Forecasting electricity consumption and production using ARIMA time series models

Barbu Paul – Gheorghe

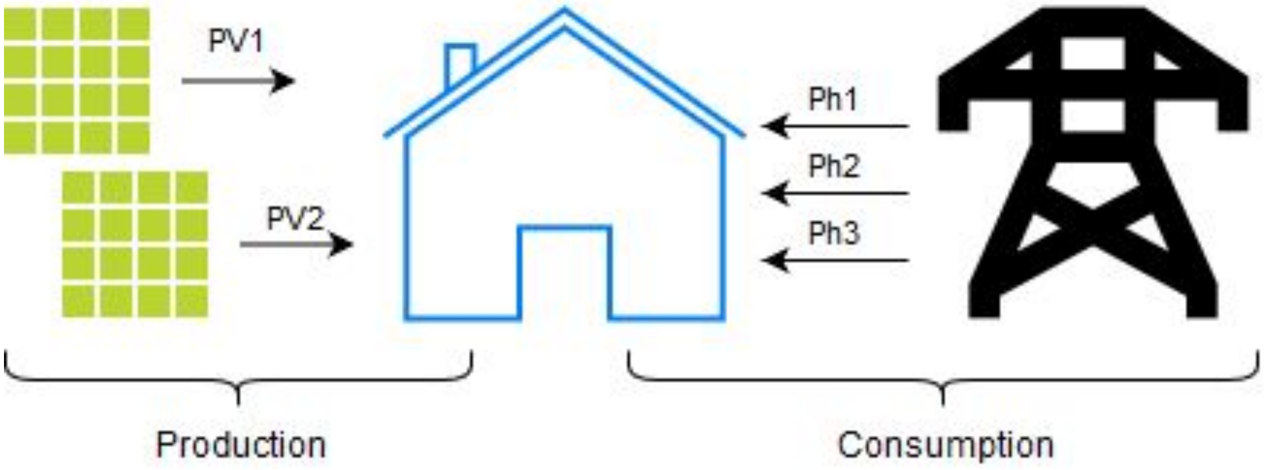
Advanced Computing Systems

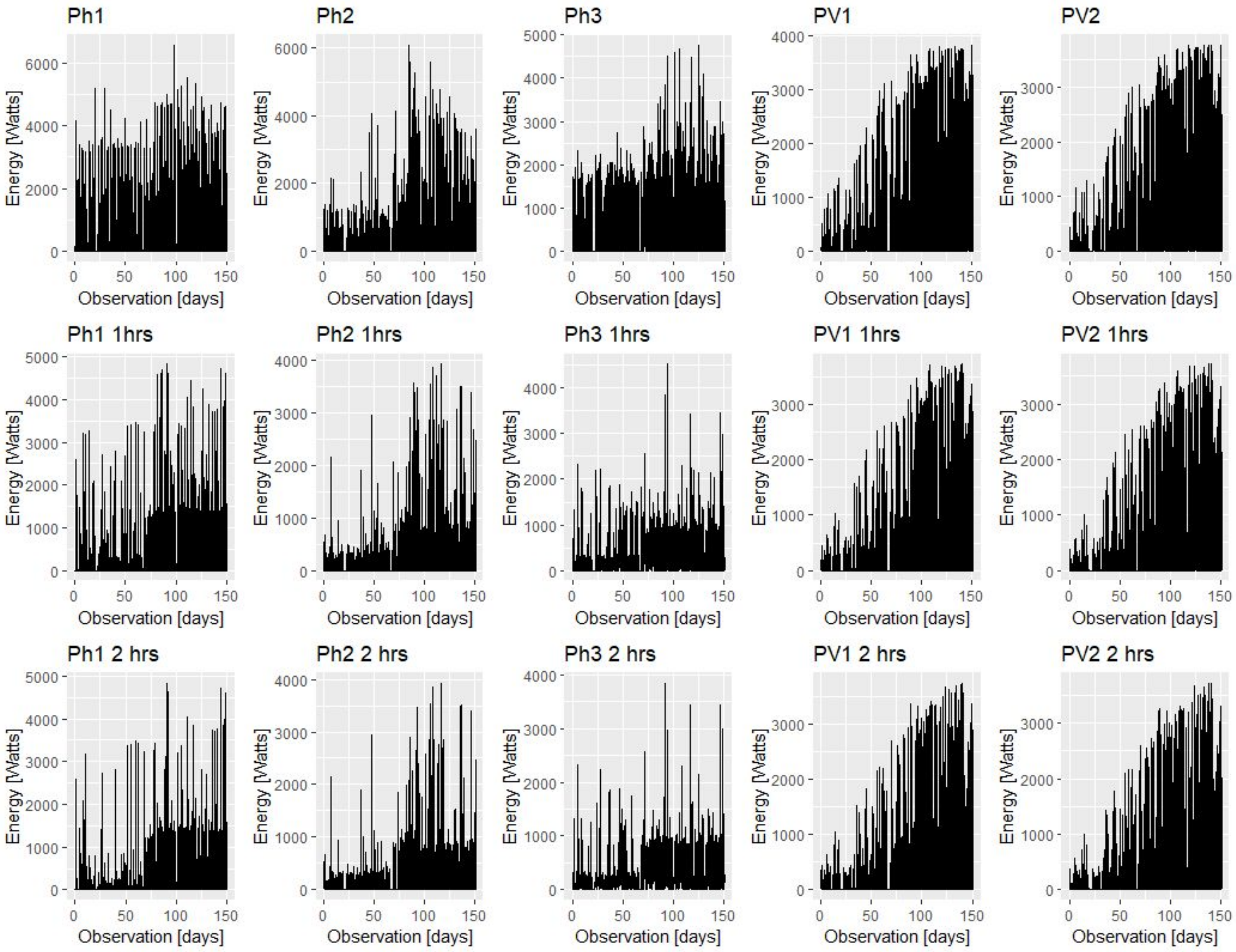
COORDINATOR: Conf. dr. ing. Árpád GELLÉRT

