Time-Series Analysis Strategy for NASA GISTEMP Data

Stefano Graziosi, Università Bocconi   OpenAI Deep Research

May 5, 2025

Introduction

This project will analyze the NASA GISTEMP monthly global surface temperature anomaly series (1880–2023) to explore climate patterns and make baseline forecasts. The data are temperature *anomalies* in $^\circ\mathrm{C}$ relative to the 1951–1980 average, providing a reliable long-term record of climate variability. We will address five key questions (Q1–Q5) using *only* methods covered in the TS_20236 syllabus—including classical decomposition, moving averages, exponential smoothing (Holt–Winters), AR/MA/ARMA (ARIMA) modeling via Box–Jenkins methodology, random walk diagnostics, Hidden Markov Models (HMM), and Dynamic Linear Models (DLM) with Kalman filtering. Throughout, we distinguish *observational signal extraction* (identifying trends/seasonality in historical data) from *predictive modeling* (forecasting future values). Each question is tackled with concrete analytical steps, and where multiple methods are applicable, we compare their pros and cons. We also cite authoritative sources (peer-reviewed articles or classic texts) for each major method to justify our approach.

1 Q1: Long-Term Trends and Warming Rate Changes

Global annual temperature anomalies ($^\circ\mathrm{C}$ relative to 1951–1980) from 1880 to 2022. The bars illustrate the long-term warming trend: early years (blue) are mostly below the baseline, mid-20th century values level off, and recent decades (red) show rapid warming.

Analytical Steps:

- **Visualize the Full Series:** Plot the monthly and annual anomaly series from 1880–2023 to inspect overall direction and variability. The upward trend in global temperature anomalies is evident (see figure above), but we will quantify it. We note qualitatively that early 20th century temperatures were cooler, the mid-20th century experienced a pause or slight cooling, and warming accelerated in the late 20th century. This suggests at least one *deceleration* (mid-century) and subsequent *acceleration* in warming that need confirmation with analysis.

- **Extract the Trend (Signal Filtering):** Apply a *classical decomposition* to separate the trend from seasonal and irregular components. In classical seasonal decomposition, a centered moving average is used to estimate the trend–cycle, assuming the seasonal pattern is roughly constant year-to-year. For monthly data, a 12-month moving average (or 13-term with end-point adjustment) will smooth out the seasonal fluctuations, yielding a trend estimate. We will examine this trend line for changes in slope. *Pros:* Simple and transparent, highlights slow variations. *Cons:* Edge effects (uncertainty at the start/end of series) and the assumption of constant seasonal pattern, which might be violated if climate seasonality changed (to be examined in Q3).

- **Quantify Trend Changes:** Using the extracted trend, identify *sub-periods* where the slope appears different. For example, one might segment the data into 1880–1930, 1930–1970, 1970–2023 (roughly reflecting early warming, mid-century stagnation, and modern warming). Compute linear fits (ordinary least squares) to the trend or annual averages in each segment to estimate warming rates (°C/decade) in each period. If the slope in 1970–2023 is significantly higher than in 1930–1970, it indicates an acceleration in warming. Likewise, a near-zero or negative slope mid-century indicates a slowdown or pause. We will formally test differences in slope by checking confidence intervals of these segment regressions. *Pros:* Easy to implement and interpret (trend in °C/decade), highlights specific periods of acceleration/deceleration. *Cons:* The choice of breakpoints can be subjective; the climate system changes gradually, so results depend on chosen cut-offs.

- **Statistical Trend Models:** To avoid arbitrarily choosing breakpoints, we can model the trend with time-varying parameters:
  - Fit an *ARIMA model with drift* to capture a stochastic trend. For instance, a random walk with drift (ARIMA(0,1,0) + constant) treats the trend as linear drift plus noise. After fitting on the full series, check if residuals around that drift show structure—e.g. if residuals pre-1970 are mostly negative and post-1970 mostly positive, it suggests the drift (trend) increased and a single linear drift is insufficient. *Pros:* Provides a statistical test for non-zero drift. *Cons:* Assumes trend is constant (no accounting for change in drift).
  - Implement a *Dynamic Linear Model (DLM)* with a local linear trend component (level and slope as state variables) that evolves over time. Using the Kalman filter, we can allow the slope to change slowly (by specifying a small process variance for the slope). The Kalman filtered state estimates will show how the trend slope varied each year. We expect to see the slope increasing in recent decades if warming accelerated. *Pros:* Captures gradual trend changes statistically, with uncertainty bounds on the trend at each time. *Cons:* More complex and requires setting state noise variances; results need careful interpretation (the model's smoothing may spread out a sharp change).
  - *Hidden Markov Model (HMM) for regimes:* As an alternative, we could attempt an HMM that assumes the time series switches between a few discrete "regimes" with different underlying mean trends. For example, a 2-state HMM might classify each time step as either "slow warming regime" or "fast warming regime," with a higher average trend in the latter. The HMM is fit via maximum likelihood (Baum–Welch algorithm) and can infer the probability of being in each regime over time. If such regimes are found (e.g. one dominating before 1970 and another after), that would objectively flag an acceleration point. *Pros:* Can detect *abrupt* shifts in trend or variance in a data-driven way. *Cons:* Assumes discrete jumps; climate trends may have changed more gradually, so an HMM might either not find a clear split or could overfit short-term fluctuations as regime changes. Also, HMMs are more often used for detecting shifts in mean level or variance rather than a continuous slope change, so results need interpretation.

