

STK4150/9150 - Environmental and Spatial Statistics

Temporal processes

Thordis Thorarinsdottir

2024-01-18

Temporal processes

Data collected over time

- Past, present, future, change
- Is the temporal aspect important?

Temporal processes

Data collected over time

- Past, present, future, change
- Is the temporal aspect important?

We focus on a statistical approach to modelling

- Variability through randomness
- Learning dynamical structure from data

Characterization of temporal processes

Denoted $\mathbf{Y}(\cdot)$ or $\{\mathbf{Y}(r) : r \in D_t\}$

- r time index
- $\mathbf{Y}(r)$ multivariate, deterministic or stochastic ($Y(r)$ if univariate)

Characterization of temporal processes

Denoted $\mathbf{Y}(\cdot)$ or $\{\mathbf{Y}(r) : r \in D_t\}$

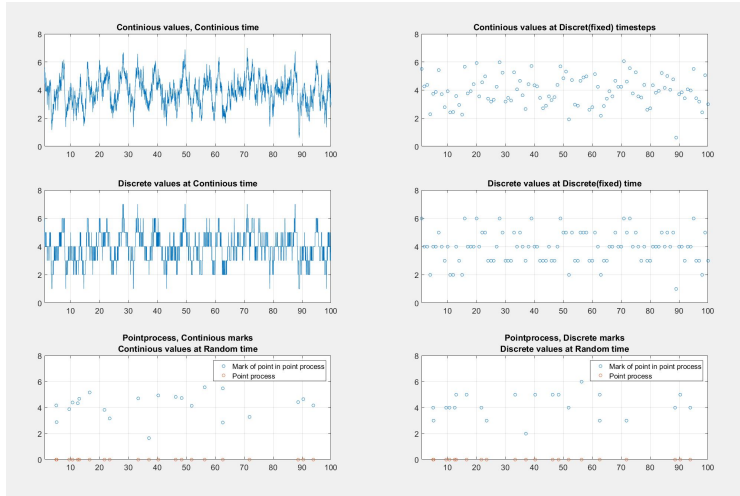
- r time index
- $\mathbf{Y}(r)$ multivariate, deterministic or stochastic ($Y(r)$ if univariate)

D_t specifies type of process

- Continuous time process, $\mathbf{Y}(t)$: $D_t \subseteq (-\infty, \infty)$ or $[0, \infty)$
- Discrete time process, $\mathbf{Y}_t$: $D_t \subseteq \mathbb{N}$ or $\mathbb{N}^+$
- Point process: Random times with the value at each time called *mark*

Continuous time may be measured or modeled discrete

Examples of temporal processes



Process model as a hierarchical model

One time series $\mathbf{Y} = \{Y_t : t = 1, \dots, T\}$

$$[\mathbf{Y}] = [Y_1, \dots, Y_T] = [Y_1][Y_2|Y_1][Y_3|Y_2, Y_1] \cdots [Y_T|Y_{T-1}, \dots, Y_1]$$

Does the past influence the present? A *first order Markov assumption* is

$$[Y_t|Y_{t-1}, \dots, Y_1] = [Y_t|Y_{t-1}]$$

Process model as a hierarchical model

One time series $\mathbf{Y} = \{Y_t : t = 1, \dots, T\}$

$$[\mathbf{Y}] = [Y_1, \dots, Y_T] = [Y_1][Y_2|Y_1][Y_3|Y_2, Y_1] \cdots [Y_T|Y_{T-1}, \dots, Y_1]$$

Does the past influence the present? A *first order Markov assumption* is

$$[Y_t|Y_{t-1}, \dots, Y_1] = [Y_t|Y_{t-1}]$$

Two processes $\mathbf{X}, \mathbf{Y}$ with

$$[\mathbf{X}] = \prod_{t=1}^T [X_t|X_{t-1}], \quad [\mathbf{Y}|\mathbf{X}] = \prod_{t=1}^T [Y_t|Y_{t-1}, X_t, X_{t-1}]$$

Does X_t and/or X_{t-1} influence Y_t ? Easier to answer than: Does $\mathbf{X}$ influence $\mathbf{Y}$?

