Skoltech
Skolkovo Institute of Science and Technology

Online ARIMA

Students: Bespalov Iaroslav,
Katser Iurii,
Kozitsin Vyacheslav,
Makarov Edgar

Short description

Skoltech
Skolkovo Institute of Science and Technology

The ubiquity of time series is a fact of modern-day life: from stock market data to social media information; from search-engine metadata to webpage analytics, modern-day data exists as a continuous flow of information indexed by timestamps.

ARIMA models are the main time-series long-term forecasting models in many practical applications. In time series trends or dependencies usually change. This requires frequent refitting of the model from scratch or performing additional training. However, fitting ARIMA models to large-scale data (e.g. hourly time series for 10 years) is time consuming. In this case, the online learning approach helps.

Objectives

Skoltech
Skolkovo Institute of Science and Technology

For a real-time application, online modeling is desirable — a self-generating and self-updating forecasting method is desirable for producing a timely forecast. The ARIMA method in time series is an effective means to address a forecast; it has a solid foundation in classical probability theory and mathematical statistics.

I. Implement the ARIMA online Gradient Descent algorithm.
II. Compare with common online learning approach:
    1. complete refitting on data of steps [1, t] each step t
    2. additional fitting (warm_start) for every new observation
    3. rolling-window refitting on [t-h, t]

ARMA/ARIMA

Skoltech
Skolkovo Institute of Science and Technology

A time series is defined as a sequence of quantitative observations at successive time. We assume time is a discrete variable, X_t denotes the observation at time t, and ε_t denotes the zero-mean random noise term at time t. The MA(q) (short for Moving Average) model considers the process: $X_t = \sum_{i=1}^{q} \beta_i \epsilon_{t-i} + \epsilon_t$, where β_i is a coefficient. Similar to MA(q) models, Autoregression model, denoted by AR(k), satisfies $X_t = \sum_{i=1}^{k} \alpha_i X_{t-i} + \epsilon_t$. In other words, it assumes each X_t is a noisy linear combination of the previous k observations. This is similar to traditional multiple regression model, but X_t is regressed on past values of X_t.

ARMA: $$X_t = \sum_{i=1}^{q} \beta_i \epsilon_{t-i} + \sum_{i=1}^{k} \alpha_i X_{t-i} + \epsilon_t$$, where ε_t are zero-mean noise term.

ARMA model describing stationary series.

ARIMA: $$\nabla^d X_t = \sum_{i=1}^{q} \beta_i \epsilon_{t-i} + \sum_{i=1}^{k} \alpha_i \nabla^d X_{t-i} + \epsilon_t$$,which are parameterized by three terms k, d, q and weights vector $\alpha \in R^k$ and $\beta \in R^q$.

ARIMA model describing non-stationary series.

Description of methods:

Skoltech
Skolkovo Institute of Science and Technology

- **Complete refitting**

This method produces complete refitting at each step t of our model on the data of steps [1, t].

- **Rolling-window refitting**

This method produces rolling-window refitting at each step t on the data of steps [t − h, t].

- **Warm start**

This method uses built-in functions of statsmodel.sarimax by using params of the previously fitted model to speed up convergence of online fitting model at each step t.

- **Gradient descent. Real online**

We now implement a more general online convex optimization solver, Online Gradient Descent, which is applicable to any convex loss functions. The projection $\prod_{\mathcal{K}}(y)$ refers to the Euclidean projection onto K, i.e., $\prod_{\mathcal{K}}(y) = \arg\min_{x \in \mathcal{K}} \|y - x\|_2$

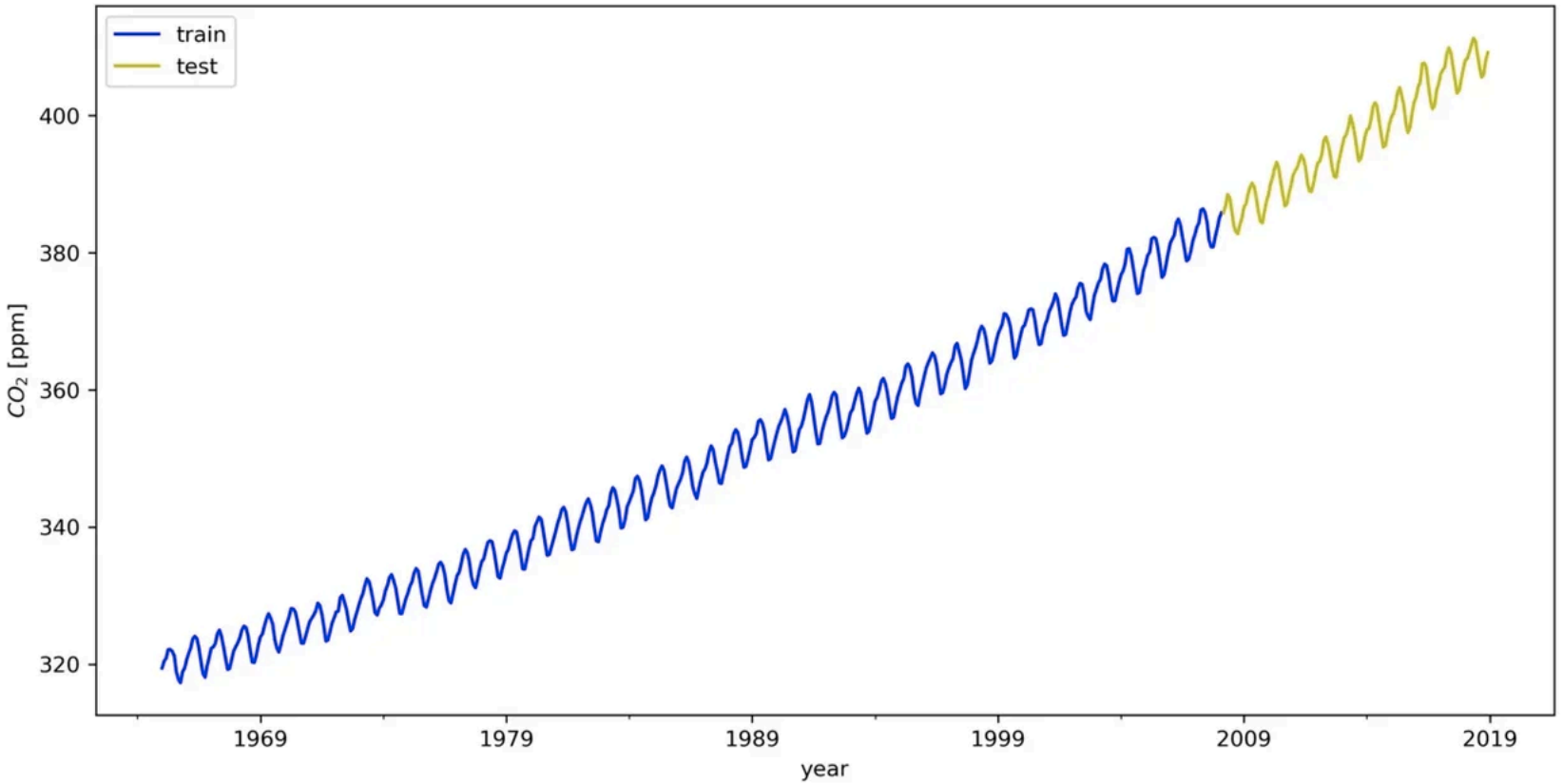
Input: parameter k, d, q; learning rate η.
Set $m = \log_{\lambda_{max}}((TLM_{max}q)^{-1})$.
for t = 1 **to** $T - 1$ **do**
  predict $\tilde{X}_t(\gamma^t) = \sum_{i=1}^{k+m} \gamma_i \nabla^d X_{t-i} + \sum_{i=0}^{d-1} \nabla^i X_{t-1}$;
  receive X_t and incur loss $\ell_t^m(\gamma^t)$;
  Let $\nabla_t = \nabla \ell_t^m(\gamma^t)$;
  Set $\gamma^{t+1} \leftarrow \prod_{\mathcal{K}}(\gamma^t - \frac{1}{\eta}\nabla_t)$;
end for

ARIMA- OGD algorithm

First dataset

Skoltech
Skolkovo Institute of Science and Technology

The first Dataset is Atmospheric CO_2 Data. After dropping NaNs and dividing Dataset for train and test we get following.



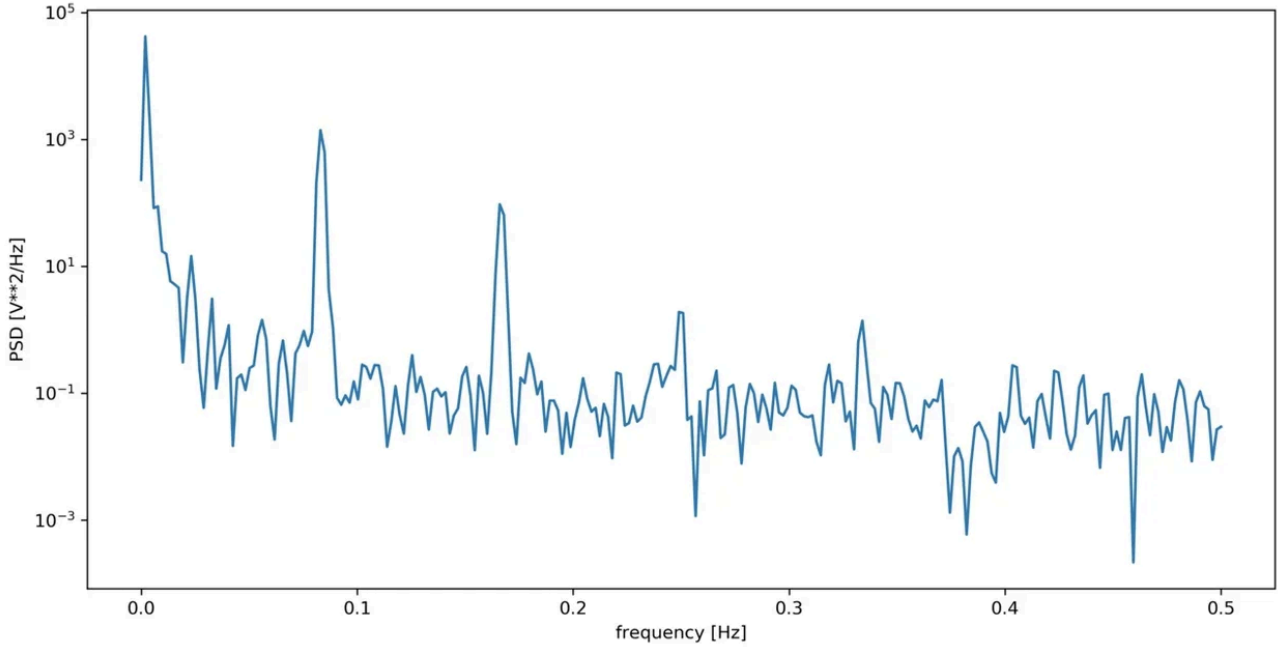
Atmospheric CO_2 Dataset

Data analysis

Skoltech
Skolkovo Institute of Science and Technology

- Estimate spectral density using Welch's method
- Selection of the period for seasonal differences to delete seasonal component using SARIMA
- Provide Augmented Dickev-Fuller test
- Selection of the period for first differences to delete trend component
- Provide Augmented Dickey-Fuller test again to approve non-stationarity of the received Dataset

