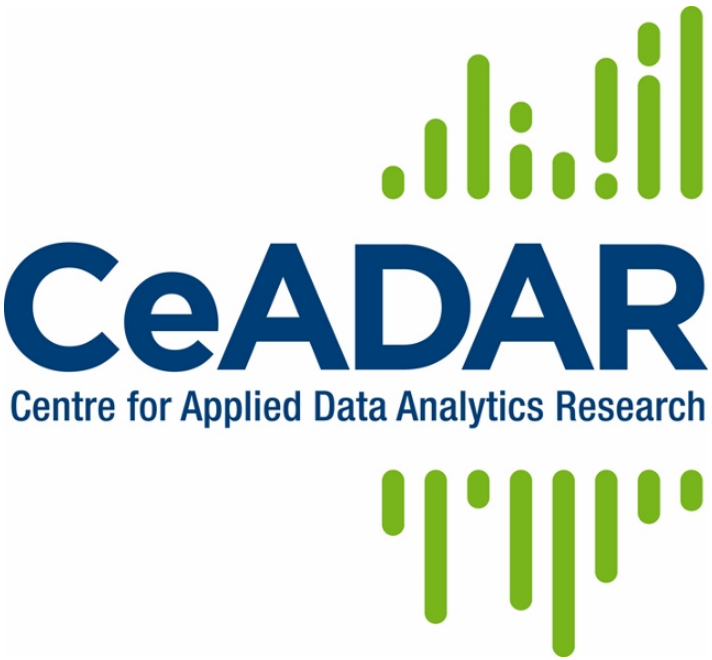


It's about Time

How to make sense of your time series data?

Bojan Božić

CeADAR @ Dublin Institute of Technology

Predict 2018

How do I know that I have time series data?

Definition

An ordered sequence of values of a variable at equally spaced time intervals.

Usage

1. Obtain an understanding of the underlying forces and structure that produced the observed data.
2. Fit a model and proceed to forecasting, monitoring or even feedback and feedforward control.

Applications

Economic Forecasting, Sales Forecasting, Budgetary Analysis, Stock Market Analysis, Yield Projections, Process and Quality Control, Inventory Studies, Workload Projections, Utility Studies, Census Analysis, and many, many more...

Life is a Sine Wave

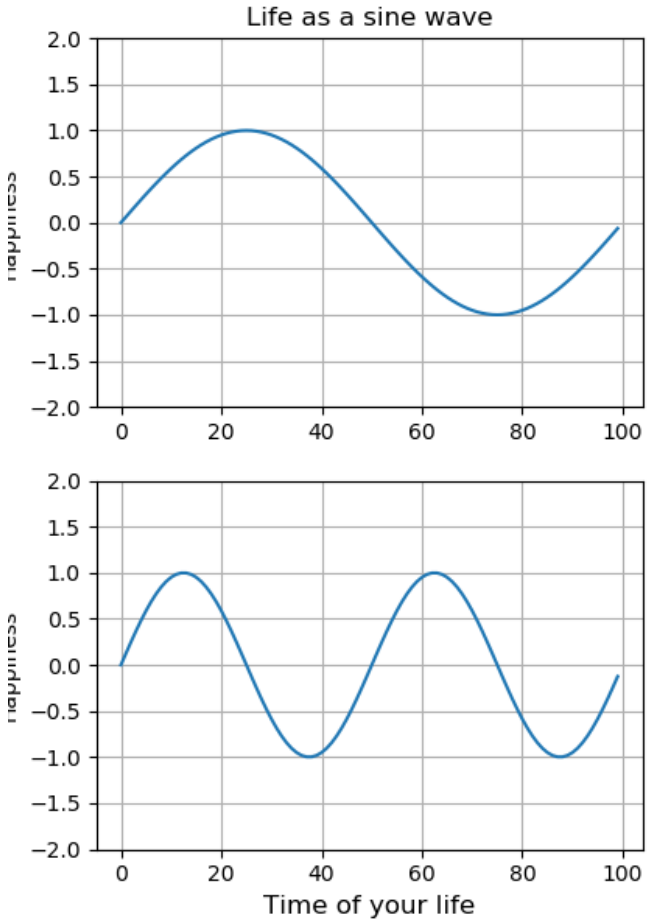


Figure 1: Life is different for everyone, but for all of us it's based on a sine wave.

