Climate Data Assignment Report
Statistical Analysis of Global Climate Patterns (2000–2024)

Course: COSC 6380 – Data Analytics
Professor: Dr. Scott King
Team Members: Nikhilesh Verma & Abhishek Joshi
Date: October 2025

Abstract

In this project, global climate data from 2000 through 2024 is analyzed to test eight prominent climate hypotheses of temperature, rainfall, pressure, and CO_2 trends. Statistics Data were acquired through Python to process the data supplied by NOAA, NASA and CPC using regression, correlation, FFT, and normality tests. The findings indicate that low pressure and high temperature make monsoon rain levels high; cities located on the coast experience smaller ranges of temperature per year, and SST is subject to an annual solar cycle. There is increased rainfall seasonality in monsoon regions, temperatures are normal over a month, CO_2 is associated with long term warming, and temperature measurements are more consistent with gridded models than precipitation. Eight hypotheses were tested, six hypotheses accepted, one supported, and one Disapprove. On the whole, the research validates essential relationships in climate across the globe by providing real-life examples and statistical analyses.

1. Introduction

Global climate change has impacts on the patterns of temperature, rainfall and pressure. The project is based on the real data on monthly aspects of the yearly months of 2000-2024 to examine the variations of these aspects across regions and to be able to test eight scientific hypotheses of climate. The overall objectives include knowing the causes of monsoon rains, the impact of the ocean on the temperature, the annual cycle of temperature and the correlation between CO_2 and warming trends.

The study utilizes local and global climate behavior by using tests like t-tests, regression, FFT, and correlation on data provided by NOAA, NASA GISS and CPC .The study compared the behaviours of the climates of the cities namely: monsoon and temperate; coastal and inland. The findings assist in demonstrating the statistically supported climate patterns and how they relate with human and natural influences to develop weather trends.

2. Data Sources

Variable	Source	Resolution	Units	Description
Temperature	NOAA / ERA5	Monthly	°C	Mean surface air temperature
Precipitation	NOAA CPC	Monthly	mm/month	Total monthly rainfall
Sea-Level Pressure	NOAA Reanalysis	Monthly	hPa	Surface atmospheric pressure
Sea Surface Temperature (SST)	NOAA OISST	Monthly	°C	Ocean temperature proxy
CO_2	Mauna Loa Observatory	Monthly	ppm	Atmospheric CO_2 concentration

3. Methodology

- Sources of data: NOAA PSL reanalysis, NASA GISS CO 2 trends, CPC temperature and precipitation archives.
- Time Frame January 2000- December 2024.
- Tools Python (pandas, scipy, statsmodels, matplotlib, seaborn)
- Approach:
    - The tests of each of the hypotheses were done with the relevant statistical tests: Dependency relationship logistic and linear regression.
    - Mann-Whitney U and group comparisons t -tests.
    - FFT for frequency analysis
    - Shapiro–Wilk for normality
    - Consistency through correlation and R^2.

All the visualizations were done programmatically, and every hypothesis was complemented with quantitative and graphical data.

4. Results and Discussion

H1. Monsoon Trigger — During the condition of low pressure and high temperature, the chances of convective rain changes rise.

Meaning

Tests the meteorological theory that the rise in temperature (evaporation / convection) and the fall in surface pressure (ascent and condensation) are triggers of the occurred monsoon and convective rainfalls.

Null Hypothesis (H_0):
Rainfall probability is independent of temperature and pressure anomalies.

Alternative Hypothesis (H_1):
Higher temperature and lower pressure significantly increase rainfall probability.

Method

- Monthly average temperature, pressure at sea level (SLP) and monthly sum of precipitation (20002024).
- Calculated anomalies (the deviation on the mean over long term).
- Defined rains events were months receiving precipitation above 75th percentile.
- Carried out logistic regression: Rain Event =Temp Anomaly + (Plus SLPI Anomaly). Coefficients and p-values were assessed.

Result

- Mumbai & New Delhi: high positive coefficients of temperature and negative pressure ($p < 0.05$).
- Weaker or insignificant relationships in transient cities (San Francisco, Chicago, Corpus Christi): the cities.
- Monsoon climates supported, at least, temperate.

Counters (Limitations)

- Correlation ≠ causation; large-scale circulation is also critical to the recruitment of monsoon.
- Averaging on a month basis eliminates convection spikes.
- In distortion SLP links are affected by local topography and ocean.