Method Comparison: We have multiple approaches to assess trends: classical decomposition vs. state-space/Kalman vs. HMM vs. piecewise regression. In practice, we might use a combination—e.g. use decomposition and visual analysis to suggest where changes occur, then confirm with a model (regression or DLM). Classical smoothing is quick and reveals the broad picture, while state-space models provide statistically rigorous estimates with uncertainties. HMMs are powerful for detecting hidden structure, but given the gradual nature of climate change, a continuous model (DLM) may be more appropriate. Each method's results will be cross-checked for consistency. By combining methods, we ensure robust conclusions about trend and acceleration in the warming signal. (For references: Chatfield (2004) and Box & Jenkins (1970) describe trend extraction and ARIMA modeling, while Petris et al. (2009) detail Kalman filtering for time-varying trends. HMM applications in time series are covered by Zucchini et al. (2017).)

2 Q2: Seasonal Pattern Analysis

Analytical Steps: Here we focus on the *seasonal cycle* in the data—how temperatures vary within each year on a regular pattern—and provide a meaningful interpretation of that pattern. Since the GISTEMP series is given as anomalies relative to a 1951–1980 monthly baseline, one might expect the *mean seasonal pattern* to be largely removed (baseline period had an average of zero anomaly for each month by definition). However, the data can still exhibit a seasonal pattern in anomalies for two reasons: (1) if the reference period is not representative of the entire period (e.g. earlier years or later years could have systematic seasonal differences), and (2) more importantly, the *seasonal variability* (amplitude of seasonal swings, or which months tend to be relatively warmer anomalies) might be discernible. We will:

- **Decompose Seasonal Components:** Use classical decomposition on the monthly data to estimate the *average seasonal indices*. In an additive decomposition, after extracting the trend as in Q1, we detrend the series and compute the average anomaly for each month across all years. This yields 12 seasonal values (one for Jan, Feb, …, Dec) that sum (approximately) to zero. These indices tell us, for example, whether anomaly values tend to be above the overall trend in certain months and below it in others. *Method:* Compute the mean of detrended anomalies for each calendar month. *Pros:* Simple and directly interprets how each month deviates from the annual mean. *Cons:* Assumes a constant seasonal effect over 140+ years (we will question this in Q3).

- **Examine Seasonal Means:** Alternatively (or additionally), calculate *seasonal aggregate anomalies* provided in the dataset: DJF (winter), MAM (spring), JJA (summer), SON (autumn) for each year. By averaging, say, all DJF values across years, we obtain an average winter anomaly, and similarly for other seasons. Comparing these can highlight if one season tends to have larger anomalies. For example, we might find that, relative to the 1951–80 baseline, recent *winter anomalies* are slightly higher than summer anomalies, implying winters have warmed more (this is a known phenomenon of greater land warming in cold season). We will quantify such differences: e.g. compute the long-term mean anomaly for each of the four seasons. If the anomalies were perfectly adjusted, these means might all be zero; any consistent differences hint at a seasonal pattern in the anomalies themselves.

- **Seasonal Subseries Plot:** Create a *seasonal subseries plot* (also called month-by-year plot): group the data by month and plot each month's values over time. This visualizes both the seasonal pattern and changes over time. For Q2's purposes, looking at each month's distribution can reveal the pattern: e.g. are anomalies in July typically higher or lower than anomalies in January? If we see that for most years July anomalies > January anomalies (or vice versa), that indicates a seasonal cycle in anomalies. Another visualization is a boxplot of anomalies by calendar month. This would show the distribution (median, IQR) of each month's anomalies across the whole record. From such a plot, we might observe, for instance, that *June–July–August* anomalies have a slightly different central value than *December–January–February* anomalies, revealing the seasonal signature.

