The main goal of this process is to acquire technical knowledge for handling time series data, building upon the work of Serpelloni et al. (2018), which addressed GPS time series. This report details the processes and methodologies employed to refine time series data, essential for the analysis and detection of displacements in a karst region, as investigated in the mentioned study. To facilitate model application and pattern detection in the north, east, and vertical dimensions, closely located GPS stations were chosen. The selected stations were BRSE, TRIE, PORD, KOPE, and MPRA, all available on the Nevada Geodetic Laboratory website.

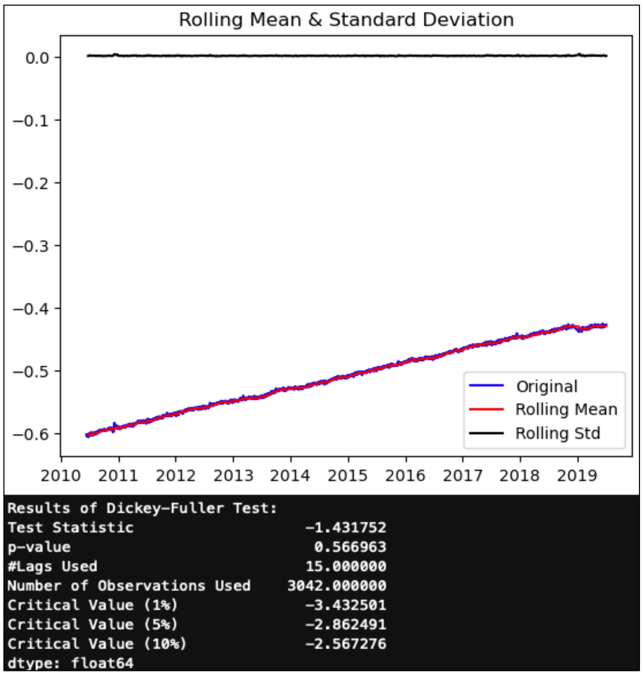
1) Pre-processing

Stationarity Assessment

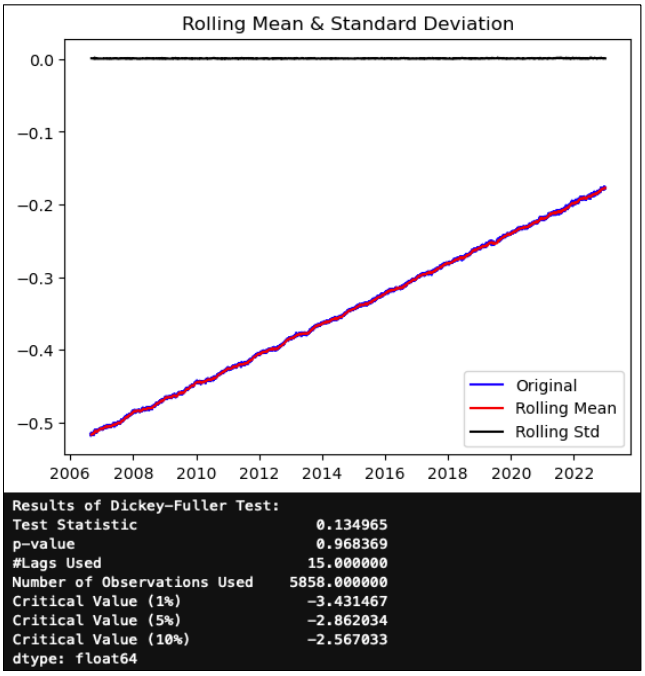
Initially, it was used the Augmented Dickey-Fuller test from the 'statsmodels.tsa.stattools' library to statistically assess the stationarity of the time series data. The main objective is to determine if there is sufficient statistical evidence to conclude whether the time series is stationary or not. Stationarity is crucial for applying statistical procedures, simplifying the modeling process, and facilitating the identification of patterns.

```
Results of Dickey-Fuller Test:
Test Statistic                  -1.431752
p-value                          0.566963
#Lags Used                      15.000000
Number of Observations Used   3042.000000
Critical Value (1%)             -3.432501
Critical Value (5%)             -2.862491
Critical Value (10%)            -2.567276
dtype: float64
```

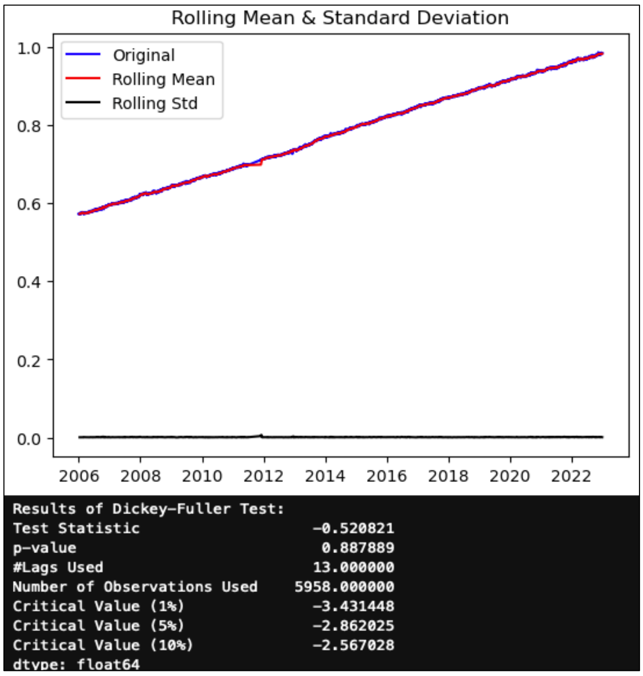
The test returns the "p-value" parameter, used to accept or reject the null hypothesis of stationarity, and the "critical values" (1%, 5%, and 10%) serve as references to compare with the value of the statistical test. Upon confirming that the series is not stationary, rejecting the null hypothesis, a logarithmic transformation was chosen to be applied to the measurements in an attempt to stabilize the variance. The choice of "log1p" (tsLog = np.log1p(instance)) proved to be more appropriate as it handles negative values in the series readings adequately.



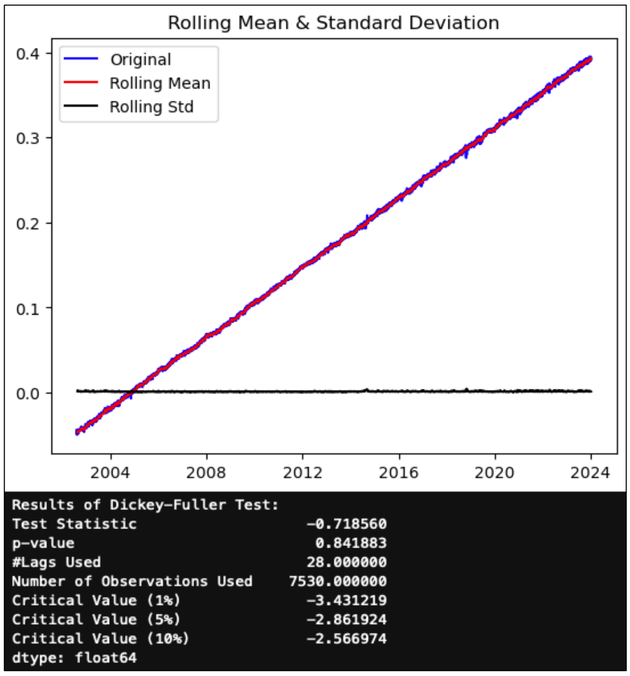
BRSE (east)



