Universitetet
i Stavanger

Time Series Analysis

24. august 2021

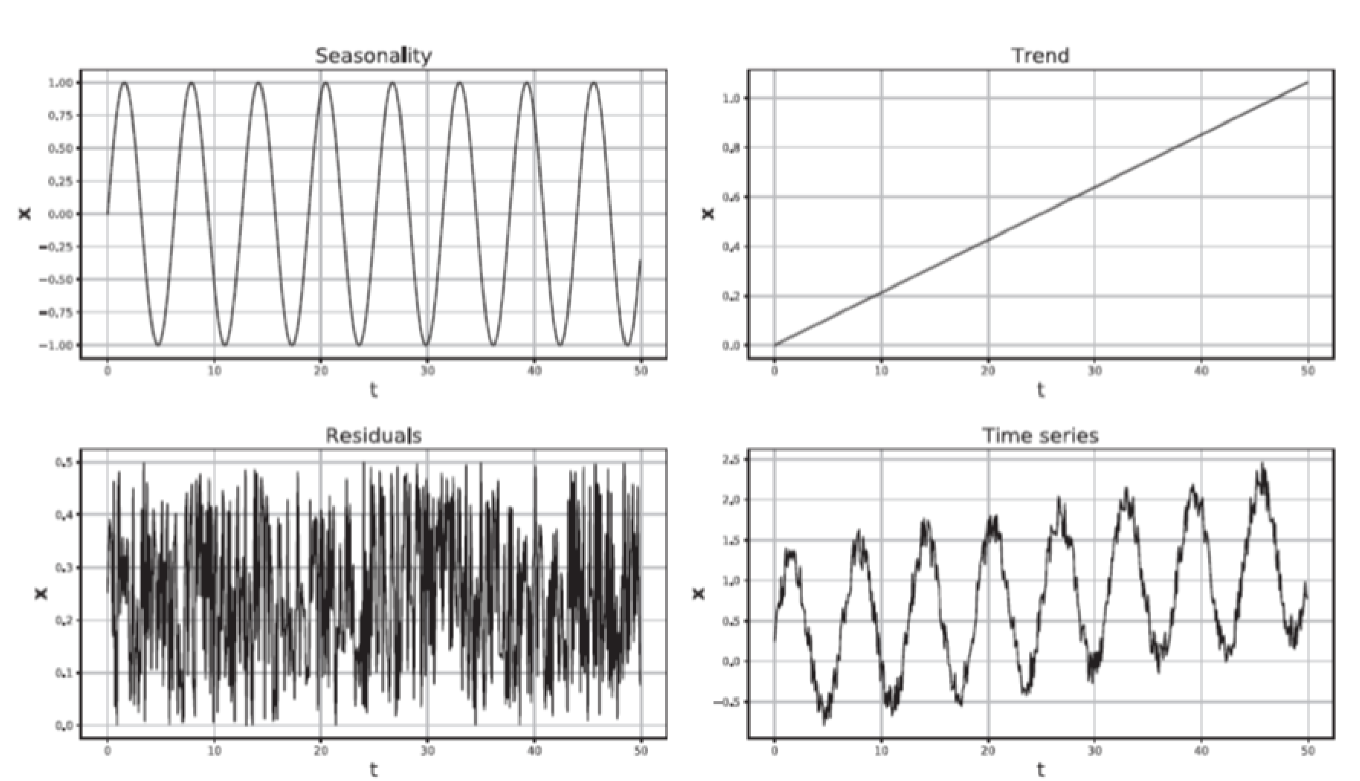
Overview

- Background foundations
- Classical approaches for time series forecasting
- Neural network models for forecasting

Background foundations

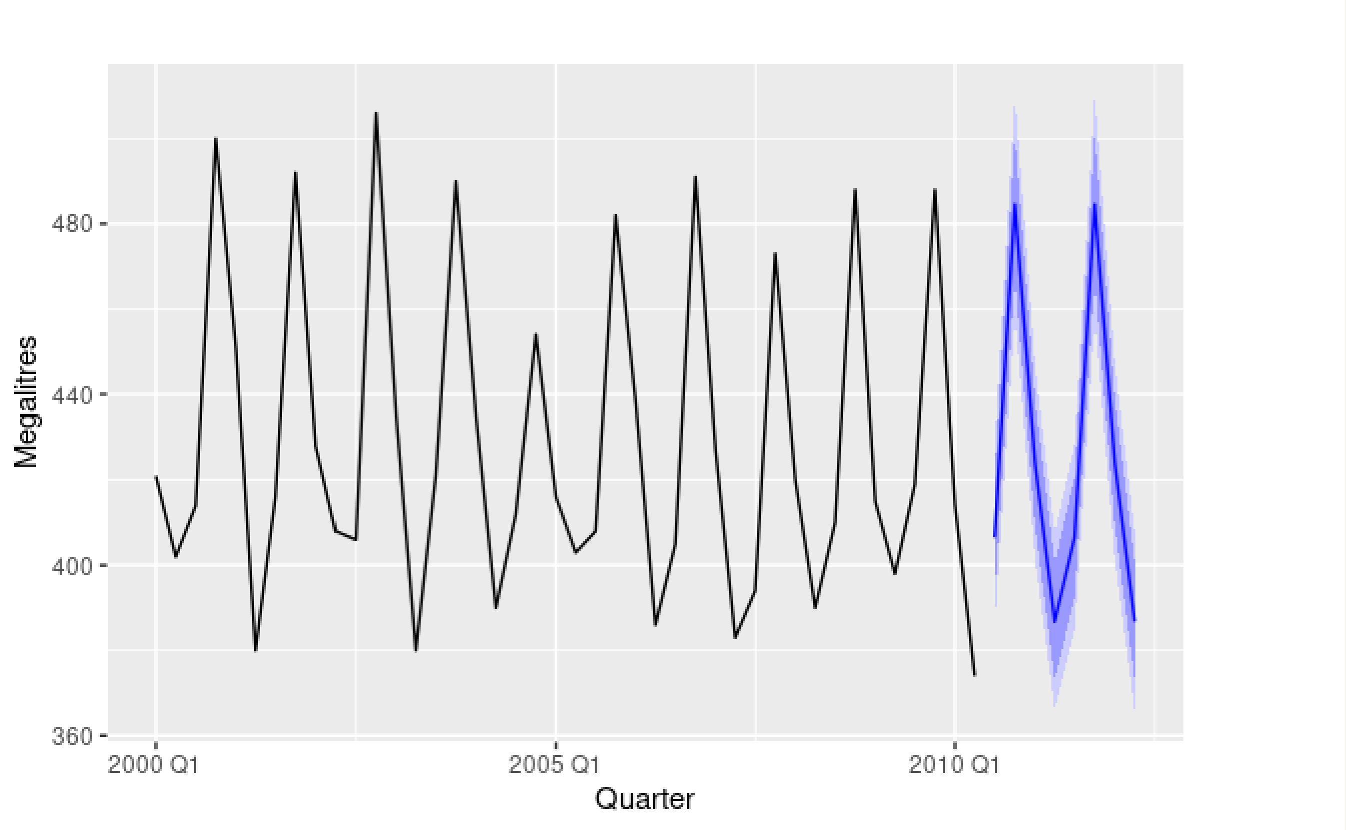
Time series

- Series of time varying values ordered in chronological sequence
- Often seen as consisting of a trend, seasonality and residual components



Torres JF, Hadjout D, Sebaa A, Martínez-Álvarez F, Troncoso A. Deep Learning for Time Series Forecasting: A Survey. Big Data. 2021 Feb 1;9(1):3-21.