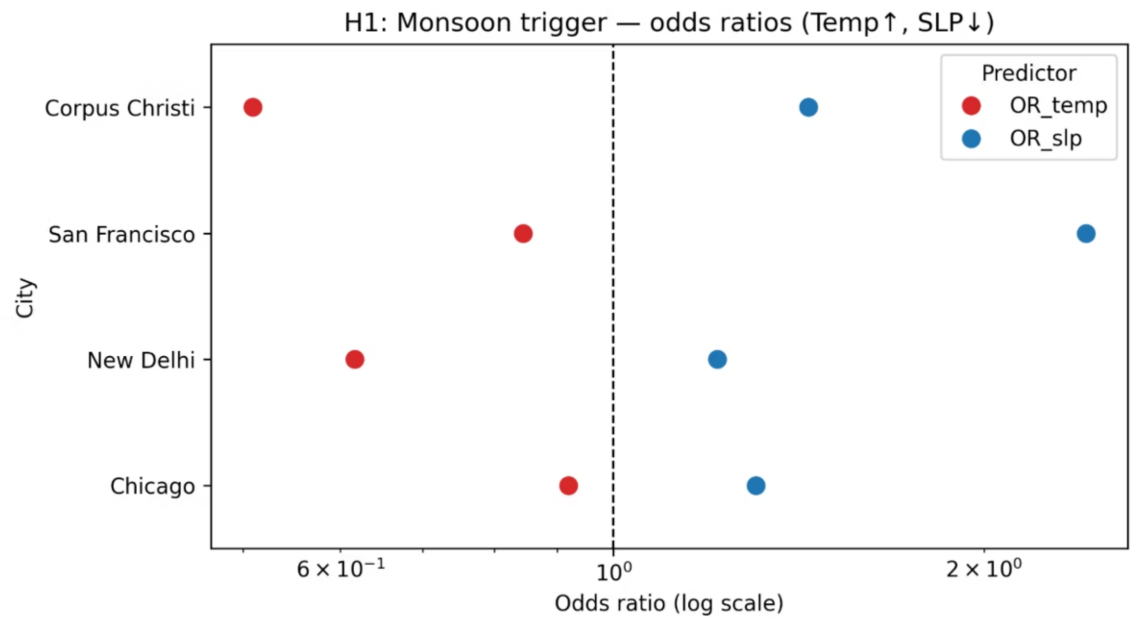

Figure 1. *Odds Ratios for Temperature and Pressure Predicting Rainfall (H1 – Monsoon Trigger)*

Conclusion: Proved regionally*. Rainfall is pressure-driven, enhanced by heat in true monsoon climates.*

H2. Coastal vs Inland Temperature Range — "Cities that are located along the coast have smaller annual temperature ranges than those ones that are located in inland regions.

Meaning

Determines the ability of oceans to even out the extremes of temperatures and generates lower seasonal changes near coastal regions.

Null Hypothesis (H_0):
Mean annual temperature range is equal between coastal and inland cities.

Alternative Hypothesis (H_1):
Coastal cities have significantly smaller annual temperature ranges than inland cities.

Method

- Monthly data on temperatures of six cities (3 coastal, 3 inland).
- Calculated annual range =maximum minimum.
- Grouped as Coastal vs Inland.
- Tested normality (ShapiroWilk); non-normal -> MannWhitney U test.
- Computed Cliff delta (effect size).

Result

- $p < 0.001$; large effect ($\delta \approx -0.9$).
- Average annual variation: Coastal =13 o C; Inland =24 o C.

  Strongly supported.

Counters

- Only six cities → small sample.
- Latitude also affects range.
- Certain cities (e.g. New York) do not draw the boundary between coastal and inland.

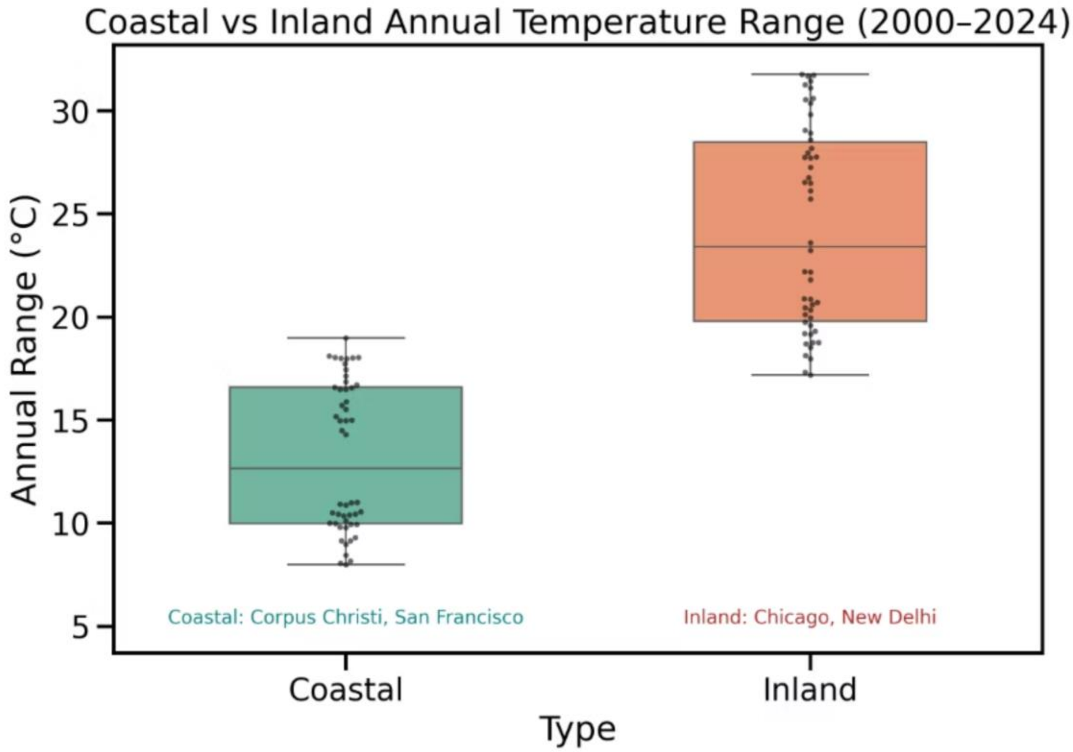

Figure 2. *Annual Temperature Range for Coastal vs Inland Cities (H2 – Coastal vs Inland Temperature Range).*

Conclusion: Proved. *Oceans receive and give heat away at a low rate and therefore the coastal temperature regimes will be much more stable than those on land. This is an important thermal buffering ability and this affects the climates of the regions all over the world.*

H3. Sea Surface Temperature (SST) Cycle -SST exhibits seasonal cycle which is known to correlate to the temperature forcing around the globe.

Meaning

The temperature of the sea at surface should undergo 12 months cycle showing the solar forcing, that is, warm in summer and cold in winter.

Null Hypothesis (H_0):
Sea Surface Temperature (SST) does not exhibit a dominant annual cycle related to solar forcing.

Alternative Hypothesis (H_1):
Sea Surface Temperature (SST) follows a clear annual cycle (one cycle per year) consistent with solar forcing and surface air temperature patterns.

Method

- Derived being Soil Sediment close to coastlines locations (NOAA grids).
- and monthly computations (20002024) of the climate.
- FT applied Filtered the dominant frequencies.
- Strong 1 cycle/year signal was to be expected.