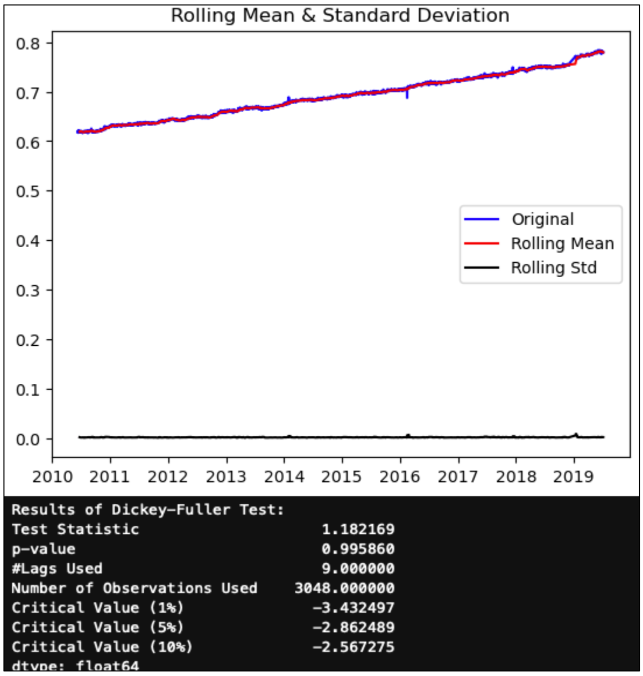
PORD (east)



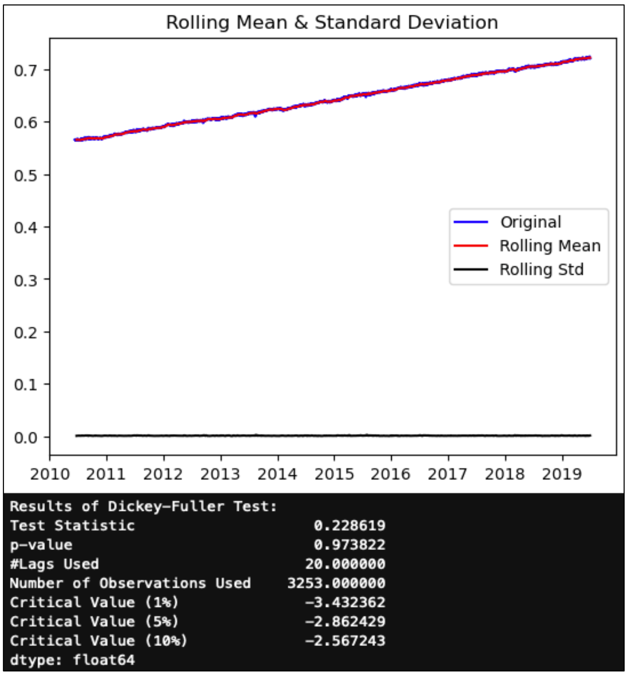
KOPE (east)



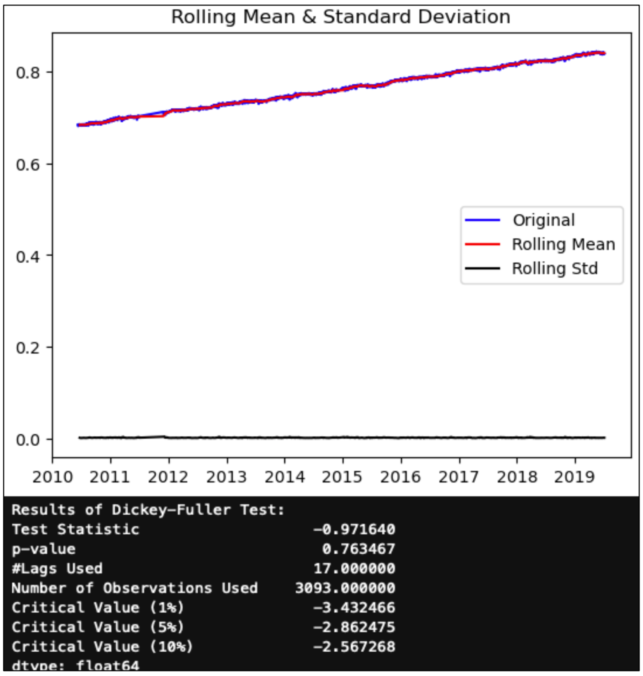
MPRA (east)



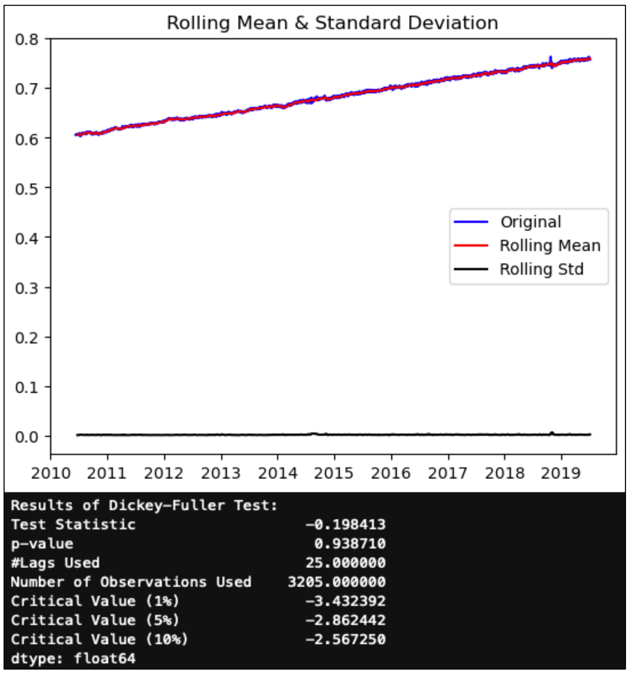
BRSE (north)



PORD (north)



KOPE (north)



MPRA (north)

Differencing for Stationarity

Subsequently, the differencing technique was applied to make the series stationary by subtracting each value from the previous one:

```
tsLog = tsLog - tsLog.shift()
```

Another possible approach would be to use the moving average by subtracting it from the series. This technique is useful for removing seasonality or short-term patterns. The expression used for this approach would be:

```
tsLog = tsLog.rolling(window=7).mean()
```

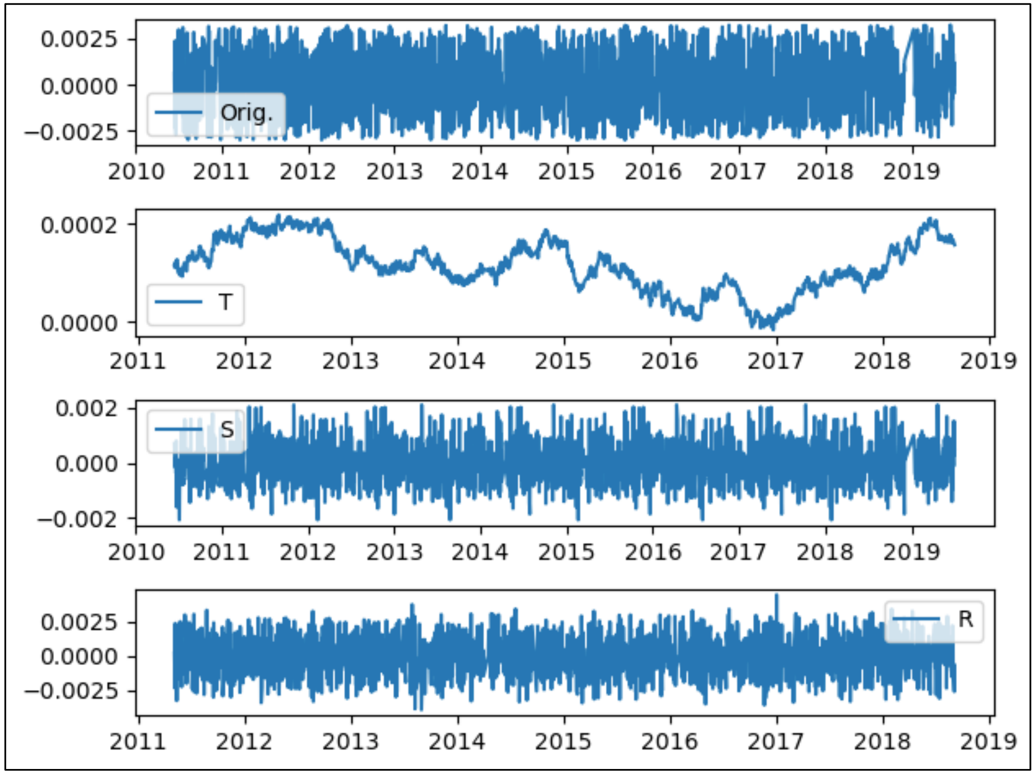
*The "window" parameter refers to the data frequency.

