Monthly water usage (ml/day)

——London Ontario, 1966-1988

Author: Di Zhao

1. Introduction

Water is the basic resource for supporting economic growth and maintaining daily life. In the past few years, with the average increase in water consumption, people's concerns have also increased (Lu, 2007). This paper mainly studies the monthly water usage in London Ontario from Jan 1966 to Dec 1988, 23 years in all (Fig 1). The monthly water usage data are time series data with data size equals to 276. Time is measured in months and water usage is measured in “ml/day”.

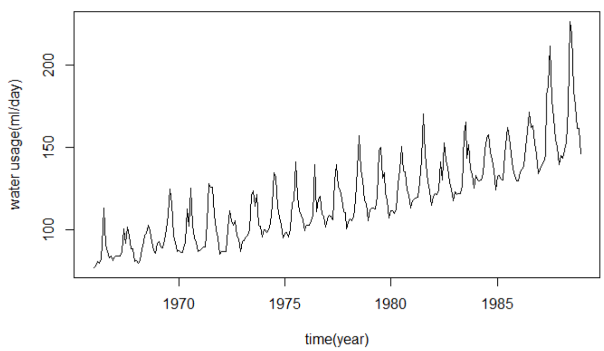


Fig 1 Time series plot of monthly water usage

2. Summary Statistics and Summary Figures

2.1 Trend

The trend of monthly water usage is increasing with time with some curvature. Thus, linear function, quadratic function or cubic function should be considered to fit the trend (Fig 2).

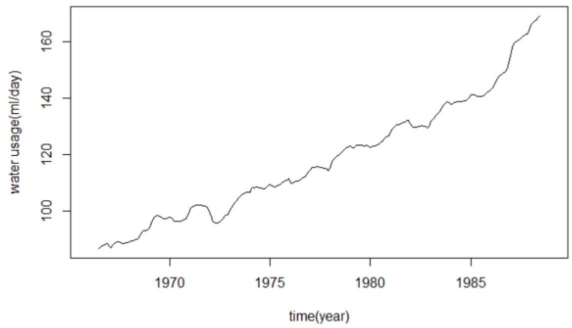


Fig 2 Trend of the monthly water usage

2.2 Variance

The variance also increases with time (Fig 1). Thus, the multiplicative model could be considered to fit the data. To simplify the time series model structure, the logarithmic transformation could scale the data with similar variance and the additive model is fitted (Fig 3).

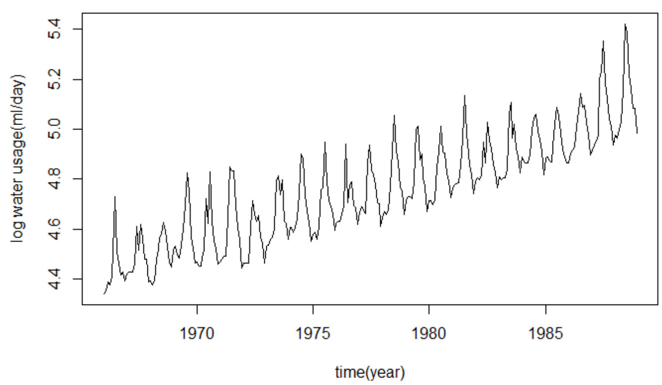
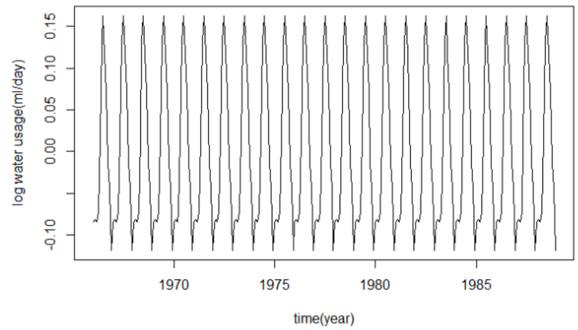