Time series model components

Assume $\{Y_t : t = 1, 2, \dots\}$, time series

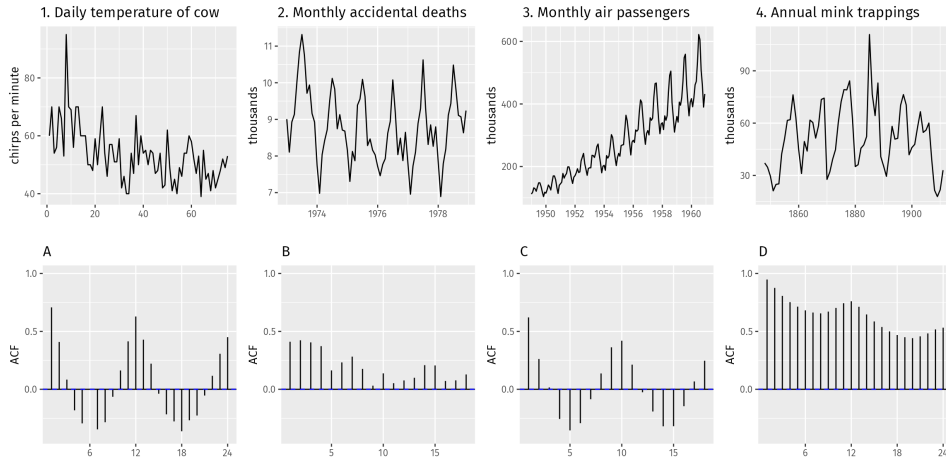
- *Mean function* $\mu_t = E(Y_t)$, deterministic
- *Autocovariance function*

$$C_Y(t, r) = \text{Cov}(Y_t, Y_r) = \mathbb{E}\{(Y_t - \mu_t)(Y_r - \mu_r)\}, \quad t, r \in D_t$$

- *Variance function* $\sigma_t^2 = \text{Var}(Y_t) = C_Y(t, t)$
- *Autocorrelation function (ACF)*

$$\rho_Y(t, r) = \frac{C_Y(t, r)}{\sqrt{C_Y(t, t)C_Y(r, r)}} \in [-1, 1]$$

Guess the ACF (Hyndman and Athanasopoulos, 2018)



Covariance functions are *nonnegative definite functions*

Consider $V = \sum_{i=1}^N a_i Y_{t_i} = \mathbf{a}^T \mathbf{Y}_{\mathbf{I}}$ for $\mathbf{I} = \{t_i\}_{i=1}^N$ any finite subset of D_t

- We have

$$\begin{aligned} \text{var}[V] &= \sum_{i=1}^N a_i^2 \text{Var}(Y_{t_i}) + 2 \sum_{1 \leq i < j \leq N} a_i a_j \text{cov}(Y_{t_i}, Y_{t_j}) \\ &= \mathbf{a}^T \mathbf{C}_{\mathbf{Y}_{\mathbf{I}}} \mathbf{a} \end{aligned}$$

where $\mathbf{C}_{\mathbf{Y}_{\mathbf{I}}}$ is the covariance matrix of $\mathbf{Y}_{\mathbf{I}}$

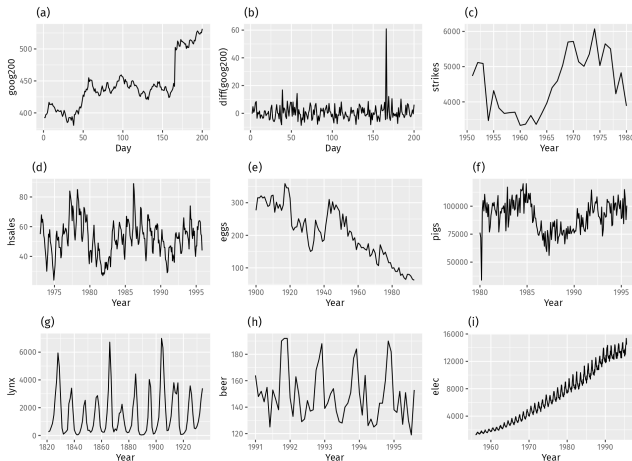
- Need $\mathbf{a}^T \mathbf{C}_{\mathbf{Y}_{\mathbf{I}}} \mathbf{a} \geq 0$ for all $\mathbf{a}$ (since $\text{var}(V) \geq 0$)
- $\mathbf{C}_{\mathbf{Y}_{\mathbf{I}}}$ specified through $C_Y(\cdot)$
- A function satisfying this requirement is a *nonnegative definite function*