Handling Non-Null Values (NaN) after Differencing

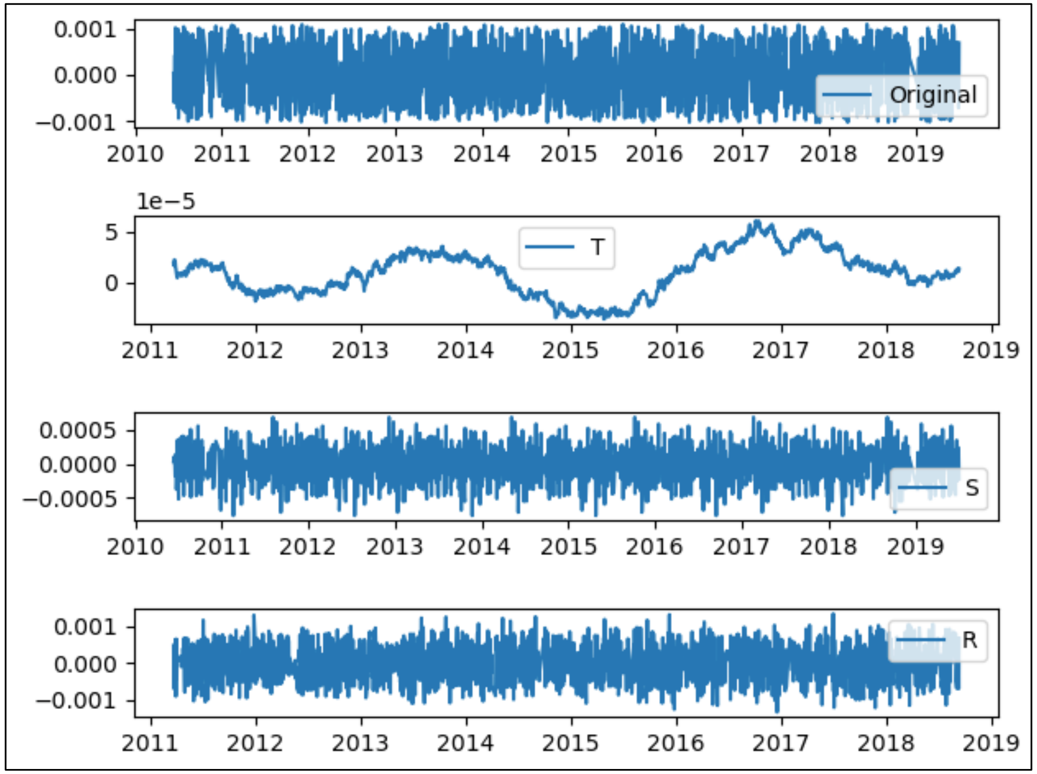
When performing differencing, it is possible to encounter non-null values (NaN). Therefore, it is crucial to properly handle these data, either by removing or replacing them with the mean. Considering that these are the final steps before processing, it would also be advisable to remove outliers to ensure that the modeling is more precise and appropriately fitted to the data. These processes aim to prepare time series for the effective application of analytical techniques, contributing to a more robust modeling and facilitating the detection of relevant patterns.

2) Modelling

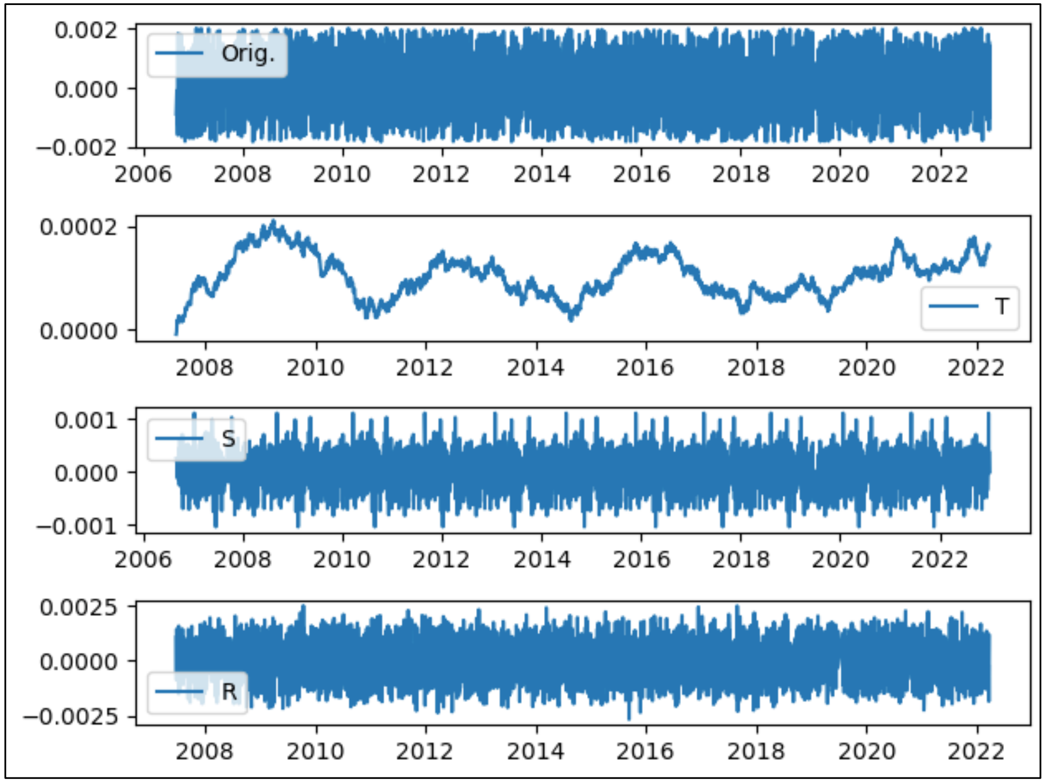
Before starting the modeling process, the individual decomposition of trend (T), seasonality (S), and residuals (R) present in the series was performed. Conducting this procedure is important because the trend helps understand the movement of data, seasonality aids in understanding cycles, while residuals help capture unidentified patterns and serve as a parameter to determine whether the model fits the data or not. Therefore, an initial decomposition was conducted and used as a reference for future analyses.