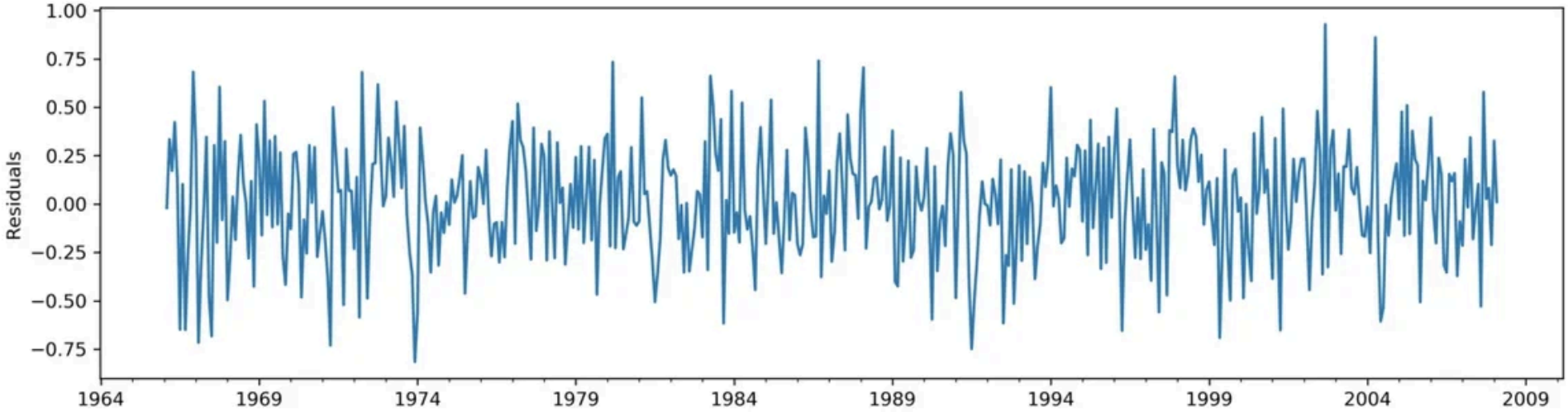
12 month - the main time period

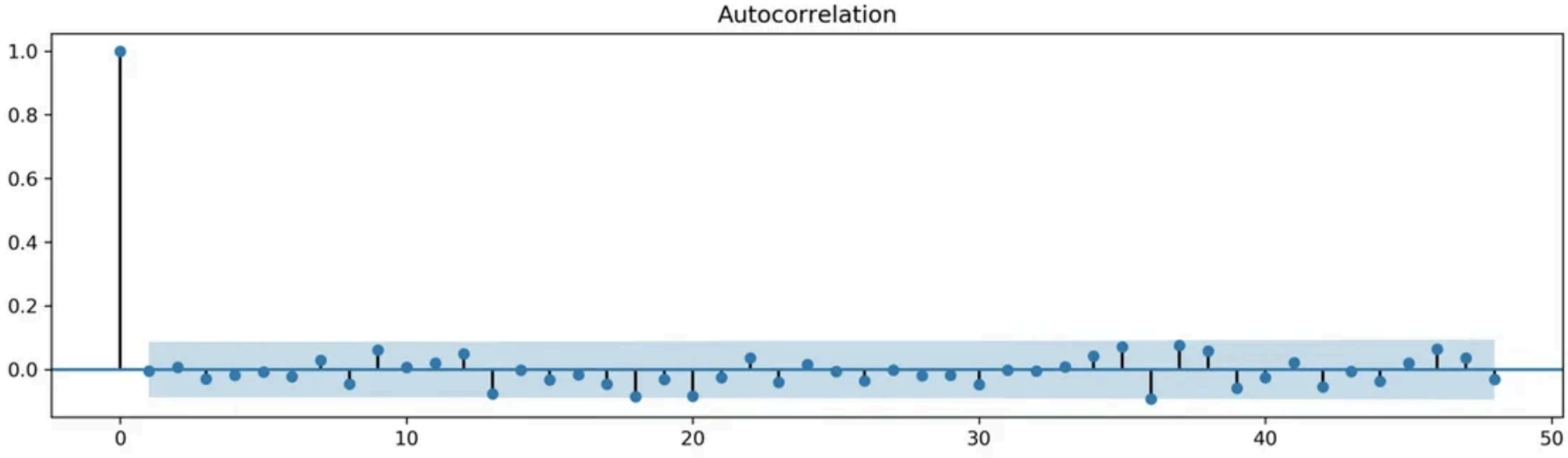


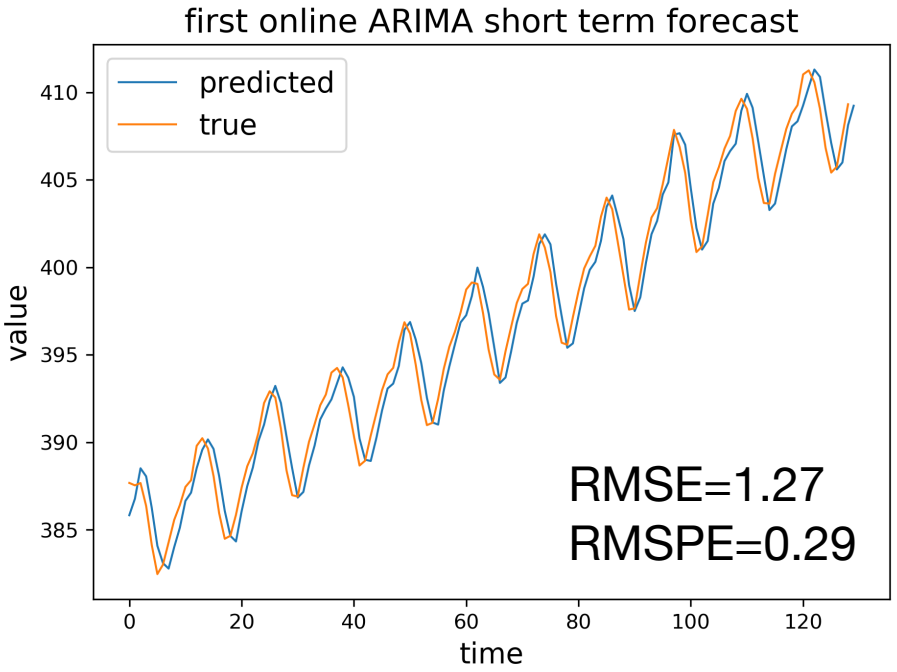
Spectral density estimation using Welch's method

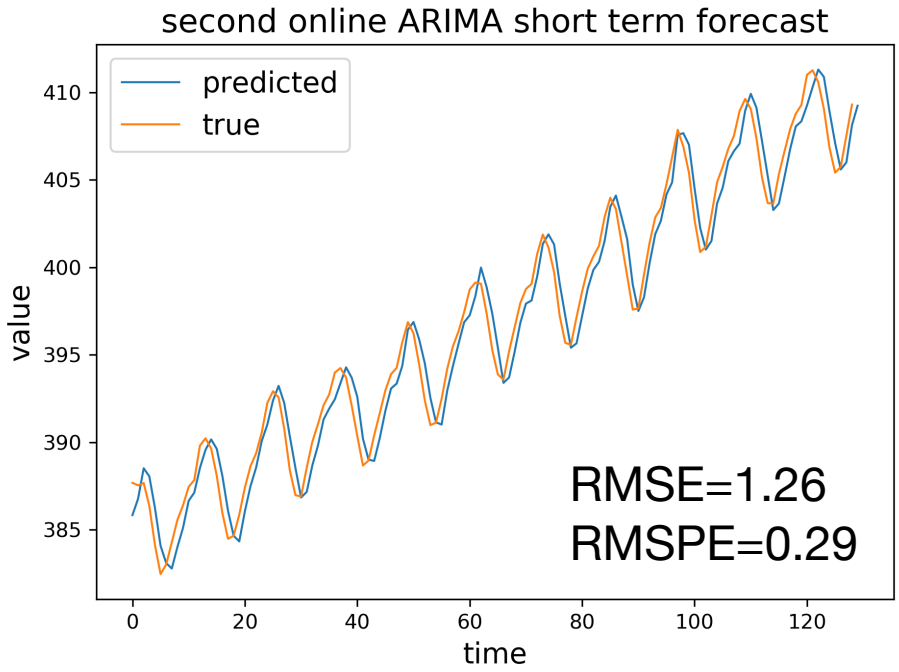
Residuals analysis

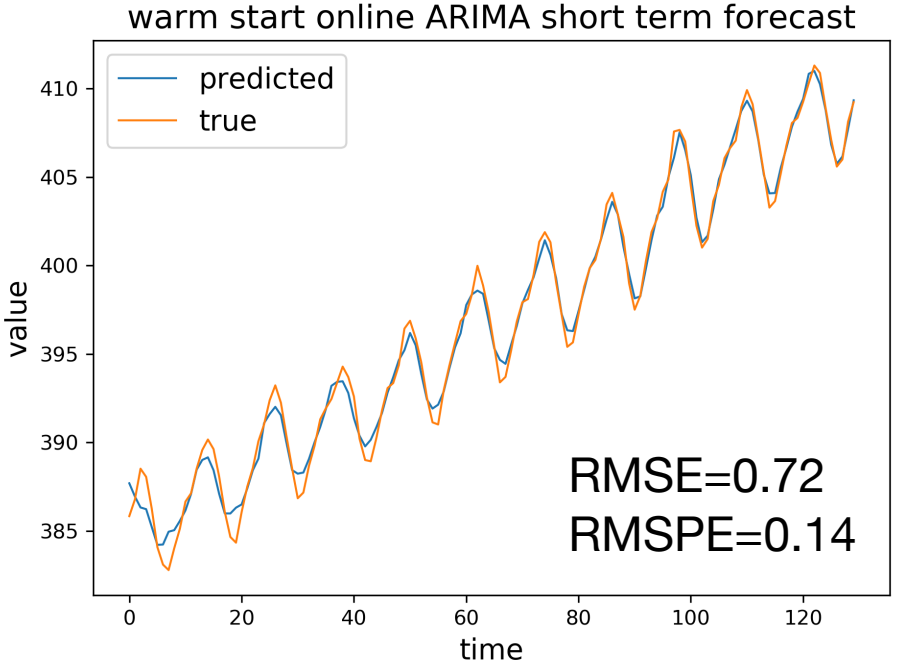
Skoltech
Skolkovo Institute of Science and Technology

