Time Series Project

Novin Shahroudi

Overview

First Time Series (WL45)

- ARIMA/SARIMA models
- Holt-winter Family
- Conclusion

Second Time Series (TS093s)

- SARIMA models
- ARIMAX with regressor (TS1422s)
- Conclusion

Time Series 1: WL45

General information:

- United States Geology Survey
- Surface-Water Monthly Statistics
- Mean value of surface water
- 993 observations
- From Jan 1930 to Sep 2012

Further details:

- Seems stationary
- It may contain some seasonality

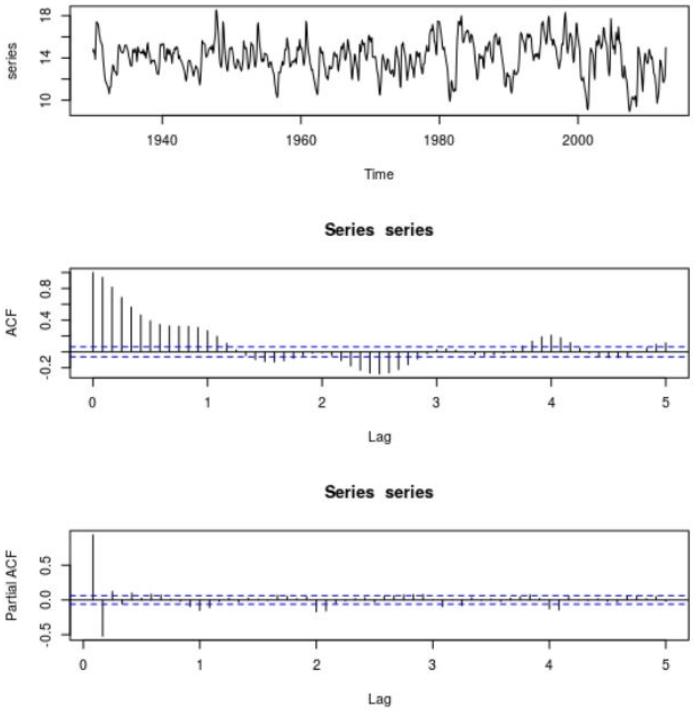


Figure 1: WL45 alongside the ACF and PACF plots

WL45 Stationarity Test

```
Dickey-Fuller = -6.841, Truncation lag parameter = 7, p-value = 0.01
```

Listing 1: Phillips-Perron Unit Root Test

```
Dickey-Fuller = -5.7226, Lag order = 9, p-value = 0.01
alternative hypothesis: stationary
```

Listing 2: Augmented Dickey-Fuller Test

ARIMA models

Based on ACF and PACF plots of the series:

- AR(3) from PACF
- Potential MA terms

Found model:

- ARIMA(3,0,3)
- ARIMA(3,0,0) not useful based on Ljung-box test

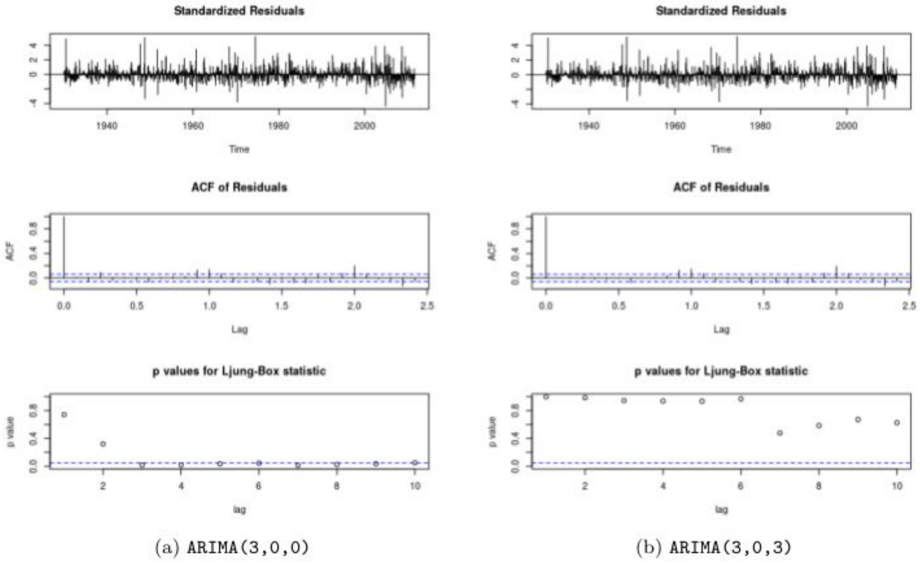
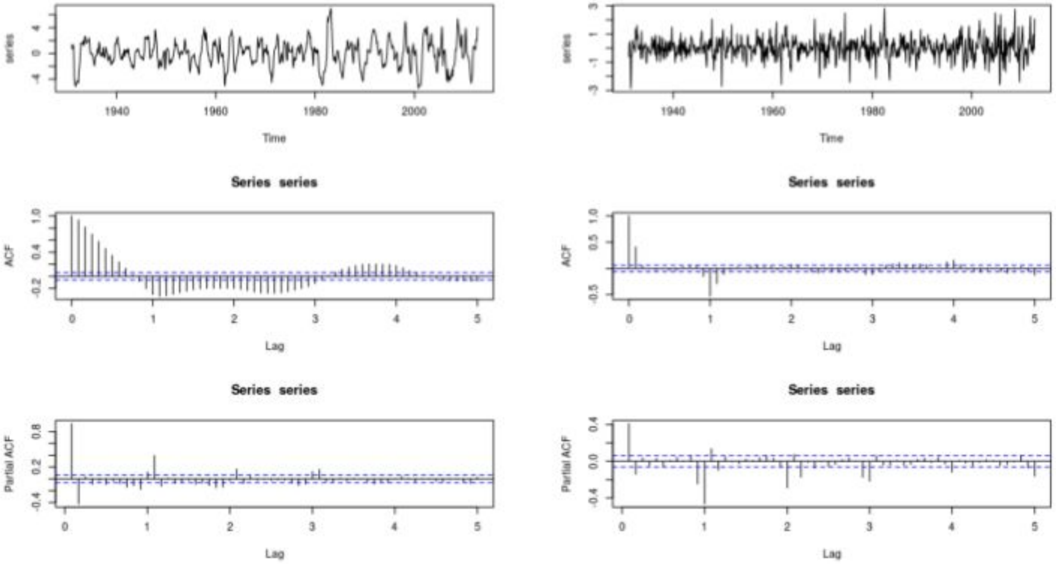


