Local prediction of weather parameters based on historical data

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

University of Ljubljana, Faculty of mathematics and physics

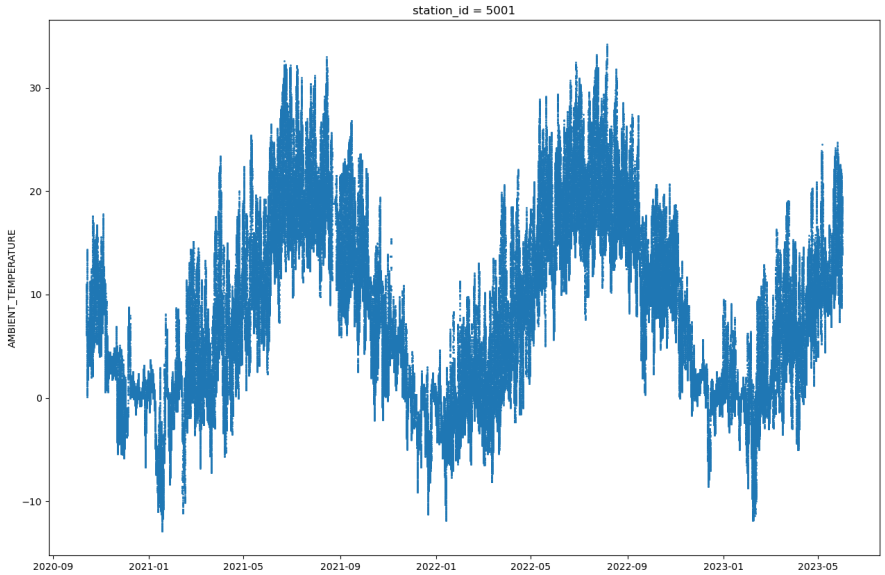
7. 6. 2024

Overview

Introduction

- Goal: short-term local weather forecast using historical data;
- data: time series of measurements of seven weather parameters with 5-minute intervals;
- programming language: Python;
- statistical models: $ARIMA$, VAR.

Ambient temperature



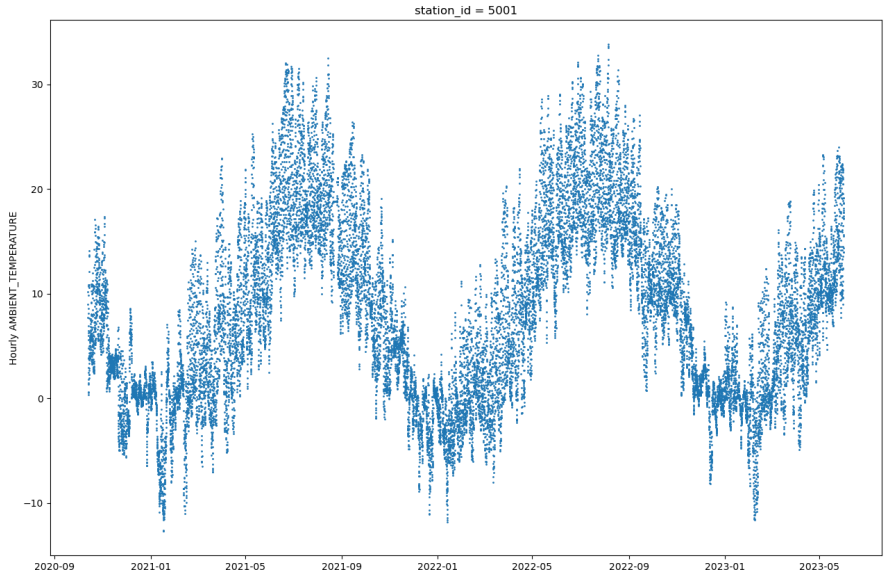
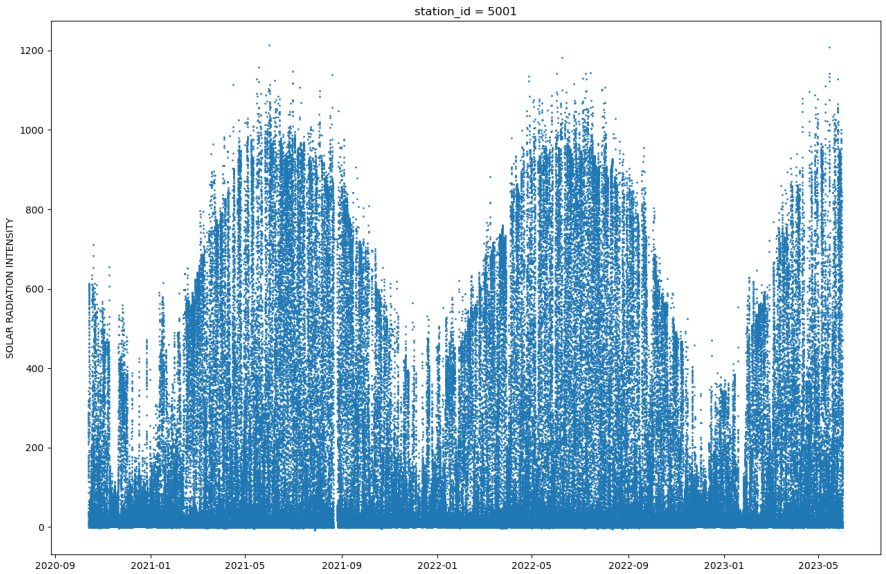


Figure: Ambient temperature

Solar radiation intensity



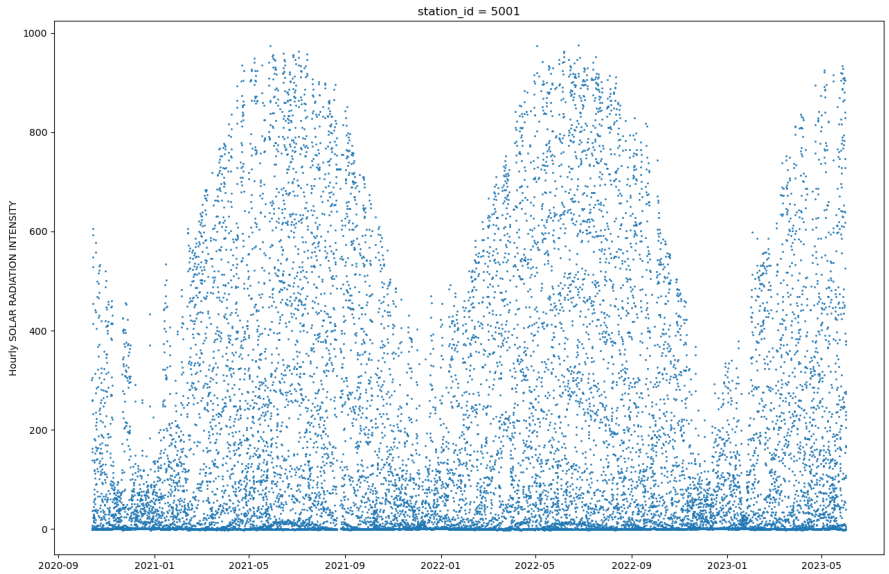
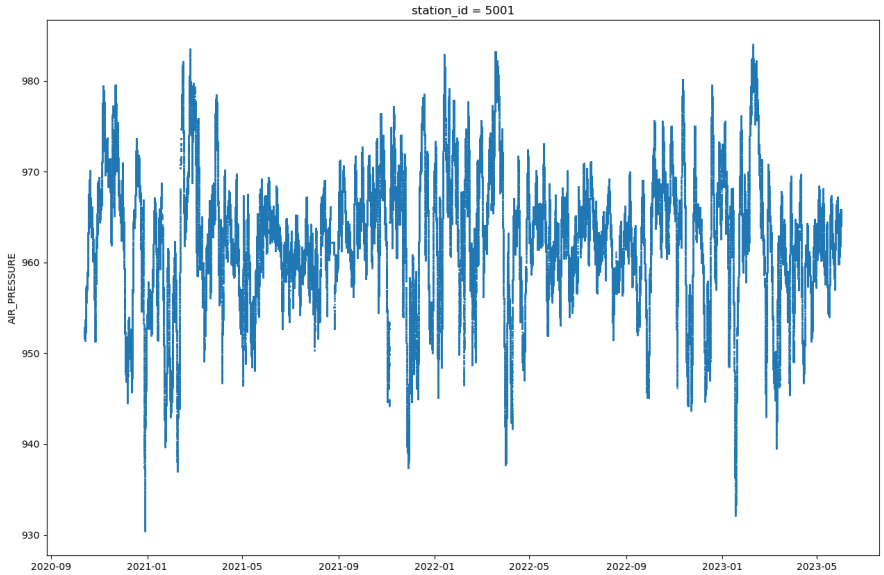