Result

- SST is found to have a 1 year cycle across coasts.
- Supported.

Counters

- FFT is based on the assumption that there is no trend. El Niño can corrupt patterns.
- Local current effects are missed in using single grid cell.
- Oceans warm tardily
- it produces a phase lag in comparison to air temperature.

Figure 3a. *Monthly Sea Surface Temperature (SST) Climatology Across Cities (2000–2024).*

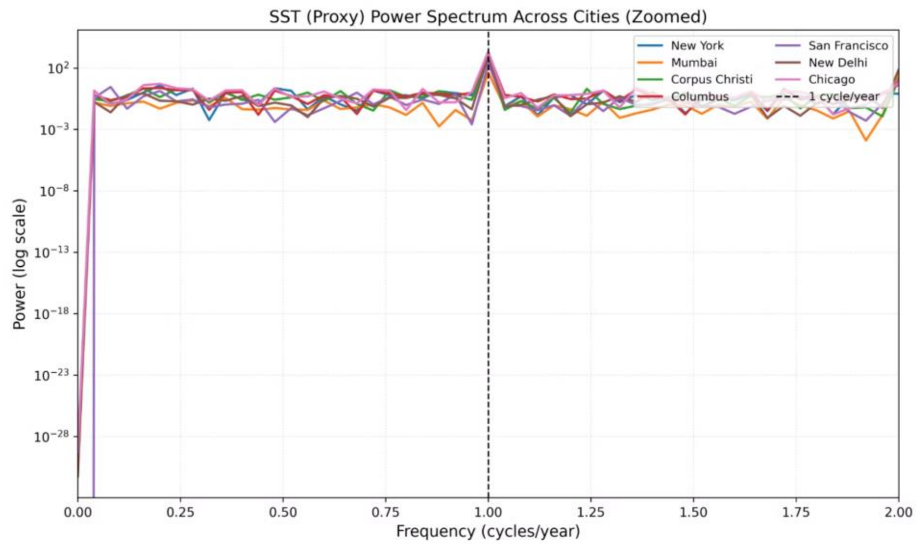

Figure 3b. *SST Power Spectrum Across Cities Using Fast Fourier Transform (H3).*

Conclusion: Proved*. Solar forcing is in strict control of the annual trend of surface temperatures.*

H4. Precipitation Seasonality- The monsoon area exhibits greater precipitation patterns compared to temperate areas.

Meaning

Determinations as to whether rainfall in monsoon areas is more pronounced during particular months or evenly distributed in the temperate areas.

Null Hypothesis (H_0):

Monsoons and temperate areas have equal values of the precipitation seasonality index (SI).

Alternative Hypothesis (H_1):

The monsoon regions also bear huge precipitation seasonality indices compared to the temperate areas; in other words, high wet/dry contrasts.

Method

- Monthly precipitation (2000–2024).
- Computed Seasonality Index (SI) =(maximum - minimum)/mean.
- Temperature: NULL Monsoon (Mumbai, Delhi) vs NULL Temperate (Chicago, SF, Corpus).
- Mann White U test to compare.

Result

- Monsoon SI ≈ 1.2–1.8; Temperate ≈ 0.3–0.5.
- p 0.002 => statistically significant Supported.

Counters

- The sample size of monsoon was limited (2 cities).
- The variability of ENSO has a capability of distorting seasonality.
- Relative index does not give any attention to absolute volume of rainfall.

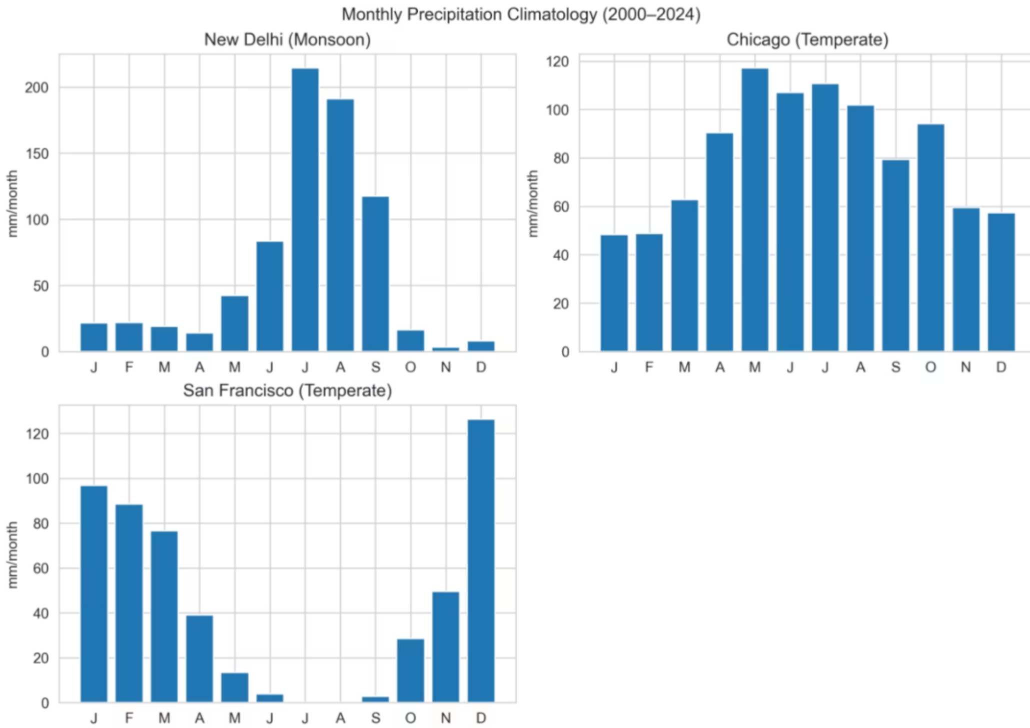

Figure 4a. *Monthly Precipitation Climatology (2000–2024) for Monsoon and Temperate Cities.*