Figure 2: Diagnosis of the fitted ARIMA models

Seasonal Difference



(a) Seasonal Difference

(b) Seasonal and Ordinary Difference

Figure 3: Result of making the series stationary with difference differences

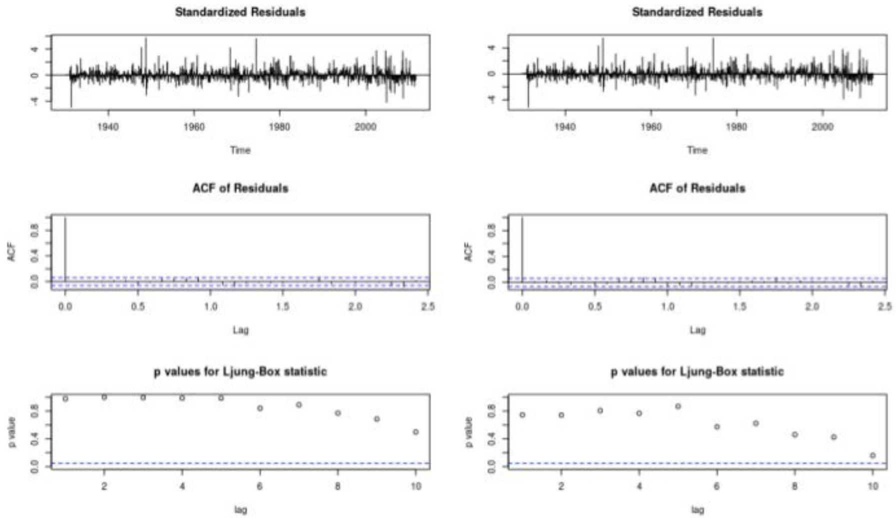
SARIMA models

Based on ACF and PACF plots of the seasonal difference:

- Up to 3 seasonal AR terms
- A seasonal MA term may needed
- 3 non-seasonal AR terms
- Some non-seasonal AR terms

Found Models:

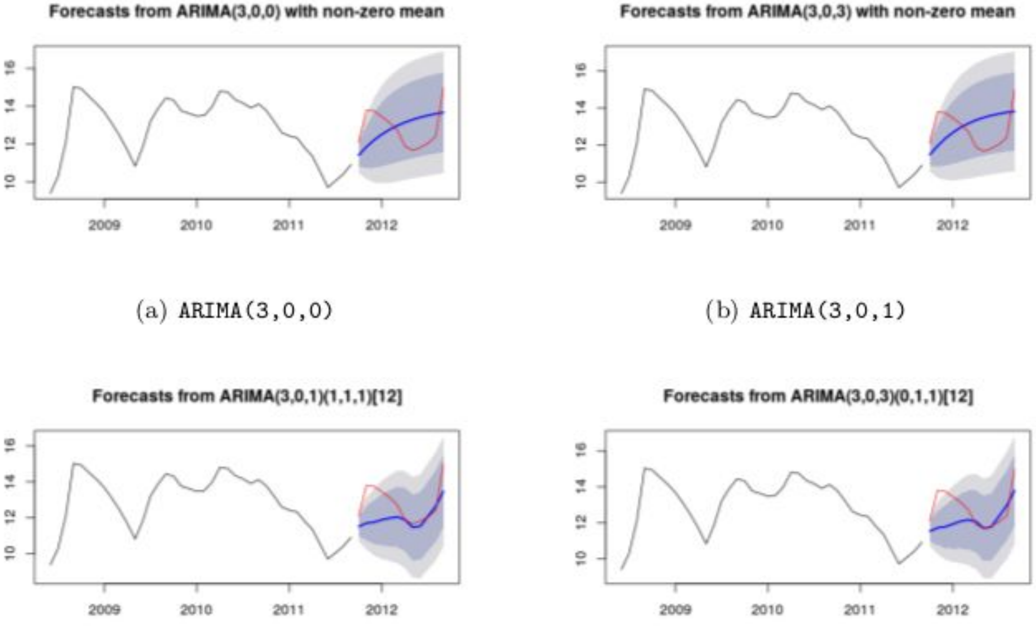
- SARIMA(3,0,1)x(1,1,1)_12
- SARIMA(3,0,3)x(0,1,1)_12



(a) SARIMA(3,0,1)x(1,1,1)_12 (b) SARIMA(3,0,3)x(0,1,1)_12

Figure 4: Diagnosis of the fitted SARIMA models

Forecast of the Models



(a) ARIMA(3,0,0)

(b) ARIMA(3,0,1)