- Aim is to estimate how a sequence of observation will continue into the future.



Hyndman, R.J., & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on 19.08.2021.

Classical approaches for time series forecasting

The basics

- Some methods used for forecasting include decomposition models, exponential smoothing models and ARIMA models.
- Model for forecasting the hourly electricity demand, ED.
- Predictor variables may be included (1)
- The demand can also form a time series which we could use in a forecasting (2)
- model to predict future demand:

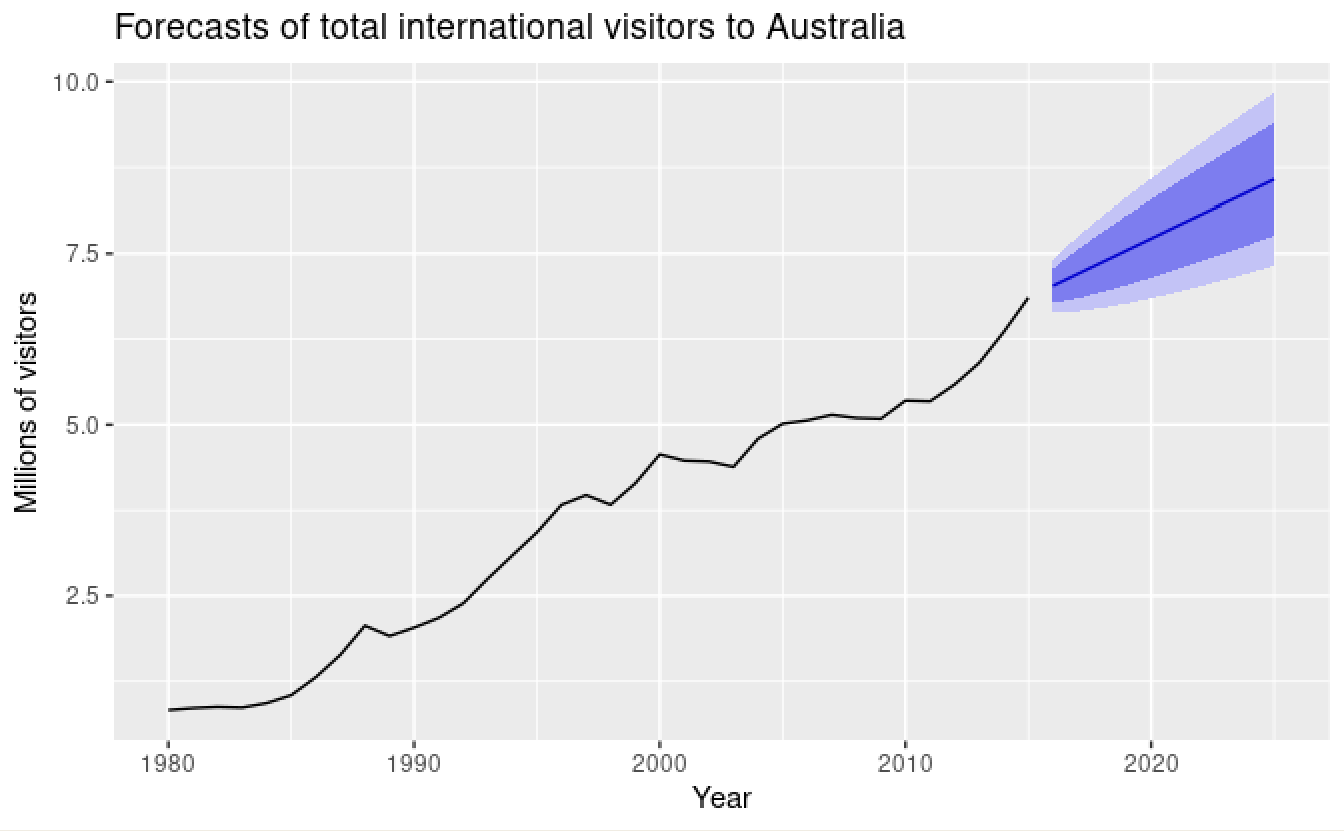
$$ED = f(\text{current temperature, population, time of day, day of week, error}).$$

$$ED_{t+1} = f(ED_t, ED_{t-1}, ED_{t-2}, \ldots).$$

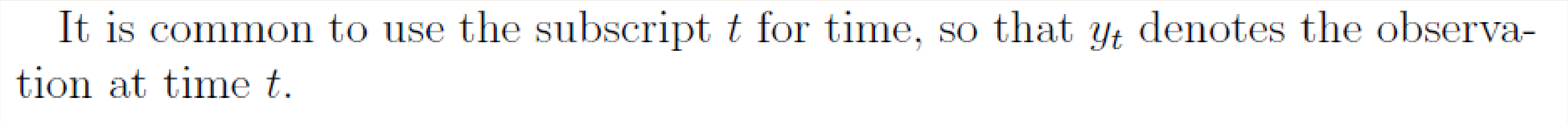
A statistical perspective

- We consider the time series values we want to predict as random variables.
- The variance of the variable will decrease the closer it is to the current observed values.
- The forcasting model aims to predict the mean in the possible range of values.
- A prediction interval gives the a range of values and the probability that the random variable will take a value in this region.

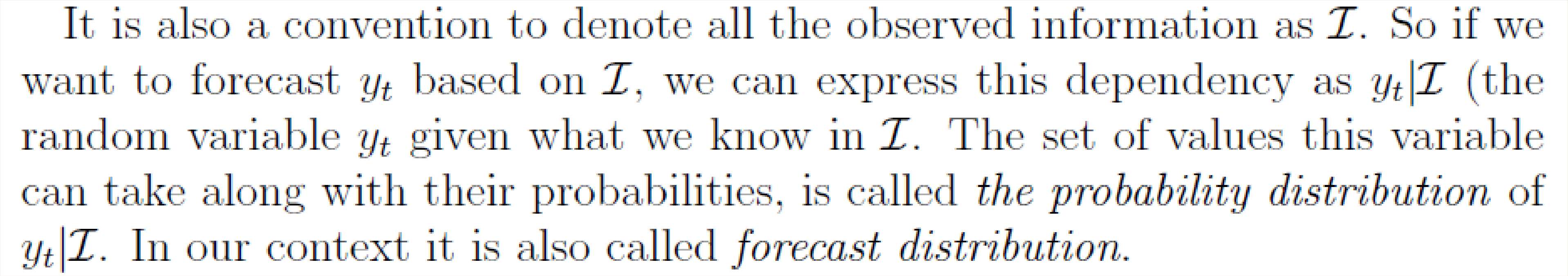
- Figure 3 shows a time series of number of visitors (black) along with a ten year forecast showing 80% ann 90% prediction intervals in dark and light shades respectively.



