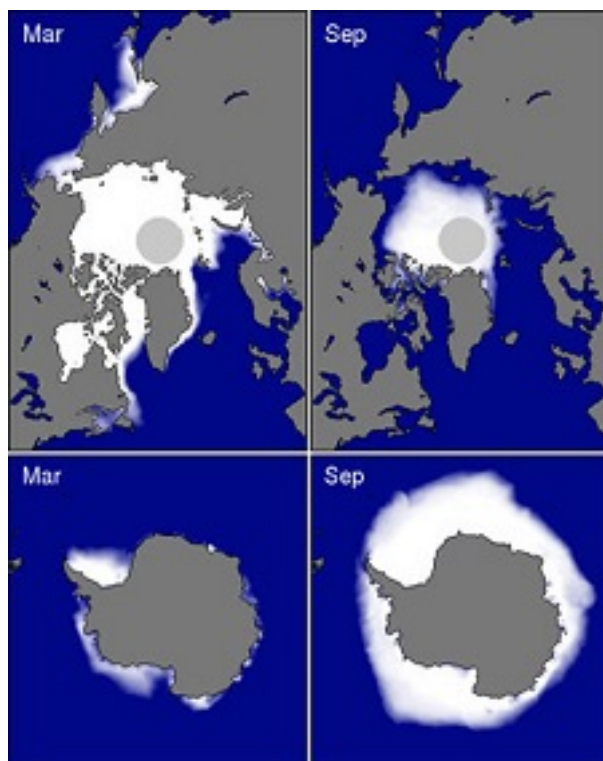


图 1: 南北极海冰地理位置对比图<sup>5</sup>, 其中上部分是北极, 下部分是南极

据表明, 这种停滞过程发生在 20 世纪 60 年代末和 70 年代初期的几年, 当时额外的淡水进入北大西洋并影响了北欧的气候。

- **换热。**在冬季, 北极的气候非常寒冷。相比之下, 海洋更加温暖。海冰覆盖了两极, 防止海洋中的热量加热上覆的大气层。这种隔热效果是海冰有助于保持北极寒冷的另一种方式。但是, 热量可以从薄冰区域, 特别是水道<sup>4</sup> 和冰盖上的小开口, 非常有效地逃逸。北冰洋与大气之间的总热交换大约有一半是通过冰洞中的开口发生的。随着更多的水道, 冰盖开口或是更薄的冰, 海冰无法有效地将海洋与大气隔绝。随后北极大气层变暖, 就会影响大气的全球环流。

相较南极海冰, 北极海冰在气候调节中的作用要更大, 因此也更受人关注。这两者的差距主要源于它们的地理位置 (见图 1), 北极的海洋是半封闭的, 几乎完全被陆地包围, 而南极洲是被海洋包围的陆地。开阔的海洋允许海冰更自由地移动, 从而南极海冰比北极海冰更易于移动。而南极的北部没有陆地边界, 海冰可以自由地向北漂浮到温暖的水域进而融化。这就使得南极冬季形成的几乎所有海冰都在夏季融化, 冬季覆盖约 1800 万平方公里的海冰到夏末就仅剩约 300 万平方公里。相比之下, 几乎被陆地包围的北极海冰虽然可以在北极盆地周围移动, 但它往往仍留在寒冷的北极水域。浮冰更容易会聚, 或相互撞击, 从而堆积使冰层变厚。更厚的冰在夏季融化期间能更长地保持冰冻。因此, 更多北极海冰在整个夏季仍然能留存下来, 在冬季存在的约 1500 万平方公里的海冰, 在夏末仍有约 700 万平方公里。

根据观测 (见图 2), 该图显示了使用美国宇航局的海冰算法从海冰指数中计算得到的 1979 年至 2018

<sup>4</sup>在流冰中裂开的一条可以航行的通道。

<sup>5</sup>Credit: National Snow and Ice Data Center, University of Colorado, Boulder, Colorado.

## Annual Mean Sea Ice Extent in the Northern and Southern Hemispheres

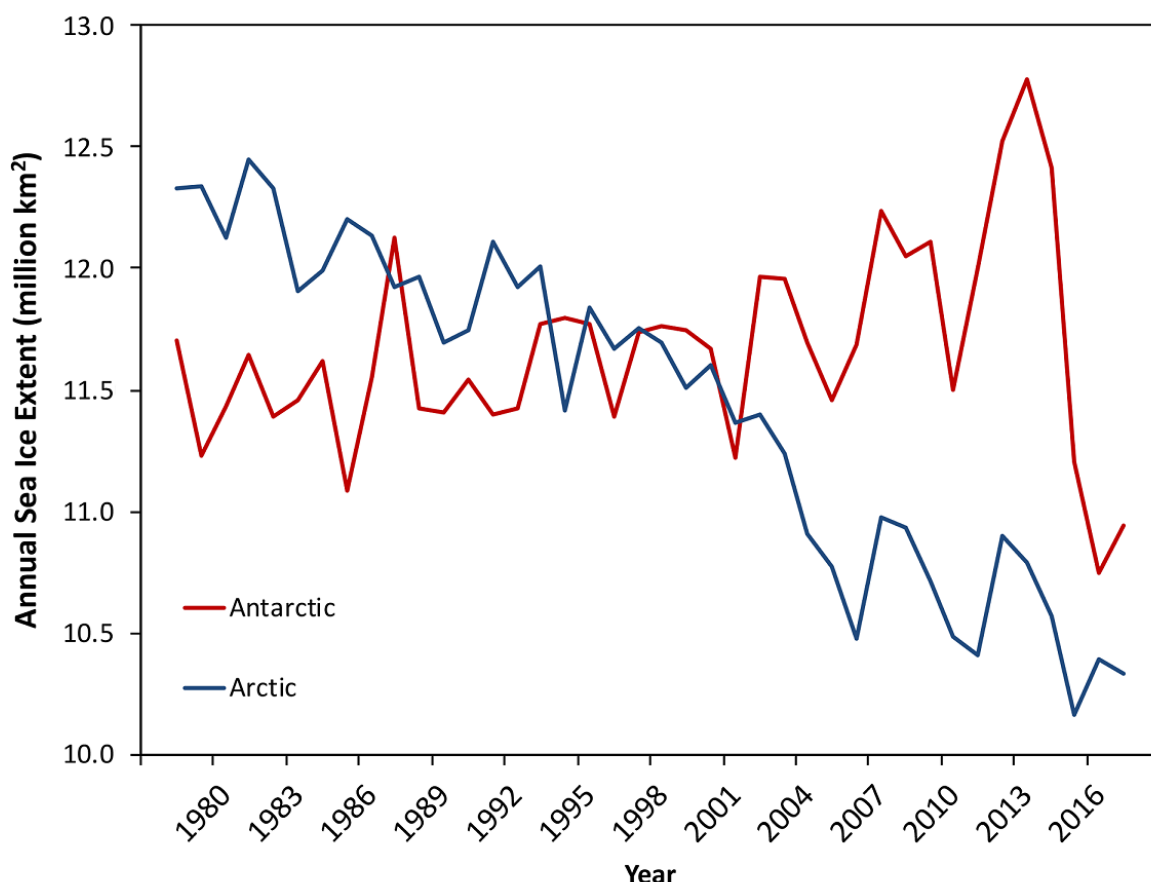


图 2: 南北极海冰范围趋势对比图<sup>7</sup>

年的平均年海冰范围。最近三十年里，北极夏季海冰的厚度与范围均显著下降，其中冰层平均每年降低 3%。这与对北极变暖的观察结果一致。这一趋势是极地地区气候变化的一个主要迹象，可能是全球变暖影响的一个指标。而自 1979 年以来，冬季北极冰面积相对于 1981 年到 2010 年的平均值每十年下降约 3%。南极冰的范围在所有月份都有一个小的增加趋势，但南极冰因其高变异性而更为显着。在从 2012 年到 2014 年每年 9 月创下历史最高纪录之后，南极在 2016 年，2017 年和 2018 年的最大范围内最低。南极海冰似乎不太受全球变暖的影响，因为按照一般的气候模型，大气变得越来越热，加热了上层海水，这从下面增加了海冰的融化。此外，热量增加也将导致降雪减少，更多以雨水的形式降落。而雨水不能像雪花那样反射热量，这就从上层加速了南极海冰的融化。但实际情况却与之不全相符，这主要是因为南极洲是一个被水环绕的大陆，而北极则是被陆地包围的海洋。南极洲周围的风和洋流将该大陆与全球天气模式隔离开来，使其保持寒冷。相比之下，北冰洋与周围的气候系统密切相关，其对气候变化更加敏感。进一步的，有研究<sup>6</sup> 表明南极海冰的变化与南极洲及周围海域的特殊地质环境有关。

<sup>6</sup> 见 <https://www.nasa.gov/feature/jpl/study-helps-explain-sea-ice-differences-at-earths-poles>

<sup>7</sup> Credit: National Snow and Ice Data Center, J.Stroeve



### 1.3 数据简介

海冰面积数据来自 kaggle 整理的数据集<sup>8</sup>，原数据由 National Snow & Ice Data Center 收集的卫星拍摄的亮度数据<sup>9</sup>生成的。数据集提供 1978 年 10 月 26 日至 2019 年 5 月 31 日每天的总海冰范围，有 7 个变量：年，月，日，海冰面积，遗失值，来源链接和所在半球。原数据由于采集航天器的更改，1987 年 12 月 3 日到 1988 年 1 月 12 日的数据存在缺失，且 1987 年 12 月 3 日前的数据由于航天器的功率限制，是隔天采集的，而 1988 年 1 月 12 日后的数据是每天采集的。

全球平均温度数据来自 kaggle 整理的数据集，原数据来自 Berkeley Earth。数据集包括 1750 年到 2015 年每月的全球陆海平均温度及不确定度，还包括全球陆地平均最高，最低温度，陆地平均温度及不确定度等数据，但我们只采用全球陆海平均温度作为研究对象。

### 1.4 章节安排与任务分工

我们将在第 2 节介绍我们对数据进行的预处理，主要是数据的增补与选择。第 3 到第 5 节我们将采用 Sarima，时间序列基本分解和 stl 分解等模型，分别对北极海冰面积，南极海冰面积和全球平均温度来进行模型拟合与预测，并在每节最后对模型拟合效果进行评估。接下来在第六节里我们利用 VAR 模型和脉冲响应分析的理论，对综合的数据进行了 VAR 建模，并对模型进行了脉冲响应分析。最后，在第 7 节里我们对结果进行总结并尝试加以解释。

本课题中，涂博海负责数据的收集与预处理，以及汇总源代码后书写报告；李传赞负责提供第 3-5 节内容的源代码及分析；朱永鑫负责提供第 6 节源代码及分析。

### 1.5 源代码

为了保证本文中数据的可验证性，我们将项目源代码上传至 github<sup>10</sup>以供参考，包括各部分的源代码，分析的文档以及生成报告的 Rmd 文件。

## 2 数据预处理

海冰面积：我们首先将海冰面积数据分为南半球和北半球两个部分，并注意到遗失值相对面积数量级差距很大，忽略遗失值。然后将数据处理为月度数据，我们将每月所有海冰面积值的平均值作为月平均值。但注意到 1987 年 12 月 3 日到 1988 年 1 月 12 日存在数据缺失，对于 1987 年 12 月月平均值，我们舍去 1987 年 12 月的值，用相邻两年 12 月海冰面积的平均值作为代替；对于 1988 年 1 月月平均值，我们先将 1 月 1 日到 1 月 12 日的值用相邻两年的对应值的平均值做增补（1987 年 1 月的值隔天记录，缺失值用相邻两天的值平均替代），再将该月的值做平均作为 1 月月平均值。

全球平均温度：注意到 1850 年前有部分数据缺失，我们提取需要部分的时间的全球陆海平均温度即可。

---

<sup>8</sup> 见 <https://www.kaggle.com/nsidcorg/daily-sea-ice-extent-data/>

<sup>9</sup> 见 <https://nsidc.org/data/nsidc-0051>

<sup>10</sup> 网址见 [https://github.com/tbh3223/TimeSeries\\_Seice](https://github.com/tbh3223/TimeSeries_Seice)

为了保持数据一致性，我们提取南北极海冰面积与全球平均温度在 1979 年 1 月到 2015 年 12 月的部分作为研究对象，且将 2013 年 1 月到 2015 年 12 月的数据留作验证集，其它数据供我们拟合模型。

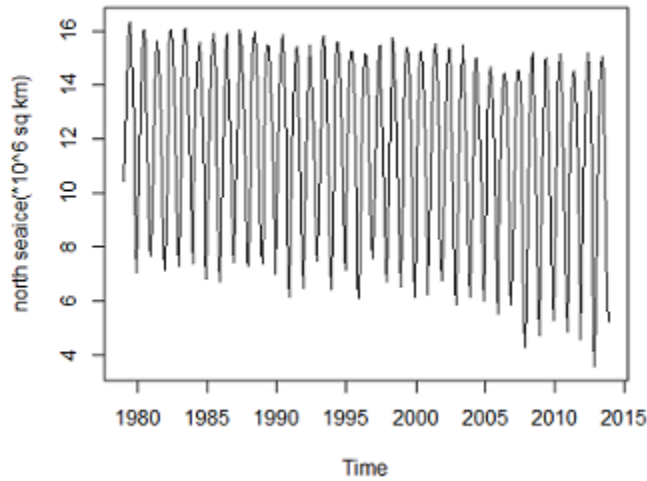


图 3: 北极海冰面积时序图

### 3 北极海冰面积的建模与预测

北极海冰面积时序图如图 3，我们将在下面进行三种模型的尝试性建模。

#### 3.1 季节性 ARIMA 模型

##### 3.1.1 模型构建

根据数据的特点，我们首先选取了季节性 ARIMA 模型进行构建，利用 R 语言的 `auto.arima`，可以方便地进行初步的模拟。模拟结果如下：

```
## Series: sinorth
## ARIMA(1,0,1)(0,1,2)[12] with drift
##
## Coefficients:
##          ar1      ma1      sma1      sma2      drift
##          0.5945  0.3230 -0.8529  0.0901 -0.0044
## s.e.    0.0533  0.0651  0.0497  0.0496  0.0008
##
## sigma^2 estimated as 0.05634: log likelihood=4.16
## AIC=3.68   AICc=3.89   BIC=27.74
```