- **Interpretation of Seasonal Pattern:** Once we identify the seasonal shape, we interpret it in physical terms. Key points might include:

  - The *amplitude* of the seasonal cycle (peak-to-trough) in the global average. The raw temperature (not anomalies) has a known seasonal cycle (global mean is typically highest around July and lowest around January, because Northern Hemisphere's larger land mass causes a global summer peak). In anomalies (deviations from a fixed baseline), any residual seasonal pattern could indicate that certain months have warmed more than others relative to the baseline.
  - We expect, given the baseline 1951–1980, anomalies in that period average to zero in each month. Before the 1950s, anomalies are mostly negative (cooler than baseline) for all months, and after 1980, anomalies are positive. However, the *seasonal pattern* could manifest as, say, winter anomalies being slightly more extreme (perhaps more negative in early period and more positive in recent period) than summer anomalies. This would align with greater warming in winters (possibly due to feedbacks like snow/ice melt).
  - If the seasonal indices from decomposition show, for example, that after removing the trend, *June* is on average +0.05°C and *December* is −0.05°C (just hypothetical), that means relative to the yearly average anomaly, June tends to run a bit higher and December a bit lower. Interpretation: maybe the reference period had relatively cooler Junes and warmer Decembers compared to the long-run average climate. We will articulate such findings clearly.

Pros and Cons of Approaches: Classical seasonal decomposition provides a clear average seasonal profile, which is easy to interpret ("seasonality" as a set of monthly corrections). However, it treats that profile as static. Direct examination of seasonal subseries gives more insight into variability and potential changes. A spectral analysis (Fourier) could also detect a clear annual cycle if present, but given the data are anomalies relative to a climatology, the annual cycle might not be strong in a spectral sense. Since the syllabus emphasizes time-domain methods, we stick to decomposition and plotting. We will use multiple views (decomposition indices, seasonal means, subseries plot) to ensure we correctly characterize the seasonal pattern.

Interpretation (Expected Outcome): We anticipate the seasonal effect in a global series to be relatively small compared to the trend and noise, because much of the seasonal cycle was removed by computing anomalies. Still, we might observe slightly higher anomalies in certain months. Likely, the analysis will show that the *seasonal pattern is linked to the distribution of land/ocean*: e.g. global anomalies might peak around *July–August* (mid-year) because that's when the Northern Hemisphere (which has more land and thus more temperature variability) is warmest, and dip slightly around *December–February*. If such a pattern emerges, we will interpret it as the effect of unequal warming—perhaps winters show a distinct behavior. Every claim will be backed by the data analysis. For instance, if winter anomalies are systematically larger in recent decades, we note that and connect it to known climate dynamics (e.g. winter polar amplification). In summary, Q2 is about describing *what the seasonal cycle looks like in the data* (even if subtle) and providing a logical explanation for it.

(Reference: Chatfield (2004) notes that identifying seasonal effects is crucial in time series analysis before modeling. We use classical decomposition as described in many texts, and our interpretation will align with climatological context as seen in NASA's documentation.)

3 Q3: Is Seasonality Constant Over Time?

This question asks whether the seasonal pattern identified in Q2 has remained stable or changed over the 1880–2023 period. In other words, we test for *evolution of seasonality*: has the magnitude or shape of the seasonal cycle in anomalies shifted? This could manifest as one season warming faster than another, altering the seasonal indices over time. We will approach this by comparing seasonal patterns in different periods and by using models that allow time-varying seasonality.

Analytical Steps:

- **Split-Sample Comparison:** Divide the dataset into at least two broad periods (e.g. 1880–1950 and 1951–2023, or perhaps thirds: 1880–1939, 1940–1979, 1980–2023) and compute the seasonal indices or monthly means for each sub-period. If seasonality is constant, each period should show a similar pattern of monthly deviations. If we observe, for example, that in the early period the difference between average July and January anomalies was X, but in the late period it's significantly different, that indicates a change in seasonal amplitude. We will quantify: for each month, the change in its mean anomaly from early to late period. A quick diagnostic is to compare the *seasonal amplitude* (max minus min of the monthly means) in each period. *Pros:* Simple and directly highlights changes. *Cons:* Only detects relatively large, long-term changes and assumes an abrupt difference between chosen periods (not sensitive to gradual continuous change).
- **Running Window Analysis:** To capture gradual changes, perform a *rolling decomposition*. For instance, take a 30-year sliding window (e.g. 1880–1909, 1890–1919, $\ldots$ up to 1994–2023) and for each window compute seasonal indices. Then track these indices over time. We might plot the seasonal amplitude or the value of a particular month's index as a function of the window center year. This would show trends such as "the winter seasonal effect has increased over time." If the seasonal pattern is truly constant, these should be flat lines (aside from random noise). If not, we'll see systematic drifts. Particularly, we suspect that if winters are warming relatively more, the winter anomaly (DJF) minus summer anomaly (JJA) gap might shrink over time.
- **Seasonal Difference by Year:** Using the provided seasonal averages per year, compute a time series of *seasonal contrast* each year. For example, define