Sibiu, 2019

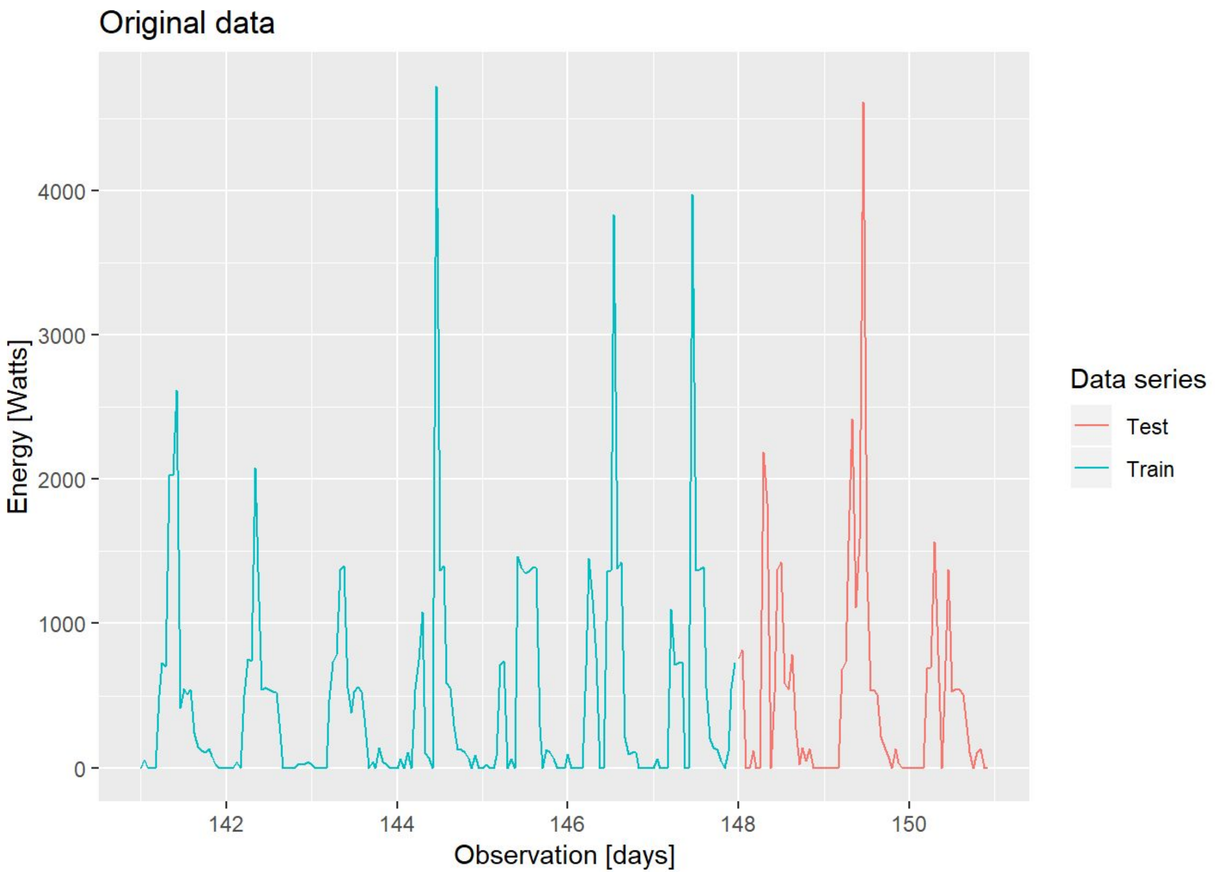
Objectives

- Furthering previous research (Feilmeier, Antonescu, Gellert)
- Forecasting electrical en. consumption & production




Ph1
Ph2
Ph3
PV1
PV2
Ph1 1hrs
Ph2 1hrs
Ph3 1hrs
PV1 1hrs
PV2 1hrs
Ph1 2 hrs
Ph2 2 hrs
Ph3 2 hrs
PV1 2 hrs
PV2 2 hrs
Energy [Watts]
Observation [days]

Last 10 days of ph1

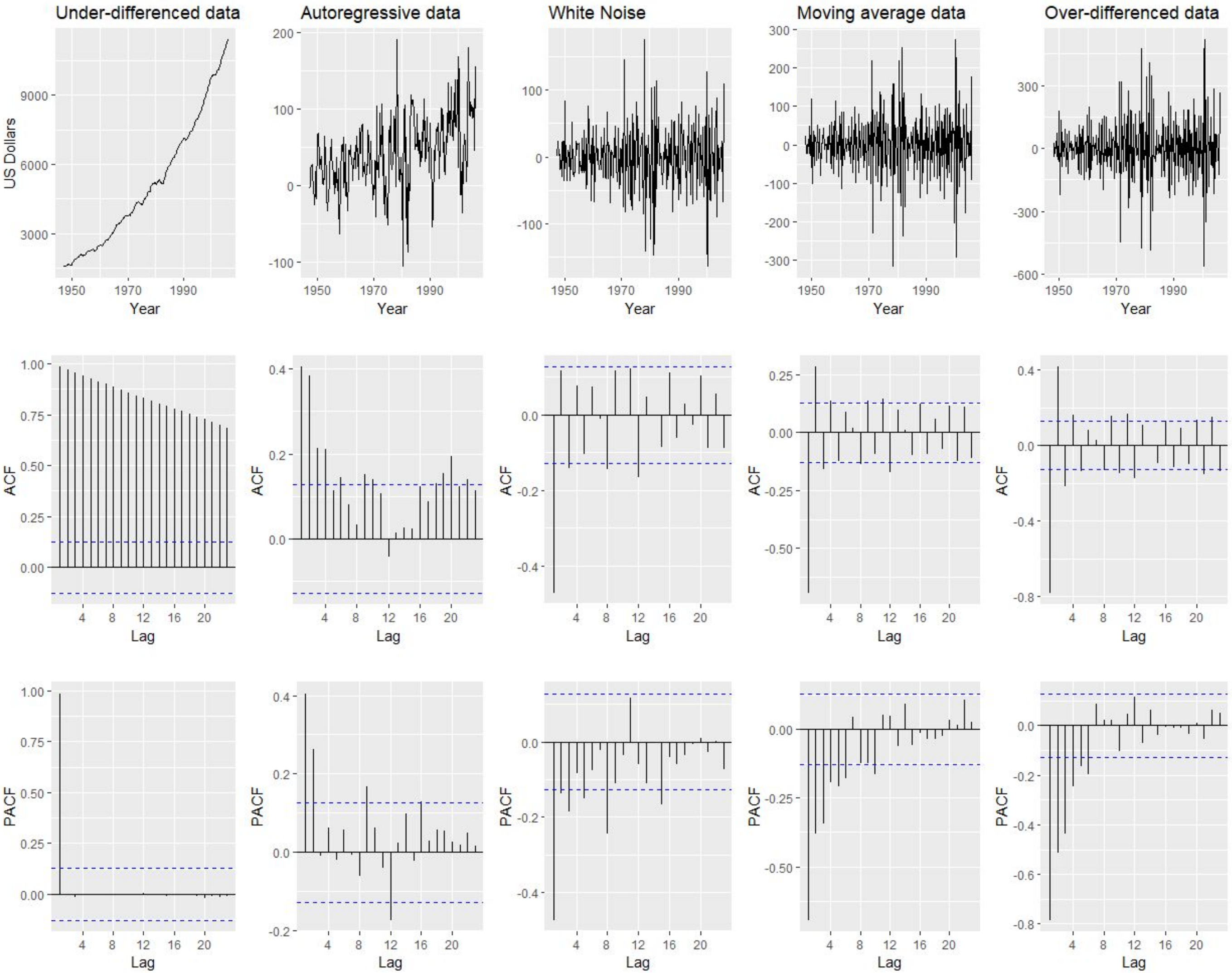


AutoRegressive Integrated Moving Average (ARIMA)

