Forecasting Monthly Average Temperature in Rome

SARIMA and SARIMAX Models with Climate Covariates

Antonella Convertini 45853A
Università degli studi di Milano
Data Science for Economics

1 Introduction

This project studies the problem of forecasting monthly average temperature in Rome using historical weather data from the ROMA CIAMPINO meteorological station (station ID IT000016239). The target variable is the monthly average temperature (TAVG, in degrees Celsius) observed from January 1951 to September 2022. The main forecasting objective is twofold:

1. Produce and evaluate one-step-ahead forecasts for TAVG over a hold-out test sample.
2. Use the best-performing model to generate a 10-year ahead forecast scenario for monthly temperatures.

The choice of this forecasting problem is motivated by several considerations:

- **Macroeconomic and social relevance.** Temperature dynamics are closely related to energy demand, health risks (e.g. heat waves), and agricultural productivity. Reliable temperature forecasts inform planning in energy, tourism, agriculture, and public health.
- **Climate change and local adaptation.** Rome is a large Mediterranean city exposed to rising temperatures and more frequent heat extremes. Understanding the temporal dynamics of local temperature, and the contribution of seasonal patterns and climate drivers, is relevant for climate adaptation and urban policy.
- **Methodological interest.** From a time-series perspective, monthly temperature exhibits strong seasonality and persistence. The data provide a useful laboratory to apply and compare benchmark models (mean, random walk, seasonal naive), ARIMA/SARIMA models, and models with exogenous regressors (SARIMAX), using out-of-sample forecast performance (RMSFE, MAFE) as evaluation criteria.

The analysis follows the schema: data exploration and transformations, model specification guided by theory and information criteria, residual diagnostics, pseudo out-of-sample evaluation on a test sample, and long-horizon scenario forecasting.

2 Data

2.1 Dataset and variables

The dataset contains monthly observations for the ROMA CIAMPINO station from January 1951 to September 2022 (861 observations). The key variables used in this project are:

- $TAVG_t$: monthly average temperature (°C), target variable.
- $PRCP_t$: monthly total precipitation.
- $DX90_t$: number of days in the month with maximum temperature above a high threshold (heat days).
- $ENSO\_index_t$: monthly ENSO (Niño 3.4) anomaly index.
- Calendar and derived features:
  - $month_t \in \{1, \dots, 12\}$: month of the year.
  - Cyclical monthly dummies encoded as

$$month\_sin_t = \sin\left(\frac{2\pi\, month_t}{12}\right), \qquad month\_cos_t = \cos\left(\frac{2\pi\, month_t}{12}\right).$$

- Lagged exogenous variables:
  - $PRCP\_lag1_t = PRCP_{t-1}$,
  - $DX90\_lag1_t = DX90_{t-1}$,
  - $ENSO\_lag1_t = ENSO\_index_{t-1}$.

The sample is split chronologically as follows:

- Training sample: January 1951 – May 2008 (689 observations).
- Test sample: June 2008 – September 2022 (172 observations).

The training sample is used for model estimation and selection, while the test sample is reserved for genuine out-of-sample forecast evaluation.

2.2 Transformations and feature construction

The target series $TAVG_t$ is used in levels; no logarithmic or per-capita transformation is necessary because it is already in degrees Celsius and takes strictly positive, moderate values. Instead, the modelling strategy focuses on handling strong seasonality and persistence through:

1. **Seasonal differencing:** in the SARIMA/SARIMAX models, a seasonal difference of order 1 at lag 12, $(1 - L^{12})$, is applied to remove annual seasonality.
2. **Non-seasonal differencing:** the best-fitting ARIMA/SARIMA specification includes a non-seasonal difference, $(1 - L)$, to reduce persistence and stabilise the mean.
3. **Exogenous regressors:** several SARIMAX specifications are estimated using different subsets of exogenous regressors:
   - Local weather: $PRCP\_lag1_t$, $DX90\_lag1_t$,
   - Seasonal cycle: $month\_sin_t$, $month\_cos_t$,
   - Climate index: $ENSO\_lag1_t$.

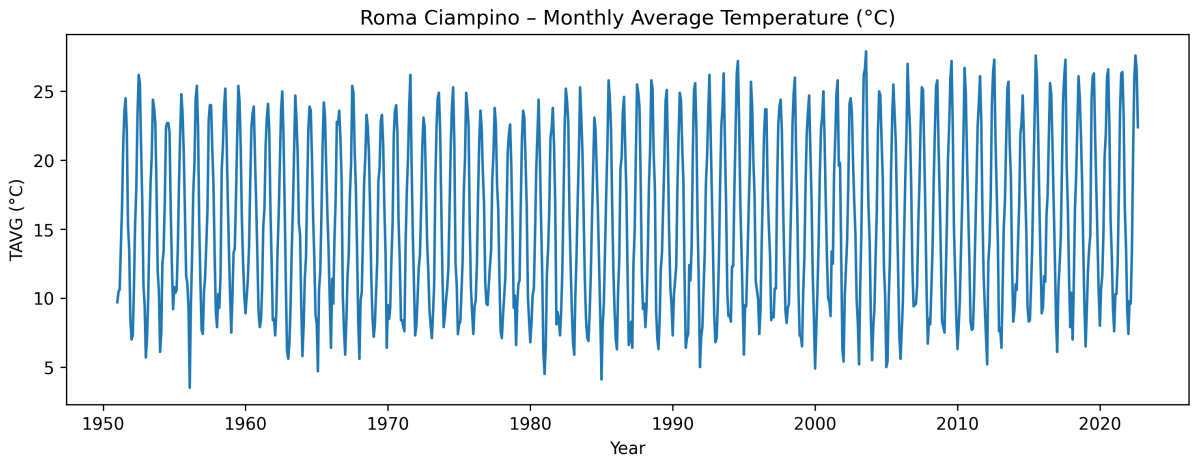

Figure 1: Monthly average temperature (TAVG) in Rome, January 1951 – September 2022. The series displays a very regular annual cycle, strong persistence and a gradual warming over time, with recent decades noticeably warmer than the early part of the sample.