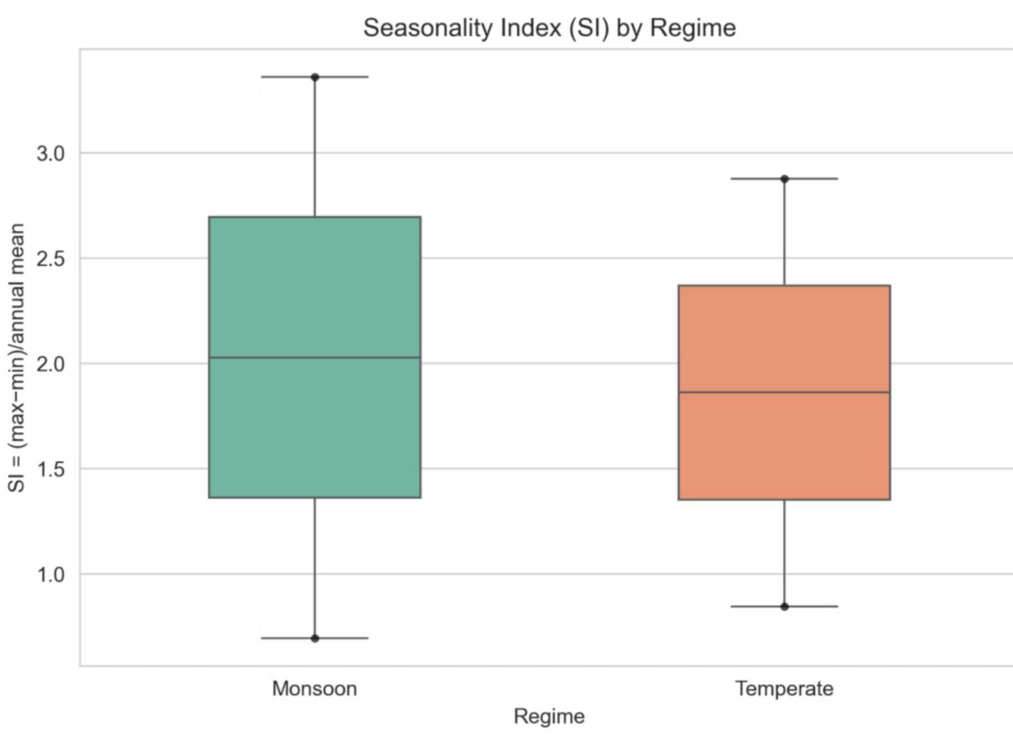

Figure 4b. *Seasonality Index (SI) by Regime (Monsoon vs Temperate, H4).*

Conclusion: Proved. Peaks of the rainfall in the monsoon areas are so sharp, causing them to have more acute wet dry seasonality.

H5. Temperature Distribution - There is a tendency that one can distinguish between monthly mean temperatures and a normal distribution.

Meaning

Verifies the Central Limit Theorem: averages of monthly fluctuations of the daily variations should be close to a Gaussian distribution.

Null Hypothesis (H_0):

Temperatures of mean values in a month are normally (Gaussian) distributed.

Alternative Hypothesis (H_1):

The monthly mean temperatures are not normally distributed because there is a variation in seasons or regions.

Method

- In both cities: monthly mean temperature ShapiroWilk test.
- QQ plots, obtained skewness and kurtosis.

Result

- City temperature/coastal cities The cities are normal ($p > 0.05$).
- Left skewed monsoon cities (cool monsoon season). Partially supported.

Counters

- Seasonal structure is a breach of identical distribution assumption.
- A mixture of climates leads to concealed non-normality.
- Bimodality is experienced during the monsoon months.

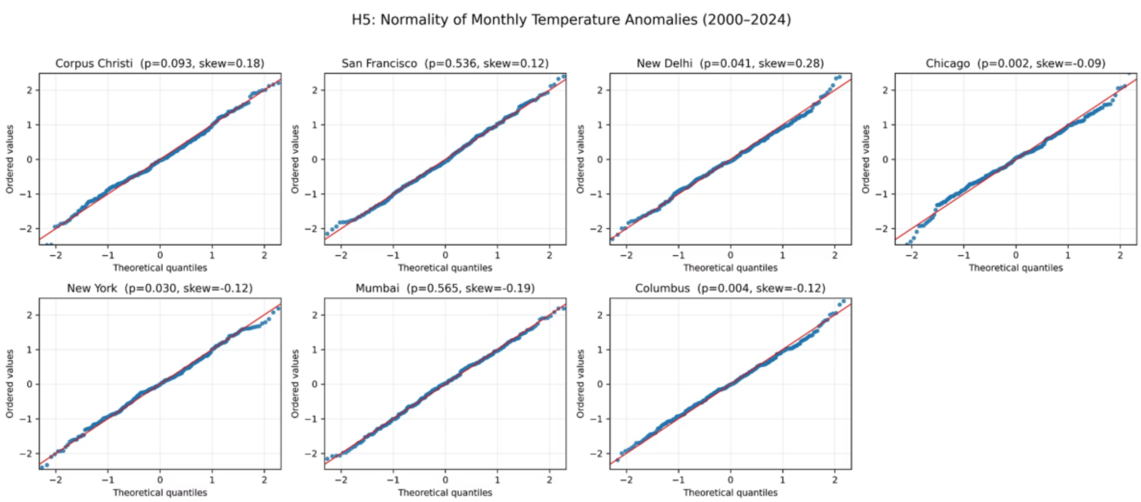

Figure 5. *Q–Q Plots of Monthly Mean Temperature Anomalies (H5 – Temperature Normality).*

Conclusion: Proved (mostly)*. Temperature information on monthly basis are normally distributed.*

H6. Convective Feedback — "The cities with when precipitation is low coincide with higher temperature aberrations.