(c) SARIMA(3,0,1)x(1,1,1)_12

(d) SARIMA(3,0,3)x(0,1,1)_12

Figure 5: Forecast of the ARIMA/SARIMA fitted models on WL45

Error measurements

	MAE	MSE	MAPE
ARIMA(3,0,0)	1.176005	1.278660	9.229816
ARIMA(3,0,1)	1.201571	1.320867	9.502182
SARIMA(3,0,1)x(1,1,1)	0.857519	1.126478	6.362497
SARIMA(3,0,3)x(0,1,1)	0.8378983	1.0854739	6.2424951

Table 1: Performance of each (S)ARIMA model based on different error measurements

Choice of SARIMA model based on Akaike

```
Coefficients:
        ar1 ar2 ar3 ma1 sar1 sma1
     0.7938 0.2630 -0.1468 0.6323 -0.1076 -0.9527
s.e. 0.1568 0.2263 0.0855 0.1495 0.0349 0.0166

sigma^2 estimated as 0.1792: log likelihood = -558.76, aic = 1131.52
```

Listing 3: SARIMA$(3, 0, 1)$x$(1, 1, 1)_{12}$ model output

```
Coefficients:
        ar1 ar2 ar3 ma1 ma2 ma3 sma1
     0.8754 0.9461 -0.8473 0.5354 -0.8383 -0.3881 -0.9626
s.e. 0.0837 0.1328 0.0755 0.0936 0.0853 0.0614 0.0160

sigma^2 estimated as 0.1799: log likelihood = -561.44, aic = 1138.88
```

Listing 4: SARIMA$(3, 0, 3)$x$(0, 1, 1)_{12}$ model output

SARIMA model formulation (1)

$$\Phi(B^m)\phi(B)\nabla_m^D\nabla^d Z_t = \Theta(B^m)\theta(B)A_t$$

$$\Phi(B^{12})\phi(B)\nabla_{12}^1\nabla^0 Z_t = \Theta(B^{12})\theta(B)A_t$$

$$(1-\Phi_1 B^{12})(1-\phi_1 B-\phi_2 B^2-\phi_3 B^3)(Z_t - Z_{t-12}) = (1+\Theta_1 B^{12})(1+\theta_1 B)A_t$$

SARIMA model formulation (2)

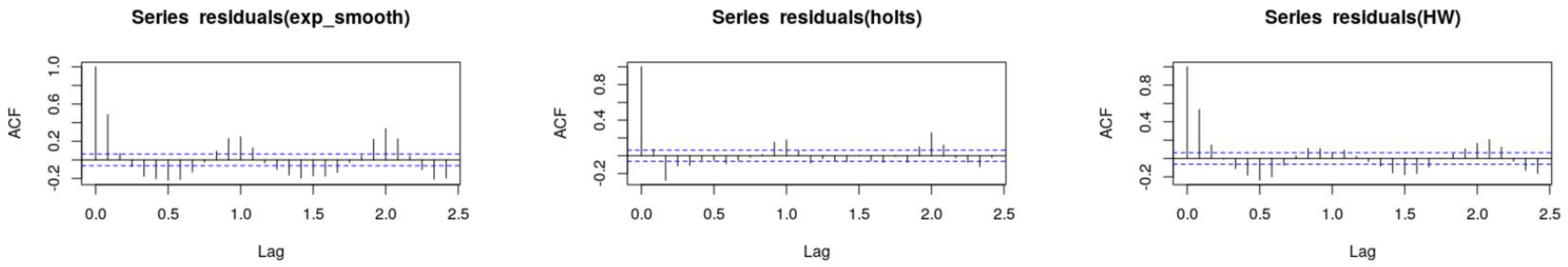
$$(1-\Phi_1 B^{12})(1-\phi_1 B)(Z_t - Z_{t-12}) = (1+\Theta_1 B^{12})(1+\theta_1 B)A_t$$

$$(1-\phi_1 B - \Phi_1 B^{12} + \Phi_1\phi_1 B^{13})(Z_t - Z_{t-12}) = (1+\theta_1 B + \Theta_1 B^{12} + \Theta_1\theta_1 B^{13})A_t$$

$$\begin{aligned} Z_t = {} & \phi_1 Z_{t-1} + \Phi_1 Z_{t-12} - \Phi_1\phi_1 Z_{t-13} + Z_{t-12} - \phi_1 Z_{t-13} - \Phi_1 Z_{t-24} - \Phi_1\phi_1 Z_{t-25} \\ & + A_t + \theta_1 A_{t-1} + \Theta_1 A_{t-12} + \Theta_1\theta_1 A_{t-13} \end{aligned}$$

$$\begin{aligned} Z_t = {} & (0.7938)Z_{t-1} + (-0.1076)Z_{t-12} - (-0.1076 \times 0.7938)Z_{t-13} + Z_{t-12} \\ & - (0.7938)Z_{t-13} - (-0.1076)Z_{t-24} - (-0.1076 \times 0.7938)Z_{t-25} \\ & + A_t + (0.6323)A_{t-1} + (-0.9527)A_{t-12} + (-0.9527 \times 0.6323)A_{t-13} \end{aligned}$$

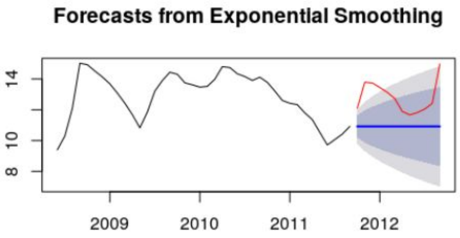
Holt-winters family



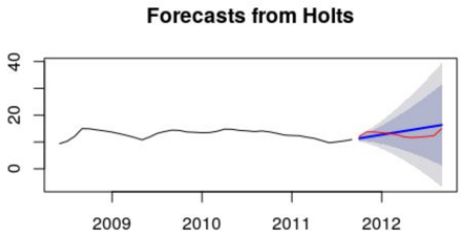
(a) `Exponential Smoothing` (b) `Holts` (c) `Holts Winter`