Figure: Solar radiation intensity

Air pressure



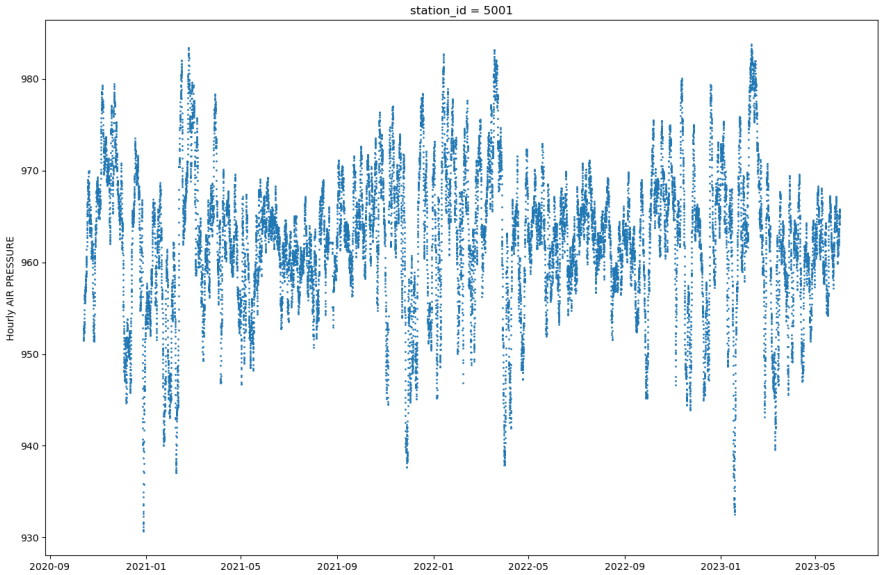
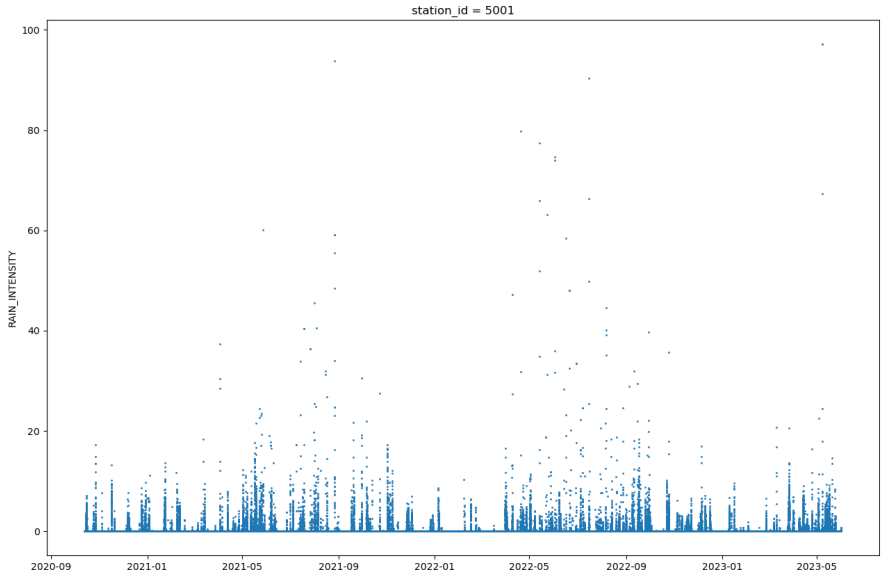


Figure: Air pressure

Rain intensity



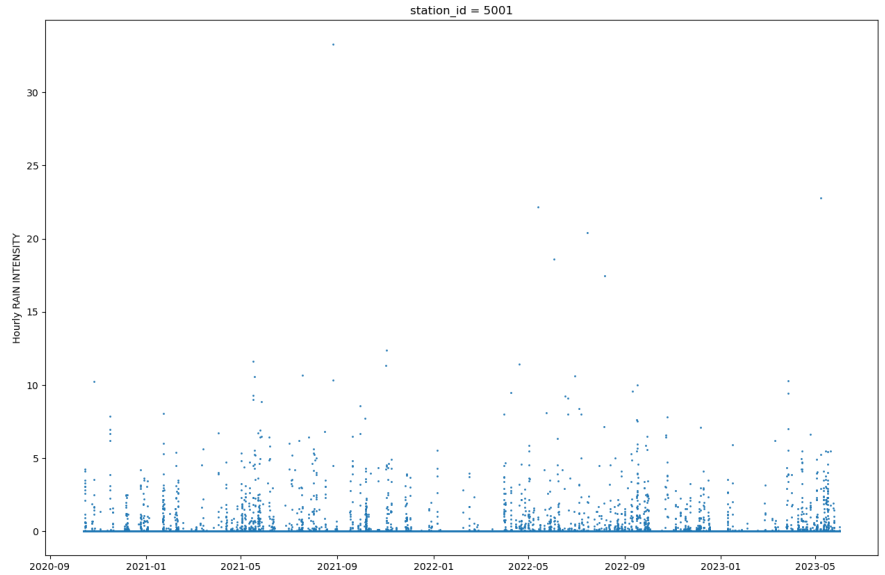
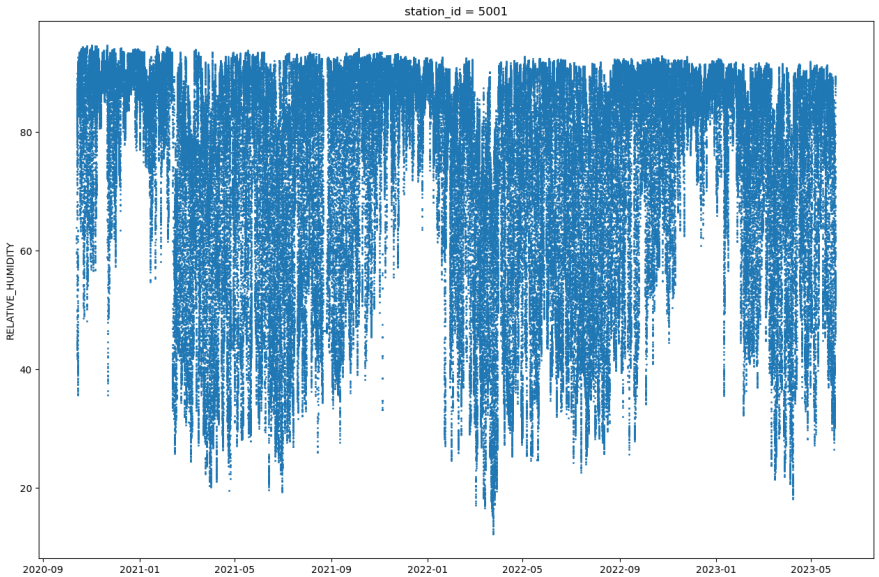


Figure: Rain intensity

Relative humidity



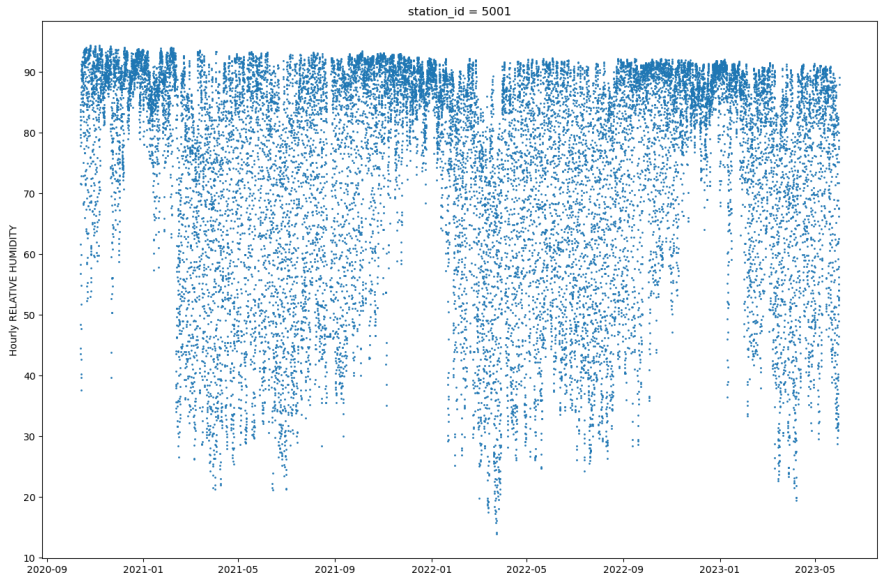
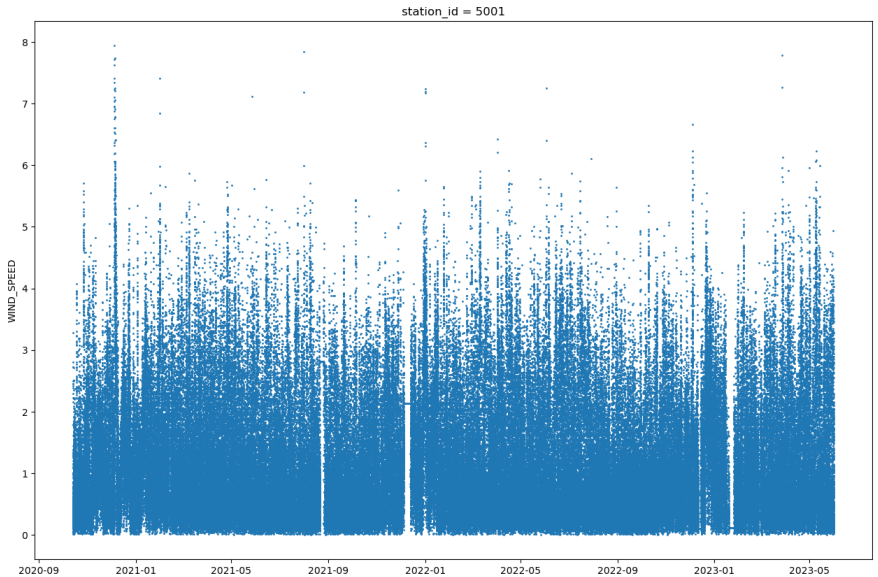