$$\Delta = (\text{JJA anomaly}) - (\text{DJF anomaly})$$

  for each year. If Δ shows a trend (e.g. decreasing over time), it means the difference between summer and winter anomalies is changing—possibly implying winters warming faster (making Δ more negative or less positive). We will plot Δ vs. time and possibly fit a regression or AR(1) to see if there's a significant trend. Additionally, one could check other contrasts like (max season – min season) per year as an annual metric of seasonal disparity.
- **Statistical Test for Changing Seasonality:** Perform an *ANOVA or interaction test* in a regression framework: Fit a linear model to anomalies with terms for seasonal factors and an interaction of seasonal factors with time. For example, model

$$Y(t) = \mu + \beta\, t + \sum_{m=1}^{12} \gamma_m D_m(t) + \sum_{m=1}^{12} \delta_m \big[t \cdot D_m(t)\big] + \varepsilon(t),$$

  where $D_m(t)$ is a dummy for month m. The δ terms allow the seasonal monthly effect γ to change linearly over time (i.e. δ captures a trend in month m's anomaly). If many δ are non-zero (significant), seasonality is not constant. In particular, if δ for, say, winter months are positive (indicating those months' anomalies increase over and above the overall trend βt), that means winter warming is accelerating relative to summer. We would likely simplify this by grouping seasons (to avoid an overfit 12-parameter interaction): e.g. allow one interaction for all winter vs. summer. But even a simplified model could test the hypothesis: H_0 = "seasonal pattern is constant" vs. H_1 = "seasonal pattern changes." *Pros:* Formal statistical test for non-constant seasonality. *Cons:* Requires specifying a functional form of change (linear here). Also, with a long series, even small changes might appear significant.

- **State-Space Approach:** Use a *Dynamic Linear Model with time-varying seasonal coefficients*. In a Bayesian or Kalman filter framework (Petris et al. 2009), we can treat the seasonal component as a set of 12 state variables (one for each month's effect) with a constraint (sum to 0 each year). We then allow those state variables to evolve slowly (random walk with small variance). Running a Kalman smoother would then yield an estimate of how each month's seasonal effect changed over time. For example, it might show that the amplitude of the seasonal cycle (difference between highest and lowest monthly state) has declined slightly over the 20th century. *Pros:* Captures gradual evolution without pre-defining a particular change point or trend; yields uncertainty for the seasonal components each year. *Cons:* Complex to implement and requires careful tuning of the state noise (to avoid overfitting noise as "change"). Given the course scope, we may outline this method conceptually rather than fully implement it.
- **Hidden Markov Model for Seasonality Regimes:** Another advanced angle is to use an HMM to see if the series switches between states with different seasonal patterns. For example, state 1 might correspond to "normal seasonality" and state 2 to "diminished seasonality." We could fit a two-state HMM where the emission distribution in each state is a seasonal time series model with distinct seasonal means. If the algorithm finds a high probability that after a certain date the system is in a new state with, say, smaller seasonal amplitude, that flags a change. However, climate changes have been mostly gradual, so an HMM might not detect a sharp transition unless one artificially forces two regimes (pre/post a certain year). We mention this for completeness, but a continuous approach (rolling analysis or DLM) is more suitable if changes are smooth.

Pros/Cons and Checks:

- Using *simple differences* (like $\Delta = \text{summer} - \text{winter}$ anomalies each year) is intuitive and directly shows if one is catching up to the other. If we find, for instance, that Δ has trended downward to near zero by 2023, that would mean winter and summer anomalies have become almost equal (winters warming more). This is easy to communicate.
- The *rolling decomposition* provides a visual, but one must ensure the window is large enough to get stable seasonal estimates yet small enough to capture changes.
- The *regression interaction test* gives a p-value to formally accept/reject constant seasonality. If the null is rejected, we conclude seasonality changed significantly. If not, we conclude it's roughly constant.
- We will also be mindful of variability: even if the *average* seasonal pattern changes, year-to-year variability (like El Niño events, volcanic eruptions) can temporarily alter the seasonal distribution of anomalies. So we will differentiate long-term change from short-term fluctuations.

Interpretation: If we find that seasonality is *not constant*, an interpretation could be:

> "The seasonal pattern has evolved, with certain seasons exhibiting stronger warming. Specifically, we observe that winters have warmed slightly more than summers over the record, leading to a reduced seasonal contrast in recent decades."