对其进行残差检验结果如图 4, 图 5。

残差检验结果显示，残差序列可视为白噪声序列，这说明拟合模型显著有效。

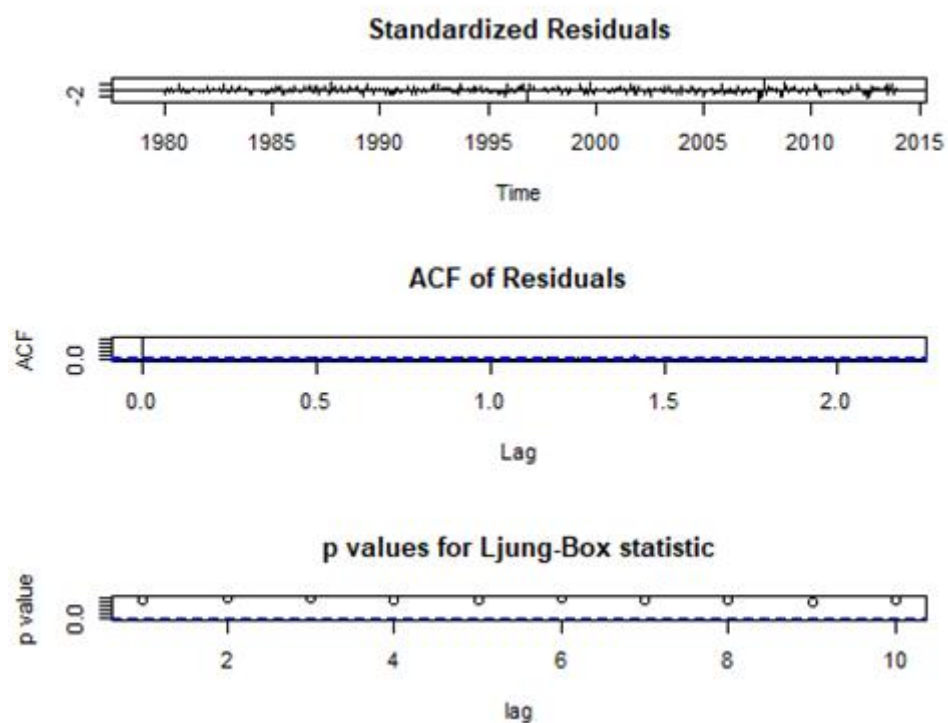


图 4: 北极海冰面积的 SARIMA 模型残差检验结果图

lag	1	2	3	4	5	6
Q.value	0.0115	0.01709	0.27839	1.30234	1.46747	1.49496
p.value	0.91461	0.99149	0.96404	0.86098	0.91679	0.95983
lag	7	8	9	10	11	12
Q.value	3.00089	3.42156	4.86298	5.07273	8.89525	8.9566
p.value	0.88492	0.90519	0.84609	0.88627	0.63156	0.70663

图 5: 北极海冰面积的 SARIMA 模型残差检验结果图

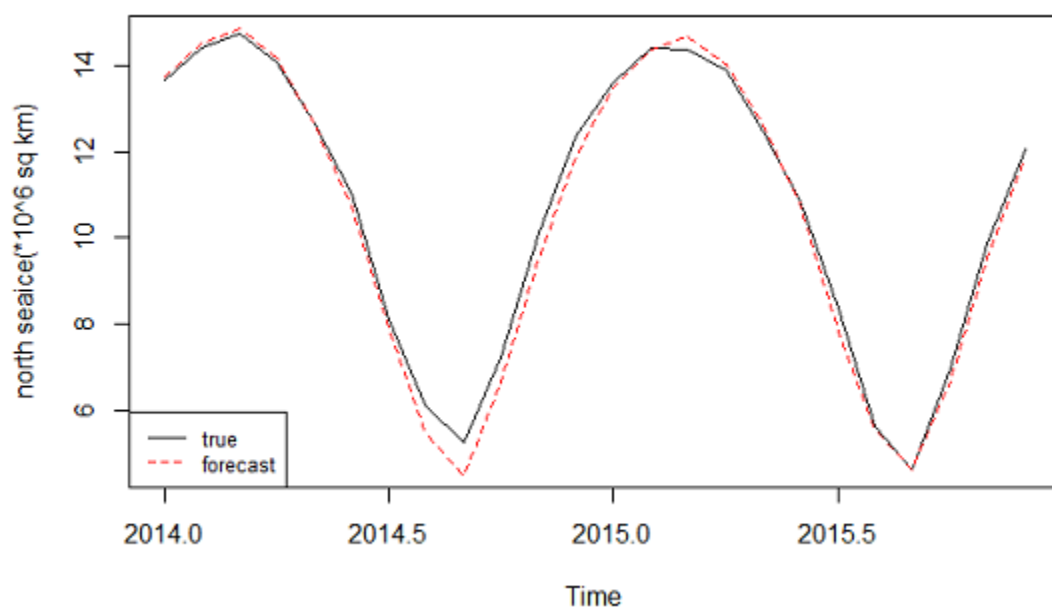


图 6: 北极海冰面积的 SARIMA 模型预测图

### 3.1.2 模型的预测

利用 forecast 对模型进行预测，得到时序图如图 6。

计算其 MSE 可得为 0.113902。

## 3.2 时间序列基本分解模型

对时间序列进行分解，分为趋势项、周期项和随机项，再分别对各项进行拟合，从而达到更加精确的预测。

### 3.2.1 趋势项 T 的估计

注意到时序图有明显趋势，先考虑一次曲线拟合，得到如下结果：

```
##
## Call:
## lm(formula = yy ~ tt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.0673 -2.7381  0.7111  2.8867  4.5333
```

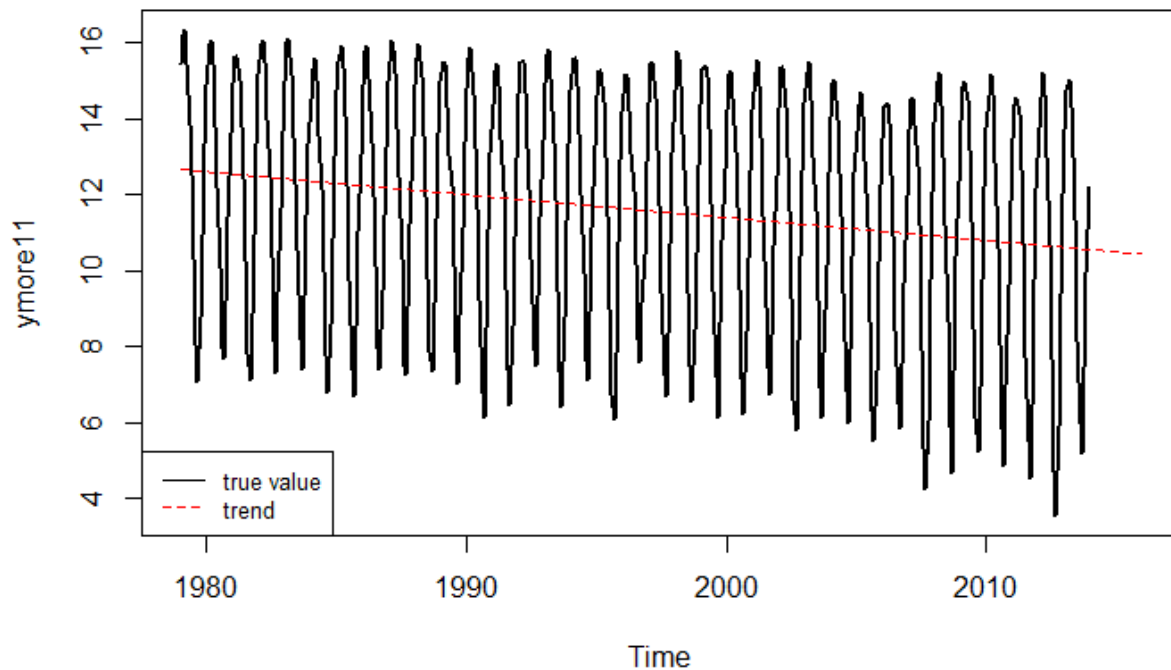


图 7: 含趋势项的时序图

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.662251   0.305663  41.425  < 2e-16 ***
## tt          -0.005011   0.001258  -3.982 8.05e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.127 on 418 degrees of freedom
## Multiple R-squared:  0.03655,    Adjusted R-squared:  0.03425
## F-statistic: 15.86 on 1 and 418 DF,  p-value: 8.052e-05
```

并可以得到含趋势项的时序图（图 7）。

检验二次项系数显著性:

```
##
## Call:
## lm(formula = yy ~ tt + I(tt^2))
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.8796 -2.7964  0.7717  2.8596  4.7024
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.242e+01  4.601e-01  26.996  <2e-16 ***
## tt          -1.594e-03  5.048e-03  -0.316    0.752
## I(tt^2)      -8.115e-06  1.161e-05  -0.699    0.485
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.128 on 417 degrees of freedom
## Multiple R-squared:  0.03768,    Adjusted R-squared:  0.03306
## F-statistic: 8.164 on 2 and 417 DF,  p-value: 0.0003329
```

由参数显著性检验认为二次项系数为零，故认为该趋势不含二次及高次项，因为如果存在高次项，投影到二次项，系数不应该显著为 0。

### 3.2.2 周期项的估计

利用过去去除趋势项的数据中相同月份的均值作周期项的估计，从而得到下表中的数据。

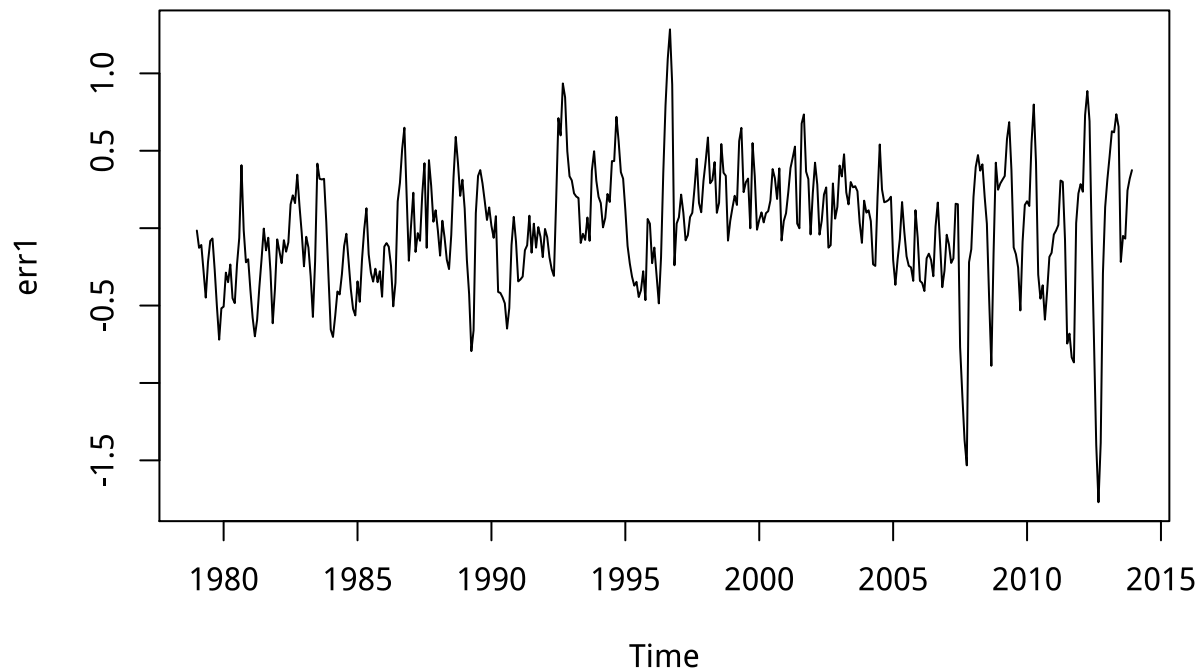
表 1: 北极海冰面积月周期项估计表

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2.7721	3.65	3.8032	3.0818	1.6683	0.1169	-2.2332	-4.5143	-5.2968	-3.3563	-0.9435	1.2517

### 3.2.3 随机项的估计

原数据减去趋势项和周期项的估计后得到随机项的估计，其时序图如下：

随机项的时序图



检验随机项是否平稳

```
##
## Augmented Dickey-Fuller Test
##
## data:  err1
## Dickey-Fuller = -5.8062, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

因为 P 值小于 0.05，拒绝原假设，可以认为序列平稳。

接下来，对随机项进行 arima 拟合，并进行残差检验，结果如下：

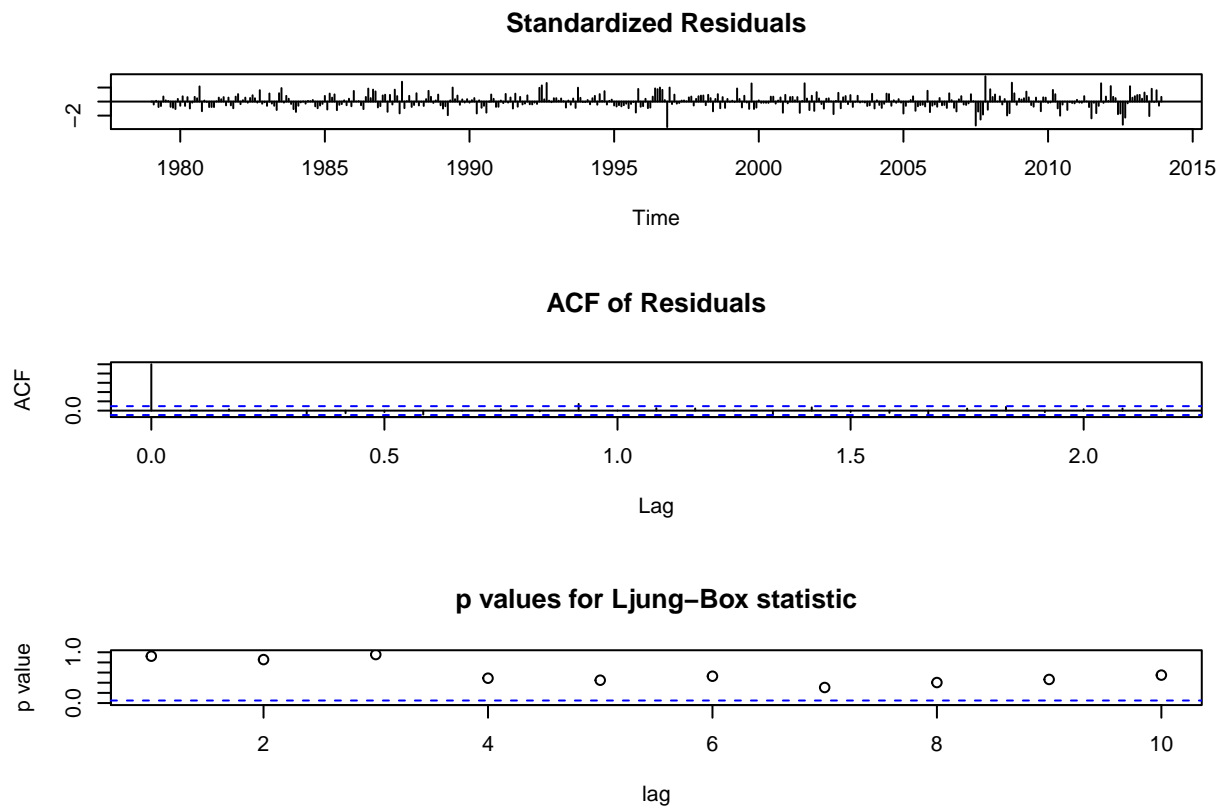
```
## Series: err1
## ARIMA(1,1,2)(0,0,2)[12]
##
## Coefficients:
##          ar1          ma1          ma2          sma1          sma2
##          0.5896   -0.6151   -0.3685    0.1261    0.1948
## s.e.    0.0533    0.0606    0.0573    0.0508    0.0451
```



lag	1	2	3	4	5	6
Q.value	0.00936	0.30881	0.32543	3.3797	4.66149	5.04621
p.value	0.92293	0.85693	0.95518	0.4964	0.45857	0.5379
lag	7	8	9	10	11	12
Q.value	8.16605	8.1667	8.55402	8.65329	17.4372	17.4372
p.value	0.31818	0.41736	0.47942	0.56529	0.09559	0.13388