Figure: Relative humidity

Wind speed



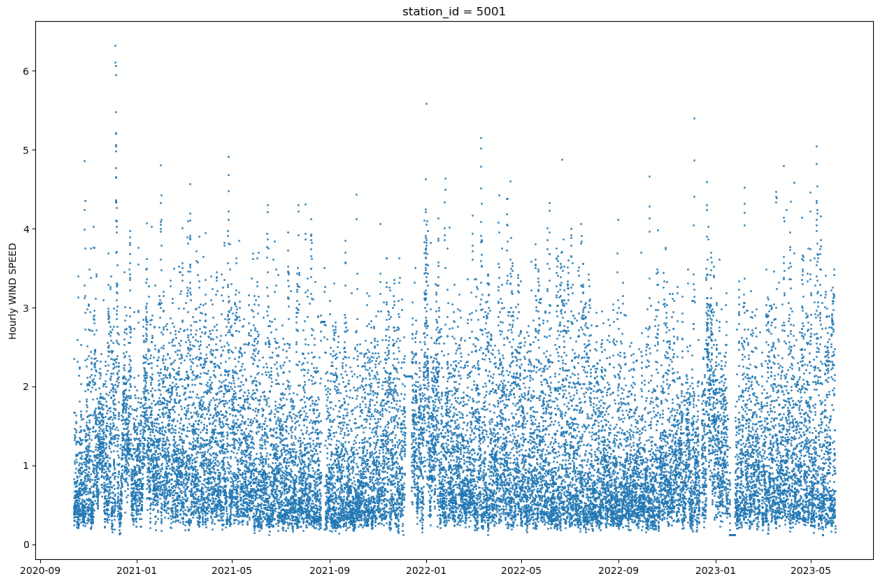


Figure: Wind speed

Wind direction

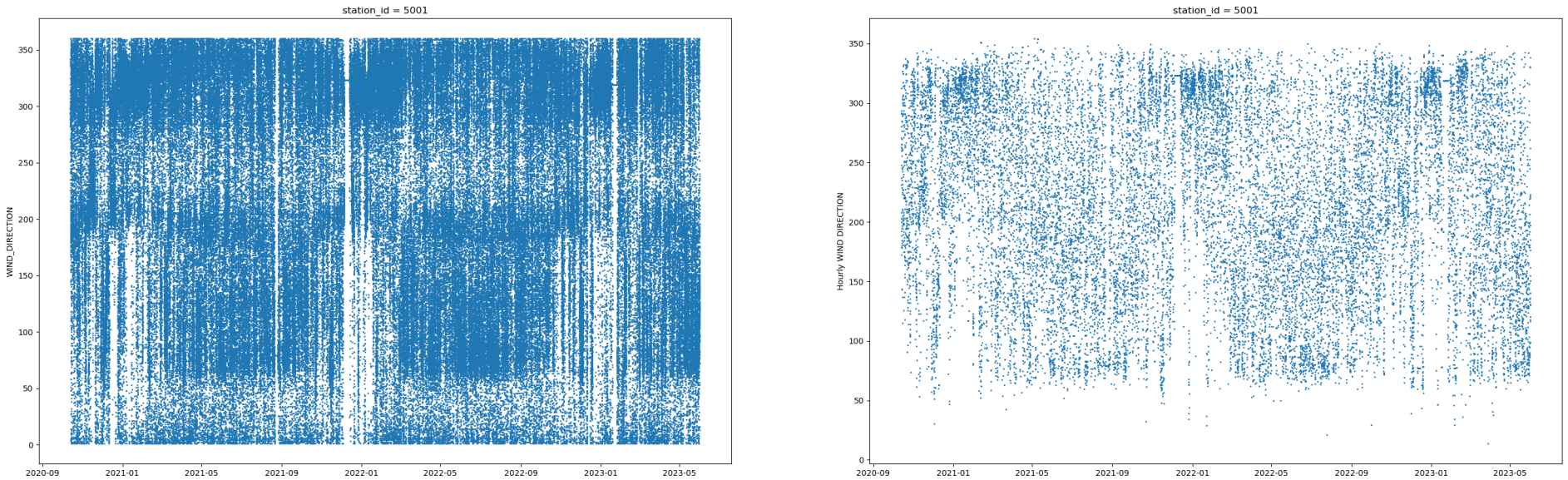


Figure: Wind direction

$ARIMA$

$ARIMA$ (**Auto- Regressive Integrated Moving Average**) is a statistical analysis model used to forecast future points in time series data. It combines three main components:

1. Auto-regressive ($AR(p)$) component: specifies that the output variable depends linearly on its own previous values.
2. Differencing ($I(d)$) component: indicates that the data values have been replaced with the difference between their values and the previous values to make the series **stationary**.
3. Moving average ($MA(q)$) component: incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

AIC and BIC

1. **Akaike Information Criterion** (AIC) balances goodness of fit and model complexity by penalizing the number of estimated parameters.
2. **Bayesian Information Criterion** (BIC) is similar to AIC but includes a stronger penalty for the number of parameters,

Methodology

- Data preparation,
- $ARIMA$ model fitting and evaluating,
- developing two simple models for comparison,
- calculating and comparing the errors,
- forecasting,
- testing on new data,
- multivariate model: Vector auto-regressive (VAR) model.

Models

For each weather parameter I fitted 6 models and compared them to the actual data. Models are:

- two best $ARIMA(p, q, r)$,
- a simple model that takes yesterday's measurement for today's value,
- a simple model that takes the average of the measurements of the last 3 days at the same time as today's value,
- VAR model,
- VAR model with sinusiodial time component.

$ARIMA$ models

1. Ambient temperature: $ARIMA(2,1,5)$ and $ARIMA(4,1,2)$.
2. Air pressure: $ARIMA(3,0,3)$ and $ARIMA(3,1,5)$.
3. Rain intensity: $ARIMA(2,0,2)$ and $ARIMA(3,0,1)$.
4. Relative humidity: $ARIMA(2,1,1)$ and $ARIMA(3,0,4)$.
5. Solar radiation intensity: $ARIMA(3,1,2)$ and $ARIMA(2,1,4)$.
6. Wind direction: $ARIMA(3,1,2)$ and $ARIMA(5,0,1)$.
7. Wind speed: $ARIMA(2,1,5)$ and $ARIMA(4,0,5)$.