Stationarity

- *Weak/second order stationarity* requires
 1. $\mathbb{E}(Y_t) = \mu$, for all $t \in D_t$
 2. $\text{Cov}(Y_t, Y_r) = C_Y(t - r)$ for all $t, r \in D_t$
- *Strong stationarity* requires
 1. For any $(t_1, \dots, t_m)$ and any τ

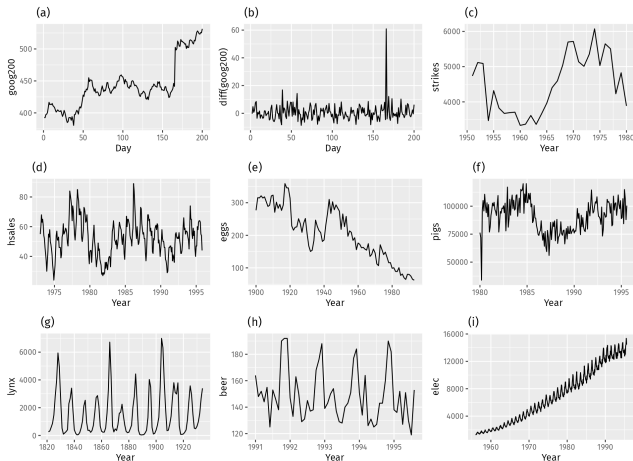
$$[Y_{t_1}, \dots, Y_{t_m}] = [Y_{t_1+\tau}, \dots, Y_{t_m+\tau}]$$

Which series is stationary? (Hyndman and Athanasopoulos, 2018)



- (a) Google stock price
- (b) Daily change in Google stock price
- (c) Annual number of strikes in the US
- (d) Monthly sales of new one-family houses in the US
- (e) Annual price of 12 eggs in the US
- (f) Monthly number of pigs slaughtered in Australia
- (g) Annual total of lynx trapped in a Canadian district
- (h) Monthly Australian beer production
- (i) Monthly Australian electricity production

Which series is stationary? (Hyndman and Athanasopoulos, 2018)



- (d), (h) and (i) have seasonality
- (a), (c), (e), (f) and (i) have trends and changing levels, (i) also has an increasing variance
- Thus, only (b) and (g) are stationary (the cycles in the lynx series are aperiodic)

Modelling temporal dependence

Assume $Y_1, \dots, Y_n$ time series

Simple model

$$Y_1 \sim \text{Gau}(\mu, \sigma^2 / (1 - \alpha^2))$$

$$Y_t = \mu + \alpha(Y_{t-1} - \mu) + \delta_t, \quad \delta_t \sim \text{Gau}(0, \sigma^2), \quad t = 2, \dots$$

This is an *autoregressive model of order 1*, AR(1)

- $Y_t \sim N(\mu, \sigma^2 / (1 - \alpha^2))$ for all t
- Independence if $\alpha = 0$

AR(1) - alternative formulation

The simple model formulation implies $\mathbf{Y} = (Y_1, \dots, Y_n)$ is multivariate Gaussian with

$$\mathbb{E}[Y_t] = \mu$$

$$\text{Var}[Y_t] = \sigma^2 / (1 - \alpha^2)$$

$$\text{Cor}[Y_t, Y_{t+\tau}] = \alpha^\tau = C_Y(\tau)$$

Can write

$$\mathbf{Y} \sim \text{Gau}(\mu \mathbf{1}, \Sigma)$$

$$\Sigma_{tt} = \text{Var}[Y_t]$$

$$\Sigma_{t,t+\tau} = \text{Cov}[Y_t, Y_{t+\tau}] = \Sigma_{tt} \text{Cor}[Y_t, Y_{t+\tau}]$$

AR(1) - extensions

The alternative formulation

$$\begin{aligned}\mathbf{Y} &\sim \text{Gau}(\mu \mathbf{1}, \Sigma) \\ \Sigma_{tt} &= \sigma^2 / (1 - a^2) \\ \Sigma_{t,t+\tau} &= \Sigma_{tt} a^\tau = \Sigma_{tt} C_Y(\tau)\end{aligned}$$

allows for extensions:

- Including covariates: $\mathbf{Y} \sim \text{Gau}(\mathbf{X}\boldsymbol{\beta}, \Sigma)$
- Other correlation structures: $\Sigma_{t,t+\tau} = \Sigma_{tt} C_Y(\tau)$
- Higher dimensional processes

Basic time series models

$\{W_t, t \in \mathbb{N}\}$ iid zero-mean random variables with $\text{var}(W_t) = \sigma_w^2$

- *White noise* $Y_t = W_t$

$$C(\tau) = \begin{cases} \sigma_w^2, & \tau = 0 \\ 0, & \tau = \pm 1, \pm 2, \dots \end{cases}$$

Basic time series models

$\{W_t, t \in \mathbb{N}\}$ iid zero-mean random variables with $\text{var}(W_t) = \sigma_w^2$

- *White noise* $Y_t = W_t$

$$C(\tau) = \begin{cases} \sigma_w^2, & \tau = 0 \\ 0, & \tau = \pm 1, \pm 2, \dots \end{cases}$$

- *Random-walk*: $Y_t = Y_{t-1} + W_t$. Not stationary

Basic time series models

$\{W_t, t \in \mathbb{N}\}$ iid zero-mean random variables with $\text{var}(W_t) = \sigma_w^2$

- *White noise* $Y_t = W_t$

$$C(\tau) = \begin{cases} \sigma_w^2, & \tau = 0 \\ 0, & \tau = \pm 1, \pm 2, \dots \end{cases}$$

- *Random-walk*: $Y_t = Y_{t-1} + W_t$. Not stationary
- *Autoregressive AR(p)*: $Y_t = \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + W_t$

Basic time series models

$\{W_t, t \in \mathbb{N}\}$ iid zero-mean random variables with $\text{var}(W_t) = \sigma_w^2$

- *White noise* $Y_t = W_t$

$$C(\tau) = \begin{cases} \sigma_w^2, & \tau = 0 \\ 0, & \tau = \pm 1, \pm 2, \dots \end{cases}$$

- *Random-walk*: $Y_t = Y_{t-1} + W_t$. Not stationary
- *Autoregressive AR(p)*: $Y_t = \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + W_t$
- *Moving average MA(q)*: $Y_t = W_t + \beta_1 W_{t-1} + \dots + \beta_q W_{t-q}$
- *ARMA(p,q)*: $Y_t = \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + W_t + \beta_1 W_{t-1} + \dots + \beta_q W_{t-q}$

Basic time series models

$\{W_t, t \in \mathbb{N}\}$ iid zero-mean random variables with $\text{var}(W_t) = \sigma_w^2$

- *White noise* $Y_t = W_t$

$$C(\tau) = \begin{cases} \sigma_w^2, & \tau = 0 \\ 0, & \tau = \pm 1, \pm 2, \dots \end{cases}$$

- *Random-walk*: $Y_t = Y_{t-1} + W_t$. Not stationary
- *Autoregressive AR(p)*: $Y_t = \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + W_t$
- *Moving average MA(q)*: $Y_t = W_t + \beta_1 W_{t-1} + \dots + \beta_q W_{t-q}$
- *ARMA(p,q)*: $Y_t = \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + W_t + \beta_1 W_{t-1} + \dots + \beta_q W_{t-q}$
- *ARIMA(p,d,q)*: $Y'_t = \alpha_1 Y'_{t-1} + \dots + \alpha_p Y'_{t-p} + W_t + \beta_1 W_{t-1} + \dots + \beta_q W_{t-q}$, where Y'_t is the d times differences series ("I" for "integrated")

Covariance functions for AR/MA processes

- AR(1): $C_Y(\tau) = \alpha_1^\tau$
- MA(1): $C_Y(\tau)$ given by

$$C_Y(\tau) = \sigma_w^2 \begin{cases} 1, & \tau = 0 \\ \beta_1, & \tau = 1 \\ 0, & \tau > 1 \end{cases}$$

In general

- AR processes: Decay exponentially
- MA processes: Zero outside lag

Simple parameter estimation

Assume second-order stationarity

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T Y_t$$

Sample mean

$$\widehat{C}(\tau) = \frac{1}{T} \sum_{t=1}^{T-\tau} (Y_{t+\tau} - \hat{\mu})(Y_t - \hat{\mu})$$