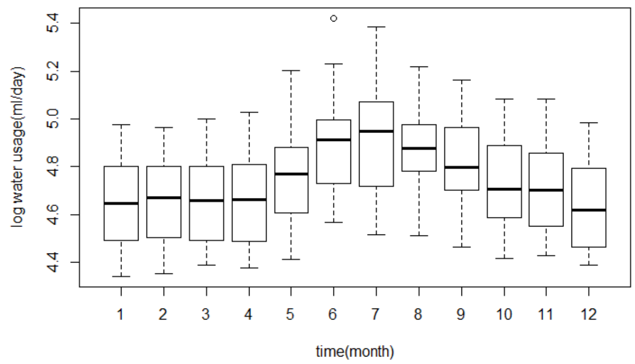
Fig 3 Logarithmic transformation of monthly water usage

2.3 Seasonal effect

A clear seasonal effect is in monthly water usage recurs every 12 months (Fig 4 (a)). Monthly water usage in summer is higher than that in other seasons (Fig 4 (b)).



(a)



(b)

Fig 4 (a) Plot of decomposed seasonal effect of logarithmic monthly water usage; (b) Box plot of seasonal effect of logarithmic monthly water

3. Analysis

3.1 Trend model

To find the best trend model, I use the Akaike information criterion (AIC) to test three possible trend models: linear model, quadratic model, and cubic model. The summary table shows the best trend model is the cubic model (Table 1) because it has the smallest AIC.

Table 1 AIC table for three trend models

	DF	AIC
Linear Model	3	-442.2131
Quadratic Model	4	-443.8495
Cubic Model	5	-445.2553

3.2 Additive model with seasonal indicators

$$m_t = \alpha_1 \times t + \alpha_2 \times t^2 + \ \alpha_3 \times t^3$$

$$T = year + \frac{month - 1}{12}$$

$$t = (T - 1977.458)/6.651545$$

$$x_t = \exp\,(m_t + \beta_{month} + z_t).$$

Where T is time, t is standard time, m_t is the trend at time t, α_i ($i \in \{1,2,3\}$) is the coefficient of trend model, x_t is the estimated monthly data usage, β_{month} is the seasonal indicator measured in months, z_t is residuals.

$$x_t = \exp\,(0.16004 \times t + 0.01395 \times t^2 + \ 0.01362 \times t^3 \ + \ \beta_{month} + z_t)$$

$$\beta_{month} = \begin{cases} 4.65395 & (month = 1) \\ 4.65727 & (month = 2) \\ 4.65561 & (month = 3) \\ 4.66630 & (month = 4) \\ 4.75234 & (\mathrm{month} = 5) \\ 4.87162 & (month = 6) \\ 4.90609 & (month = 7) \\ 4.85500 & (\mathrm{month} = 8) \\ 4.80903 & (month = 9) \\ 4.72391 & (month = 10) \\ 4.69384 & (month = 11) \\ 4.61914 & (month = 12) \end{cases}$$