Figure 6: Residuals of the HW-family models studied with auto-correlation function

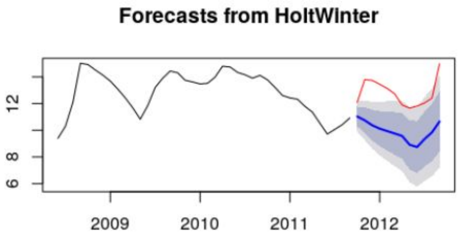
Forecasts for Holt-winters



(a) Exponential Smoothing



(b) Holt's Method



(c) Holt-Winters

Figure 7: Forecast of the HW-family models

Holt-winters models performance

	MAE	MSE	MAPE
Exponential Smoothing	1.894033	2.130382	14.307822
Holt's Method	1.854273	2.163968	14.874172
Holt-Winters	2.891893	2.983005	22.401574

Table 2: Performance of each HW-family based on different error measurements

Holt's Method

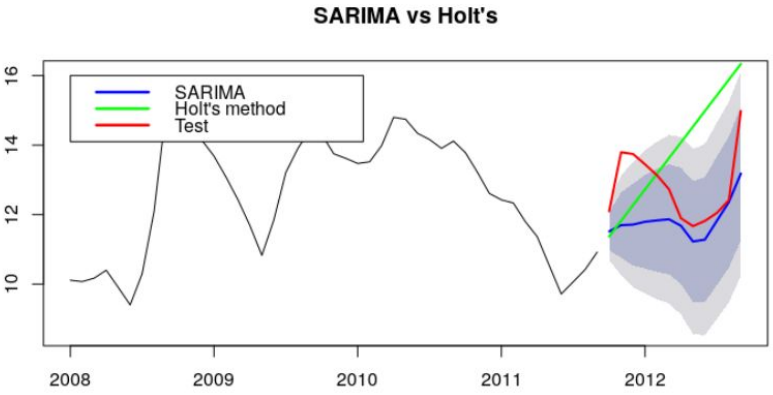
```
Smoothing parameters:
 alpha: 0.8415971
 beta : 0.005600746
 gamma: 1
```

$$\hat{Z}_{t+p|t} = a_t + b_t p$$
$$a_t = \alpha z_t + (1-\alpha)\hat{z} = \alpha z_t + (1-\alpha)(a_{t-1} + b_{t-1})$$
$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}$$

$$\hat{Z}_{t+p|t} = a_t + b_t p$$
$$a_t = 0.841 z_t + (1-0.841)(a_{t-1} + b_{t-1}),$$
$$b_t = 0.005(a_t - a_{t-1}) + (1-0.005)b_{t-1}$$

Conclusion for first time series

- SARIMA performed better among all experimented models
- Among Holt-winters family Holts worked best
- The performance of models evaluated using the error measurements
- AIC was only needed to be used for SARIMA models comparison
- Other transformations might have helped



Time Series 2: TC093s

General information:

- Traffic Accidents with casualties on the roads
- 336 observations
- From Jan 1990 to Dec 2017

Further details:

- Non-stationary
- With obvious seasonality

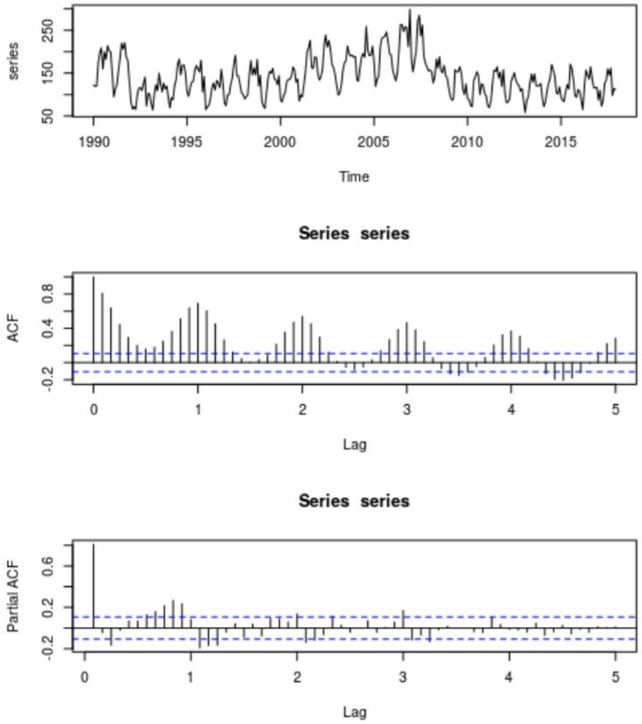


Figure 9: TC093s.txt plots

Stationarity tests

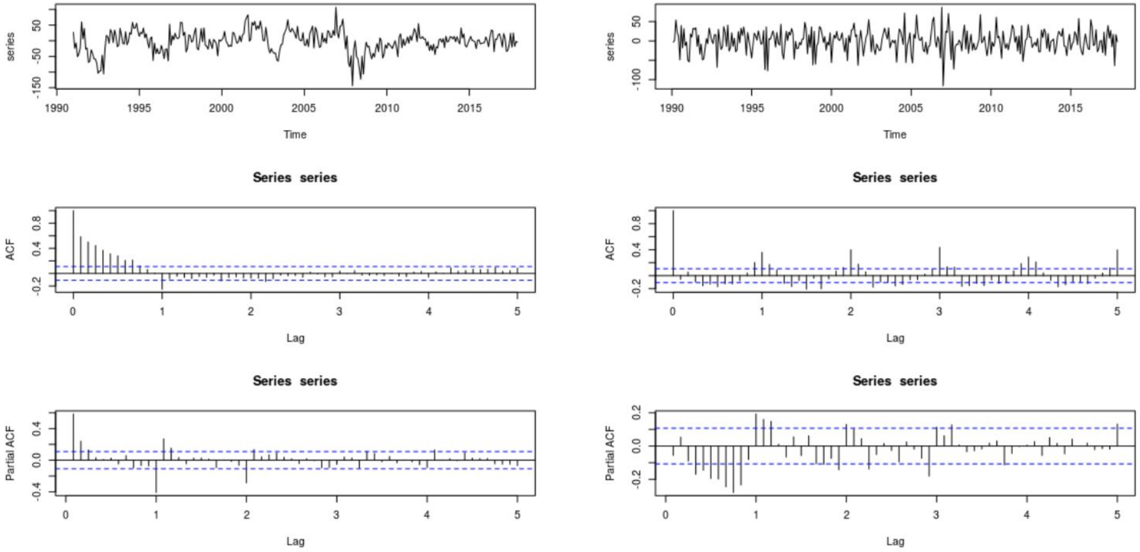
```
Dickey-Fuller = -6.1154, Truncation lag parameter = 5, p-value = 0.01
```

Listing 6: Phillips-Perron Unit Root Test

```
Dickey-Fuller = -4.1344, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

Listing 7: Augmented Dickey-Fuller Test

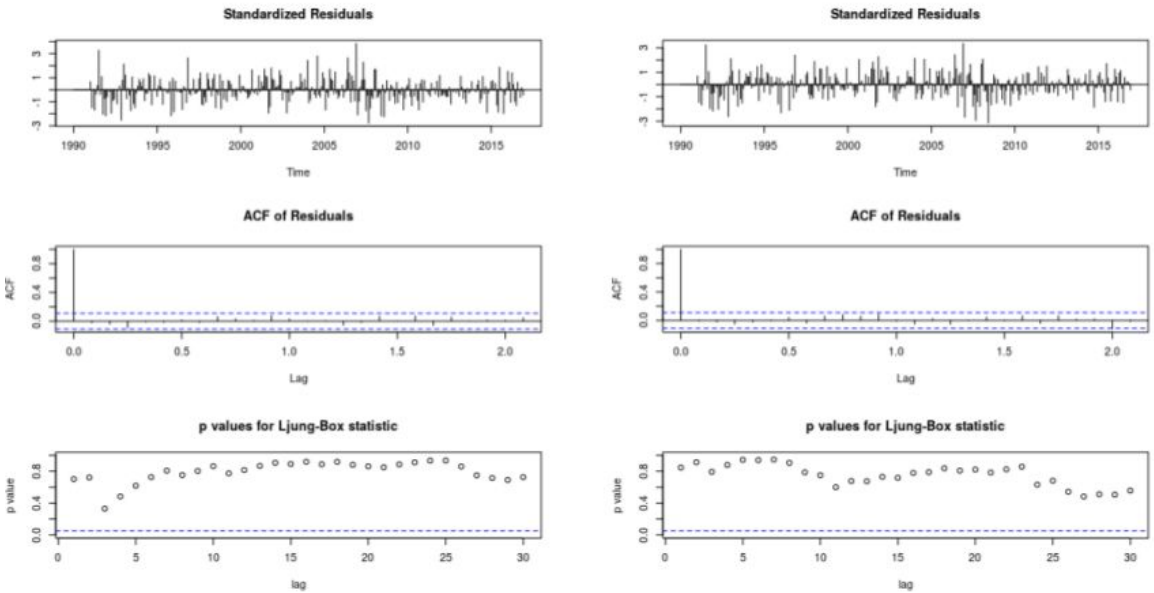
Seasonal difference



(a) Seasonal Difference (b) Seasonal and Ordinary Difference

Figure 10: Result of making the series stationary with difference differences

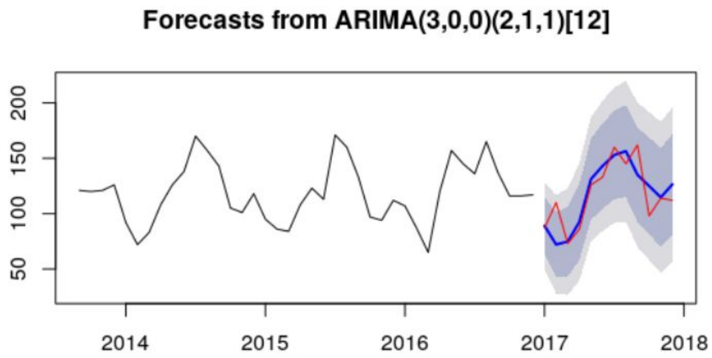
Fitted SARIMA models



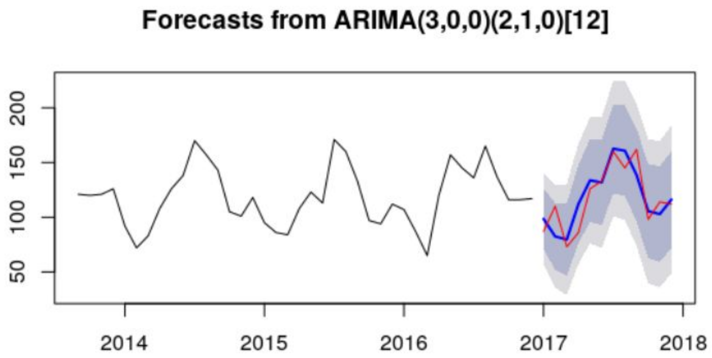
(a) `SARIMA(3,0,0)x(2,1,1)` (b) `SARIMA(3,0,0)x(2,1,0)`

Figure 11: Diagnosis of the fitted SARIMA models

Forecast results



(a) `SARIMA(3,0,0)x(2,1,1)`



(b) `SARIMA(3,0,0)x(2,1,0)`

Figure 12: Forecast of the fitted SARIMA fitted models on TC093s

Model performance based on error

	MAE	MSE	MAPE
SARIMA(3,0,0)x(2,1,1)	12.59178	16.96720	10.70865
SARIMA(3,0,0)x(2,1,0)	12.11275	14.93572	11.07459
ARIMAX(3,0,0)x(2,1,1)	18.51041	20.80500	16.46115

Table 3: Performance of each (S)ARIMA(X) models based on different error measurements

Choice of model based on Akaike