Forecast of ambient temperature

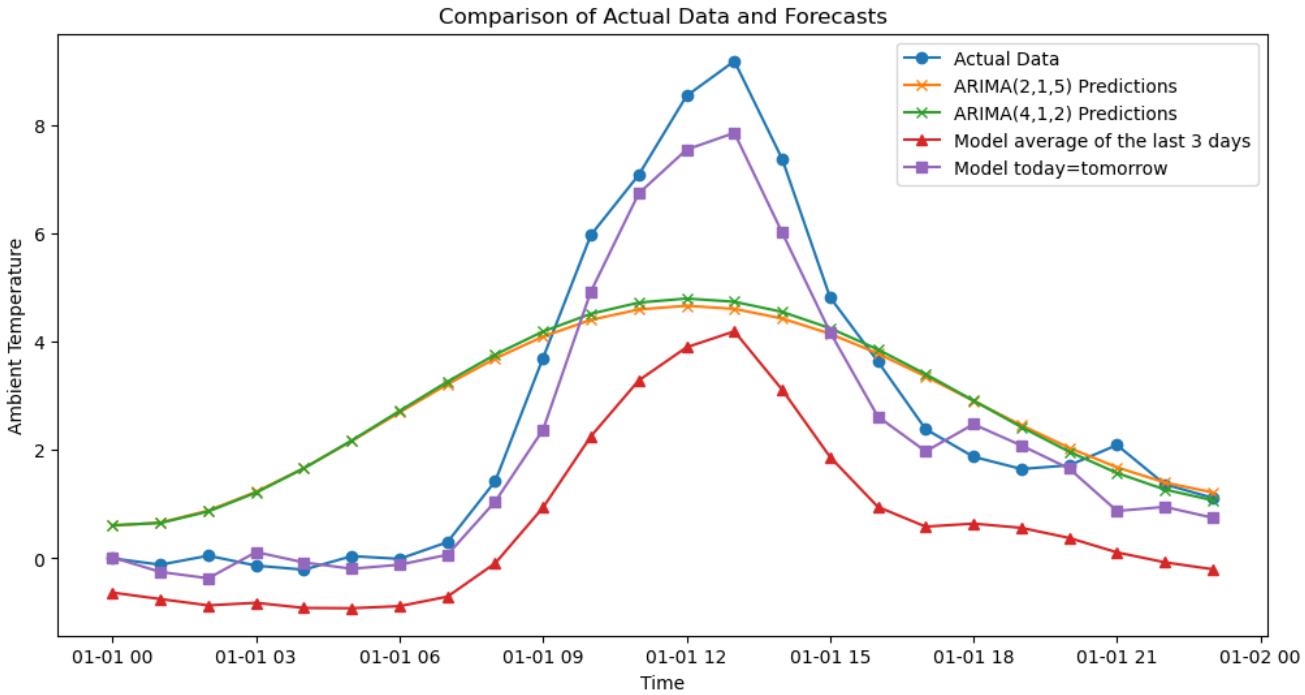


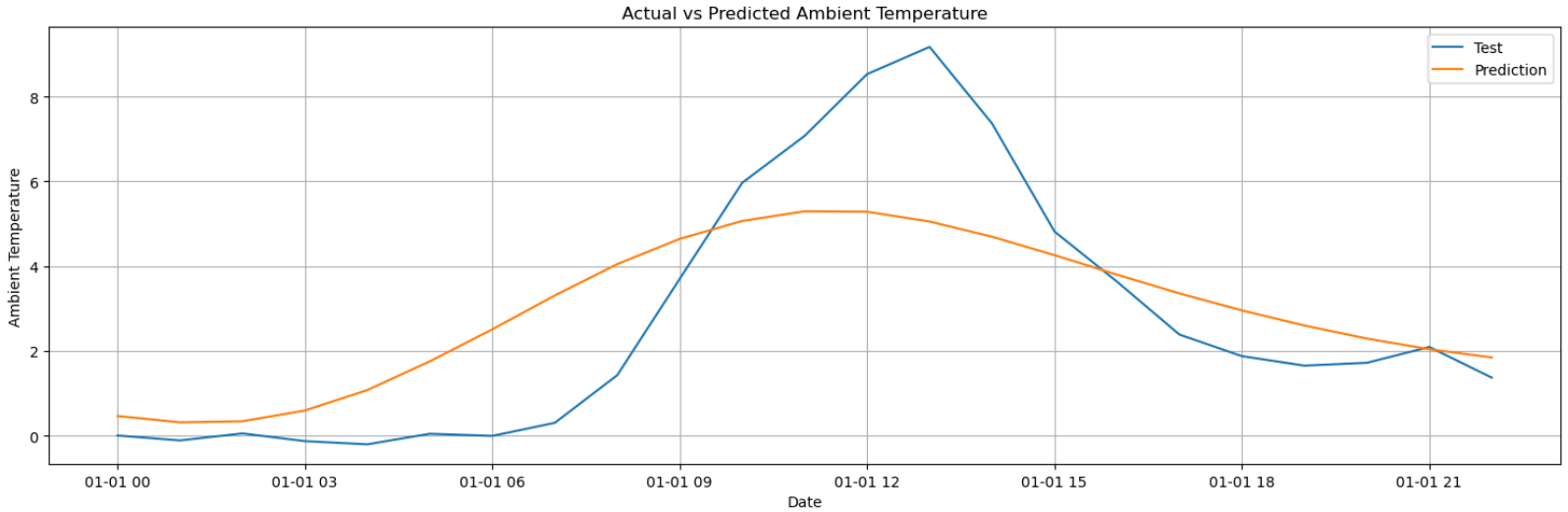
Figure: Models' forecasts and actual data

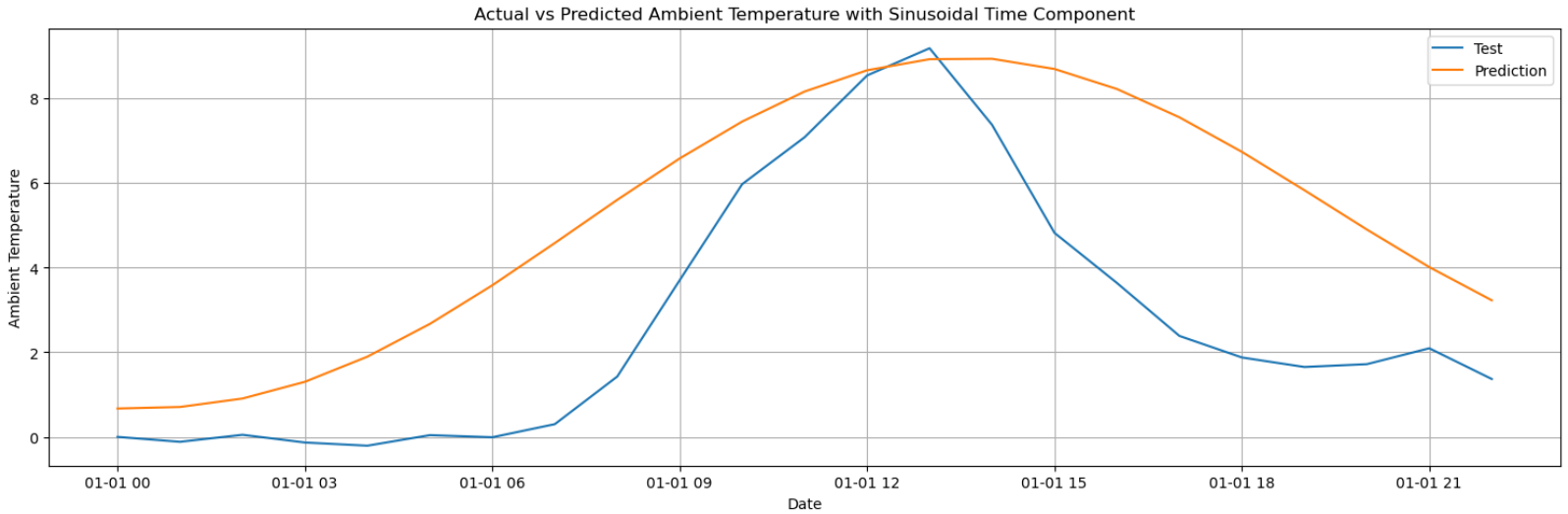
Forecast of ambient temperature

Model	**Forecast Absolute Error[C]**
$ARIMA(2,1,5)$	1.490556
$ARIMA(4,1,2)$	1.474775
Model average of the last 3 days	1.995486
Model today=tomorrow	0.560069

Table: Forecast Absolute Errors

Forecast of ambient temperature: VAR models





Forecast of ambient temperature: VAR models

Model	**Absolute Error**
VAR model with sinusoidal time component	2.458093476878387
VAR model	1.339075022424107
ARIMA(2,1,5)	1.490556
ARIMA(4,1,2)	1.474775

Table: Forecast Absolute Errors

$\Rightarrow$ Transforming time with the sinus function worsened the forecast results, but from the plots it is visible that it captured the higher temperature around noon better.