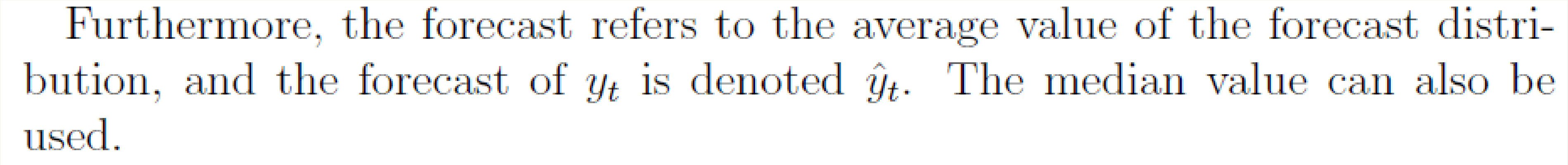
Hyndman, R.J., & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on 19.08.2021.



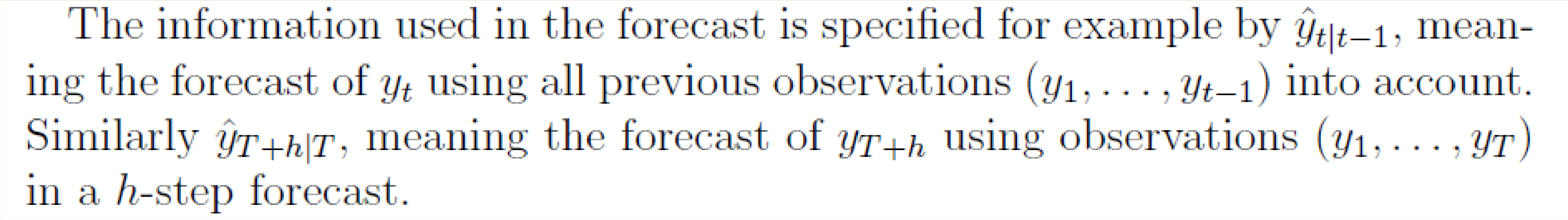
It is common to use the subscript t for time, so that y_t denotes the observation at time t.



It is also a convention to denote all the observed information as $\mathcal{I}$. So if we want to forecast y_t based on $\mathcal{I}$, we can express this dependency as $y_t|\mathcal{I}$ (the random variable y_t given what we know in $\mathcal{I}$. The set of values this variable can take along with their probabilities, is called *the probability distribution* of $y_t|\mathcal{I}$. In our context it is also called *forecast distribution.*



Furthermore, the forecast refers to the average value of the forecast distribution, and the forecast of y_t is denoted $\hat{y}_t$. The median value can also be used.



The information used in the forecast is specified for example by $\hat{y}_{t|t-1}$, meaning the forecast of y_t using all previous observations $(y_1, \ldots, y_{t-1})$ into account. Similarly $\hat{y}_{T+h|T}$, meaning the forecast of y_{T+h} using observations $(y_1, \ldots, y_T)$ in a h-step forecast.

Time series decomposition

- It is usual to decompose according to an additive (1)
- Or alternatively multiplicative assumption (2)

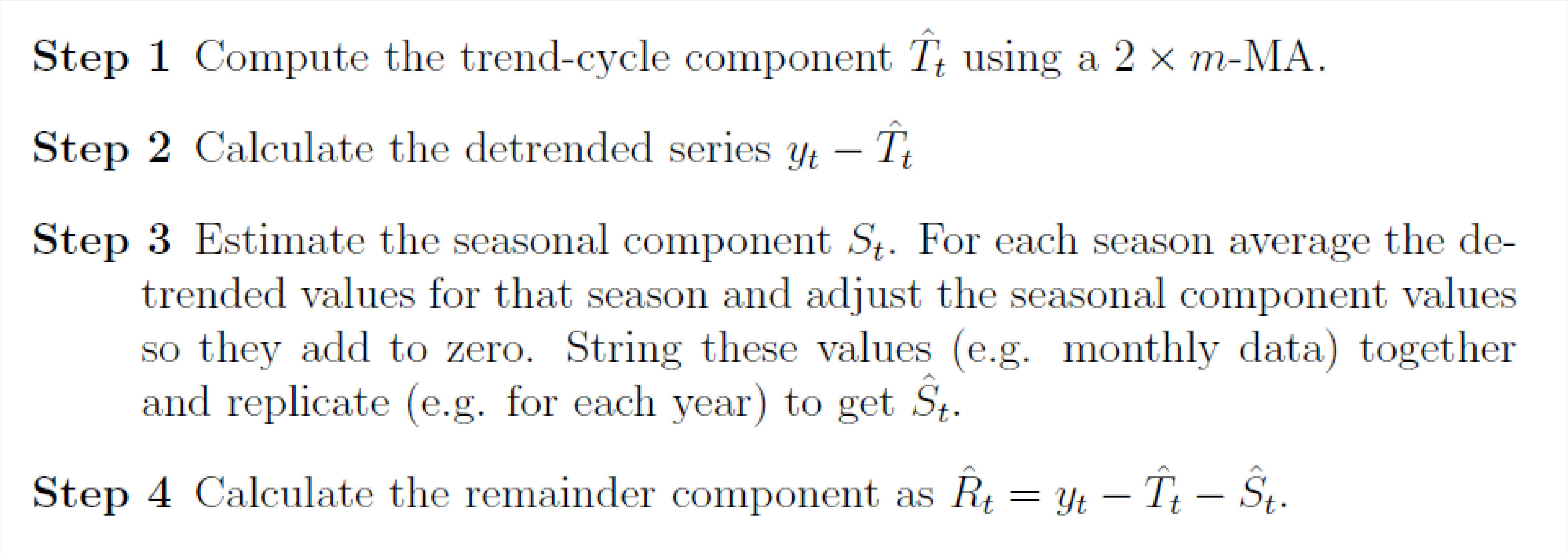
$$y_t = S_t + T_t + R_t$$

$$y_t = S_t \times T_t \times R_t$$

- When we want to decompose a time series, a classical way is to estimate the trend-cycle by using a moving average.

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^{k} y_{t+j},$$

- The classical decomposition techniques are 100 years old, but serves the purpose of illustrating the fundamental concepts.