3.3 Residual model

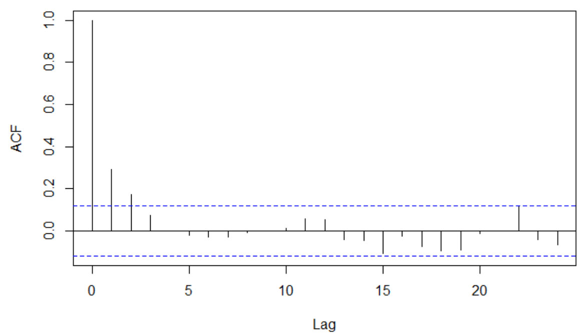
I use the Augmented Dickey-Fuller Test (adf test) to test if the time series is stationary. The p-value = 0.01 < 0.05, reject H_0, thus there is strong evidence that the time series is stationary.

```
(RStudio, Version 1.2.1335)

p-value smaller than printed p-value

Augmented Dickey-Fuller Test

data:  resid(fit.seas)
Dickey-Fuller = -5.9494, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

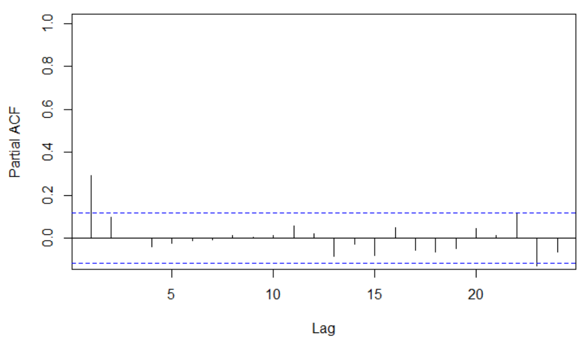
The autocorrelation factor plot (ACF plot) tells us the residual model should include MA (2), and partial autocorrelation factor plot (PACF) tells us the residual model should include AR (1) theoretically (Fig 5). However, to make sure I use the AIC to test best ARMA model, and I obtain ARMA (1,1) is the best ARMA model for residuals.

$$z_t = 0.5283 z_{t-1} + w_t - 0.2582 w_{t-1} - 0.0001$$

$$w_t \sim N(0, 0.002139885)$$



(a)



(b)

Fig 5 (a) ACF plot of residuals of the additive model with seasonal indicator; (b) PACF plot of residuals of the additive model with seasonal indicator monthly water usage.

3.4 Final model & examination

Final model:

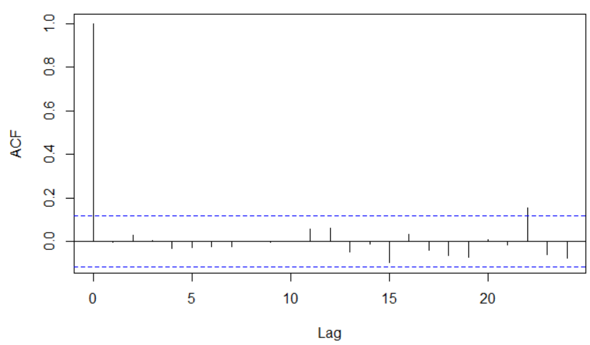
$$x_t = \exp\left(0.16004 \times t + 0.01395 \times t^2 + 0.01362 \times t^3 + \beta_{month} + z_t\right)$$

$$\beta_{month} = \begin{cases} 4.65395 & (month = 1) \\ 4.65727 & (month = 2) \\ 4.65561 & (month = 3) \\ 4.66630 & (month = 4) \\ 4.75234 & (\text{month} = 5) \\ 4.87162 & (month = 6) \\ 4.90609 & (month = 7) \\ 4.85500 & (\text{month} = 8) \\ 4.80903 & (month = 9) \\ 4.72391 & (month = 10) \\ 4.69384 & (month = 11) \\ 4.61914 & (month = 12) \end{cases}$$

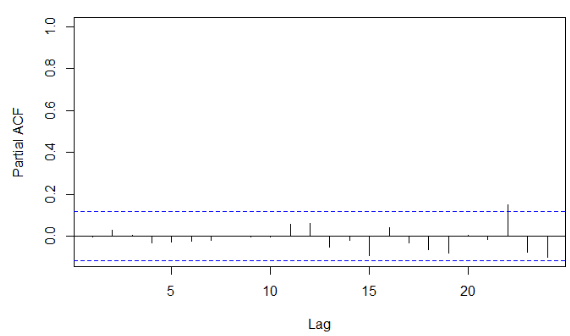
$$z_t = 0.5283 z_{t-1} + w_t - 0.2582 w_{t-1} - 0.0001$$

$$w_t \sim N(0, 0.002139885)$$

I use ACF plot and PACF plot to examination the final model fit (Fig 6 (a), Fig (b)). The residuals of the final model have no autocorrelation and seasonal pattern, thus the final model fits the data well.



(a)



(b)

Fig 6 (a) ACF plot of residuals of the final model; (b) PACF plot of residuals of the final model.

4 Conclusion

4.1 Summary

