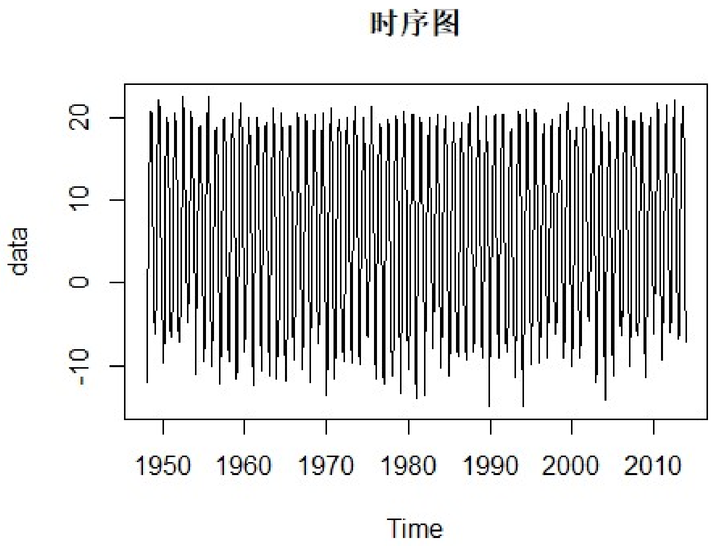
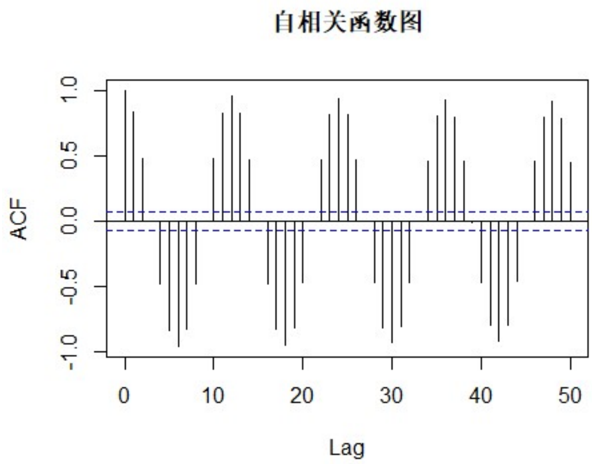
```r
library(dplyr)
library(uroot)
library(forecast)
library(purrr)
library(modelsummary)
library(ggplot2)

data <- read.table("D:/预删除文件夹/大三下/时间序列/月度温度数据.txt") %>%
  select(-1) %>%
  t() %>%
  as.vector() %>%
  ts(start = c(1948,01),frequency = 12)
plot(data,main="时序图")
```




```r
acf(as.vector(data),lag.max = 50,main="自相关函数图")
```

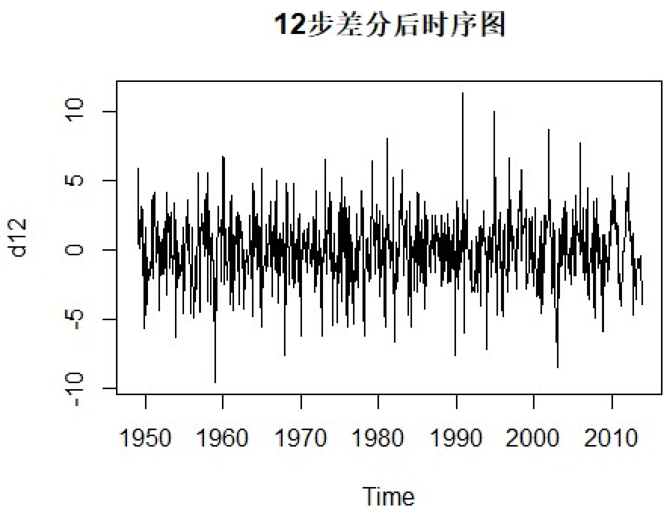


```r
hegy.test(data,c(1,0,0))

##
##  HEGY test for unit roots
##
## data:  data
##
##          statistic p-value
## t_1        -5.3086       0 ***
## t_2        -7.2144       0 ***
## F_3:4       0.2466  0.7953
## F_5:6      41.0435       0 ***
## F_7:8      62.0727       0 ***
## F_9:10     52.1435       0 ***
## F_11:12    55.0785       0 ***
## F_2:12     45.5517       0 ***
## F_1:12      43.936       0 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Deterministic terms: constant
## Lag selection criterion and order: fixed, 0
## P-values: based on response surface regressions

hegy.test(data,c(1,0,1))

##
##  HEGY test for unit roots
##
## data:  data
##
##          statistic p-value
## t_1        -6.7687       0 ***
## t_2        -7.8149       0 ***
## F_3:4      60.3168       0 ***
## F_5:6      64.5678       0 ***
## F_7:8      72.0734       0 ***
## F_9:10     63.7933       0 ***
## F_11:12    64.7878       0 ***
## F_2:12     71.4513       0 ***
## F_1:12     68.1468       0 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Deterministic terms: constant + seasonal dummies
## Lag selection criterion and order: fixed, 0
## P-values: based on response surface regressions
```