- Data has to be stationary (statistical properties constant over time)
- ARIMA(p, d, q)
- ARIMA(p, d, q)(P, D, Q)[m]
- AutoCorrelation Function (ACF)
- Partial AutoCorrelation Function (PACF)
- Residuals

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

- phi - AR coefficients
- theta - MA coefficients
- epsilon - errors



Under-differenced data
Autoregressive data
White Noise
Moving average data
Over-differenced data
US Dollars
Year
ACF
PACF
Lag

I(d)

- Integrated - the opposite of differencing
- The algorithm differences the data, hence it has to be "integrated" when we start

$$y'_t = y_t - y_{t-1}$$

$$\begin{aligned} y''_t &= y'_t - y'_{t-1} \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2}. \end{aligned}$$

AR(p)

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

- p - how many previous values affect this one
- phi - in what proportion does the previous value affect the present one
- ACF - decaying or sinusoidal - we do have correlation between several lags
- PACF - critical at lag p - up to which lag the correlation is significant

MA(q)

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

- q - how many previous errors affect this value
- epsilon - in what proportion does the previous errors affect the present value
- ACF - critical at lag q - up to which lag the correlation is significant
- PACF - decaying or sinusoidal - we do have correlation between several lags

ARIMAX

$$y'_t = c + \beta x_t + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

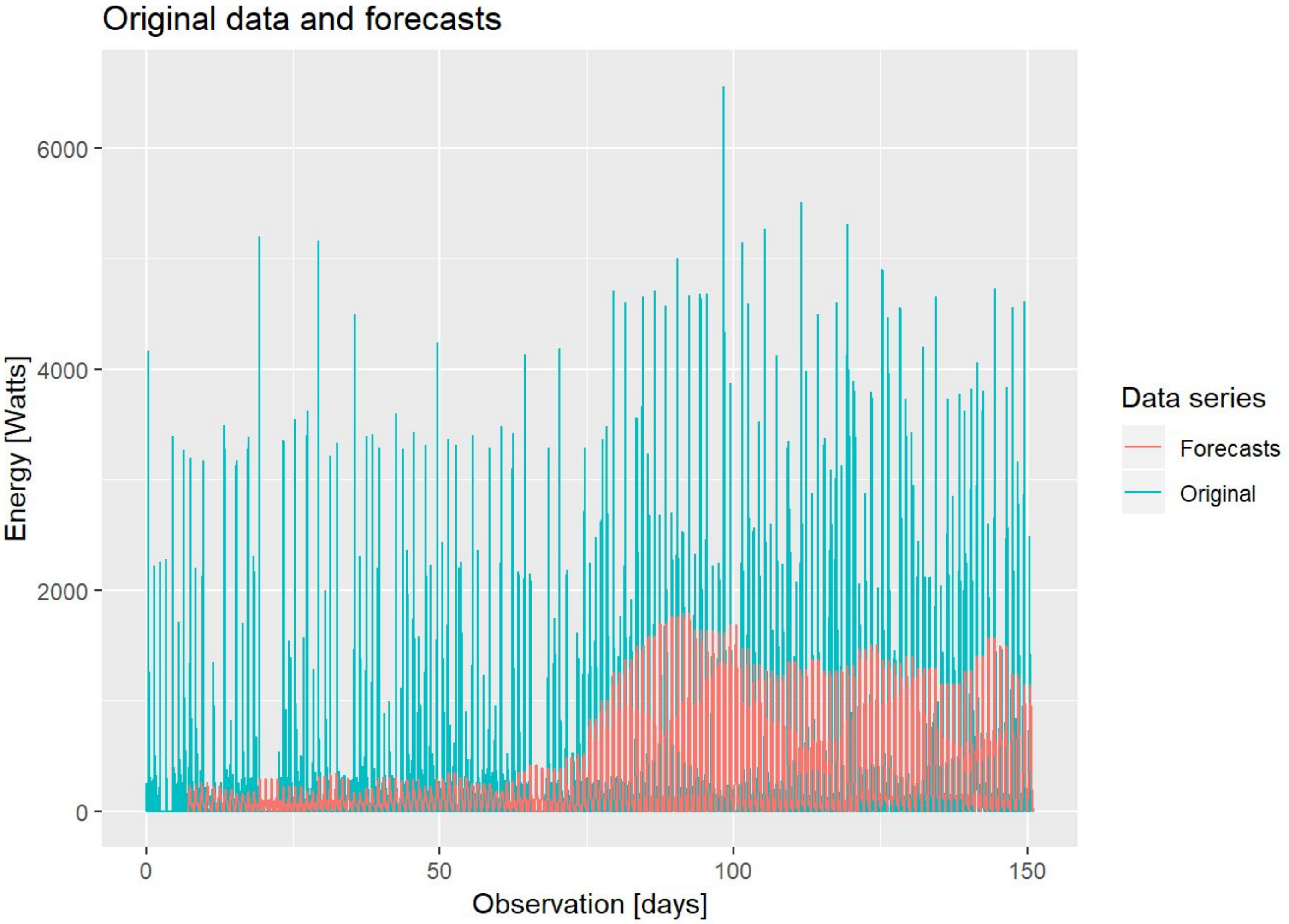
- x - external regressors
- beta - external regressors coefficient

External regressors:

- Seasonal dummies: quarter, month, events, etc.
- Dummies: trend, **time**, etc.
- Fourier terms

ph1

- 7 training days
- 2 forecast days
- ARIMA(1, 0, 0)
  - 2 Fourier terms
  - 11:00 - 12:00
- MAE = 218.04