Conversely, if results show no significant change:

> "The seasonal cycle in global anomalies appears remarkably stable over time – each month's deviation has increased in absolute value due to overall warming, but the *pattern* (which month is relatively highest/lowest) stayed consistent."

Either outcome will be explained. We will tie any change to known climate mechanisms: e.g. greater winter warming could be due to Arctic amplification in DJF, whereas summers are moderated by ocean thermal inertia.

Importantly, any finding of non-constant seasonality must acknowledge that our baseline period is fixed (1951–1980). As the base for anomalies, it can cause an apparent change: for instance, because anomalies are zero in the base period, earlier years (cooler) might show a different pattern simply because the baseline had a particular seasonal curve. We will double-check by possibly re-baselining the anomalies to another period (e.g. a 30-year period early on vs later) as a sensitivity check. If the seasonal changes persist regardless of baseline, it's a real physical change; if not, it might be an artifact of baseline choice.

(Reference: Many time series texts discuss checking the stability of seasonal patterns. For example, Chatfield (2004) notes that classical methods assume constant seasonality and can break down if this assumption fails. By applying methods above, we directly test that assumption. HMM usage for regime shifts in time series is discussed by MacDonald & Zucchini (2009), and DLM approaches by Petris et al. (2009).)

4 Q4: Exploratory Forecasting from January 2019 (5-Year Horizon)

In this step, we *pretend we are at January 2019* with data through December 2018, and we aim to forecast future global temperature anomalies up to 5 years ahead (through 2023). We will produce (a) one-step-ahead *recursive forecasts* for 5 years (which essentially means forecasting each month from Jan 2019 to Dec 2023 sequentially, using only information available prior to each forecast), and/or (b) a single *5-step-ahead forecast* (interpreted as a 5-year ahead prediction, since 5 years = 60 months). The key is to use *only exploratory univariate time-series tools* (no incorporation of external predictors or fancy machine learning beyond the syllabus), demonstrating how well we can project short-term trends and seasonality.

Step 0: Preparation – Using Data up to 2018:

We first truncate the dataset at December 2018. All model fitting and parameter selection will be done on this training period, as if 2019–2023 are unknown. We also note from Q1–Q3 insights that by 2018 the series has a strong upward trend and seasonality. We will leverage those insights: for example, if we found seasonality roughly constant and a continuing trend, our forecasts should account for both.

Step 1: Stationarize if Using ARIMA (Box–Jenkins Identification):

According to the *Box–Jenkins methodology*, we examine if the series is stationary and identify needed differencing. The anomaly series up to 2018 is clearly non-stationary (due to trend). Also, there is an annual seasonal periodicity. So, we will likely apply *differencing*: one non-seasonal difference to remove the long-term trend, and one seasonal difference (lag 12) to remove any persistent seasonal effect. After differencing, we will check the stationarity (e.g. the autocorrelation function (ACF) should decay to zero rather than staying high). If the differenced series looks stationary, we proceed to identify ARMA terms.

Step 2: Model Identification (ARIMA vs Exponential Smoothing):

We consider multiple forecasting approaches and compare:

- **ARIMA (Seasonal ARIMA):** Using ACF and PACF plots of the differenced series, determine plausible orders (p, q) for AR and MA terms (and (P, Q) for seasonal AR, MA). For example, the ACF might show a significant spike at lag 1 and 12, suggesting an AR(1) and seasonal AR(1) or MA terms. Given climate data often have some autocorrelation (e.g. due to ENSO cycles ~ 2–5 years), an AR(1) or AR(2) could be useful. We will likely consider a model like *ARIMA*$(0, 1, 1)(0, 1, 1)[12]$ or *ARIMA*$(1, 1, 0)(0, 1, 1)[12]$, which are common for trend + seasonal data (this is akin to Holt–Winters in ARIMA form). We will formally estimate candidate models using maximum likelihood and check diagnostics: residual ACF (should be white noise) and Ljung–Box test p-values. The *Box–Jenkins iterative approach* of identification → estimation → diagnostics → refinement will be used until we have a satisfactory model. *Pros:* ARIMA can model both trend (via differencing or drift term) and seasonality (via seasonal AR or differencing) and provide prediction intervals analytically. It's a statistically grounded approach with well-defined assumptions. *Cons:* Requires careful identification and may be sensitive to the non-stationarity (if the trend is changing, a simple ARIMA with constant drift might slightly under-forecast an accelerating trend).
- **Exponential Smoothing (Holt–Winters method):** We will also apply *Holt–Winters seasonal exponential smoothing*, treating the series as having a level, a trend, and seasonal components that are updated recursively. Specifically, we can use *additive Holt–Winters* (appropriate since anomalies can go positive/negative around zero) with parameters (α, β, γ) for level, trend, season updates. This method will produce a forecast by extrapolating the last estimated trend and repeating the seasonal pattern. Notably, Holt–Winters is known to be equivalent to certain ARIMA models; for instance, *non-seasonal Holt's linear smoothing* is optimal for ARIMA$(0, 2, 2)$, and the Holt–Winters with seasonality can be seen as a special case of a state-space model (ETS). We will fit Holt–Winters by minimizing in-sample error or using built-in formulas for optimal parameters. *Pros:* Simple to implement, automatically captures trend and seasonality, and often produces good forecasts for seasonal data. It doesn't require the data to be explicitly differenced; it inherently filters out the trend and season. *Cons:* Lacks the rigorous diagnostic framework of ARIMA; one must check that the chosen smoothing parameters are reasonable. Also, it assumes trend and seasonal patterns remain the same in the forecast period (which might be fine for short horizons).
- **Baseline (Naïve) Methods:** As a benchmark, consider two naive forecasts:
  - *Seasonal Naïve:* Forecast that each month in 2019–2023 will have the same anomaly as it did in the corresponding month of the last observed year (e.g. forecast Jan 2019 anomaly = Jan 2018 anomaly). This method basically says anomalies repeat year-to-year (which might be too simplistic given the trend).