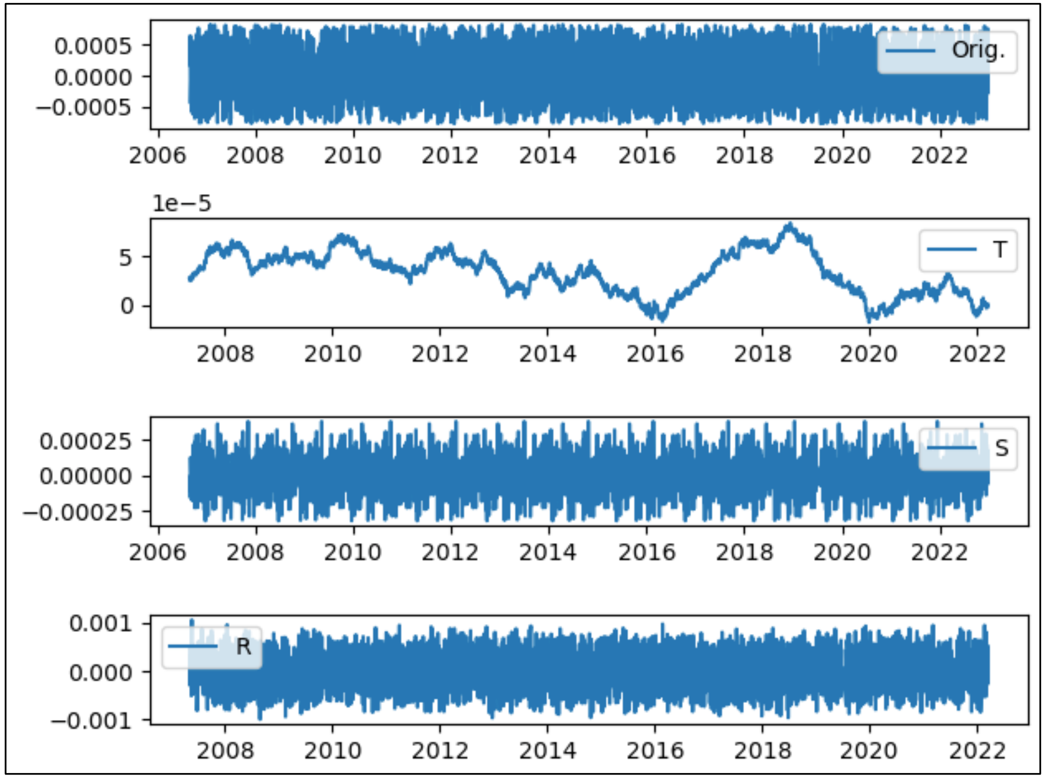
BRSE (east)



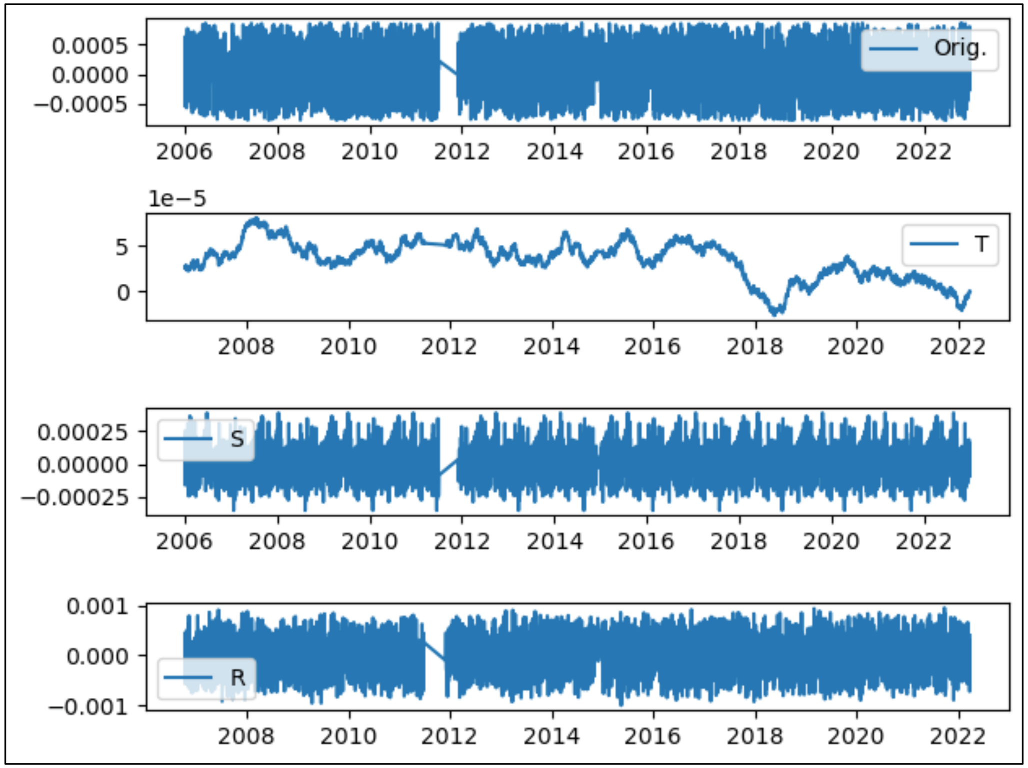
BRSE (north)



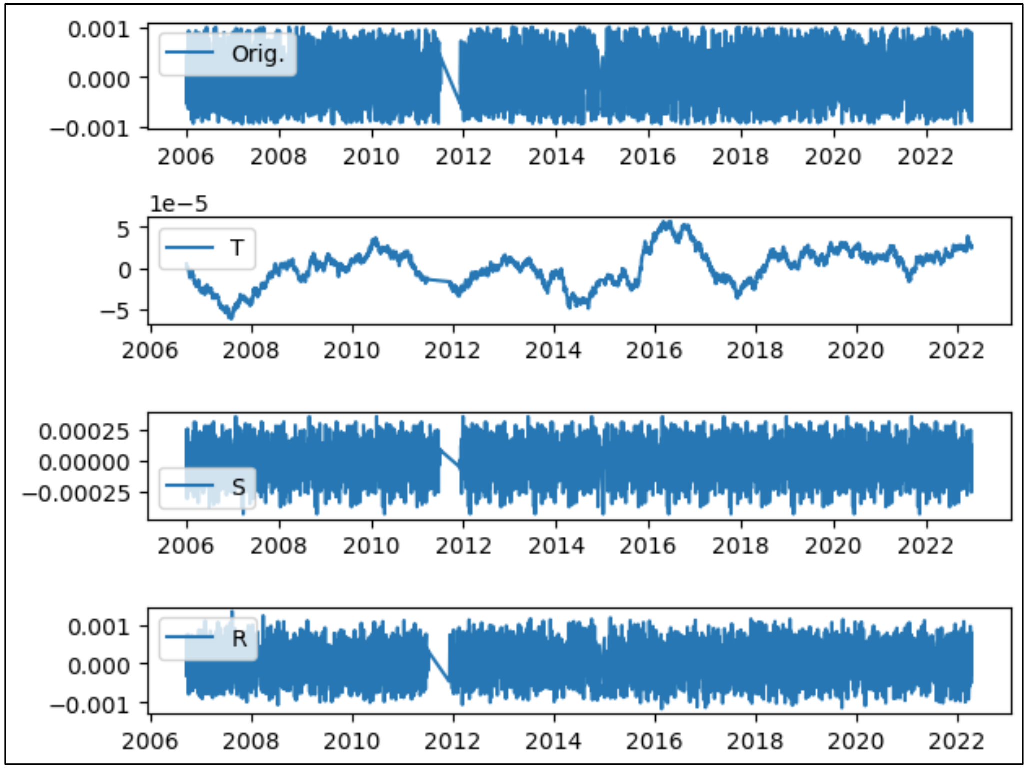
PORD (east)