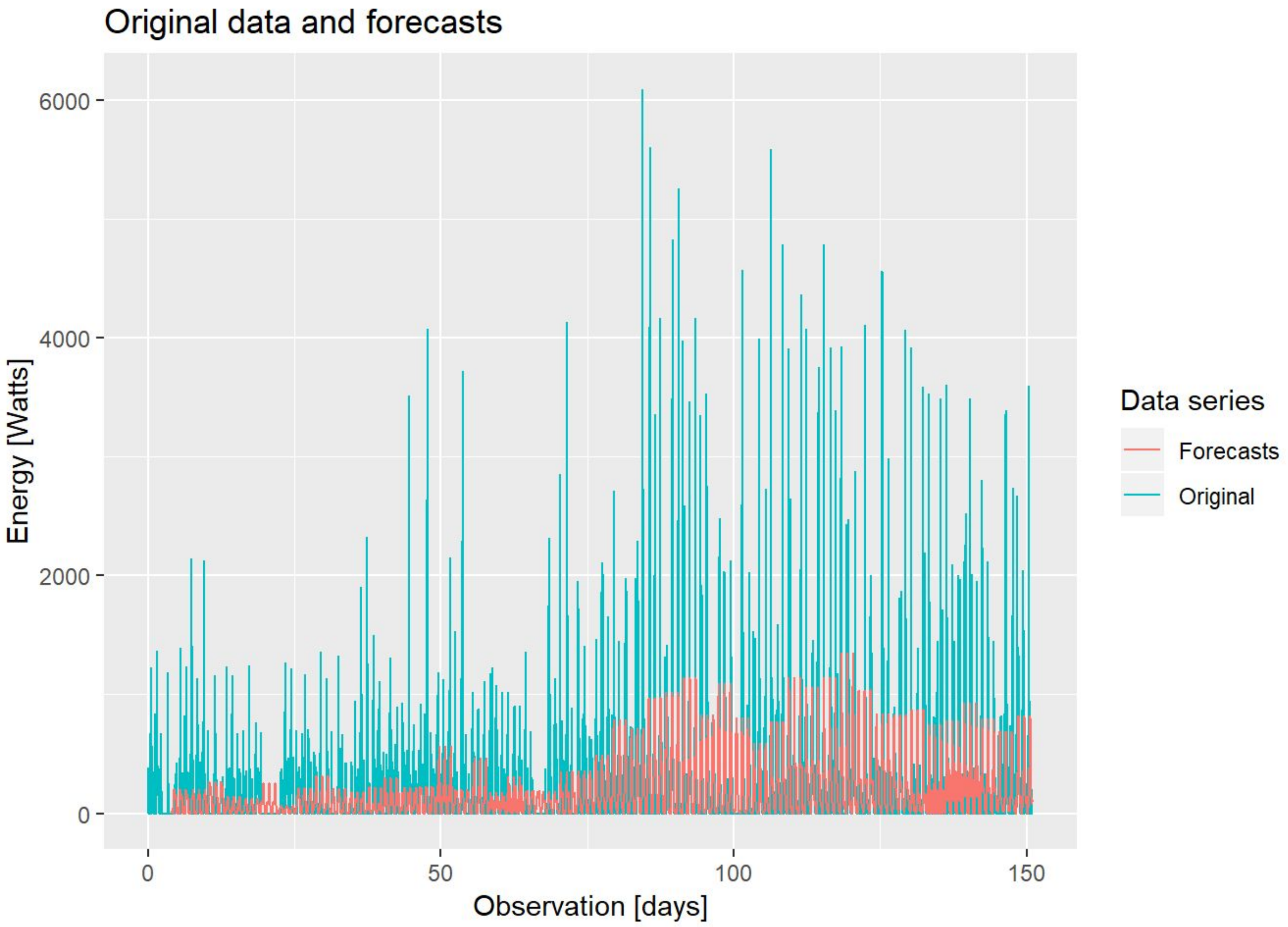
5min vs subsampled data

Time series	Model	K	Dummies (start:length)	Train days	Test days	Running time [s]	RMSE	MAE
2hrs ph1	Arima(1, 0, 0)	2	-	7	2	6.15	467.94	220.87
2hrs ph1	Arima(1, 0, 0)	2	12:6	7	2	6.80	463.00	218.81
2hrs ph1	Arima(1, 0, 0)(1, 0, 0)	2	-	7	2	4.83	469.44	223.40
2hrs ph1	Arima(1, 0, 0)(1, 0, 0)	2	12:6	7	2	5.08	464.96	222.06
1hrs ph1	Arima(1, 0, 0)	2	-	7	2	5.25	454.62	217.44
1hrs ph1	Arima(1, 0, 0)	2	11:1	7	2	5.97	450.00	215.51
1hrs ph1	Arima(1, 0, 0)(1, 0, 0)	2	-	7	2	6.70	455.76	218.36
1hrs ph1	Arima(1, 0, 0)(1, 0, 0)	2	11:1	7	2	5.89	452.21	217.18
ph1	mean	-	-	7	1	7.65	576.53	352.93
ph1	naive	-	-	7	1	7.75	655.62	286.68
ph1	snaive	-	-	7	1	2,906.22	587.99	231.19
ph1	Arima(1, 0, 0)	2	-	7	2	8.42	460.42	223.15
ph1	Arima(1, 0, 0)	2	11:1	7	2	10.59	452.44	218.04