Time Series Decomposition

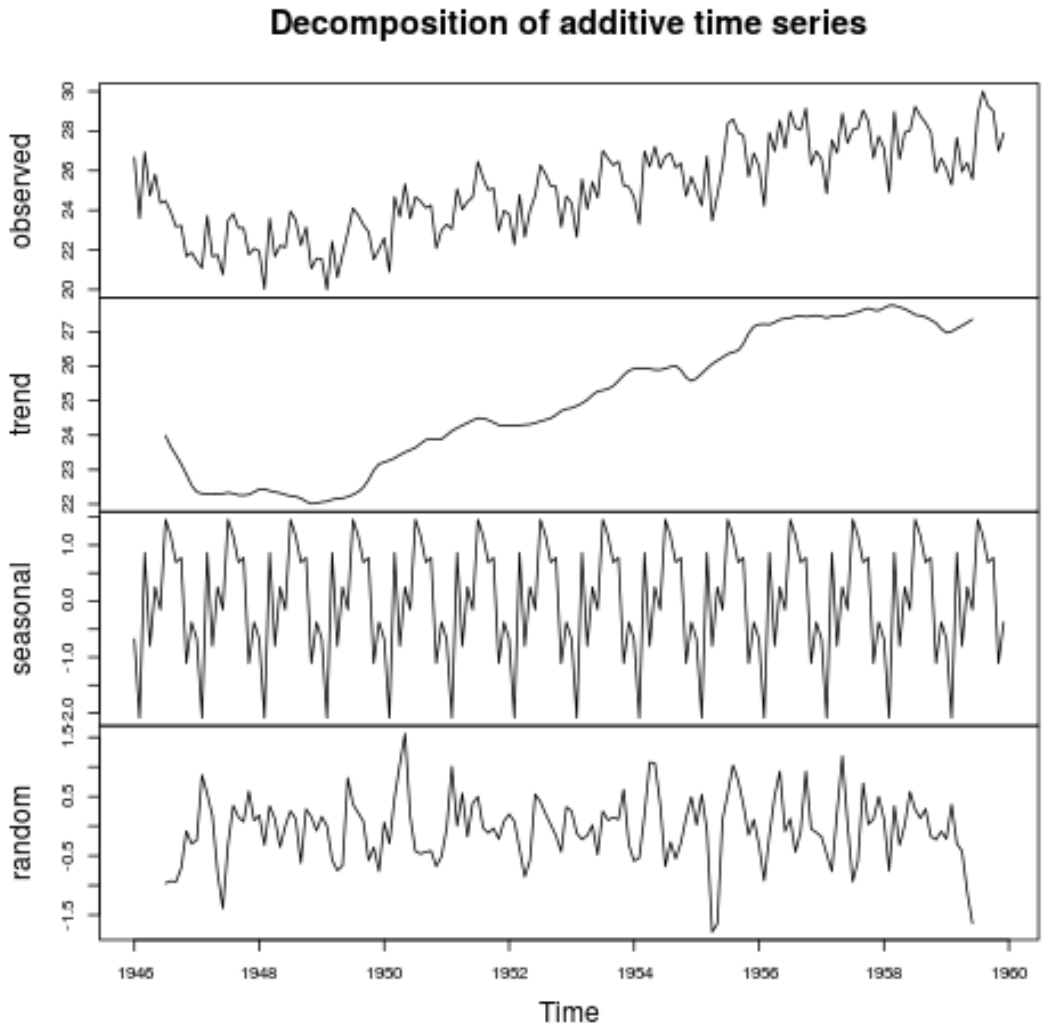
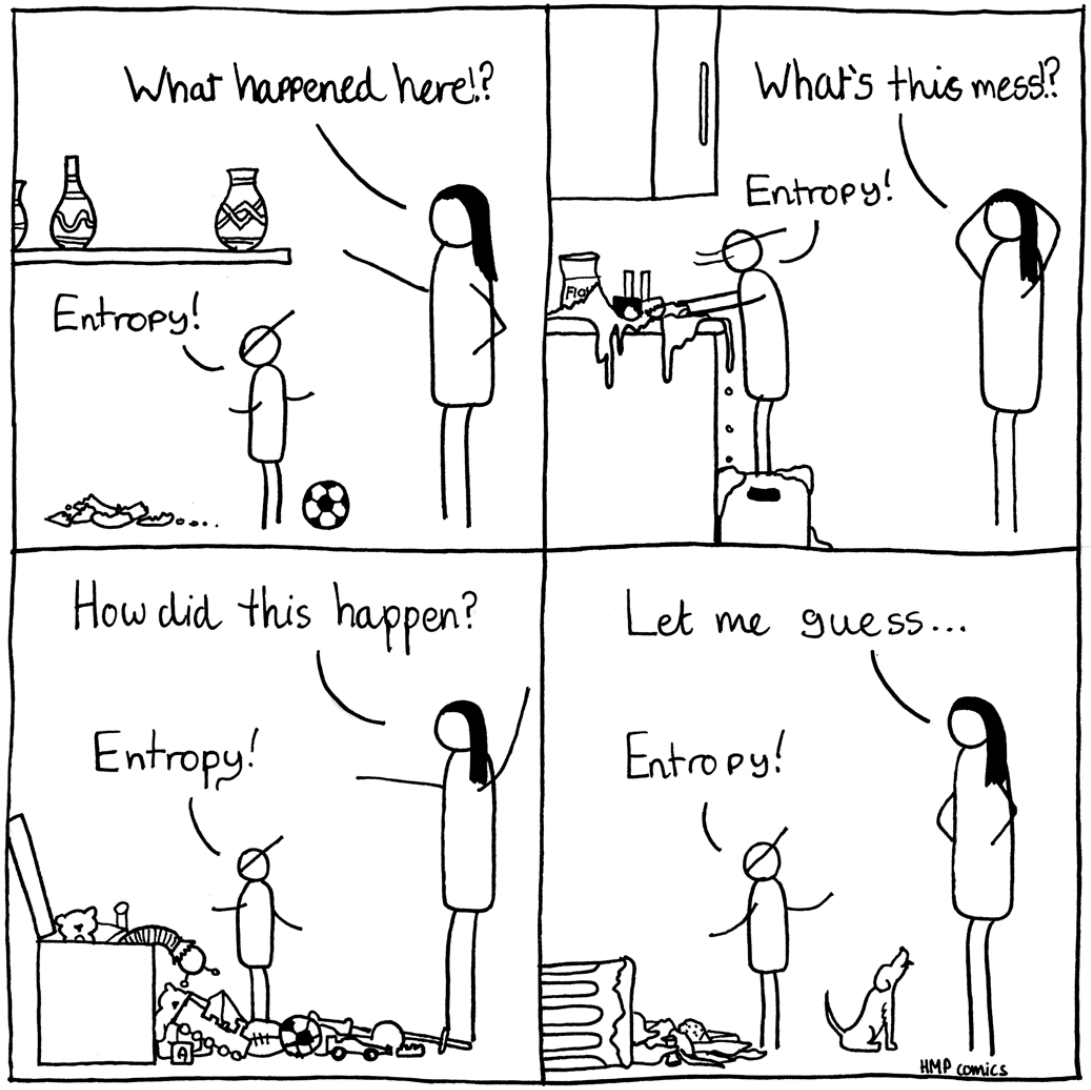


