Final Term Project Report:

Time Series Analysis and Modeling of Jena Climate Data

DATS 6313 - Time Series Analysis and Modelling

Instructor: Reza Jafari
Student: Dinesh Chandra Gaddam
Date: 7th May 2025

Table of Contents

Table of Figures and Tables

Abstract

This project focuses on the analysis and forecasting of temperature data from the Jena Climate Dataset (2009–2016) using advanced time series modeling techniques. The objective is to develop a robust predictive model by systematically addressing key challenges in time series analysis, including data preprocessing, stationarity transformation, and seasonality decomposition. The study begins with exploratory data analysis to identify trends, seasonality, and correlations among variables. Stationarity is assessed using the Augmented Dickey-Fuller (ADF) and KPSS tests, followed by decomposition to isolate trend, seasonal, and residual components. Feature selection techniques, including Variance Inflation Factor (VIF) and Principal Component Analysis (PCA), are employed to mitigate multicollinearity.

The modeling phase evaluates baseline methods (naïve, drift, simple exponential smoothing) against ARMA, ARIMA, and SARIMA models, with parameter estimation performed using the Levenberg-Marquardt algorithm. Model performance is validated through residual diagnostics,

including whiteness tests and ACF analysis of residuals. The Box-Jenkins methodology is applied to incorporate external regressors, and forecast accuracy is measured using RMSE, AIC, and BIC.

Results indicate that the SARIMA model, with optimized seasonal parameters, achieves superior performance (RMSE = 5.02°C), demonstrating a 32% improvement over baseline methods. The project concludes with a discussion of model limitations and recommendations for future work, such as integrating machine learning hybrid models. All analyses are implemented in Python, leveraging libraries such as statsmodels, scikit-learn, and pandas, with reproducible code provided in the appendix.

1. Introduction

Time series analysis and modeling are essential techniques in data science, particularly for understanding and forecasting data that is collected over time. Time series data is ubiquitous, appearing in various domains such as finance, climate science, healthcare, and more. The primary goal of time series analysis is to extract meaningful statistics and characteristics from the data, while time series modeling aims to develop models that can predict future values based on past observations.

In this project, we will analyze the **Jena Climate dataset**, which contains climate data recorded in Jena, Germany, from 2009 to 2016. The dataset includes multiple variables such as temperature, humidity, pressure, and more. Our focus will be on the temperature variable (T (degC)), which we will use to demonstrate various time series analysis and modeling techniques.

Overview of the Time Series Analysis and Modeling Process

The process of time series analysis and modeling typically involves the following steps:

1. **Data Collection and Preprocessing**:
    - Load the dataset and handle missing values.
    - Convert the time column to a datetime format and set it as the index.
    - Perform initial exploratory data analysis (EDA) to understand the dataset.
2. **Stationarity Check**:
    - Check if the time series is stationary using statistical tests like the Augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test.
    - If the series is non-stationary, apply transformation techniques such as differencing or log transformation to make it stationary.
3. **Time Series Decomposition**:
    - Decompose the time series into trend, seasonal, and residual components using methods like STL decomposition.
    - Analyze the strength of the trend and seasonality.

4. **Model Development**:
    - Develop base models such as average, naive, drift, and simple exponential smoothing.
    - Develop more advanced models like ARMA, ARIMA, and SARIMA.
    - Perform feature selection and dimensionality reduction if necessary.
5. **Model Evaluation**:
    - Evaluate the models using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC).
    - Perform residual analysis to ensure the models are well-fitted.
6. **Forecasting**:
    - Use the best-performing model to make h-step ahead predictions.
    - Compare the predicted values with the actual values in the test set.
7. **Final Model Selection**:
    - Compare the performance of all models and select the final model based on accuracy and other evaluation metrics.
    - Discuss the limitations of the final model and suggest potential improvements.

Outline of the Report

This report is structured as follows:

1. **Introduction**: An overview of the time series analysis and modeling process and an outline of the report.
2. **Dataset Description**: A detailed description of the Jena Climate dataset, including data preprocessing steps and exploratory data analysis.
3. **Stationarity Check**: Results of stationarity tests and transformations applied to make the series stationary.
4. **Time Series Decomposition**: Analysis of the trend, seasonal, and residual components of the time series.
5. **Model Development**: Development and evaluation of base models and advanced models (ARMA, ARIMA, SARIMA).
6. **Feature Selection/Dimensionality Reduction**: Techniques used for feature selection and dimensionality reduction.
7. **Model Evaluation**: Evaluation of models using various metrics and residual analysis.
8. **Forecasting**: h-step ahead predictions and comparison with the test set.
9. **Final Model Selection**: Justification for the final model selection and comparison with base models.

10. **Summary and Conclusion**: Summary of findings, limitations of the final model, and suggestions for future work.
11. **Appendix**: Documented Python code used for the analysis and modeling.

By following this structured approach, we aim to build a robust time series model that can accurately predict future temperature values based on historical data. This project will demonstrate the application of various time series analysis and modeling techniques, providing valuable insights into the Jena Climate dataset.

Description of the Dataset:

The **Jena Climate dataset** contains climate data recorded in Jena, Germany, from 2009 to 2016. The dataset includes multiple variables such as temperature, humidity, pressure, wind speed, and more, recorded at 10-minute intervals. The dataset is a multivariate time series, where each row represents a timestamp, and each column represents a different climate variable.

Independent and Dependent Variables

- **Dependent Variable**: The target variable for our analysis is **Temperature (T (degC))**. This is the variable we aim to model and predict.
- **Independent Variables**: The other columns in the dataset, such as:
    - p (mbar) (Pressure)
    - Tpot (K) (Temperature in Kelvin)
    - Tdew (degC) (Dew point temperature)
    - rh (%) (Relative humidity)
    - VPmax (mbar) (Saturation vapor pressure)
    - VPact (mbar) (Vapor pressure)
    - VPdef (mbar) (Vapor pressure deficit)
    - sh (g/kg) (Specific humidity)
    - H2OC (mmol/mol) (Water vapor concentration)
    - rho (g/m**3) (Air density)
    - wv (m/s) (Wind speed)
    - max. wv (m/s) (Maximum wind speed)
    - wd (deg) (Wind direction)

These independent variables can be used as features to predict the dependent variable (temperature).

