Shuo Chen
2020012894
chenshuo20@mails.tsinghua.edu.cn

Homework Assignment 3
Environmental Simulation Practice

1 Introduction

This experiment provides a dataset of sewage pipe network length for each month from 1978 to 2010, with a total of 396 records. It is required to use appropriate time series tools to forecast the length of the urban sewage pipe network for a period of time in the future.

2 Smoothing of Time Series and Seasonal Decomposition
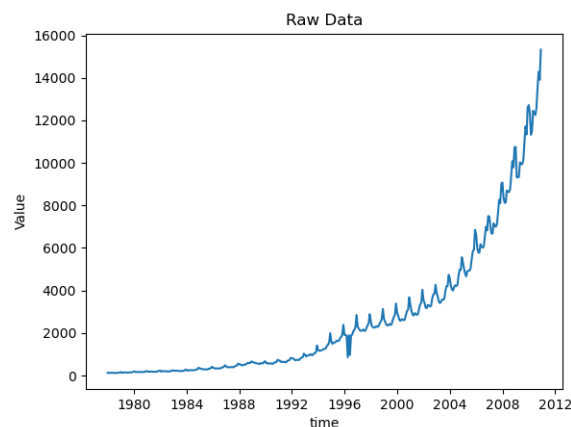
The raw data is shown below.

Abbildung 1: Raw Data

2.1 Smoothing of Time Series

We use lowess method to make the curve smooth. The details can be found in the code. And the result is
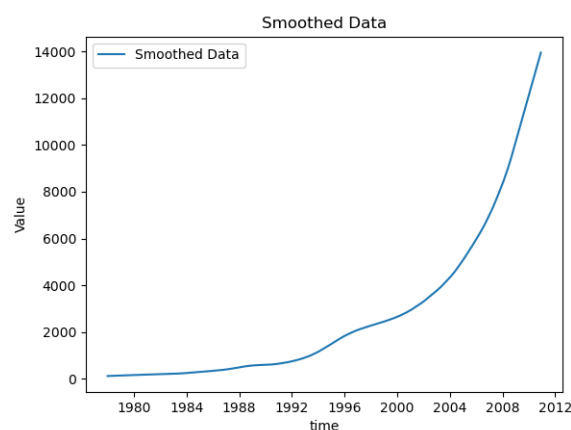

Abbildung 2: Caption

We find that the raw data almost obedient to exponential growth. So we take a logarithm for all values. And we use the ADF test. The original sequence is differenced once, that is, a first-order lag. ADF test is performed on the differenced sequence, and the test result shows that the sequence is stationary at this point.

	ADF	p-value
Raw Data	-3.612	0.006
Smoothed Data	-2.661	0.081

In this case, it means that we have conducted the Augmented Dickey-Fuller (ADF) test on a time series data set and the test result indicates that the null hypothesis of non-stationarity of the time series can be rejected at a significance level of 0.05, which means the time series is likely stationary.

2.2 Seasonal Decomposition

There are several reasons why we should do seasonal decomposition. Firstly, by decomposing the time series into its individual components, we can gain a better understanding of the underlying patterns and trends in the data, which can help to inform forecasting and decision-making. Secondly, the decomposition can help to identify the sources of variation in the time series, which can be useful for detecting anomalies or outliers in the data. Lastly, the decomposition can provide a basis for modeling and forecasting the time series using different techniques.
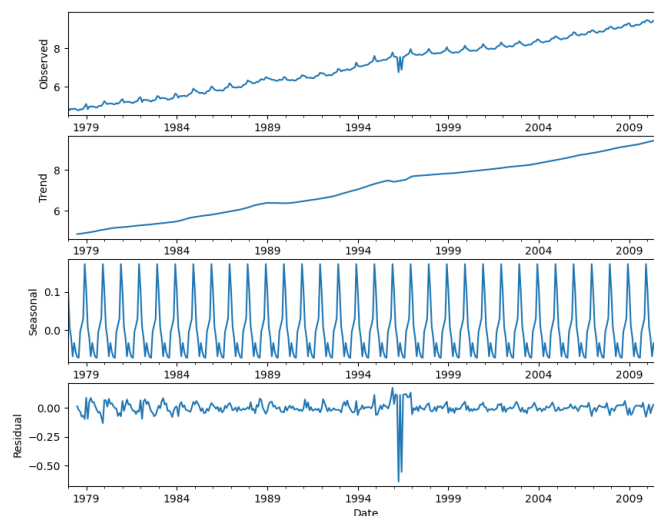
And we get the result:

Abbildung 3: Seasonal Decompose

3 Forcast by ARIMA Model

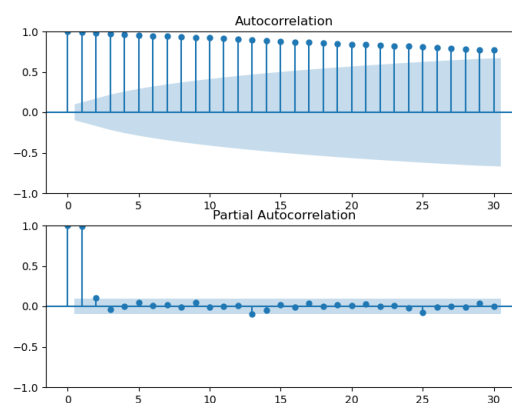We can use ACF and PACF plots to choose candidate models, and the plots are shown below:

Abbildung 4: ACF and PACF

And we choose the $(p, d, q) = (1, 1, 1)$ to forecast.
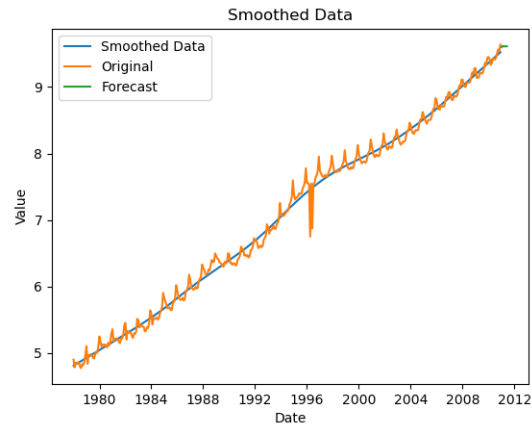And we get the result:

Abbildung 5: Forecast

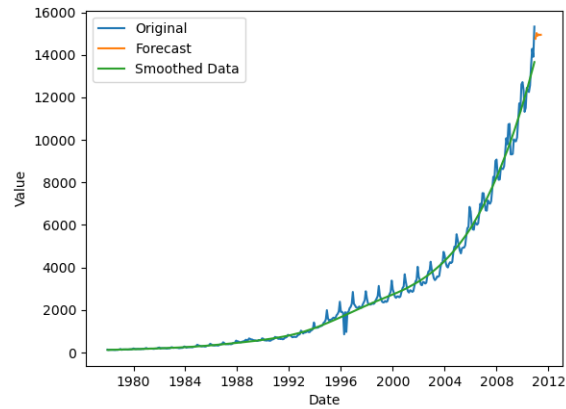| Time	Value
2011-01-01	14762.192644
2011-02-01	15026.554904
2011-03-01	14901.537499
2011-04-01	14960.251894
2011-05-01	14932.586828
2011-06-01	14945.602131

Abbildung 6: True Forecast Result