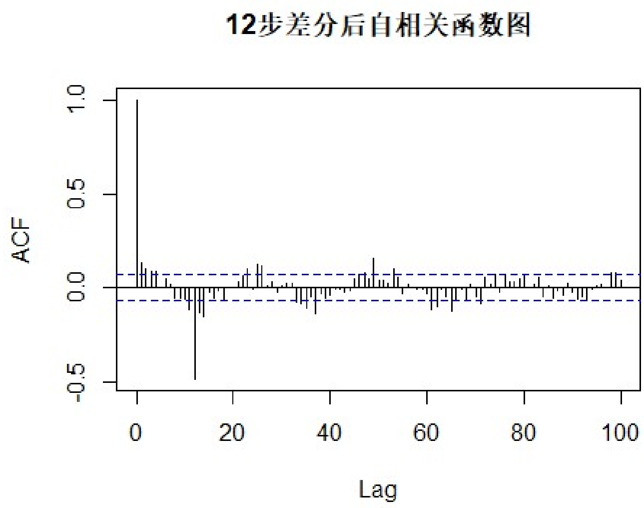
从时序图中可以看到没有明显的趋势，但存在周期波动特征，结合自相关函数图来看，大致可以确定周期为 12，若假定有确定的季节趋势，HEGY 检验结果显示不

存在单位根，若假定无没有确定的季节趋势，HEGY 检验结果显示存在单位根，由此可以作两种尝试，一是考虑提取确定性季节趋势，二是不提取，下面首先考虑不提取季节特征

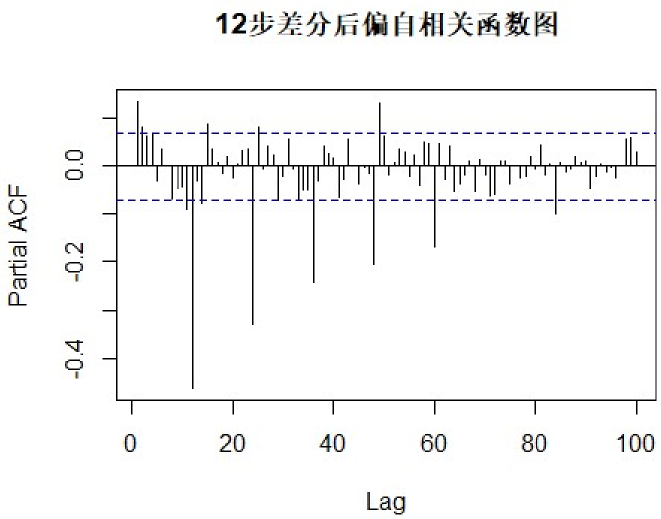
```
d12 <- diff(data,lag = 12)
plot(d12,main="12 步差分后时序图")
```




```
acf(as.vector(d12),lag.max = 100,main="12 步差分后自相关函数图")
```




```
acf(as.vector(d12),type = "partial",lag.max = 100,main="12 步差分后偏自相关函数图")
```




```
hegy.test(d12,c(1,0,0))

##
##  HEGY test for unit roots
##
## data:  d12
##
##          statistic p-value
## t_1       -14.8265       0 ***
## t_2       -12.3119       0 ***
## F_3:4     193.6718       0 ***
## F_5:6     168.4906       0 ***
## F_7:8     182.4369       0 ***
## F_9:10     162.681       0 ***
## F_11:12    161.442       0 ***
## F_2:12    199.6249       0 ***
## F_1:12    192.7838       0 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Deterministic terms: constant
## Lag selection criterion and order: fixed, 0
## P-values: based on response surface regressions
```

