

ASSIGNMENT

Robusta Metoder

MS1415

Bruna G. Palm

Department of Mathematics and Natural Sciences
Blekinge Institute of Technology, Sweden

April 23, 2025

Introduction

For this assignment, you may work individually or in pairs (groups of two students). The goal is to apply and compare different statistical models using a provided dataset, with the objective of identifying the most suitable approach for the data. Your results should be presented in a concise written report (in PDF format), to be submitted via the course page on Canvas before the stated deadline. Throughout the report, you are expected to clearly explain your steps, justify your methodological choices, and support your conclusions with appropriate validation. Aim to keep your writing clear, concise, and focused.

Grades

This assignment is graded with the following grades: G/Ux/U.

Allowed programming language, packages, and tools

Feel free to use the language you are most comfortable with.

Upon submission of your report

Before you upload your report make sure:

1. That you have included your **names** and **email addresses** in the report.
2. The report follows a logical and a well-structured **format** with written explanations.
3. That you have carefully checked the report for **spelling and grammatical errors**.
4. That your report is written in **English**.

Failure to comply with any of the aspects above could result in a failing grade, and that you have to revise the report and submit it again on a later deadline.

The data set

The dataset provided contains monthly records of the useful water volume of the Caconde Reservoir, located in Caconde, São Paulo, Brazil. This volume is defined as the percentage of the reservoir's capacity between its minimum and

maximum operational levels. In this assignment, we will model the proportion of this percentage, which is scaled to lie within the interval $(0, 1)$. The available data spans from January 2015 to February 2024. It is sourced from publicly accessible records, specifically from <http://www.ipeadata.gov.br>, and is also available on the course page on Canvas. Figure 1 displays the time series of the monthly useful water volume for the Caconde Reservoir.

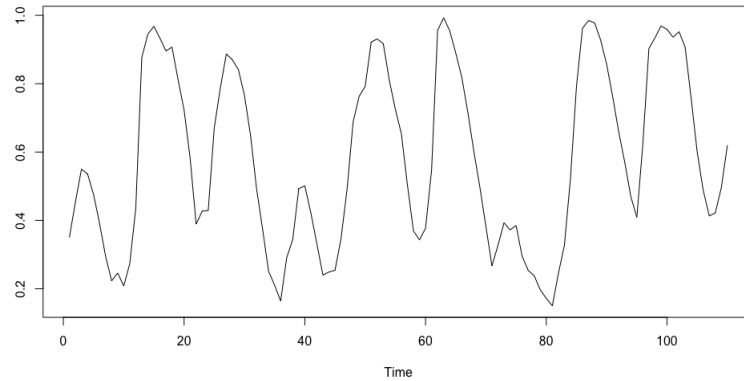


Figure 1: Monthly useful water volume of the Caconde Reservoir.

Your Task

The objective of this assignment is to apply time series modeling techniques to predict the useful water volume of the Caconde Reservoir.

Your task can be summarized as follows:

1. Use visual exploration tools to understand the dataset. Investigate aspects such as time dependence, seasonality, trends, and data behavior.
2. Test the series for stationarity.
3. Split the dataset into training and validation sets by reserving the last 6 observations for forecasting evaluation.
4. Estimate an Autoregressive (AR) model using the dynamic regression approach discussed in class. Select the appropriate number of covariates based on the data time dependence.
5. Estimate an ARMA model including seasonal components (e.g., sine and/or cosine terms) to account for seasonality.
6. Estimate a SARMA model.
7. Compare the three models. Which model is the most suitable for this dataset? Keep in mind that a model can only be used for prediction if it has been properly validated. To compare the forecasting performance, evaluate both in-sample and out-of-sample predictions using visual tools and performance metrics.
8. Are there any other models that could be applied?

Good luck!