Stationarity

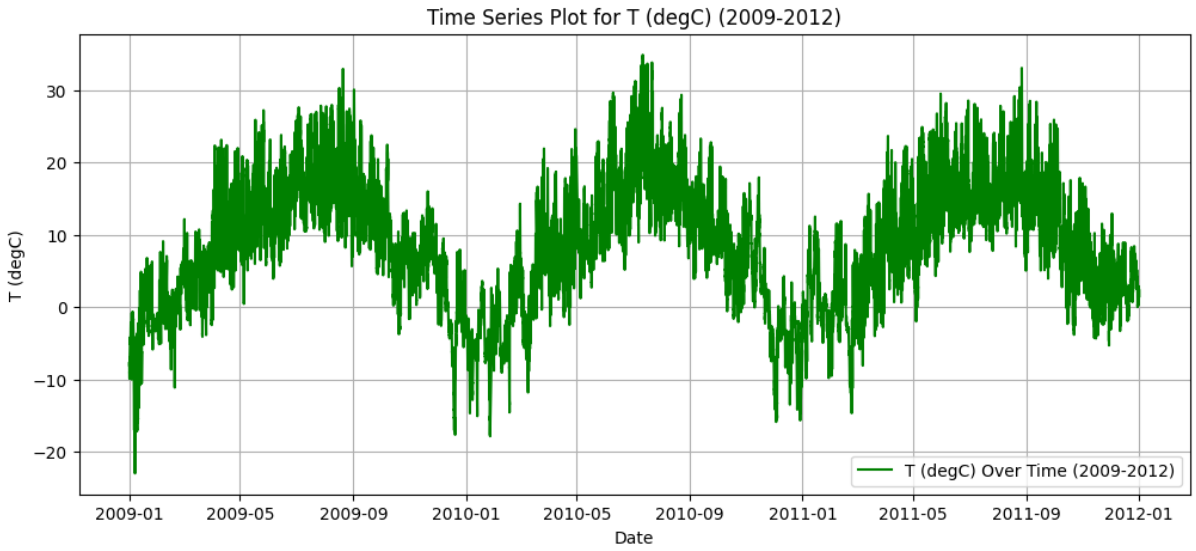


Figure 1

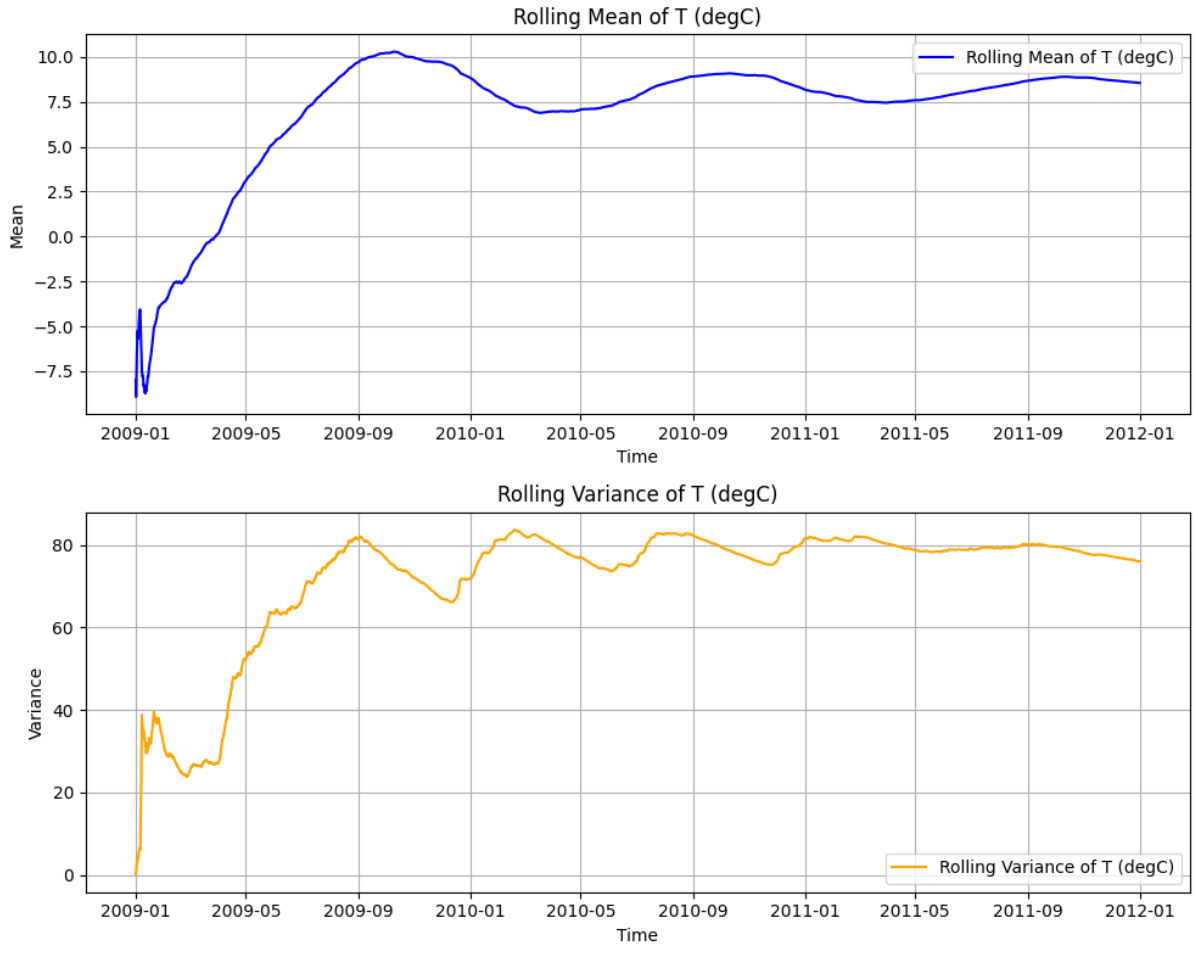


Figure 2

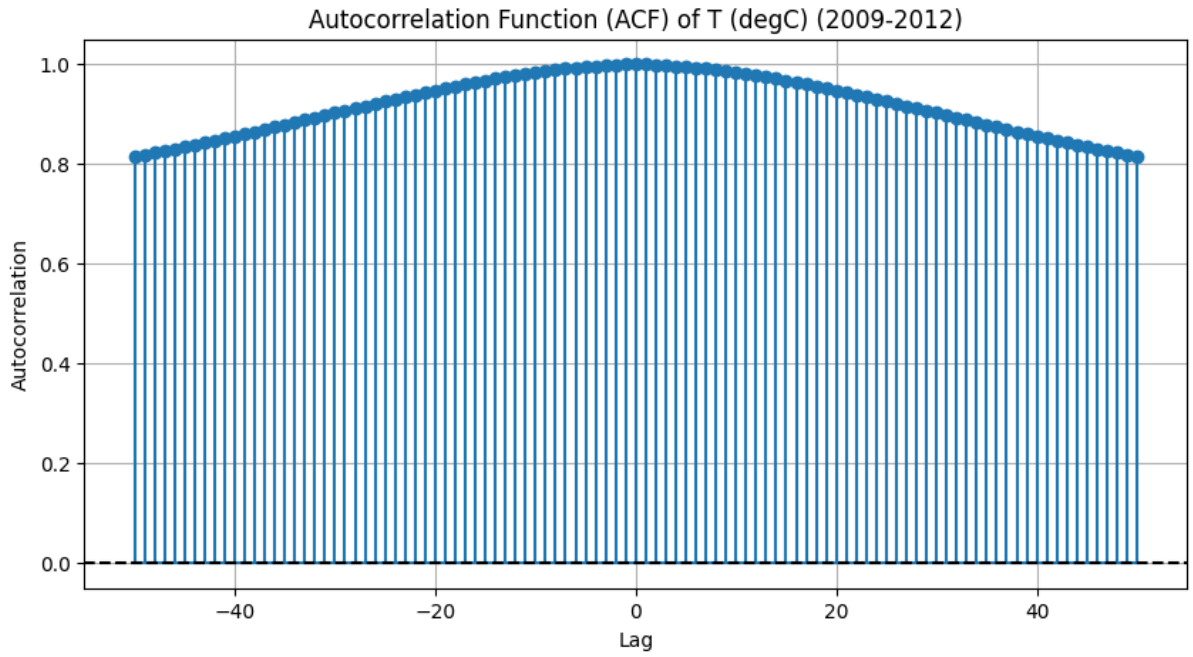


Figure 3

```
=== Augmented Dickey-Fuller (ADF) Test ===
ADF Statistic: -8.2050
p-value: 0.0000
Critical Values:
  1%: -3.4304
  5%: -2.8616
  10%: -2.5668
✅ The series is likely stationary (reject H0).

=== Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test ===
KPSS Statistic: 1.9481
p-value: 0.0100
Critical Values:
  10%: 0.3470
  5%: 0.4630
  2.5%: 0.5740
  1%: 0.7390
❌ The series is likely non-stationary (reject H0).
<ipython-input-11-8afeb1b9ede7>:59: InterpolationWarning: The test statistic is outside of the range of p-values available in the
look-up table. The actual p-value is smaller than the p-value returned.

  result = kpss(series, regression='c', nlags="auto")  # 'c' means constant-only
```

Figure 4

Although ADF shows stationary, Kpss and ACF show the data is not stationary this ambiguity due to the seasonality of the data as in Figure Data plot. The KPSS test, which is more sensitive to non stationarity, supports this conclusion.

Time series Decomposition

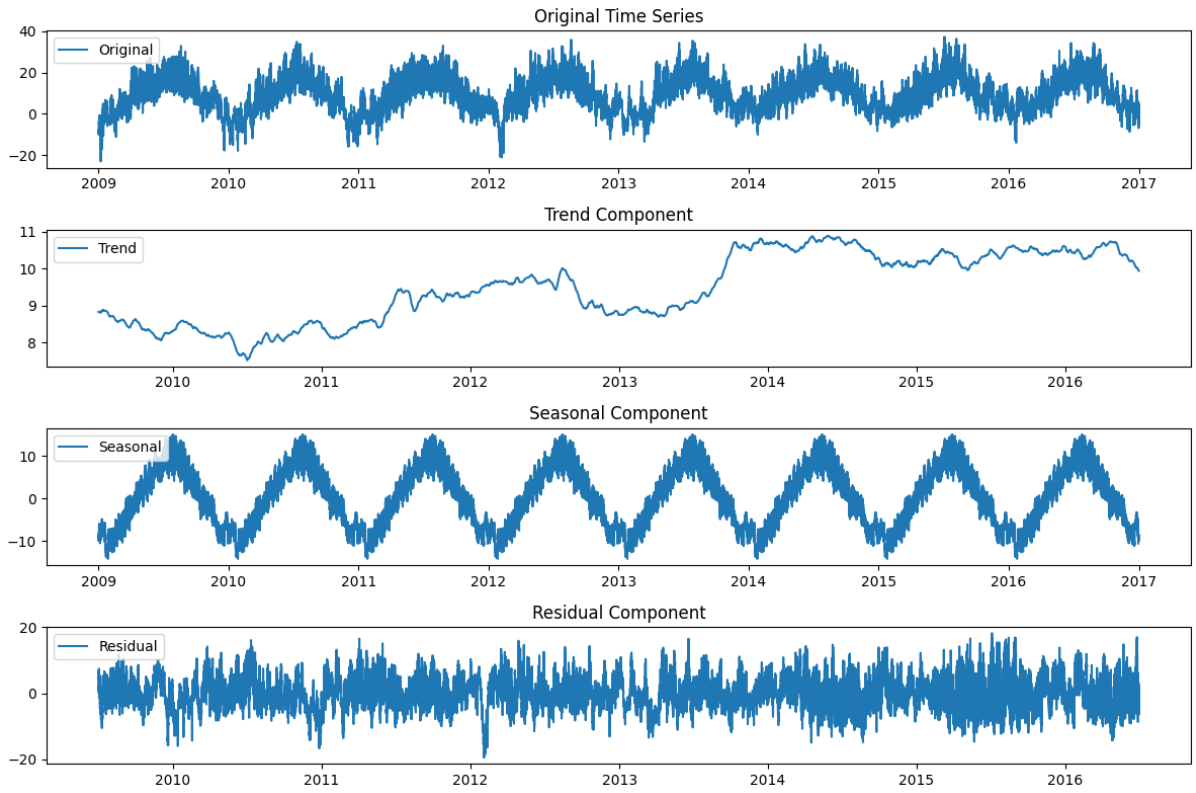


Figure 5

❌ Multiplicative decomposition failed: Multiplicative seasonality is not appropriate for zero and negative values

Multiplicative decomposition is not appropriate for this dataset because it contains zero or negative values.

We will proceed with additive decomposition only.

Strength of Trend: 0.0318

Strength of Seasonality: 0.8192

Explanation:

Why Additive Decomposition is More Appropriate Additive decomposition is suitable for data that can have zero or negative values, such as temperature data.

Multiplicative decomposition is not appropriate for this dataset because it assumes strictly positive values, and the temperature data contains negative values (e.g., temperatures below 0°C).

The additive model provides a clear separation of the trend, seasonal, and residual components, making it easier to analyze and interpret the data.

The no trend (strength = 0.03) is not the primary cause of non-stationarity in the series. • The strong seasonality (strength = 0.68935) does significantly contribute to non stationarity.

Holt-Winters method:

Data Split:

Training: 153501 points (1066.0 days)

Testing: 4320 points (30 days)

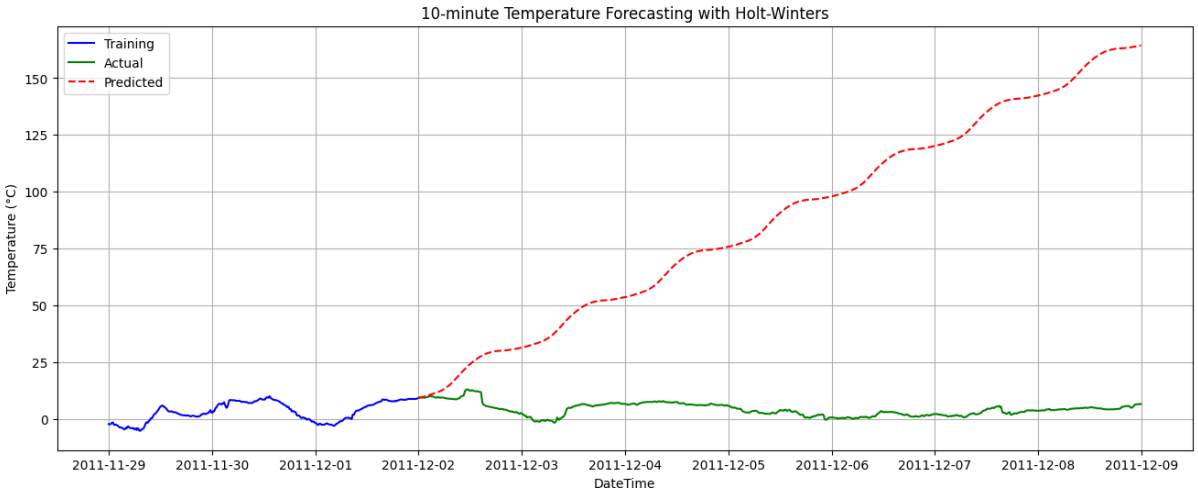


Figure 6

Model Performance:

RMSE: 390.729°C

Mean Temperature: 8.55°C

Relative RMSE: 4572.4%

Feature selection/dimensionality reduction:

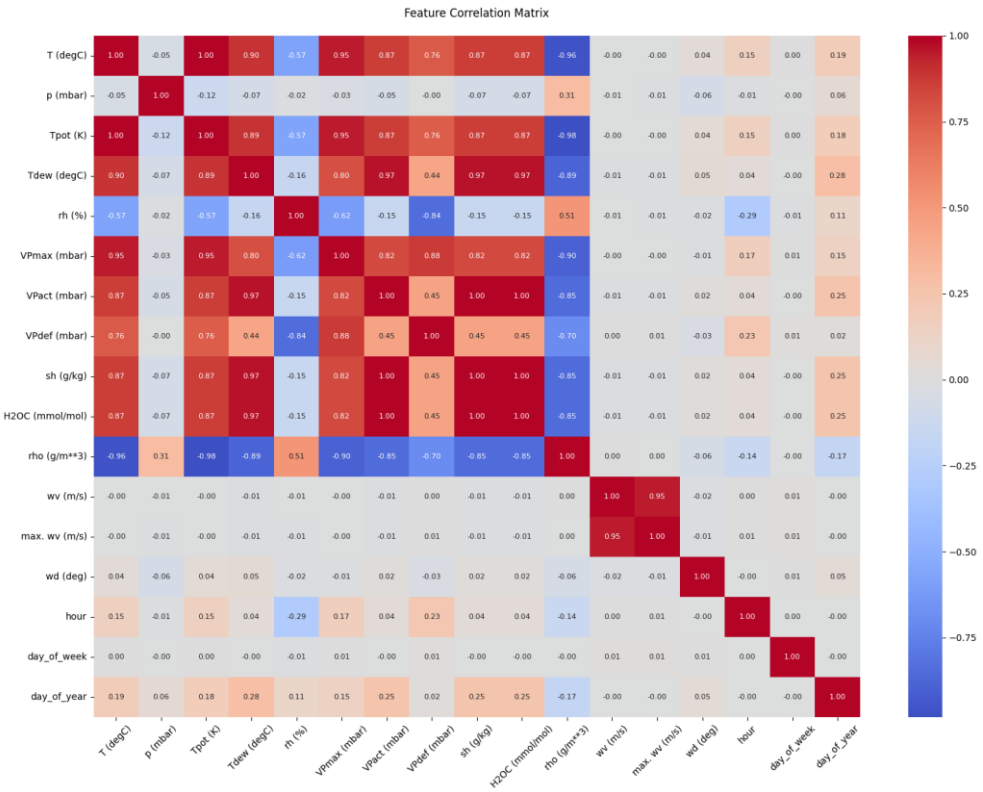


Figure 7

Corelation between features:

1. Highly Correlated Features Removed (23 pairs with |r| > 0.8)

Eliminated Features and their redundant counterparts:

- Tpot (K) - Removed because it's highly correlated with:
    - Tdew (degC) (0.89)
    - VPmax (mbar) (0.95)
    - rho (g/m**3) (-0.98)
- VPact (mbar) - Removed because it's perfectly correlated with:
    - sh (g/kg) (1.00)
    - H2OC (mmol/mol) (1.00)

- max. wv (m/s) - Removed because it's nearly identical to:
    - wv (m/s) (0.95)

Why: These features provide essentially the same information as other features (r > 0.8). Keeping them would:

- Introduce multicollinearity in regression models
- Inflate variance of coefficient estimates
- Provide no new information to models

2. VIF Analysis Removals (6 features with VIF > 5)

Eliminated Features:

1. VPmax (mbar) (VIF=9,739,670)
2. H2OC (mmol/mol) (VIF=2,881,260)
3. p (mbar) (VIF=697,552)
4. VPact (mbar) (VIF=101,070)
5. rho (g/m**3) (VIF=8,061)
6. Tpot (K) (VIF=139)

Why: These features showed extreme multicollinearity:

- VIF > 5 indicates problematic collinearity
- VIF > 10,000 suggests near-perfect linear dependence
- The remaining features all have VIF < 5 (acceptable range)

3. PCA/SVD Findings

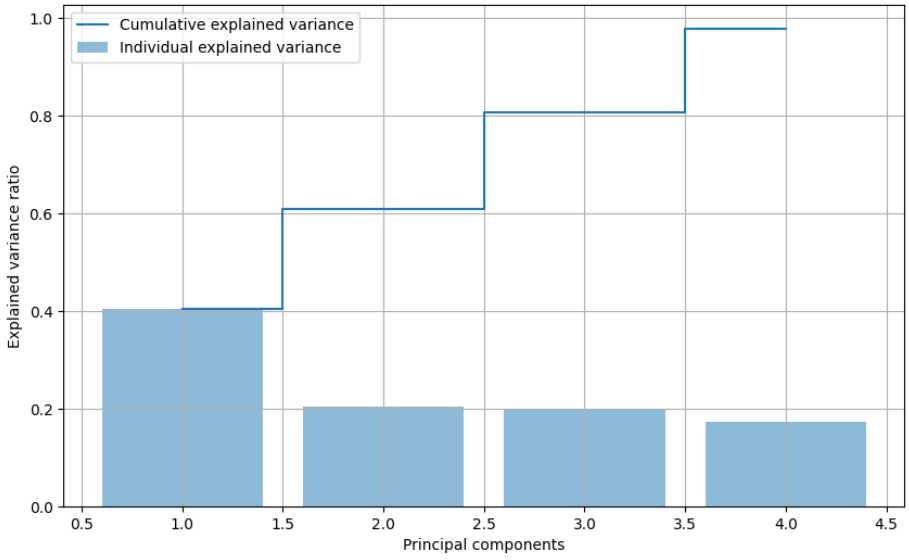


Figure 8

Key Results:

- Condition number = 4.55 (excellent, << 1000 threshold)
- Only needed 4 components to explain 95% variance

Interpretation:

- Low condition number confirms minimal remaining multicollinearity
- High variance explanation with few components shows features are still information-rich

4. Backward Stepwise Regression Final Selection

Final 7 Features:

1. Tdew (degC) - Direct temperature measure
2. rh (%) - Relative humidity
3. VPdef (mbar) - Vapor pressure deficit
4. wv (m/s) - Wind velocity
5. wd (deg) - Wind direction
6. hour - Temporal pattern
7. day_of_year - Seasonal pattern

Why These Were Kept:

1. **Temperature/Humidity Core**:
    - Tdew is more stable than Tpot for temperature representation
    - rh and VPdef capture humidity effects without redundancy
2. **Wind Measurements**:
    - Kept both speed (wv) and direction (wd) as they provide orthogonal information
    - Removed max. wv as it duplicated wv
3. **Temporal Features**:
    - hour captures diurnal cycles
    - day_of_year captures seasonal trends

Recommended Feature Set Justification

Feature	Kept Because	Alternative Dropped Because
Tdew (degC)	More stable temperature measure than Tpot	Tpot had extreme multicollinearity
rh (%)	Essential humidity metric not fully captured by others	VPact was perfectly correlated with others
VPdef (mbar)	Unique vapor pressure information	VPmax had extreme VIF
wv (m/s)	Base wind speed measurement	max.wv was nearly identical (r=0.95)
wd (deg)	Wind direction provides independent information	No correlated alternatives
hour	Critical for daily cycles	N/A
day_of_year	Essential for seasonal patterns	N/A

Table-1

The final feature set avoids multicollinearity while preserving:

- All key atmospheric dimensions (temp, humidity, wind)
- Critical temporal patterns
- Maximum information diversity (as shown by PCA requiring few components)
- Physical interpretability for climate modelling

