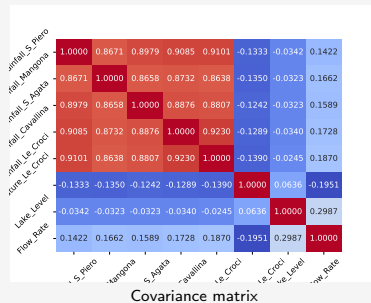# Capgemini test: forecasting water levels

Stefano Petrucci

petrucci.ste@gmail.com

# Data set

- Chosen *Lake_Bilancino*
- Two target variables:
    - Lake level
    - Flow rate
- Most variables missing before 01/01/2004
  $\Rightarrow$ removed since $\sim 8\%$ of the total
- Missing data not replaced[a]
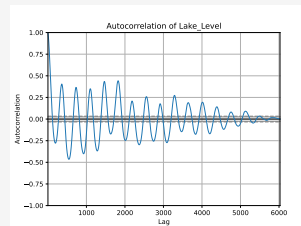

Covariance matrix

---

[a]Applied an interpolation only when computing the autocorrelation.

Forecasting strategy

1. Simple AutoRegressive (AR) model
2. More complex AutoRegressive Integrated Moving Average (ARIMA) model
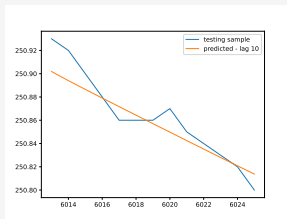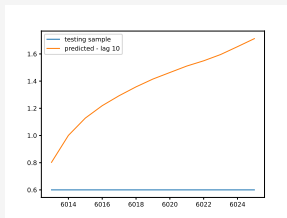3. Multivariate analysis (not implemented)
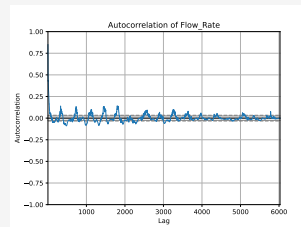
Flow rate - autocorrelation

Lake level - autocorrelation

Predictions - 13 samples
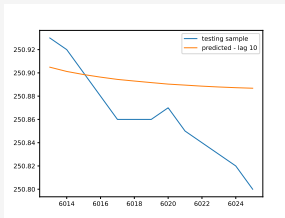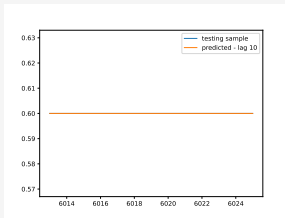
AutoRegressive model

Lake level

Flow rate

ARIMA model

Lake level

Flow rate

Conclusions

- Implemented a toy script to compute AR and ARIMA algorithms
- Both algorithms showed better performance on Lake_level