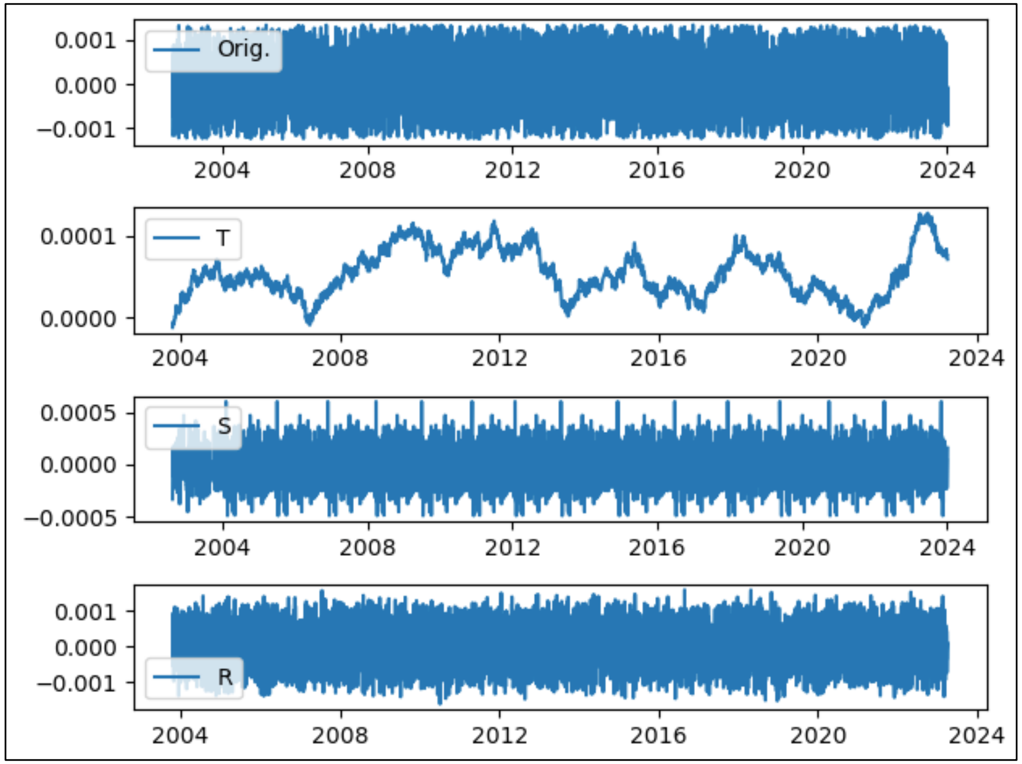
PORD (north)



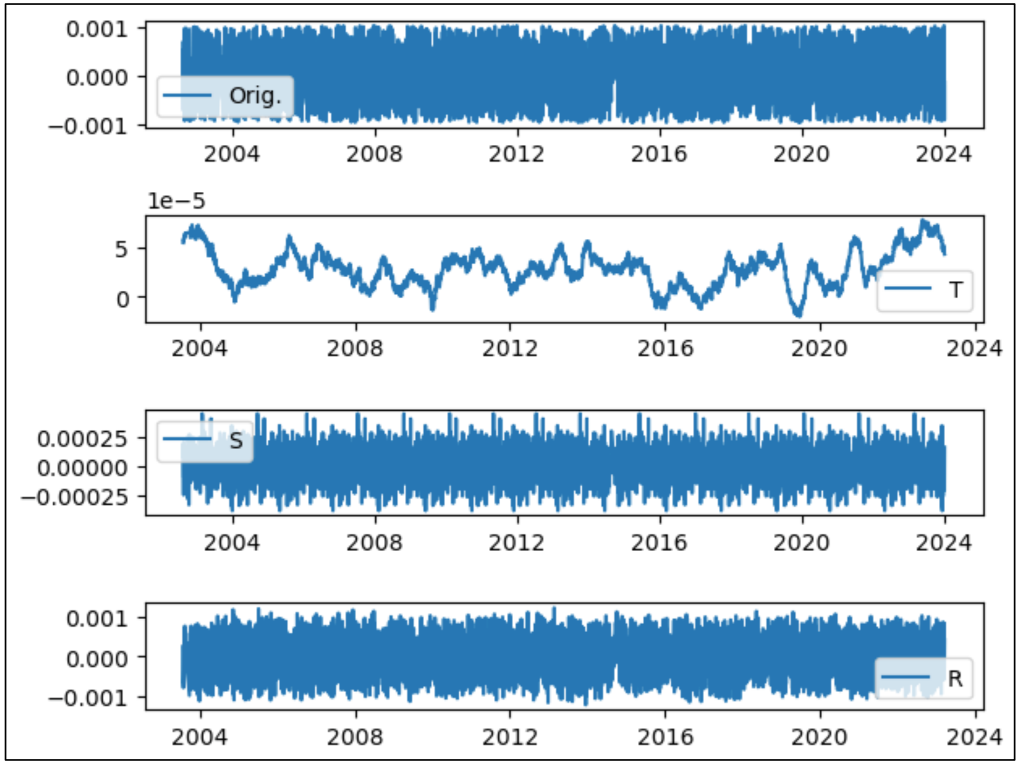
KOPE (east)



KOPE (north)



MPRA (east)



MPRA (north)

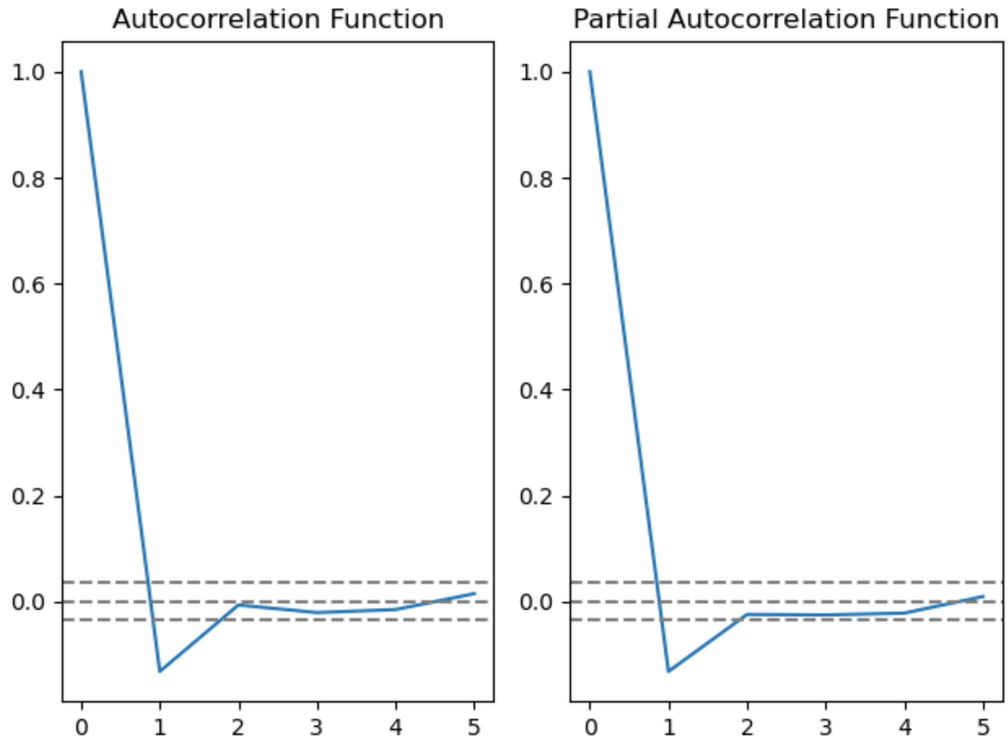
While exploring modeling concepts for time series, the model known as Autoregressive Integrated Moving Average (ARIMA) emerged, being often recommended for analysis and forecasting in time series. ARIMA is based on identifying autocorrelations present in data at different time instances, using fundamental parameters: p, d, and q for model construction.