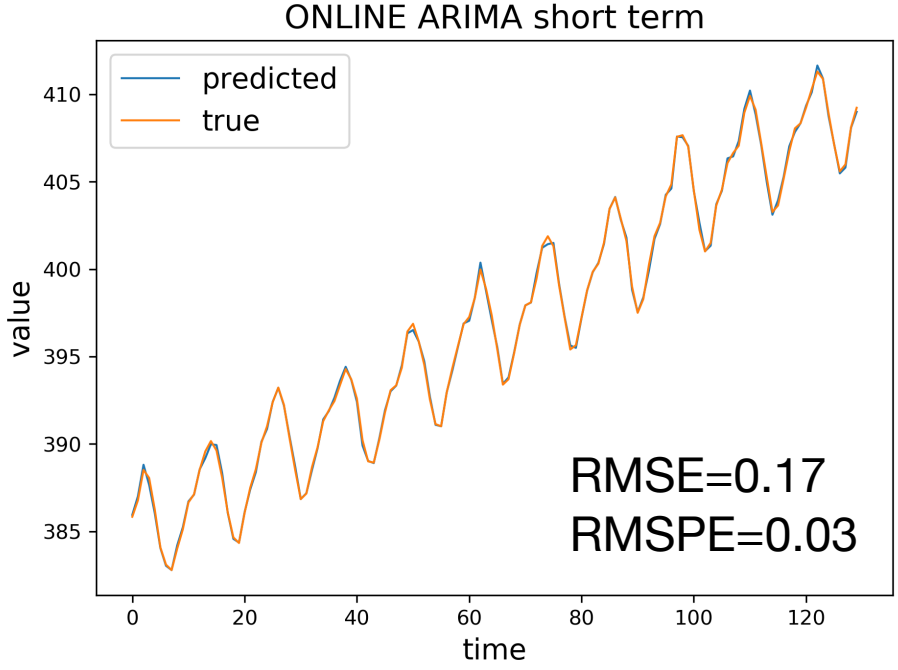


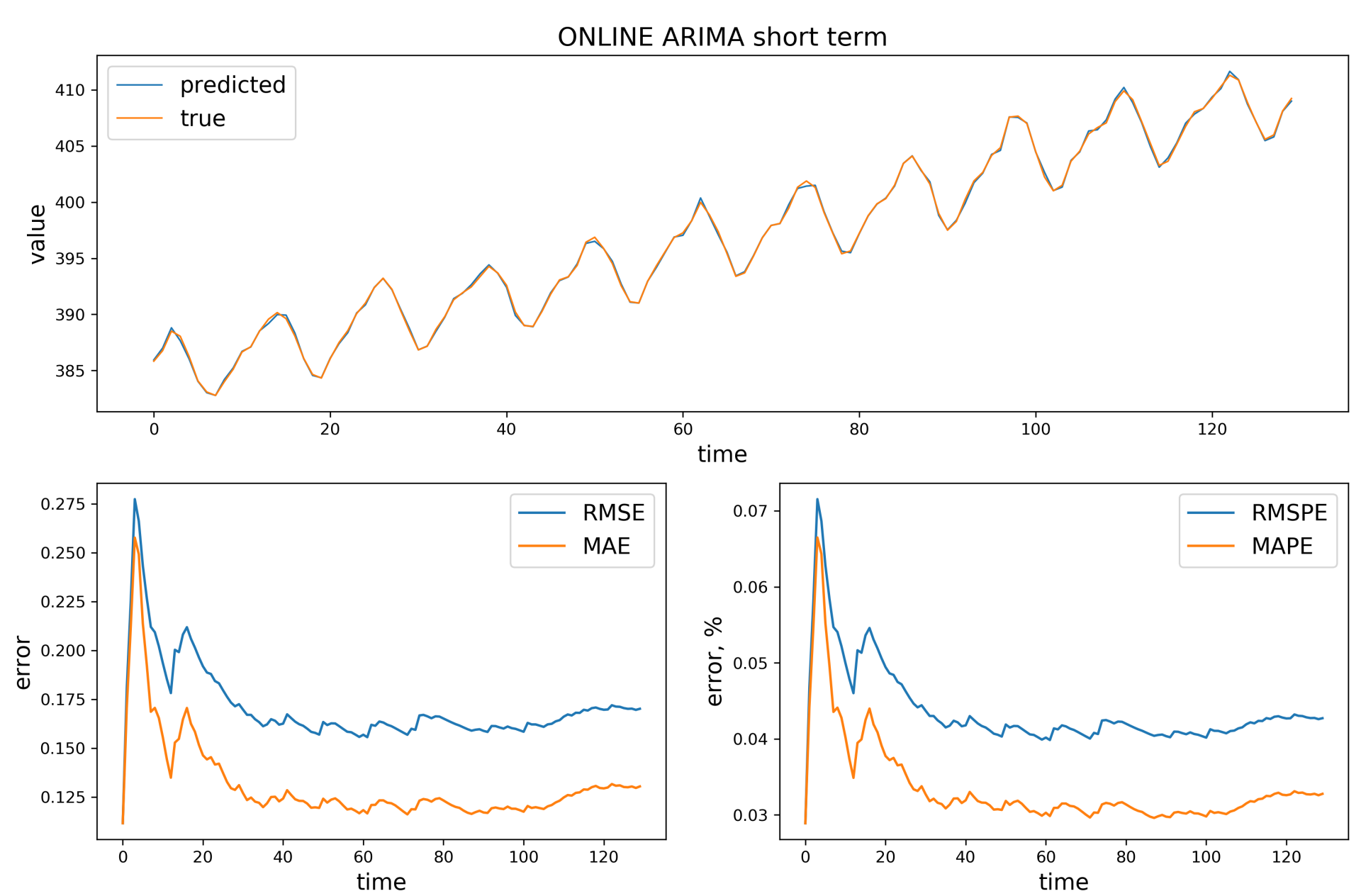


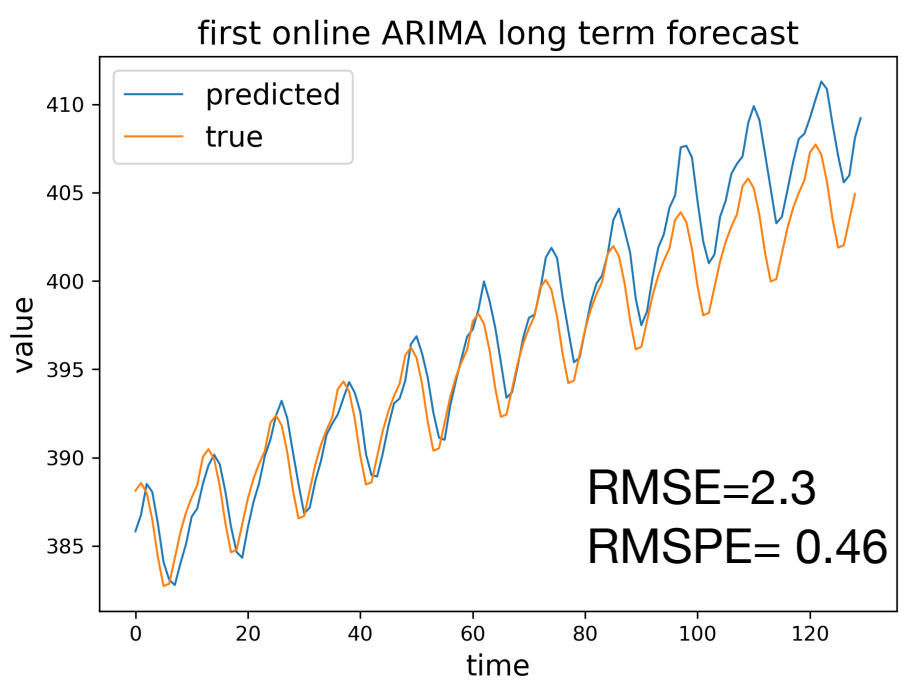

first online ARIMA short term forecast
predicted
true
RMSE=1.27
RMSPE=0.29
value
time

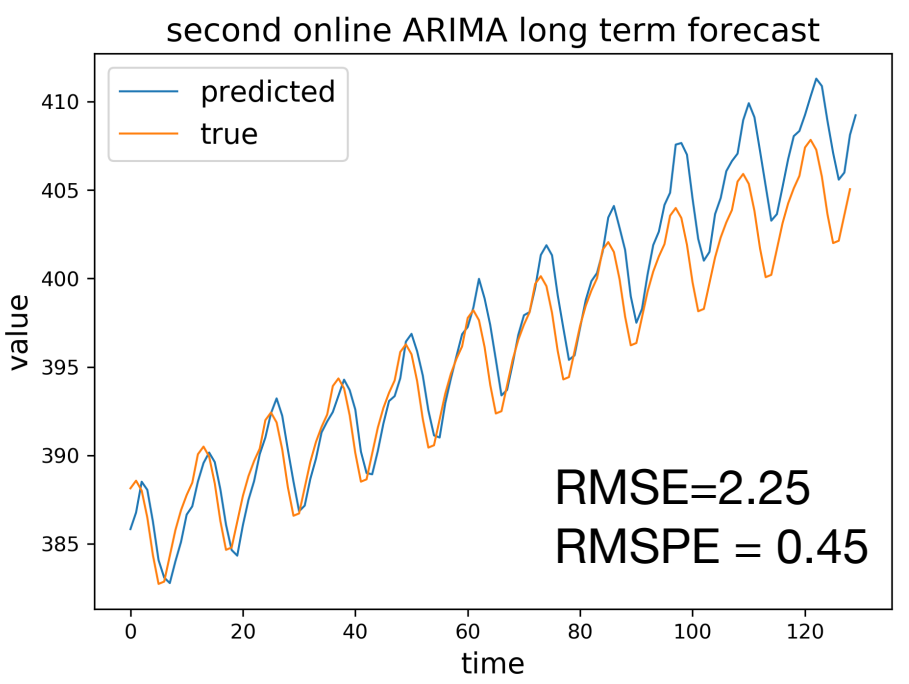

second online ARIMA short term forecast
predicted
true
RMSE=1.26
RMSPE=0.29
value
time

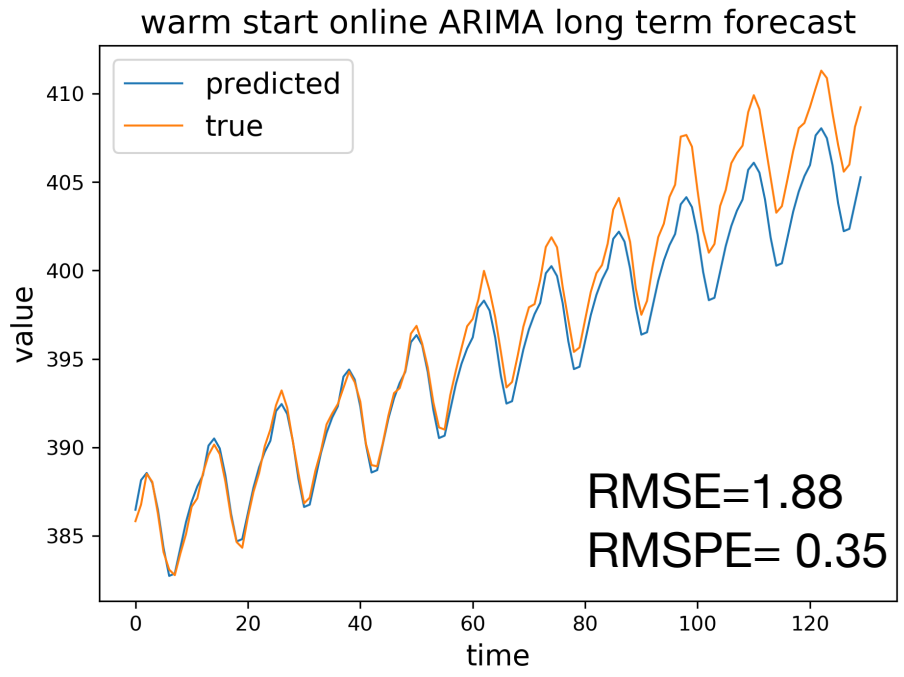

warm start online ARIMA short term forecast
predicted
true
RMSE=0.72
RMSPE=0.14
value
time

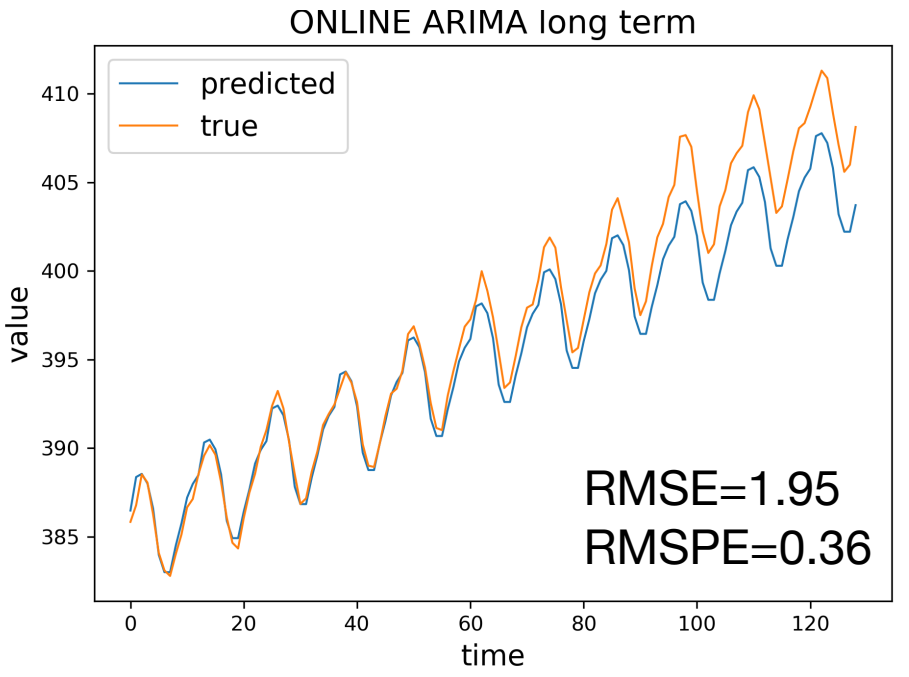

ONLINE ARIMA short term
predicted
true
RMSE=0.17
RMSPE=0.03
value
time

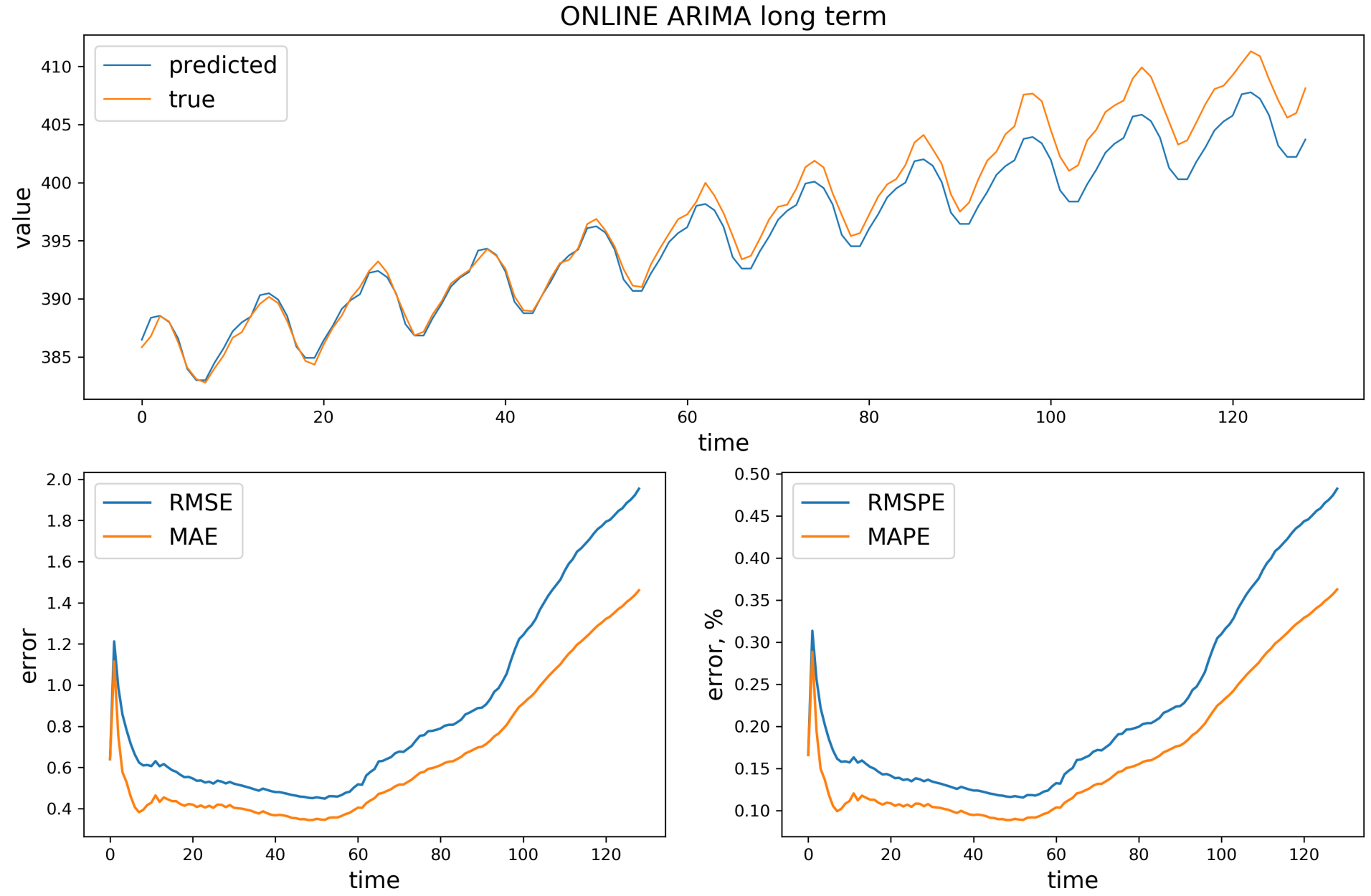

ONLINE ARIMA short term
predicted
true
value
time
RMSE
MAE
error
time
RMSPE
MAPE
error, %
time


first online ARIMA long term forecast
predicted
true
value
time
RMSE=2.3
RMSPE= 0.46


second online ARIMA long term forecast
predicted
true
value
time
RMSE=2.25
RMSPE = 0.45


warm start online ARIMA long term forecast
predicted
true
value
time
RMSE=1.88
RMSPE= 0.35