- *Drift Method (Random walk with drift):* Use the last observed trend to extrapolate. For example, compute the linear trend from 2010–2018 and assume that continues: add that increment each year. In practice, the *random walk with drift* model would forecast that each future anomaly = last anomaly + (drift $\times$ number of steps ahead). For monthly data, drift per month is small, but over 60 months accumulates. This is essentially ARIMA(0,1,0)+drift. *Pros:* Very simple and often surprisingly hard to beat for short-term forecasts (random walk forecast uses the latest value as best guess). *Cons:* Here it likely underestimates warming because it ignores that the drift itself might be increasing and also ignores seasonality completely (except seasonal naive which ignores trend). These methods are mainly for baseline comparison to ensure our more complex models add value.

- **Dynamic Linear Model (State-Space Forecasting):** If we set up a DLM with trend and season components (as per Q1 and Q3 analysis) and we have estimated it up to end of 2018, we can use the *Kalman filter* equations to forecast ahead. The DLM approach would naturally provide a forecast distribution. For example, a local linear trend + seasonal DLM would produce forecasts similar to Holt–Winters but with Kalman filter handling of uncertainty. We might mention this approach to demonstrate understanding, although implementing it fully might be beyond an exploratory scope. *Pros:* Conceptually unifies the trend/seasonal model and gives optimal forecasts under the state-space assumptions. *Cons:* Complexity and not strictly necessary if simpler methods suffice in performance.

Step 3: Generate Forecasts (2019–2023):

We will produce the forecasts using the chosen method(s). For (a) one-step-ahead recursive forecasts, the process is: forecast Jan 2019, then when moving to Feb 2019, incorporate Jan 2019 (if actual Jan 2019 was known, we would use it, but since we pretend we don't know actual, we might instead feed the Jan 2019 forecast into the model for Feb—effectively this is just the multi-step forecast logic). In practice, for ARIMA/Holt–Winters, generating a 60-month ahead forecast yields the same result as iteratively one-step forecasting, because we are not updating the model with new actuals (since in a true forecast scenario we don't have them). However, to mimic how one would do it in real time, we can simulate: forecast Jan 2019, then when the (hypothetical) Jan 2019 comes in, update, forecast Feb 2019, etc. But since we have no new info, this is equivalent to the direct multi-step forecast. We will present the forecast values for each year 2019, 2020, … 2023, likely as annual averages or selected months for brevity. For (b) a single 5-step (5-year) ahead forecast, we interpret that as the endpoint forecast for Dec 2023 (or the average of year 2023). We can provide that as well—basically the same output but focusing on the farthest horizon.

If multiple models are used (ARIMA vs Holt–Winters), we will compare their forecasts. Do they both show continuing warming at similar rates? Minor differences might occur: e.g. Holt–Winters might extrapolate the last few years' strong trend which could yield slightly higher 2023 forecast than an ARIMA with constant drift estimated over a longer period (which might dilute the recent acceleration). We will tabulate or plot the forecast trajectories.

Step 4: Method Comparison and Justification:

We will clearly list the chosen forecasting approach and why:

- *Preferred method:* A *Seasonal ARIMA* (or Holt–Winters) model that captures the upward trend and seasonality is likely to be our primary tool. This method is consistent with the Box–Jenkins forecasting principles and the methods taught in class. It allows us to produce both point forecasts and interval forecasts (to be discussed in Q5).

