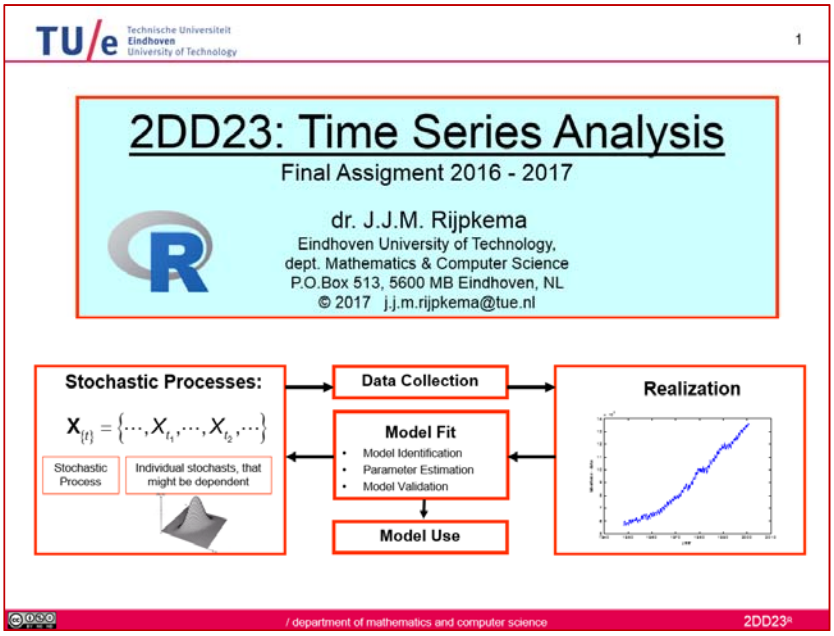


## General Information:

For a successful completion of the elective course 2DD23: “Time Series Analysis and Forecasting”, participants who have registered for the 2016-2017 edition of the course must solve, in groups of at maximum two, an assignment on the analysis of time series, consisting of two parts: **part 1** is related to univariate time series and **part 2** is related to multivariate time series.

Groups should hand in their results in the form of a final report in English. The size of the report is limited **to a maximum of 8 pages for the main body of text, not including the appendices.** Reports should be submitted to j.j.m.rijpkema@tue.nl in pdf-format, named: **2DD23Report_2017_groupXX.pdf**, with XX the relevant group number. A deadline for submission is agreed upon per group and is one of:

Thursday, July 6th - 2017, 16.00
Thursday, July 27th - 2017, 16.00
Monday, August 14th - 2017, 16.00

The final report should contain, for both of the assignment, at least the following elements:

- Introduction, problem formulation and research hypotheses.
- An Exploratory Data Analysis of available data.
- Plan and a reasoned choice of the statistical techniques and models and of its necessary assumptions.
- Results of the analysis (s) and reasoned conclusions.
- Checking the assumptions of the techniques used.
- Conclusions and recommendations on problem formulation and analysis.
- Appendix 1: a selection of relevant (!) statistical output.
- Appendix 2: R scripts used for the analyses.

Models are expected to be fitted “non automatically” (unless otherwise stated!), which means that all steps needed for the selection of the model and the fitting of the parameters should be explicitly mentioned, performed and discussed by the group. Mere use of the final results from an ‘automatic modeler’ is insufficient!

Should results give rise to this, an additional oral discussion, within 10 days after handing in the report might be scheduled by mutual agreement.

# ASSIGNMENT Part 1: Univariate Time Series

**Data set: groupXX_data_1.Rdata** (with XX the relevant group number)

One focus of research on business forecasting has been empirical validation. The International Institute of Forecasters (www.forecasters.org) has collected a large number of datasets that are used by hundreds of researchers for validation and comparison of various forecasting methods. From the total of 3003 datasets that are available for the M3-competition an individual M3-selection is made for each group, participating in the 2DD23 final assignment.

In this first part of the final assignment 2DD23 you are asked to compare, based on the individual M3-selection, properties and performances of decompositional methods, exponential smoothing models and Box Jenkins type of models, as discussed in the course 2DD23.

More specifically you are asked to:

- Perform an exploratory data analysis for the M3-selection indicated, based on adequate representations and "fingerprint analyses", both in the time and the frequency domain.
- Perform a classical seasonal decomposition analysis on series that exhibit seasonality and give an interpretation for the results obtained.
- Fit an adequate exponential smoothing model and discuss the steps taken and the choices made in detail. Express the final model obtained in specific formular form and given it an explicit interpretation. Furthermore, verify and validate the final model selected and discuss your findings.
- Fit an adequate Box-Jenkins model and discuss the steps taken and the choices made in detail. Express the final model obtained in specific formular form and given it an explicit interpretation. Furthermore, verify and validate the final model selected and discuss your findings.
- Explain how forecasts can be generated from these models for the first 6 (for annual data) or 18 (for monthly data) datapoints after the end of the series. Generate specific results (including prediction limits) and discuss your findings.
- Compare results obtained for exponential smoothing and Box-Jenkins models!

# ASSIGNMENT Part 2: Multivariate Time Series

**Data set: groupXX_data_2.RData** (with XX the relevant group number)

The Gapminder Organisation (http://www.gapminder.org/ ) has as its mission "to unveil the beauty of statistics for a fact based world view'. In line with this mission they have collected time series data on a large number of indicators in the field of economics, education, energy, environment, health, infrastructure, population and work for individual countries all over the world. Furthermore, they have developed a dynamical tool, gapminder (http://www.gapminder.org/downloads/), to visualize combinations of indicators as they evolve over time for individual countries. Based on these tools and the available data Hans Rosling, the founder of the Gapminder Organisation, discusses in a TED-talk video entitled "the magic washing machine" (http://www.gapminder.org/videos/hans-rosling-and-the-magic-washing-machine/) the impact of technological development on society and its relation with energy consumption in the world. He formulates a conjecture that energy consumption is associated to the stage of economic development of a country and the size of the population.

In this second part of the final assignment 2DD23 you are asked to explore this conjecture in more detail, by analyzing the selection of Gapminder indicators for an individual country:

- Energy use per capita in tonnes of oil equivalent (toe).
- Gross Domestic Product (GDP) per capita in constant 2000 US$, which may be considered an indicator for the stage of economic development of a country.

More specifically you are asked to:

- Perform an exploratory data analysis on the indicators mentioned for the selected country, based on adequate representations and "fingerprint analyses", both in the time and the frequency domain. Results for individual indicators and for their mutual relation should be discussed!
- Fit an adequate univariate Box-Jenkins model to the indicator for the energy use per capita and discuss the steps taken and the choices made. Express the final model obtained in specific formular form and given it an explicit interpretation. Furthermore, verify and validate the final model selected and discuss your findings.
- Fit an adequate dynamic regression model to the indicator for the energy use per capita, where GDP per capita is used as an independent predictor variable. Discuss the steps taken and the choices made! Express the final model obtained in specific formular form and given it an explicit interpretation. Furthermore, verify and validate the final model selected and discuss your findings.
- Compare results obtained from the univariate Box Jenkins models and from the dynamic regression model. Check, if exponential smoothing models for the energy indicator outperform the Box Jenkins and/or the dynamic regression models fitted and comment on the results obtained!