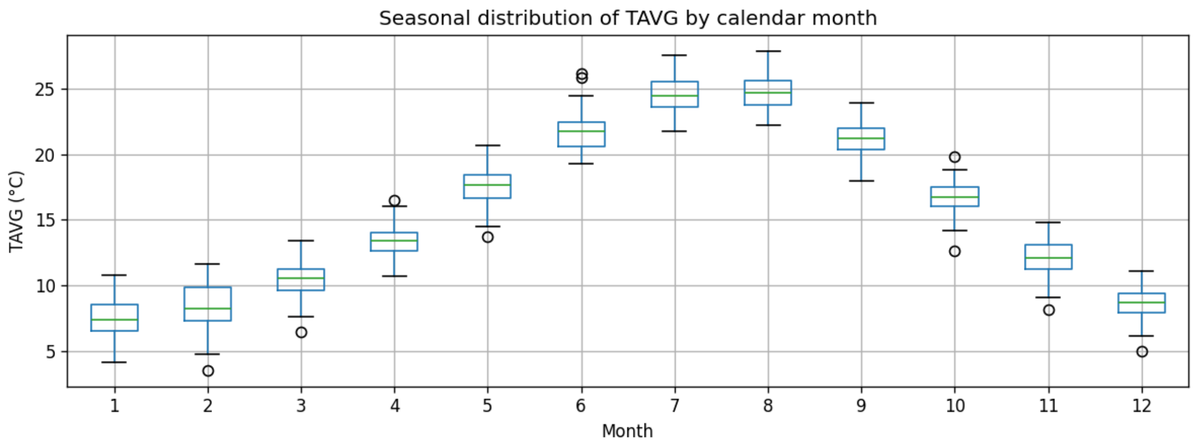

Figure 2: Seasonal distribution of monthly average temperature (TAVG) by calendar month. Winter months cluster around 7–9°C, while July–August are close to 25°C, with an amplitude of about 17°C between winter lows and summer highs.

2.3 Descriptive statistics and time-series features

Over the full sample, TAVG has the following summary statistics:

$$\text{mean} = 15.58^\circ\text{C}, \quad sd = 6.26, \quad \text{min} = 3.5, \quad \text{max} = 27.9.$$

The mean monthly pattern shows a typical Mediterranean climate:

- Winter months (January–February) average around 7.5°C to 8.5°C.
- Peak summer months (July–August) average around 24.6°C–24.7°C.
- The amplitude between winter lows and summer highs is approximately 17°C.

To visualise the seasonal distribution more explicitly, Figure 2 shows boxplots of TAVG by calendar month.

2.4 Stationarity analysis

To assess stationarity, an Augmented Dickey–Fuller (ADF) test is applied to the TAVG series. The test regression is of the standard form (with intercept):

$$\Delta y_t = \alpha + \gamma y_{t-1} + \sum_{j=1}^{p} \phi_j \Delta y_{t-j} + \varepsilon_t, \tag{1}$$

where $y_t = TAVG_t$. The null hypothesis H_0 is $\gamma = 0$ (unit root), while the alternative H_1 is $\gamma < 0$ (stationarity).

The estimated ADF statistic is approximately -3.82 with a p-value around 0.003, which leads to rejection of the unit-root null at conventional significance levels. However, given the strong seasonality and slow-moving trend visible in the plot, the modelling strategy still considers SARIMA specifications with both non-seasonal and seasonal differencing, and chooses the best specification based on information criteria and forecast performance.

The autocorrelation functions provide additional information. Figure 3 shows the ACF of TAVG in levels, first difference, and seasonal difference.

The ACF of the seasonal difference is much flatter: apart from a negative spike at lag 12, most autocorrelations lie inside the approximate 95% confidence band. This supports the use of a SARIMA model with one non-seasonal and one seasonal difference.