The model includes two parts: an additive model and residual model. The additive model includes trend (cubic function) and the seasonal effect (seasonal indicators). The residual model includes ARMA (1,1) model.

4.2 In-sample simulation

Final model fits in-sample data well (Fig 7).

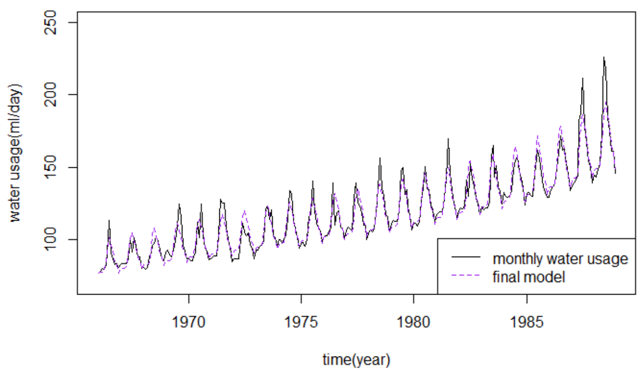


Fig 7 The superimposed plot of simulated data from final model and monthly water usage.

4.3 Future prediction

The final model predicts that, in the future five years, the water usage will keep increasing over the next five years with increasing variance as time goes on (Fig 8).

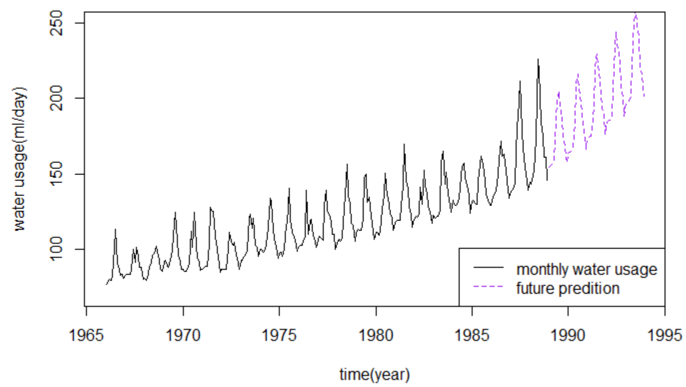
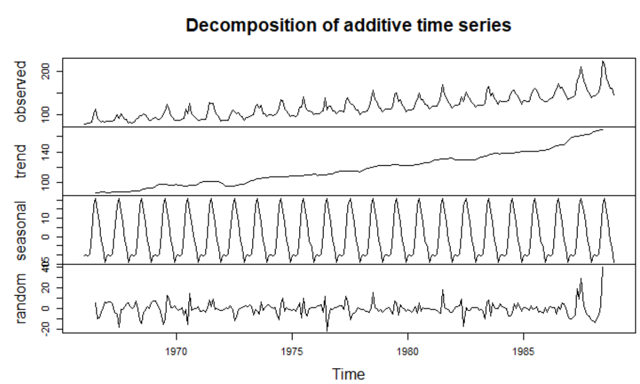


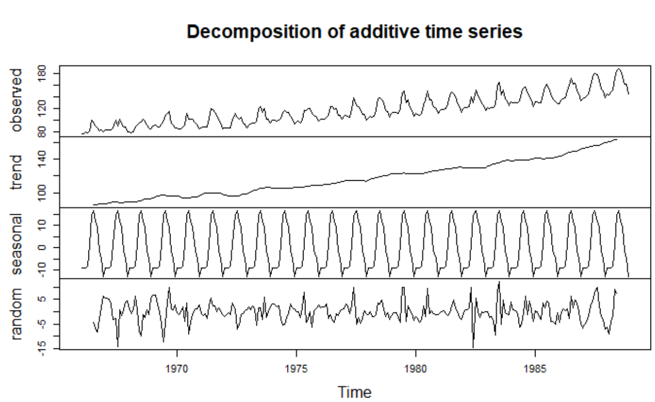
Fig 8 The superimposed plot of the future 5-year prediction data from final model and monthly water usage.

5 Concern

There might be extreme values in these time series data. After rescaling these values, the trend would be more smoothy (Fig 9).



(a)



(b)

Fig 9 (a) Original data decomposed plot; (b) Rescaled data decomposed plot

6 Possible cause in reality

6.1 Possible reason for trend increases with time

- More and more water energy is used in industries.
- People pay more attention to hygiene and take more shower than before.
- People are richer, and the charge of water is affordable.

6.2 Possible reason for Variance increases with time

- More and more air conditions or swimming pools are using in summer. Thus, water usage in summer is gradually greater than that in the rest of the seasons, along with the time.
- Industries are developing every day. Water circulation systems are more used to cool down machines especially in summer.
- Because of global warming, more water is tended to evaporate. When summer comes, farmers use more water to irrigate the fields, and people use more water to water the garden.

Reference

Lu, T. (2007), Research of domestic water consumption: a field study in Harbin, China

Time Series Data Library (citing: Hipel and McLeod (1994)), Monthly water usage (ml/day), London Ontario, 1966-1988. Retrieved from: https://datamarket.com/data/set/22qu/monthly-water-usage-mlday-london-ontario-1966-1988#!ds=22qu&display=table