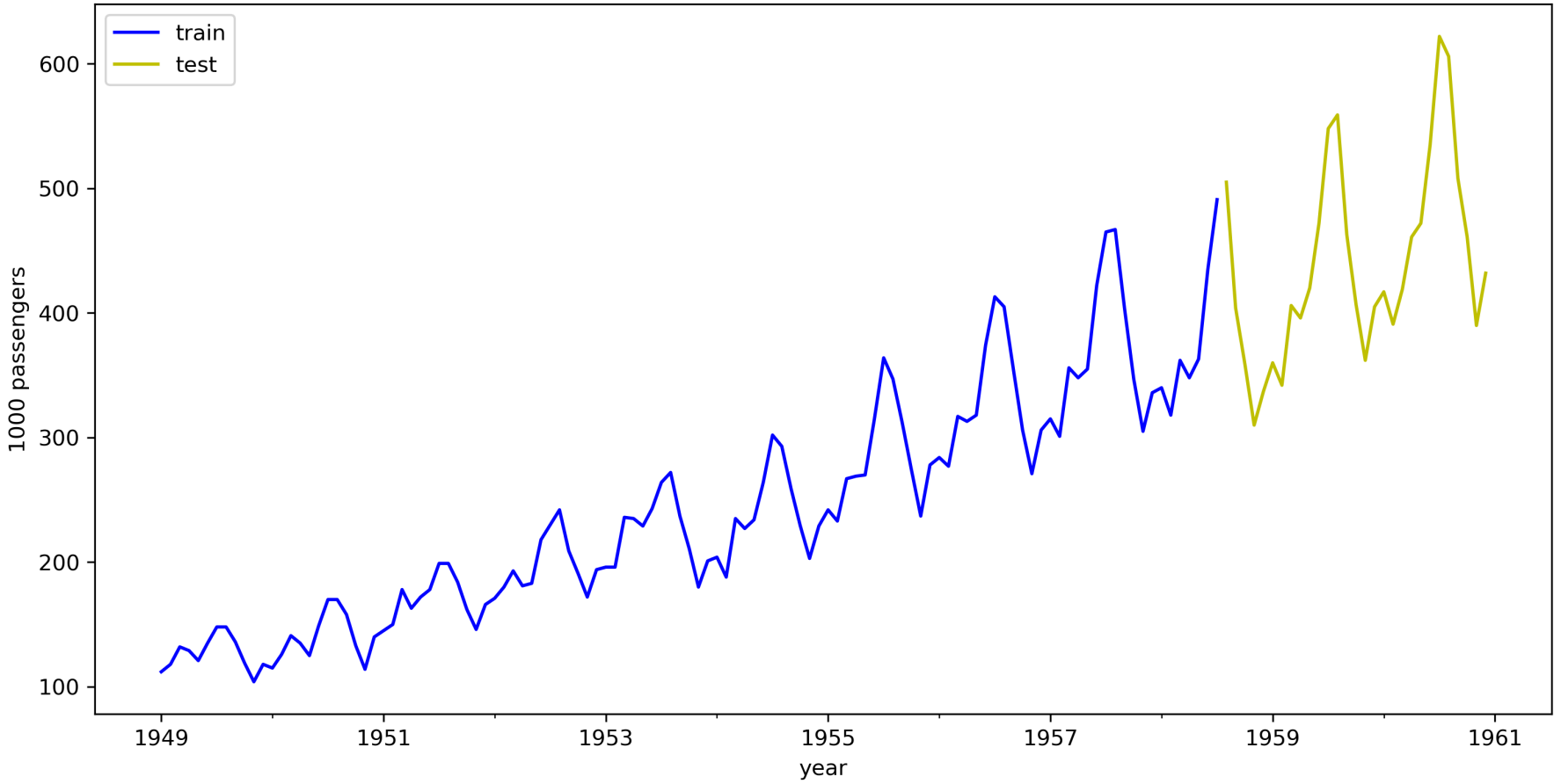

ONLINE ARIMA long term
predicted
true
value
time
RMSE=1.95
RMSPE=0.36

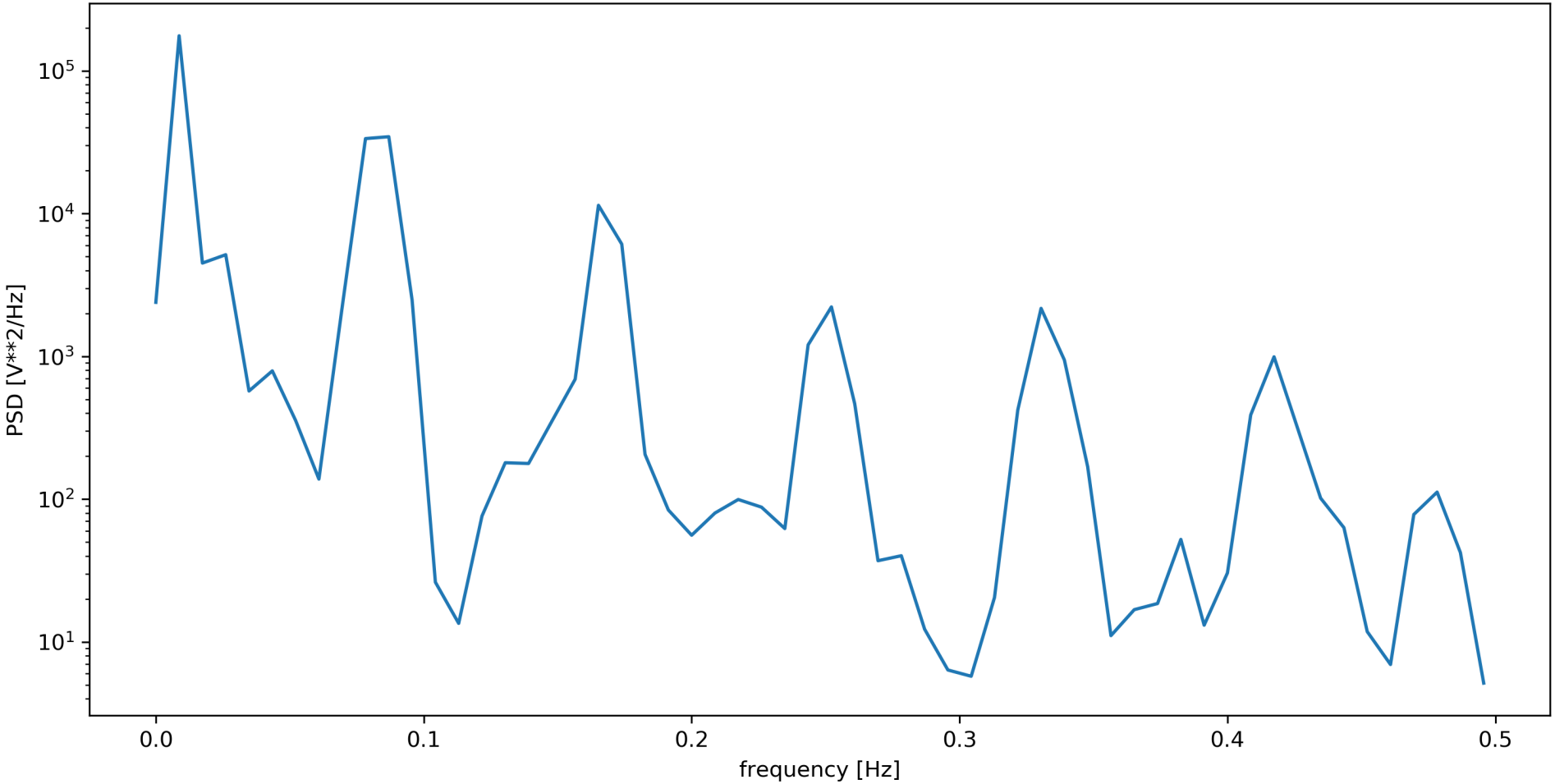

ONLINE ARIMA long term
predicted
true
value
time
RMSE
MAE
error
time
RMSPE
MAPE
error, %
time

Second Dataset

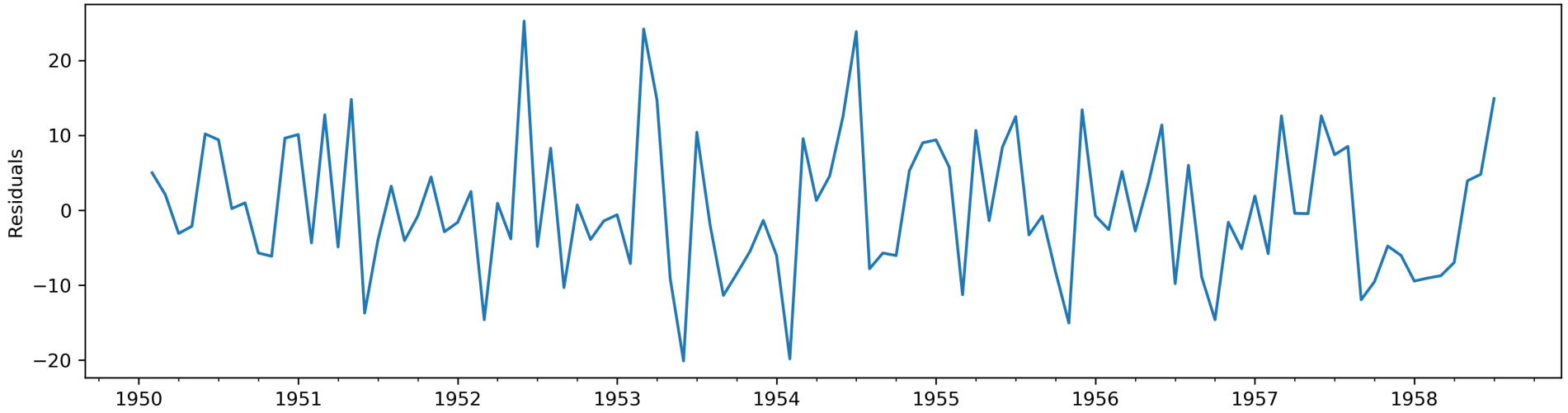
Skoltech
Skolkovo Institute of Science and Technology

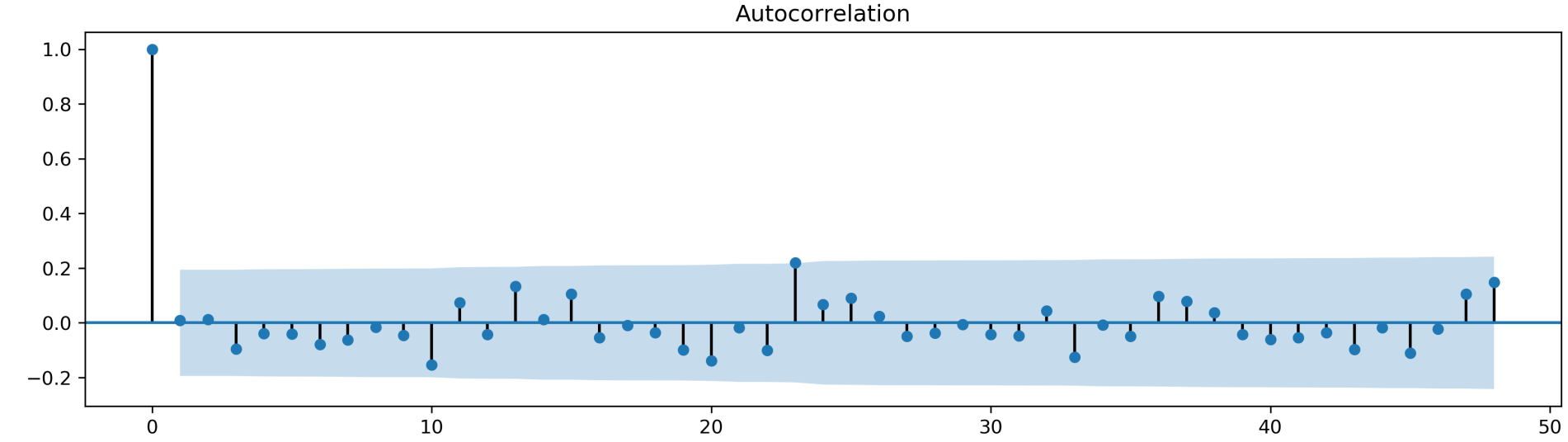
The second Dataset is International airline passengers: monthly totals in thousands. After dropping NaNs and deviding Dataset for train and test second dataset looks like at the figure 3. Then we provide all the procedures described earlier for the first Dataset.