Forecast of solar radiation intensity

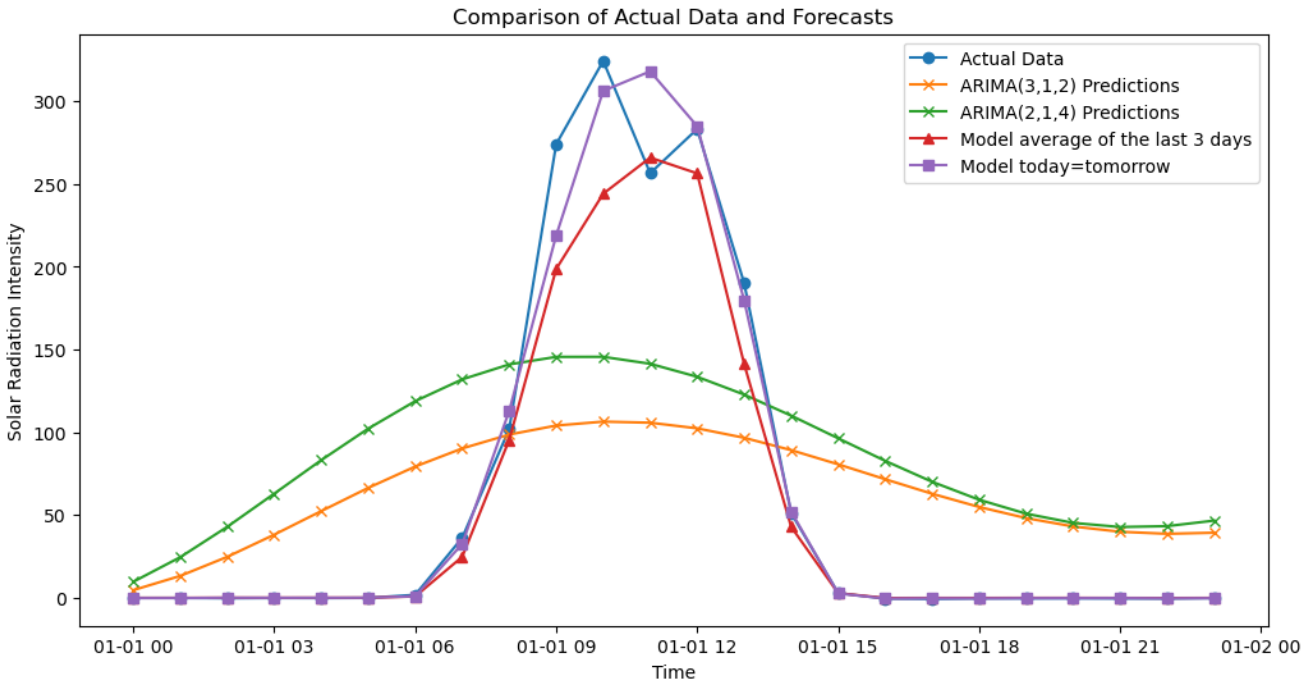


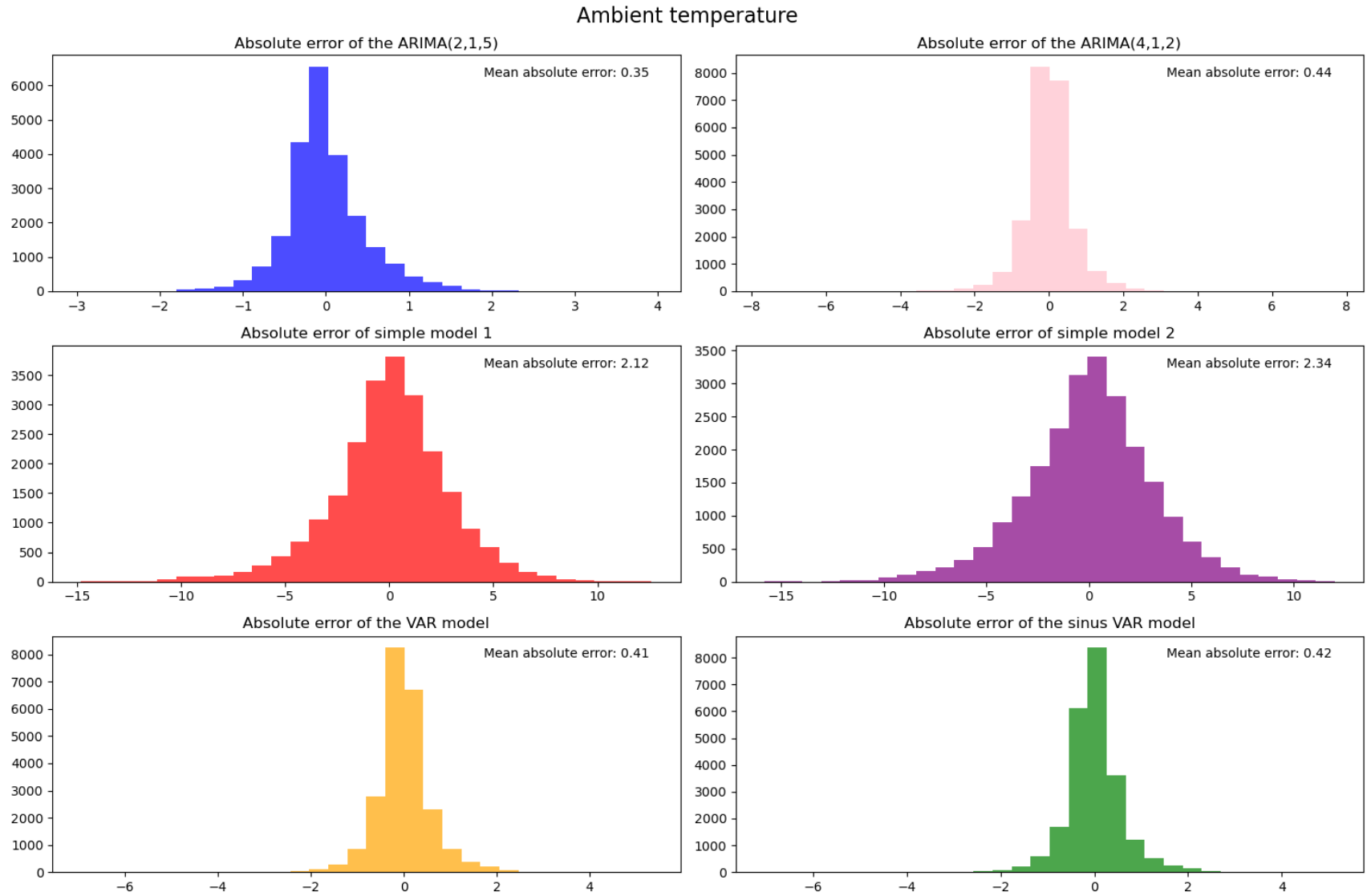
Figure: Models' forecasts and actual data

Forecast of solar radiation intensity

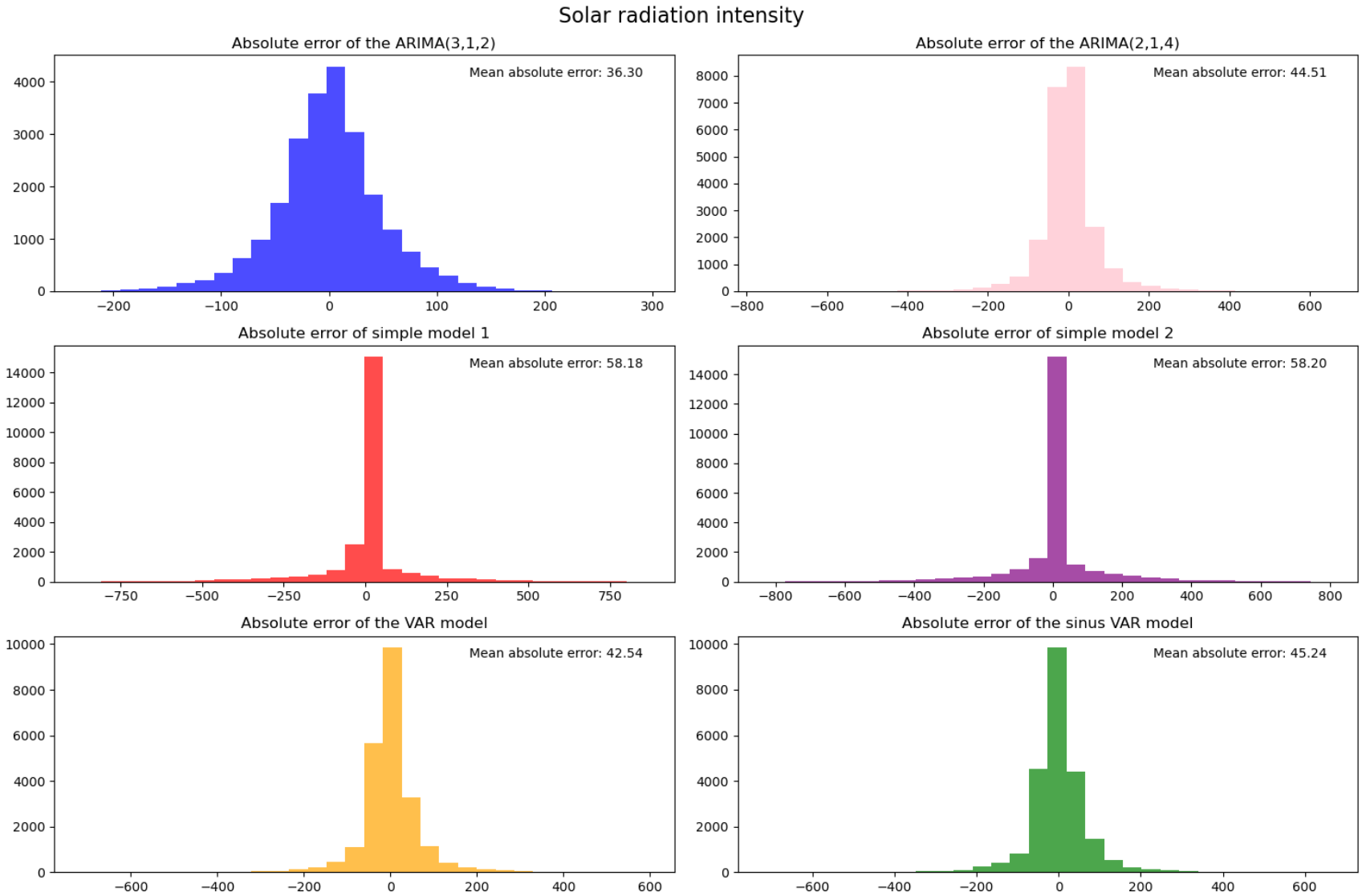
Model	**Forecast Absolute Error** $[\frac{W}{m^2}]$
$ARIMA(3,1,2)$	69.423578
$ARIMA(2,1,4)$	75.599554
Model average of the last 3 days	11.315625
Model today=tomorrow	6.901389