图 8: box-test 检验结果图

```
##
## sigma^2 estimated as 0.05538:  log likelihood=12.7
## AIC=-13.39  AICc=-13.19  BIC=10.83
```



残差检验结果显示，残差序列基本可视为白噪声序列，但是为了更加精确地拟合预测，检验一下是否存在 arch 效应。其 box-test 和 arch-test 检验结果如下图 8，图 9。

接受原假设，认为有异方差。

然后我们进行了多次产生尝试，发现尽管我们已经去除了周期项，但是周期的信息仍然存在，对模型的拟合产生了很大的影响，所以我们在本次模型构建中，先对残差项进行了 12 阶差分，再对新序列拟合 arma-garch 模型。拟合结果如下：

```
##
```

lag	1	2	3	4	5	6
LM.value	1.51602	9.87434	10.044	23.736	23.9926	25.8952
p.value	0.21822	0.00717	0.0182	9E-05	0.00022	0.00023
lag	7	8	9	10	11	12
LM.value	26.7705	27.5218	28.866	28.8076	32.0697	32.163
p.value	0.00037	0.00057	0.00068	0.00134	0.00074	0.00131

图 9: box-test 检验结果图

```
## *-----*
## *          GARCH Model Fit          *
## *-----*
##
## Conditional Variance Dynamics
## -----
## GARCH Model   : sGARCH(1,1)
## Mean Model    : ARFIMA(1,0,1)
## Distribution   : norm
##
## Optimal Parameters
## -----
##      Estimate Std. Error  t value Pr(>|t|)
## mu      -0.003790   0.039885 -0.095028 0.924293
## ar1      0.508142   0.062336  8.151670 0.000000
## ma1      0.317879   0.069303  4.586816 0.000005
## omega    0.011588   0.016568  0.699421 0.484289
## alpha1   0.064609   0.041437  1.559204 0.118948
## beta1    0.817881   0.201965  4.049616 0.000051
##
## Robust Standard Errors:
##      Estimate Std. Error  t value Pr(>|t|)
## mu      -0.003790   0.040088 -0.094546 0.924675
## ar1      0.508142   0.076507  6.641814 0.000000
## ma1      0.317879   0.063866  4.977288 0.000001
## omega    0.011588   0.032134  0.360602 0.718397
## alpha1   0.064609   0.072999  0.885059 0.376125
## beta1    0.817881   0.397331  2.058437 0.039548
##
## LogLikelihood : -99.91645
##
## Information Criteria
```

```

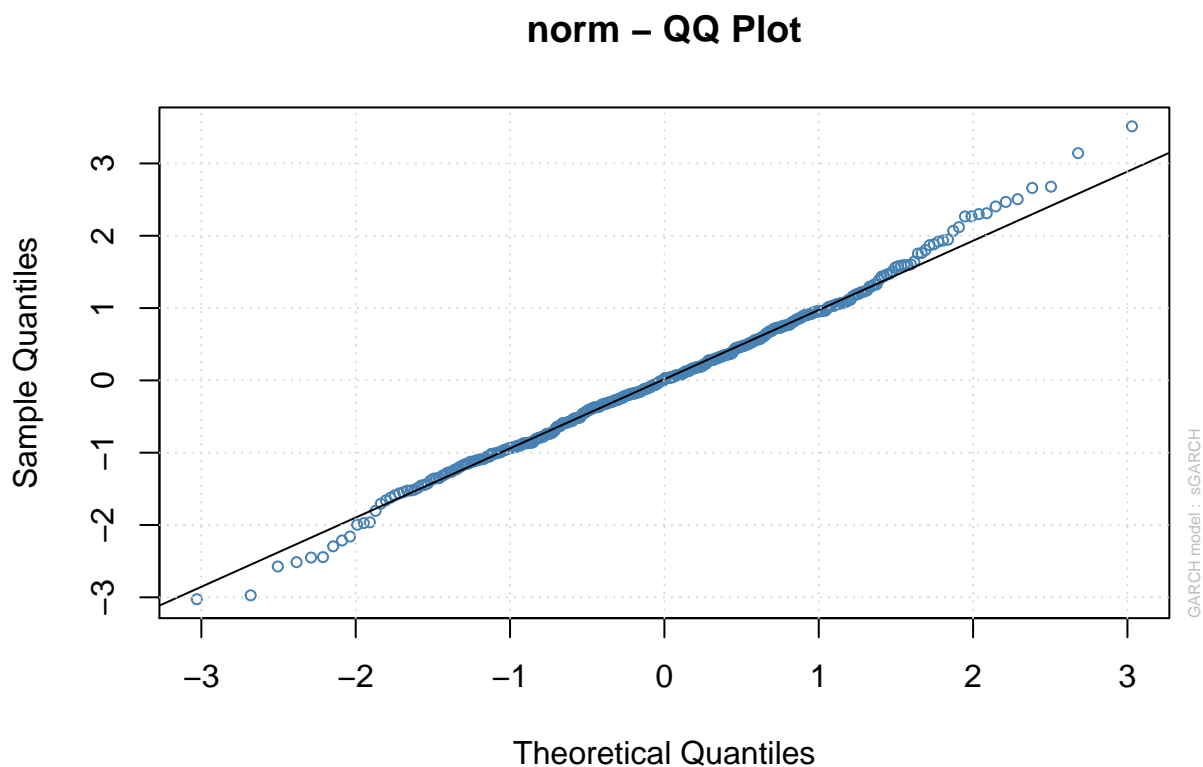
## -----
##
## Akaike          0.51920
## Bayes           0.57819
## Shibata         0.51877
## Hannan-Quinn    0.54254
##
## Weighted Ljung-Box Test on Standardized Residuals
## -----
##
##                statistic p-value
## Lag[1]          0.6117  0.4342
## Lag[2*(p+q)+(p+q)-1] [5]  1.5565  0.9969
## Lag[4*(p+q)+(p+q)-1] [9]  2.9733  0.8946
## d.o.f=2
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## -----
##
##                statistic p-value
## Lag[1]          0.5701  0.450234
## Lag[2*(p+q)+(p+q)-1] [5]  9.4607  0.012647
## Lag[4*(p+q)+(p+q)-1] [9] 16.3236  0.001699
## d.o.f=2
##
## Weighted ARCH LM Tests
## -----
##
##      Statistic Shape Scale P-Value
## ARCH Lag[3]   0.02206 0.500 2.000 0.88192
## ARCH Lag[5]   1.35209 1.440 1.667 0.63205
## ARCH Lag[7]   7.89307 2.315 1.543 0.05542
##
## Nyblom stability test
## -----
## Joint Statistic:  0.7498
## Individual Statistics:
## mu      0.03389
## ar1     0.10832
## ma1     0.02986
## omega   0.37120

```

```

## alpha1 0.28310
## beta1 0.32691
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:      1.49 1.68 2.12
## Individual Statistic: 0.35 0.47 0.75
##
## Sign Bias Test
## -----
##              t-value   prob sig
## Sign Bias      0.2728 0.7852
## Negative Sign Bias 0.7681 0.4429
## Positive Sign Bias 0.5908 0.5550
## Joint Effect    0.9986 0.8016
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## -----
##   group statistic p-value(g-1)
## 1    20      16.41      0.6297
## 2    30      27.44      0.5479
## 3    40      35.14      0.6467
## 4    50      43.47      0.6960
##
##
## Elapsed time : 0.1439168

```



### 3.2.4 模型的预测

利用趋势项、周期项、随机项的预测，随机项进行了二期滚动预测，趋势项和周期项直接预测，再相加得到最后的预测结果，可视化结果如图 10:

计算其 MSE 可得为 0.078740。

## 3.3 LOESS 分解模型

尽管我们已经有了两个模型，但预测的效果不尽满意，这个模型中我们对分解的方法进行改进，采用 LOESS 分解法分解，基本思路是一样的，但这里的问题是 LOESS 分解得到的趋势并不是一个简单的线性趋势，那么我们又对趋势进行了建模拟合，在下面我们可以看到，这种花费大量时间尝试所得模型，对我们的预测很有帮助。

先对测试集数据进行 LOESS 分解，如下图所示:

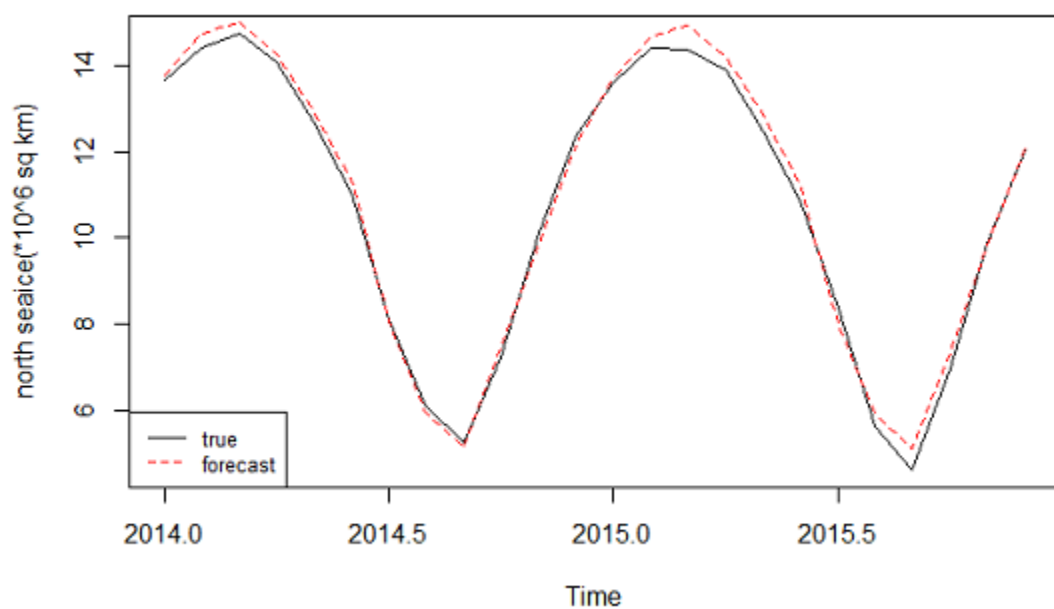
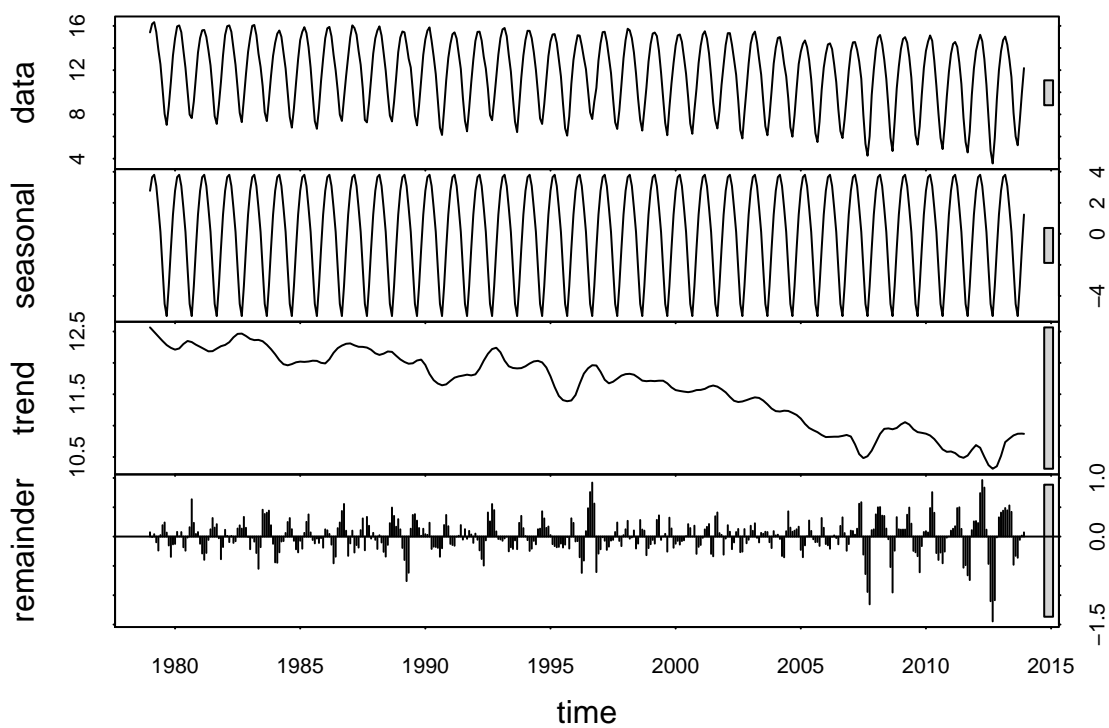


图 10: 北极海冰面积的基本分解模型预测图



January	February	March	April	May	June
2.774199	3.651674	3.80439	3.082539	1.668562	0.1166301
July	August	September	October	November	December
-2.233975	-4.51552	-5.298458	-3.358438	-0.9461913	1.2486

图 11: 月周期估计值

### 3.3.1 趋势项 T 的估计

在进行多项式拟合的过程中，发现二次曲线模型显著，利用其进行拟合。

```
##
## Call:
## lm(formula = trand4 ~ tt + I(tt^2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.50229 -0.13092  0.03738  0.12680  0.44704
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.232e+01  2.553e-02 482.547 < 2e-16 ***
## tt          -1.118e-03  2.801e-04  -3.992 7.74e-05 ***
## I(tt^2)      -8.108e-06  6.443e-07 -12.585 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1736 on 417 degrees of freedom
## Multiple R-squared:  0.9128, Adjusted R-squared:  0.9124
## F-statistic: 2183 on 2 and 417 DF, p-value: < 2.2e-16
```

### 3.3.2 周期项的估计

利用过去去除趋势项的数据中相同月份的均值作周期项的估计，从而得到下表中的数据。

### 3.3.3 随机项的估计

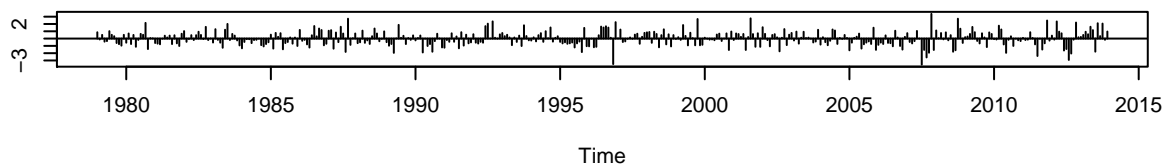
原始数据减去趋势项和月周期项得到随机项，对其做平稳性检验如下：

```
##
## Augmented Dickey-Fuller Test
##
## data: err4
## Dickey-Fuller = -6.7221, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

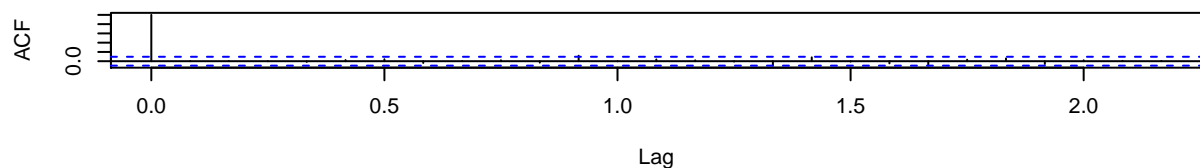
拒绝原假设，可以认为序列平稳，下面用 arima 模型去拟合：