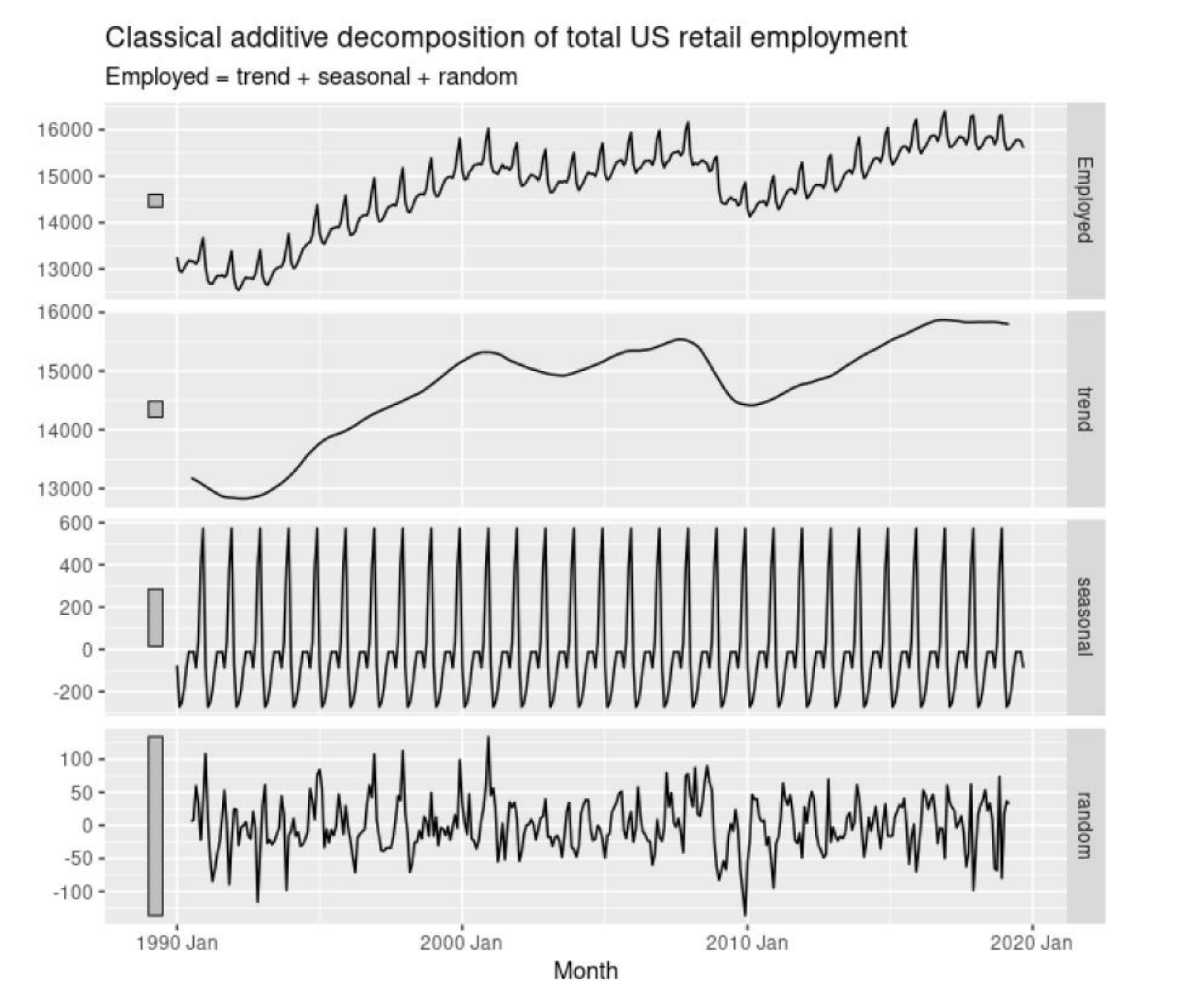


Step 1 Compute the trend-cycle component $\hat{T}_t$ using a $2 \times m$-MA.

Step 2 Calculate the detrended series $y_t - \hat{T}_t$

Step 3 Estimate the seasonal component S_t. For each season average the detrended values for that season and adjust the seasonal component values so they add to zero. String these values (e.g. monthly data) together and replicate (e.g. for each year) to get $\hat{S}_t$.

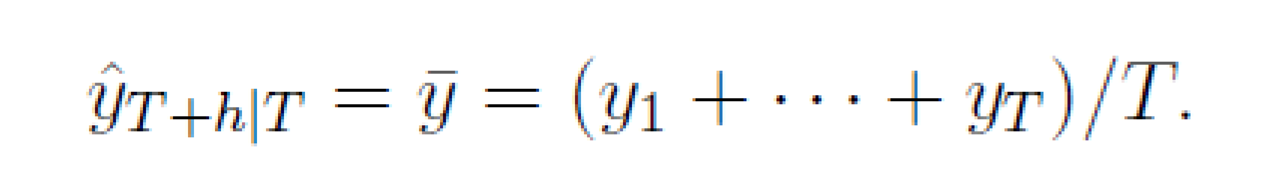
Step 4 Calculate the remainder component as $\hat{R}_t = y_t - \hat{T}_t - \hat{S}_t$.



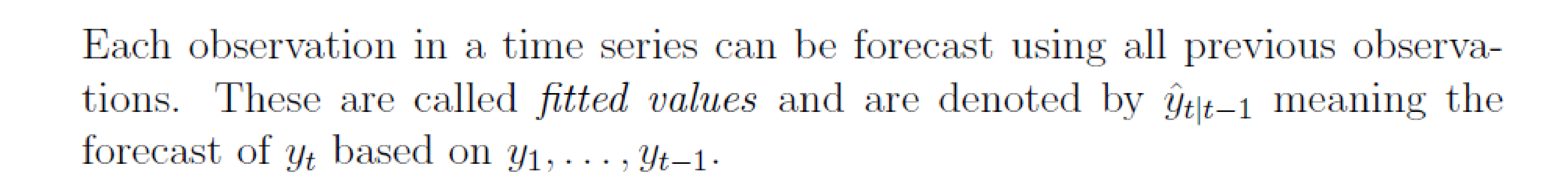
Hyndman, R.J., & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on 19.08.2021.

Simple forecasting methods

- In the average method the forecast of all future values are equal to the average of the historical data, i.e.



$$\hat{y}_{T+h|T} = \bar{y} = (y_1 + \cdots + y_T)/T.$$



Each observation in a time series can be forecast using all previous observations. These are called *fitted values* and are denoted by $\hat{y}_{t|t-1}$ meaning the forecast of y_t based on $y_1, \ldots, y_{t-1}$.

Residuals

- The residuals are the leftovers after fitting a model. (1)
- There are a number of metrics that are useful to evaluate the model performance.
  - The forecast error (2)
  - Mean absolute error (MEA) (3)
  - Root mean squared error (RMSE) (4)

$$e_t = y_t - \hat{y}_t.$$

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T},$$

$$MAE = \mathrm{mean}(|e_t|)$$

$$RMSE = \sqrt{\mathrm{mean}(e_t^2)}$$

Regression models

- In the design of time series regression models we assume there is a linear relationship between the time series of interest y (the forecast variable or regressand) and other time series x (the predictor variable(s) or regressor(s)).
- An example of a simple relationship where there is only a single predictor variable x (1)
- When there are two or more predictor variables we have a multiple regression model (2)

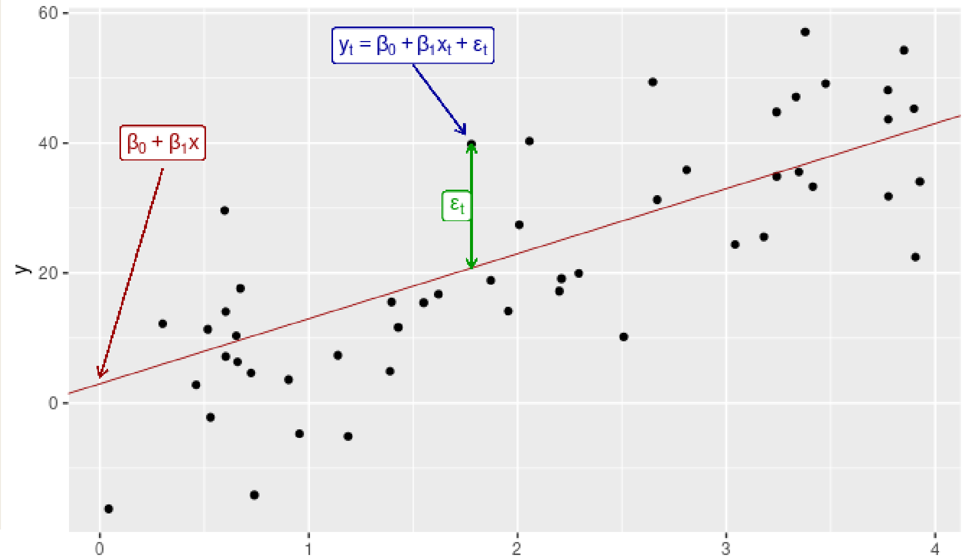
$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t.$$

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_k x_{k,t} + \varepsilon_t,$$

Least squares estimation

- The coefficients, β 0 , β 1 , . . . , β k , of the regression models are estimated by minimising the sum og squared errors

$$\sum_{t=1}^{T} \varepsilon_t^2 = \sum_{t=1}^{T} (y_t - \beta_0 - \beta_1 x_{1,t} - \beta_2 x_{2,t} - \cdots - \beta_k x_{k,t})^2.$$



Hyndman, R.J., & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on 19.08.2021.