```
Coefficients:
        ar1 ar2 ar3 sar1 sar2 sma1
     0.5196 0.2069 0.1890 -0.1035 -0.0254 -0.8525
s.e. 0.0563 0.0621 0.0559 0.0806 0.0761 0.0618

sigma^2 estimated as 403.6: log likelihood = -1387.96, aic = 2789.92
```

Listing 8: `SARIMA(3,0,0)x(2,1,1)` model output

```
Coefficients:
        ar1 ar2 ar3 sar1 sar2
     0.4931 0.2130 0.1664 -0.7492 -0.3965
s.e. 0.0559 0.0614 0.0558 0.0554 0.0552

sigma^2 estimated as 460.8: log likelihood = -1403.85, aic = 2819.7
```

Listing 9: `SARIMA(3,0,0)x(2,1,0)` model output

SARIMA model formulation

$$\Phi(B^m)\phi(B)\nabla_m^D\nabla^d Z_t = \Theta(B^m)\theta(B)A_t$$

$$\Phi(B^{12})\phi(B)\nabla_{12}^1\nabla^0 Z_t = \Theta(B^{12})\theta(B)A_t$$

$$(1-\Phi_1 B^{12}-\Phi_2 B^{24})(1-\phi_1 B-\phi_2 B^2-\phi_3 B^3)(Z_t - Z_{t-12}) = (1+\Theta_1 B^{12})A_t$$

$$\boxed{(1+0.1035B^{12}+0.0254B^{24})(1-0.5196B-0.2069B^2-0.1890B^3)(Z_t - Z_{t-12}) = (1-0.8525B^{12})A_t}$$

ARIMAX - Regression time series: TC1422s

General information:

- Freight traffic on railways
- 326 observations
- From Jan 1991 to Feb 2018

Further details:

- Non-stationary
- With possibility of having seasonality
- The only series with matching dates in the dataset available for regressing with TC093s

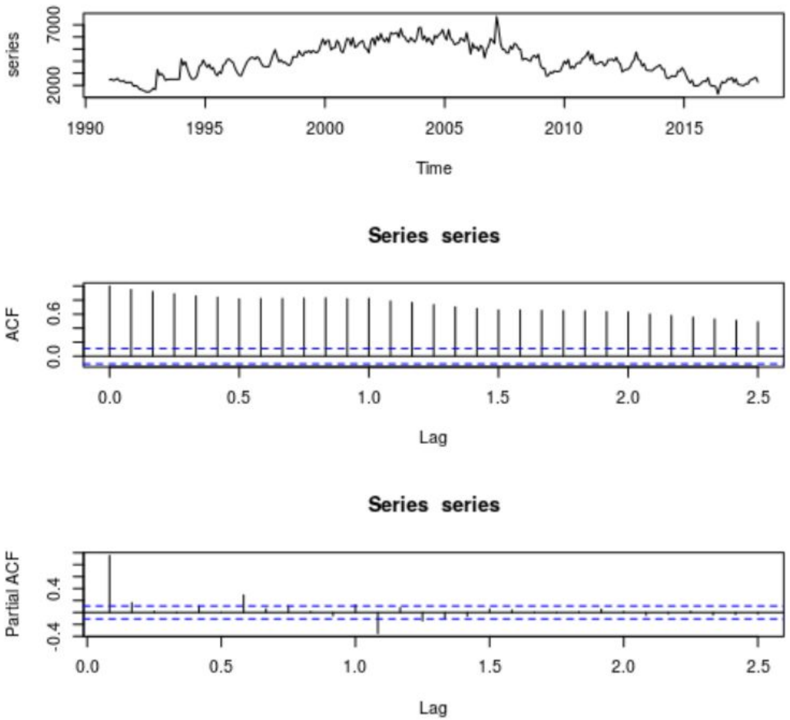


Figure 13: TC093s.txt plots

Steps taken

- Align the beginning and ending dates of the two series
- Cross correlate the series for checking any obvious relation
- Applying ARIMA on regressor and then target serie

Stationarity test

```
Dickey-Fuller = -2.3124, Truncation lag parameter = 5, p-value = 0.4448
```

Listing 10: Phillips-Perron Unit Root Test