Empirical autocovariance function

Can divided by $T - \tau$ for unbiased estimate

Dividing by T gives a *nonnegative definite function*

$$\hat{\rho}(\tau) = \frac{\widehat{C}_Y(\tau)}{\widehat{C}_Y(0)} \in [-1, 1]$$

General parameter estimation

- Method-of-moments
 - Set of linear equations for AR models
- Maximum likelihood
 - Usually need numerical optimization for maximizing the likelihood

Example (Hyndman and Athanasopoulos, 2018)

```
library(forecast)
library(ggplot2)
library(fpp)
autoplot(elecequip)
```

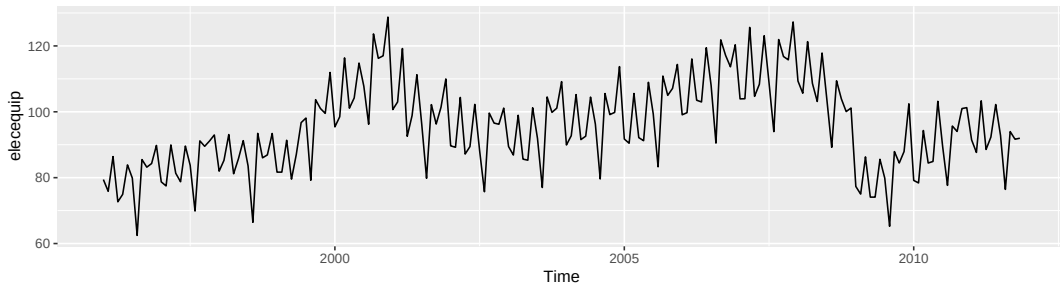


Figure 1: Electrical equipment orders index in the Euro area

Example cont. (Hyndman and Athanasopoulos, 2018)

```
elecequip %>% stl(s.window='periodic') %>% seasadj() -> eeadj  
autoplot(eeadj)
```

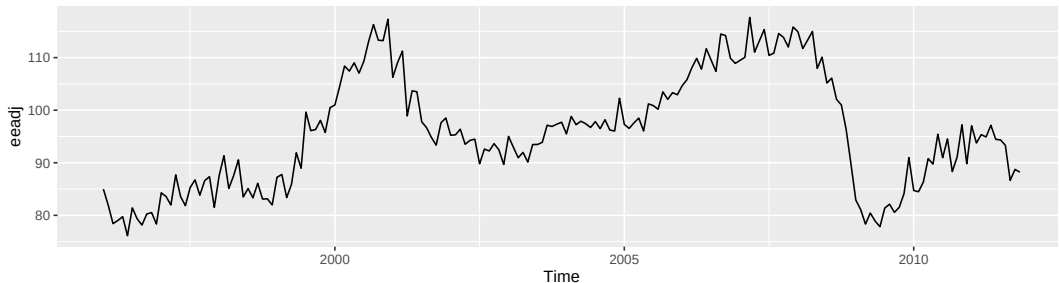


Figure 2: Seasonally adjusted electrical equipment orders index in the Euro area

Example cont. (Hyndman and Athanasopoulos, 2018)

```
eeadj %>% diff() %>% ggtsdisplay(main="")
```

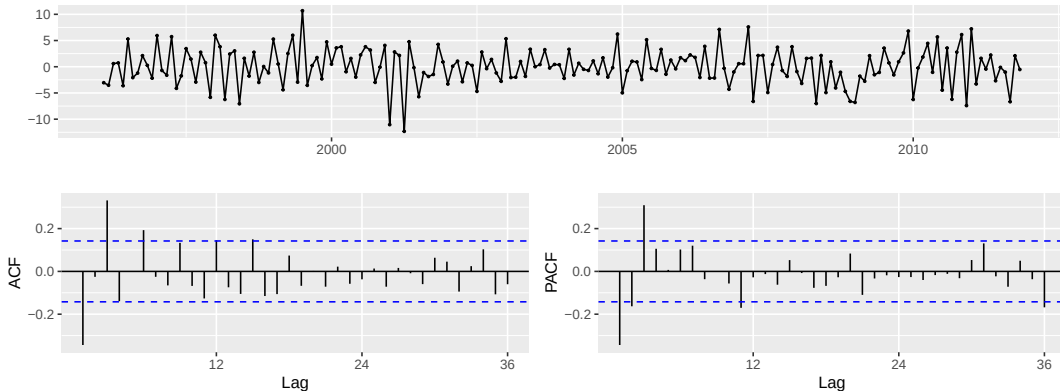


Figure 3: Time plot, ACF and PACF plots for the differenced seasonally adjusted data.

