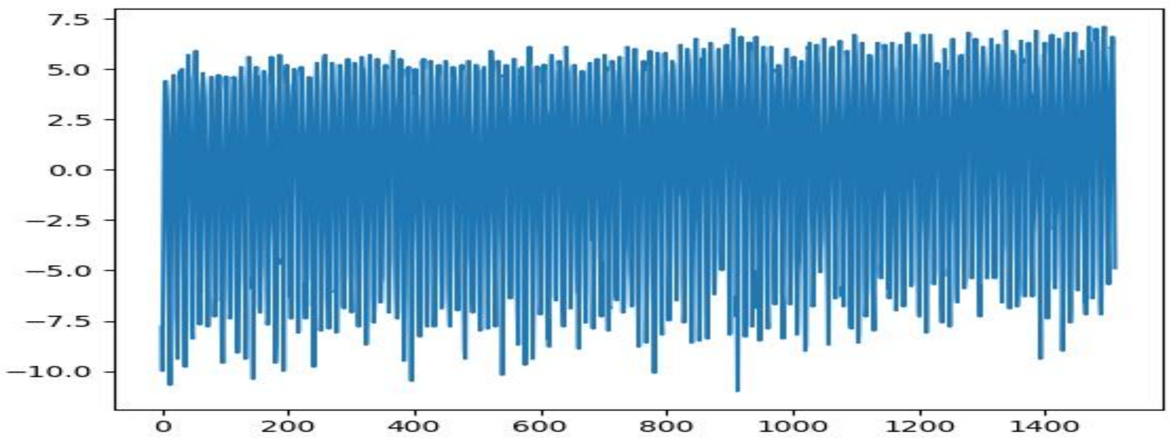
Hong Kong Temperature

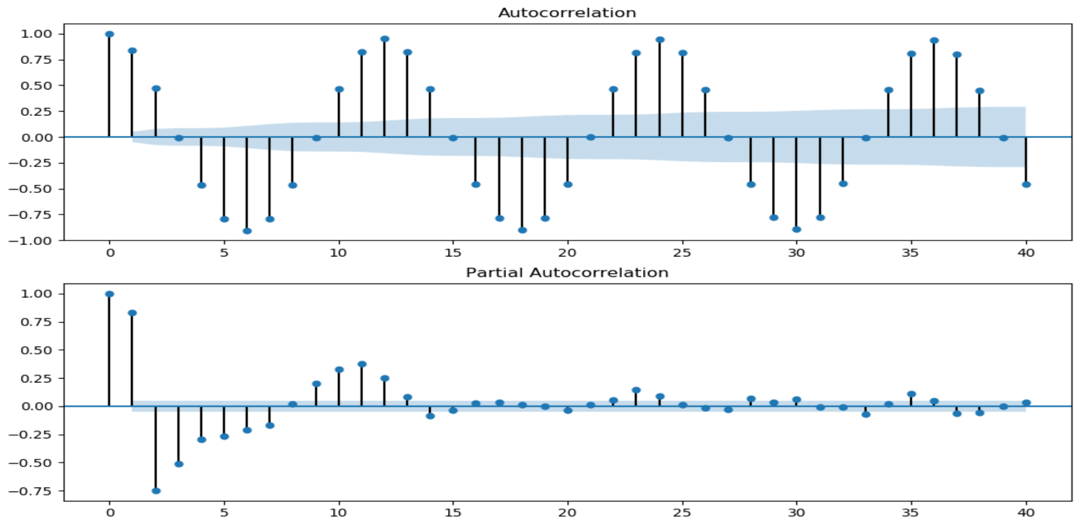
Weather has always been an important topic for human beings. Therefore, in this project, we are going to build time series model on monthly historical temperature data of Hong Kong and forecast the temperature of different months of 2018. The temperature data is downloaded from Hong Kong Observatory. The link is http://www.hko.gov.hk/cis/monthlyElement_uc.htm. We used the historical data from 1885 Jan to 2017 Dec and forecast temperature of different months of 2018.

The first step is to build time series model.