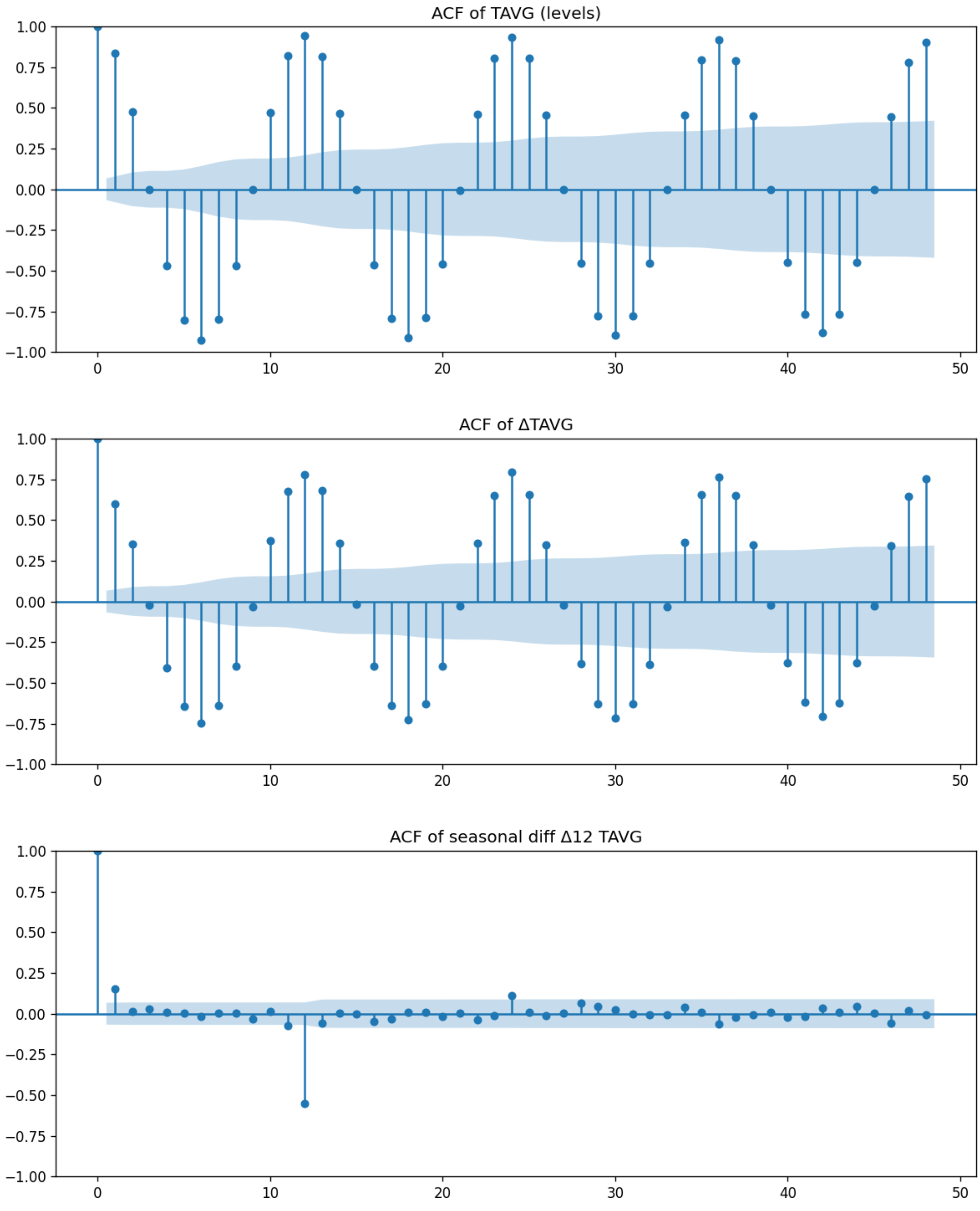

Figure 3: Autocorrelation functions (ACF) of TAVG: (top) levels, (middle) first difference $\Delta TAVG_t$, (bottom) seasonal difference $(1 - L^{12})TAVG_t$. Levels show very strong persistence and clear annual peaks at lags 12, 24, 36, ... After first differencing, autocorrelations decay faster but seasonal spikes remain. The seasonal difference largely removes the deterministic annual component, with most autocorrelations inside the 95% band, suggesting that a combination of non-seasonal and seasonal differencing is appropriate.

3 Methods and Models

3.1 Benchmark models

Three simple benchmark models are considered.

3.1.1 Constant mean model

The constant mean model assumes:

$$y_t = \mu + \varepsilon_t, \qquad \varepsilon_t \sim \text{i.i.d. } (0, \sigma^2), \tag{2}$$

where $y_t = TAVG_t$. The parameter μ is estimated by the sample mean over the training sample. The h-step-ahead forecast is simply

$$\hat{y}^{\text{Mean}}_{T+h|T} = \hat{\mu}, \qquad h \geq 1. \tag{3}$$

3.1.2 Random walk with drift

The random walk with drift is specified as:

$$y_t = \mu + y_{t-1} + \varepsilon_t. \tag{4}$$

Estimated by OLS, the h-step-ahead forecast is

$$\hat{y}^{\text{RW}}_{T+h|T} = \hat{\mu}h + y_T. \tag{5}$$

3.1.3 Seasonal naive model

For monthly data with seasonal period $s = 12$, the seasonal naive model sets

$$\hat{y}^{\text{SNAIVE}}_{t+h|t} = y_{t+h-s}, \tag{6}$$

so that at horizon h, the forecast repeats the observation from the same month one year earlier.

3.2 SARIMA specification

Inspection of the ACF/PACF and the strong annual cycle suggest that a natural starting point for TAVG is a seasonal ARIMA model. Based on inspection of ACF/PACF and information criteria, a range of SARIMA$(p, d, q) \times (P, D, Q)_{12}$ models is estimated. Table 1 reports the top five specifications ranked by AIC.

The best specification according to both AIC and BIC is:

$$\text{SARIMA}(0, 1, 2) \times (0, 1, 1)_{12}.$$

In backshift notation, this model can be written as

$$(1 - L)(1 - L^{12})y_t = (1 + \theta_1 L + \theta_2 L^2)(1 + \Theta_1 L^{12})\varepsilon_t, \tag{7}$$

where ε_t is a white-noise error.

Table 1: Top SARIMA specifications by AIC (seasonal period 12).

p	d	q	P	D	Q	AIC	BIC
0	1	2	0	1	1	2221.991	2244.46
2	1	1	0	1	1	2223.848	2250.82
1	1	1	0	1	1	2225.024	2247.50
2	1	2	0	1	1	2225.601	2257.06
0	1	2	1	1	1	2228.891	2255.85

3.3 SARIMAX models with exogenous regressors

To exploit additional information beyond TAVG dynamics, several SARIMAX specifications are considered. The general SARIMAX model is:

$$(1-L)(1-L^{12})y_t = \beta_0 + \beta^\top x_t + (1+\theta_1 L + \theta_2 L^2)(1+\Theta_1 L^{12})\varepsilon_t, \tag{8}$$

where x_t collects exogenous regressors.

The following exogenous sets are considered:

- **SARIMAX PRCP:** $x_t = (PRCP\_lag1_t)$.
- **SARIMAX MonthCyc:** $x_t = (month\_sin_t, month\_cos_t)$.
- **SARIMAX FourExog:** $x_t = (PRCP\_lag1_t, DX90\_lag1_t, ENSO\_index_t, ENSO\_lag1_t)$.
- **SARIMAX ENSO lag1:** $x_t = (ENSO\_lag1_t)$.

These models are designed to test whether local precipitation, heat extremes, seasonal cycle in deterministic form, and global ENSO conditions improve forecast accuracy over purely univariate SARIMA.