- **Autoregressive Component (AR):**
  The parameter *p* in ARIMA is associated with the autoregressive component (AR). This component reflects the correlation between current observations and observations in previous time periods. In other words, it models the relationship or dependence between different moments in time.

- **Integrated Component (I):**
  The parameter *d* refers to the integrated component (I). It represents the number of non-seasonal differences necessary to make the time series stationary. Stationarity is crucial to ensuring constant statistical properties over time.

- **Moving Average Component (MA):**
  The *q* parameter is associated with the Moving Average (MA) component. This component models the correlation between current observations and the errors (residuals) of observations in previous time periods (lags).

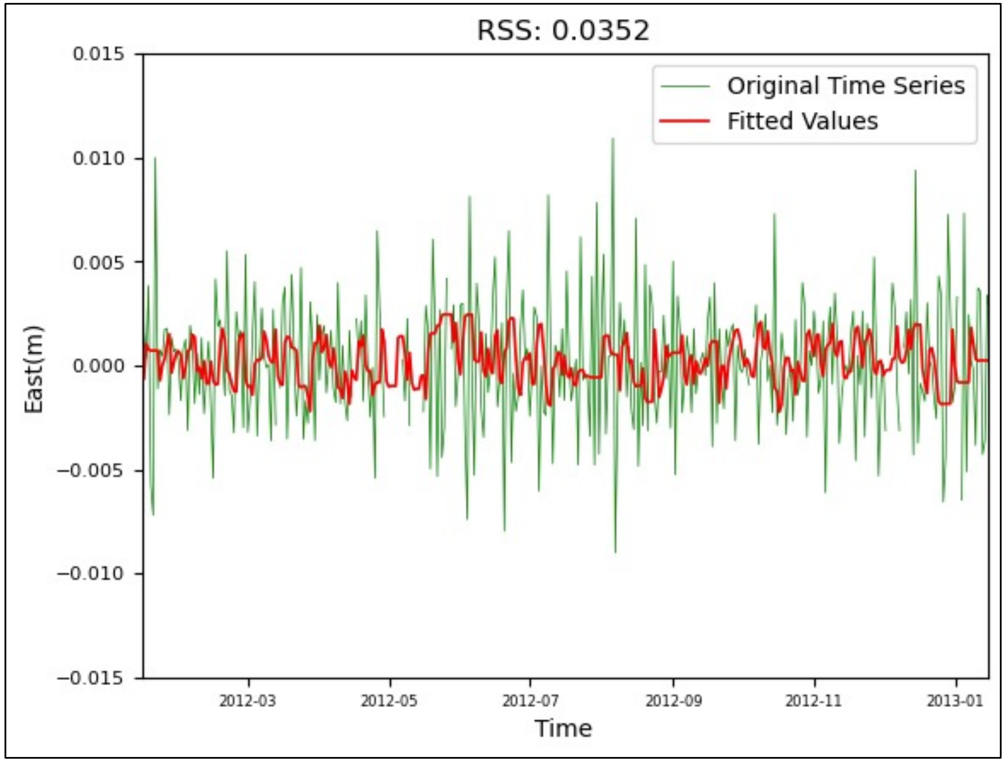
It is important to note that the ARIMA modeling process involves the analysis of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to identify the appropriate values of *p*, *d*, and *q*. These values are determined after the individual decomposition of the trend, seasonality, and residuals mentioned earlier.



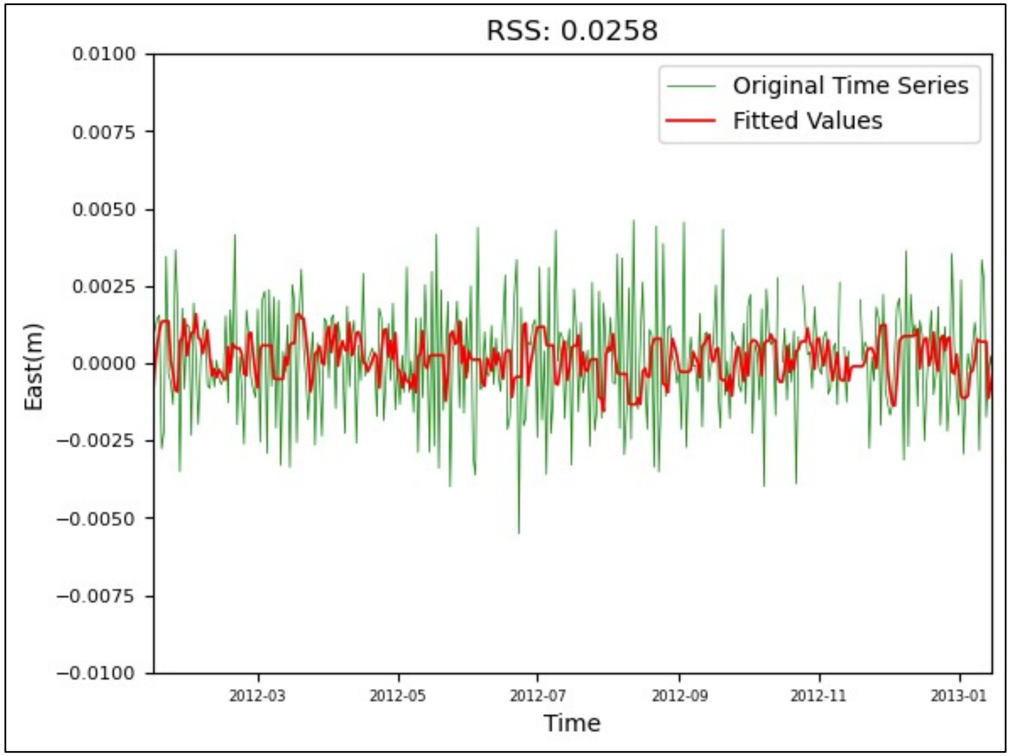
ACF and PACF plots example

This approach provides a robust framework for handling a variety of temporal patterns present in time series. After producing the ACF and PACF plots, the model was estimated and the graphs were obtained for the analysis of the time series. This procedure was applied to all series mentioned at the beginning of the report (BRSE, TRIE, PORD, KOPE, and MPRA) for analysis in the east, north, and vertical directions.