- Predictions of y t can be obtained by using the estimated coefficients in (24) with error term set to zero

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1,t} + \hat{\beta}_2 x_{2,t} + \cdots + \hat{\beta}_k x_{k,t}.$$

ARIMA models

- In autoregressive models we forecast the variable of interest using a linear combination of past values of the variable (1).
- In a moving average model past forecast errors are used (2)

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}.$$

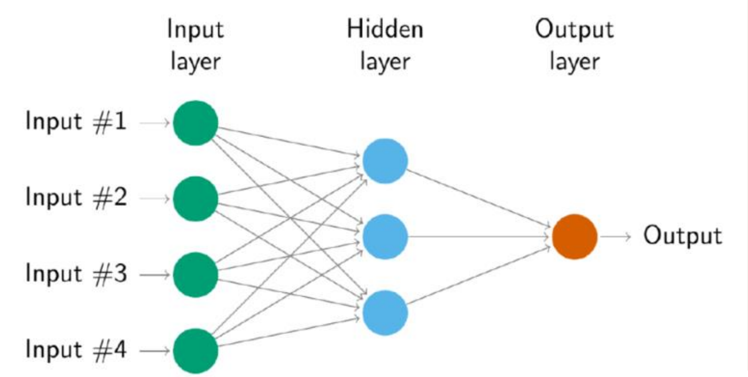
- If we combine the autoregressive and moving average models we get the nonseasonal AutoRegressive Integrated Moving Average (ARIMA) model. (1)
- After the model order has been identified, the next step is to estimate the parameters φ 1 , . . . , φ p , θ 1 , . . . , θ q . One approach to do this is to use maximum likelihood estimation (MLE). This corresponds to find the parameter values that minimizes (2)

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

$$\sum_{t=1}^{T} \varepsilon_t^2.$$

Neural network models for forecasting

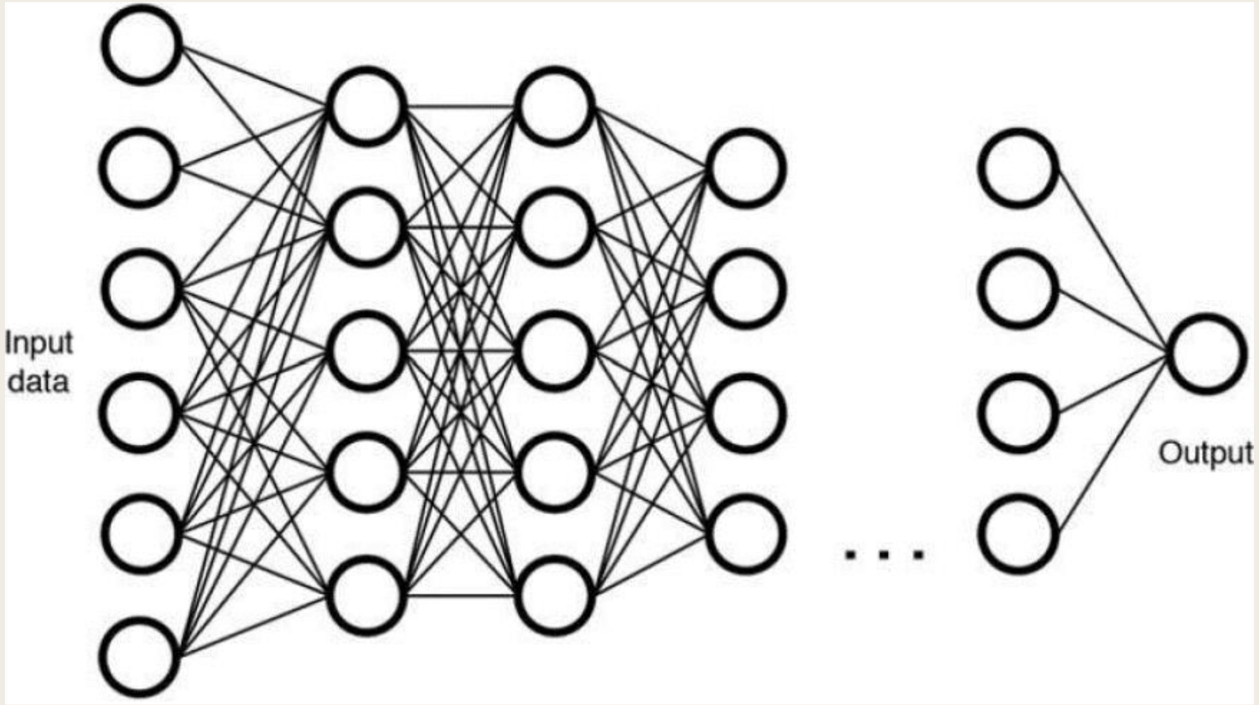
- With time series, lagged values of the time series can be used as inputs to the neural network.
- We look at a model for using the network for autoregressive modelling.
- We use the term NNAR(p, k) to indicate there are p lagged inputs and k nodes in the hidden layer.
- For example, in a N N AR(9, 5) model, the last nine observations (y t−1 , y t−2 , . . . , y t−9) are used as inputs for forecasting the output y t .



Hyndman, R.J., & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on 19.08.2021.

Deep feed forward neural network

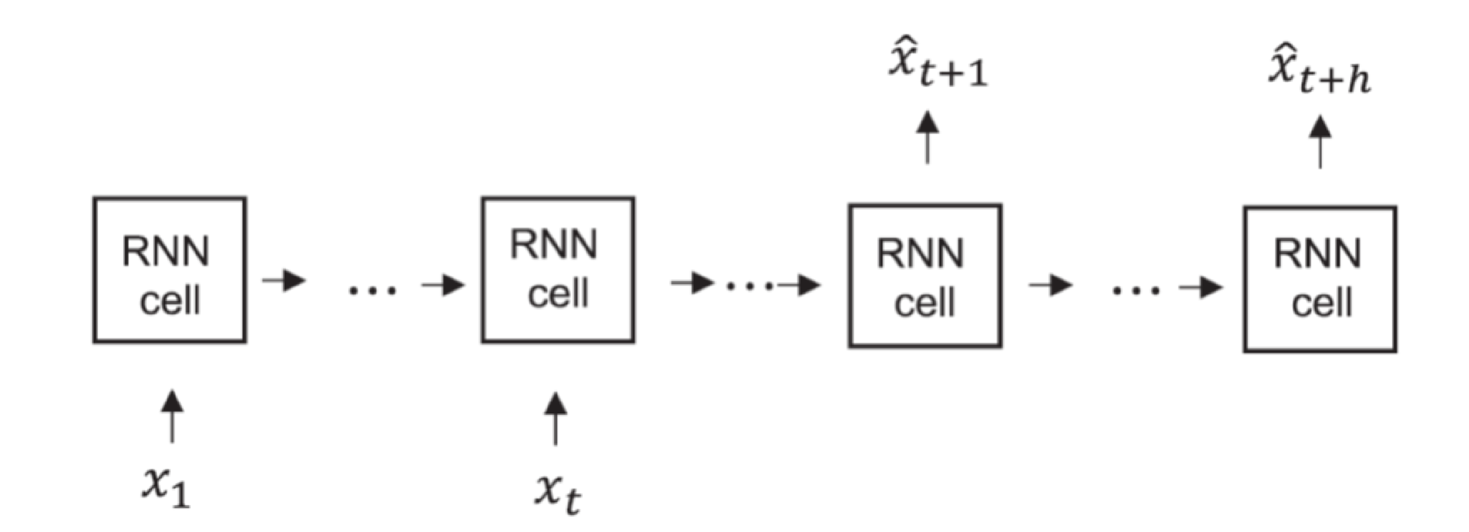
- Deep feed forward neural networks have gained wide acceptance as an efficient way of learning models. It is speculated that one can solve a problem more efficiently going for deep models rather than wide models.