3.4 Estimation strategy and model selection

All models are estimated on the training sample (1951–2008). The estimation strategy follows the lecture guidelines:

1. **Train–test split.** The series is split into a training sample for estimation and a test sample for out-of-sample evaluation. The train–test split is strictly chronological to respect the time-series structure.
2. **SARIMA order selection.** A grid of SARIMA$(p,d,q)\times(P,D,Q)$ models with p,q,P,Q in a reasonable range is estimated. For each specification, AIC and BIC are computed. The model with the lowest BIC (and very similar AIC) is selected as the baseline SARIMA, as shown in Table 1.
3. **SARIMAX selection.** For the chosen differencing and ARMA orders, several SARIMAX models with different exogenous sets are estimated. Their out-of-sample performance on the test sample is compared using RMSFE and MAFE.

For test evaluation, each model is estimated once on the training sample and then used to generate one-step-ahead forecasts over the test period. Forecasts are genuine one- step-ahead predictions, using only information available up to time t when forecasting y_{t+1}.

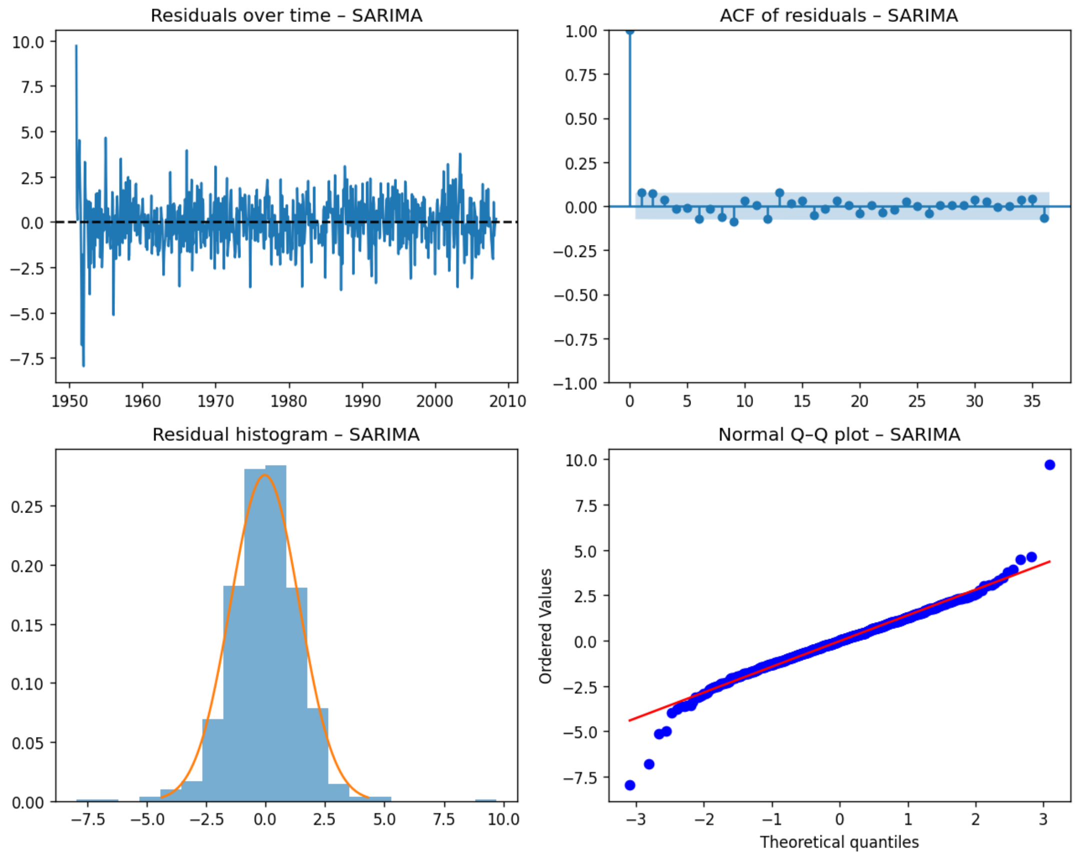

Figure 4: Residual diagnostics for the selected SARIMA$(0, 1, 2) \times (0, 1, 1)_{12}$ model. Top-left: residuals over time, fluctuating around zero with roughly constant variance. Top-right: residual ACF, with no large remaining autocorrelations or seasonal spikes. Bottom-left: residual histogram with overlaid normal density, showing an approximately symmetric, bell-shaped distribution. Bottom-right: normal Q–Q plot, with residual quantiles close to the 45° line except for moderate tail deviations. Overall, these diagnostics support the adequacy of the SARIMA specification.

3.5 Diagnostics

Model diagnostics follow the standard checks:

- **Residual autocorrelation.** For the selected SARIMA and SARIMAX models, residual ACF and PACF are inspected. The correlograms do not show strong remaining autocorrelation or clear seasonal patterns, suggesting that the ARIMA structure adequately captures the dynamics.
- **Residual distribution.** Histograms and Q–Q plots of residuals indicate approximately symmetric distributions, with some deviations in the tails, which is typical for weather data.
- **Stability.** Parameter estimates are stable across sub-periods of the training sample, and the models' forecast behaviour is consistent when extended towards the end of the sample.

Figure 4 shows the residual diagnostics for the selected SARIMA model. Similar diagnostics for the SARIMAX models are reported in the Appendix.

Overall, diagnostics support the adequacy of the SARIMA$(0, 1, 2) \times (0, 1, 1)_{12}$ specification and its SARIMAX extensions as reasonable approximations for TAVG dynamics.