Table 2.6: Models for the `Ph1` data series

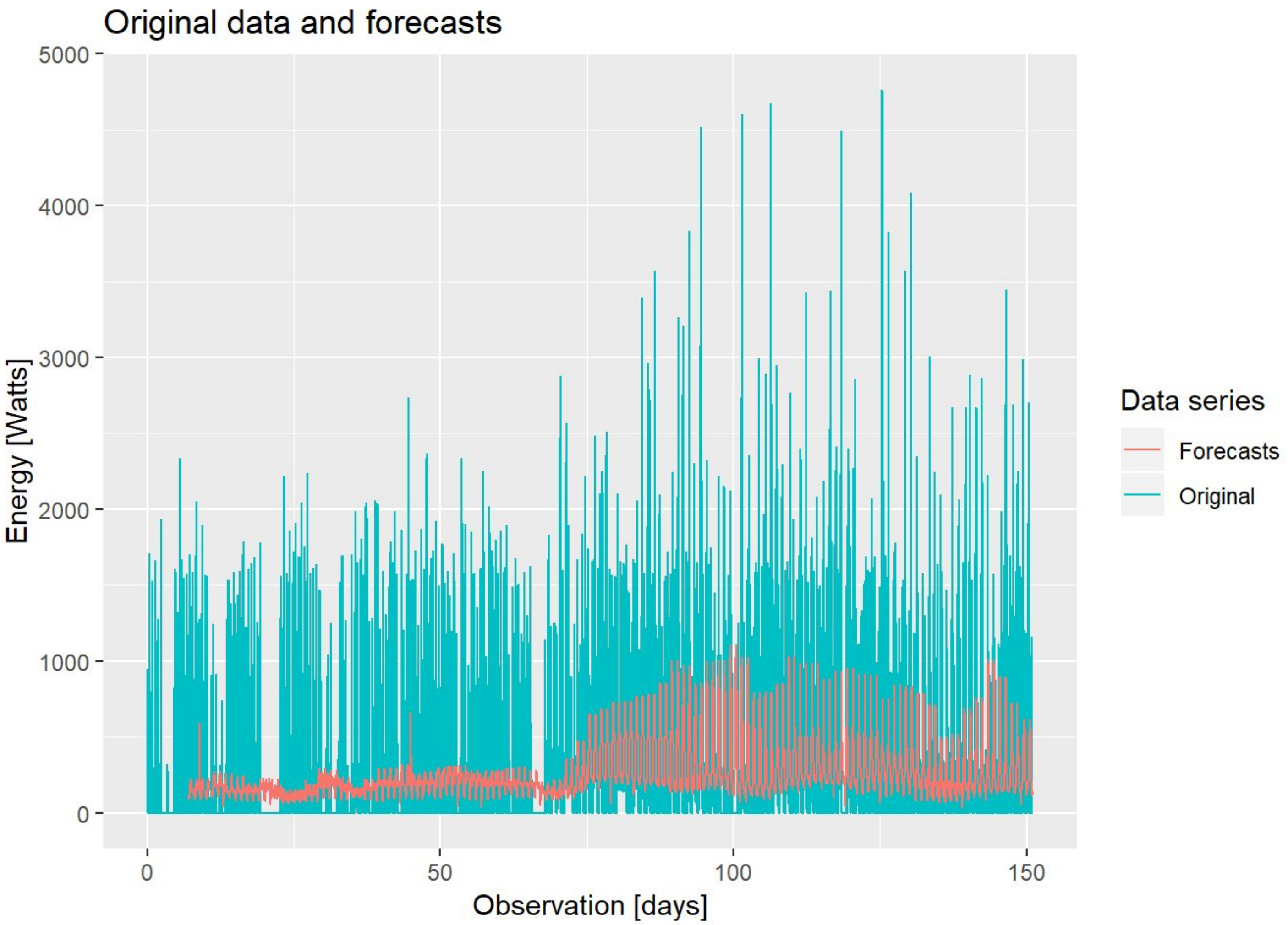
ph2

- 4 training days
- 3 forecast days
- ARIMA(1, 0, 0)
    - 3 Fourier terms
    - 08:00 - 11:00
- MAE=161.40



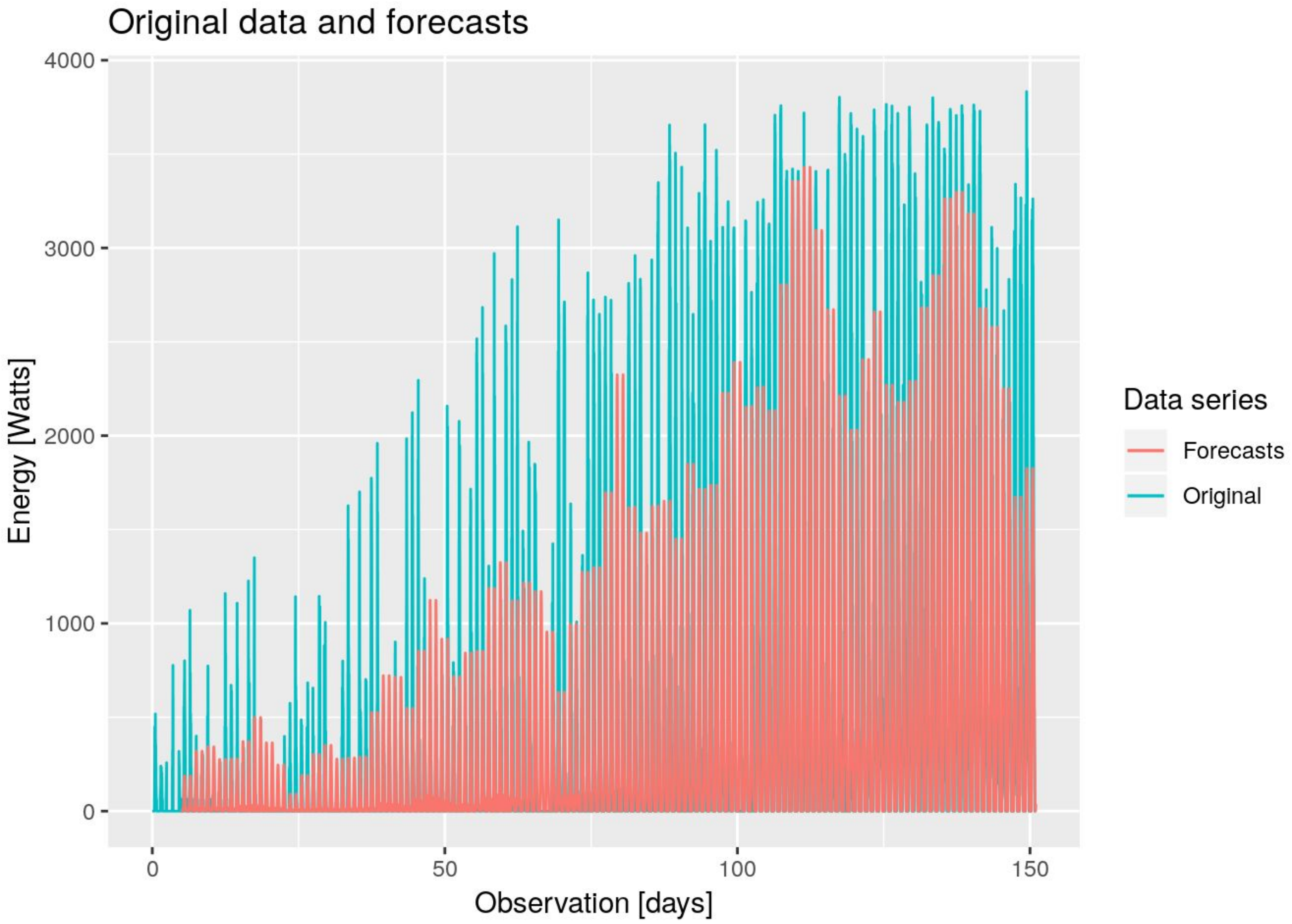
ph3

- 7 training days
- 2 forecast days
- ARIMA(1, 0, 0)
  - 2 Fourier terms
  - 09:00 - 11:00
- MAE = 183.96