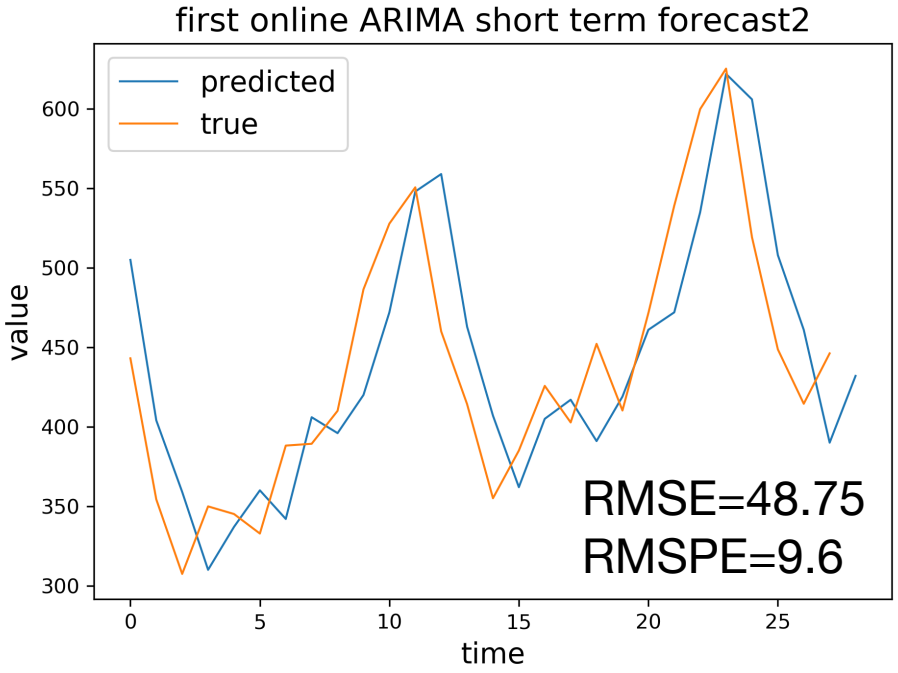


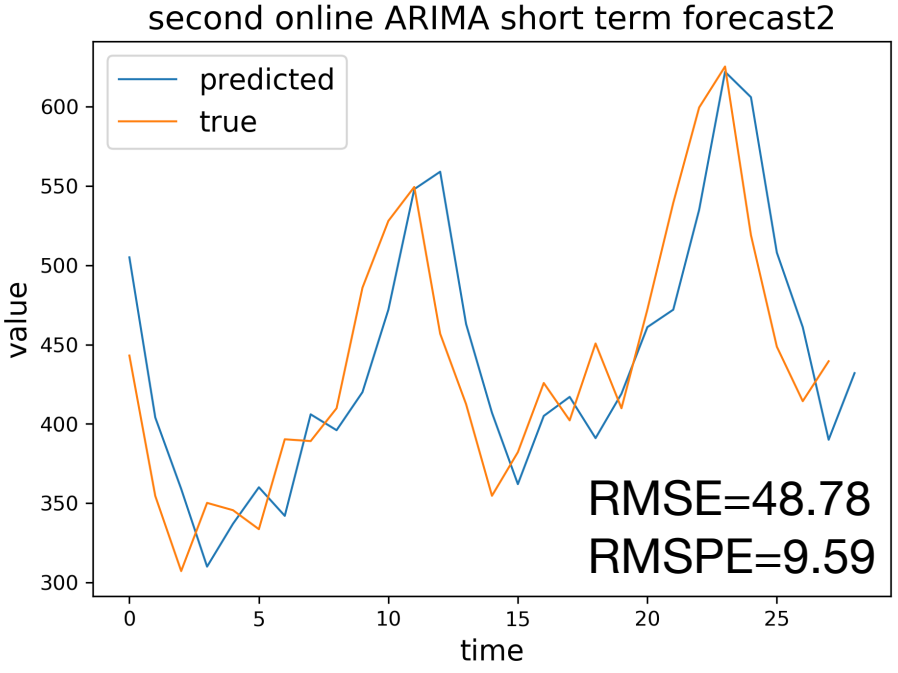


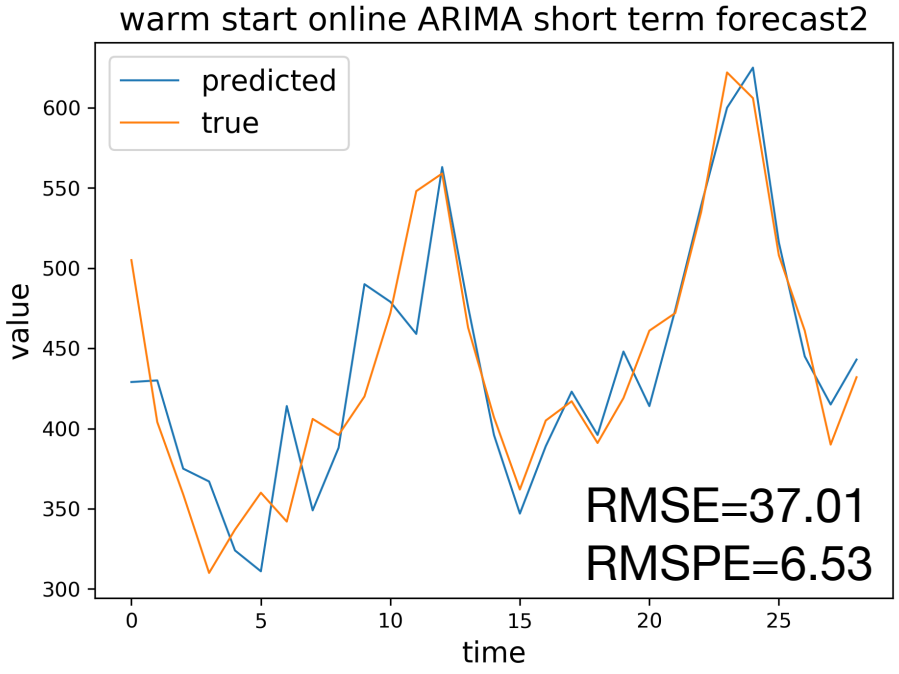
Spectral density estimation using Welch's method