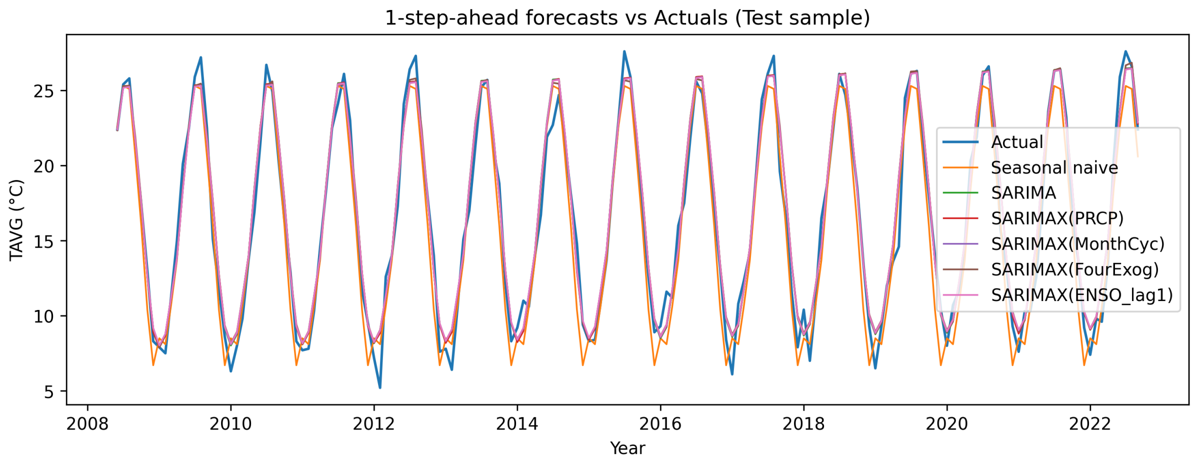

Figure 5: One-step-ahead forecasts of TAVG vs actuals on the test sample (June 2008 – September 2022). The blue line shows realised monthly average temperature. Seasonal naive, SARIMA, SARIMAX(PRCP), SARIMAX(MonthCyc), SARIMAX(FourExog) and SARIMAX(ENSO_lag1) forecasts track the annual cycle with different precision. Benchmark models (mean and random walk with drift, not shown here for clarity) perform substantially worse. SARIMA and SARIMAX forecasts follow both the seasonal cycle and inter-annual variation very closely, staying near the realised series.

4 Results

4.1 Forecasts vs actuals on the test sample

Figure 5 summarises the core forecasting result of the paper. It plots the observed TAVG in the test period (June 2008 – September 2022) together with one-step-ahead forecasts from all models: constant mean, random walk with drift, seasonal naive, SARIMA and the SARIMAX specifications.

Visual inspection shows:

- The seasonal naive model already captures the annual cycle but tends to underreact to unusually warm or cool months.
- SARIMA and SARIMAX models track both the seasonal pattern and year-to-year fluctuations more accurately, with smaller forecast errors around peaks and troughs.

4.2 Forecast evaluation

Forecast accuracy is measured by Root Mean Squared Forecast Error (RMSFE) and Mean Absolute Forecast Error (MAFE). For a test sample of size T_{test} these are

$$\text{RMSFE} = \sqrt{\frac{1}{T_{\text{test}}} \sum_{t=1}^{T_{\text{test}}} (y_t - \hat{y}_t)^2}, \qquad \text{MAFE} = \frac{1}{T_{\text{test}}} \sum_{t=1}^{T_{\text{test}}} |y_t - \hat{y}_t|. \tag{9}$$

Table 2 summarises one-step-ahead forecast performance on the test sample.

The main findings are:

- All ARIMA-based models dramatically outperform the simple benchmarks. The seasonal naive model already reduces RMSFE from about 6.62–7.24 (Mean, RW with drift) to 1.65, reflecting the importance of seasonality.

Table 2: One-step-ahead forecast performance on the test sample (June 2008 – September 2022).

Model	RMSFE	MAFE
Mean	6.62	5.76
RW with drift	7.24	6.31
Seasonal naive	1.65	1.34
SARIMA	1.25	0.97
SARIMAX PRCP	1.25	0.96
SARIMAX MonthCyc	1.26	0.98
SARIMAX FourExog	1.27	0.98
SARIMAX ENSO lag1	1.26	0.98

- The SARIMA$(0, 1, 2) \times (0, 1, 1)_{12}$ model further reduces RMSFE to about 1.25 and MAFE to about 0.97. This indicates that exploiting within-year dynamics and ARMA structure substantially improves forecast accuracy over purely seasonal extrapolation.
- Among SARIMAX models, SARIMAX PRCP produces the lowest RMSFE (1.2490) and MAFE (0.9643), but the improvement over the pure SARIMA model is extremely small. Other exogenous specifications (MonthCyc, FourExog, ENSO lag1) do not improve performance and in some cases slightly worsen it.

Overall, the best-performing model in out-of-sample one-step-ahead forecasts is SARIMAX with lagged precipitation (SARIMAX PRCP), but the gains relative to the simpler SARIMA model are marginal. This suggests that most of the predictive power is contained in the autoregressive and seasonal structure of TAVG itself, with local precipitation and ENSO exerting only a limited incremental effect at the monthly horizon.

4.3 Interpretation of results

The results can be interpreted as follows:

- **Strong predictability via seasonality.** The seasonal naive model already performs quite well, indicating that the annual cycle is a dominant driver of monthly temperature. Knowing last year's temperature for the same month provides a strong baseline prediction.
- **Value of ARIMA dynamics.** The SARIMA model improves substantially upon the seasonal naive benchmark by capturing additional short-run dynamics within the year and between seasons.
- **Limited incremental role of exogenous variables.** Exogenous regressors such as lagged precipitation, ENSO, and deterministic monthly dummies have only a small effect on one-step-ahead predictive accuracy when added on top of the SARIMA structure.
- **Model robustness.** Diagnostic checks and the comparison of SARIMA and SARIMAX models indicate that the main conclusions are robust to reasonable variations in model specification.

4.4 Ten-year ahead forecast scenario

Using the selected SARIMA and SARIMAX models, a 10-year ahead forecast scenario is generated from October 2022 to September 2032. Figure 6 shows the projected paths together with the full historical series.

