UNIVERZA V LJUBLJANI
FACULTY OF MATHEMATICS AND PHYSICS

Financial mathematics – Master's study

Regina Blagotinšek

Local prediction of weather parameters based on historical data

Project report

Mentor: Dr. Matjaž Depolli,
Supervisor: Prof. Dr. Sergio Cabello Justo

Ljubljana, May 22, 2024

Contents

Local prediction of weather parameters based on historical data

Abstract

Keywords: time series, ARIMA, weather forecast, error evaluations

1. Introduction

summarize the instructions(relevant parts), it is done in python, summarize what was done: data visualization, ARIMA models fits, errors evaluation, forecast, comparison with 2 simple models. this was the training part. next part was testing. new data, how well do the chosen models fit it and how good are the forecasts of new data. dont mention multivariate here, that will be the last section as what to do next and will be done later.

2. Theoretical Background

2.1. **ARIMA (Auto-Regressive Integrated Moving Average) model.** just a quick ARIMA theory section, shrink it into 1 page? what is ARIMA? importance of stationarity seasonality? (SARIMA) residuals ACF, PACF AIC, BIC

3. Methodology and Implementation

3.1. **Data.** The data I worked with is in the form of time series with 5 minute time steps of weather parameters' measurements. The parameters are ambient temperature, solar radiation intensity, air pressure, relative humidity, wind speed, wind direction and rain intensity. To reduce the dimension of the data set, I aggregated the 5 minute time steps into 1 hour time steps by taking the hourly means.There were also some missing measurements and since the ARIMA model requires equal time intervals I filled the gaps by taking the next available measurement.

3.2. **ARIMA model fitting and evaluating.** For each of the parameters I was searching for the best ARIMA(p,d,q) with two different methods. First one is with the function auto_arima function from the pmdarima Python library. It searches through a range of potential models and selects the best one based on the AIC value. The second method was also done by searching for the optimal p, d and q parameters and checking AIC and BIC values.

3.3. **Errors.** There are many ways to evaluate the errors of the model. In case of modelling time series, investigating the residuals is important. I looked at ACF and PACF plots of all the ARIMA models. A good way to compare the models is also to compare the histograms of their residuals. I also considered the absolute errors and the mean absolute error. Another insightful comparison, especially for the solar intensity radiation was hourly absolute error for 1 day, since there is considerably less solar radiation in the night, the error in the night should be lower than the error during the daytime. For evaluating the fitted ARIMA models, I compared them to 2 simple models and compared their forecasts, which will be explained in the next subsection.

3.4. **Forecast.** I did the forecasts of all the weather parameters for 1 day (24 steps). I compared the absolute errors of ARIMA models' predictions with the actual measurements and 2 simple models' predictions. The first simple model forecasts the tomorrow's weather with today's measurements. The second simple model forecast the next day's weather with the average of the last 3 days' measurements. The ARIMA model is considered 'good' if it is predicting better or at least closely to this two simple models.

3.5. **Testing on new data.** The final way of evaluating the fitted ARIMA models was testing them on new data. New data was again measurements of the 7 weather parameters in 5 minute intervals. I aggregated it to hourly values and fit the suggested ARIMA models to it. For each parameter I tested the better ARIMA fitted to the first data to see if it also fits well here. I evaluated that by checking the AIC and BIC values, calculating the absolute errors and mean absolute error of fitted values and actual values and also by testing the forecasts. The forecasts were tested the similar way as in the prevoius step (2 simple models, 1 ARIMA model and actual values).

4. Results

In this chapter I will make an overview of all the weather parameters and lastly a multivariate model for all of them combined. For purposed of this report, I will only present the main results of my work. Detailed results can be found on the repository of this project.

4.1. **Ambient temperature.** The plot on the left side below is of the data in 5 minute time intervals and on the right side is are the average hourly measurements.

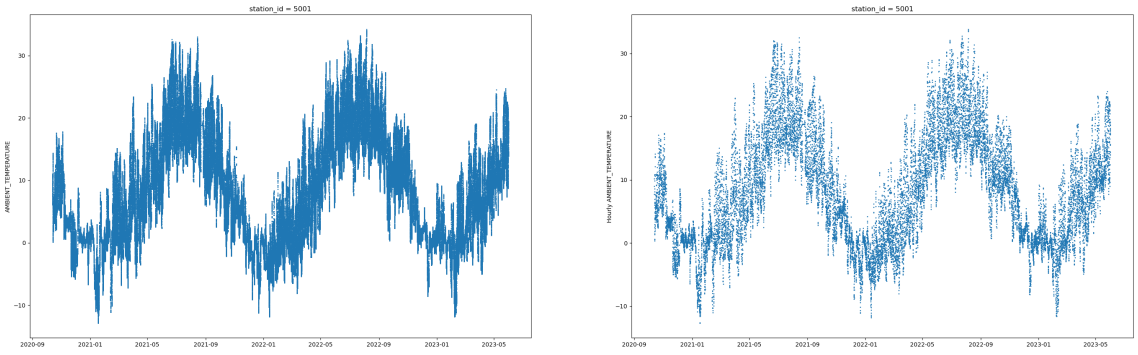

Figure 1. Ambient temperature

The best ARIMA models for ambient temperature are ARIMA(2,1,5) and ARIMA(4,1,2). Based on AIC, ARIMA(4,1,2) is preferred.

residuals: ACF, PACF plots (only from the better one (4,1,2))

- error: plot of absolute error (only (4,1,2))

forecast: plot of compared forecasts to actual data, forecast absolute error plot and absolute mean error.

both ARIMA models are better than last-3-day-model but worse than tomorrow=today. not so bad!

4.2. **Solar radiation intensity.**

4.3. **Relative Humidity.**

References

[1] T. Bertok, *Slučajne matrične igre*, Delo diplomskega seminarja (2020) 4–24.

[2] O. A. Camarena, *Matrix games* (2021). Pridobljeno 22. 4. 2022 z naslova: https://www.matem.unam.mx/ omar/math340/matrix-games.html.

[3] J. Berg & A., Engel, *Matrix games, mixed strategies, and statistical mechanics*, Institute for theoretical physics (1998) 1–4.

[4] L. Ein-Dor & I., Kanter, *Matrix games with nonuniform payoff distributions*, Physica A (2001) 80-88.