Table: Forecast Absolute Errors

Ambient temperature



Solar radiation intensity



Solar radiation intensity $ARIMA$ model error

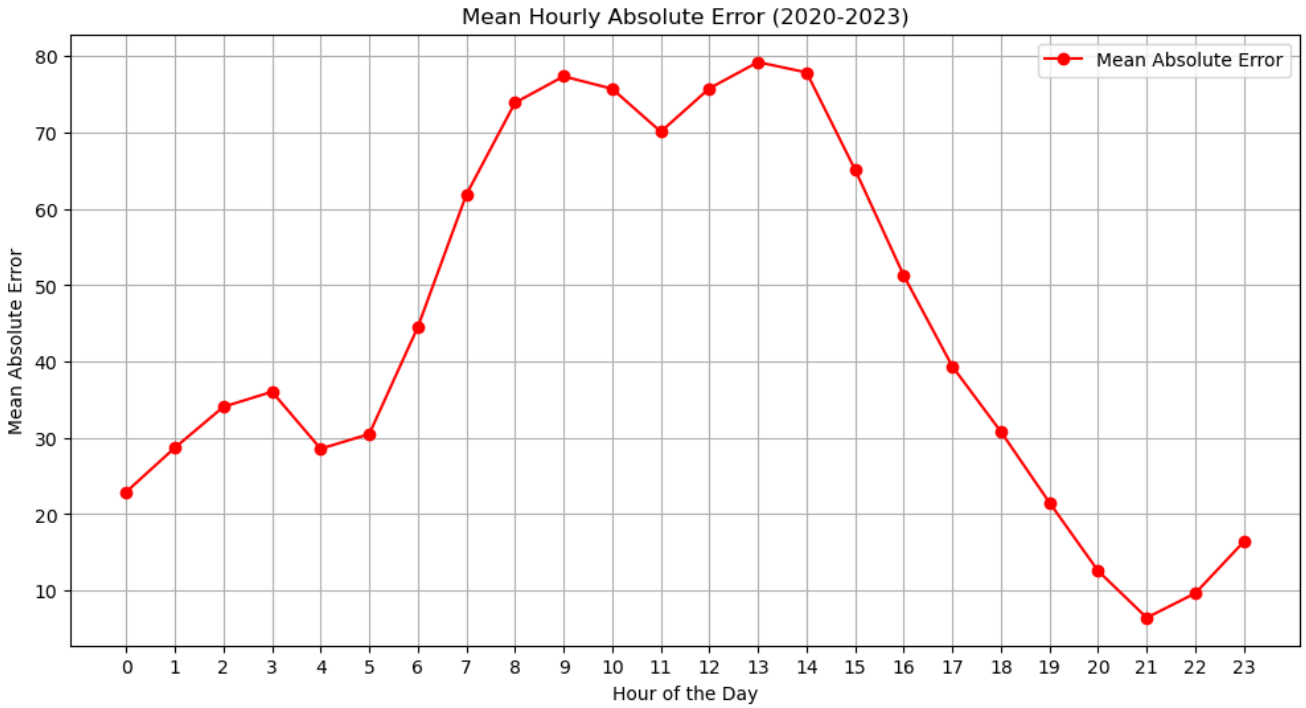
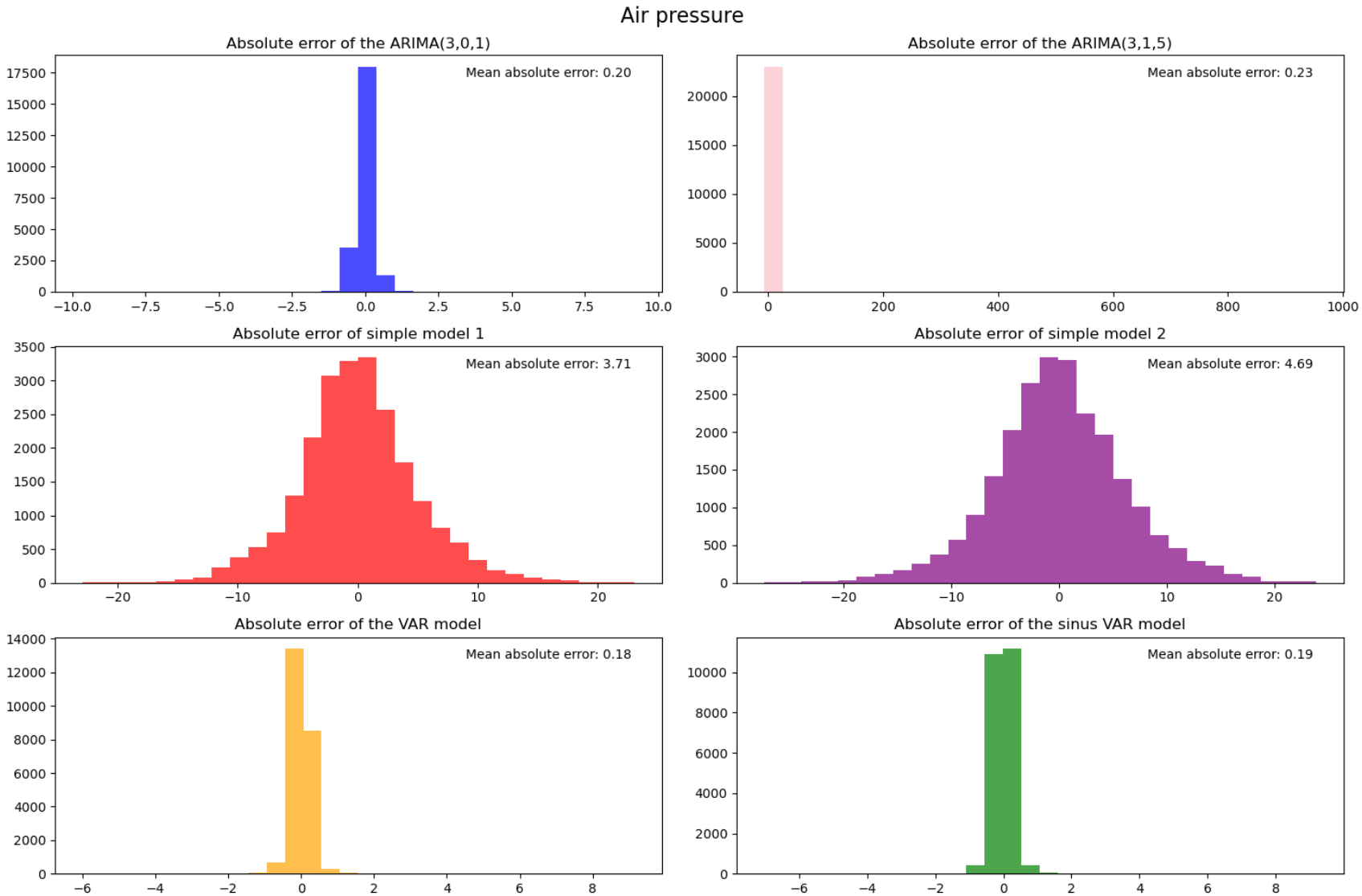


