Exponential Smoothing Partial Analysis

This data packages involved are global average surface temperature anomaly, atmospheric carbon dioxide level, UK outbound tourism number in the GMAF international passenger survey, and UK inland energy consumption. The following is a detailed analysis based on the model results:

1. Data preparation and preprocessing

Before performing exponential smoothing modeling on the data, a data check was performed and all data sets had no missing values from the first period to the last period.

2. Model and parameters selection

On each data set, we calculated different smoothing parameters, including smoothing level, trend, and seasonal component, and used these parameters to fit the model. The following are the results of each data:

CH4: The smoothing level is 0.999999985, the smoothing trend is 0.03099, and the seasonal component is very small (close to 0). The prediction results of the model show that the MSE is 1.3155, the RMSE is 1.1470, the MAE is 0.8859, and the MAPE is 0.05%, indicating that the model is very accurate in predicting CH4 data.

ET12: The smoothing level is 0.1990, the smoothing trend is 0.0, and the seasonal component is 0.1605. The prediction effect of the model is good, with an MSE of 0.6538, a RMSE of 0.8086, a MAE of 0.6171, and a MAPE of 3.58%, showing that the model can capture the long-term trend of energy consumption well, but there are still some errors in some seasonal fluctuations.

GMAF: The smoothing level is 0.7829, the smoothing trend is close to 0, and the seasonal component is 0.2171. Although the model fits well, due to the large fluctuations in the data, the MSE is 220088.2971, the RMSE is 469.1357, the MAE is 276.2873, and the MAPE is 17.07%, indicating that the prediction accuracy is low, especially in capturing seasonal fluctuations. There are obvious errors.

MST: The smoothing level is 0.4259, the trend component is close to 0, and the seasonal component is 0.0515. The prediction error of the MSTA model is small, with an MSE of 0.0138, a RMSE of 0.1174, a MAE of 0.0903, and a MAPE of 118.64%, but the high MAPE value shows that the model has a large error in capturing short-term fluctuations.

3. Model Evaluation

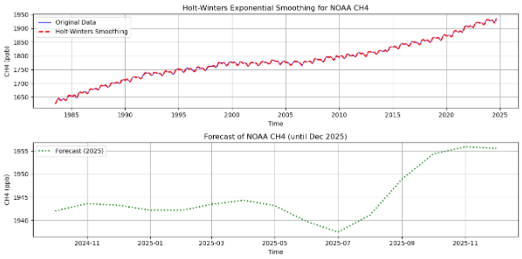
For different data sets, exponential smoothing performs differently in dealing with trends and seasonal changes. The CH4 and ET12 data are more accurate, especially CH4, which has a very low MAPE, indicating that exponential smoothing performs very well on this data. However, GMAF and MSTA have poorer predictions, especially the high MAPE value of MSTA, indicating that exponential smoothing is insufficient in capturing seasonal fluctuations and short-term changes.

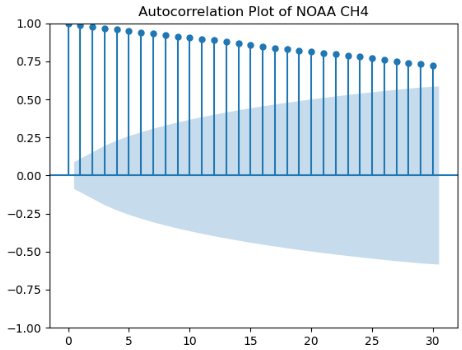
4. Image output result analysis

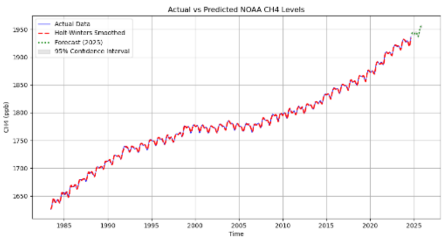
According to the forecast graphs for each data, we can observe that the exponential smoothing method tracks the long-term trend well, especially in the CH4 and ET12 data. However, when dealing with GMAF and MSTA, although the long-term trend can be captured, the fluctuations of seasonal changes cannot be accurately predicted.

5.Conclusion

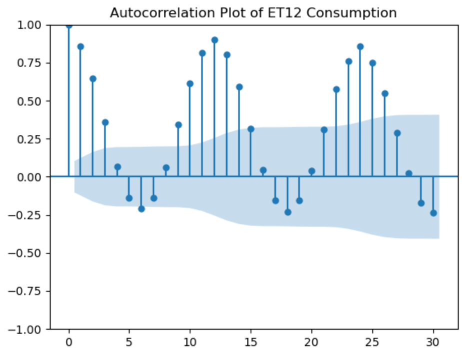
Exponential smoothing has certain advantages in dealing with long-term trends and seasonal fluctuations, especially for relatively stable data such as CH4 and ET12. However, for data with large fluctuations (such as GMAF) and data with large short-term fluctuations (such as MSTA), the forecast error of this method is large.