Torres JF, Hadjout D, Sebaa A, Martínez-Álvarez F, Troncoso A. Deep Learning for Time Series Forecasting: A Survey. Big Data. 2021 Feb 1;9(1):3-21.

Recurrent neural networks

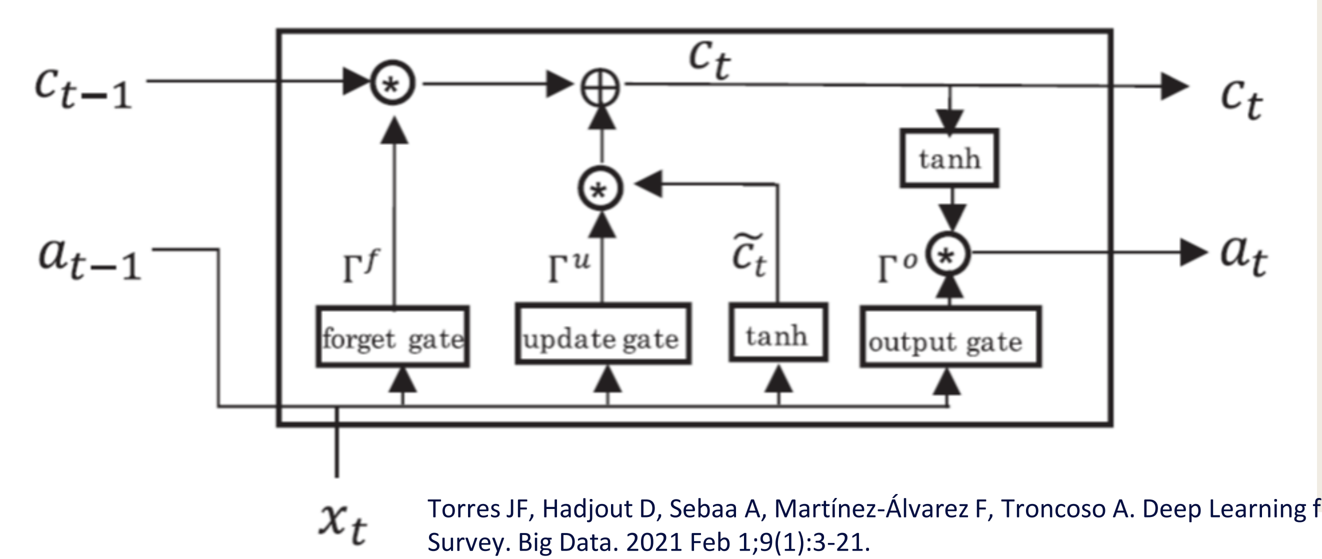
- Recurrent neural networks (RNNs) are designed to handle sequential data and are therefore well suited for handling time series as the time dependencies can be handled.



Torres JF, Hadjout D, Sebaa A, Martínez-Álvarez F, Troncoso A. Deep Learning for Time Series Forecasting: A Survey. Big Data. 2021 Feb 1;9(1):3-21.

Long short-term memory

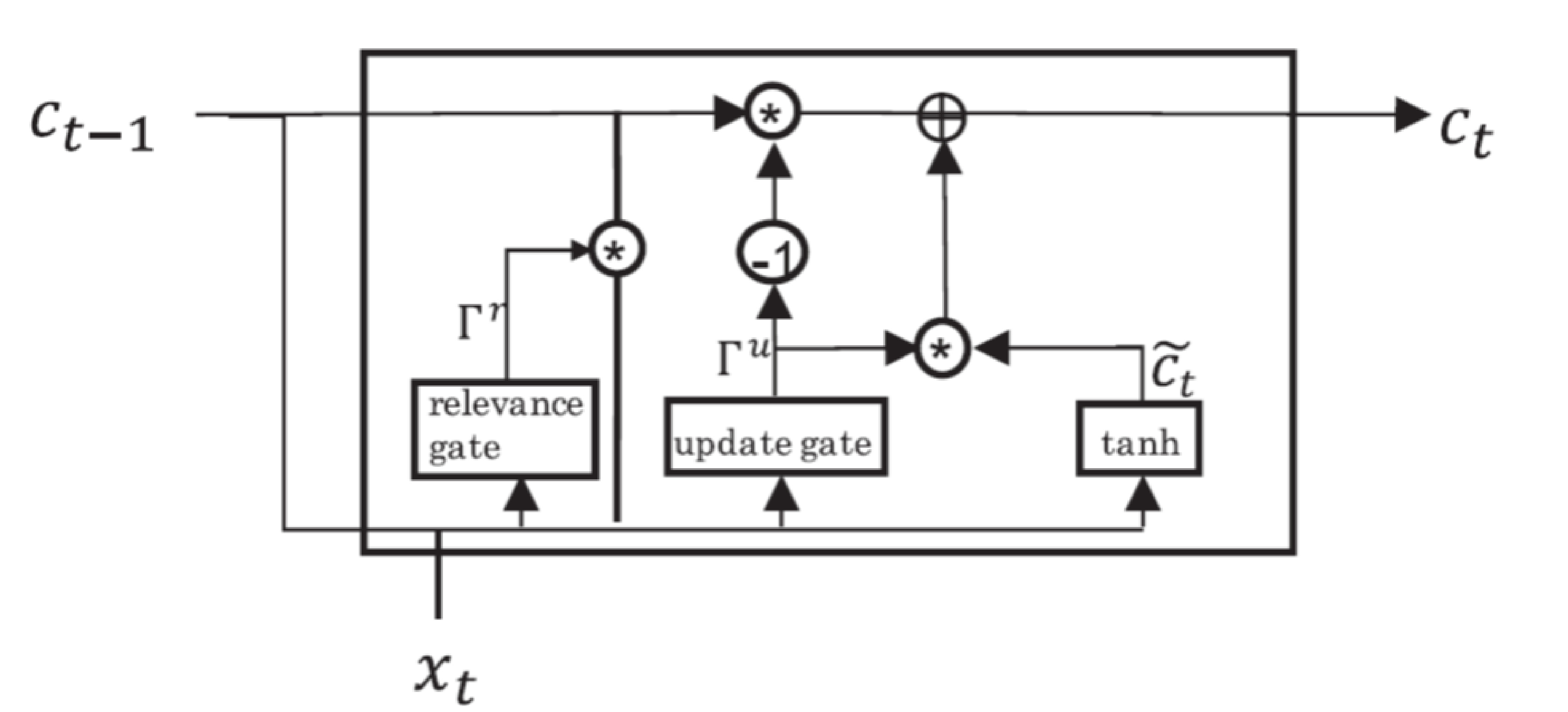
- The standard basic RNNs suffers from the vanishing gradient problem, which prevents learning for networks with a high number of layers. Therefore these networks might be considered as having a short term memory making it problematic to deal with long sequences.
- A hidden unit in an LSTM is shown in figure 11. The LSTM uses three gates, the Γf forget gate, Γu update gate, and Γo output gate.