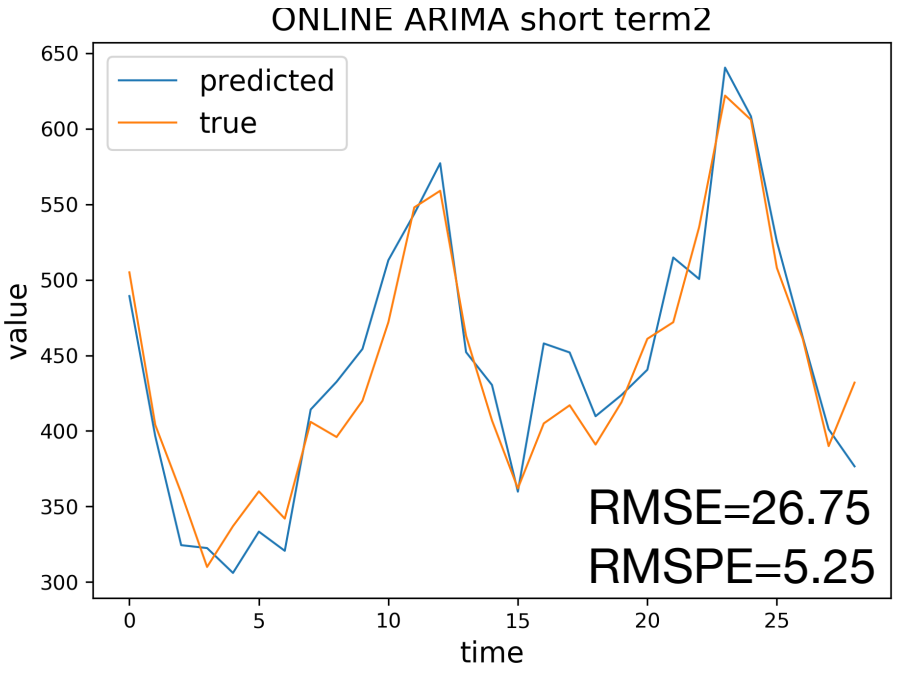

Residuals
20
10
0
−10
−20
1950
1951
1952
1953
1954
1955
1956
1957
1958

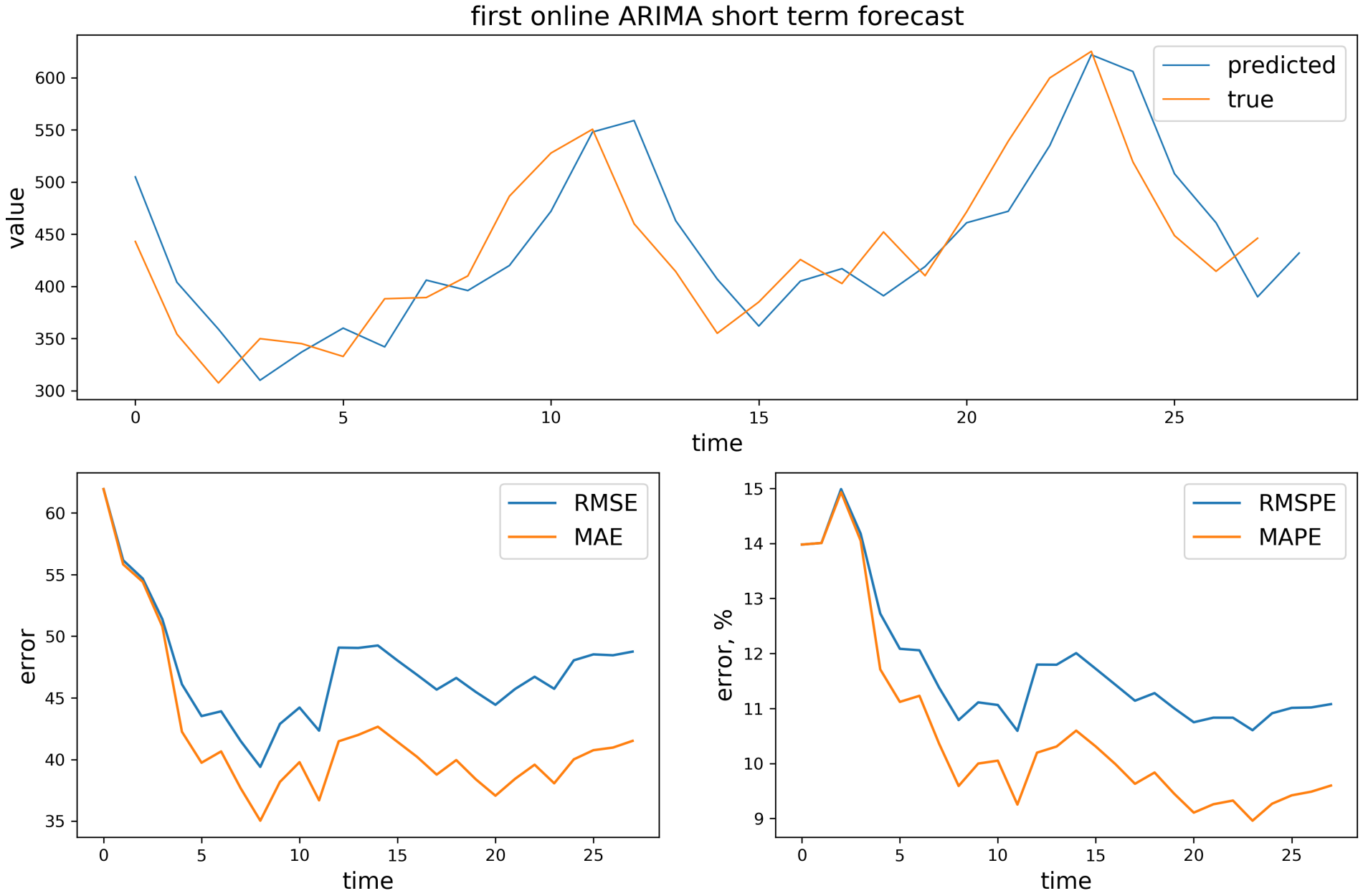

Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50

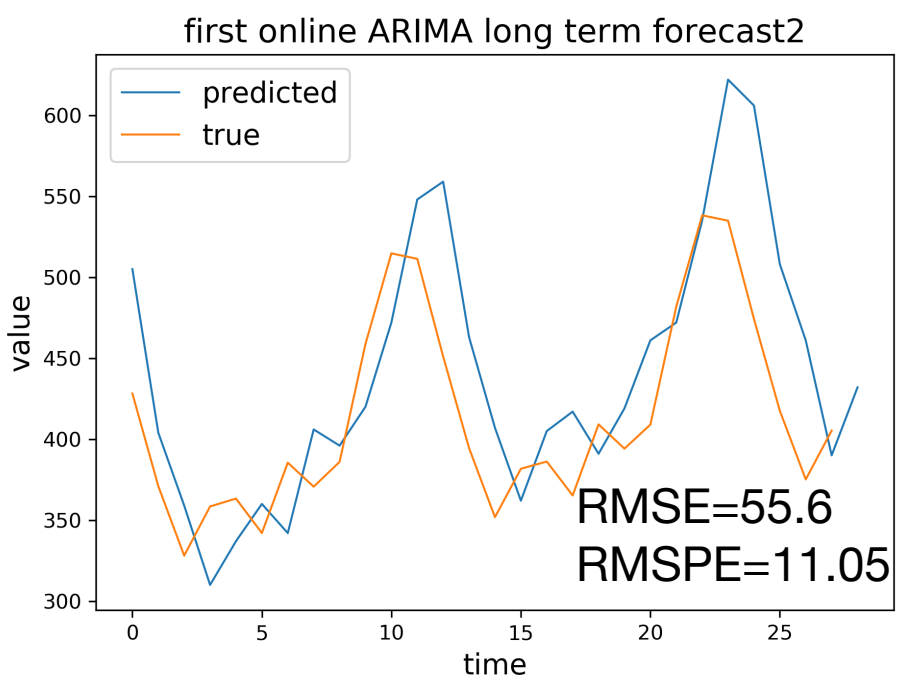

first online ARIMA short term forecast2
predicted
true
RMSE=48.75
RMSPE=9.6
value
time

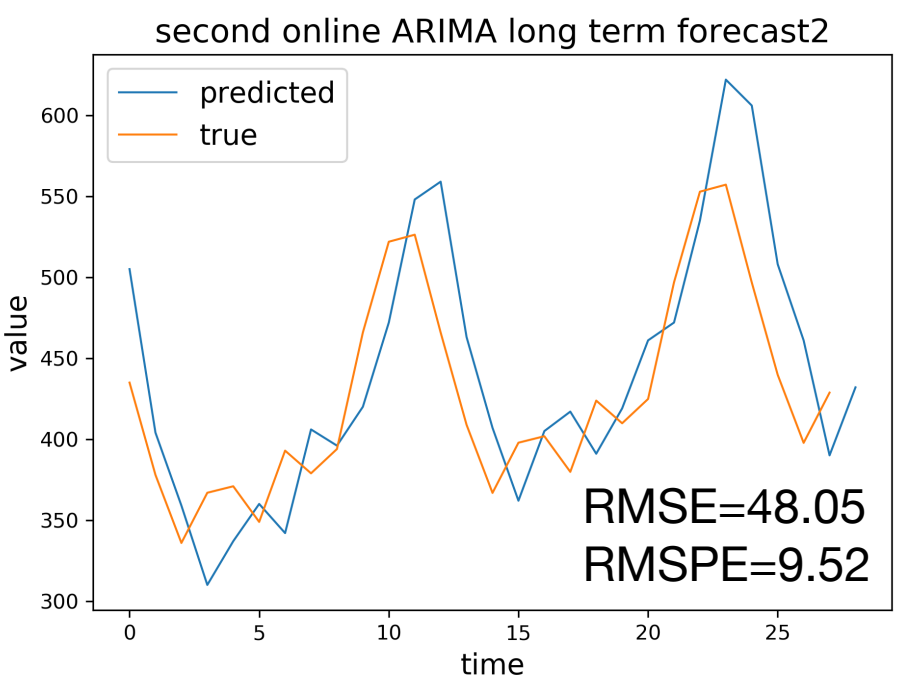

second online ARIMA short term forecast2
predicted
true
RMSE=48.78
RMSPE=9.59
value
time

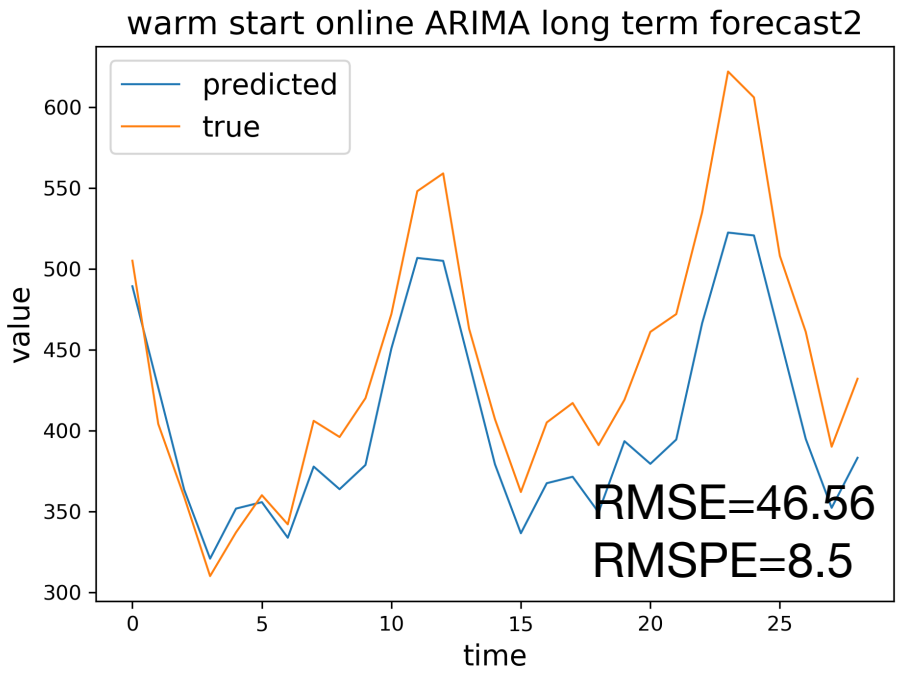

warm start online ARIMA short term forecast2
predicted
true
RMSE=37.01
RMSPE=6.53
value
time

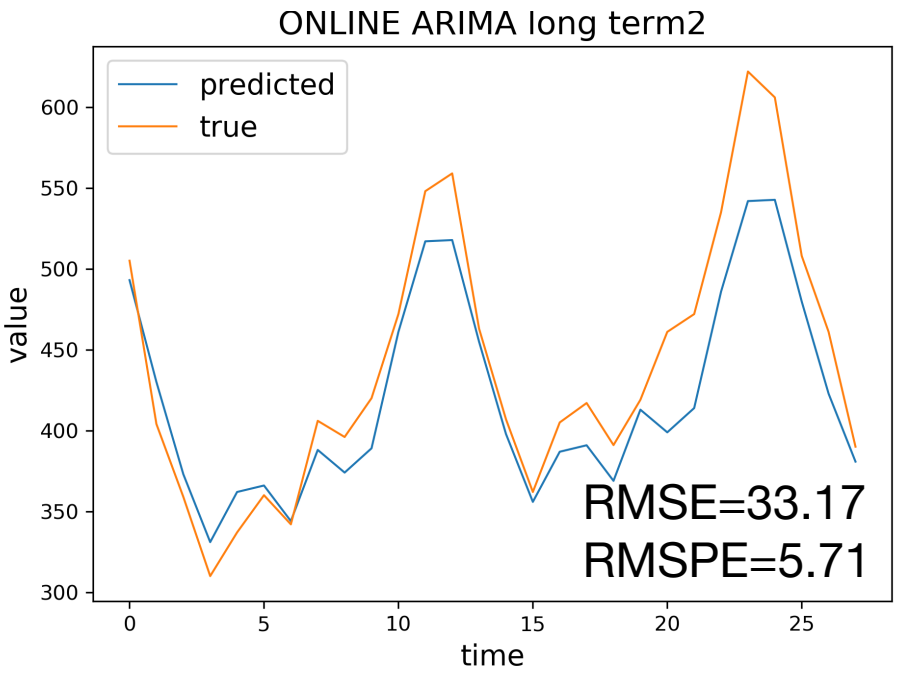

ONLINE ARIMA short term2
predicted
true
RMSE=26.75
RMSPE=5.25
value
time

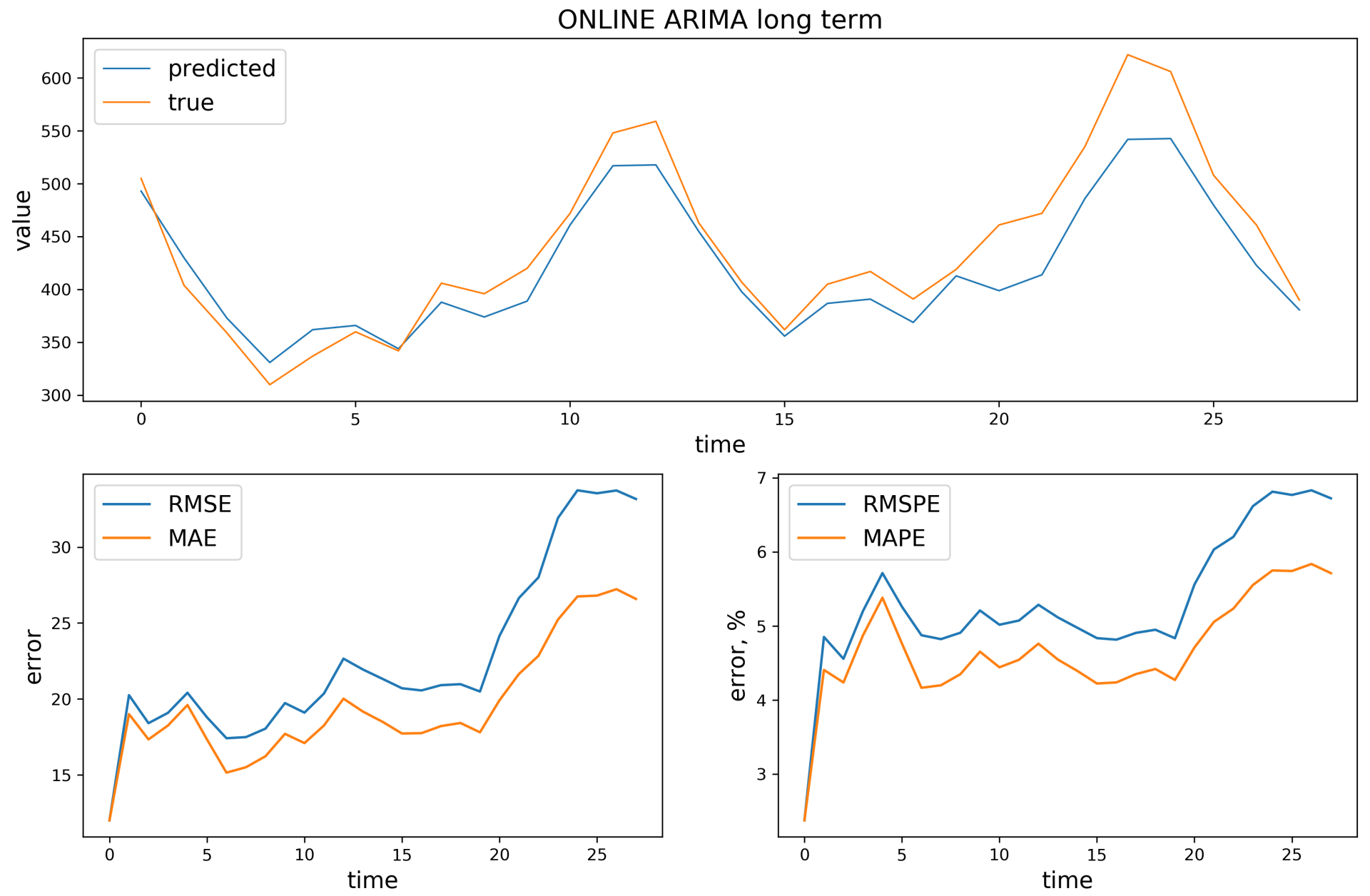

first online ARIMA short term forecast
predicted
true
value
time
RMSE
MAE
error
time
RMSPE
MAPE
error, %
time


first online ARIMA long term forecast2
predicted
true
value
time
RMSE=55.6
RMSPE=11.05


second online ARIMA long term forecast2
predicted
true
value
time
RMSE=48.05
RMSPE=9.52


warm start online ARIMA long term forecast2
predicted
true
value
time
RMSE=46.56
RMSPE=8.5


ONLINE ARIMA long term2
predicted
true
value
time
RMSE=33.17
RMSPE=5.71


ONLINE ARIMA long term
predicted
true
value
time
RMSE
MAE
error
time
RMSPE
MAPE
error, %
time

Computational efficiency

Skoltech
Skolkovo Institute of Science and Technology

Method	Time in msec/sample			
	short-term		long-term	
	fit	pred	fit	pred
Complete refit	180.9	14.3	134	13.2
Roll-window refit	130.3	12.4	127.9	10.7
Warm start	2.15	3.61	0.094	0.197
Online ARIMA-GD	0.0074	0.0065	0.0079	0.0086

Conclusion

Skoltech
Skolkovo Institute of Science and Technology

We empirically compared our four online ARIMA algorithms, in which the promising results on two different datasets validate that all of the algorithms are effective and promising for time series prediction. T Online ARIMA method show great results, and efficient.