可以看到 12 步差分后，自相关函数在周期点上呈现截尾特征，偏自相关函数在周期点上呈现拖尾特征，据此确定 P=0,并分别尝试 Q=1,2,3，对于一个周期内来说，自相关函数呈现拖尾特征，偏自相关函数呈现截尾特征，据此确定 q=0,并分别尝试 p=2,3

```
model1 <- Arima(data,c(2,0,0),c(0,1,3))
model2 <- Arima(data,c(3,0,0),c(0,1,3))
model3 <- Arima(data,c(2,0,0),c(0,1,2))
```

```r
model4 <- Arima(data,c(3,0,0),c(0,1,2))
model5 <- Arima(data,c(2,0,0),c(0,1,1))
model6 <- Arima(data,c(3,0,0),c(0,1,1))
models <- list("SARIMA(2,0,0)(0,1,3)"=model1,
               "SARIMA(3,0,0)(0,1,3)"=model2,
               "SARIMA(2,0,0)(0,1,2)"=model3,
               "SARIMA(3,0,0)(0,1,2)"=model4,
               "SARIMA(2,0,0)(0,1,1)"=model5,
               "SARIMA(3,0,0)(0,1,1)"=model6)
modelsummary(models,stars = T,gof_map = "bic")
```

	SARIMA(2,0,0)(0,1,3)	SARIMA(3,0,0)(0,1,3)	SARIMA(2,0,0)(0,1,2)	SARIMA(3,0,0)(0,1,2)	SARIMA(2,0,0)(0,1,1)	SARIMA(3,0,0)(0,1,1)
ar1	0.147***	0.137***	0.139***	0.130***	0.138***	0.130***
	(0.037)	(0.037)	(0.036)	(0.036)	(0.036)	(0.036)
ar2	0.106**	0.092*	0.098**	0.086*	0.098**	0.085*
	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)
sma1	-0.955***	-0.961***	-0.957***	-0.965***	-0.969***	-0.973***
	(0.037)	(0.037)	(0.041)	(0.041)	(0.027)	(0.027)
sma2	-0.067	-0.059	-0.017	-0.011		
	(0.049)	(0.049)	(0.041)	(0.041)		
sma3	0.068+	0.062				
	(0.039)	(0.039)				
ar3		0.095**		0.097**		0.098**
		(0.036)		(0.036)		(0.036)
BIC	3219.5	3219.2	3215.8	3215.1	3209.3	3208.5

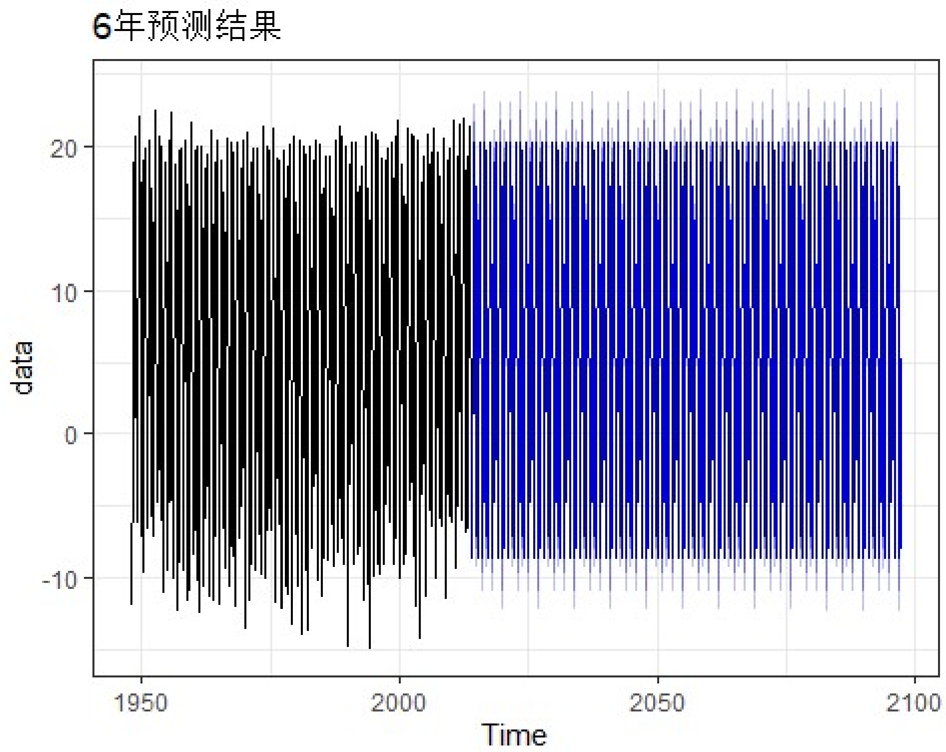
- p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```r
map(models,~.$residuals) %>%
  map(Box.test,lag=12)