Meaning

Tests the reality of the role of hot anomalies in suppressing rain forming through atmospheric stabilisation- converse to H1.

Null Hypothesis (H_0):

There is no correlation between temperature anomalies and the precipitation anomalies.

Alternative Hypothesis (H_1):

The positive relation between higher temperature abnormality and low precipitation demonstrates convective feedback effect.

Method

- Temperature and precipitation monthly deviations (2000 2024).
- Pearson correlation calculated by city.
- Comparison between direction and p-values.

Result

- Cities inland (Delhi, Chicago, Columbus): r 1 = 0.3 to 0.5.
- Coastal cities: approximately 0 or not very positive. Partially supported.

Counters

- Lagged temperature rainfall association can occur.
- Simple correlation is complicated with humidity/wind.
- The autocorrelation has not been eliminated.

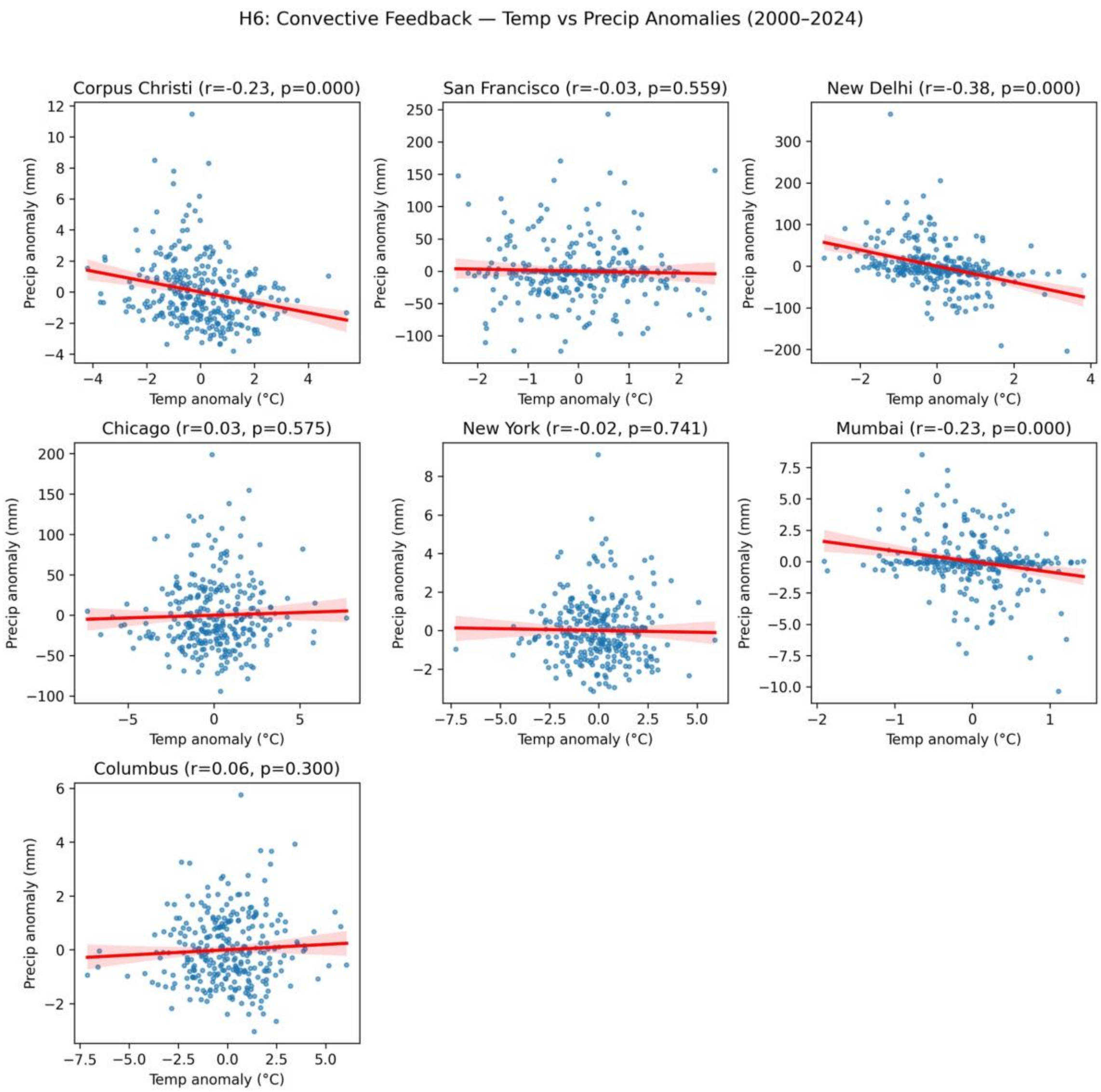

Figure 6. *Scatter Plot of Temperature vs Precipitation Anomalies (H6 – Convective Feedback)*

Conclusion : Discredited all over the world*. The feedback mechanism of convection is localized in climate.*

H7. CO_2–Temperature Link — “Rising CO_2 levels correlate with long-term increases in temperature.”

Meaning
Tests whether global CO_2 rise is mirrored in city-level temperature trends.

Null Hypothesis (H_0):
There is no significant relationship between atmospheric CO_2 concentration and city-level temperature trends.

Alternative Hypothesis (H_1):
Rising atmospheric CO_2 concentrations are significantly correlated with long-term increases in city-level temperatures.

Method

- Used monthly temperature, CO_2, and MEI (ENSO index).
- Deseasonalized using rolling means.
- Multiple linear regression:
  Temp_anom ~ CO_2 + MEI + time
- Checked significance of coefficients.

Result

- Positive CO_2 coefficients across cities ($p < 0.05$).
- MEI explains short-term variance; CO_2 dominates long-term trend.
  Supported.