Figure 2: A decomposed population time series.

Case Study: Tractor and Farm Equipment Manufacturing Company

- Established after WWII
- Slow but consistent growth since inception
- Struggles to keep inventory and production costs down because of variability in demand
- Pressure to reduce production costs
- Want to understand impact of marketing on sales

Why is your bank password safe?



This is why we don't teach our children about entropy until much later...

Case Study Decomposition

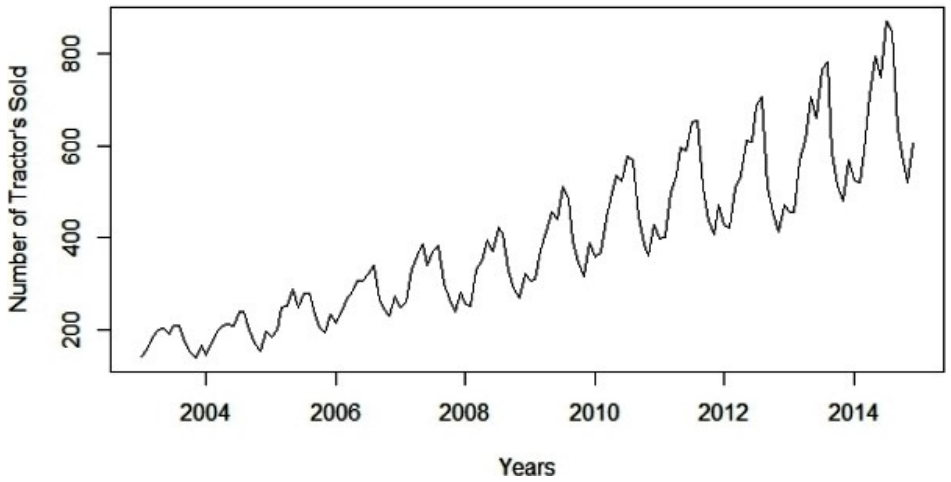


Figure 3: Original sales time series.

- **Trend** – overall direction of the series (upwards, downwards)
- **Seasonality** – monthly or quarterly patterns
- **Cycle** – long term business cycles
- **Remainder** – random noise left