Torres JF, Hadjout D, Sebaa A, Martínez-Álvarez F, Troncoso A. Deep Learning for Time Series Forecasting: A Survey. Big Data. 2021 Feb 1;9(1):3-21.

Gated recurrent units

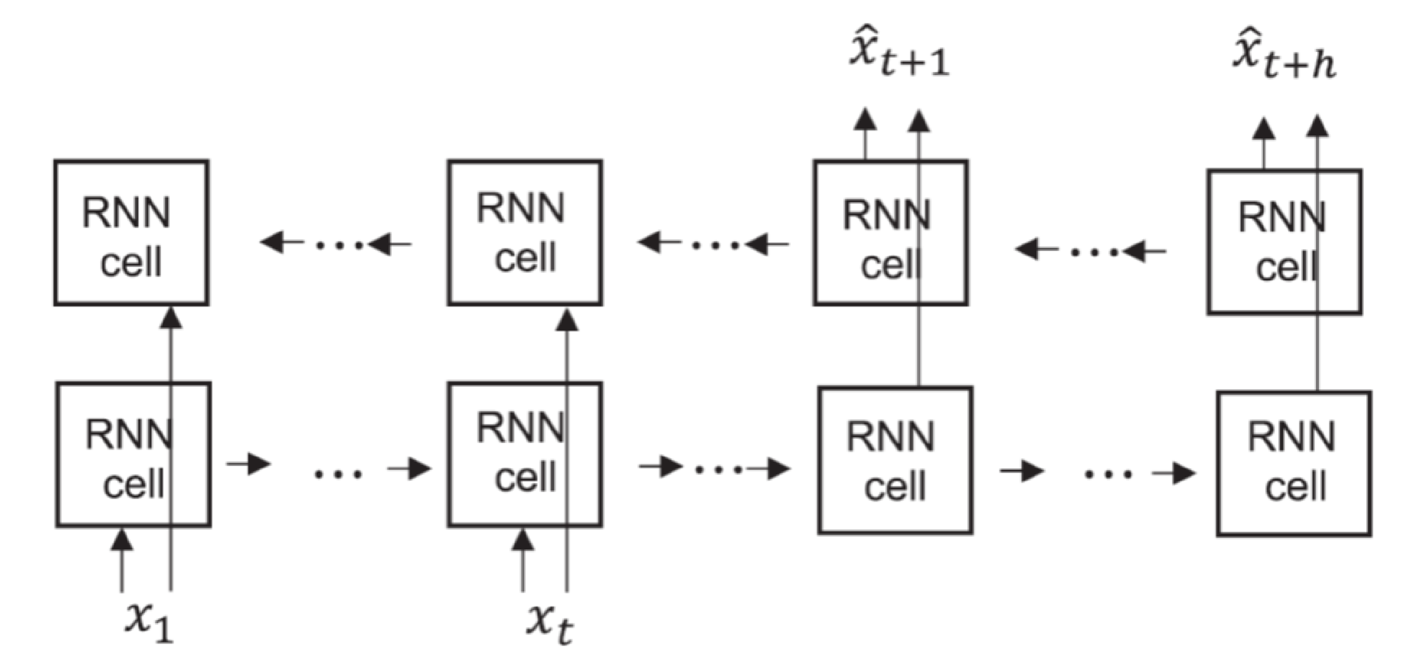
- Gated recurrent units (GRU) are simplifications of LSTM which are less computational expensive although not as powerful.
- The GRU only has two gates, a Γu update gate and a a Γr relevance gate.



Torres JF, Hadjout D, Sebaa A, Martínez-Álvarez F, Troncoso A. Deep Learning for Time Series Forecasting: A Survey. Big Data. 2021 Feb 1;9(1):3-21.

Bidirectional RNN

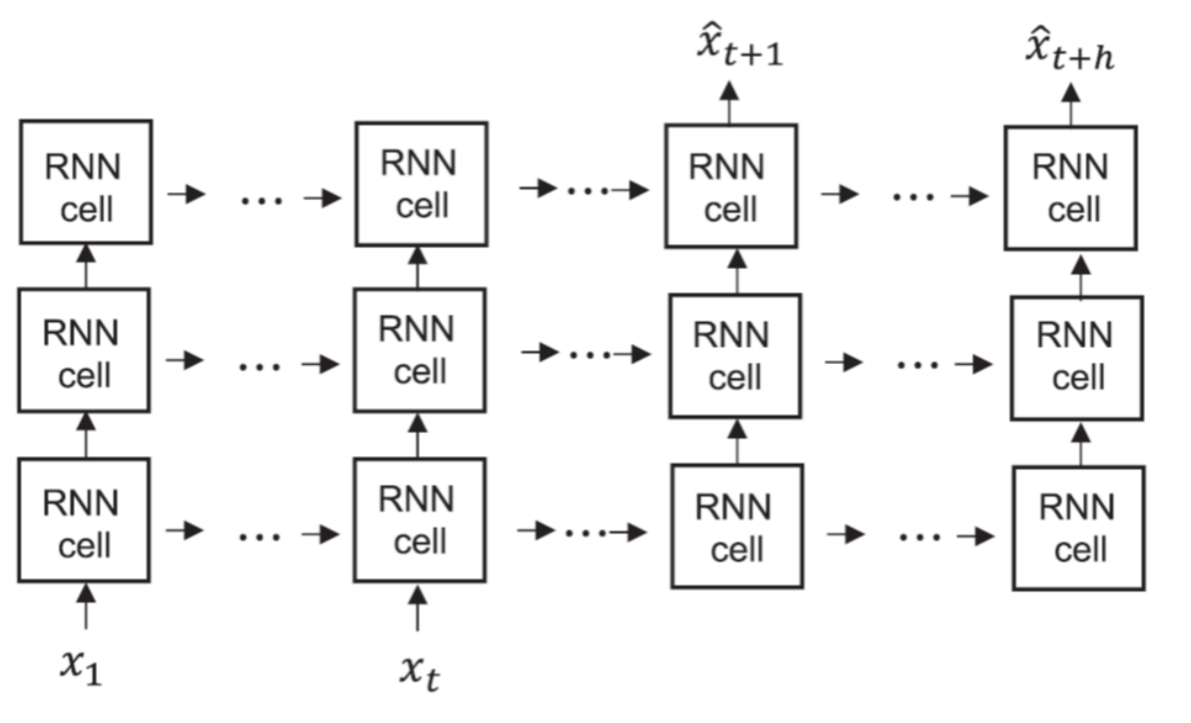
- In some cases, like natural language processing, information both preceeding and following the current instant of time is needed. Bidirectional recurrent neural networks (BRNNs) address this.