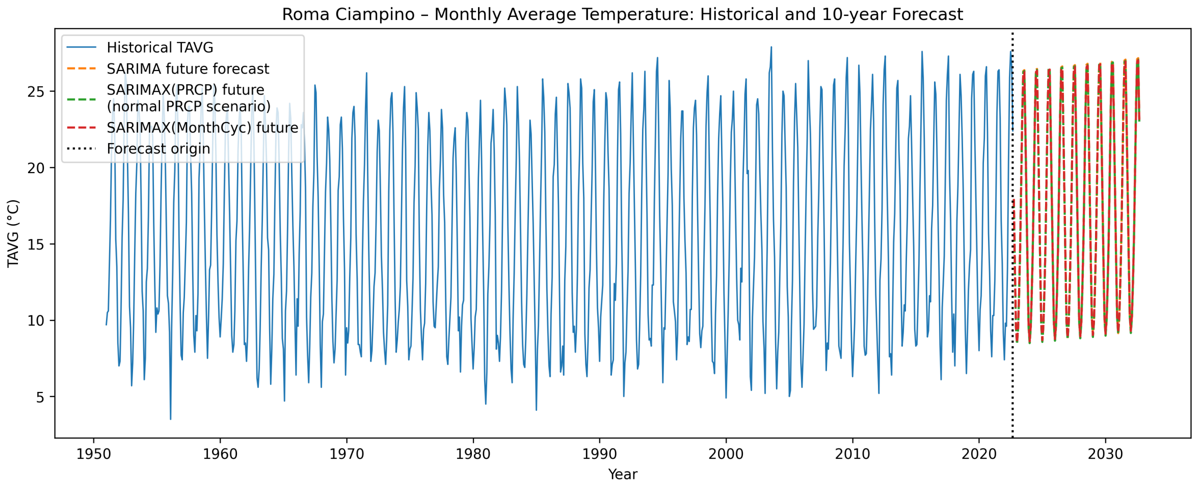

Figure 6: Roma Ciampino – Monthly average temperature: historical series and 10-year forecast. The solid blue line shows historical TAVG. Dashed lines report 10-year-ahead forecasts from SARIMA, SARIMAX(PRCP) under a "normal precipitation" scenario, and SARIMAX(MonthCyc). The vertical dotted line marks the forecast origin (end of estimation sample). All models project a persistent and regular seasonal cycle with summer peaks around 26–27°C and winter lows around 8–10°C; differences between models are modest.

Key features of the forecast scenario:

- All models predict a very persistent and regular seasonal cycle, with winter lows around 8–10°C and summer peaks around 26–27°C.
- Average forecasted TAVG over the 10-year horizon is around 17.2°C across models, broadly consistent with recent historical averages.
- Differences between SARIMA and SARIMAX forecasts are small, reflecting the limited incremental role of exogenous regressors in the selected specifications.

These forecasts should be interpreted as *statistical extrapolations* under the assumption that the historical dynamics captured by this framework continue to hold. They do not explicitly incorporate future climate-change scenarios or structural breaks in the underlying process.

5 Conclusion

This project applied SARIMA and SARIMAX models to forecast monthly average temperature in Rome, using a long historical sample (1951–2022). The main findings can be summarised as follows:

- Simple benchmark models (constant mean, random walk with drift) perform poorly for this strongly seasonal series. A seasonal naive model that repeats last year's monthly value already achieves a large reduction in RMSFE and MAFE.
- A SARIMA$(0,1,2)\times(0,1,1)_{12}$ model, selected using information criteria and supported by residual diagnostics, yields substantial further gains in forecast accuracy, demonstrating the importance of capturing both seasonal and non-seasonal dynamics.

- Adding exogenous regressors (lagged precipitation, deterministic monthly cycle, ENSO) through SARIMAX specifications yields only marginal improvements over the univariate SARIMA model, with SARIMAX PRCP performing best but only slightly better than SARIMA. This indicates that most of the predictive information at the monthly horizon is intrinsic to the temperature series itself.
- The resulting 10-year ahead forecasts display a stable seasonal pattern with modest warming relative to past decades, but they should be interpreted cautiously as purely statistical projections, not as physically based climate-change scenarios.

Limitations and possible extensions. Several extensions are possible:

- Incorporating more detailed climate covariates (e.g. radiative forcing, large-scale atmospheric indices) and testing non-linear or regime-switching models.
- Explicitly modelling conditional heteroskedasticity (ARCH/GARCH-type models) to improve density forecasts and prediction intervals.

Despite these limitations, the project shows that relatively simple SARIMA/SARIMAX models can deliver accurate short-run forecasts of local monthly temperatures and provide a useful baseline for more advanced climate and policy analysis.

Appendix: Additional diagnostic plots

This appendix reports residual diagnostics for the SARIMAX specifications discussed in the main text. Each figure shows residuals over time (top left), the residual autocorrelation function (top right), a histogram with a normal density (bottom left), and a normal Q–Q plot (bottom right).