Model

$$Y_t = f(Trend_t, Seasonality_t, Remainder_t)$$

Case Study Trend

Moving Average

$$MA = \frac{\sum_{i=-m}^{m} Y_{t+i}}{2m}$$

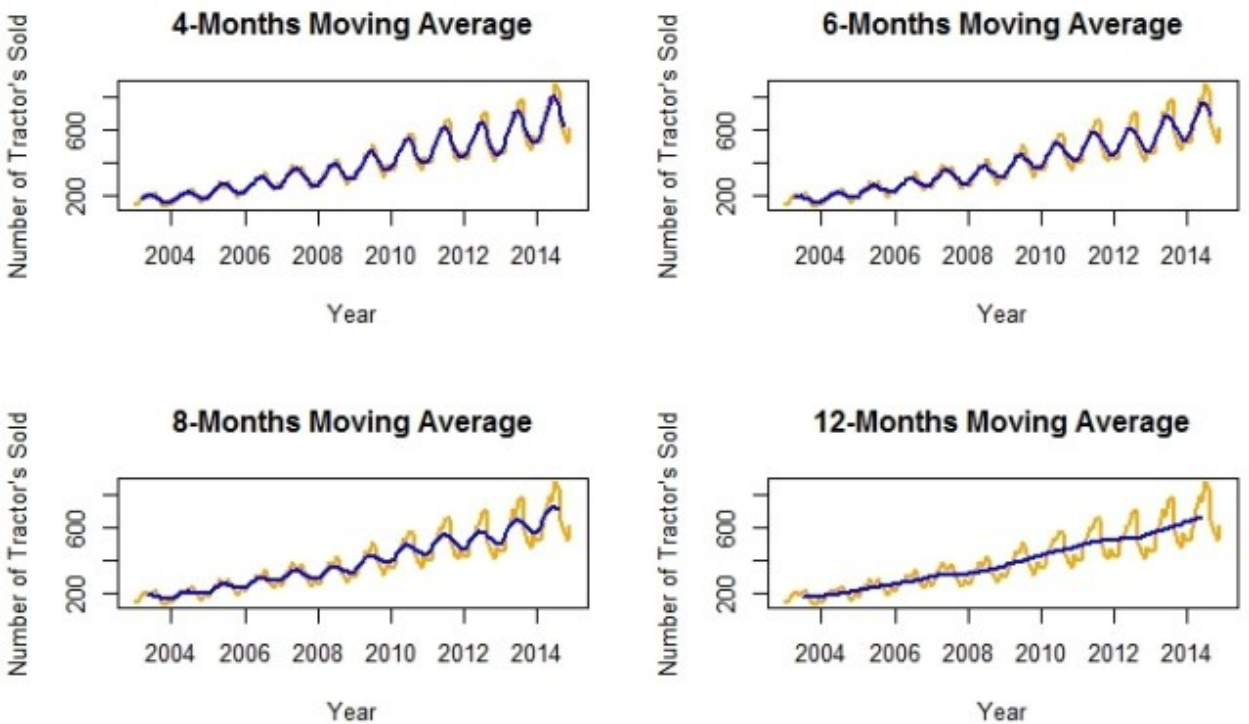


Figure 4: Smoothing the trend with moving average.

Case Study Seasonality

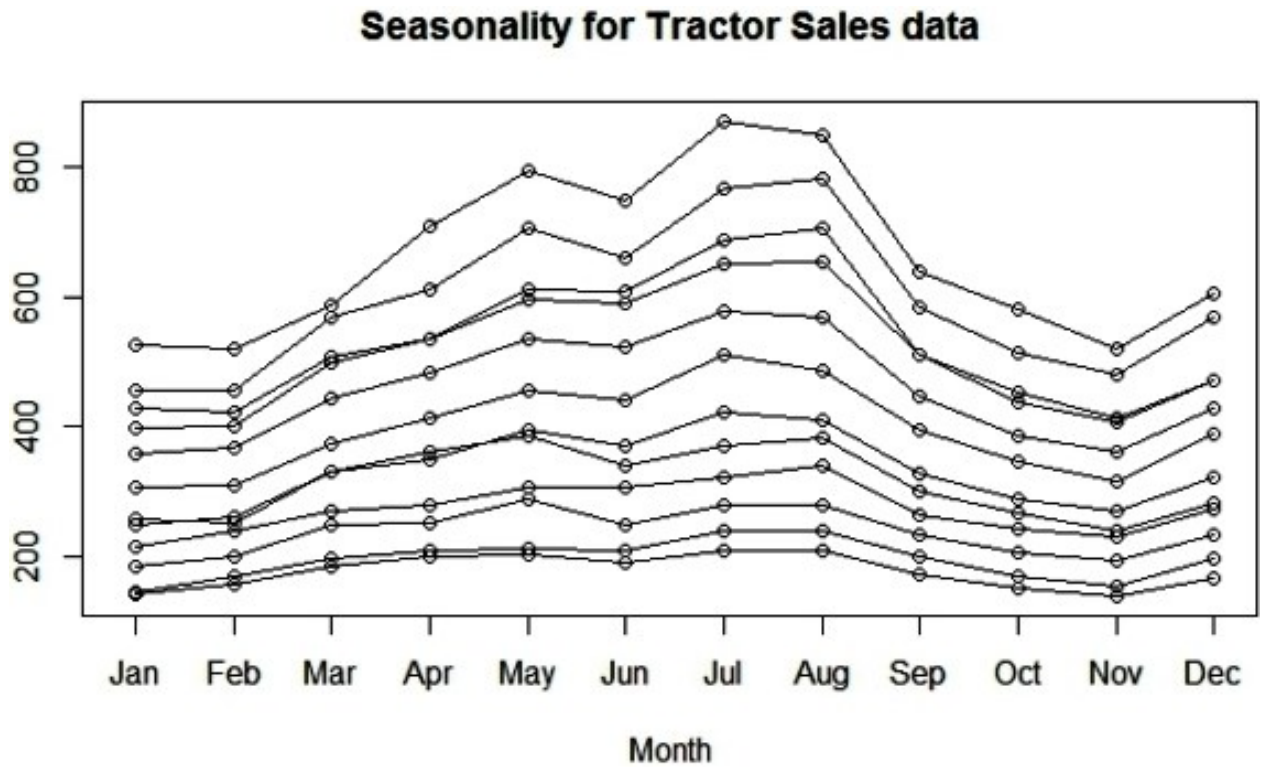


Figure 5: Looking into seasonality.