```
Dickey-Fuller = -1.3798, Lag order = 6, p-value = 0.8381
alternative hypothesis: stationary
```

Listing 11: Augmented Dickey-Fuller Test

TC1422s Staionarized

Experiments:

- Ordinary difference
- Seasonal difference (period=12)
- Ordinary + seasonal difference (overkill)
- Based on ACF: some non-seasonal MA terms and possibly some seasonal
- Based on PACF: up to 3 non-seasonal AR terms and up to 2 seasonal terms

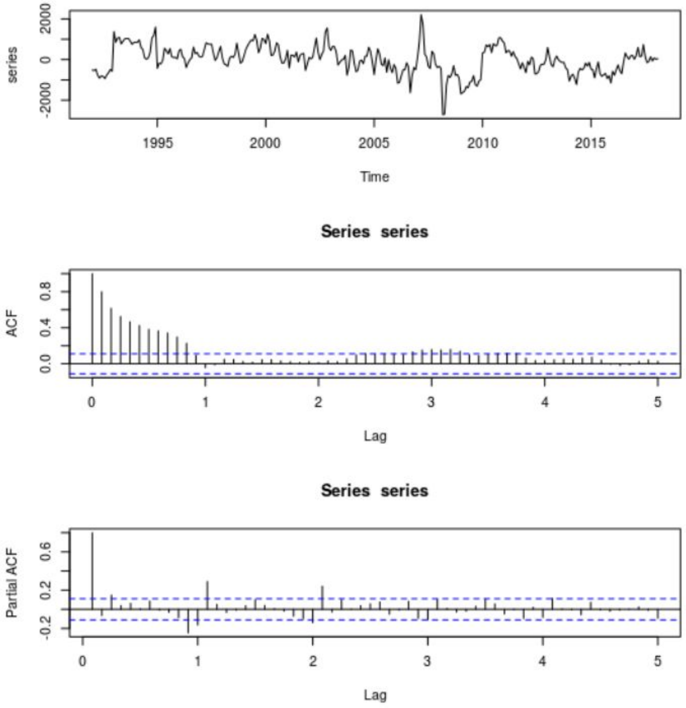


Figure 14: Seasonal Difference of `period=12` for the TC1422s

Appropriate lag from regressor

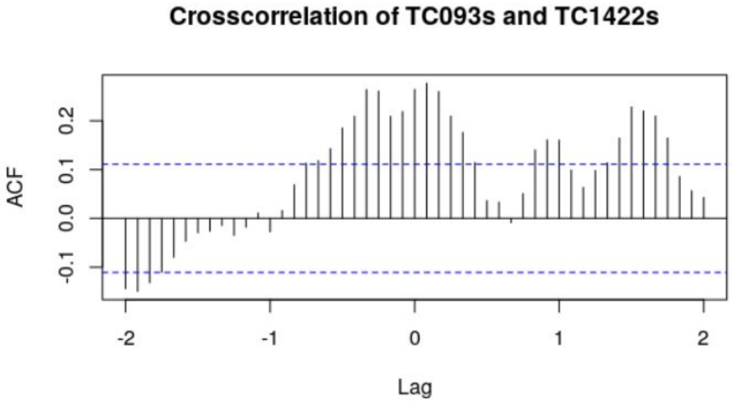


Figure 15: Cross-correlation of two time series TC093s and TC1422s

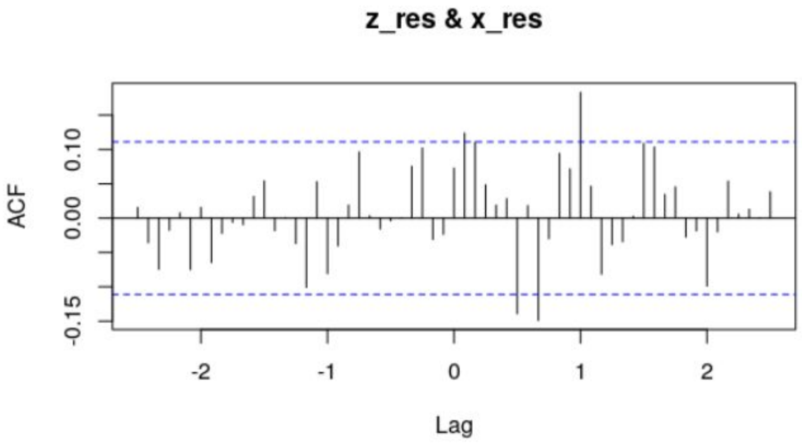


Figure 16: Cross-correlation of residuals of the fitted model to X and Z (pre-whitening method)

ARIMAX Fitted model

- Ljung-box for horizon >= 12
- ACF lag 12
- Best found model: ARIMAX(3,0,1)x(1,1,0)_12

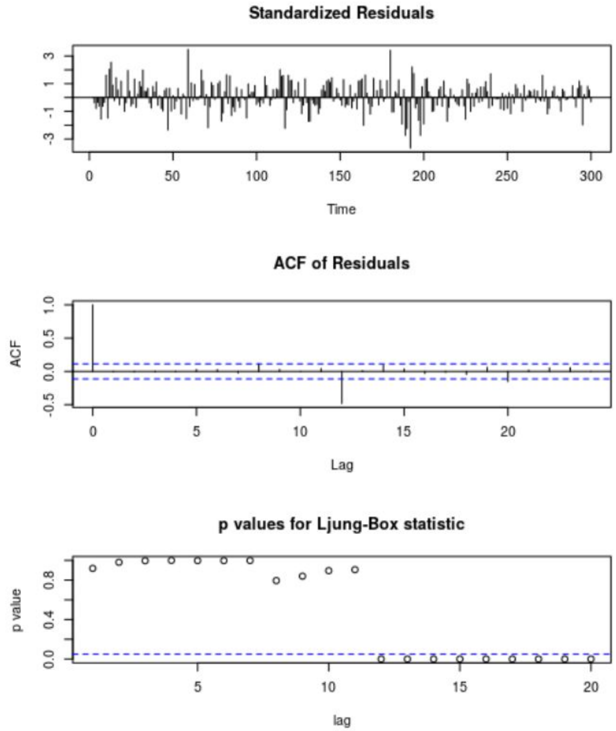


Figure 17: Residual obtained from fitting ARIMA to the Z with X as a regressor

ARIMAX formulation (1)

$$Z_t = \beta X_t + \varepsilon_t \qquad ARIMAX general form$$

$$\Phi(B^{12})\phi(B)\nabla^1_{12}\varepsilon_t = \Theta(B^{12})\theta(B)A_t \qquad \varepsilon_t \text{ fitted to SARIMA model better}$$