- *Comparison:* We will mention how Holt–Winters' forecasts compare. Often, for a short horizon like 5 years, different reasonable methods will produce qualitatively similar forecasts because the inertia of the climate system is large. We expect all methods to forecast continued positive anomalies (given the strong recent trend). The difference might be in the slope: a method using only data up to 2018 might slightly underestimate 2019–2023 because the warming trend was accelerating (e.g. if one simply took a long-term average drift). Holt–Winters, which gives more weight to recent data via smoothing, might forecast a bit higher anomalies.

- We will also note that these are purely statistical forecasts. They effectively extrapolate the past into the future. For a physical system like climate, one might also consider external scenario information (like projected greenhouse gas emissions), but since the task restricts us to time-series tools, we acknowledge that limitation. In our exploratory framework, the statistical forecast is an "unconditional" forecast assuming past patterns continue.

(References: The Box–Jenkins ARIMA forecasting framework is documented by Box et al. and many others. Chatfield (2001, 2004) compares exponential smoothing and ARIMA, noting Holt–Winters can be viewed as ARIMA(0,2,2) with seasonality. We will cite these to justify model choice. Also, Chatfield (2000) emphasizes comparing forecasts to a simple benchmark like a random walk forecast to ensure our model adds skill.)

5 Q5: Forecast Uncertainty and Its Limitations

Assessing Uncertainty:

- **Model-Based Prediction Intervals:** For the chosen forecasting model (e.g. ARIMA), we will calculate standard *forecast intervals* at 1 to 60 months ahead. ARIMA models yield a forecast error variance that typically grows with the horizon. For example, in a random walk with drift model, the forecast variance increases linearly with lead time (h). We will use the appropriate formula or software to get 80% or 95% prediction intervals for each future point. These intervals assume the model is correct and only account for the inherent randomness (the unexplained residual variance) propagating forward. Exponential smoothing methods can also generate intervals; since Holt–Winters can be seen as an ARIMA, it has analogous formulas. We will report something like: *"The model predicts December 2023 anomaly to be* $\approx 1.0\,^{\circ}\mathrm{C}$ *with a 95% prediction interval of [0.6 °C, 1.4 °C]."* This communicates uncertainty.
- **Inspection of Interval Width:** We will examine how wide the intervals are. If using ARIMA with drift, because the model essentially is highly persistent (close to a random walk with drift), the uncertainty of the level can grow quite large over 5 years. However, global temperature anomalies have a lot of inertia; one might find the interval is not extremely large ($\sim \pm 0.2\breve{\ }0.3\,^{\circ}\mathrm{C}$ by 5 years). We will highlight that even the upper bound of our forecast still indicates warming, just with some margin.
- **Limitations of These Intervals:** A crucial point: *model-based intervals likely underestimate true uncertainty* because they condition on the chosen model. They capture *random variation* (the "error term" uncertainty) but often ignore *parameter uncertainty and structural uncertainty*. For example, our ARIMA parameters (like drift) were estimated; if we had estimation error, the real forecast uncertainty is larger. Chatfield (1993) notes that out-of-sample prediction intervals tend to be too narrow in practice, meaning they do not capture the required proportion of future observations. One reason is that models often don't account for the possibility of regime change or model misspecification. We will explicitly cite that "Ignoring model uncertainty can lead to forecast intervals that are too narrow." In our climate context, a structural change could be an unexpected event (major volcanic eruption, a sudden shift in ocean cycles) or a change in emissions trajectory—none of which our purely statistical model knows about. Thus, the interval we give is conditional on "business as usual" climate dynamics continuing.
- **Assessing Uncertainty Qualitatively:** We will supplement numeric intervals with qualitative discussion. For instance, we might say: *"The forecasts have limited uncertainty in the sense that we are quite confident the global mean will remain above the 1951–1980 average (most likely considerably above, given all recent years were +0.8 °C or more). However, the uncertainty in the precise anomaly for a given month is non-negligible. Natural variability (e.g. El Niño/La Niña events) can cause a particular year to be warmer or cooler by* $\sim 0.1\breve{\ }0.2\,^{\circ}\mathrm{C}$. *Our methods' intervals account for typical variability, but if an extraordinary event occurred (like a large volcanic eruption cooling the planet briefly), the actual anomaly could fall outside our predicted range."* This kind of scenario-based uncertainty is *not* captured by standard time-series models, which assume the future will behave statistically like the past. We will explain that limitation.
- **Why Uncertainty May Be Limited (in another sense):** There's an interesting nuance: On one hand, our prediction intervals might be too *narrow* by not including all uncertainties, but on the other hand, one could argue our ability to assess uncertainty is *limited*. We only have the time series itself to estimate variability. We cannot easily put probabilities on unforeseen shocks. Also, climate has a *trend*—so unlike forecasting, say, stock prices, here the uncertainty of direction (warming vs cooling) is low; we are quite sure of the direction (upwards). The main uncertainty is in the rate of increase and short-term fluctuations. We will clarify that because the warming trend is strong and consistent, the predictions of continued warming are actually fairly robust in sign—thus the forecasts are *qualitatively certain* (we confidently predict anomalies will stay high or rise) even if *quantitatively* the exact values have error. This is somewhat unique: a random walk without drift would have uncertainty growing with $\sqrt{h}$, but a drift (trend) reduces relative uncertainty in sign.