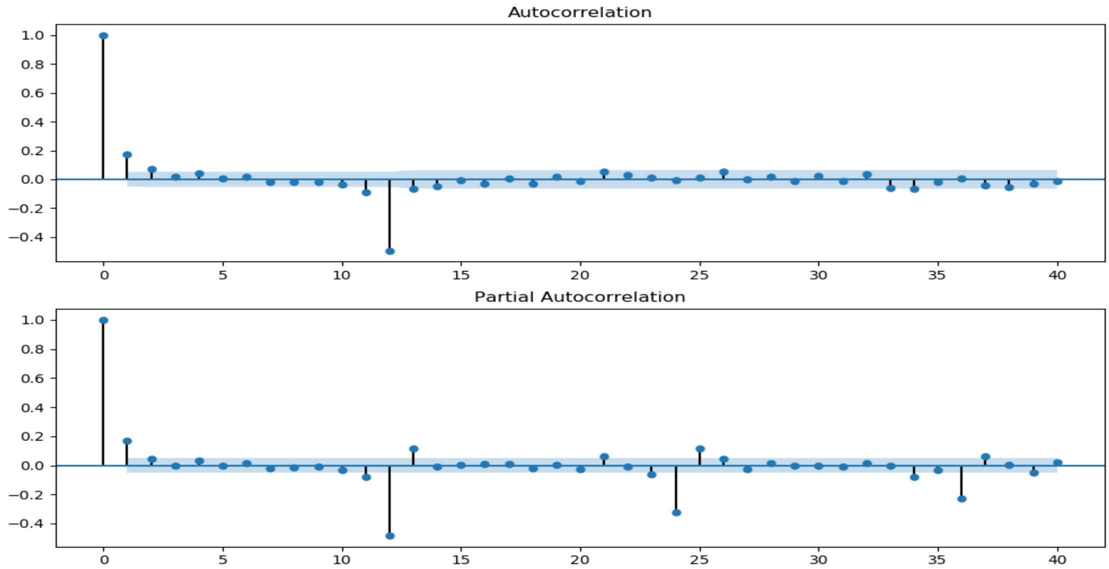
We first load the data, and then we subtract its mean from it. We plot it



We calculate its ACF and PACF



It is obvious the ACF shows seasonal pattern. Since we are using monthly data, we take lag=12, which is consistent with the ACF. After that, we plot the ACF again.



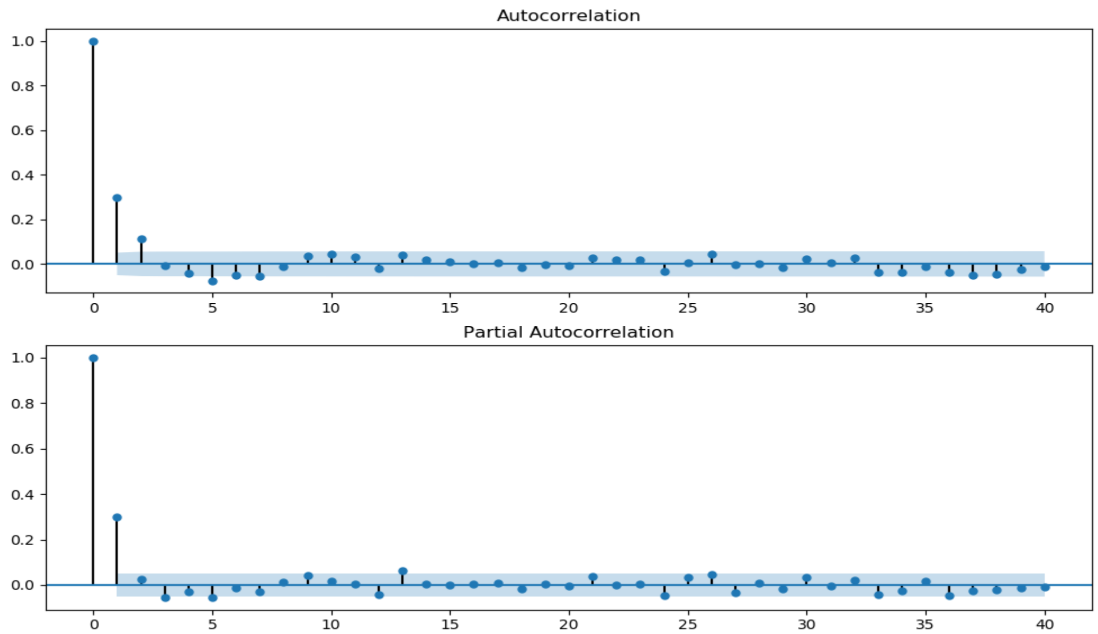
We see that the PACF shows seasonal exponential decay. The period is 12. Therefore, we suspect seasonal ARIMA(0,0,0)*(0,1,1,12). We fit the model and check the ACF of residuals. We got the following result.