Torres JF, Hadjout D, Sebaa A, Martínez-Álvarez F, Troncoso A. Deep Learning for Time Series Forecasting: A Survey. Big Data. 2021 Feb 1;9(1):3-21.

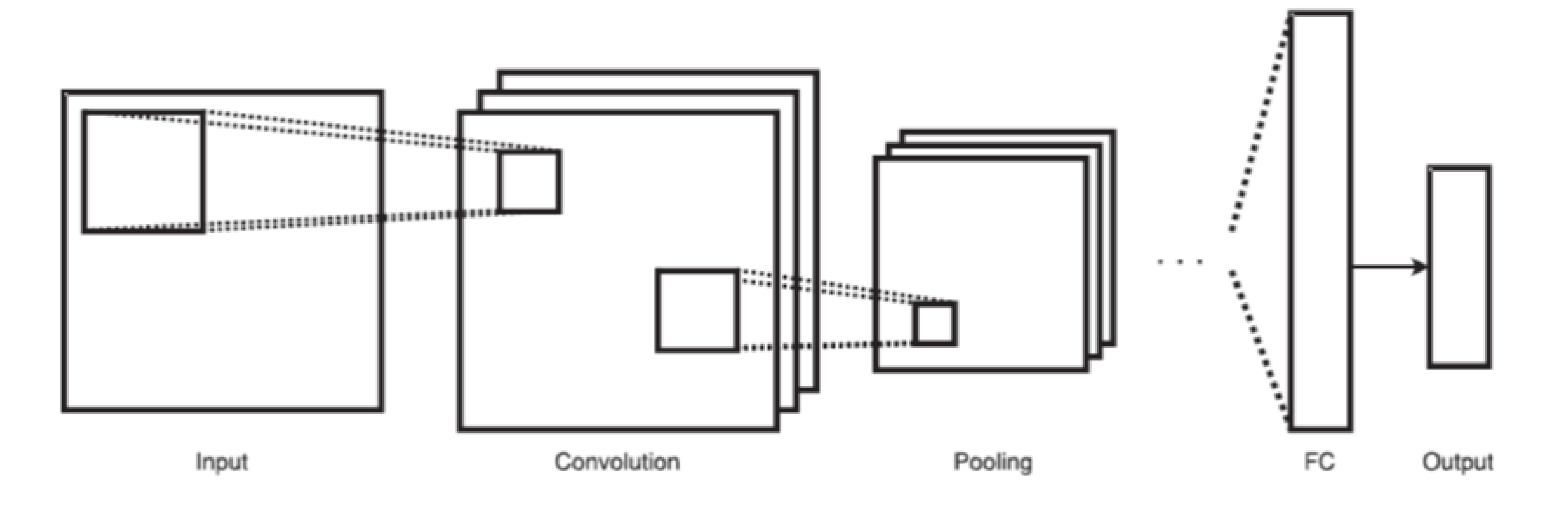
Deep recurrent neural network

- A deep recurrent neural network (DRNN) has more than one layer and is also called a stacked RNN. The hidden units can be standard RNN, GRU, or LSTM units and it can be unidirectional or bidirectional.



Convolutional neural network

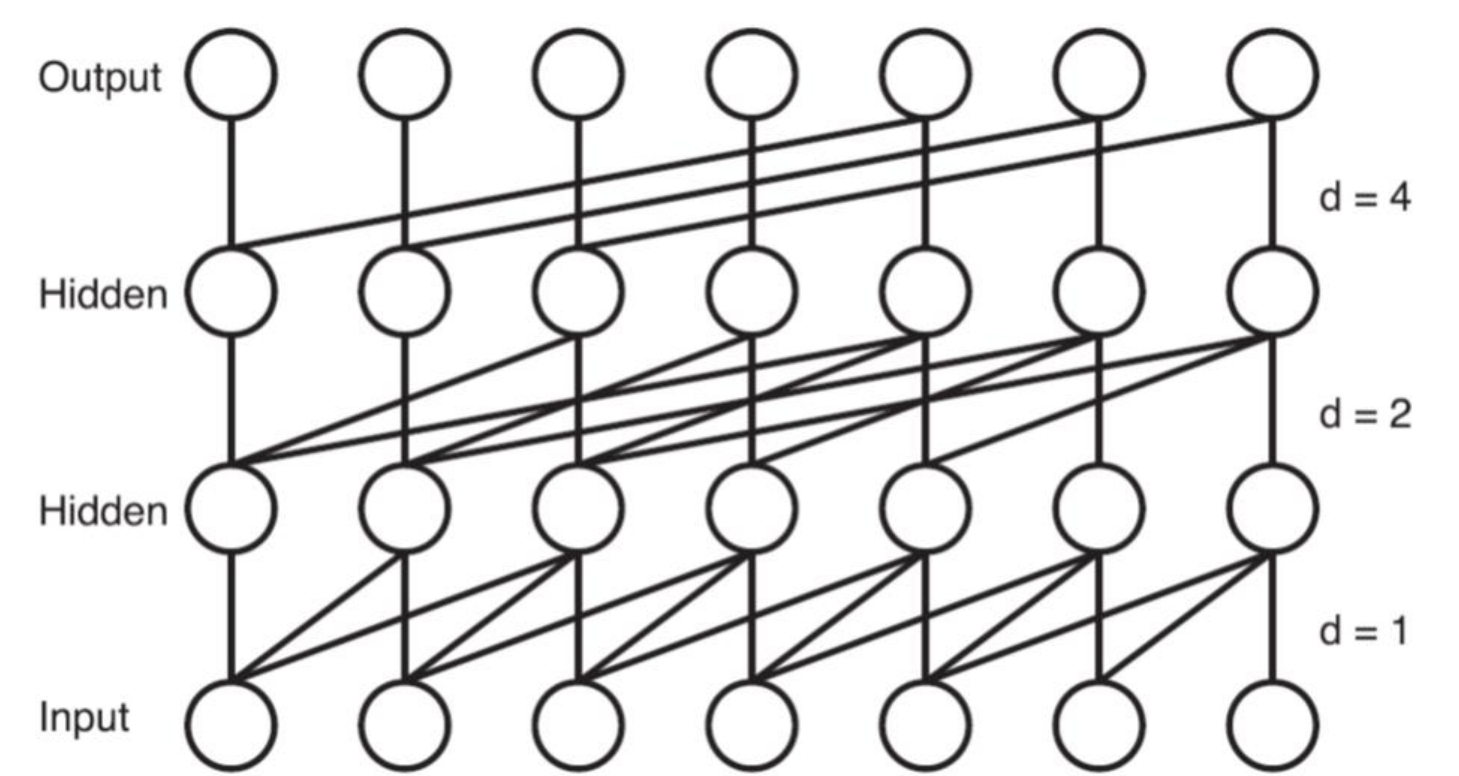
- A convolutional neural network (CNN) can be used for the feature extraction part



Torres JF, Hadjout D, Sebaa A, Martínez-Álvarez F, Troncoso A. Deep Learning for Time Series Forecasting: A Survey. Big Data. 2021 Feb 1;9(1):3-21.

Temporal CNN

- CNNs are typically used for images, but there has emerged variants for data sequences like the temporal convolutional networks (TCN).



Torres JF, Hadjout D, Sebaa A, Martínez-Álvarez F, Troncoso A. Deep Learning for Time Series Forecasting: A Survey. Big Data. 2021 Feb 1;9(1):3-21.