Example cont. (Hyndman and Athanasopoulos, 2018)

```
(fit <- Arima(eeadj, order=c(3,1,1)))
```

Series: eeadj

ARIMA(3,1,1)

Coefficients:

	ar1	ar2	ar3	ma1
	0.0519	0.1191	0.3730	-0.4542
s.e.	0.1840	0.0888	0.0679	0.1993

$\sigma^2 = 9.737$: log likelihood = -484.08

AIC=978.17 AICc=978.49 BIC=994.4

Example cont. (Hyndman and Athanasopoulos, 2018)

```
checkresiduals(fit)
```

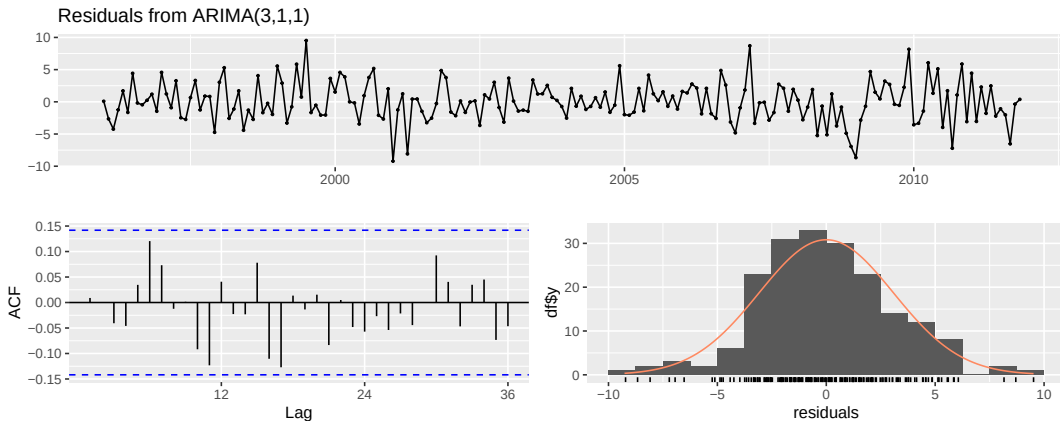


Figure 4: Residual plots for the ARIMA(3,1,1) model.

Example cont. (Hyndman and Athanasopoulos, 2018)

```
autoplot(forecast(fit))
```

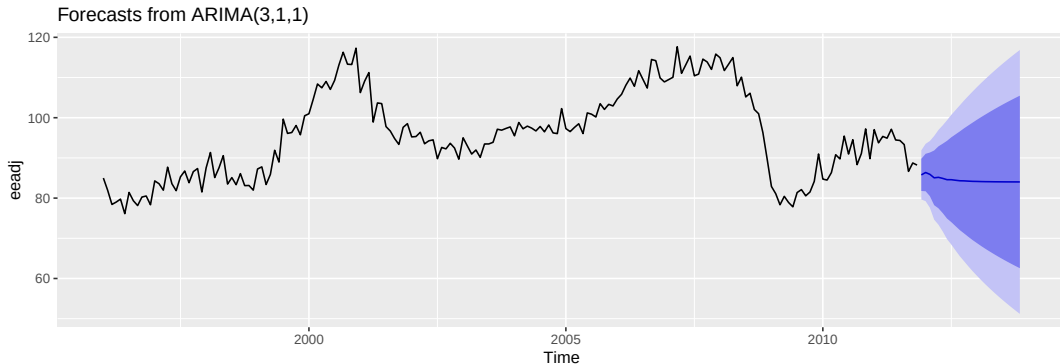


Figure 5: Forecasts for the seasonally adjusted electrical orders index.

References

Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. [OTexts.com/fpp2](https://otexts.com/fpp2). Accessed on 16.01.2024.

R Markdown

<https://vimeo.com/178485416>