```
arima(x = z1, order = c(3, 0, 1), seasonal = c(1, 1, 0), xreg = x1)

Coefficients:
        ar1 ar2 ar3 ma1 sar1 x1
     0.4540 0.1442 0.1626 -1.0000 -0.0572 0.0084
s.e. 0.3752 0.2111 0.0894 0.0105 0.3798 0.0034

sigma^2 estimated as 691.5: log likelihood = -1398.96, aic = 2811.91
```

Listing 12: Fitted ARIMAX Model output

ARIMAX formulation (2)

$$\begin{aligned}\varepsilon_t = {} & \phi_1(\varepsilon_{t-1} - \varepsilon_{t-13}) + \phi_2(\varepsilon_{t-2} - \varepsilon_{t-14}) + \phi_3(\varepsilon_{t-3} - \varepsilon_{t-15}) - \Phi_1(\varepsilon_{t-12} - \varepsilon_{t-24}) - \\ & \Phi_1\phi_1(\varepsilon_{t-13} - \varepsilon_{t-25}) - \Phi_1\phi_2(\varepsilon_{t-14} - \varepsilon_{t-26}) - \Phi_1\phi_3(\varepsilon_{t-15} - \varepsilon_{t-27}) + A_t + \theta_1 A_t + \varepsilon_{t-12}\end{aligned}$$

$$\begin{aligned}Z_t = {} & \beta X_t + \phi_1(\varepsilon_{t-1} - \varepsilon_{t-13}) + \phi_2(\varepsilon_{t-2} - \varepsilon_{t-14}) + \phi_3(\varepsilon_{t-3} - \varepsilon_{t-15}) - \Phi_1(\varepsilon_{t-12} - \varepsilon_{t-24}) - \\ & \Phi_1\phi_1(\varepsilon_{t-13} - \varepsilon_{t-25}) - \Phi_1\phi_2(\varepsilon_{t-14} - \varepsilon_{t-26}) - \Phi_1\phi_3(\varepsilon_{t-15} - \varepsilon_{t-27}) + A_t + \theta_1 A_t + \varepsilon_{t-12}\end{aligned}$$

$$\begin{aligned}Z_t = {} & 0.0084 X_t + 0.4540(\varepsilon_{t-1} - \varepsilon_{t-13}) + 0.1442(\varepsilon_{t-2} - \varepsilon_{t-14}) + 0.1626(\varepsilon_{t-3} - \varepsilon_{t-15}) - \\ & (-0.0572)(\varepsilon_{t-12} - \varepsilon_{t-24}) - (-0.0572 \times 0.4540)(\varepsilon_{t-13} - \varepsilon_{t-25}) - \\ & (-0.0572 \times 0.1442)(\varepsilon_{t-14} - \varepsilon_{t-26}) - (-0.0572 \times 0.1626)(\varepsilon_{t-15} - \varepsilon_{t-27}) + \\ & A_t + (-1.0000)A_t + \varepsilon_{t-12}\end{aligned}$$

Conclusion of second time series

	MAE	MSE	MAPE
SARIMA(3,0,0)x(2,1,1)	12.59178	16.96720	10.70865
SARIMA(3,0,0)x(2,1,0)	12.11275	14.93572	11.07459
ARIMAX(3,0,0)x(2,1,1)	18.51041	20.80500	16.46115

- SARIMA performed better
- ARIMAX could have been better with better regressor
- Other transformations might have helped

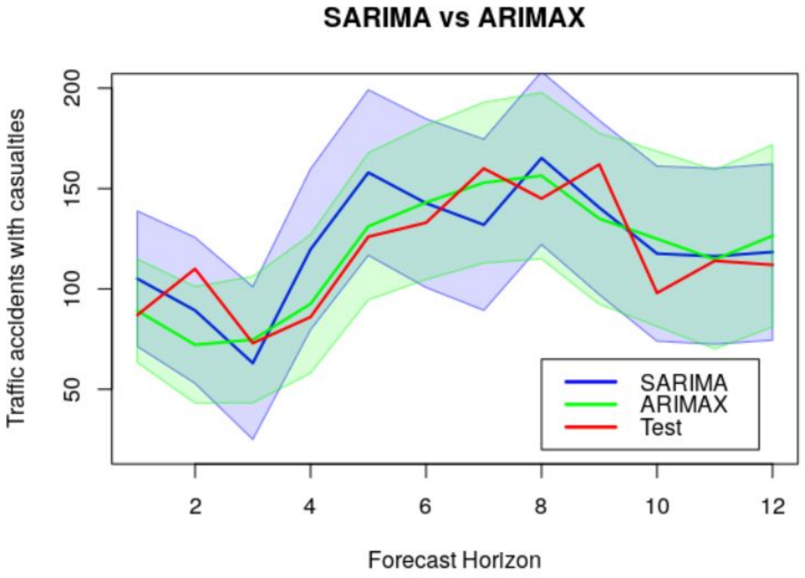


Figure 18: Result of the forecast with the SARIMA(3,0,0)x(2,1,1) and ARIMAX(3,0,1)x(1,1,0)

Thank you!

novin@ut.ee