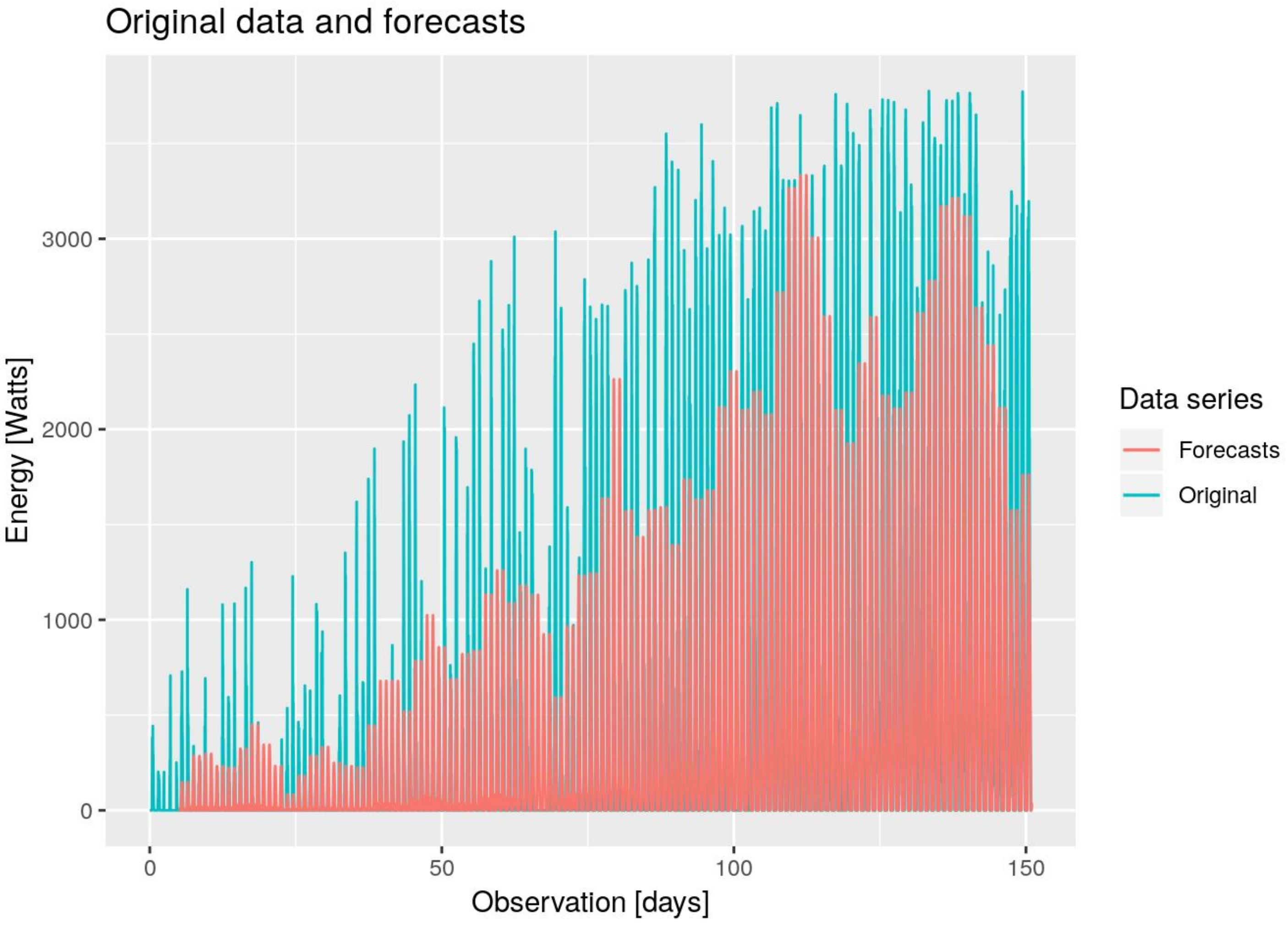
PV1

- 5 training days
- 2 forecast days
- ARIMA(1, 0, 0)
  - 3 Fourier terms
  - 11:00 - 17:00
- MAE = 225.91



PV2

- 5 training days
- 2 forecast days
- ARIMA(1, 0, 0)
  - 3 Fourier terms
- MAE = 220.00



ARIMA vs Markov vs Neural

Benchmark vs ARIMA on the ph3 data

Method	naive	meanf	snaive	ARIMA (1, 0, 0) 2 Fourier terms	ARIMA (1, 0, 0) 2 Fourier terms 09:00 to 11:00 dummies
MAE	215.26	216.88	218.67	187.96	183.96

Benchmark vs ARIMA on the PV1 data

Method	naive	meanf	snaive	ARIMA (1, 0, 0) 2 Fourier terms	ARIMA (1, 0, 0) 2 Fourier terms 16:00 to 20:00 dummies	ARIMA (1, 0, 0) 3 Fourier terms 11:00 to 17:00 dummies
MAE	445.33	516.38	262.57	238.26	237.10	225.91

Conclusions

- Only short periods of time can be predicted
- PV1, PV2 data sets are time-of-day correlated
- ARIMA is time consuming, simpler methods exist
- ARIMA models are not very suited for high frequency data

Future work

- Downsampling with interval averages instead of snapshots
- Automating the predictions for industry use
- Alternative algorithms: bats, tbats, ETS.

Critical values for ACF & PACF

Z distribution and p values

95% confidence interval - Z=1.96

(P)ACF critical value $\pm 1.96/\sqrt{T}$ (two std. dev. - assume an almost stationary series)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233