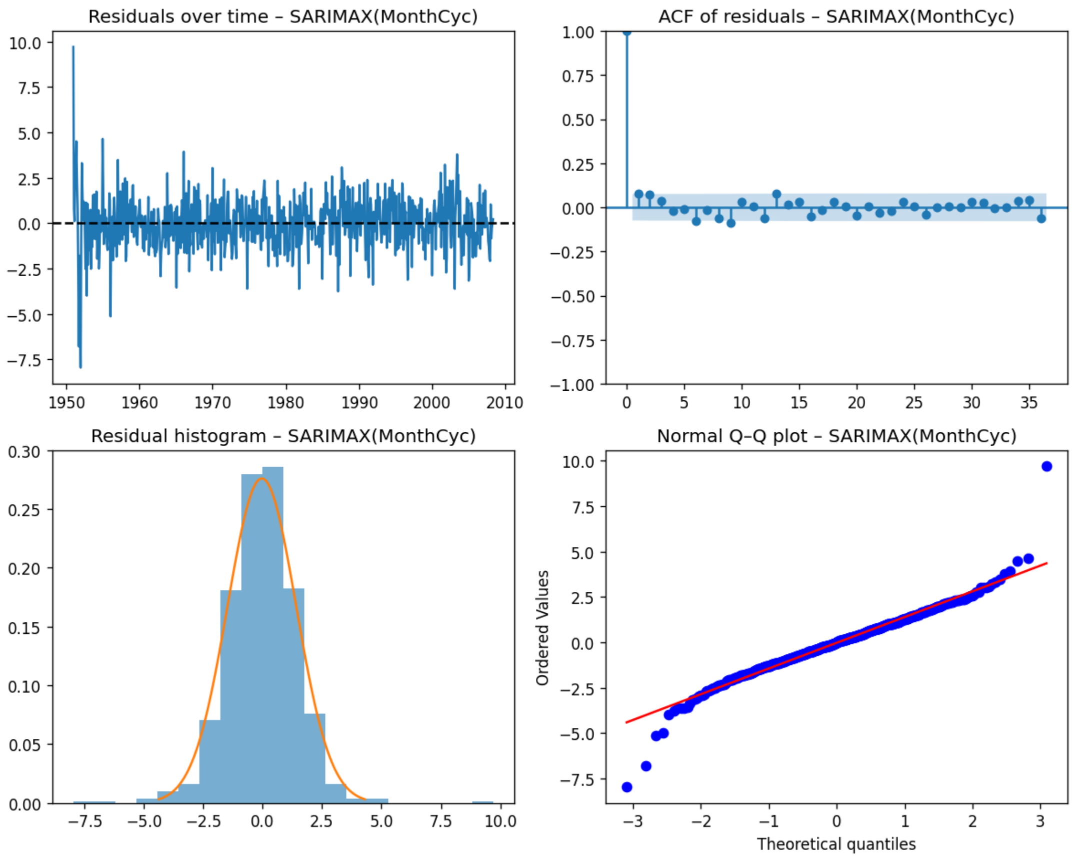

Figure 7: Residual diagnostics for SARIMAX model with cyclical monthly dummies (MonthCyc).

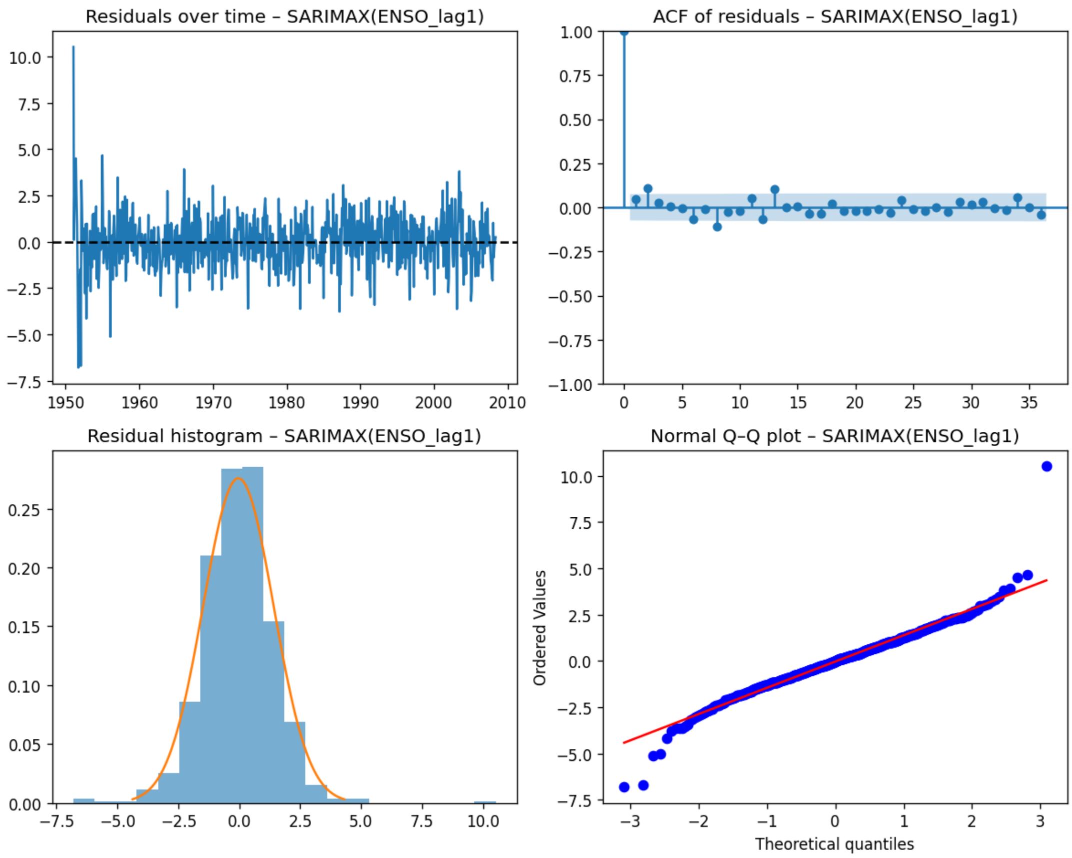

Figure 8: Residual diagnostics for SARIMAX model with ENSO regressor (ENSO_lag1).