## $`SARIMA(2,0,0)(0,1,3)`
##
##  Box-Pierce test
##
## data:  .x[[i]]
## X-squared = 13.401, df = 12, p-value = 0.3406
##
##
## $`SARIMA(3,0,0)(0,1,3)`
##
##  Box-Pierce test
##
## data:  .x[[i]]
## X-squared = 5.3528, df = 12, p-value = 0.9451
```

```
## 
## 
## $`SARIMA(2,0,0)(0,1,2)`
## 
##  Box-Pierce test
## 
## data:  .x[[i]]
## X-squared = 14.485, df = 12, p-value = 0.2708
## 
## 
## $`SARIMA(3,0,0)(0,1,2)`
## 
##  Box-Pierce test
## 
## data:  .x[[i]]
## X-squared = 5.9537, df = 12, p-value = 0.9184
## 
## 
## $`SARIMA(2,0,0)(0,1,1)`
## 
##  Box-Pierce test
## 
## data:  .x[[i]]
## X-squared = 14.745, df = 12, p-value = 0.2557
## 
## 
## $`SARIMA(3,0,0)(0,1,1)`
## 
##  Box-Pierce test
## 
## data:  .x[[i]]
## X-squared = 5.9795, df = 12, p-value = 0.9171
```

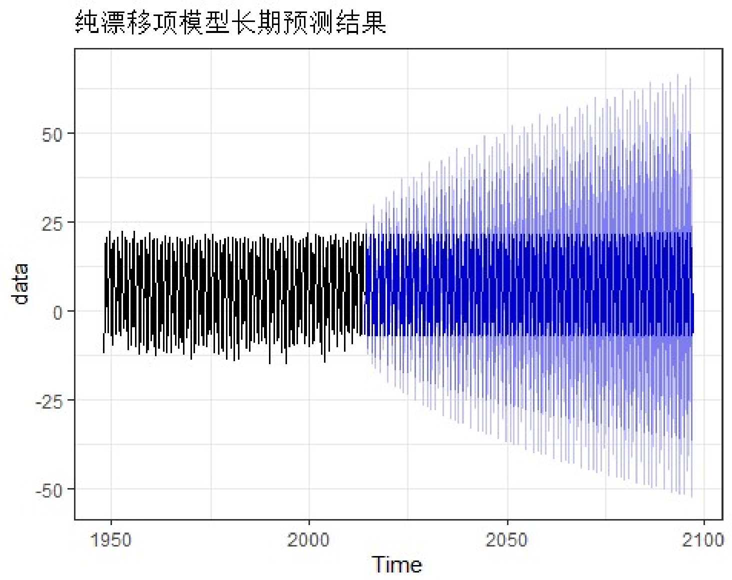
可以看到所有模型的残差序列都可以认为是纯随机的，可以认为这些模型都较好地提取了序列信息。还可以注意到其中一些模型的部分系数不显著，删去不显著的项之后 BIC 有所上升，综合来看，SARIMA(3,0,0)(0,1,1)模型各项系数显著，且 BIC 最小，残差纯随机性检验的 p 值也很大，故选定该模型进行拟合，拟合结果如下图所示

```
forecast(model6,1000) %>%
  autoplot() +
  labs(title = "6 年预测结果") +
  theme_bw()
```



下面考虑提取确定性季节特征，首先看仅含有漂移项的模型

```
model_drift <- Arima(data,c(0,0,0),c(0,1,0),include.constant = T)
model_drift %>%
  forecast(1000) %>%
  autoplot() +
  labs(title = "纯漂移项模型长期预测结果") +
  theme_bw()
```