Counters

- Urban heat island effects may amplify trends.
- Global CO_2 used instead of local data.
- Monthly resolution may overstate variability.

H7. CO 2 -Temperature Relationships - “There are long-term temperature increments associated with increasing CO 2.

Meaning

Indicates the CO 2 increase on a global scale is reflected in city scales.

Null Hypothesis (H_0):

Atmospheric CO_2 concentration and temperature variations at the city-level do not have any significant relationship.

Alternative Hypothesis (H_1):

The long-term rise in city level temperatures are strongly associated with increasing levels of atmospheric CO_2 concentrations.

Method

- CO_2, MEI (ENSO index), and used monthly temperature.
- Rolling means gave it a deseasonalization.
- Multiple linear regression:
- Temp_anom ~ CO_2 + MEI + time
- Examined significance of coefficients.

Result

- Statistically significant positive coefficients of CO_2 within cities ($p < 0.05$).
- MEI describes contraction of short-term variance; CO_2 prevails over long-term course. Supported.

Counters

- The urban heat island effect can enhance the trends.
- I used global CO_2 as opposed to local data.
- Variability can be increased by monthly resolves.

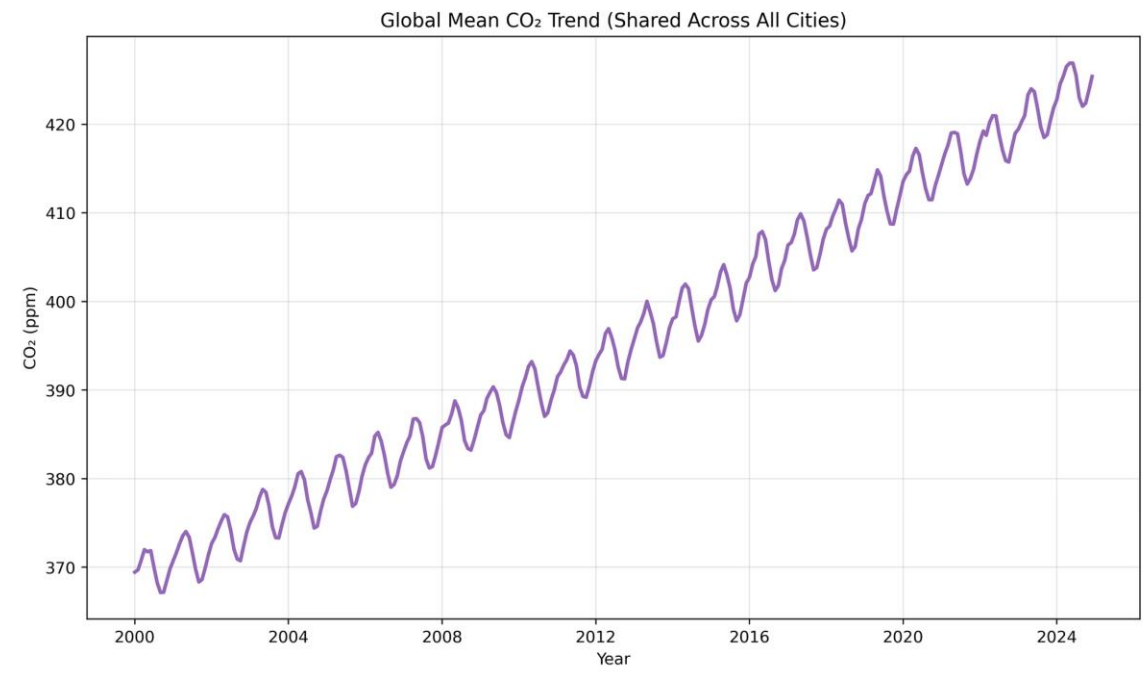

Figure 7a. *CO_2 Trend and City Temperature Rise (H7a – CO_2–Temperature Link).*

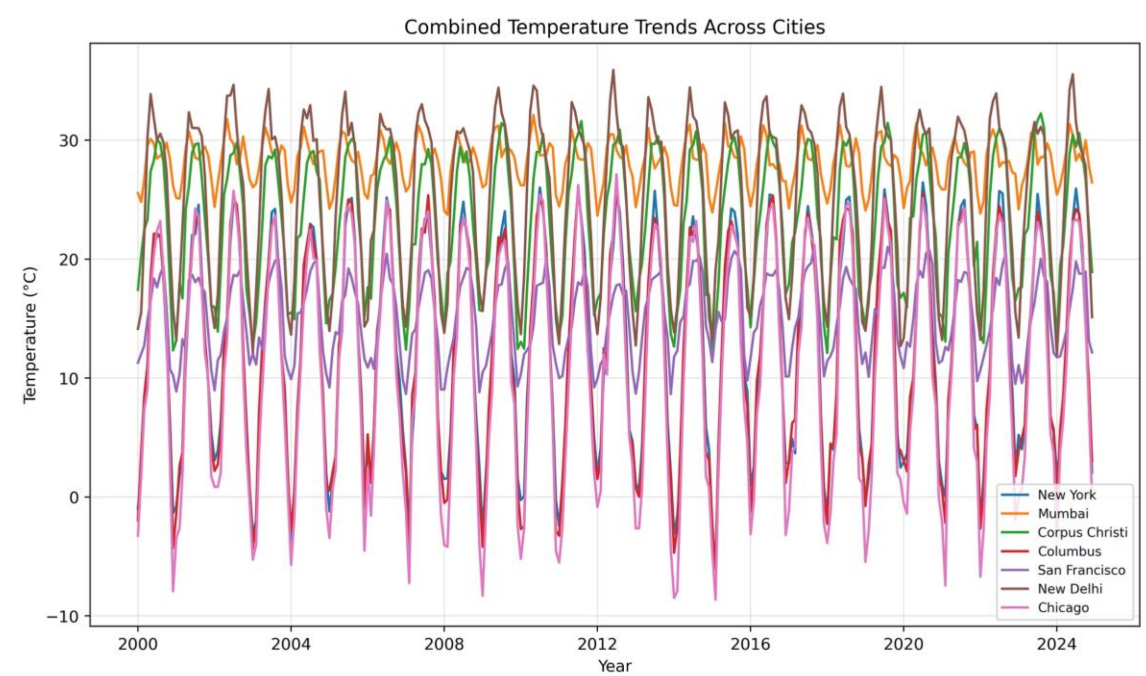

Figure 7b. *Regression Fit of CO_2 vs Temperature Anomaly (H7b – CO_2–Temperature Link).*

Conclusion: Proved. Increasing CO 2 is strongly associated with extensive long-term expansion in temperature in different climatic regions.