Further research of this topic will include practical task analysis and improvements of implemented algorithms attempts for online: time reduction without reduced forecasting accuracy.

Thank you for attention!

Questions?

In [16]:
```
print("Dickey-Fuller criterion: p = {0}, used lag = {1}".format(round(sm.tsa.stattools.adfuller(train['CO2 [ppm]'])[1],4)
                                                                ,sm.tsa.stattools.adfuller(train['CO2 [ppm]'])[2]))
```

```
Dickey-Fuller criterion: p = 0.999, used lag = 14
```

ADF test confirms the non-stationarity hypothesis with a high degree of confidence. Let us find seasonal difference and conduct ADF test again:

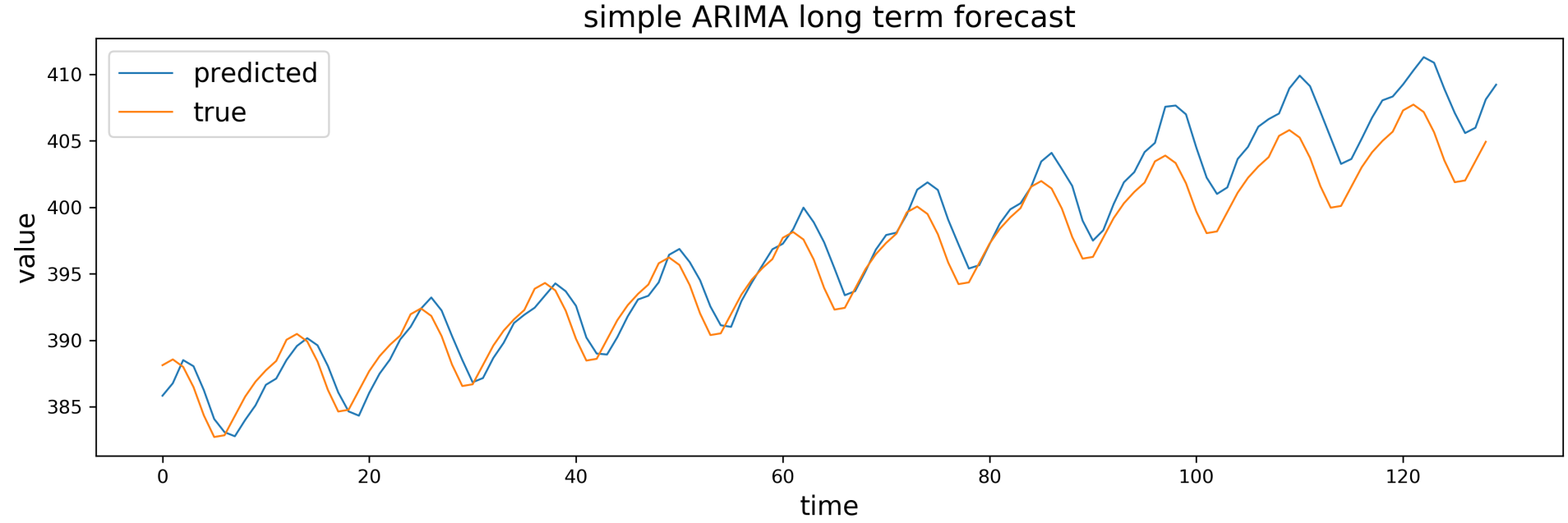
In [17]:
```
train['CO2 [ppm]_sdiff'] = train['CO2 [ppm]'] - train['CO2 [ppm]'].shift(12)
print("Dickey-Fuller criterion: p = {0}, used lag = {1}".format(round(sm.tsa.stattools.adfuller(train['CO2 [ppm]_sdiff'][12:
                                                                ,sm.tsa.stattools.adfuller(train['CO2 [ppm]_sdiff'][12:])[2]))
```

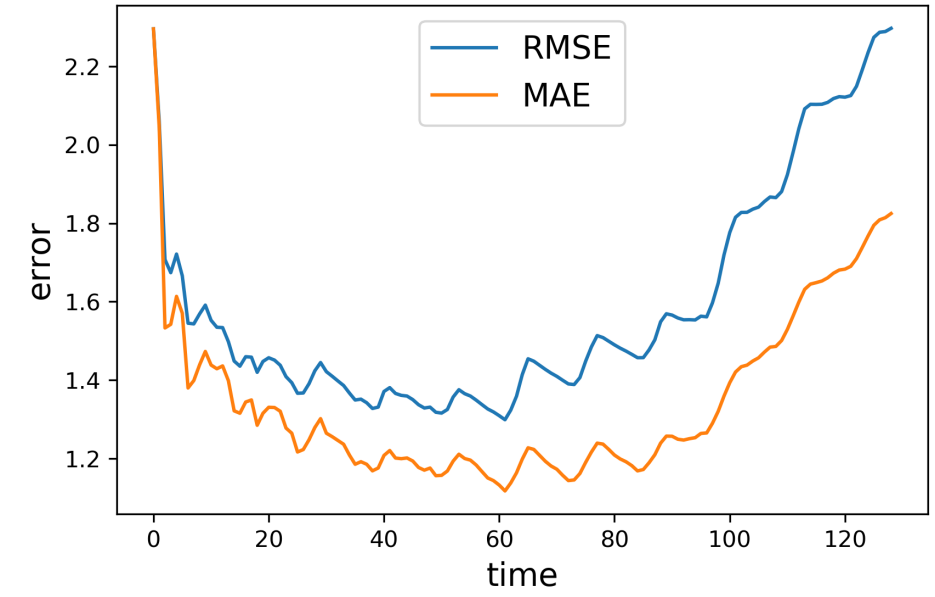
```
Dickey-Fuller criterion: p = 0.0005, used lag = 13
```

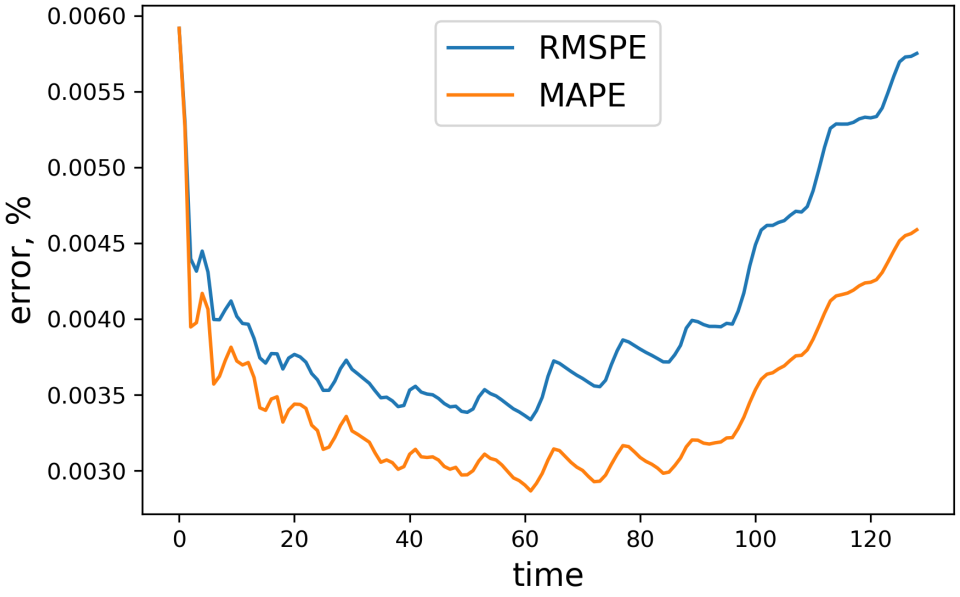
ADF test confirms the non-stationarity hypothesis with a low degree of confidence. Let us find first difference of the time series to delete trend from the data:

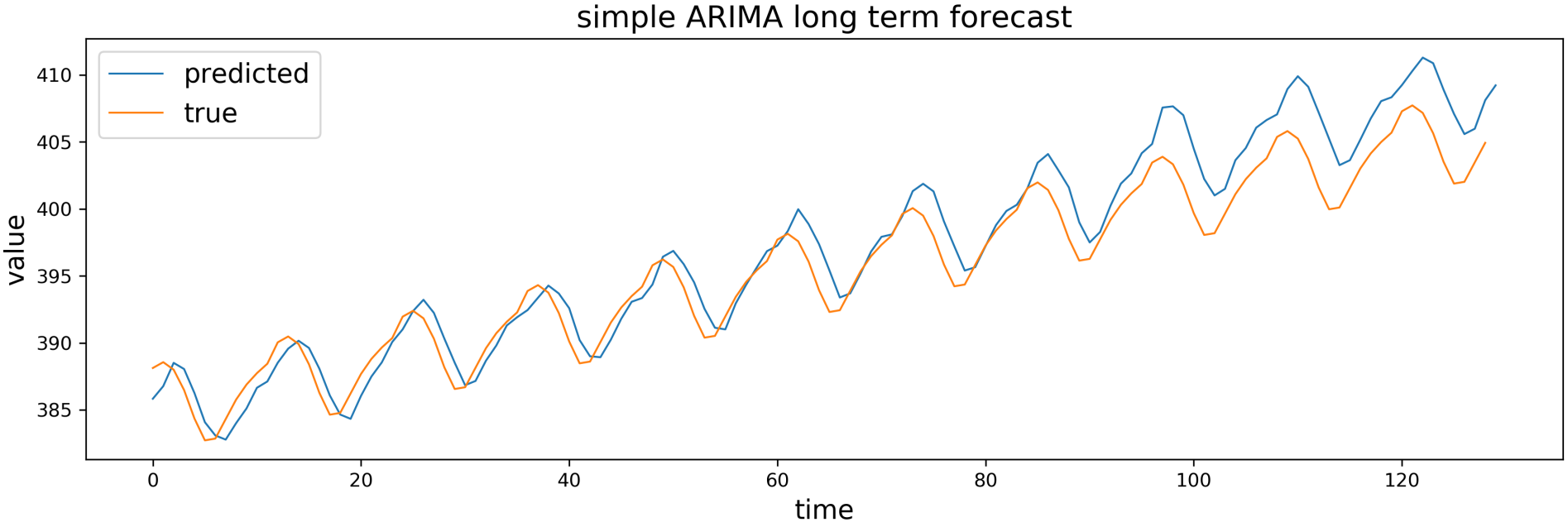
In [18]:
```
train['CO2 [ppm]_sdiff_diff'] = train['CO2 [ppm]_sdiff'] - train['CO2 [ppm]_sdiff'].shift(1)
print("Dickey-Fuller criterion: p = {0}, used lag = {1}".format(round(sm.tsa.stattools.adfuller(train['CO2 [ppm]_sdiff_diff'
                                                                ,sm.tsa.stattools.adfuller(train['CO2 [ppm]_sdiff_diff'][13:])
```

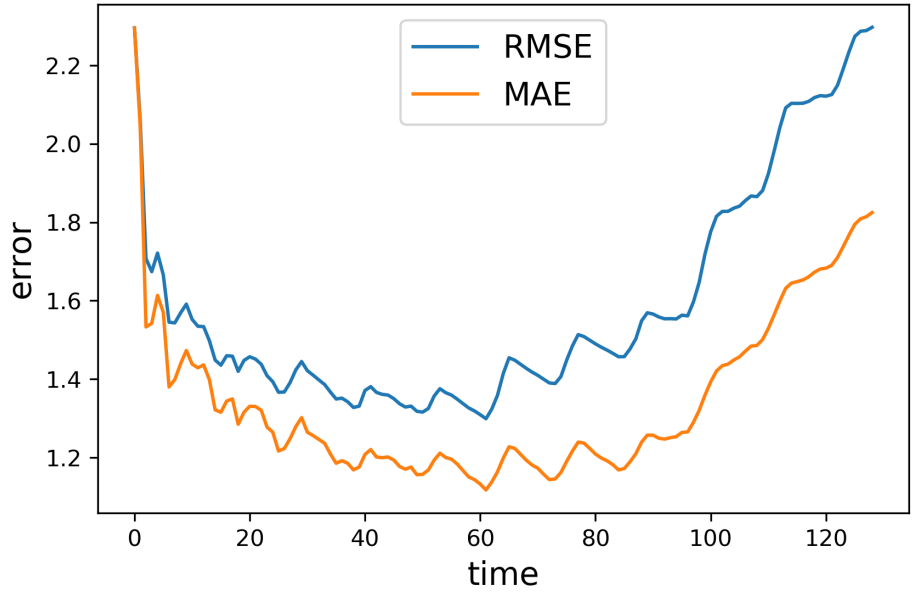
```
Dickey-Fuller criterion: p = 0.0, used lag = 12
```