```
## Series: err4
## ARIMA(2,0,3)(0,0,2)[12] with zero mean
##
## Coefficients:
##          ar1      ar2      ma1      ma2      ma3      sma1      sma2
##          1.3280  -0.5832  -0.3765  -0.1085   0.1382   0.0983   0.1563
## s.e.    0.1572   0.1109   0.1668   0.0869   0.0794   0.0513   0.0468
##
## sigma^2 estimated as 0.05371: log likelihood=20.77
## AIC=-25.53   AICc=-25.18   BIC=6.79
```

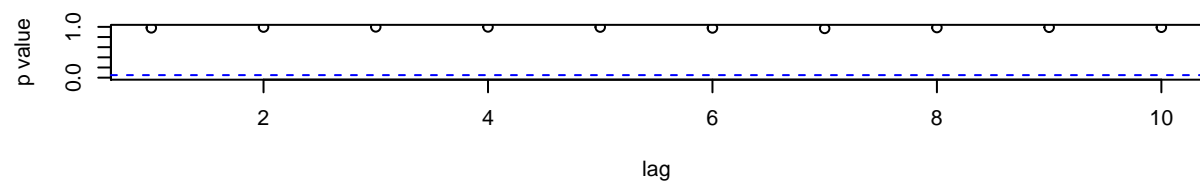
**Standardized Residuals**



**ACF of Residuals**



**p values for Ljung-Box statistic**





box-test result

lag	1	2	3	4	5	6
Q.value	0.00157	0.00571	0.00617	0.04612	0.18851	0.57519
p.value	0.96836	0.99715	0.99987	0.99974	0.99923	0.9968
lag	7	8	9	10	11	12
Q.value	1.47096	1.59455	2.41569	2.51759	9.22807	9.22896
p.value	0.9833	0.99102	0.98307	0.99062	0.60085	0.68326

arch-test result

lag	1	2	3	4	5	6
LM.value	2.1911	9.30926	9.31215	21.4518	21.8866	23.4664
p.value	0.13881	0.00952	0.02542	0.00026	0.00055	0.00065
lag	7	8	9	10	11	12
LM.value	24.2103	24.9105	25.1124	25.0386	28.0268	28.0097
p.value	0.00105	0.00161	0.00285	0.00527	0.00321	0.00551

图 12: box-text 和 arch-test 检验值

对其做 box-text 和 arch-test, 结果如图 12。

可见存在 arch 效应, 利用 arma-garch 模型进行拟合, 残差项正态性良好, 可以进行应用。下面进行拟合:

```
##
## *-----*
## *          GARCH Model Fit          *
## *-----*
##
## Conditional Variance Dynamics
## -----
## GARCH Model   : sGARCH(1,1)
## Mean Model    : ARFIMA(2,0,1)
## Distribution   : norm
##
## Optimal Parameters
## -----
##      Estimate Std. Error  t value Pr(>|t|)
## mu      -0.013877    0.039481 -0.35149 0.725219
## ar1       0.472144    0.245439  1.92367 0.054396
## ar2       0.022136    0.177201  0.12492 0.900584
## ma1       0.358981    0.233822  1.53527 0.124717
## omega     0.010730    0.012084  0.88793 0.374579
## alpha1    0.067633    0.037753  1.79146 0.073219
## beta1     0.824523    0.149630  5.51040 0.000000
##
## Robust Standard Errors:
```

```

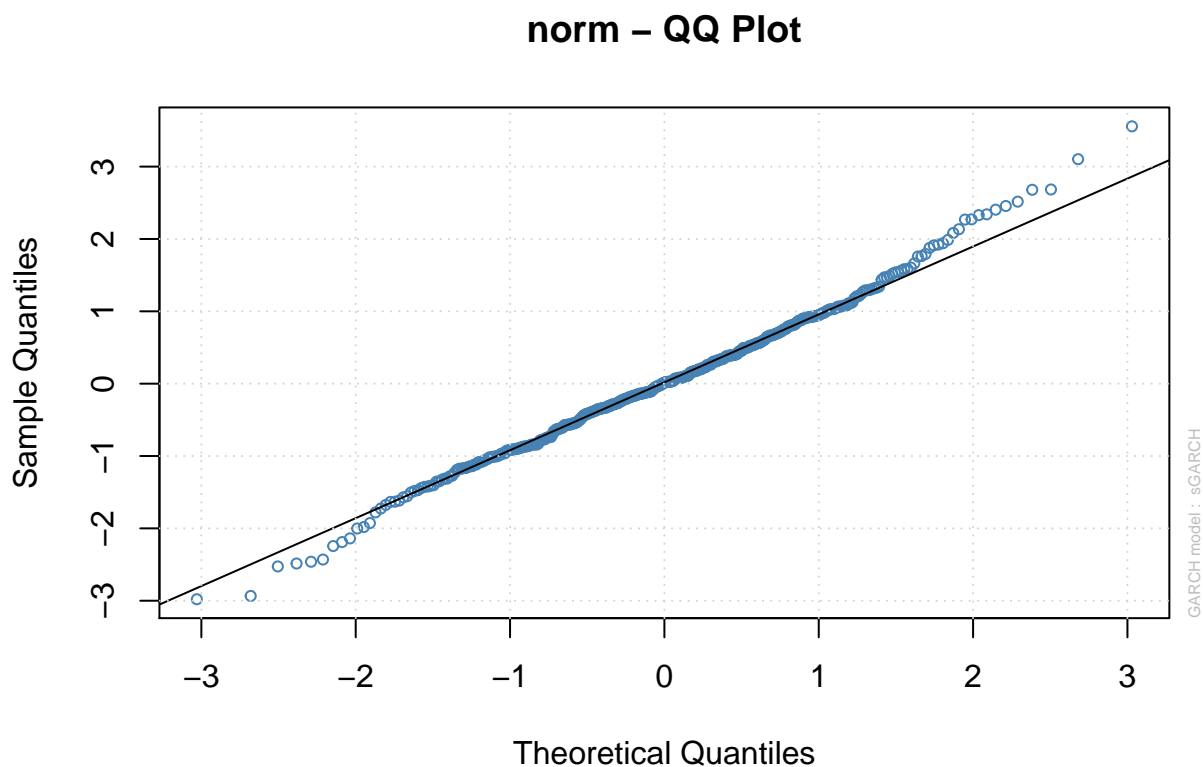
##      Estimate Std. Error t value Pr(>|t|)
## mu      -0.013877    0.039283 -0.35327 0.723889
## ar1      0.472144    0.285968  1.65104 0.098731
## ar2      0.022136    0.200219  0.11056 0.911964
## ma1      0.358981    0.261563  1.37244 0.169925
## omega    0.010730    0.018202  0.58949 0.555536
## alpha1   0.067633    0.054498  1.24103 0.214596
## beta1    0.824523    0.230622  3.57521 0.000350
##
## LogLikelihood : -99.33424
##
## Information Criteria
## -----
##
## Akaike          0.52125
## Bayes           0.59007
## Shibata         0.52067
## Hannan-Quinn    0.54848
##
## Weighted Ljung-Box Test on Standardized Residuals
## -----
##
##              statistic    p-value
## Lag[1]              0.7673 3.811e-01
## Lag[2*(p+q)+(p+q)-1] [8]    2.6988 9.997e-01
## Lag[4*(p+q)+(p+q)-1] [14]   30.0811 2.863e-11
## d.o.f=3
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## -----
##
##              statistic    p-value
## Lag[1]              0.7388 0.3900400
## Lag[2*(p+q)+(p+q)-1] [5]    10.3017 0.0077017
## Lag[4*(p+q)+(p+q)-1] [9]    17.2805 0.0009831
## d.o.f=2
##
## Weighted ARCH LM Tests
## -----
##              Statistic Shape Scale P-Value

```

```

## ARCH Lag[3]  0.005048 0.500 2.000 0.94336
## ARCH Lag[5]  1.332593 1.440 1.667 0.63742
## ARCH Lag[7]  7.927413 2.315 1.543 0.05446
##
## Nyblom stability test
## -----
## Joint Statistic:  1.5889
## Individual Statistics:
## mu      0.03575
## ar1     0.12137
## ar2     0.08818
## ma1     0.04340
## omega   0.34434
## alpha1  0.26567
## beta1   0.29975
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:      1.69 1.9 2.35
## Individual Statistic:  0.35 0.47 0.75
##
## Sign Bias Test
## -----
##
##          t-value   prob sig
## Sign Bias      0.7083 0.4791
## Negative Sign Bias  1.0922 0.2754
## Positive Sign Bias  0.4053 0.6854
## Joint Effect      1.5827 0.6633
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## -----
##   group statistic p-value(g-1)
## 1    20      18.08      0.5172
## 2    30      28.32      0.5007
## 3    40      32.59      0.7560
## 4    50      59.16      0.1518
##
##
## Elapsed time : 0.150914

```



### 3.3.4 模型的预测

对上述模型进行预测得到图 13。

计算其 MSE 为 0.061675

### 3.4 预测结果对比与模型评论

	model1	model2	model3
MSE	0.113902	0.07874	0.061675

其中 model1 分别表示北极海冰面积的 SARIMA 模型，基本分解模型和 LOESS 分解模型。由 MSE 可见 LOESS 分解模型要优于另外两个模型。

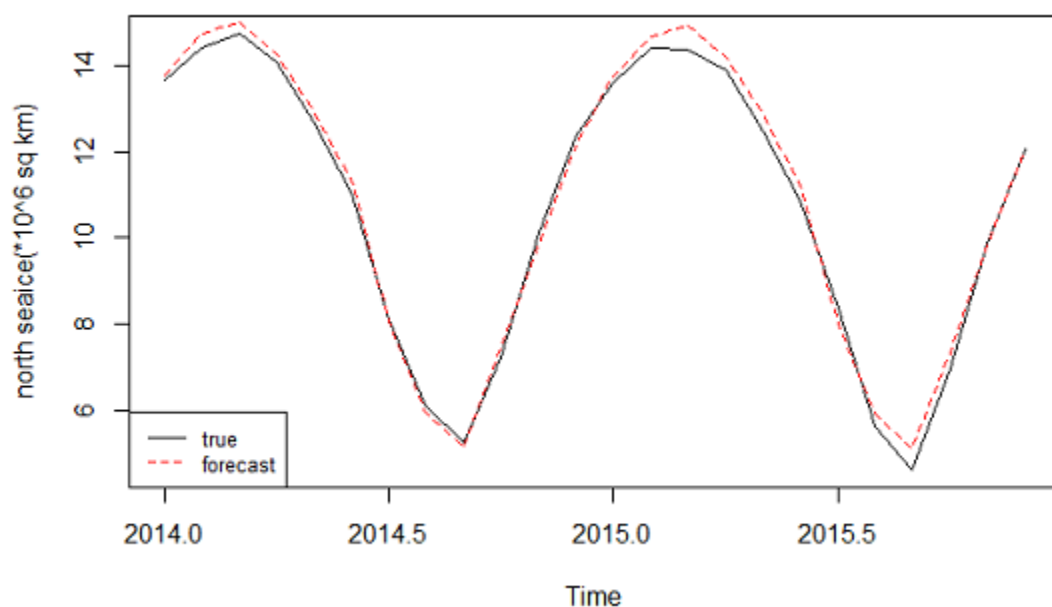


图 13: 北极海冰面积的 LOESS 分解模型预测图

## 4 南极海冰面积的建模与预测

从北极海冰面积预测的结果来看，直接用 `auto.arima` 拟合预测远没有自己建模预测的好，接下来就类比北极海冰的后两个模型，建立南极海冰和温度的预测模型。

南极海冰面积时序图如图 14。

### 4.1 时间序列的基本分解模型

#### 4.1.1 趋势项 $T$ 的估计

时序图趋势不明显，考虑线性拟合

```
##
## Call:
## lm(formula = yy ~ tt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.543  -5.709   0.325   5.308   7.696
##
## Coefficients:
```

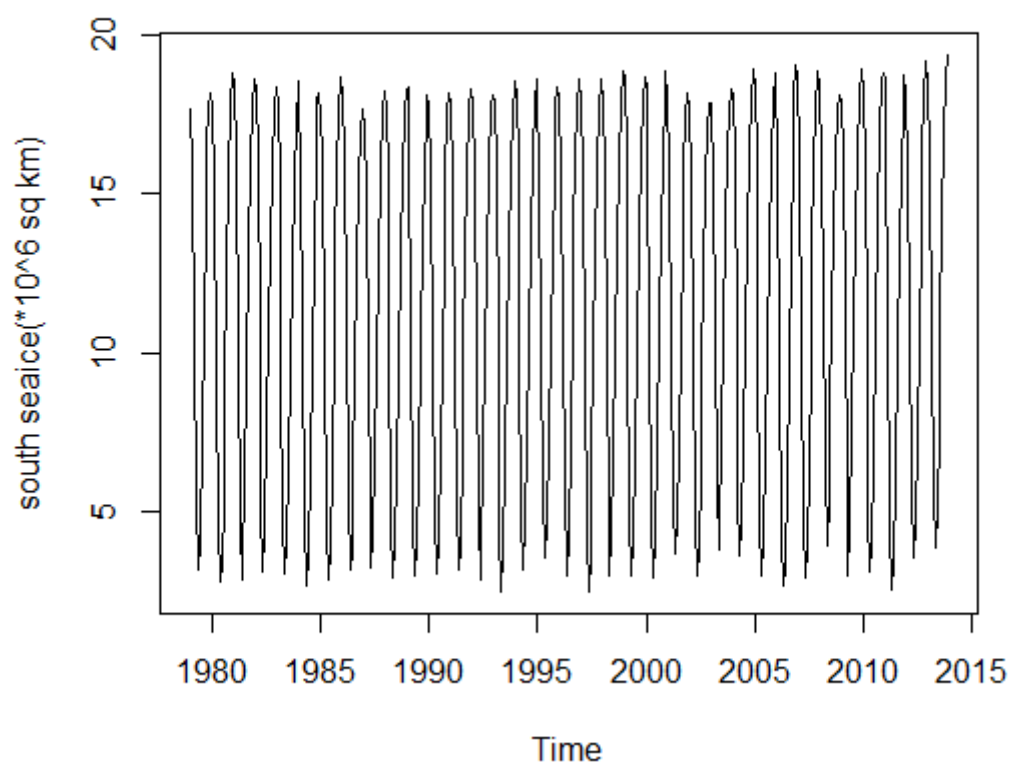


图 14: 南极海冰面积时序图

January	February	March	April	May	June
-6.598122	-8.527741	-7.578035	-4.771089	-1.437343	4.383283
July	August	September	October	November	December
6.156491	6.931734	1.777373	6.536328	4.307105	-1.171046

图 15: 南极海冰月周期项估计图

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.067121   0.544826   20.31  <2e-16 ***
## tt          0.002578   0.002243    1.15   0.251
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.573 on 418 degrees of freedom
## Multiple R-squared:  0.003152,    Adjusted R-squared:  0.0007671
## F-statistic: 1.322 on 1 and 418 DF,  p-value: 0.251
```