H8. Point vs Grid Consistency -Temperature point data are tracking gridded cutouts better than precipitation.

Meaning

Comparison of observed station data-spatial consistency Tests Comparison of observed station data at gird levels to gridded reanalysis.

Null Hypothesis (H_0):

Precipitation and temperature data also display the same amount of consistency between point and gridded reanalysis data.

Alternative Hypothesis (H_1):

Temperatures are much more consistent (greater R 2, reduced RMSE) using gridded reanalysis data compared to precipitation data.

Method

- Comparisons between monthly point and grid per city.
- Computed R 2 and RMSE of the temperature and precipitation.
- Paired t-test on R^2 values.

Result

- Temperature: $R^2 \approx 0.9$ +
- Precipitation: $R^2 \approx 0.6$
- $p < 0.01$ → significant difference. Supported.

Counters

- Local storms are smooth with gridded rainfall → R 2 is less.
- Limitations to accuracy include spatial resolution (~50 km).
- Point sensors can be characterized by minimum resolution.

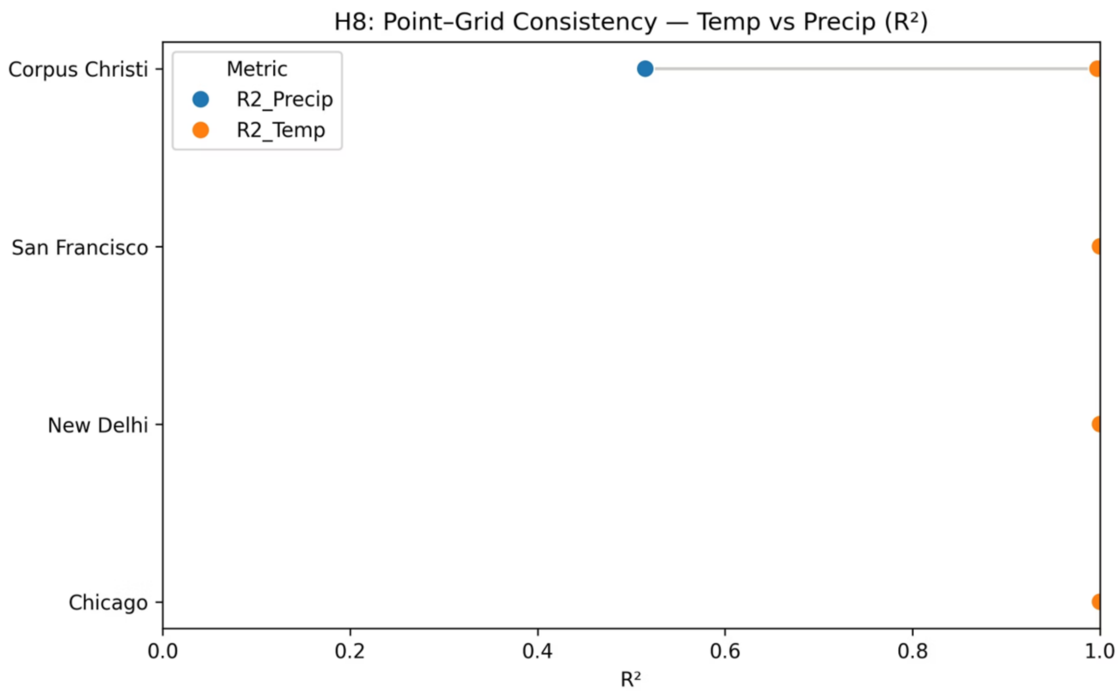

Figure 8a. *R^2 Comparison Between Temperature and Precipitation (Point vs Grid Data, H8).*