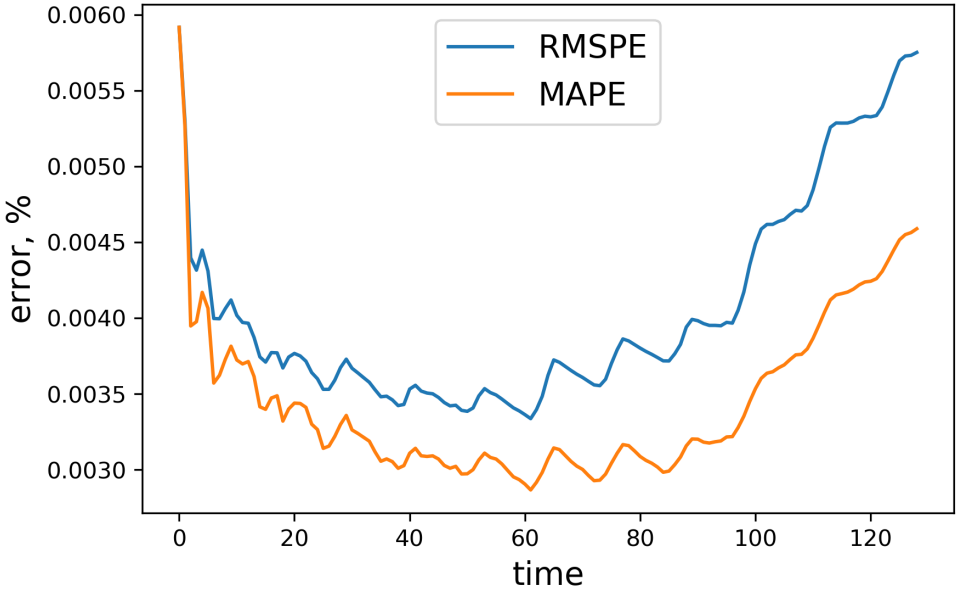

simple ARIMA long term forecast
predicted
true
value
time

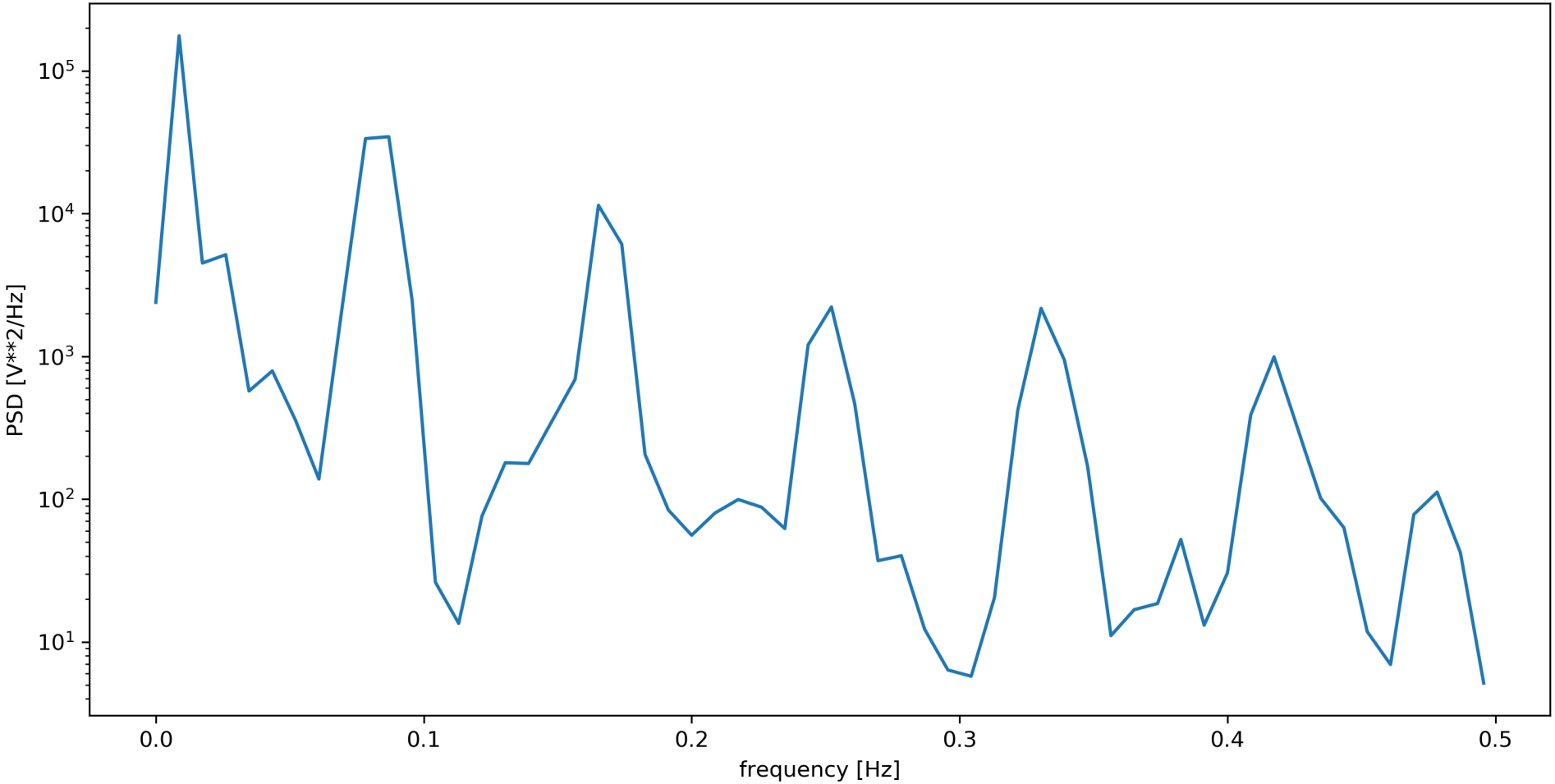

RMSE
MAE
error
time

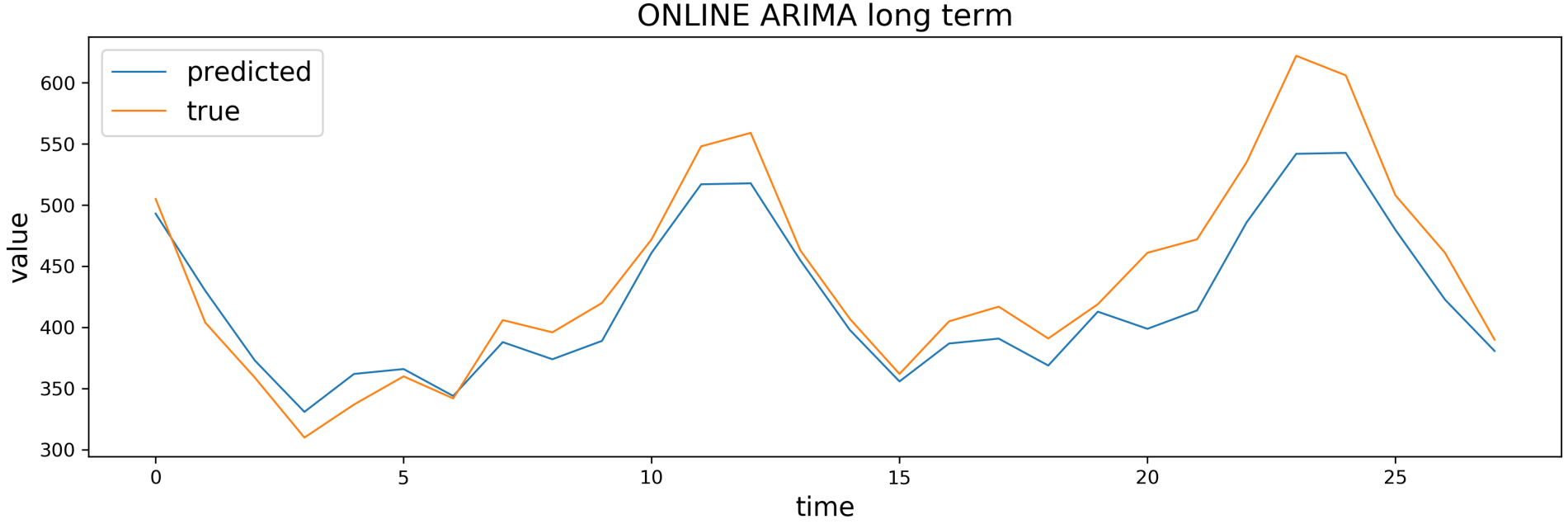

RMSPE
MAPE
error, %
time

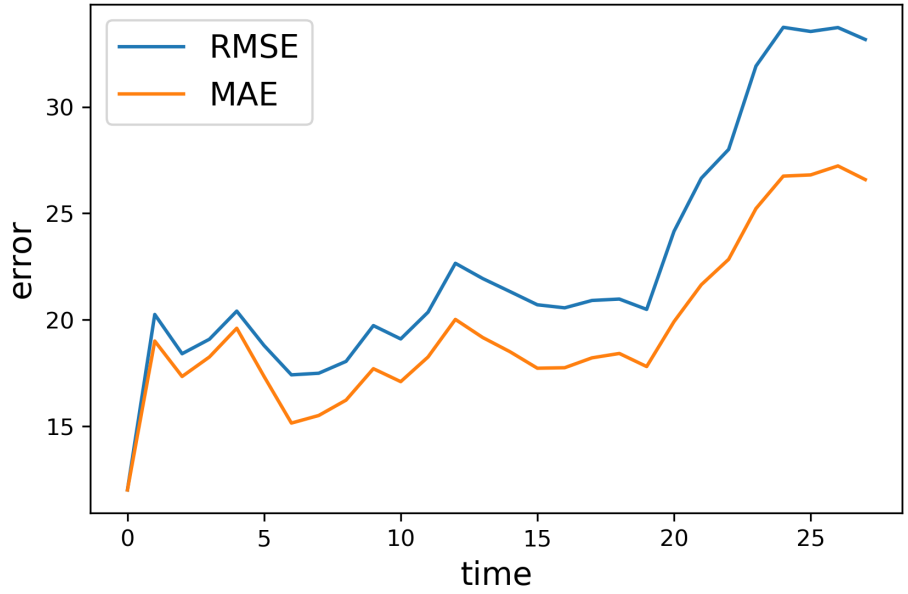

simple ARIMA long term forecast
predicted
true
value
time

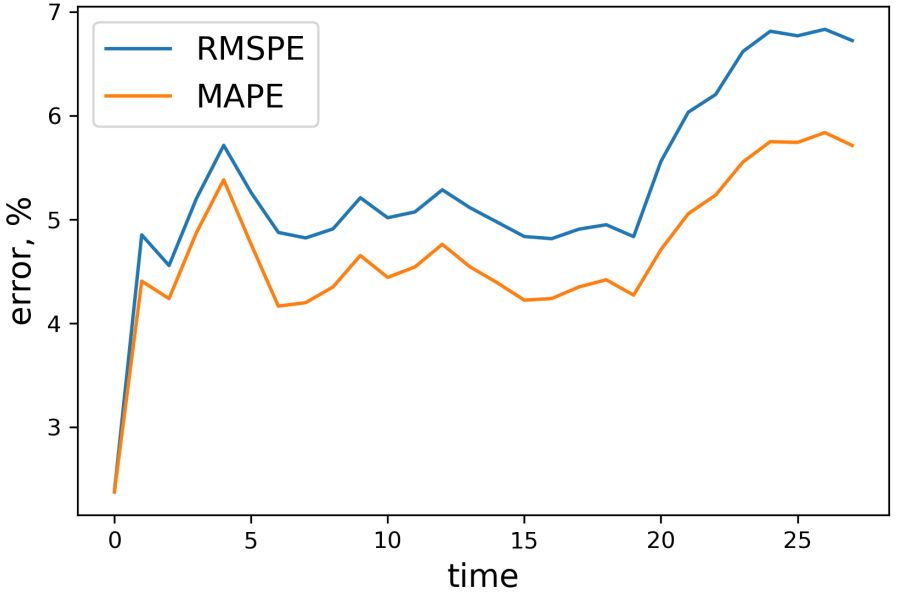

RMSE
MAE
error
time


RMSPE
MAPE
error, %
time


PSD [V**2/Hz]
10^5
10^4
10^3
10^2
10^1
0.0
0.1
0.2
0.3
0.4
0.5
frequency [Hz]


ONLINE ARIMA long term
predicted
true
value
time


RMSE
MAE
error
time


RMSPE
MAPE
error, %
time