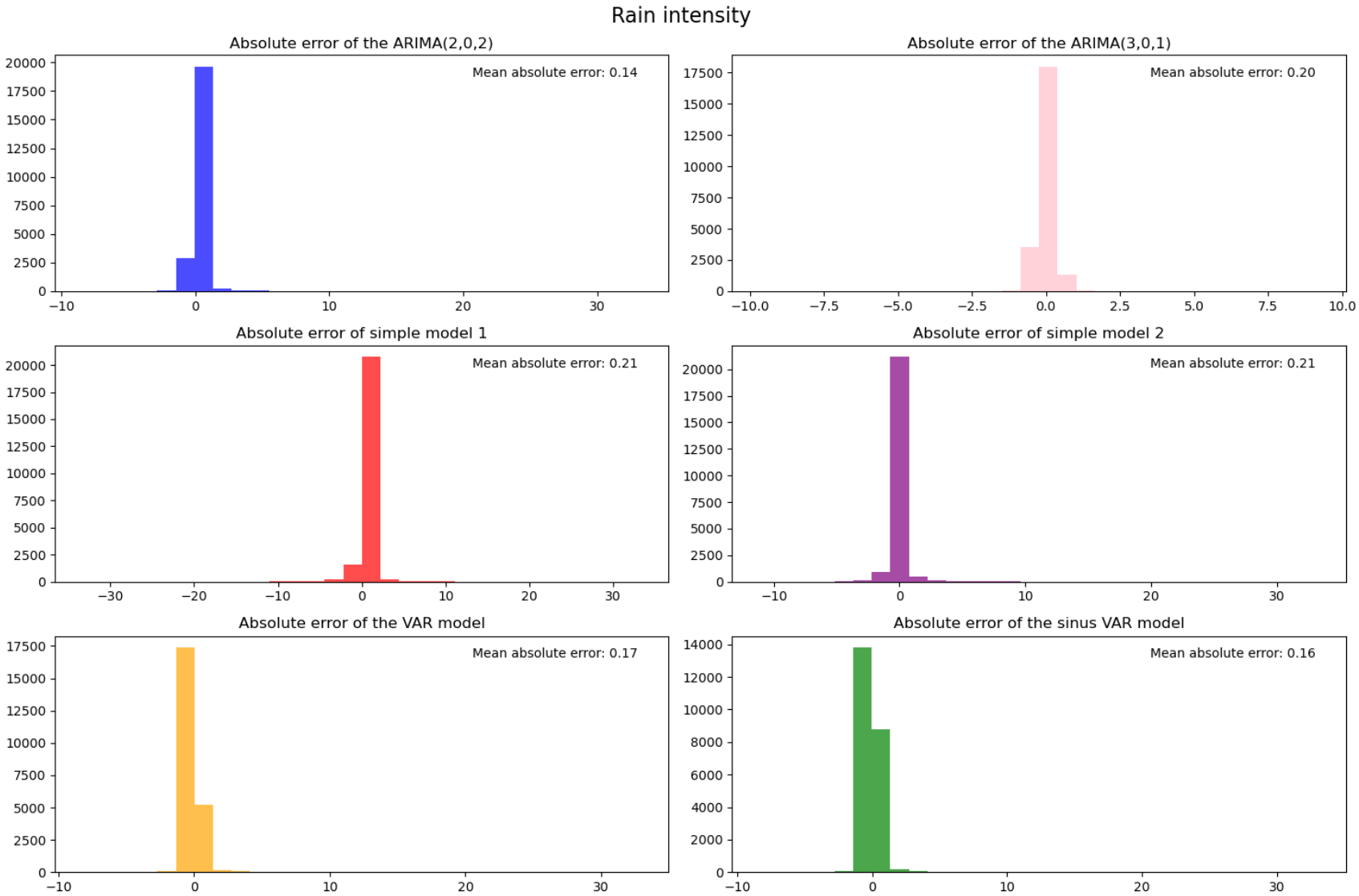
Figure: Absolute error of the solar radiation intensity $ARIMA$ model

Since the radiation is significantly lower in the night, the average absolute errors should also be and are lower.