我们发现一次项系数并不显著，利用均值进行趋势项的拟合。

#### 4.1.2 周期项的估计

利用过去去除趋势项的数据中相同月份的均值作周期项的估计，从而得到图 15 中的数据。

#### 4.1.3 随机项的拟合

对随机项检验平稳性

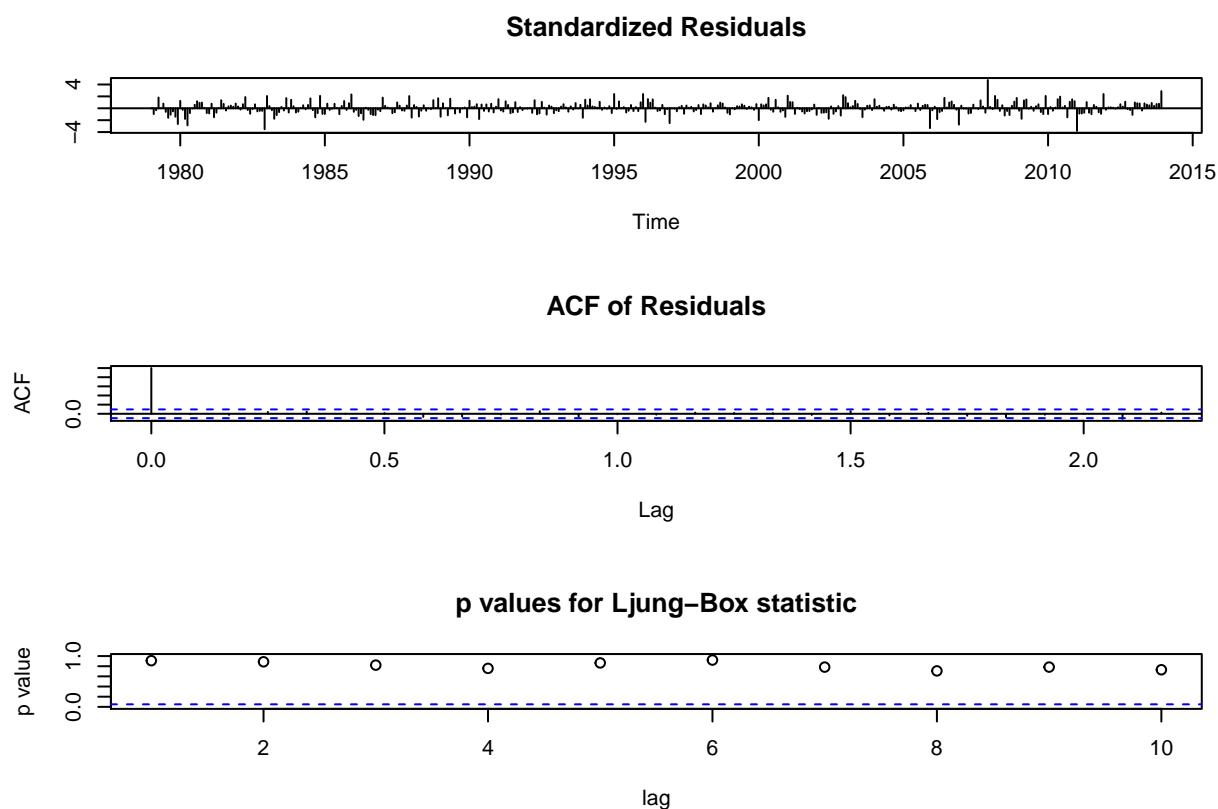
```
##
## Augmented Dickey-Fuller Test
##
## data:  err1
## Dickey-Fuller = -6.8977, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

拒绝原假设，可以认为序列平稳。

用 auto.arima 建模并进行残差检验如下：

```
## Series: err1
## ARIMA(1,1,2)(2,0,2)[12]
```

```
##
## Coefficients:
##          ar1          ma1          ma2          sar1          sar2          sma1          sma2
##          0.5604 -0.7701 -0.1987  0.0327 -0.6088  0.0356  0.4139
## s.e.    0.0675  0.0797  0.0751  0.1719  0.1561  0.1967  0.1803
##
## sigma^2 estimated as 0.09237:  log likelihood=-93.76
## AIC=203.53   AICc=203.88   BIC=235.83
```



对残差进行 box-test 和 arch-test 结果如图 16

#### 4.1.4 模型的预测

利用趋势项、周期项、随机项进行预测。随机项进行了二期滚动预测，趋势项和周期项直接预测，再相加得到最后的预测结果，可视化结果如图 17:

计算其 MSE 值为 0.486908

## 4.2 LOESS 分解模型

考虑用 stl 进行分解，分解结果如下:



box-test result

lag	1	2	3	4	5	6
Q.value	0.01144	0.03273	0.10761	0.66028	0.85685	0.85915
p.value	0.91481	0.98377	0.99091	0.95614	0.97327	0.99039
lag	7	8	9	10	11	12
Q.value	4.21081	6.91533	7.96299	8.82632	10.75288	10.75290
p.value	0.75520	0.54579	0.53788	0.54866	0.46419	0.55021

arch-test result

lag	1	2	3	4	5	6
LM.value	0.06398	0.22398	0.23119	0.23858	0.41305	0.59080
p.value	0.80032	0.89405	0.97240	0.99343	0.99496	0.99655
lag	7	8	9	10	11	12
LM.value	0.59345	0.67138	0.78672	0.96354	1.84520	3.48877
p.value	0.99903	0.99959	0.99979	0.99985	0.99897	0.99100

图 16: box-test 和 arch-test 检验结果

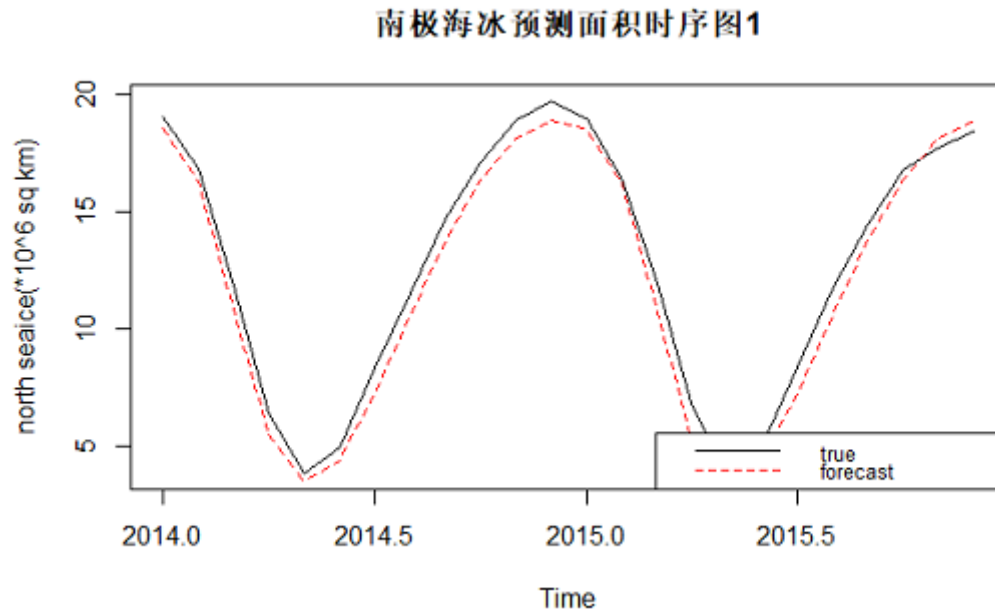
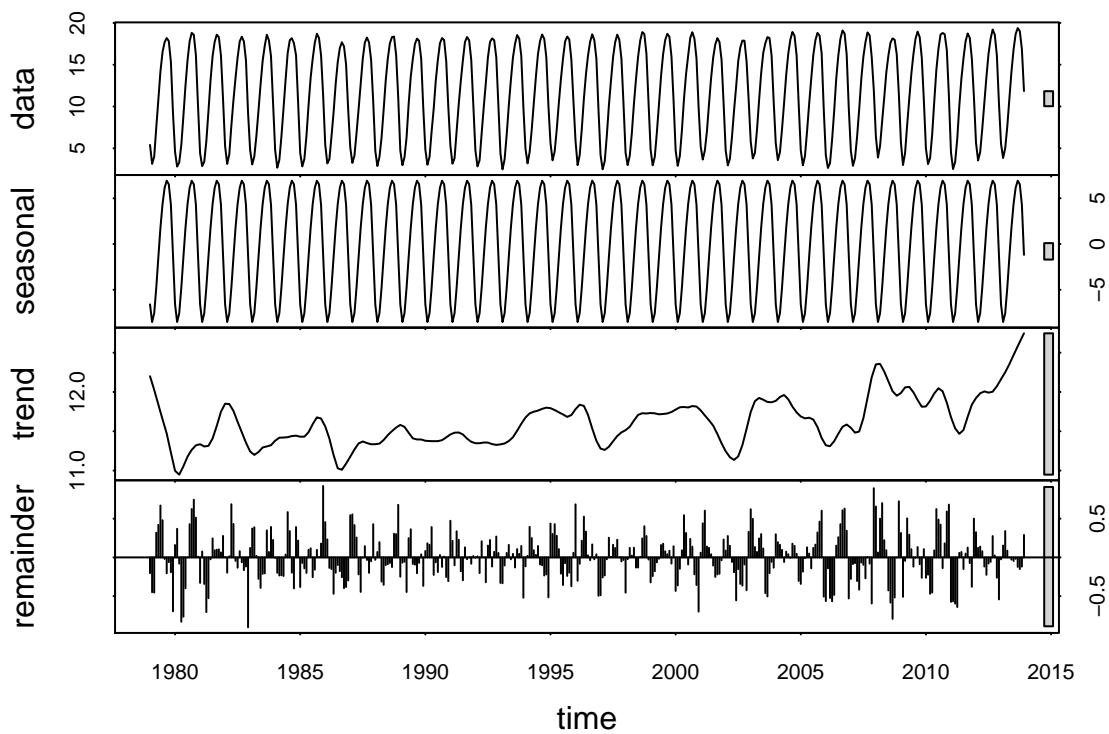


图 17: 南极海冰面积的基本分解模型预测图



#### 4.2.1 趋势项 $T$ 的估计

考虑用一次项拟合趋势项结果如下：

```
##
## Call:
## lm(formula = trand4 ~ tt + I(tt^2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.52876 -0.12127 -0.00854  0.15986  0.74188
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.146e+01  3.364e-02 340.659  < 2e-16 ***
## tt          -9.403e-04  3.690e-04  -2.548   0.0112 *
## I(tt^2)       5.936e-06  8.488e-07   6.993 1.07e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.2287 on 417 degrees of freedom
## Multiple R-squared: 0.446, Adjusted R-squared: 0.4433
## F-statistic: 167.8 on 2 and 417 DF, p-value: < 2.2e-16
```

#### 4.2.2 周期项的估计

我们将对应月的所有值的平均作为月周期项，结果如下：

表 3: 南极海冰面积月周期项估计表

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2.8037	3.6751	3.8218	3.0939	1.6739	0.1158	-2.2408	-4.5285	-5.3175	-3.3837	-0.9775	1.2111

#### 4.2.3 随机项的拟合

检验随机项是否平稳

```
##
## Augmented Dickey-Fuller Test
##
## data: err4
## Dickey-Fuller = -19.976, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

拒绝原假设，可以认为序列平稳

可以利用 `auto.arima` 进行建模，模型参数及残差检验如下：

```
## Series: err4
## ARIMA(3,0,0)(2,1,0)[12]
##
## Coefficients:
##          ar1          ar2          ar3          sar1          sar2
##          0.7822    -0.1965    0.1033    -0.4997    -0.4277
## s.e.    0.0497     0.0623    0.0497     0.0472     0.0469
##
## sigma^2 estimated as 0.1303: log likelihood=-164.16
## AIC=340.32   AICc=340.53   BIC=364.39
```

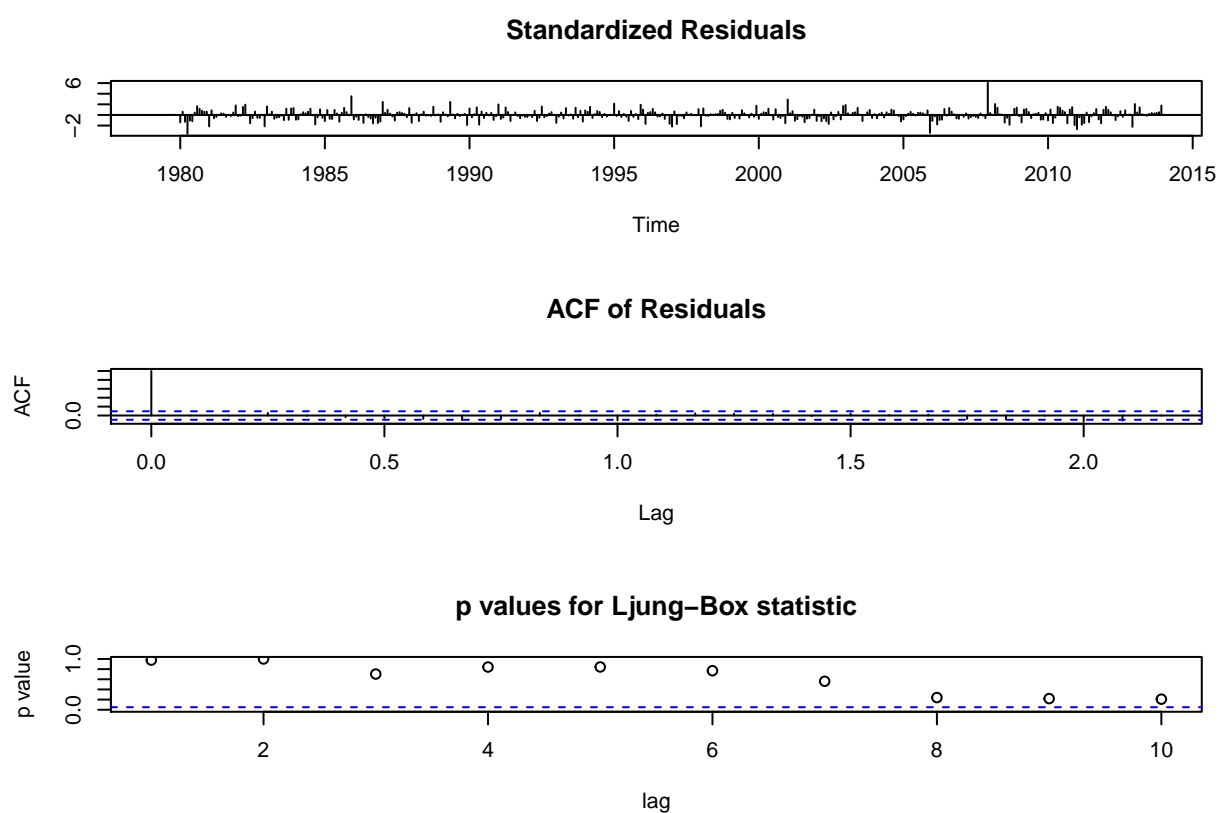
box-test result

lag	1	2	3	4	5	6
Q.value	0.01144	0.03273	0.10761	0.66028	0.85685	0.85915
p.value	0.91481	0.98377	0.99091	0.95614	0.97327	0.99039
lag	7	8	9	10	11	12
Q.value	4.21081	6.91533	7.96299	8.82632	10.75288	10.75290
p.value	0.75520	0.54579	0.53788	0.54866	0.46419	0.55021

arch-test result

lag	1	2	3	4	5	6
LM.value	0.06398	0.22398	0.23119	0.23858	0.41305	0.59080
p.value	0.80032	0.89405	0.97240	0.99343	0.99496	0.99655
lag	7	8	9	10	11	12
LM.value	0.59345	0.67138	0.78672	0.96354	1.84520	3.48877
p.value	0.99903	0.99959	0.99979	0.99985	0.99897	0.99100