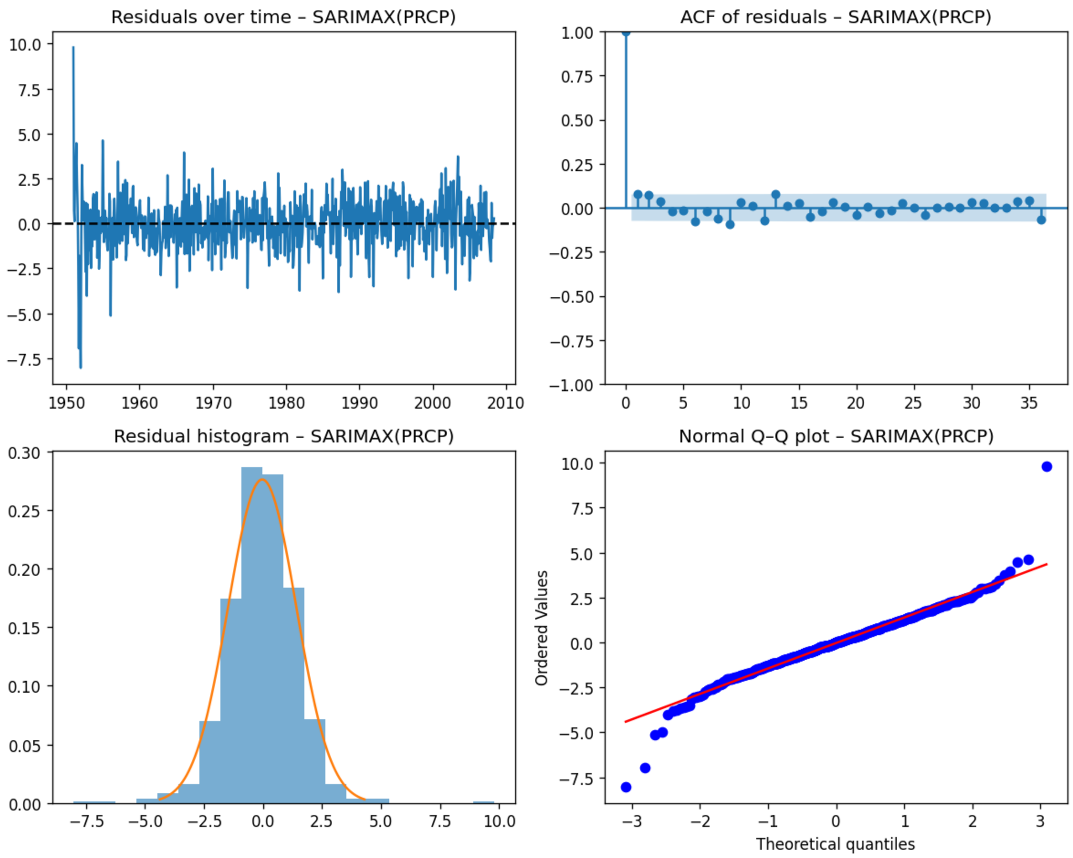

Figure 9: Residual diagnostics for SARIMAX model with lagged precipitation (PRCP_lag1).

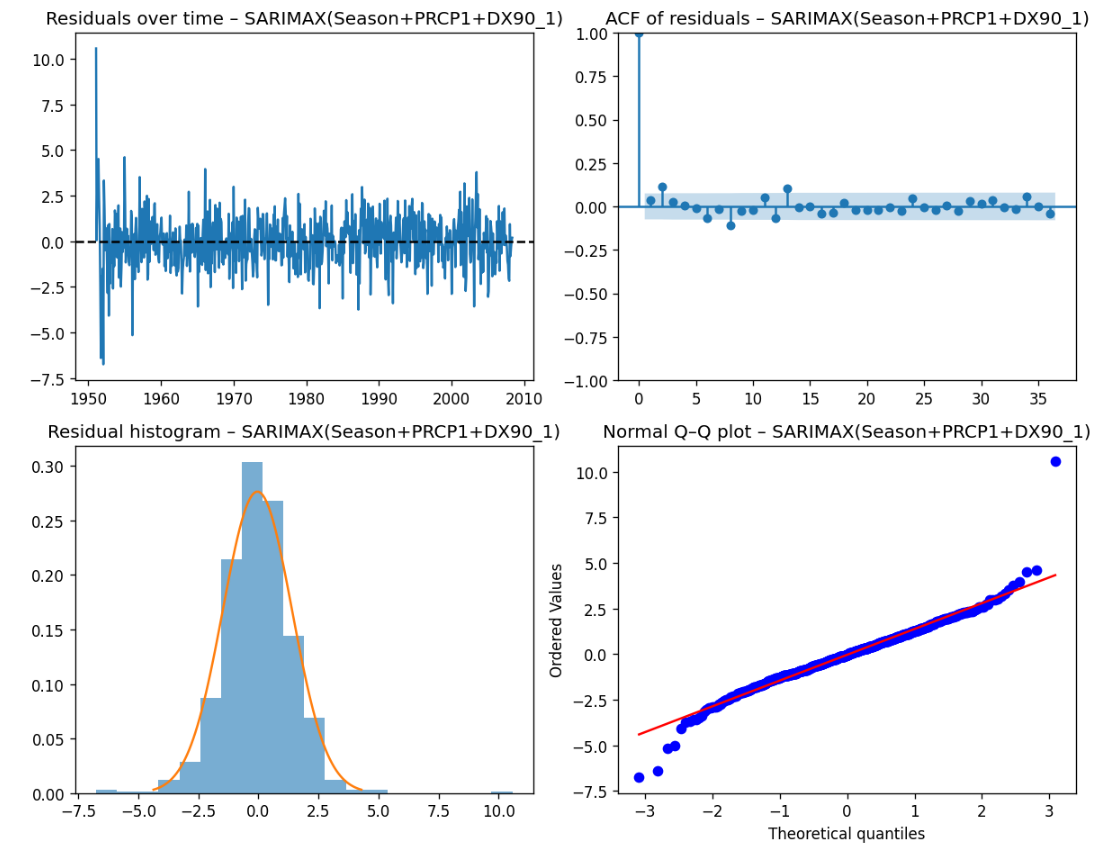

Figure 10: Residual diagnostics for SARIMAX model with a richer set of exogenous regressors.