Statespace Model Results

Dep. Variable:	y	No. Observations:	1512
Model:	SARIMAX(0, 1, 1, 12)	Log Likelihood	-2085.858
Date:	Tue, 08 May 2018	AIC	4175.716
Time:	21:51:41	BIC	4186.359
Sample:	0 - 1512	HQIC	4179.679
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ma.S.L12	-0.9102	0.010	-89.030	0.000	-0.930	-0.890
sigma2	0.9316	0.026	35.612	0.000	0.880	0.983

Ljung-Box (Q):	105.43	Jarque-Bera (JB):	146.47
Prob(Q):	0.00	Prob(JB):	0.00
Heteroskedasticity (H):	1.12	Skew:	-0.08
Prob(H) (two-sided):	0.20	Kurtosis:	4.52

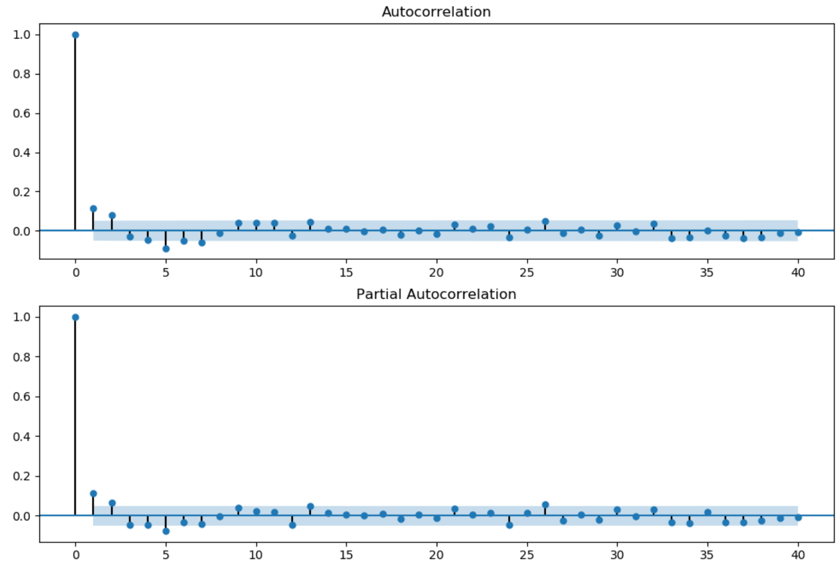


Note that the p-value of Ljung-Box test is 0. Namely, the residuals are not white noise. We should adjust our model.

Let's look back to our model.

The first ACF term is larger than 0.2, which is significant. Therefore, we suspect the data comes from (1,0,0)*(0,1,1,12). We fit the model and check the ACF of residuals. We got the following result.

```
                         Statespace Model Results
==========================================================================================
Dep. Variable:                                          y    No. Observations:        1512
Model:                SARIMAX(1, 0, 0)x(0, 1, 1, 12)    Log Likelihood           -2047.210
Date:                                 Tue, 08 May 2018    AIC                      4100.420
Time:                                         22:38:46    BIC                      4116.384
Sample:                                              0    HQIC                     4106.365
                                                - 1512
Covariance Type:                                   opg
==================================================================================
                coef      std err          z        P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         0.2278       0.023      10.108       0.000       0.184       0.272
ma.S.L12     -0.9248       0.010     -96.231       0.000      -0.944      -0.906
sigma2        0.8836       0.024      36.264       0.000       0.836       0.931
=======================================================================================
Ljung-Box (Q):                        40.04   Jarque-Bera (JB):              183.02
Prob(Q):                               0.47   Prob(JB):                        0.00
Heteroskedasticity (H):                1.09   Skew:                           -0.12
Prob(H) (two-sided):                   0.36   Kurtosis:                        4.69
```



Note that the p-value of Jyung-box test is 0.47, which is very satisfactory. Moreover, the p-values of all coefficients are all close to 0, which means these coefficients are significant. This is our final model.

$$(1 - 0.2278B)(1 - B^{12})(X_t - \mu) = (1 + 0.9248B^{12})a_t$$

Then we forecast temperature of different months of 2018.

The following is our forecast:

16.25943111 16.8828221 19.15698154 22.82703633 26.19575051 28.32051208 28.92710004 28.75853273 28.00080662 25.74414145 22.10149286 17.91269989

We also showed the actual temperature data from Jan to Apr 2018.

16.1

16.0

20.8

23.6

Indeed, we got very close estimate for the first 4 months.