Differences

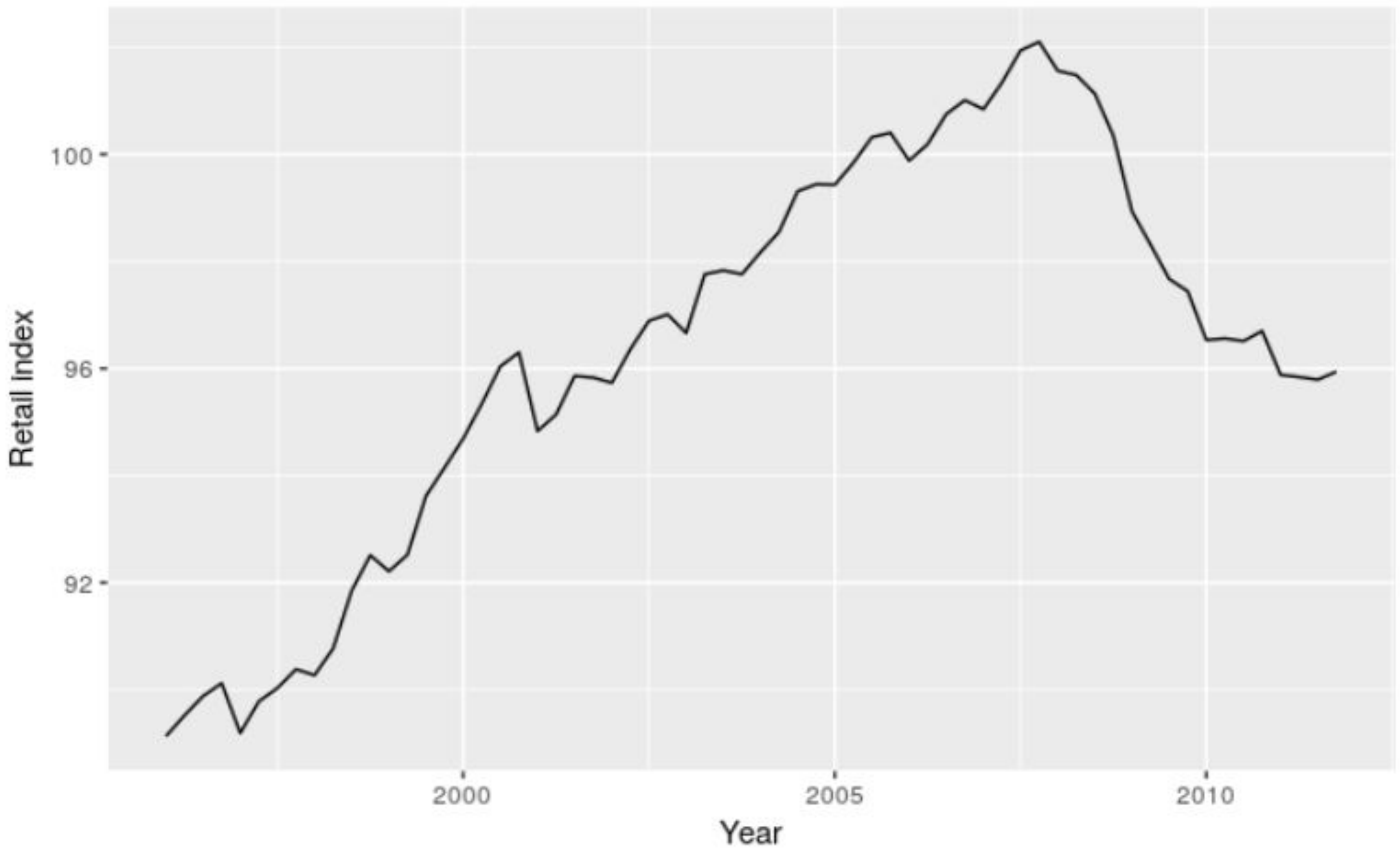
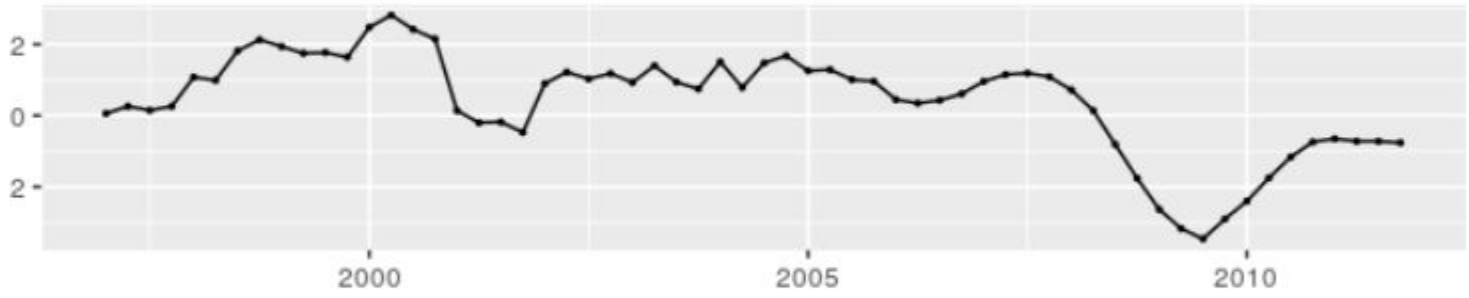


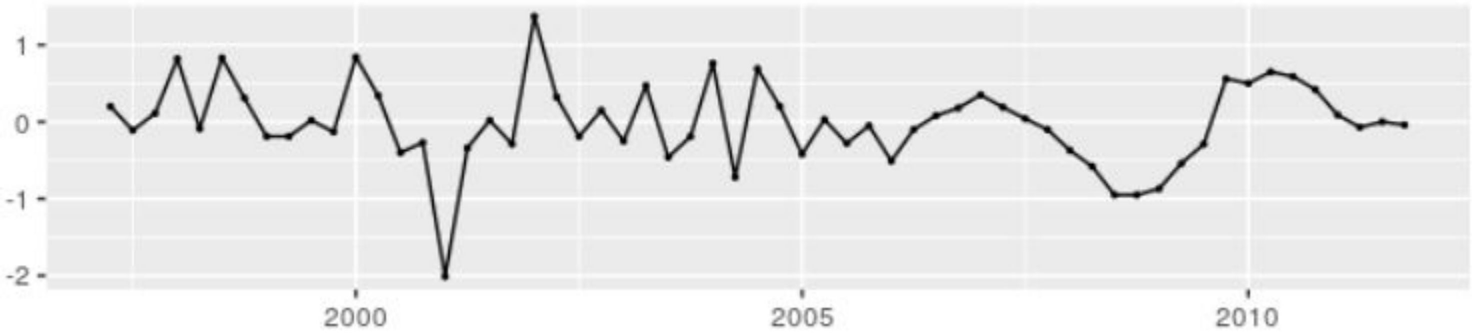
Figure 8.17: Quarterly retail trade index in the Euro area (17 countries), 1996–2011, covering wholesale and retail trade, and the repair of motor vehicles and motorcycles. (Index: 2005 = 100).