Case Study Irregular Remainder

Multiplicative Time Series

$$Y_t = Trend_t * Seasonality_t * Remainder_t$$

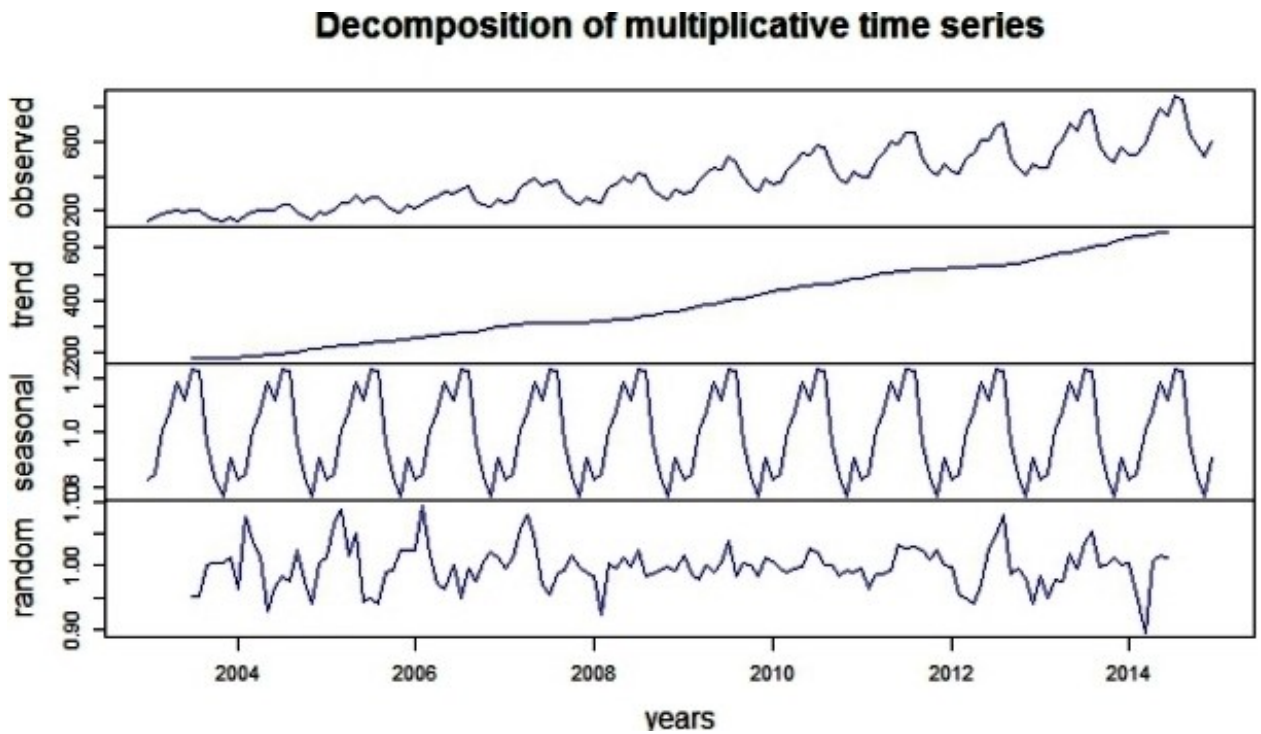


Figure 6: Decomposing a multiplicative TS model.

Observations

- **Trend:** 12-months moving average looks quite similar to a straight line hence you could have easily used linear regression to estimate the trend in this data.
- **Seasonality:** seasonal plot displays a fairly consistent month-on-month pattern. The monthly seasonal components are average values for a month after removal of trend. Trend is removed from the time series using the following formula: $Seasonality_t * Remainder_t = \frac{Y_t}{Trend_t}$
- **Irregular Remainder (random):** is the residual left in the series after removal of trend and seasonal components. Remainder is calculated using the following formula: $Remainder_t = \frac{Y_t}{Trend_t \times Seasonality_t}$

ARIMA and Sugar Cane Juice

Figure 7: A sugar cane juice producer in India.

ARIMA Modelling

- **1st Pass Integrated:**

  No differencing (d=0):

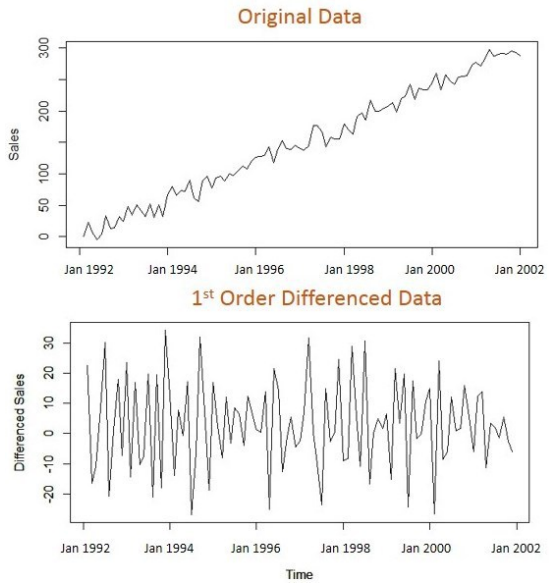
  $Y'_t = Y_t$

  1st differencing (d=1):

  $Y'_t = Y_t - Y_{t-1}$

  2nd differencing (d=2):

  $Y'_t = Y_t - Y_{t-1} -$
  $(Y_{t-1} - Y_{t-2}) =$
  $Y_t - 2 \times Y_{t-1} + Y_{t-2}$