- **Comparing Methods' Uncertainty:** If we tried both ARIMA and Holt–Winters, we can compare their interval estimates. They should be similar if mapped to the same confidence level, since Holt–Winters can be tied to an ARIMA model. If there's discrepancy, we'll discuss it. Perhaps we'll mention that a fully Bayesian approach (put priors on parameters, etc.) could widen intervals to account for parameter uncertainty, but that's beyond our scope. Instead, we might do a simple sensitivity: try fitting the model on an earlier cutoff (say up to 2010) and forecasting 2011–2018, see how much actuals deviate—that gives an empirical sense of forecast error which we can extrapolate. This is more of a check than a requirement, but it can illustrate whether our interval calibration is reasonable.

Summary of Uncertainty: We will provide at least one interval or error bar visualization for the 5-year forecast and state: *"For example, by 2023 our forecasted anomaly is* $\approx 1.0\,^{\circ}\mathrm{C}$ *with an uncertainty of* $\pm 0.3\,^{\circ}\mathrm{C}$ *(95% interval). This interval may not fully capture all uncertainties—for instance, it assumes no major changes in the climate forcing. As Chatfield (1993) argued, practical prediction intervals often end up too narrow because they ignore model uncertainties. We acknowledge that limitation. Nevertheless, the forecast confidently indicates continued warming, and the probability of a return to 1951–1980 average levels (0 °C anomaly) in the next 5 years is extremely low under these models."* This addresses both the numeric assessment and the deeper explanation of limits.

In essence, the uncertainty analysis will show that while we can put error bounds on our forecasts (and we should, per good practice), those bounds have to be interpreted with caution. They likely underestimate the true uncertainty if something fundamentally shifts. However, given the inertia of climate, our 5-year horizon might not be hugely uncertain—we're not forecasting random fluctuations but an ongoing trend plus known variability. We will likely conclude that the biggest source of uncertainty in these forecasts is *internal variability* (like year-to-year fluctuations), rather than uncertainty about the existence of the trend itself. And because our models don't include physics, they might not anticipate, say, a leveling off or acceleration beyond what's in the data.

(References: We will cite Chatfield (1993) on the importance of interval forecasts and common issues. Also Chatfield (2000) noted that one should present forecasts as intervals to illustrate uncertainty. The concept that model-based intervals ignore some uncertainties is widely noted. We can also reference that state-space forecasts from Kalman filter naturally give error bands but still conditional on the model. These citations back up our cautionary statements.)

Conclusion (Summary)

In this strategy document, we outlined a comprehensive approach to analyze the NASA GISTEMP time series. For **long-term trends (Q1)**, we will extract the warming trend using moving-average smoothing and advanced models, identifying a mid-century slowdown and late-century acceleration. For the **seasonal pattern (Q2)**, we will describe the annual cycle of anomalies using decomposition, noting how global seasonal effects manifest in the anomaly data. We then test **seasonality stability (Q3)**, using split-period analyses and possibly state-space models to determine if winters have warmed faster than summers (i.e. changing seasonal amplitude). Transitioning to **forecasting (Q4)**, we will fit suitable time-series models (like Holt–Winters and ARIMA) to data up to 2018 and generate 5-year forecasts, comparing methods to ensure robust predictions. Finally, we will **assess forecast uncertainty (Q5)** by providing prediction intervals and discussing their limitations, acknowledging that while our models give a reasonable range, real-world unpredictability and structural changes could widen the true uncertainty. Throughout, we ensure to use only the tools taught in the course, to justify each analytical step with sound reasoning and literature support, and to maintain a clear distinction between analyzing past data (filtering signals) and predicting future data (forecasting). This ensures the final project is methodologically rigorous and well-grounded in time-series analysis principles while addressing the practical questions about global temperature trends and forecasts in a scientifically interpretable way.