3) ARIMA results



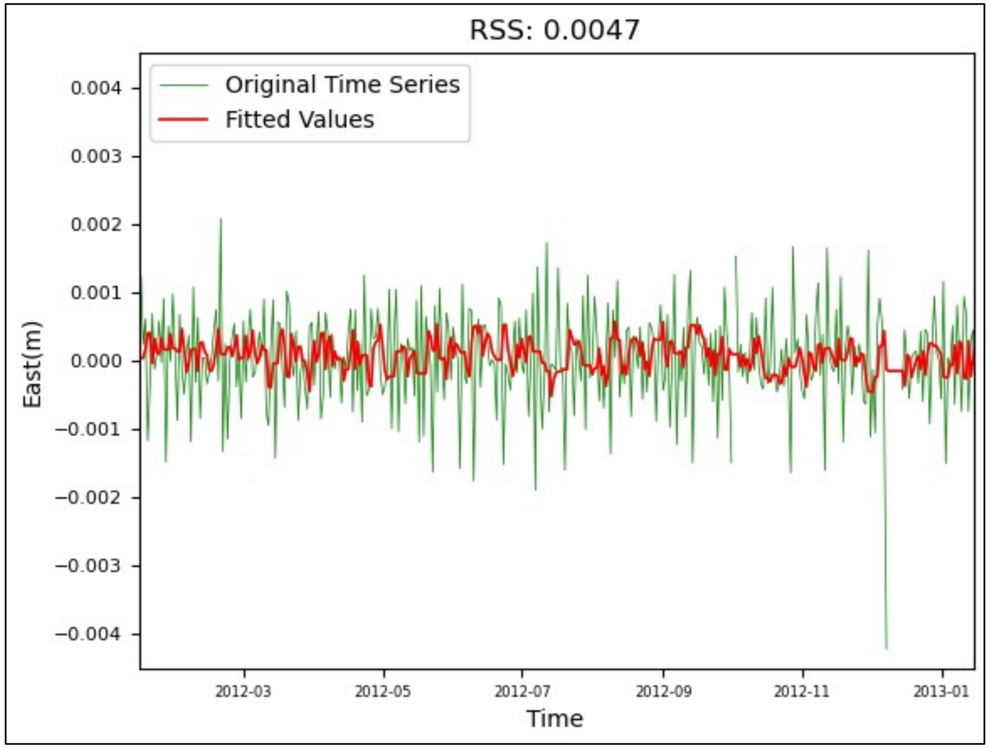
RSS: 0.0352
Original Time Series
Fitted Values
East(m)
Time
0.015
0.010
0.005
0.000
−0.005
−0.010
−0.015
2012-03
2012-05
2012-07
2012-09
2012-11
2013-01



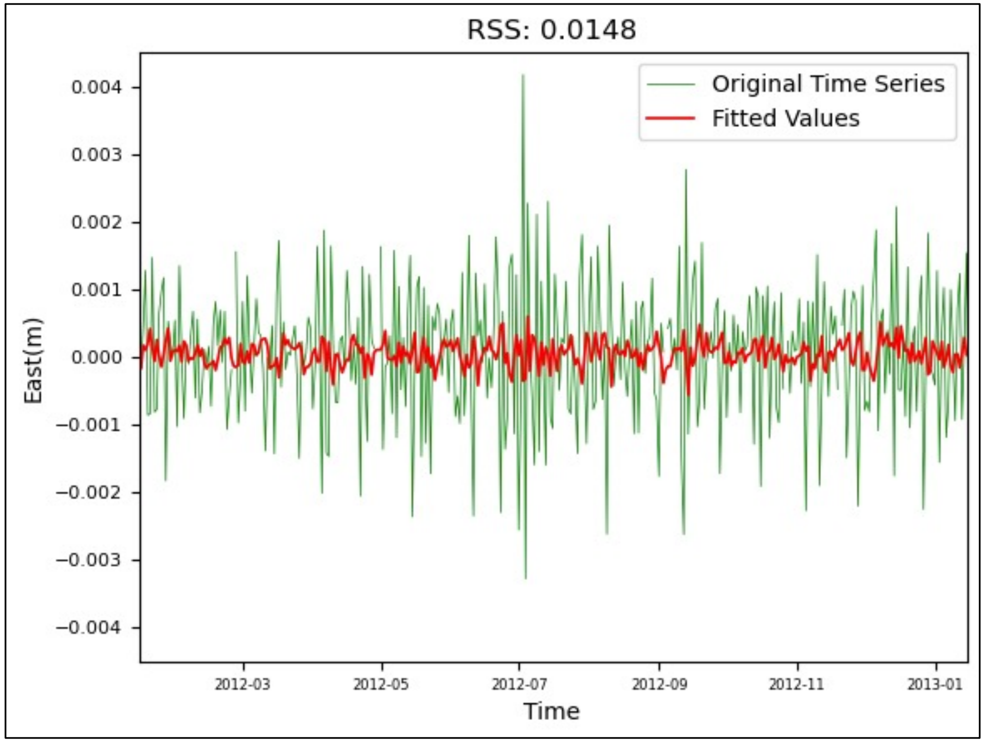
RSS: 0.0258
Original Time Series
Fitted Values
East(m)
Time
0.0100
0.0075
0.0050
0.0025
0.0000
−0.0025
−0.0050
−0.0075
−0.0100
2012-03
2012-05
2012-07
2012-09
2012-11
2013-01

BRSE

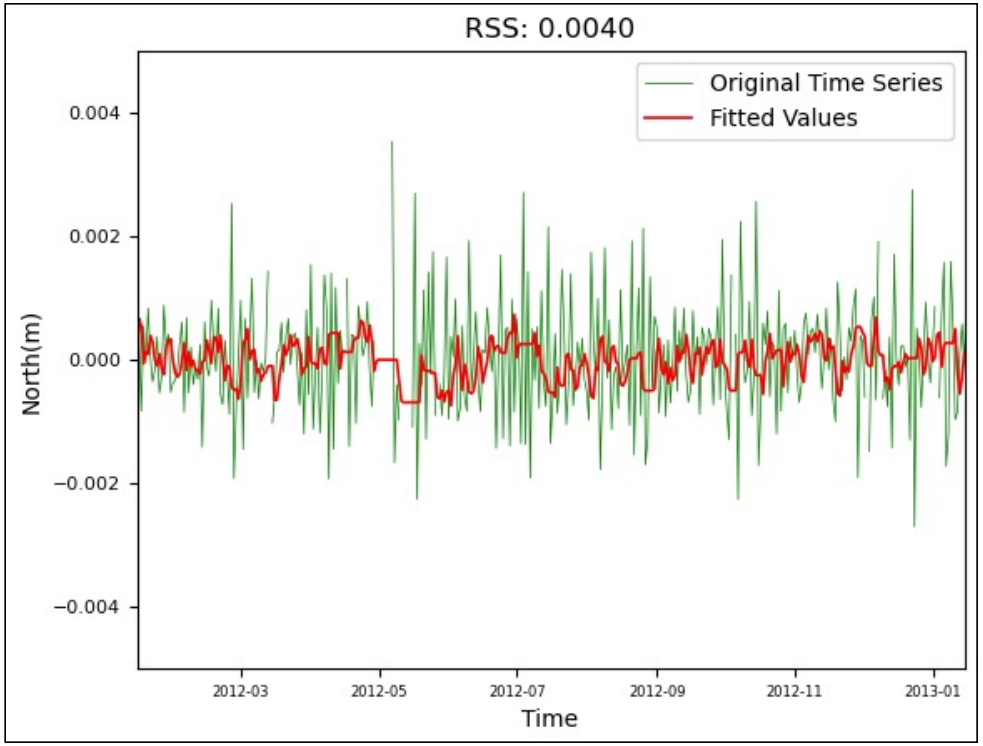
PORD


RSS: 0.0047
Original Time Series
Fitted Values
East(m)
Time
0.004
0.003
0.002
0.001
0.000
−0.001
−0.002
−0.003
−0.004
2012-03
2012-05
2012-07
2012-09
2012-11
2013-01