图 18: box-test 和 arch-test 检验的结果



对数据进行 box-test 和 arch-test 检验结果如图 18

可见不存在 arch 效应，利用 arima 模型进行拟合，残差项随机性良好，可以进行应用。

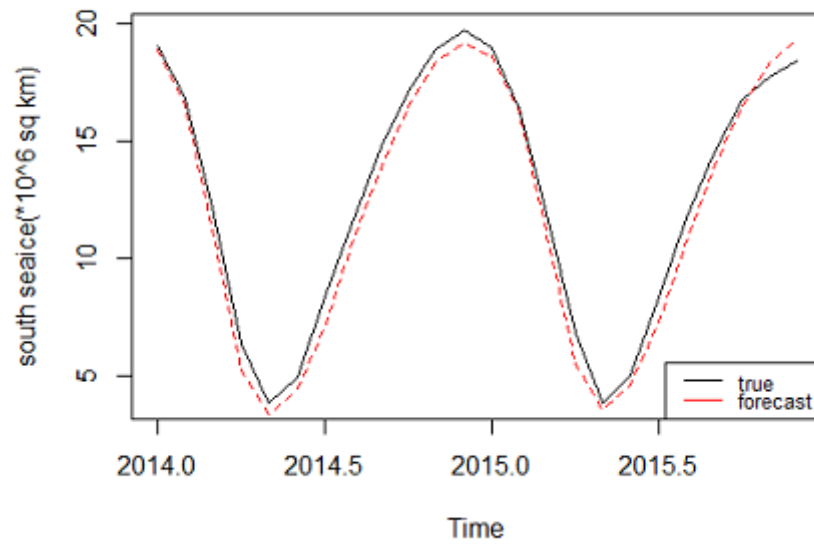


图 19: 南极海冰面积的 LOESS 模型预测图

#### 4.2.4 模型的预测

对上述模型预测如图 19

计算其 MSE 为 0.363085

#### 4.3 预测结果对比与模型评论

	model1	model2
MSE	0.486908	0.363085

其中 model1 分别表示南极海冰面积的基本分解模型和 LOESS 分解模型。由 MSE 可见 LOESS 分解模型要优于基本分解模型。

## 5 全球平均温度的建模与预测

### 5.1 时间序列基本分解模型

#### 5.1.1 趋势项 $T$ 的估计

首先对 `tem` 进行趋势项分解，注意到时序图有一定趋势，利用一次函数拟合。

```
##
## Call:
## lm(formula = yy ~ tt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.7189 -3.9721  0.2321  3.9921  6.0600
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.696849   0.404167  21.518  <2e-16 ***
## tt           0.002481   0.001664   1.491   0.137
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.134 on 418 degrees of freedom
## Multiple R-squared:  0.005294,    Adjusted R-squared:  0.002914
## F-statistic: 2.224 on 1 and 418 DF,  p-value: 0.1366
```

没有通过显著性检验，用平均值估计趋势项

#### 5.1.2 周期项的估计

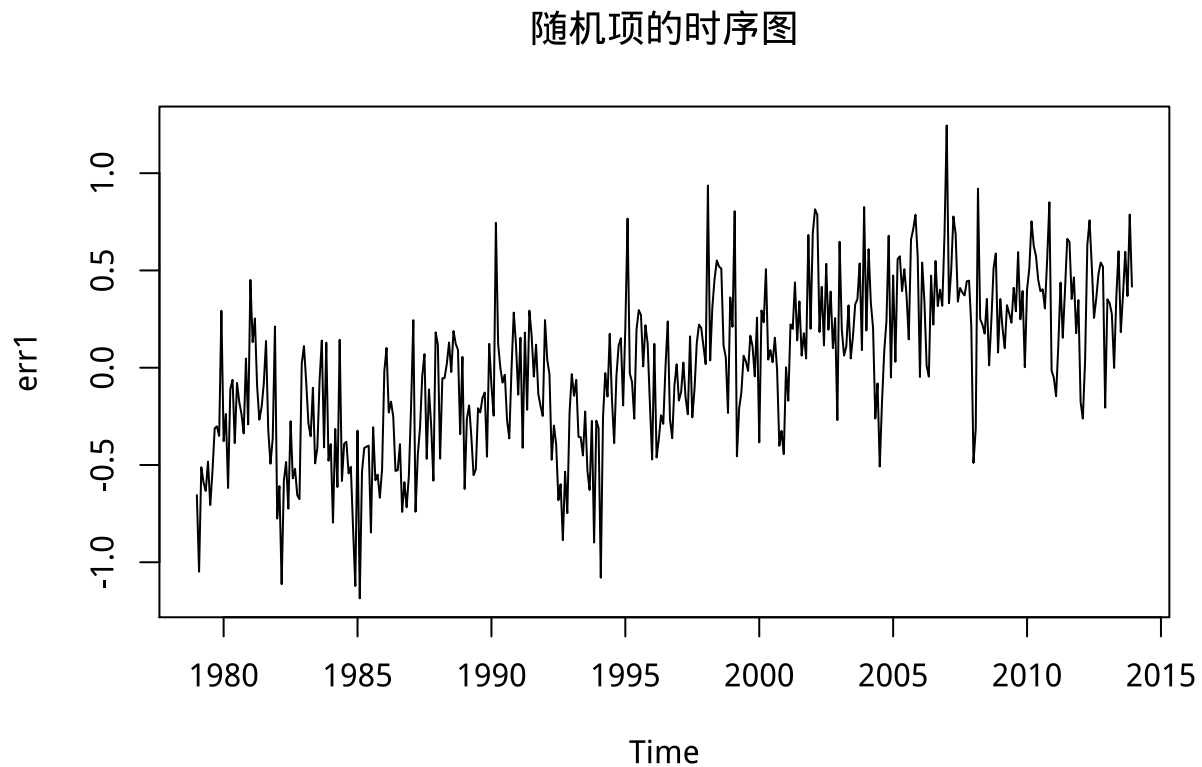
利用过去去除趋势项的数据中相同月份的均值作周期项的估计，从而得到下表中的数据：

表 5: 全球平均温度月周期项估计表

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
-5.8855	-5.3297	-3.2337	-0.1739	2.6131	4.7503	5.6009	5.1455	3.3393	0.6683	-2.5826	-4.9119

### 5.1.3 随机项的拟合

画出随机项的时序图



检验随机项是否平稳

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: err1  
## Dickey-Fuller = -5.3691, Lag order = 7, p-value = 0.01  
## alternative hypothesis: stationary
```

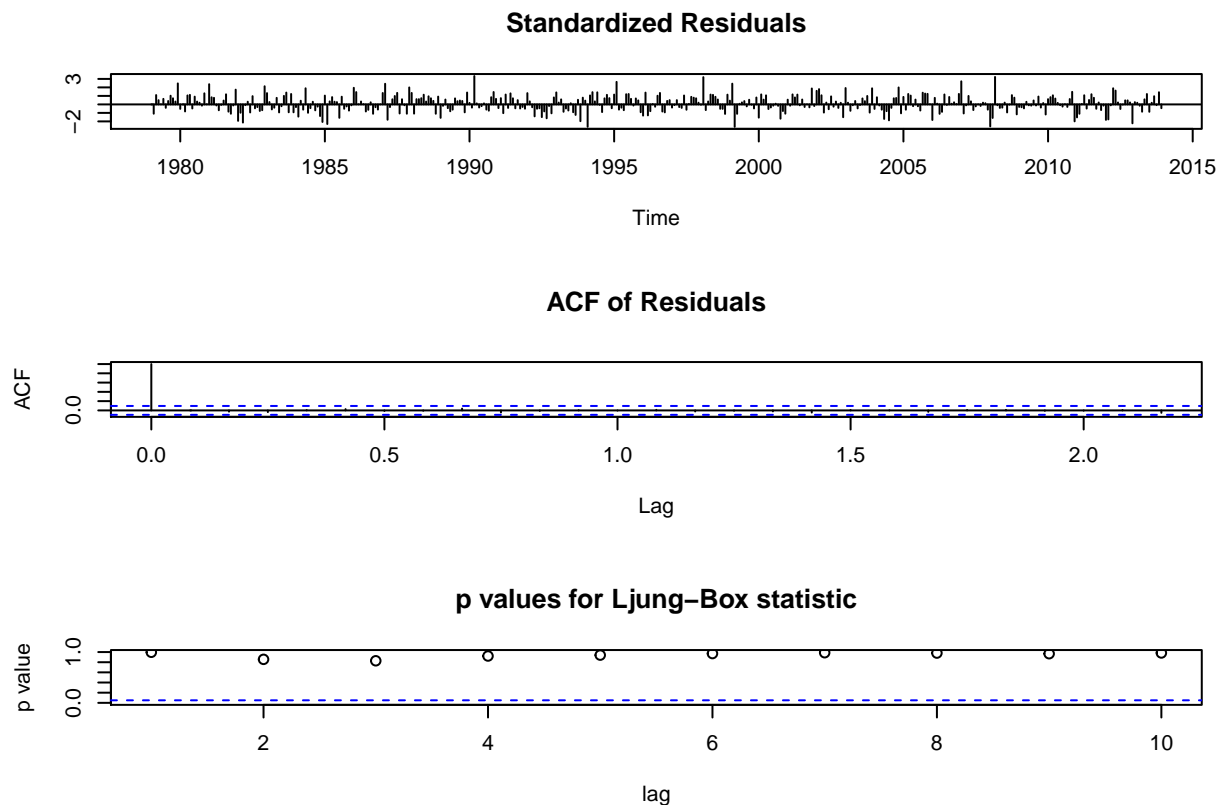
拒绝原假设，可以认为随机项序列平稳。

```
## Series: err1  
## ARIMA(2,1,3)(2,0,1)[12]  
##  
## Coefficients:  
##          ar1      ar2      ma1      ma2      ma3      sar1      sar2      sma1  
##      -0.1134  0.6749  -0.6050  -0.7339  0.3749  -0.6161  -0.0178  0.5280
```

box-test result						
lag	1	2	3	4	5	6
Q.value	0.000026	0.306693	0.878821	0.905647	1.231712	1.301804
p.value	0.995973	0.857832	0.830536	0.923749	0.941799	0.971558
lag	7	8	9	10	11	12
Q.value	1.329680	1.991262	2.862802	2.918849	2.933639	2.943293
p.value	0.987628	0.981279	0.969487	0.983270	0.991574	0.995931
arch-test result						
lag	1	2	3	4	5	6
LM.value	0.022483	0.810835	1.219226	2.375976	3.293787	4.589064
p.value	0.880809	0.666698	0.748397	0.666973	0.654793	0.597489
lag	7	8	9	10	11	12
LM.value	5.541574	6.335175	7.026628	7.885481	8.262051	12.880235
p.value	0.594174	0.609742	0.634347	0.640022	0.689655	0.377806

图 20: box-test 和 arch-test 的检验结果

```
## s.e.    0.1308  0.1196  0.1514  0.1695  0.1256  0.6300  0.0786  0.6289
##
## sigma^2 estimated as 0.08322:  log likelihood=-70.45
## AIC=158.9  AICc=159.34  BIC=195.24
```



对残差进行 box-test 和 arch-test 的检验, 结果如图 20



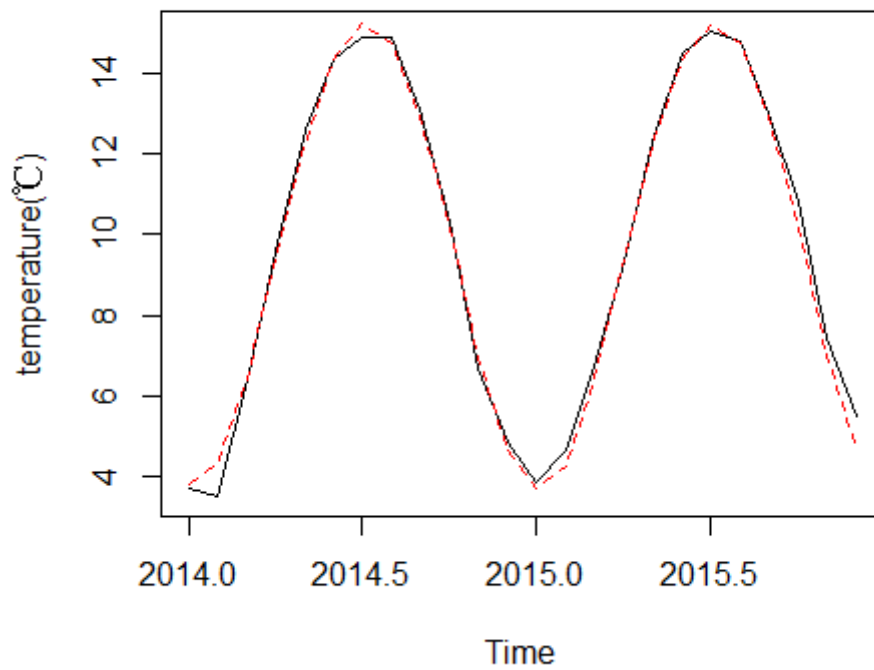


图 21: 全球平均温度的基本分解模型预测图

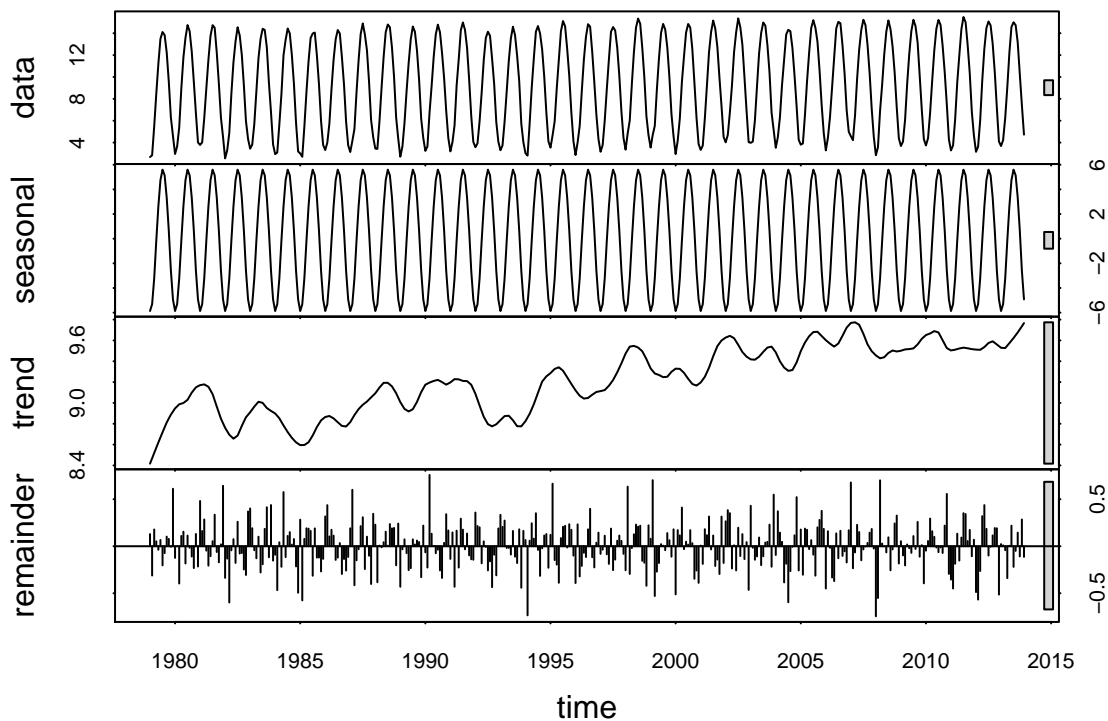
#### 5.1.4 模型的预测

利用以上模型可以预测得到图 21

可以计算得其 MSE 为 0.1092166

## 5.2 LOESS 分解模型

先对数据进行 STL 分解，结果如下图所示：



### 5.2.1 趋势项 $T$ 的估计

在进行多项式拟合的过程中，发现二次曲线模型显著，利用其进行拟合。