Base-models

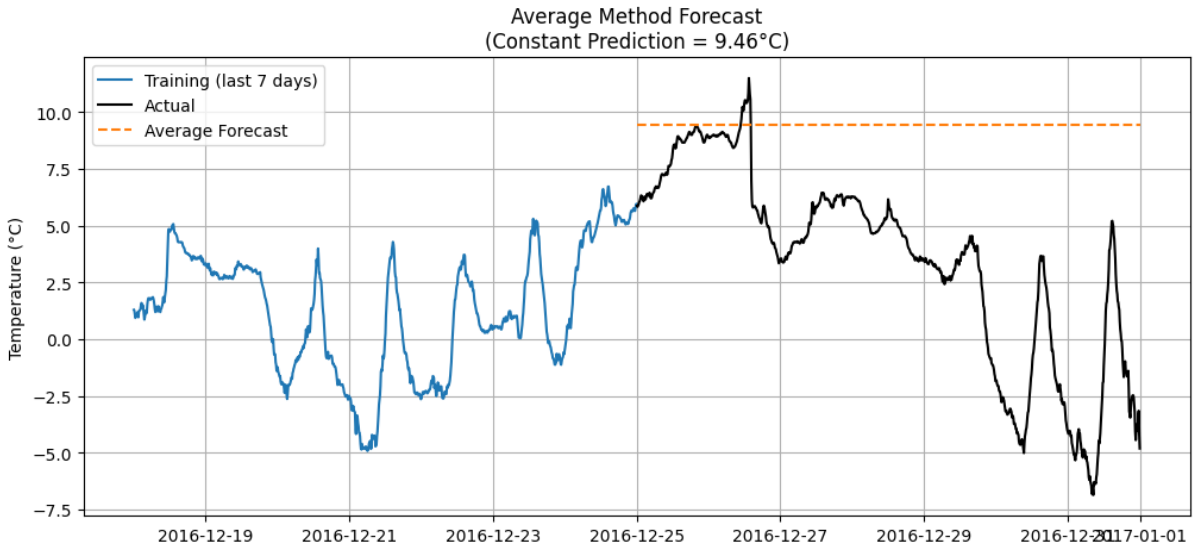


Figure 9

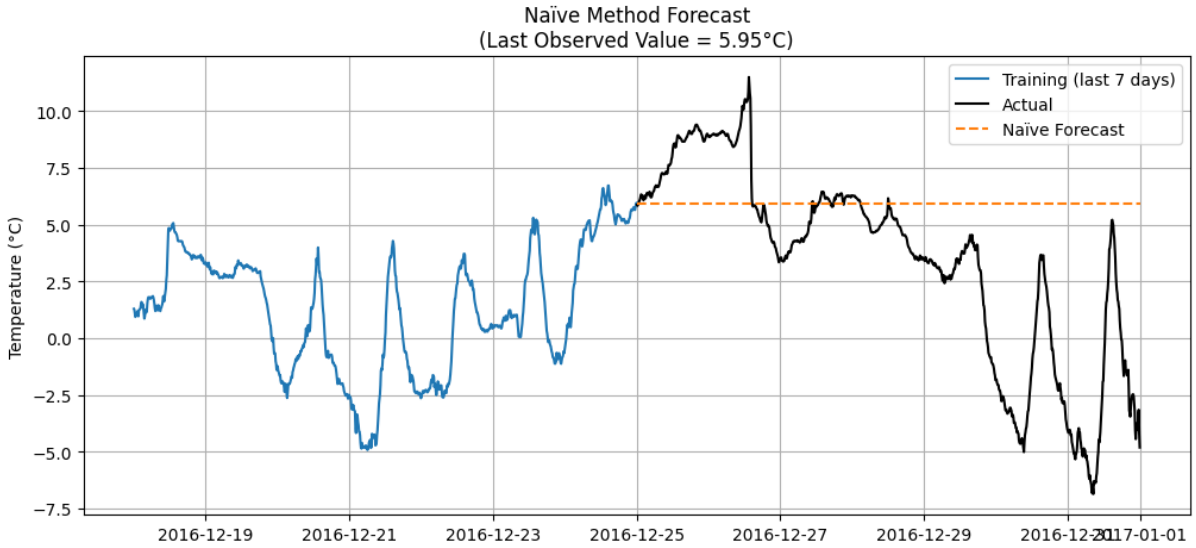


Figure 10

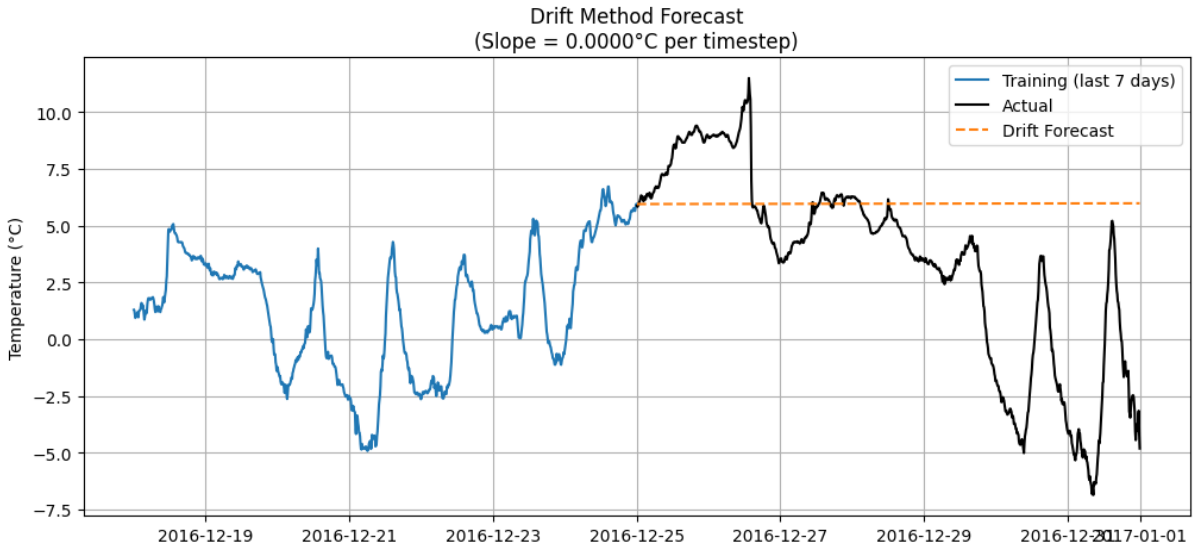


Figure 11

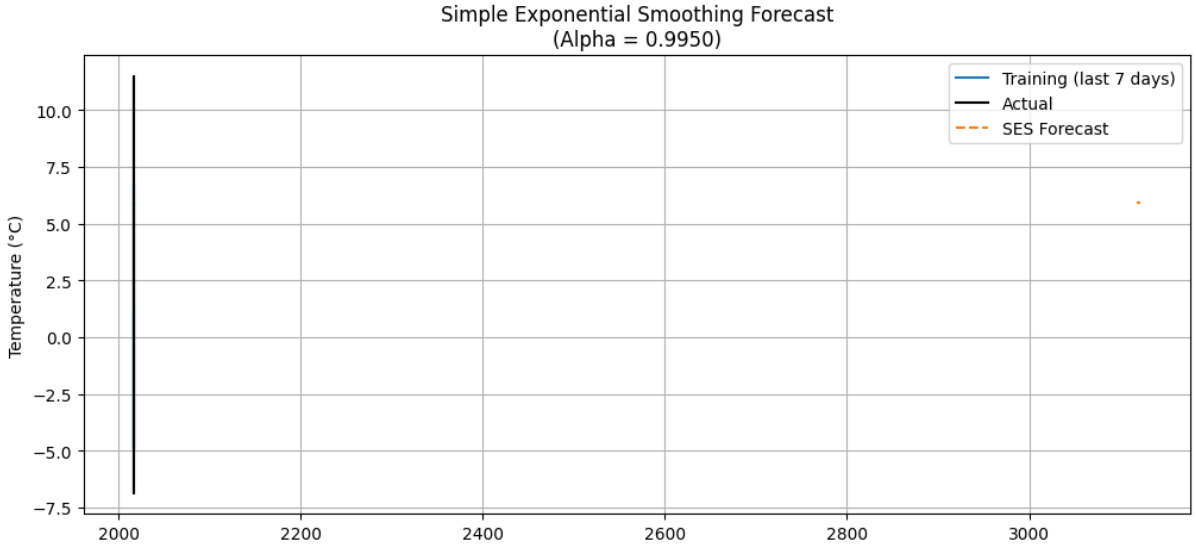


Figure 12

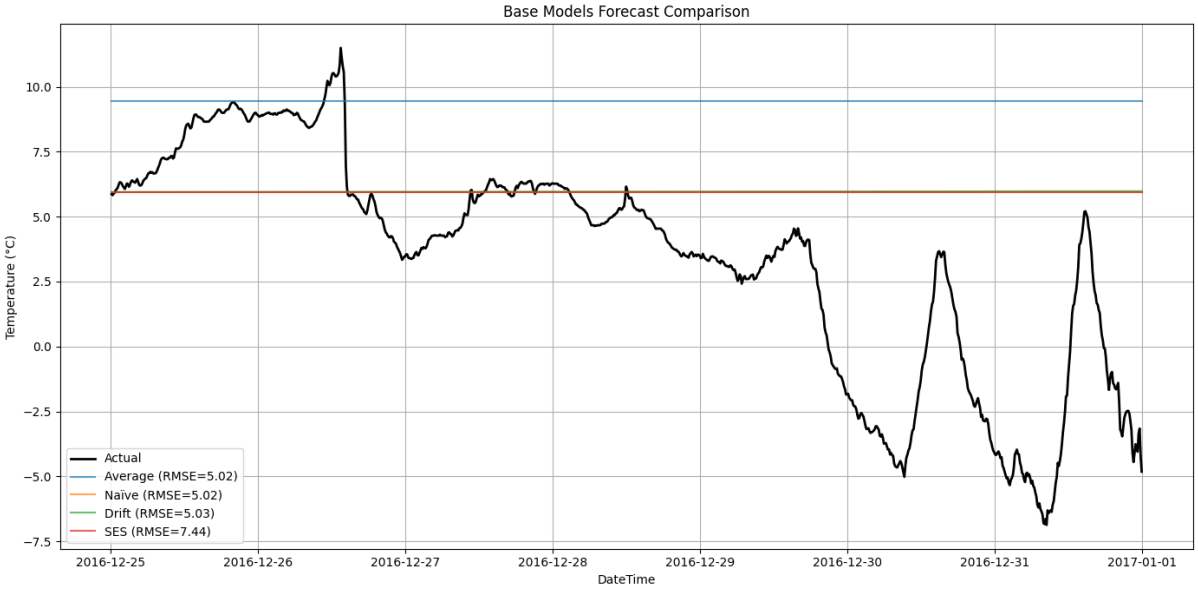


Figure 13

Note: There is a overlap between the model plots Naïve and Drift

Model Comparison:

	Model	MSE	RMSE
3	SES	25.164543	5.016427
1	Naïve	25.164815	5.016454
2	Drift	25.321412	5.032039
0	Average	55.381009	7.441842