```r
Box.test(model_drift$residuals,lag = 12)

##
##  Box-Pierce test
##
## data:  model_drift$residuals
## X-squared = 238.12, df = 12, p-value < 2.2e-16
```

可以看到从点估计来说，该模型也在一定程度拟合了温度的变化特征，但纯随机性检验显示该模型的残差序列极其显著地不随机，且显而易见的是该模型的预测方差会不断变大，这对预测来说也是不好的，下面在之前挑出的 SARIMA(3,0,0)(0,1,1)模型基础上引入漂移项

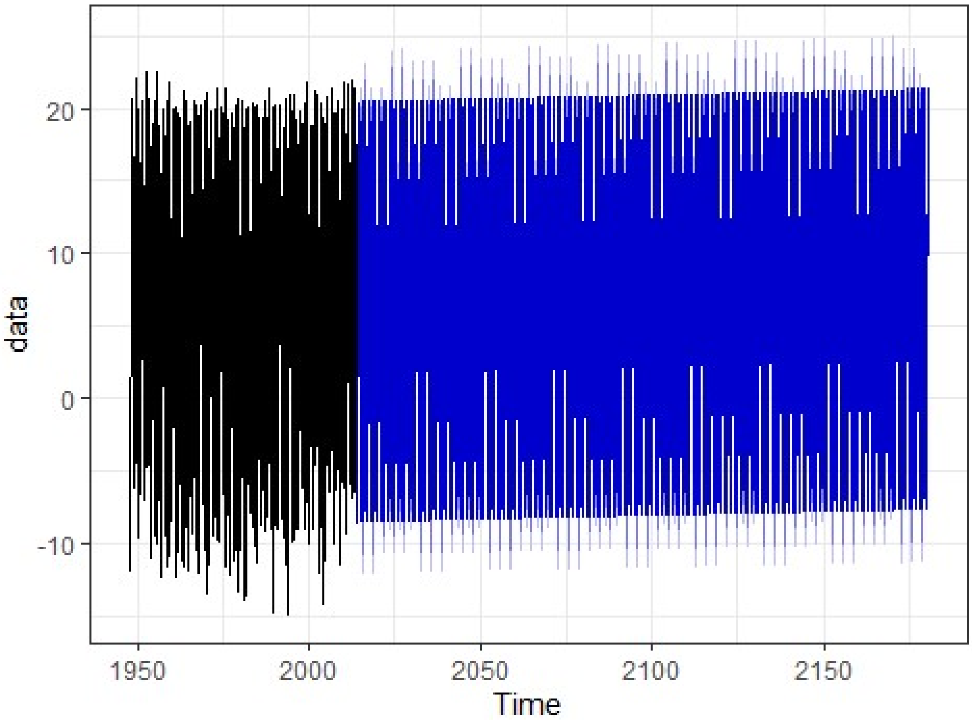
```r
model_drift2 <- Arima(data,c(3,0,0),c(0,1,1),include.drift = T)
Box.test(model_drift2$residuals)

##
##  Box-Pierce test
##
## data:  model_drift2$residuals
## X-squared = 0.09526, df = 1, p-value = 0.7576

modelsummary(list("non_drift"=model6,"drift"=model_drift2),stars = T,go
f_map = c("bic","rmse"),fmt = 4)
```

	non_drift	drift
ar1	0.1295***	0.1296***
	(0.0362)	(0.0361)
ar2	0.0852*	0.0848*
	(0.0361)	(0.0360)
ar3	0.0976**	0.0967**
	(0.0358)	(0.0358)
sma1	-0.9726***	-0.9799***
	(0.0274)	(0.0343)
drift		0.0004
		(0.0004)
BIC	3208.5	3214.2
RMSE	1.80	1.79

- p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```r
model_drift2 %>%
  forecast(2000) %>%
  autoplot() +
  labs(title = "含漂移项模型长期预测结果") +
  theme_bw()
```

含漂移项模型长期预测结果



可以看到模型的残差也可以认为是纯随机的，各项系数依旧显著且变化不大，但漂移项并不显著，且因为引入了漂移项，BIC 有所上升，不过 RMSE 有所下降。从预测结果来看，预测方差没有越发膨胀，总体而言预测值呈现上升趋势