```
##
## Call:
## lm(formula = trand4 ~ tt + I(tt^2))
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.36852	-0.10981	-0.00696	0.12385	0.36548

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	8.755e+00	2.343e-02	373.728	< 2e-16 ***
tt	2.110e-03	2.570e-04	8.211	2.77e-15 ***
I(tt^2)	3.282e-07	5.911e-07	0.555	0.579

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

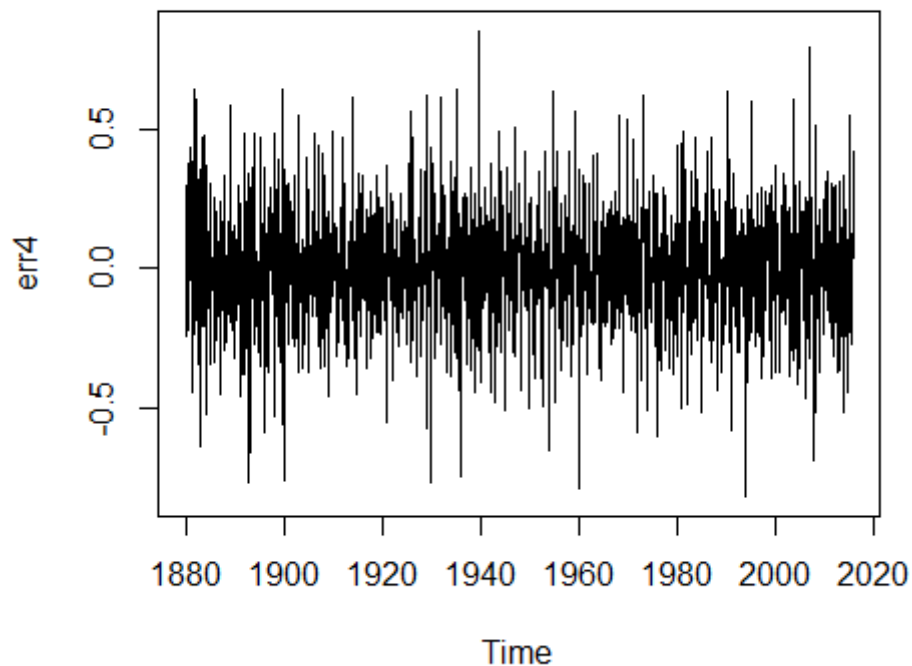


图 22: 全球平均温度的基本分解模型预测图

```
##
## Residual standard error: 0.1593 on 417 degrees of freedom
## Multiple R-squared:  0.7469, Adjusted R-squared:  0.7457
## F-statistic: 615.2 on 2 and 417 DF,  p-value: < 2.2e-16
```

### 5.2.2 周期项的估计

利用过去去除趋势项的数据中相同月份的均值作周期项的估计，从而得到下表中的数据：

表 6: 全球平均温度月周期项估计表

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
-5.8728	-5.3193	-3.2255	-0.1679	2.6168	4.7517	5.6001	5.1424	3.334	0.6607	-2.5924	-4.924

### 5.2.3 随机项的拟合

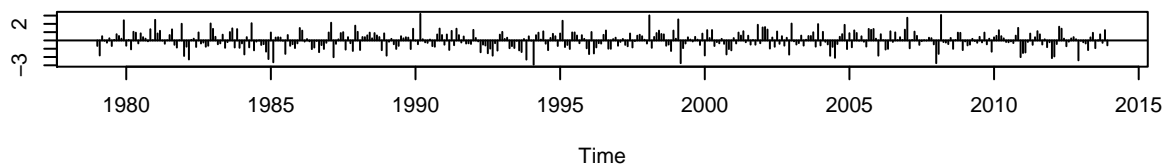
随机项时序图见图 22

```
##
## Augmented Dickey-Fuller Test
##
## data: err4
## Dickey-Fuller = -5.3769, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

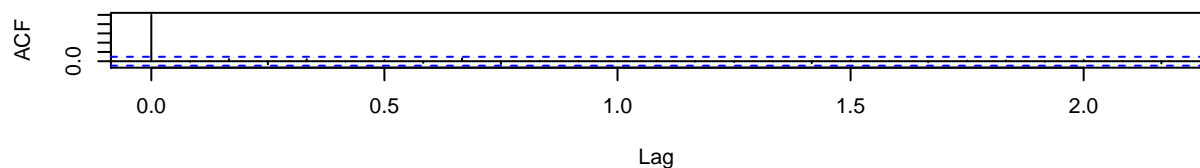
拒绝原假设，可以认为随机项序列平稳。

```
## Series: err4
## ARIMA(1,0,1)(0,0,1)[12] with zero mean
##
## Coefficients:
##          ar1      ma1      sma1
##      0.8125 -0.5529 -0.0745
## s.e.  0.0588  0.0833  0.0478
##
## sigma^2 estimated as 0.08057: log likelihood=-65.68
## AIC=139.36   AICc=139.46   BIC=155.52
```

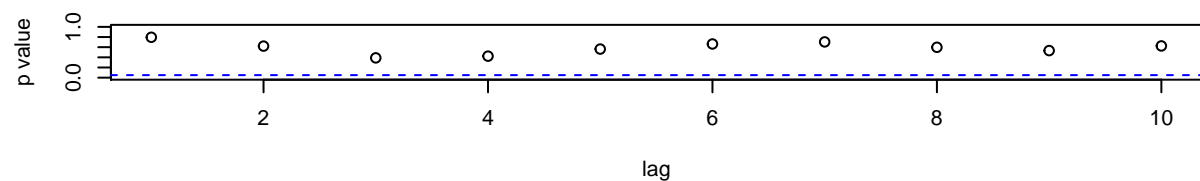
**Standardized Residuals**



**ACF of Residuals**



**p values for Ljung-Box statistic**



box-test result						
lag	1	2	3	4	5	6
Q.value	2.55E-05	0.306693	0.878821	0.905647	1.231712	1.301804
p.value	0.995973	0.857832	0.830536	0.923749	0.941799	0.9715581
lag	7	8	9	10	11	12
Q.value	1.32968	1.991262	2.862802	2.918849	2.933639	2.943293
p.value	0.987628	0.981279	0.969487	0.98327	0.991574	0.9959313
arch-test result						
lag	1	2	3	4	5	6
LM.value	0.022483	0.810835	1.219226	2.375976	3.293787	4.589064
p.value	0.880809	0.666698	0.748397	0.666973	0.654793	0.5974894
lag	7	8	9	10	11	12
LM.value	5.541574	6.335175	7.026628	7.885481	8.262051	12.880235
p.value	0.594174	0.609742	0.634347	0.640022	0.689655	0.3778055

图 23: box-test 和 arch-test 的检验结果

残差检验中第二张图的 ACF.test 显示残差没有明显的自相关。而 Ljung-Box test 显示所有 P 值都大于 0.1,

下面对数据进行 box-test 和 arch-test 的检验, 结果如图 23. 可见不存在 arch 效应, 利用 arma 模型进行拟合, 残差项通过了白噪声和 arch 检验, 可以进行应用。

#### 5.2.4 模型的预测

利用以上模型可以预测得到图 24, 可以计算得其 MSE 为 0.09610202

### 5.3 预测结果对比与模型评论

	model1	model2
MSE	0.10922	0.09606

其中 model1 分别表示温度的基本分解模型和 LOESS 分解模型。由 MSE 可见 LOESS 分解模型要优于基本分解模型。

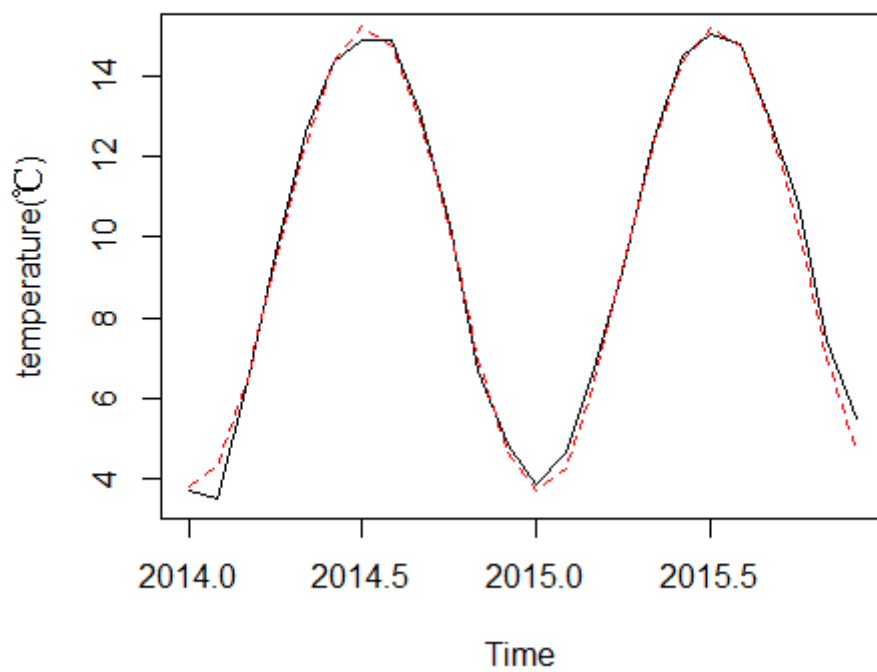


图 24: 全球平均温度的 LOESS 分解模型预测图 (实线真实值, 虚线预测值)

## 6 多元数据的 VAR 模型构建与预测

下面将南北极海冰面积和全球平均温度数据综合为一个多元时间序列，构建  $P$  阶的向量自回归模型 (VAR) 为：

$$y_t = \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + u_t$$

### 6.1 模型滞后阶数的确定

使用 `vars` 包中的 `VARselect` 函数对全球温度，南北极海冰面积的多元时间序列定阶。三个信息准则的阶数如下表 8：

表 8: VAR 信息准则

AIC(n)	HQ(n)	SC(n)	FPE(n)
2	2	2	2

根据信息准则，判定 VAR 模型的滞后阶为 2。

### 6.2 拟合模型

使用 `vars` 包中的 `VAR` 函数对数据进行拟合，其拟合结果如下

```
##
## VAR Estimation Results:
## =====
##
## Estimated coefficients for equation tem:
## =====
## Call:
## tem = tem.l1 + sisouth.l1 + sinorth.l1 + tem.l2 + sisouth.l2 + sinorth.l2 + const + trend + sd1
##
##      tem.l1    sisouth.l1    sinorth.l1      tem.l2    sisouth.l2
## 0.262609945 -0.073322995 -0.087578916 0.208817752 0.028782416
##   sinorth.l2      const      trend      sd1      sd2
## 0.071871549 5.332151503 0.001155846 0.029294264 1.065759761
##      sd3      sd4      sd5      sd6      sd7
## 3.178116092 5.645708274 7.294478706 8.151321779 7.962351180
##      sd8      sd9      sd10      sd11
## 6.841125184 4.999526527 2.997955934 0.999317037
```