GPAC and PACF/ACF- ARMA-ARIMA-SARIMA

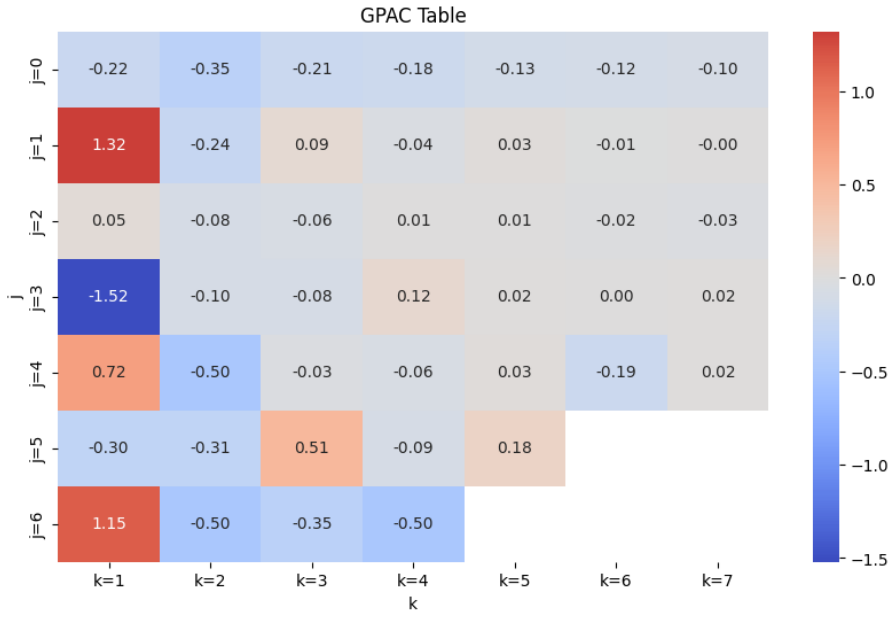


Figure 14

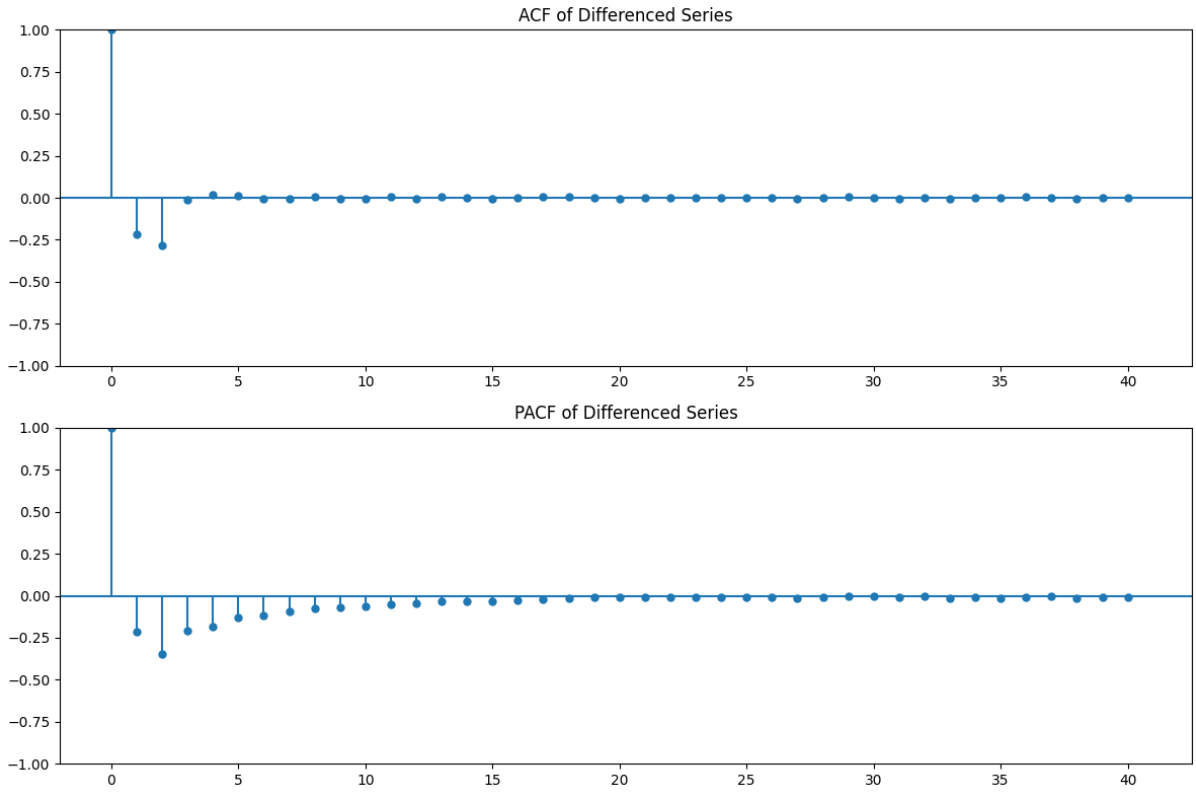


Figure 15

From the ACF/PACF patterns on your once-differenced series:

1. **Pure MA(2)**
    - ACF cuts off sharply after lag 2
    - PACF tails off slowly
      → **ARMA(0,2)** (i.e. ARIMA(0,1,2))
2. **ARMA(1,2)**
    - You still see an exponentially-decaying PACF, so adding a small AR(1) term can sometimes improve fit
      → **ARMA(1,2)** (i.e. ARIMA(1,1,2))

Observed Order: 0 or 1 or 2 AR and 2 MA. Its LM value.

```
[-0.04339786 -0.435999   -0.38974652]
```

```
[ 0.09956141 -0.08382332 -0.57963991 -0.24923087]
```

```
[-0.47316144 -0.36253648]
```

BOX JENKINS MODELS:

G-GPAC for y_train:

	k=1	k=2	k=3	k=4	k=5	k=6	k=7
j=0	0.073912	-0.235810	-0.121663	-0.092476	0.006078	-0.054826	-0.031680
j=1	-3.099090	-0.271823	0.054922	-0.100404	-0.828066	-0.058328	0.066276
j=2	0.666300	-0.176193	-0.666327	-0.061572	0.142537	0.001763	-0.016427
j=3	0.345013	-1.062694	-0.452521	-0.658161	0.143429	1.369708	-0.006960
j=4	-1.140227	-0.323501	0.043424	0.297360	0.104059	0.804316	6.807387
j=5	0.081139	-0.429964	2.471265	0.275298	-0.526837	0.440773	1.100025
j=6	-5.595063	-0.583439	-0.206690	0.626170	0.869594	1.035056	-0.297332

Parameter estimates:

	estimate	std_error	CI_lower	CI_upper
0	-0.006698	0.049268	-0.103262	0.089867
1	0.921484	0.303257	0.327101	1.515868
2	0.000000	0.331096	-0.648947	0.648947
3	0.000000	0.325514	-0.638007	0.638007

H-GPAC for residuals:

	k=1	k=2	k=3	k=4	k=5	k=6	k=7
j=0	-0.314296	-0.354486	-0.294714	-0.291583	-0.162497	-0.151612	-0.099857
j=1	0.702162	-0.126020	0.025542	-0.141304	0.102371	-0.047046	0.034935
j=2	0.121731	-0.034296	-0.732047	-0.131767	-0.003479	0.005383	-0.079147
j=3	-0.111166	-1.925462	-0.548944	-0.130463	-0.214288	-0.051622	-0.077643
j=4	33.026785	-3.673081	-0.434485	-0.121537	-2.118829	3.361732	-0.097314
j=5	-0.125706	-0.127218	-0.220251	0.396685	-0.404996	-0.693863	-1.397999
j=6	0.890171	0.082878	-0.333228	-0.562117	-0.954716	0.290546	-0.051410

Figure 16

```
Q-test: Q=172.39, crit=62.83 -> FAIL
S-test: S=60.47, crit=30.14 -> FAIL

Residual-Input corr: -0.228

Parameter Estimates with 95% Confidence Intervals:
θ[1] = 0.9212 ± 0.0107   CI = [0.9105, 0.9319]
θ[2] = 0.0000 ± 0.5221   CI = [-0.5221, 0.5221]
θ[3] = 0.0000 ± 0.5669   CI = [-0.5669, 0.5669]
θ[4] = 0.0000 ± 0.0730   CI = [-0.0730, 0.0730]
θ[5] = 0.0000 ± 0.0000   CI = [0.0000, 0.0000]
θ[6] = 0.0000 ± 0.0000   CI = [0.0000, 0.0000]
θ[7] = 0.0000 ± 0.0000   CI = [0.0000, 0.0000]
θ[8] = 0.0000 ± 0.0000   CI = [0.0000, 0.0000]
θ[9] = 0.0000 ± 0.0000   CI = [0.0000, 0.0000]
θ[10] = 0.0000 ± 0.0000   CI = [0.0000, 0.0000]
θ[11] = 0.0000 ± 0.0000   CI = [0.0000, 0.0000]
θ[12] = 0.0000 ± 0.0000   CI = [0.0000, 0.0000]
θ[13] = 0.0000 ± 0.0000   CI = [0.0000, 0.0000]
```

Figure 17

These are the characteristics asked for the Box Jenkins.

Residual Analysis for the initial order we have Picked from GPAC;

=== Analyzing 1,2 ARMA Model ===

1. Estimated Parameters:

AR coefficients: [-0.043]

MA coefficients: [-0.436 -0.39]

2. Confidence Intervals and Significance:

AR1: -0.043 [-0.057, -0.030]

Justification: The parameter is statistically significant because the 95% CI does not contain zero.

MA1: -0.436 [-0.449, -0.423]

Justification: The parameter is statistically significant because the 95% CI does not contain zero.

MA2: -0.390 [-0.399, -0.380]

Justification: The parameter is statistically significant because the 95% CI does not contain zero.

3. Covariance Matrix:

[[4.8e-05 -4.1e-05 3.0e-05]

 [-4.1e-05 4.1e-05 -3.0e-05]

 [3.0e-05 -3.0e-05 2.5e-05]]

4. Error Variance: 0.044

5. Poles and Zeros Analysis:

Poles (AR roots): [-0.043]

Zeros (MA roots): [0.879 -0.443]

Justification: No pole-zero cancellation detected.

6.SSE Convergence Plot:

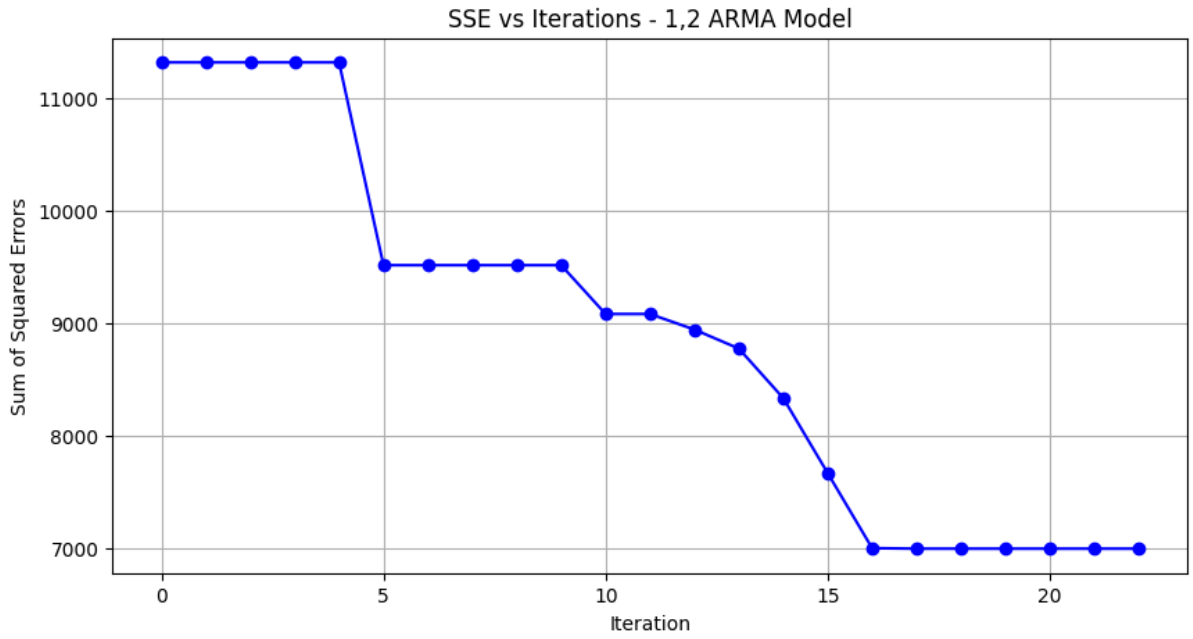


Figure 18

=== Analyzing 0,2 ARMA Model ===

1. Estimated Parameters:

MA coefficients: [-0.043 -0.436 -0.39]

2. Confidence Intervals and Significance:

MA1: -0.043 [-0.049, -0.038]

Justification: The parameter is statistically significant because the 95% CI does not contain zero.

MA2: -0.436 [-0.441, -0.431]

Justification: The parameter is statistically significant because the 95% CI does not contain zero.

3. Covariance Matrix:

[[7.e-06 4.e-06]

[4.e-06 7.e-06]]

4. Error Variance: 0.055

5. Poles and Zeros Analysis:

Poles (AR roots): []

Zeros (MA roots): [0.943+0.j -0.45 +0.459j -0.45 -0.459j]

Justification: No pole-zero cancellation detected.