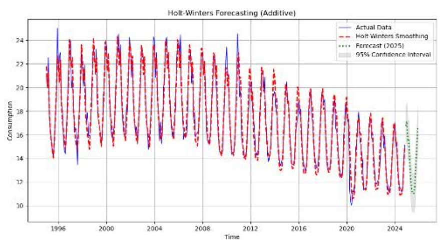




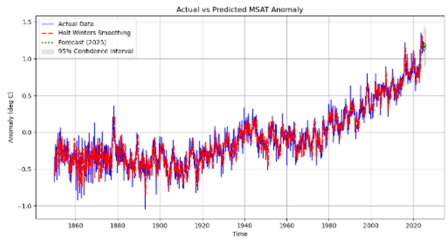


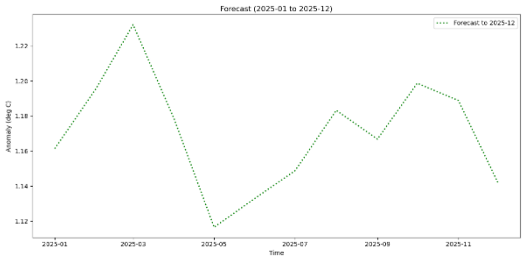
CH4 Exponential Smoothing Outcome

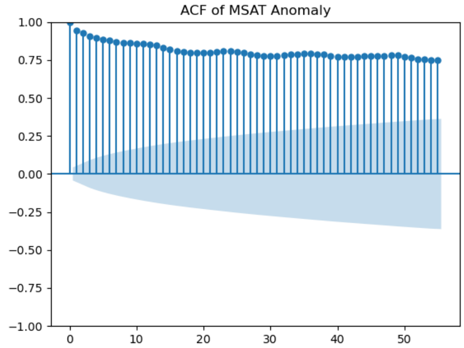




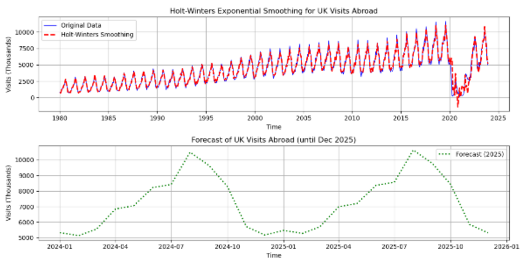
ET12_Exponential Smoothing Outcome

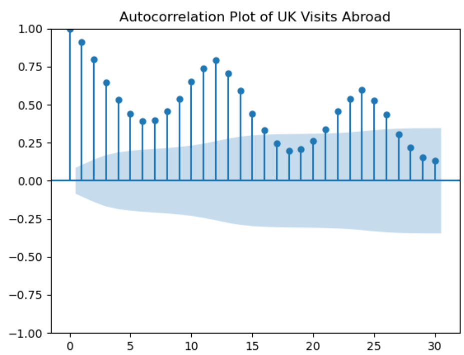






MAST Exponential Smoothing Outcome





GMAF Exponential Smoothing Outcome

2. ARIMA

(a) Data preparation and its impact:

Before applying the ARIMA model to the MSTA (global mean surface temperature anomaly) data, we use the ADF test to check the stationarity of the data. The test result shows that the p-value is 0.9519 (the series is non-stationary). To solve this problem, we perform a difference process and obtain a stationary series with an ADF test p-value of 6.41659043505176e-23, which confirms that the data after the first difference is stationary and can be used for the next step of ARIMA modeling.

(b) Preliminary analysis and conclusions:

By visually inspecting the time series (including autocorrelation and partial autocorrelation plots), we can see the presence of trend and cyclical components. The first-order difference processing removes the trend, while the autocorrelation and partial autocorrelation plots show a pattern suitable for ARIMA modeling (and consistent with the minimum value of the Akaike Information Criterion results). These analyses provide a basis for the preprocessing of the ARIMA model.

(c) ARIMA model selection, testing and evaluation:

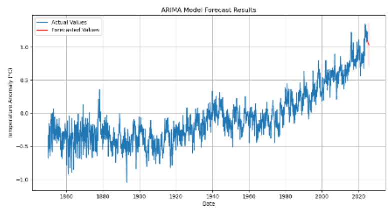
We selected the ARIMA model by gradually searching for the minimum value of the Akaike Information Criterion (AIC). The best model finally selected was ARIMA(1, 1, 2), whose parameters are as follows:

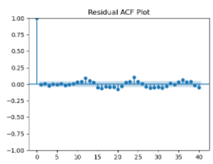
- AR(1) coefficient: 0.7803
- MA(1) coefficient: -1.2406
- MA(2) coefficient: 0.2667
- Intercept: 0.0001

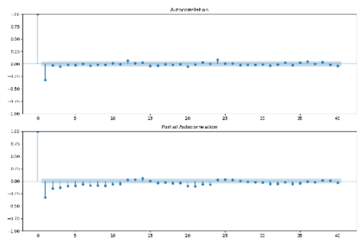
The AIC value of this model is -3010.802, which shows a better fit than other ARIMA models. The Ljung-Box test results show that there is no significant autocorrelation in the residuals (p value = 0.91), indicating that the model effectively captures the temporal structure of the data.