- **2nd Pass AutoRegressive:**

  $Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t$

- **3rd Pass MovingAverage:**

  $Y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$

White Noise and ARIMA

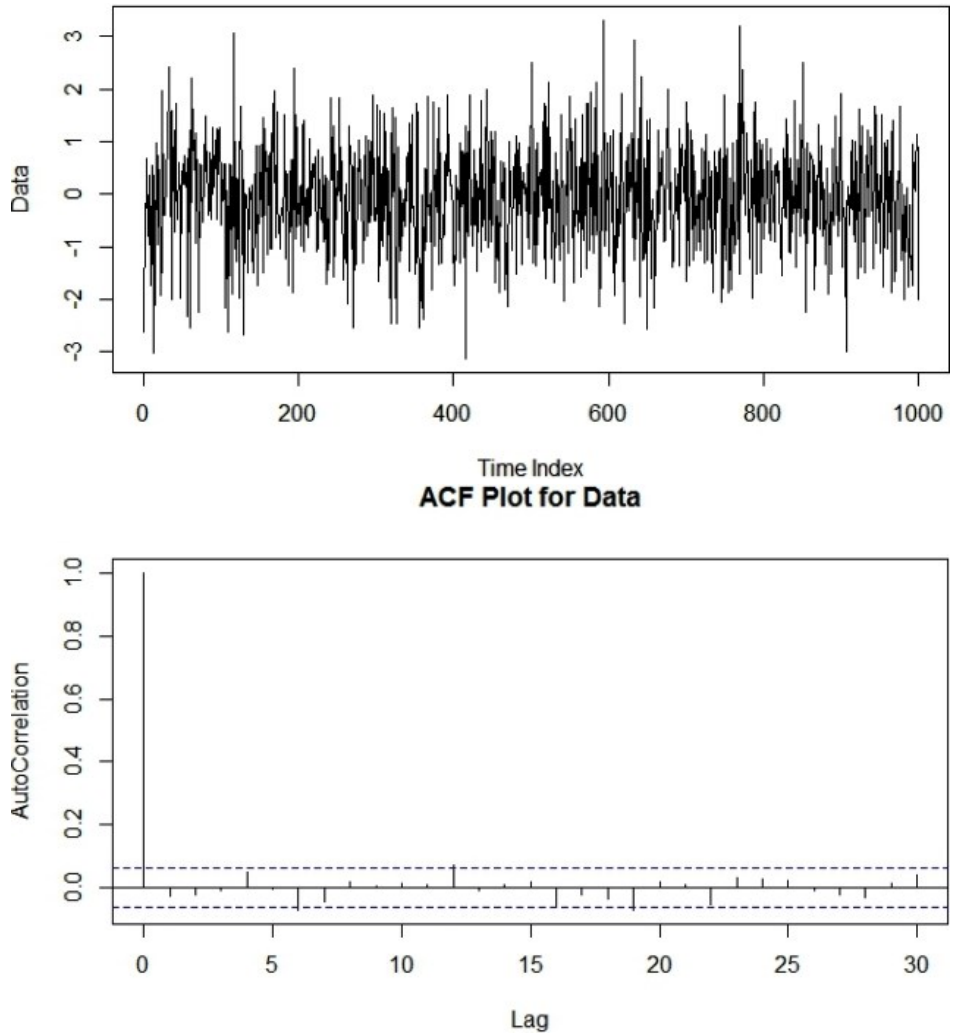
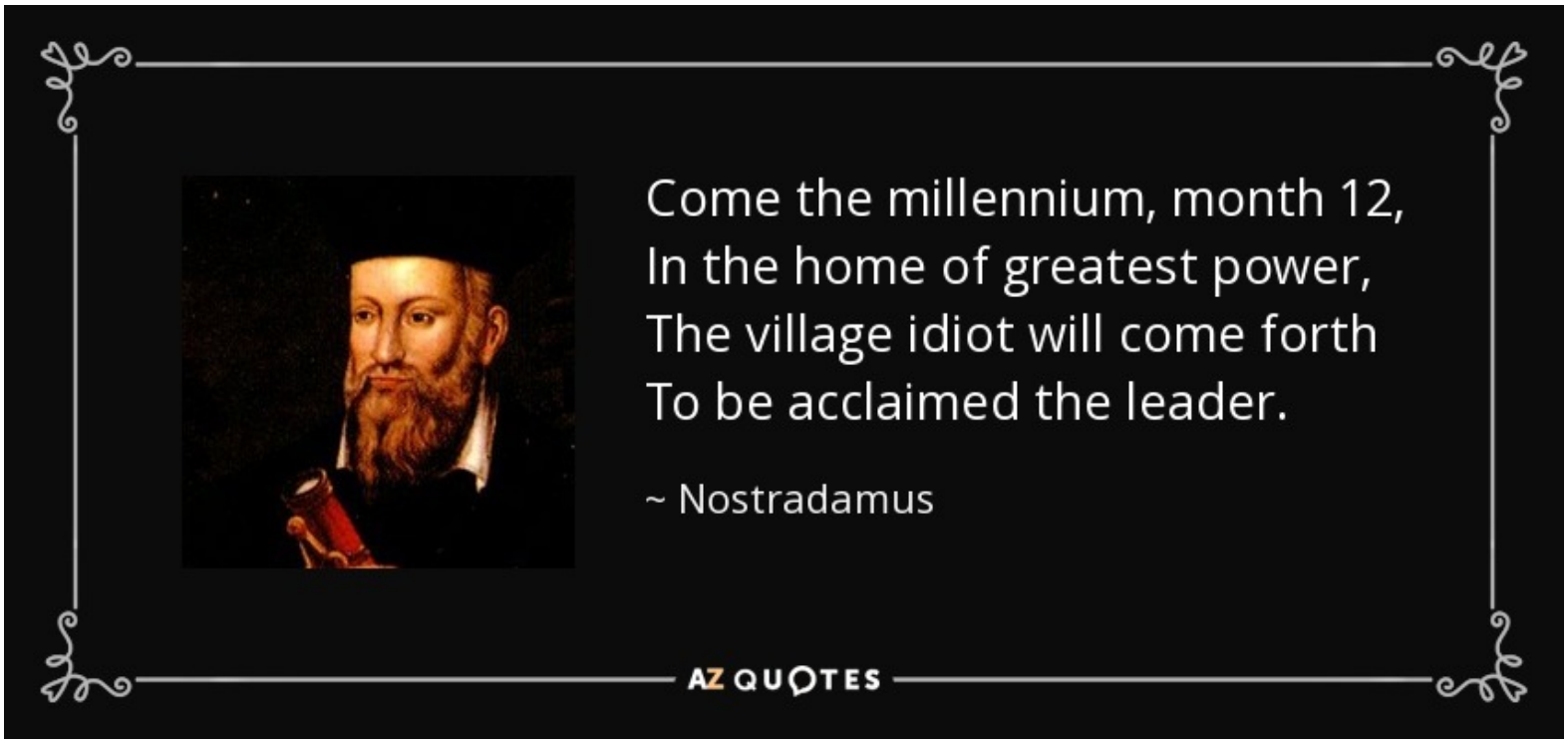