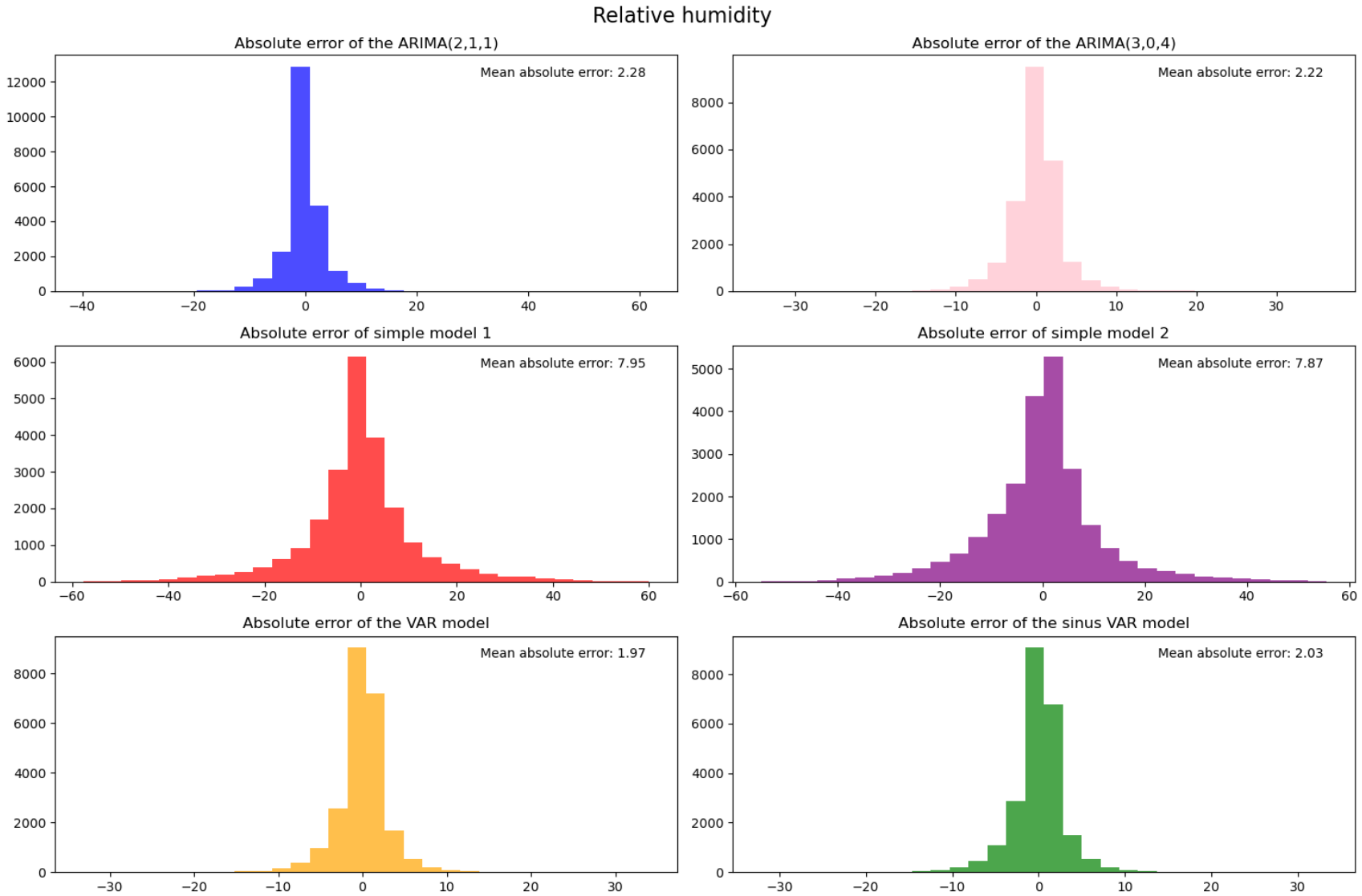
Air pressure



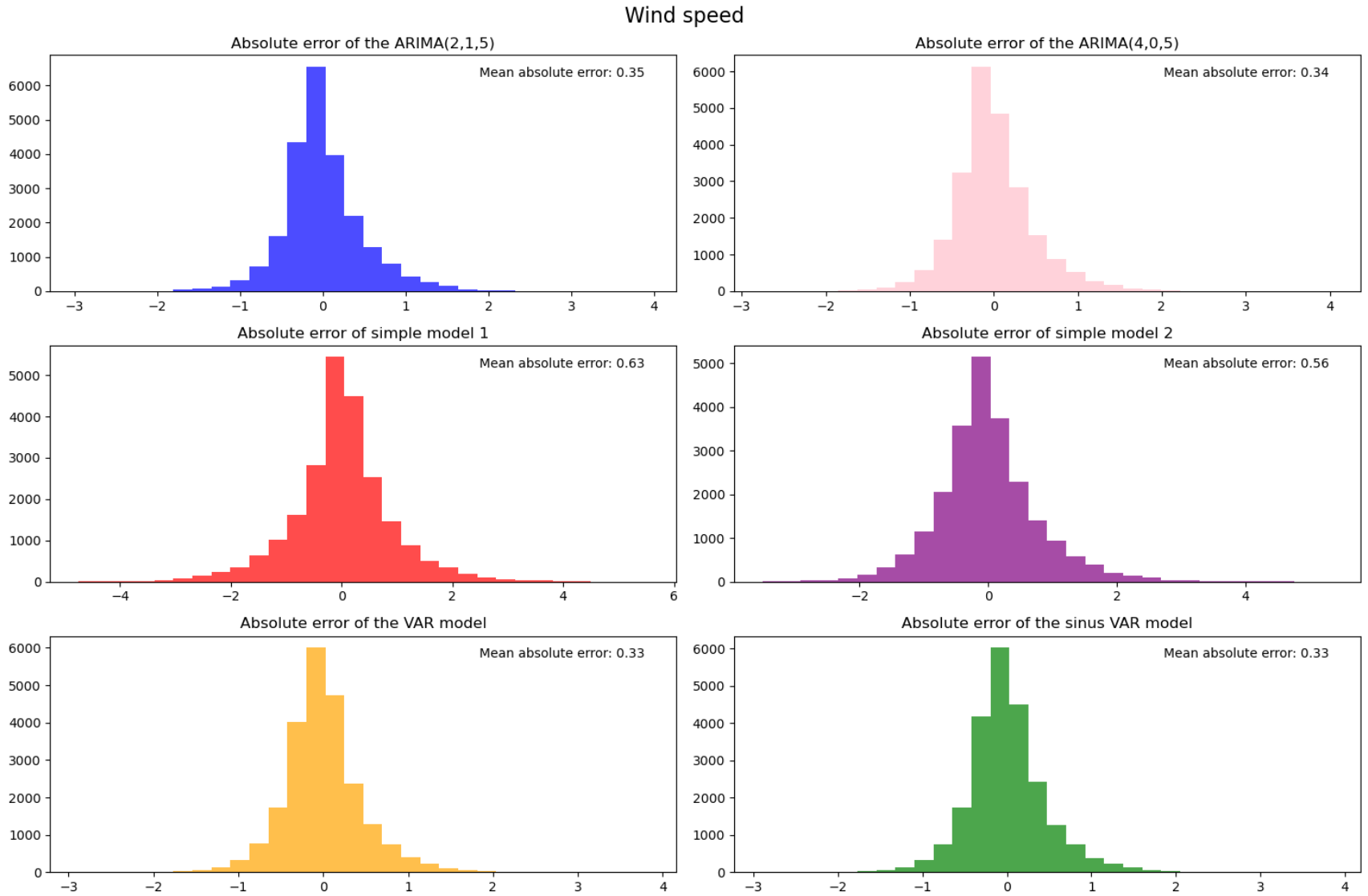
Rain intensity



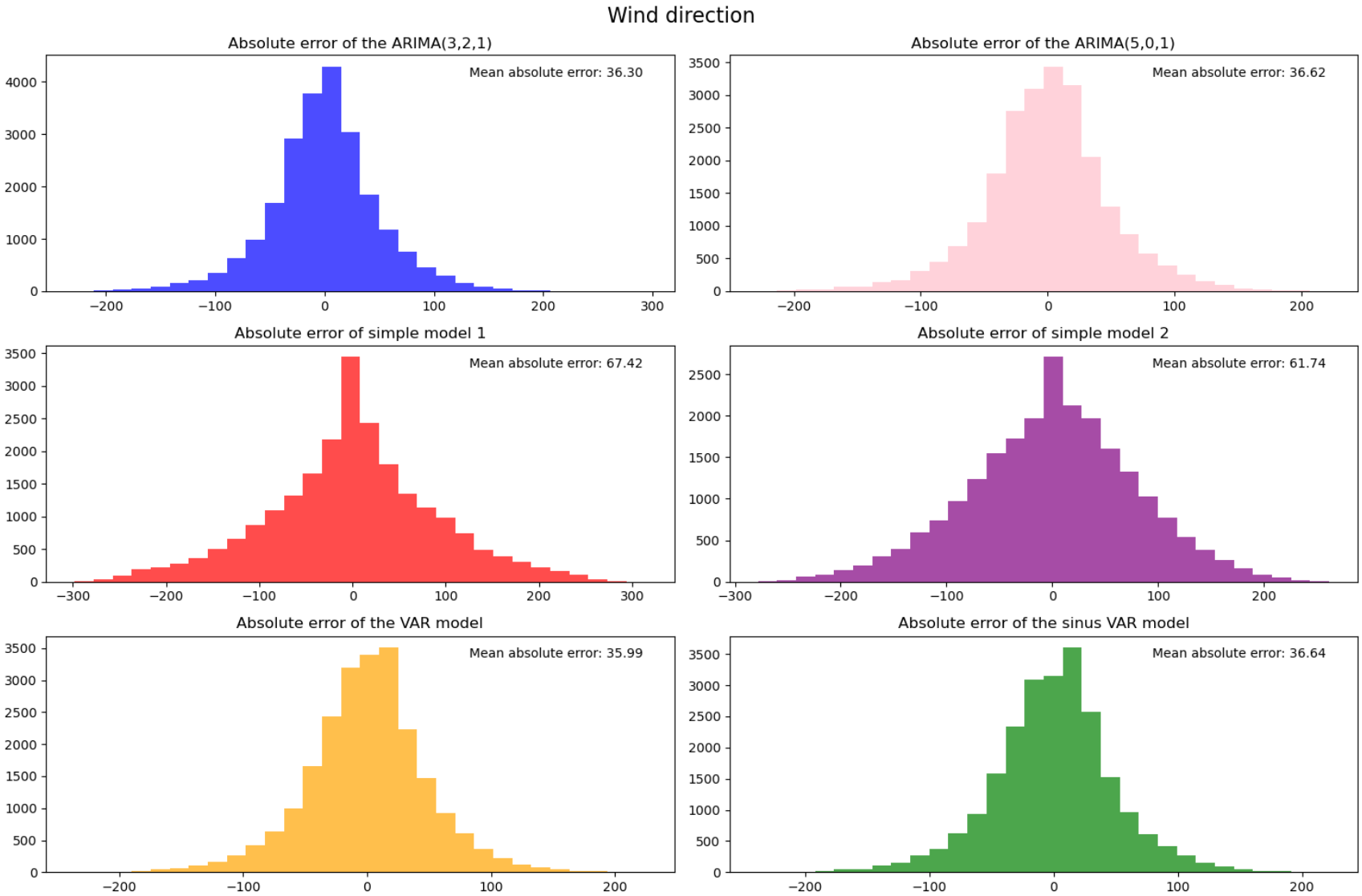
Relative humidity



Wind speed


Wind speed
Absolute error of the ARIMA(2,1,5)
Mean absolute error: 0.35
Absolute error of the ARIMA(4,0,5)
Mean absolute error: 0.34
Absolute error of simple model 1
Mean absolute error: 0.63
Absolute error of simple model 2
Mean absolute error: 0.56
Absolute error of the VAR model
Mean absolute error: 0.33
Absolute error of the sinus VAR model
Mean absolute error: 0.33

Wind direction



Main results

- $ARIMA$ and VAR models fit the data better than simple models.
- Simple models **mostly** made better forecasts than $ARIMA$ and VAR models.

Further research proposal

- Longer time series,
- $SARIMA$ instead of $ARIMA$,
- more training data sets,
- multivariate $ARIMA$ model instead of VAR model.

Sources I

- Bojan Basrak, *ESSENTIALS OF TIME SERIES (Lecture notes), with examples in R*, Department of Mathematics, University of Zagreb.
- Hyndman, R.J., & Athanasopoulos, G., *Forecasting: principles and practice, 3rd edition*, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on 20. 5. 2024.
- Param Raval , *Vector auto-regressive model in Python*,https://www.analyticsvidhya.com/blog/2021/08/vector-autoregressive-model-in-python/, Accessed on 1. 6. 2024.
- Rabi Kumar Singh , *Auto ARIMA on Multivariate Time Series*, https://www.kaggle.com/code/jurk06/auto-arima-on-multivariate-time-series, Accessed on 20. 5. 2024.

Sources II

- Robert Nau , *Statistical forecasting: notes on regression and time series analysis*, Fuqua School of Business Duke University, https://people.duke.edu/~rnau/411home.htm, Accessed on 20. 5. 2024.
- Shahanaj Parvin and Murshida Khanam, *Comparison Between ARIMA and VAR Model Regarding the Forecasting of the Price of Jute Goods in Bangladesh*.
- Param Raval , *How to Build ARIMA Model in Python for time series forecasting?*,https://www.projectpro.io/article/how-to-build-arima-model-in-python/544, Accessed on 20. 5. 2024.