```
##
##
## Estimated coefficients for equation sisouth:
## =====
## Call:
## sisouth = tem.l1 + sisouth.l1 + sinorth.l1 + tem.l2 + sisouth.l2 + sinorth.l2 + const + trend +
##
##      tem.l1      sisouth.l1      sinorth.l1      tem.l2      sisouth.l2
## 0.0200941790 0.7711196002 0.1365059304 -0.0620160118 -0.1139955065
##      sinorth.l2      const      trend      sd1      sd2
## -0.1421790098 4.2990272280 0.0006434985 -1.5485813726 0.0444887163
##      sd3      sd4      sd5      sd6      sd7
## 1.8920259192 3.8422589858 5.3078592257 6.4863744513 7.1492419593
##      sd8      sd9      sd10      sd11
## 7.4865336645 7.2309777883 6.2297708094 3.9586735823
##
##
## Estimated coefficients for equation sinorth:
## =====
## Call:
## sinorth = tem.l1 + sisouth.l1 + sinorth.l1 + tem.l2 + sisouth.l2 + sinorth.l2 + const + trend +
##
##      tem.l1      sisouth.l1      sinorth.l1      tem.l2      sisouth.l2
## -0.117970641 0.037985628 0.988533647 -0.037499571 -0.030377730
##      sinorth.l2      const      trend      sd1      sd2
## -0.316082772 5.386723779 -0.001146635 -0.134213206 -0.234221108
##      sd3      sd4      sd5      sd6      sd7
## -0.529187074 -0.951567577 -1.242240805 -1.222757311 -2.150362671
##      sd8      sd9      sd10      sd11
## -2.420578492 -1.701210415 0.085532543 -0.012069682
```

### 6.3 模型的诊断性检验

使用 vars 包中的 stability 函数进行稳定性检验，参数选择“OLS-CUSUM”（残差累积和），在该检验生成的曲线图（见图 25）中，残差累积和曲线以时间为横坐标，图中绘出两条临界线，累积和未超出这两条临界线，则说明参数具有稳定性。

Granger 因果检验结果图如图 26。

从 Granger 检验结果来看，全球平均温度是南北极海冰变化的 Granger 原因，北极海冰是全球平均温度和南极海冰变化的 Granger 原因。由此我们也可以证实文章初的论点，即：全球平均温度的变化会直接作



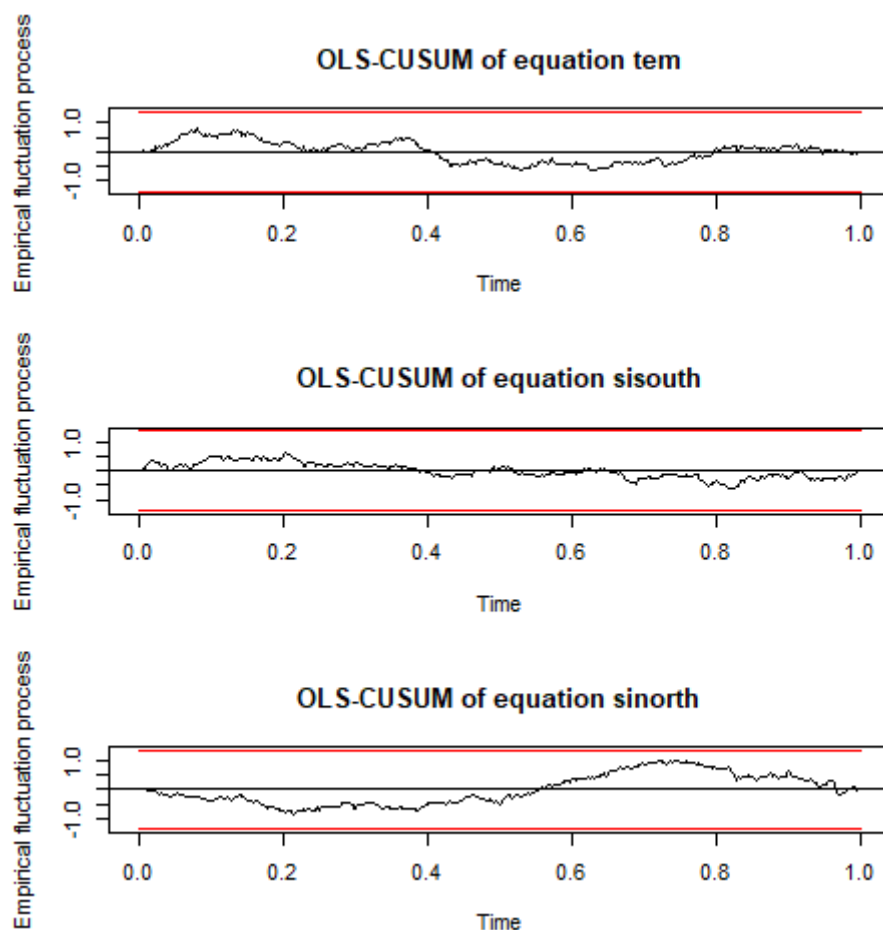


图 25: 模型稳定性检验图

原假设 <sup>↗</sup>	P 值 <sup>↗</sup>	结论 <sup>↗</sup>
全球温度不是南北极海冰的 Granger 原因 <sup>↗</sup>	0.008107 <sup>↗</sup>	在 0.01 水平下拒绝原假设 <sup>↗</sup>
南极海冰不是全球温度和北极海冰的 Granger 原因 <sup>↗</sup>	0.367 <sup>↗</sup>	不能拒绝原假设 <sup>↗</sup>
北极海冰不是全球温度和南极海冰的 Granger 原因 <sup>↗</sup>	0.08265 <sup>↗</sup>	在 0.1 水平下拒绝原假设 <sup>↗</sup>
南北极海冰不是全球温度的 Granger 原因 <sup>↗</sup>	0.2483 <sup>↗</sup>	不能拒绝原假设 <sup>↗</sup>

图 26: Granger 因果检验结果图

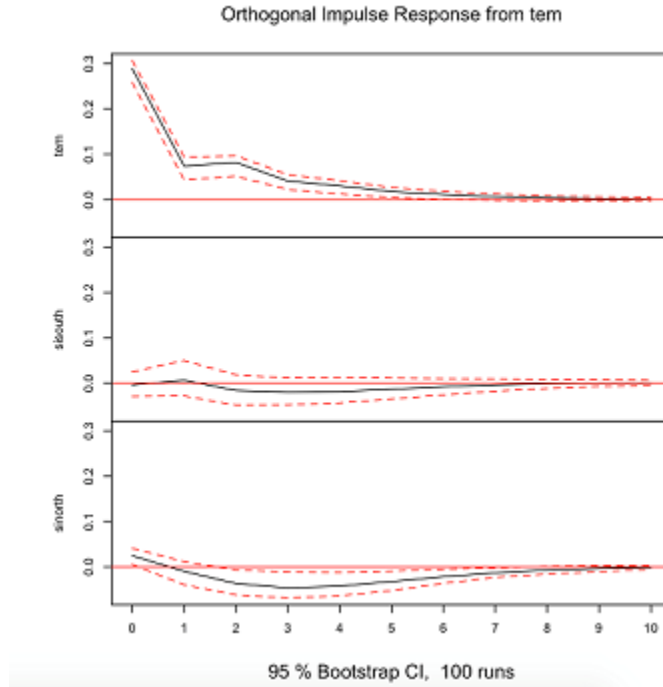


图 27: 全球平均温度对南北极海冰面积的冲击

用于海冰大小，北极海冰在气候调节中的作用要比南极海冰更大，南极海冰不太受其余两者影响。

## 6.4 脉冲响应分析

### 6.4.1 来自全球平均温度的脉冲反应

全球平均温度对其他变量的冲击图见图 27。可以看出，全球温度对南北极海冰面积都有不同程度的影响。其中南极海冰收到的影响在第二期开始加大，并在第四期达到最大后衰减趋于平稳。而北极海冰面积受的影响一开始为正，后快速变为负，并在第三期达到最大，而后开始衰减趋于平稳。所以，全球平均温度对海冰的面积以负向影响为主并慢慢趋于 0。

### 6.4.2 来自南极海冰的脉冲反应

南极海冰面积对其他变量的冲击图见图 28。由图可见，南极海冰面积对其余两个变量的影响较小，对全球平均温度有些许负向作用，但随后快速衰减为 0。

### 6.4.3 来自北极海冰的脉冲反应

北极海冰面积对其他变量的冲击图见图 29。可以看到，北极海冰面积对全球平均温度和南极海冰面积都有影响。其中对全球平均温度以负向影响为主，但是很快趋于 0。而对南极海冰面积一开始是正向影响，且影响比温度大，后转为负向影响并衰减为 0。

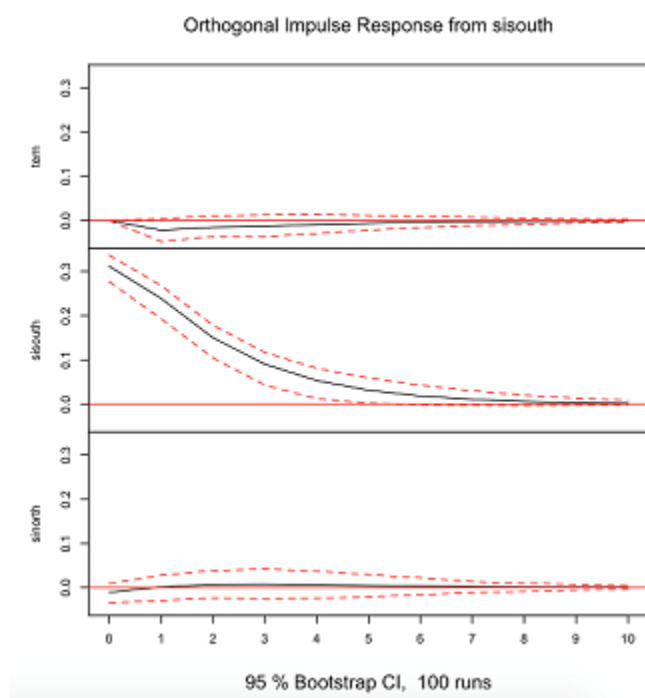


图 28: 南极海冰对其它变量的冲击

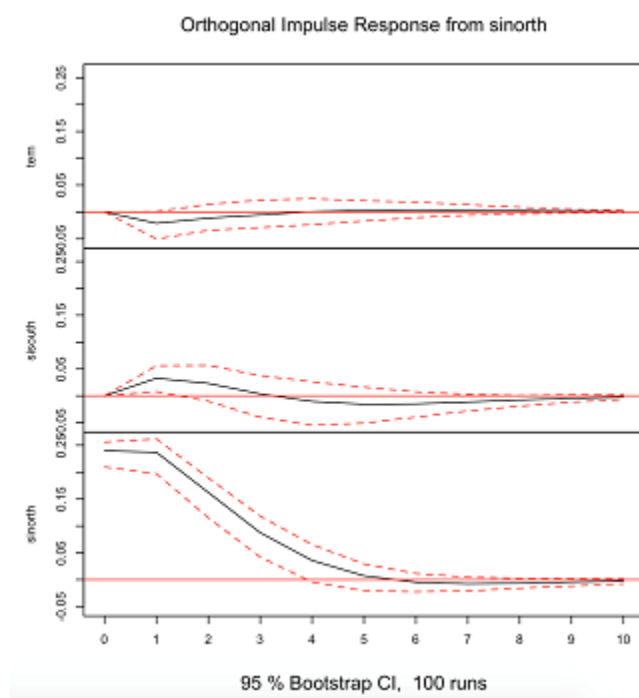


图 29: 北极海冰对其它变量的冲击

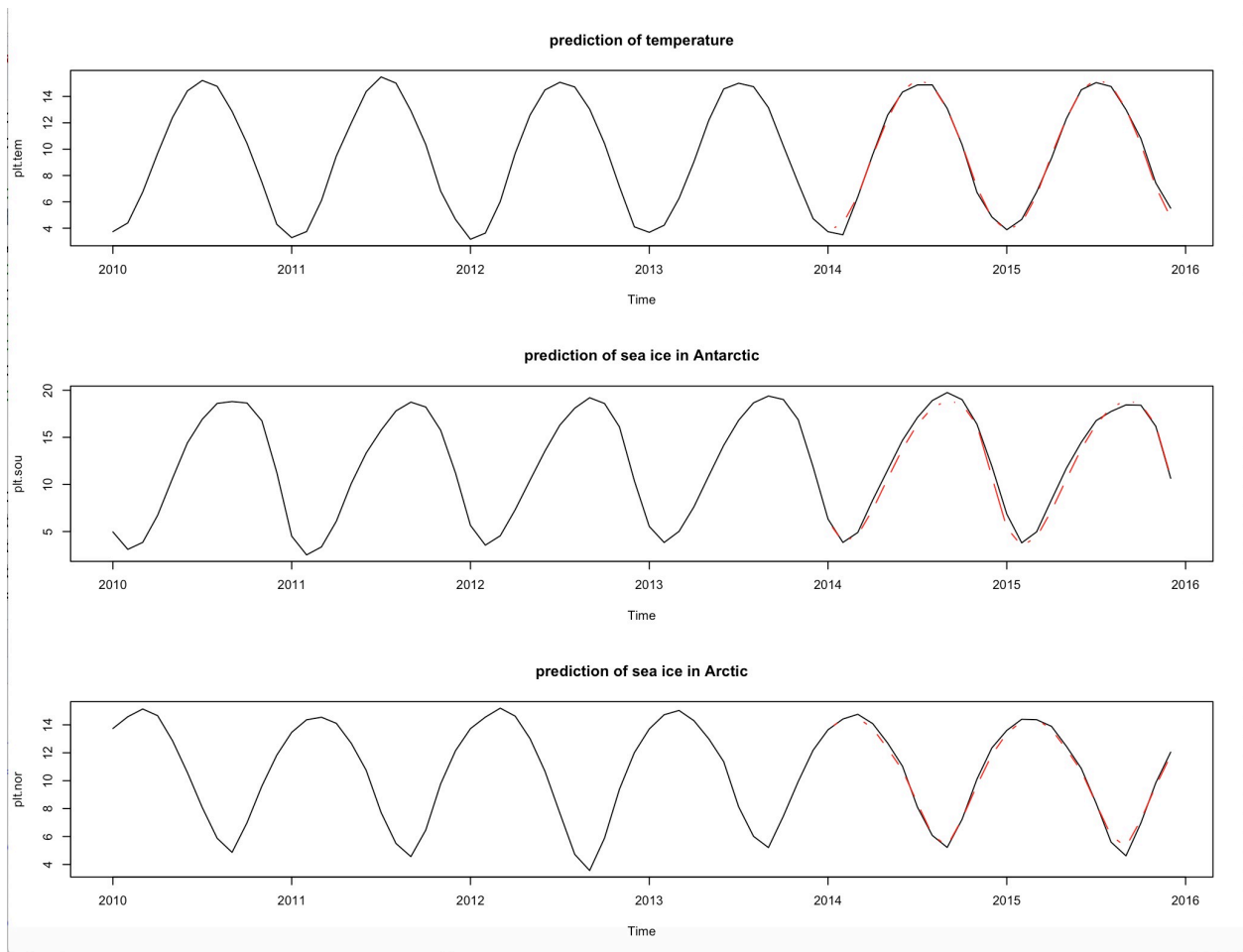


图 30: var 模型对全球平均温度，南北极海冰面积的预测

#### 6.4.4 脉冲响应结果分析

全球平均温度影响南北极海冰面积，且脉冲持续效应时间较长，负影响表示海冰面积有缩小的趋势。北极海冰大小又反过来影响全球平均温度，其减少对温度上升有促进作用，并且由于环流等因素影响也会影响南极海冰的生成和融化，且影响很大与全球温度相当。而南极海冰对温度和北极海冰的影响都很小。

### 6.5 预测与评估

接下来使用 VAR 模型预测接下来 24 期（即 2013 年到 2015 年两年）的数值。预测图见图 30。

可以看出大致与原数据趋势相同，我们计算 MSE 得图 31。

全球温度	南极海冰	北极海冰
0.09225231	0.5199779	0.08172167

图 31: var 模型预测的 MSE

## 7 总结

全球温度用 VAR 模型预测的 MSE 为 0.09225231，略小于用 LOESS 分解模型预测的 MSE 0.09606。南极海冰面积用 VAR 模型预测的 MSE 为 0.5199779，大于海冰面积的基本分解模型与 LOESS 分解模型。北极海冰面积用 VAR 模型预测的 MSE 为 0.08172167，小于 SARIMA 模型的 MSE，但大于基本分解模型与 LOESS 分解模型的 MSE。我们猜想所用信息越多，预测就会更加准确，但 VAR 模型并没有体现出很大的优势。这可能是因为我们在用季节性 VAR 模型拟合时，对数据的残差项  $u_t$  并没有进行合理建模，而在对数据单独建立模型时，我们对时间序列的分解更加彻底，拟合得更加充分。也可能是因为一些变量，比如南极海冰面积，并没有表现出与其它变量很强的相关性，从而预测效果不好。

我们从 Granger 因果性的角度，初步验证了全球平均温度，南北极海冰面积之间的相关关系——全球平均温度是南北极海冰变化的 Granger 原因，北极海冰是全球平均温度和南极海冰变化的 Granger 原因。由此我们也可以证实文章初的论点，即：全球平均温度的变化会直接作用于海冰大小，北极海冰在气候调节中的作用要比南极海冰更大，南极海冰不太受其余两者影响。