Quaterly data

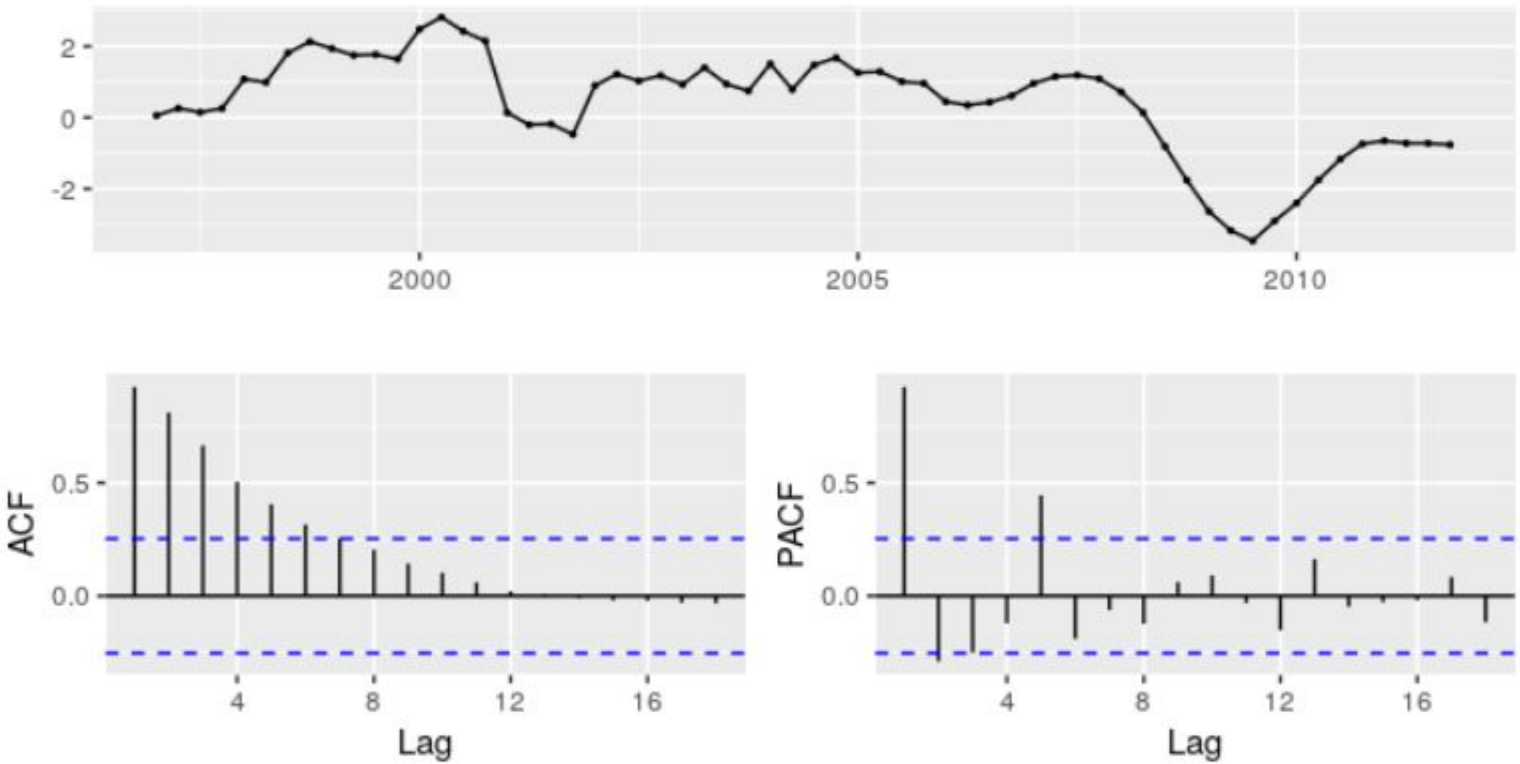
D=1



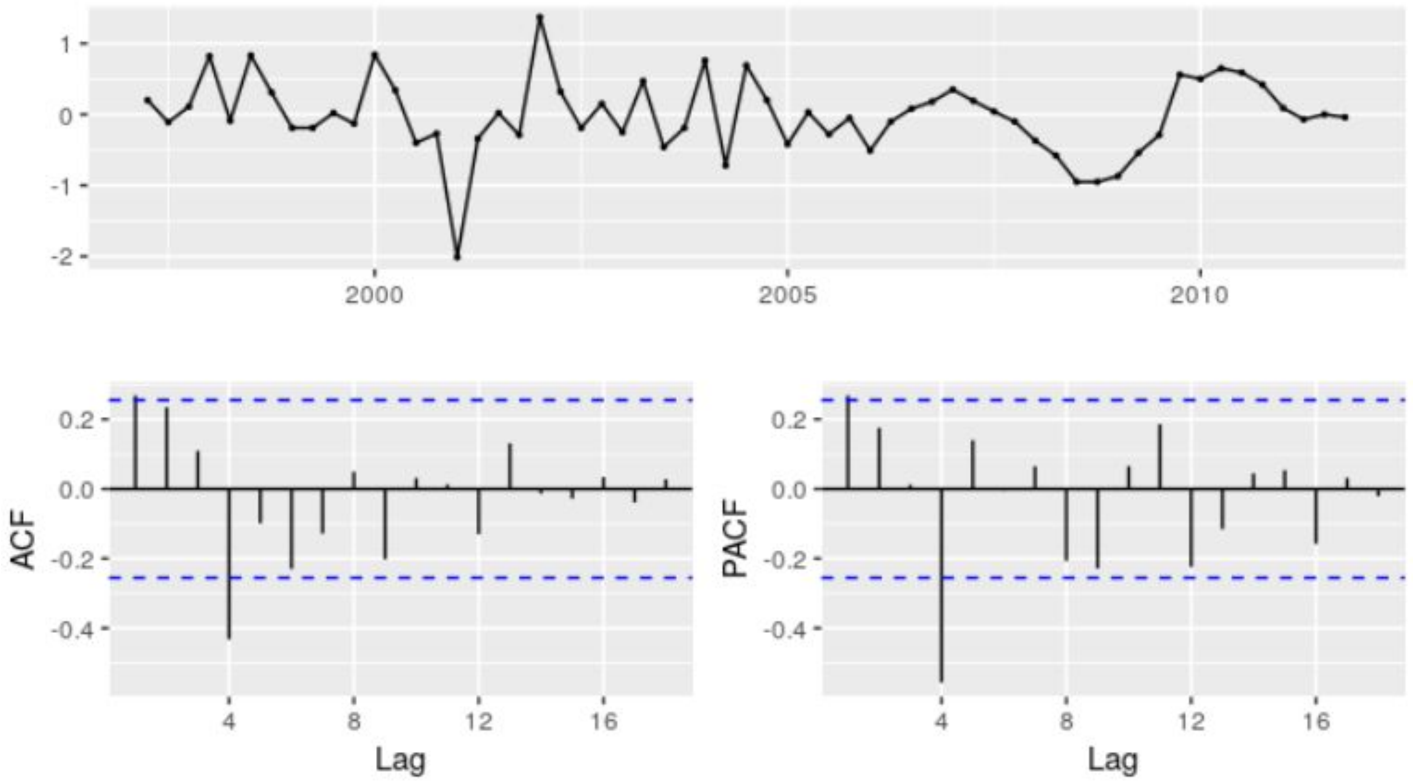
D=1 & d=1



AR(2)



MA(1) and SMA(1)


1
0
-1
-2
2000
2005
2010
ACF
0.2
0.0
-0.2
-0.4
4
8
12
16
Lag
PACF
0.2
0.0
-0.2
-0.4
4
8
12
16
Lag