Figure 8: Checking for juice.

Ever wanted to play Nostradamus?



Step-through Time Series Forecasting

Step 1: Plot the data

Step 2: Difference data to make it stationary on mean

1st Differencing (d=1): $Y_t^{'} = Y_t - Y_{t-1}$

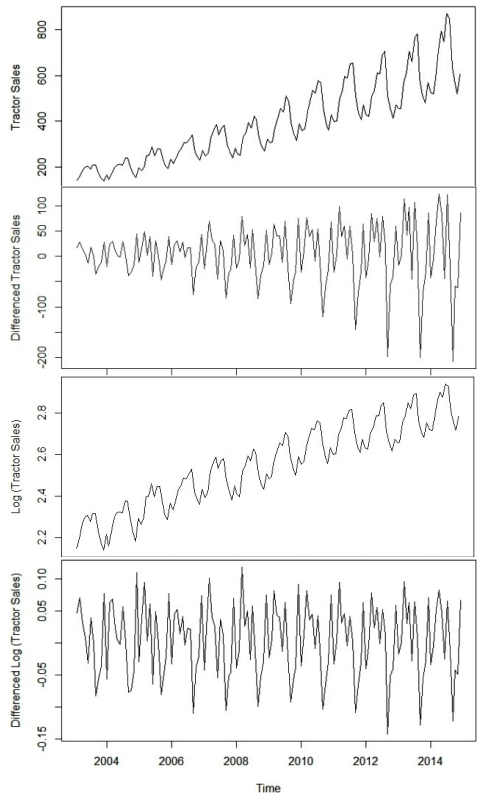
Step 3: Log transform data to make it stationary on variance

Log of sales: $Y_t^{new} = log_{10}(Y_t)$

Step 4: Difference log transform data to make it stationary on both mean and variance

1st Differencing (d=1) of log of sales:

$Y_t^{new'} = log_{10}(Y_t) - log_{10}(Y_{t-1})$



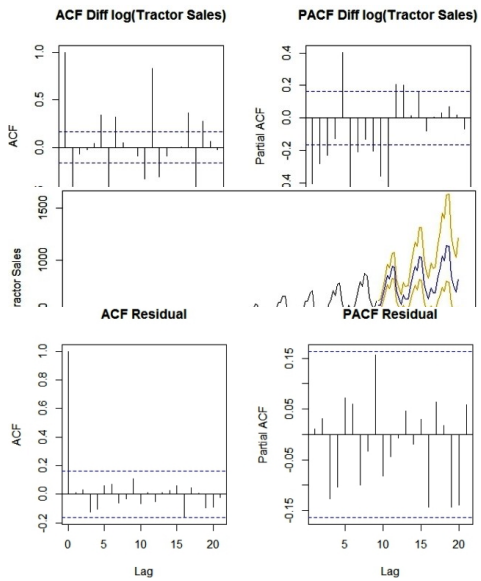
Step-through Time Series Forecasting

Step 5: Plot ACF and PACF to identify potential AR and MA model

Step 6: Identification of best fit ARIMA model

Step 7: Forecast sales using the best fit ARIMA model

Step 8: Plot ACF and PACF for residuals of ARIMA model to ensure no more information is left for extraction



Akaike Information Criterion (AIC)

Definition

Selects the best fit model by balancing between goodness-of-fit and number of parameters used in the model. Alternative: Bayesian Information Criterion (BIC).