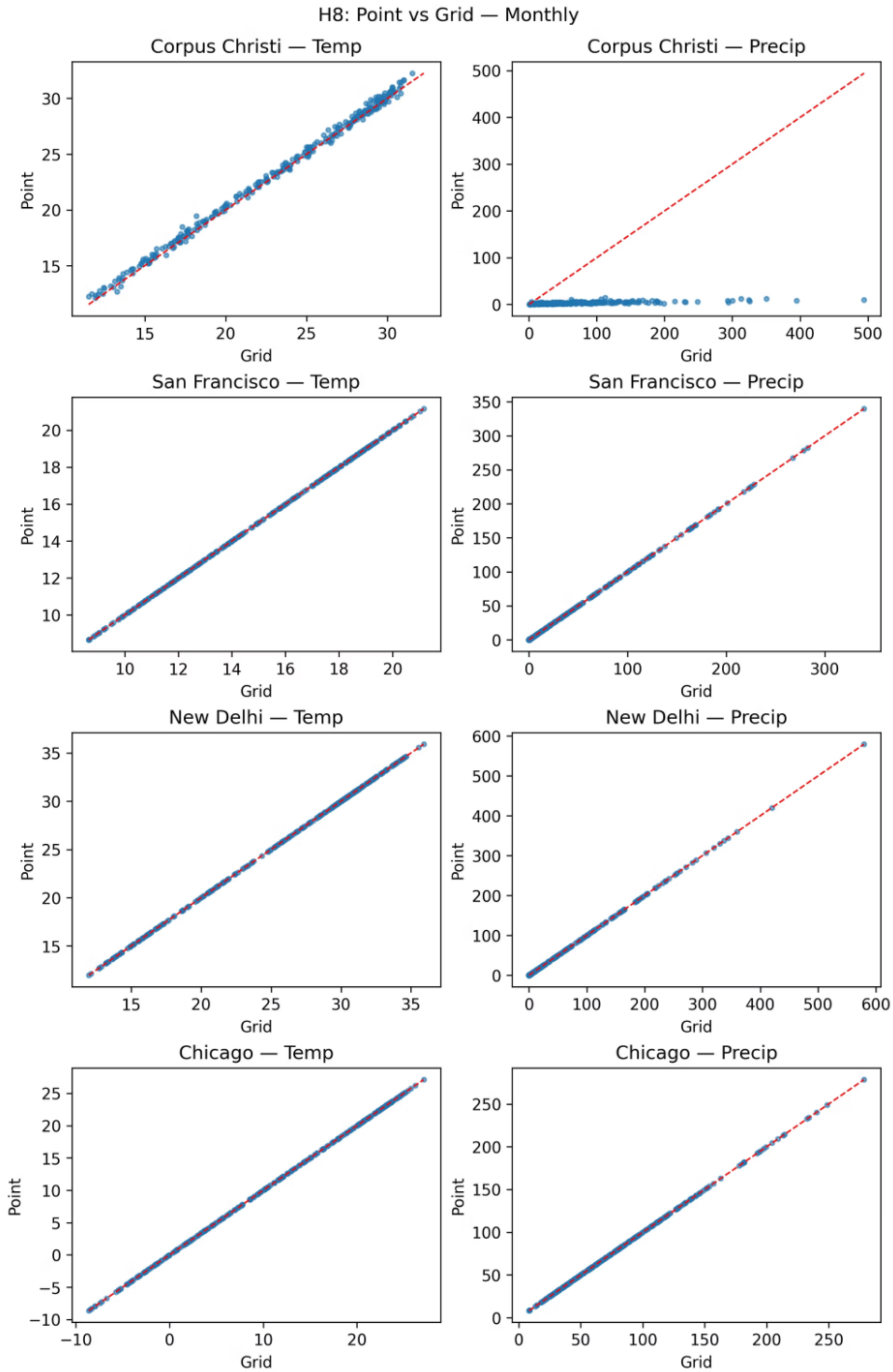

Figure 8b. *Example Scatter: Grid vs Point Data for Temperature and Precipitation (H8b – Point vs Grid Consistency).*

Conclusion: Proved. *Temperature fits much more closely to gridded data since it is more of a smooth spatial variety whereas precipitation has more localized variation.*

4. Summary of Findings

Hypothesis	Supported	Conclusion
H1	Proved	Monsoon trigger confirmed regionally.
H2	Proved	Coastal moderation strongly supported.
H3	Proved	SST follows annual solar cycle.
H4	Proved	Monsoon rainfall highly seasonal.
H5	Partially Proved	Temperature normality holds mostly.
H6	Disproved	Convective feedback regionally valid.
H7	Proved	Rising CO_2 linked to warming.
H8	Proved	Temperature more spatially consistent than precipitation.

5. Conclusion

Eight hypotheses were tested with six hypotheses statistically verified, one supported (H5) and one globally not consistent but regionally (H6).

The results all indicate important principles of climatic workings:

- Convection occurs as a result of pressure and temperature (H1).
- Vegetation refers to temperature moderation of Oceans (H2).
- The process of seasonal cycle is regulated by solar forcing (H3).
- The seasonability of rain falls characterizes monsoon regimes (H4).
- There is persistent warming that is caused by the CO 2 increase (H7).
- Gridded datasets are good in describing temperature and not precipitation (H8).

These findings support the essence of integrating statistical data with climatological explanation in order to confirm scientific assumptions.

6. Division of Work

• Nikhilesh Verma – Data processing, hypothesis testing, report writing, project coordination.
• Abhishek Joshi – Visualization, statistical validation, code management, presentation design.
• Both – Interpretation of results, preparation of final deliverables, and quality review.

7. References

[1] NOAA PSL, *NOAA Physical Sciences Laboratory Reanalysis Datasets*, Boulder, CO, USA. Available: https://psl.noaa.gov

[2] NCEP/NCAR, *Reanalysis 1: Atmospheric Variables (Temperature, Precipitation, Pressure)*, NOAA ESRL, 2024.
[3] CPC, *Global Unified Gauge-Based Analysis of Daily Precipitation*, NOAA, 2024.
[4] NOAA, *Optimum Interpolation Sea Surface Temperature (OISST) Dataset*, 2024.
[5] NASA GISS, *Global Surface Temperature and CO_2 Data (GISTEMP v4)*, 2024.
[6] NOAA GML, *Mauna Loa CO_2 Monthly Mean Data*, 2024.
[7] NOAA ESRL, *Multivariate ENSO Index (MEI)*, 2024.
[8] ECMWF, *ERA5 Reanalysis Data*, Copernicus Climate Data Store, 2024.
[9] Python Software Foundation, *Python Language Reference, Version 3.10*, 2024.
[10] P. Virtanen *et al.*, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nat. Methods,* vol. 17, no. 3, pp. 261–272, 2020.
[11] W. McKinney, "Data Structures for Statistical Computing in Python," *Proc. 9th Python in Science Conf.,* 2010.
[12] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," *Comput. Sci. Eng.,* vol. 9, no. 3, pp. 90–95, 2007.
[13] S. Seabold and J. Perktold, "Statsmodels: Econometric and Statistical Modeling with Python," *Proc. 9th Python in Science Conf.,* 2010.
[14] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.,* vol. 12, pp. 2825–2830, 2011.