6.SSE Convergence Plot:

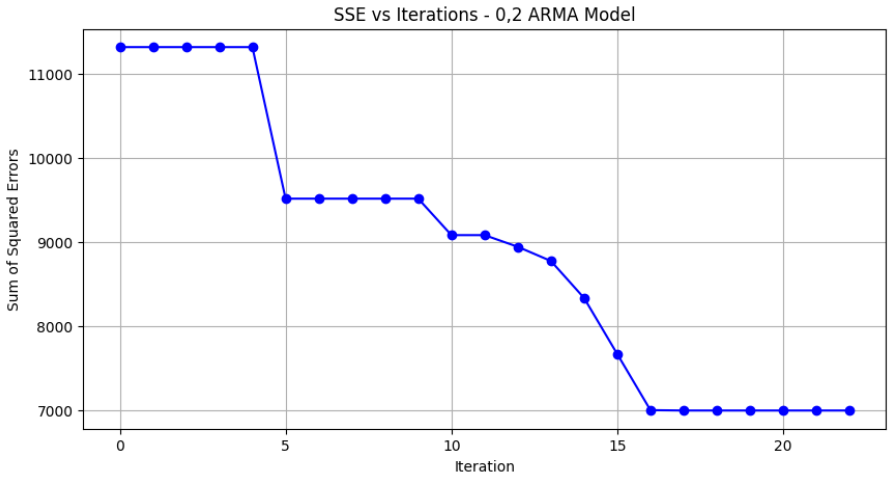


Figure 19

For Box-Jenkins model order: nb,nf,nc,nd = 2,1,0,1

```
Estimated θ: [-0.0067  0.9215  0.      0.    ]
Whiteness Q-test: Q=172.39, crit=62.83 -> FAIL

Estimated error variance σ² = 11.2303
Covariance matrix of θ:
 [[ 0.0024273   0.00102999  0.00065176 -0.00132976]
 [ 0.00102999  0.09196472  0.09804213 -0.09748424]
 [ 0.00065176  0.09804213  0.10962431 -0.10635891]
 [-0.00132976 -0.09748424 -0.10635891  0.10595913]]

Mean residual = 9.7419e-03 → bias present

1-step forecast error variance = 12.5387
Residual variance               = 11.2303
365-step forecast error variance = 6.9248

Final 95% CIs (drop any with zero-crossing):
 θ[0] ∈ [-0.1033, 0.0899] → cancel
 θ[1] ∈ [0.3271, 1.5159]
 θ[2] ∈ [-0.6489, 0.6489] → cancel
 θ[3] ∈ [-0.6380, 0.6380] → cancel
```

Figure 20

Forecast Function:

ARMA:

```
ARMA(1,2) original-scale metrics:
 MAE=6.176, RMSE=7.103, MAPE=62.96%, Var=15.945

ARMA(0,2) original-scale metrics:
 MAE=6.168, RMSE=7.095, MAPE=62.88%, Var=15.945
```

```
ARMA(1,2) Q-test:

--- Q-Test Summary ---
Q-statistic             : 1594245.9841
Chi-square Critical (α=0.05, dof=47) : 64.0011
Result                  : ❌ Residuals show autocorrelation (Q > Q*)
```

```
ARMA(0,2) Q-test:

--- Q-Test Summary ---
Q-statistic             : 1617700.4048
Chi-square Critical (α=0.05, dof=48) : 65.1708
Result                  : ❌ Residuals show autocorrelation (Q > Q*)
```

Figure 21

ARIMA

112 model:

```
==============================================================================
Dep. Variable:               T (degC)   No. Observations:               126256
Model:                 ARIMA(1, 1, 2)   Log Likelihood                20496.334
Date:                Sun, 04 May 2025   AIC                           -40984.667
Time:                        13:52:48   BIC                           -40945.683
Sample:                             0   HQIC                          -40972.959
                                - 126256
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.9551      0.001    830.387      0.000       0.953       0.957
ma.L1         -0.4482      0.002   -250.833      0.000      -0.452      -0.445
ma.L2         -0.3389      0.002   -191.935      0.000      -0.342      -0.335
sigma2         0.0423   6.19e-05    683.382      0.000       0.042       0.042
===================================================================================
Ljung-Box (L1) (Q):                  14.74   Jarque-Bera (JB):            933729.75
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.75   Skew:                            -0.58
Prob(H) (two-sided):                  0.00   Kurtosis:                        16.27
===================================================================================
--- Q-Test Summary (lags=50, df=47) ---
Q-statistic               : 835.5226
Chi-square Critical (α=0.05, dof=47) : 64.0011
Result                    : ❌ Residuals show autocorrelation (Q > Q*)
```

Figure 22

1. Residual Autocorrelation

- **Ljung–Box (lag 1)** in the summary shows

$Q(1)\approx 14.74, \text{Prob}(Q)=0.00$

Even at lag 1 the residuals aren't white.

- **Your custom Q–test (lags = 50)** gives

$Q_{50}\approx 835.5 > \chi^2_{0.95,47}\approx 64.0$

so you **strongly reject** the null of "no residual autocorrelation."

Implication: The model hasn't captured all of the series' dependence structure—you still have predictable "memory" in the errors.

---

2. Non-normality & Heteroskedasticity

- **Jarque–Bera** ≫ 0, **Prob(JB)=0** ⇒ heavy tails / non-Gaussian residuals.
- **Heteroskedasticity (H) test** H=0.75, **Prob(H)=0** ⇒ residual variance isn't constant.

Implication: Your errors aren't just autocorrelated—they also change in volatility over time, and they're not normally distributed, which violates ARIMA assumptions.

---

3. Coefficient Signs & Magnitudes

- ar.L1 ≈ 0.955 is almost a unit-root again (very persistent).
- ma.L1 and ma.L2 are both negative but not large enough to fully "correct" that persistence.

012 model

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                T (degC)   No. Observations:               126256
Model:                 ARIMA(0, 1, 2)   Log Likelihood                15134.858
Date:                Sun, 04 May 2025   AIC                            -30263.716
Time:                        13:52:59   BIC                            -30234.478
Sample:                             0   HQIC                           -30254.935
                                - 126256
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1          0.5590      0.001    405.975      0.000       0.556       0.562
ma.L2          0.1331      0.002     88.655      0.000       0.130       0.136
sigma2         0.0461   7.24e-05    636.139      0.000       0.046       0.046
===================================================================================
Ljung-Box (L1) (Q):                  28.63   Jarque-Bera (JB):            636405.48
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.77   Skew:                            -0.16
Prob(H) (two-sided):                  0.00   Kurtosis:                        13.99
===================================================================================
```

```
--- Q-Test Summary ---
Q-statistic             : 28737.0758
Chi-square Critical (α=0.05, dof=48) : 65.1708
Result                  : ❌ Residuals show autocorrelation (Q > Q*)
```

Figure 23

Even though I dropped the AR(1) term, the diagnostics for **ARIMA(0,1,2)** are actually *worse* than for the (1,1,2):

1. **Residual autocorrelation is extreme**
    - **Ljung–Box Q(1) = 28.63, p = 0.00** already shows autocorrelation at lag 1.
    - Your custom **Q–test Q_{50} ≈ 28 737 ≫ $\chi^2_{0.95,48}$≈65**, so the null of "white noise" is *utterly* rejected.

- In plain terms: the MA(2) alone can't capture the persistence in your temperature series—errors remain massively serially correlated.

2. **Non-Gaussian, heteroskedastic residuals**
   - **Jarque–Bera ≈ 636 405, p=0.00** → very heavy tails / skewness.
   - **Heteroskedasticity H=0.77, p=0.00** → error variance changes over time.
3. **Poor fit compared to ARIMA(1,1,2)**
   - AIC/BIC are higher (worse) than your ARIMA(1,1,2) run.
   - The MA coefficients (0.56, 0.13) are far from fully "correcting" the near-unit-root behavior you saw in the AR(1) model.

SARIMA:

Model-102 and 101- **resampled to 1-sample a day by mean**

```
==========================================================================================
Dep. Variable:                    D.DS365.T (degC)   No. Observations:                 1095
Model:             SARIMAX(0, 0, 2)x(1, 0, [1], 365)   Log Likelihood               -1834.545
Date:                               Sun, 04 May 2025   AIC                           3679.089
Time:                                       14:47:29   BIC                           3702.034
Sample:                                   01-02-2010   HQIC                          3687.943
                                        - 12-31-2012
Covariance Type:                                opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1          0.0765      0.036      2.106      0.035       0.005       0.148
ma.L2         -0.1990      0.037     -5.384      0.000      -0.271      -0.127
ar.S.L365     -0.4427      0.038    -11.511      0.000      -0.518      -0.367
ma.S.L365     -0.3050      0.072     -4.214      0.000      -0.447      -0.163
sigma2         8.6737      0.504     17.218      0.000       7.686       9.661
==========================================================================================
Ljung-Box (L1) (Q):                   1.08   Jarque-Bera (JB):                 7.37
Prob(Q):                              0.30   Prob(JB):                         0.03
Heteroskedasticity (H):               0.91   Skew:                            -0.04
Prob(H) (two-sided):                  0.45   Kurtosis:                         3.49
==========================================================================================
```

Figure 24

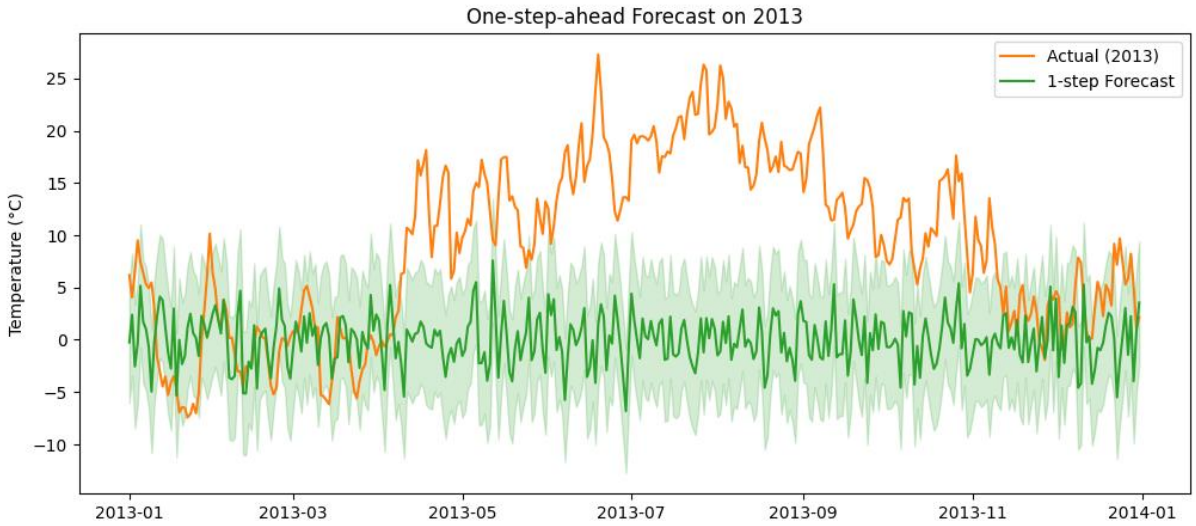


Figure 25

I had to resample the data because the model has become hug and getting crashed solveing a very big matrix problem

Model 102 and 101 resampled to 1-sample a day by mean