Formula

$$AIC = 2K - 2ln(L)$$

Case Study Regression with ARIMA Errors

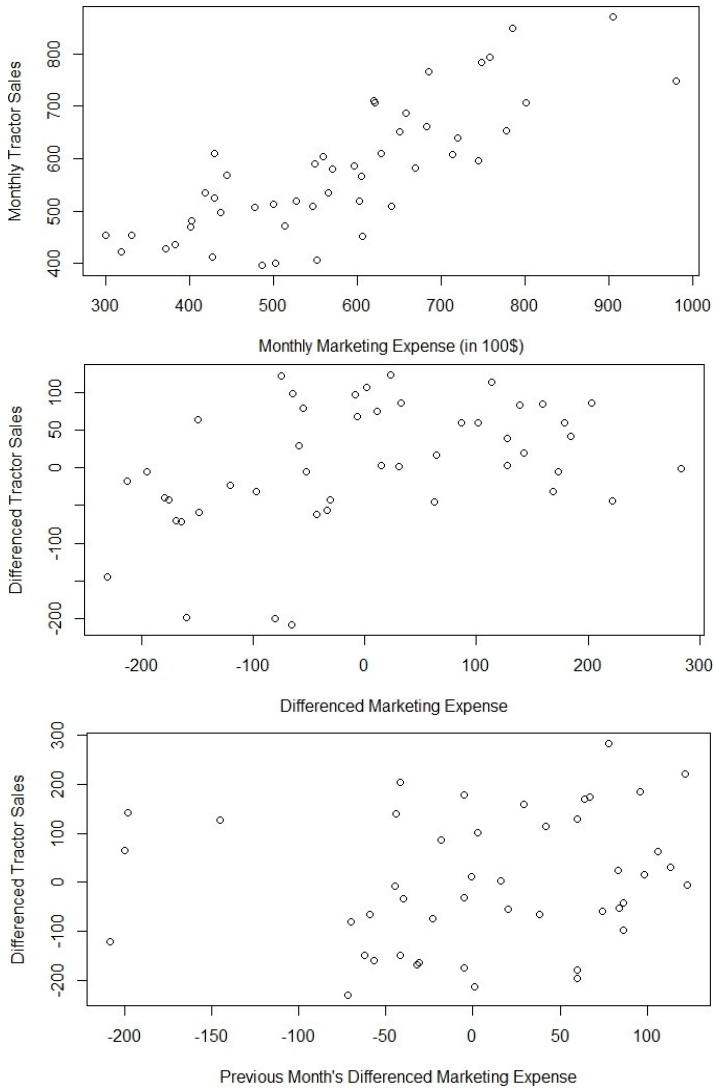


Figure 9: Marketing Expense, Differenced, and Previous Months

CeADAR Demonstrator Applied Time Series Analysis and Prediction

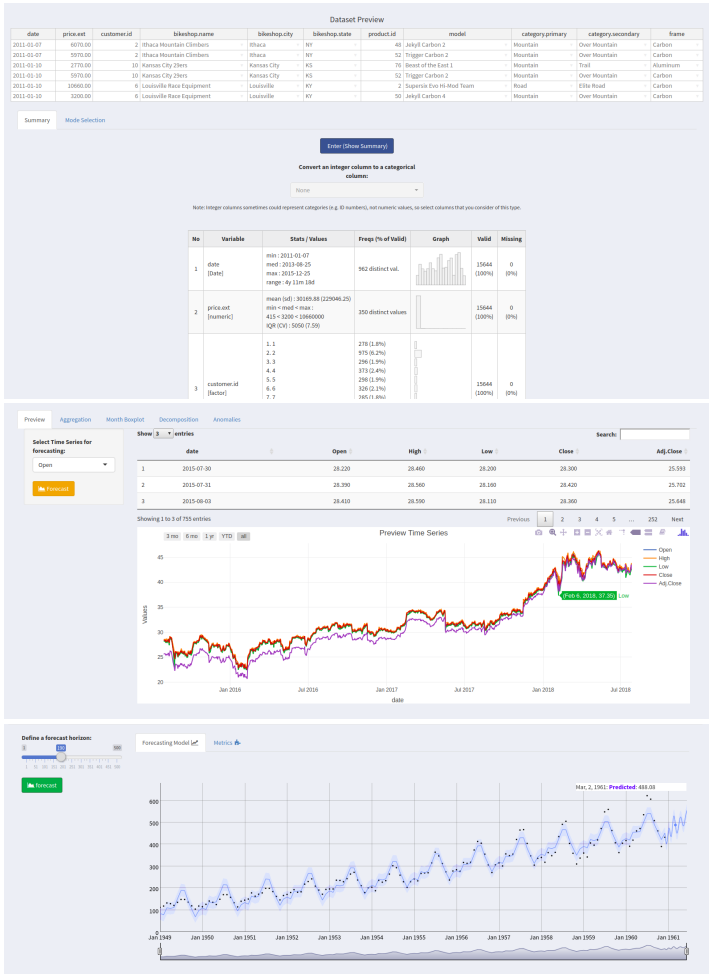


Figure 10: CeADAR Applied Time Series and Prediction demonstrator.

Contact

Bojan Božić
CeADAR Research Fellow
Email: bojan.bozic@dit.ie
Twitter: @bojan_bozic

TimeSeries use case and data: ucanalytics.com
Images: AZ Quotes, Michael Sloan, coolperthnights.com