KOPE

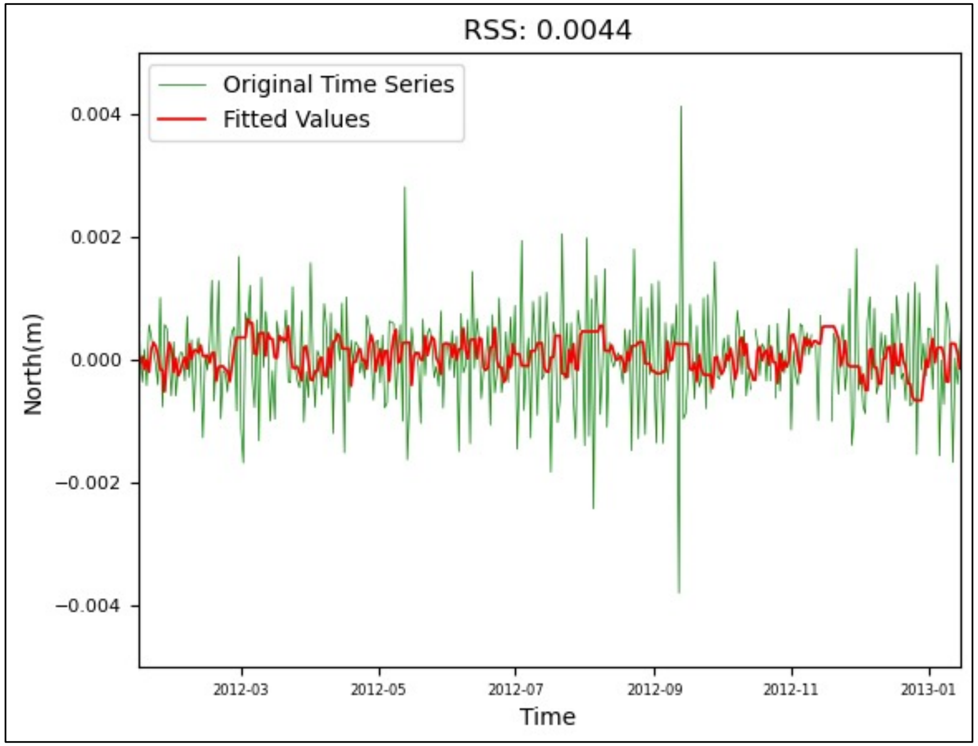

RSS: 0.0148
Original Time Series
Fitted Values
East(m)
Time
0.004
0.003
0.002
0.001
0.000
−0.001
−0.002
−0.003
−0.004
2012-03
2012-05
2012-07
2012-09
2012-11
2013-01

MPRA



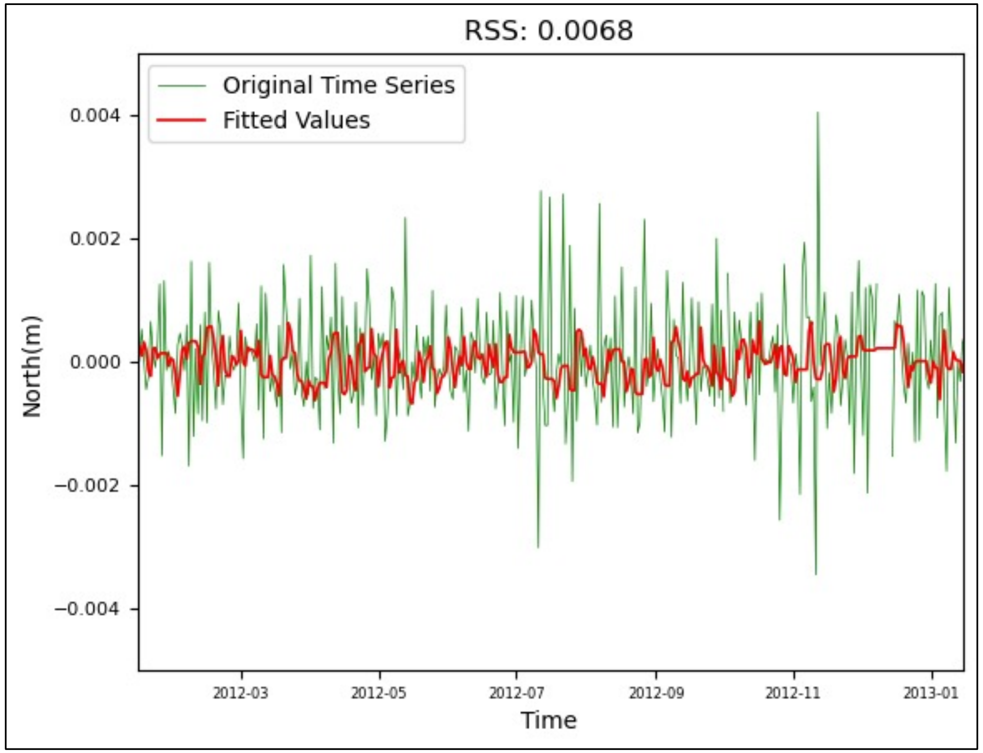
RSS: 0.0040
Original Time Series
Fitted Values
North(m)
0.004
0.002
0.000
−0.002
−0.004
2012-03
2012-05
2012-07
2012-09
2012-11
2013-01
Time

BRSE

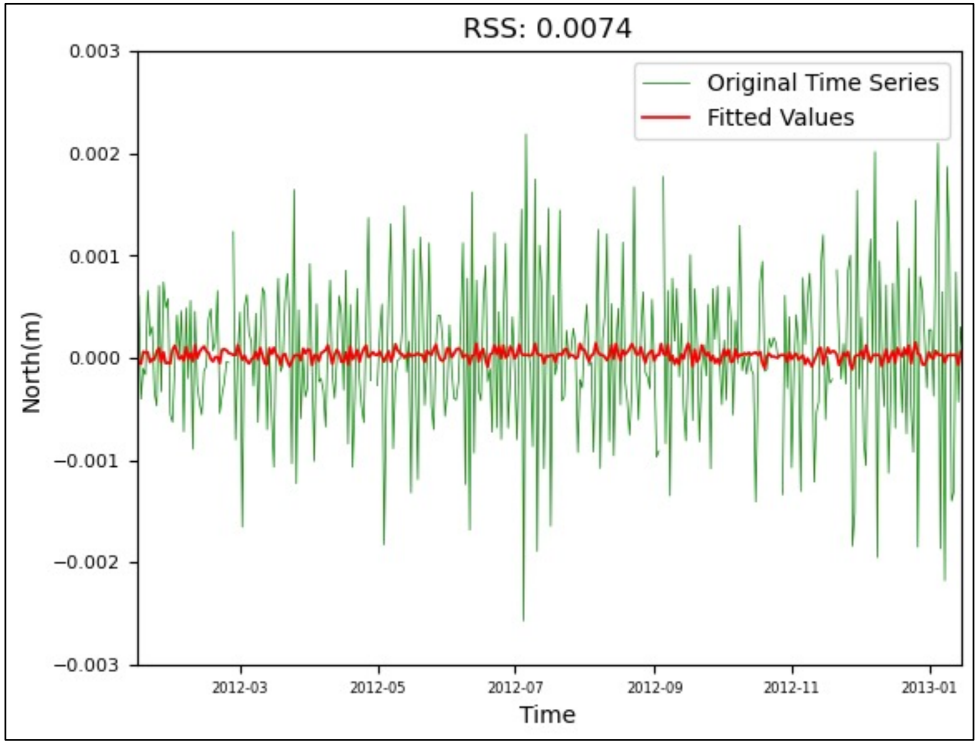


RSS: 0.0044
Original Time Series
Fitted Values
North(m)
0.004
0.002
0.000
−0.002
−0.004
2012-03
2012-05
2012-07
2012-09
2012-11
2013-01
Time

PORD


RSS: 0.0068
Original Time Series
Fitted Values
0.004
0.002
0.000
−0.002
−0.004
North(m)
2012-03
2012-05
2012-07
2012-09
2012-11
2013-01
Time

KOPE


RSS: 0.0074
Original Time Series
Fitted Values
0.003
0.002
0.001
0.000
−0.001
−0.002
−0.003
North(m)
2012-03
2012-05
2012-07
2012-09
2012-11
2013-01
Time

MPRA