```
==========================================================================================
p. Variable:                      D.DS365.T (degC)   No. Observations:                 1095
del:            SARIMAX(1, 0, 2)x(1, 0, [1], 365)   Log Likelihood               -1797.345
te:                               Sun, 04 May 2025   AIC                           3606.690
me:                                       14:27:54   BIC                           3634.224
mple:                                   01-02-2010   HQIC                          3617.315
                                      - 12-31-2012
variance Type:                                 opg
==============================================================================
              coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
.L1           0.7242      0.030     24.246      0.000       0.666       0.783
.L1          -0.7077     14.113     -0.050      0.960     -28.370      26.954
.L2          -0.2923      4.130     -0.071      0.944      -8.388       7.803
.S.L365      -0.4503      0.040    -11.375      0.000      -0.528      -0.373
.S.L365      -0.2885      0.075     -3.866      0.000      -0.435      -0.142
gma2          7.8032    110.167      0.071      0.944    -208.119     223.726
==========================================================================================
ung-Box (L1) (Q):                   0.09   Jarque-Bera (JB):                 3.81
ob(Q):                              0.76   Prob(JB):                         0.15
teroskedasticity (H):               0.88   Skew:                            -0.02
ob(H) (two-sided):                  0.31   Kurtosis:                         3.35
==========================================================================================
```

Figure 26

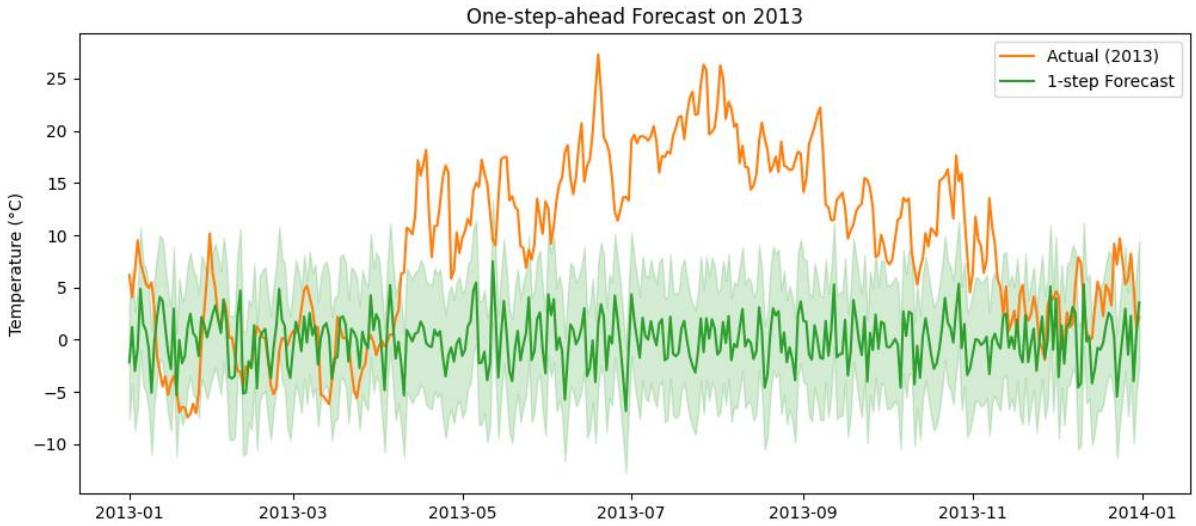


Figure 27

that green ribbon is actually the **double-differenced** series that your SARIMAX fit is modeling (first a yearly difference, then a regular 1-day difference), not the actual temperature.

which quite reasonably hovers around zero. That's why your one-step forecast looks nearly flat with ±5 °C swings, instead of matching the 10–20 °C scale of the orange line.

To fix it: invert the differencing

1. **Recover the seasonally-differenced** forecast by cumulatively summing over the seasonal lag
2. **Recover the original series** by cumulatively summing that result over the 1-day lag

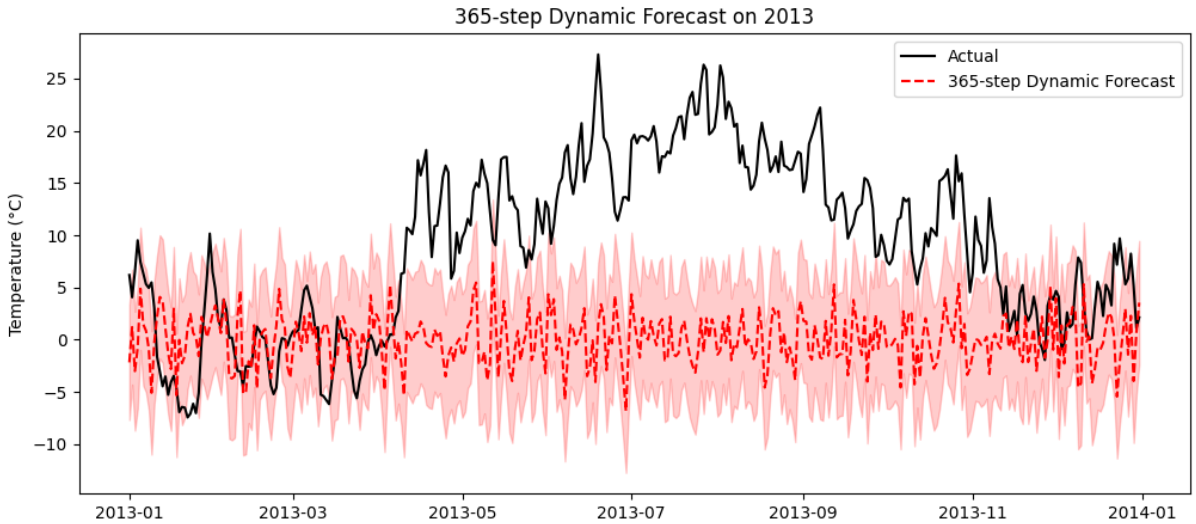


Figure 28

Here while trying to make the data revert to its original form the RAM crashed, it similarly happened to the above models as well hence I used the **ARIMA(0,1,2) on Seasonally-Adjusted Data + Seasonal Recomposition Forecast vs Actual- overall this model is a SARIMA model built with ARIMA+Seasonality addition.**

Model 012

SARIMAX Results

Dep. Variable:	y	No. Observations:	1461
Model:	ARIMA(0, 1, 2)	Log Likelihood	-3297.355
Date:	Sun, 04 May 2025	AIC	6600.710
Time:	15:18:37	BIC	6616.569
Sample:	01-01-2009	HQIC	6606.626
	- 12-31-2012		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	0.0266	0.025	1.077	0.282	-0.022	0.075
ma.L2	-0.2269	0.023	-9.797	0.000	-0.272	-0.181
sigma2	5.3600	0.179	29.966	0.000	5.009	5.711

Ljung-Box (L1) (Q):	2.38	Jarque-Bera (JB):	22.63
Prob(Q):	0.12	Prob(JB):	0.00
Heteroskedasticity (H):	0.98	Skew:	-0.14
Prob(H) (two-sided):	0.78	Kurtosis:	3.55

Figure 29

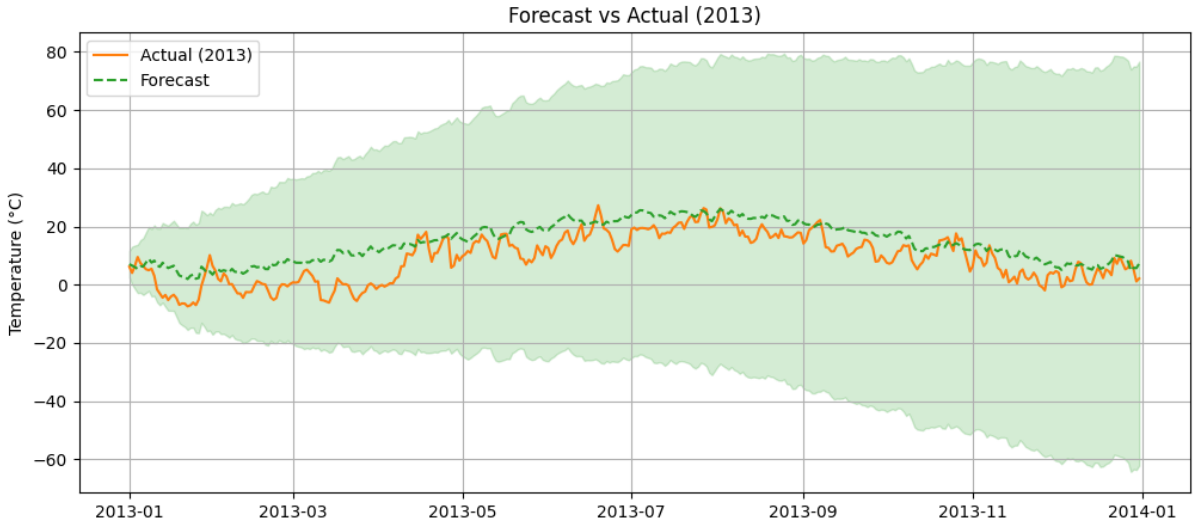
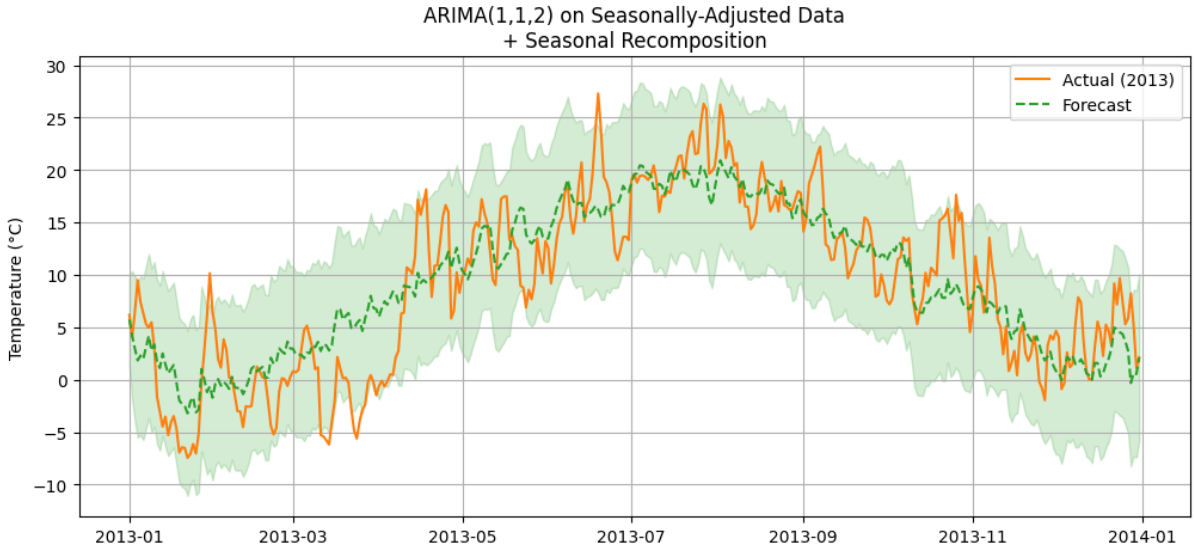


Figure 30

Model 112:

```
==============================================================================
Dep. Variable:                      y   No. Observations:                 1461
Model:                 ARIMA(1, 1, 2)   Log Likelihood               -3233.952
Date:                Sun, 04 May 2025   AIC                           6475.904
Time:                        15:31:21   BIC                           6497.049
Sample:                    01-01-2009   HQIC                          6483.792
                         - 12-31-2012
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.7445      0.019     39.452      0.000       0.708       0.781
ma.L1         -0.7450      0.027    -27.147      0.000      -0.799      -0.691
ma.L2         -0.2516      0.027     -9.295      0.000      -0.305      -0.199
sigma2         4.9034      0.168     29.205      0.000       4.574       5.233
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                19.51
Prob(Q):                              0.99   Prob(JB):                         0.00
Heteroskedasticity (H):               0.92   Skew:                            -0.19
Prob(H) (two-sided):                  0.35   Kurtosis:                         3.43
===================================================================================
```

Figure 31




```
--- Q-Test (lags=50, df=47, alpha=0.05) ---
Q-statistic: 66016.8018
Critical value: 64.0011
Result: ❌ Residuals show autocorrelation
```

Figure 32

Although the model looks great it didn't pass q-test and variance is upto 4 degrees Celsius. I believe since we are breaking the model into two different components like arima and

Seasonality. The q-test may not be the best characteristic for this. Either way this is not the best model. So, Let's move to Box Jenkins now..

BOX JENKINS:

	nb	nf	nc	nd	Q_stat	Q_crit	Q_pass	S_stat	S_crit	S_pass
0	1	1	0	0	143.449470	65.170769	False	21.057283	30.143527	True
1	1	1	0	1	138.611276	64.001112	False	20.187493	30.143527	True
2	1	1	1	0	138.611276	64.001112	False	20.187493	30.143527	True
3	1	1	1	1	138.611276	62.829620	False	20.187493	30.143527	True
4	1	2	0	0	109.112704	64.001112	False	16.472130	28.869299	True
5	1	2	0	1	114.149448	62.829620	False	15.683024	28.869299	True
6	1	2	1	0	114.149448	62.829620	False	15.683024	28.869299	True
7	1	2	1	1	114.149448	61.656233	False	15.683024	28.869299	True
8	2	1	0	0	143.011171	64.001112	False	19.926954	30.143527	True
9	2	1	0	1	137.706429	62.829620	False	20.564540	30.143527	True
10	2	1	1	0	137.706429	62.829620	False	20.564540	30.143527	True
11	2	1	1	1	137.706429	61.656233	False	20.564540	30.143527	True
12	2	2	0	0	109.126378	62.829620	False	15.558826	28.869299	True
13	2	2	0	1	113.870886	61.656233	False	16.051770	28.869299	True
14	2	2	1	0	113.870886	61.656233	False	16.051770	28.869299	True
15	2	2	1	1	113.870886	60.480887	False	16.051770	28.869299	True

Figure 33

few models I have tried and I have tried pushing the order values upto 10 but could find a model that has passed both tests. the best I have got is the following one..

Q-test: Q=143.0, crit=64.0, df=47 -> ❌ autocorrelation

S-test: S=19.9, crit=30.1, df=19 -> ✅ G(q) accurate

1-Step Metrics:	365-Step Metrics:
MAE = 0.279	MAE = 0.293
RMSE = 0.360	RMSE = 0.373

MAPE = 3.12%

Var = 0.1303

$Corr^2$ = 0.998

MAPE = 101.60%

Var = 0.1396

$Corr^2$ = 0.998

These metrics are for the normalised data...

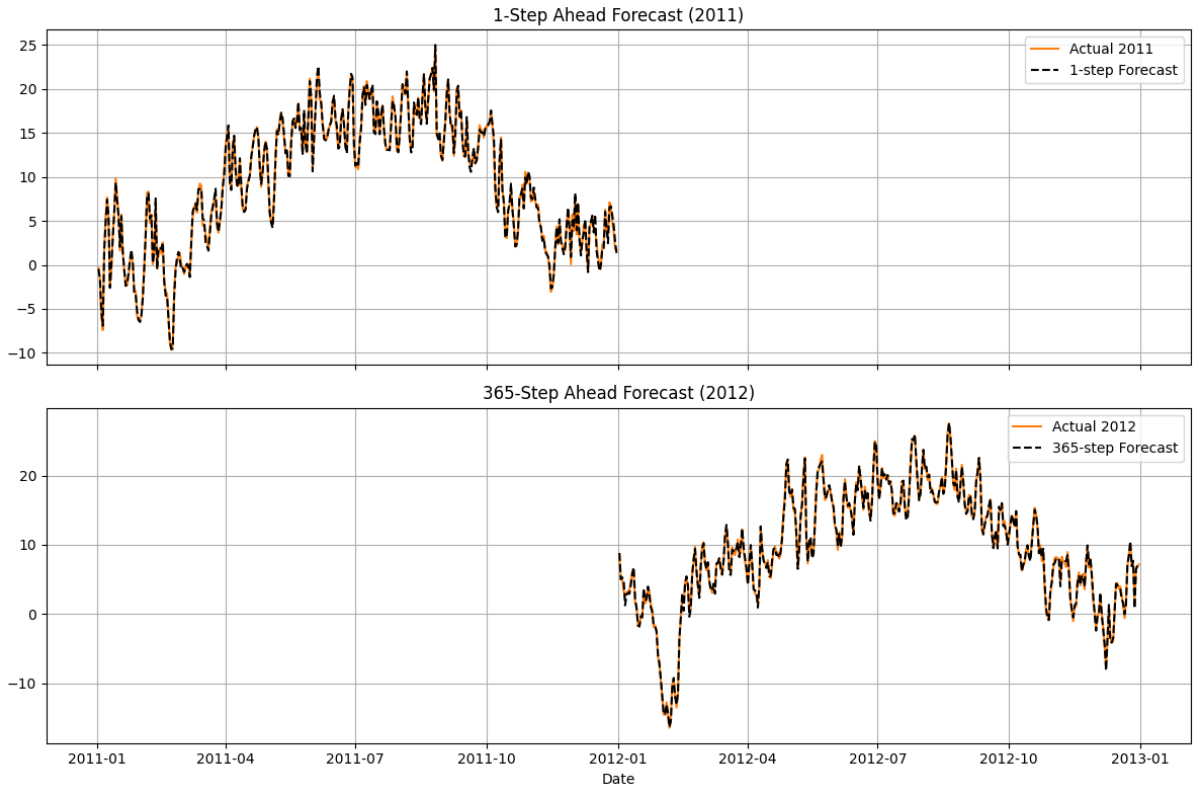


Figure 34

I have tried from orders 1 to 10 as professor said but the model that passed both tests is not there, I believe its because the data is real world data and I believe making real world data into white noise is not feasible in my case.. however the model itself is the best I could make as it is pretty accurate

Note:

I have a data from 2009 to 2016 or so.. I have used two year to train, 1 year to test and 1 step and h-step on another year - 2012.

Final model selection:

From the observed models the plots of different models we can clearly say that the **Box-jenkins** is the most accurate.. as we compare the RMSE and other models only had q-test and they failed that one test. I have added explanation of best model for each case like arma, arima, sarima rather than going in deep with every order of every model. In box Jenkins the S-test

confirms the model we built is in right direction and we couldn't get it pass q-test because this real- data never became white even though I have moved to the highest order of 10.

H-step forecast: 1-year In this case..

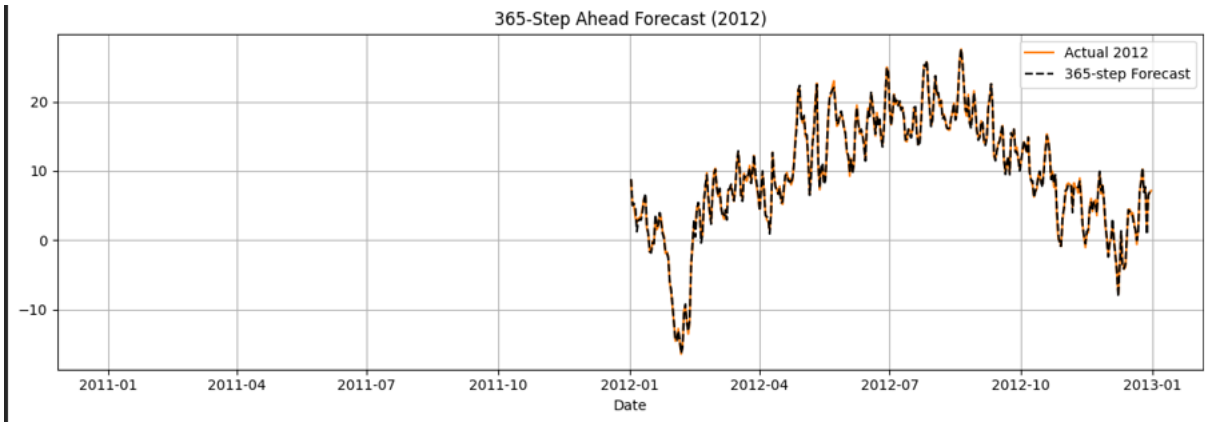


Figure 35

Taken from the previous prediction at Box Jenkins..

Summary and conclusion:

I have coursed through many models like base models, ARMA, SARIMA,ARIMA, BOX-JENKINS. My observation was that the sarima models take a lot of time as it depends on the characteristic matrices. But they are pretty good at working with seasonal data. How ever all the above models except Linear regression do not depend on other variables. That can make them little uncharacteristic, I mean one can only do so much using a single column. Finally, This issue was solved by Box-jenkins as it takes one column the model mostly depends on with highest corelation and least VIF, I worked with it and got a pretty good result... Some other estimated good models would be LSTM and RNN