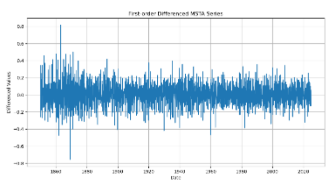
(d) Comparison with exponential smoothing method

The exponential smoothing method analysis shows that the MAPE of the MAST data set is 118.64%, which has a large prediction error, while the MSE of ARIMA is 0.0139, RMSE: 0.1179, MAE: 0.0903, and MAPE: 114.68%. The results show that ARIMA (1, 1, 2) can capture the long-term trend of the data more effectively. The residual diagnostics of the ARIMA model (the p-value of the Ljung-Box test is 0.91, and the p-value of the Jarque-Bera test is 0.00) confirm that the model fits well, further supporting its effectiveness compared to the exponential smoothing method.









Forecast description: The prediction results of the ARIMA model show a clear upward trend in 2025. This prediction is consistent with the long-term warming trend of the MSTA series.

Regression prediction analysis

Data Preparation and Analysis

In the regression model, we used four key time series data: MSTA (global mean surface temperature anomaly), CH4 (atmospheric carbon dioxide level), GMAF (number of UK outbound tourists from the International Passenger Survey), and ET12 (UK inland energy consumption). The data preparation steps include:

•Handling missing values, especially missing data in the CH4, GMAF, and ET12 columns.

•Calculating the variance inflation factor (VIF) of each variable to check if there is a multicollinearity problem. The const term has a larger VIF value of 2175.39, indicating that there may be multicollinearity.

Regression results analysis

This section selected a multiple linear regression model (OLS) to fit the relationship between MSTA (temperature anomaly) and the other three indicators (CH4, GMAF, ET12). The regression results show:

R-squared=0.678， which means that the regression model can explain 67.8% of MSTA fluctuations.

P(F-statistic)=2.98e-87， Indicates that the model is statistically significant.

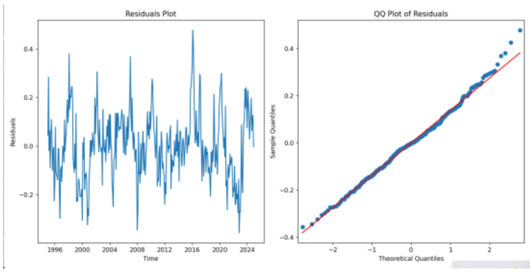
p-value： The p-values for CH4 and GMAF were very small (both < 0.05), indicating that they had significant effects on MSTA, while the p-value for ET12 was 0.103, indicating that its effect was not significant.

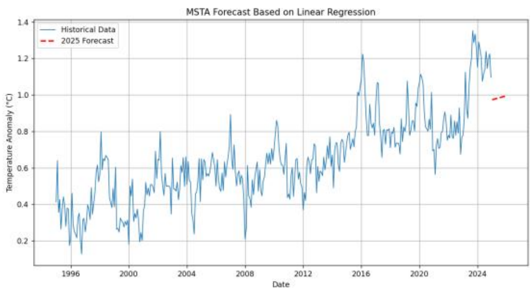
Prediction effect evaluation:

The prediction results of the model show a good fit, and the temperature anomaly gradually increases, which is consistent with the trend of global temperature rise. For December 2025, the predicted value is 0.9936°C. The calculated evaluation indicators are as follows: MSE: 0.0189, RMSE: 0.1375, MAE: 0.1062, MAPE: 20.64%. These indicators show that although the regression model can fit the data well, its prediction errors still exist, especially the high MAPE, indicating that the model has large prediction errors in some months.

Output Graphics

The residual sequence diagram and QQ diagram show the distribution of the residuals of the regression model. The overall distribution is close to normal distribution. The forecast diagram of MSTA in 2025 shows the model's forecast trend of future temperature anomalies.





Conclusion

Based on the results of the multivariate linear regression model, CH4 and GMAF have significant effects on MSTA, while ET12 has a weaker effect. Although the model is able to predict the long-term trend of temperature anomalies well, the MAPE is high.

Appendix

Excel Preprocessing

CH4： No preprocessing in excel.

ET12： From ET_1.2_JAN_25.csv -sheet: Month select column Month and Unadjusted total.

GMAF: Select column:Date, GMAF from previous sheet.

MSAT: Select Time ,Anomaly (deg C) from previous sheet.

Python Code Explanation

ExponentialSmoothing_CH4_35621168

The forecast model of CH4 data was constructed using additive seasonal exponential smoothing to capture the seasonality and trend of the data.

ExponentialSmoothing_ET12_35621168

Additive seasonal exponential smoothing is used to fit the ET12 data to its trend and seasonal variations.

ExponentialSmoothing_GMAF_35621168

Additive seasonal exponential smoothing is used to fit the GMAF data to capture its seasonality and trend.

ExponentialSmoothing_MSAT_35621168

Use Holt-Winters to build a time series model suitable for MSAT data with trend and seasonality.

ARIMA_MSAT_35621168

Fitting the MSAT data using the ARIMA(1,1,2) model

MultipleLinearregression_35621168

The least squares method (OLS) was used to fit the multivariate linear regression model to the MAST data and the fitting results were returned.