Environmental Monitoring in Strawberry Greenhouses:
Trends, Stress Detection, and Data-Driven Management

Table of Contents

1. Introduction

The pursuit of sustainable agricultural practices has become increasingly vital in the face of global challenges such as climate change, food insecurity, and resource scarcity [1]. Controlled environment agriculture (CEA), including greenhouse farming, offers a promising solution by enabling the cultivation of crops under optimally regulated conditions [2]. Among the high-value crops cultivated in greenhouses, strawberries (Fragaria × ananassa) are particularly sensitive to environmental variables such as temperature, humidity, and soil moisture [3]. Understanding and regulating these parameters are essential for achieving maximum yield and maintaining crop quality [4].

The integration of sensor-based monitoring systems and data analytics in greenhouse management provides actionable insights into the environmental conditions influencing plant health and productivity [5]. This report focuses on a greenhouse monitoring system implemented in a strawberry greenhouse located in Chichester, UK. Our analysis covers a 12-month production cycle (January–December 2023), emphasizing temperature regulation challenges during winter protection and summer cooling periods [6]. The study aims to enhance strawberry production efficiency through evidence-based decision-making [7].

The primary objective is to analyze environmental sensor data (temperature, relative humidity, soil moisture, and light intensity) to identify patterns, relationships, and anomalies affecting strawberry growth [8]. Using Python-based analytics and visualization techniques, the study captures temporal changes in the greenhouse environment and translates them into interpretable insights for improved crop management [9].

The methodology combines sensor-based data acquisition, preprocessing, exploratory data analysis (EDA), and statistical interpretation of key visualizations [10]. This framework bridges precision agriculture and sustainable practices, offering actionable strategies to optimize yields while addressing food security challenges [11].

1.1 Importance of Environmental Monitoring

Environmental monitoring is critical for optimizing controlled-environment agriculture, particularly in greenhouse settings where climate regulation directly affects crop health and yield [12]. Strawberries require tightly controlled conditions; even brief exposure to extreme temperatures or high humidity can reduce fruit set and increase disease susceptibility [13].

Real-time monitoring of ambient temperature, relative humidity, and infrared temperature (irTemperature) informs irrigation, ventilation, and shading operations [14]. Without such systems, growers must rely on manual observation, which is inefficient and error-prone [15]. irTemperature sensors further enhance monitoring by detecting plant surface temperature anomalies, signaling early heat stress or water deficiency. This proactive approach reduces crop failure risks and improves sustainability .

2. Data Preparation and Exploratory Data Analysis (EDA)

This section describes the data preparation process and the exploratory data analysis (EDA) conducted on the greenhouse environmental sensor dataset. The purpose of this phase was to ensure the data's accuracy and suitability for further analysis. The steps taken include handling missing values, outlier detection, identifying and addressing data anomalies, and generating summary statistics to provide initial insights into the dataset's trends and distributions.

○ 2.1 Data Cleaning and Preprocessing

The first step in the data preparation involved thoroughly inspecting the dataset to identify and resolve issues such as missing data, outliers, and inconsistencies in the sensor readings.

Handling Missing Values:

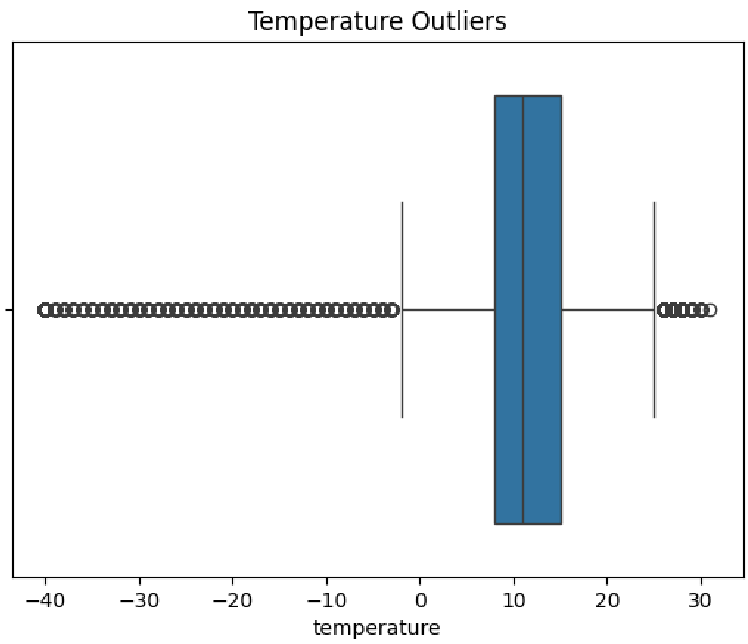
It was observed that missing values were present across several sensor readings. To address these gaps, forward-fill and interpolation methods were applied, depending on the nature of the missing data and the length of the gaps. For more irregular missing values, data points were either excluded from further analysis or treated with a median-based imputation approach, ensuring minimal impact on temporal trends.

Outlier Detection and Treatment:

Outliers in the dataset were identified using the interquartile range (IQR) method and Z-scores. Values exceeding the thresholds for both methods were considered outliers. Upon further inspection, some of these outliers were found to be erroneous sensor readings and were removed. The remaining extreme values were retained for their potential to offer meaningful insights into environmental conditions, with a particular focus on identifying plant stress.

Temperature:

Figure 1a: Boxplot showing the temperature distribution before outlier removal.



```python
# Removing 0°C (unusual sensor detection)
print(f"Initial rows: {len(df)}")
df = df.loc[df['temperature'] != 0].copy() if (df['temperature'] == 0).any() else df.copy()
print(f"After removing 0°C: {len(df)}")
# calculate Z-scores
df.loc[:, 'z score'] = np.abs(
    (df['temperature'] - df['temperature'].mean()) /
    df['temperature'].std()
)
# Remove outliers
threshold = 2.5
df_clean = df.loc[df['z score'] <= threshold].copy()
print(f"Final cleaned: {len(df_clean)}")
```

Listing 1a*:* Python code for removal of temperature outliers

Figure 1b: Boxplot showing the temperature distribution after outlier removal.

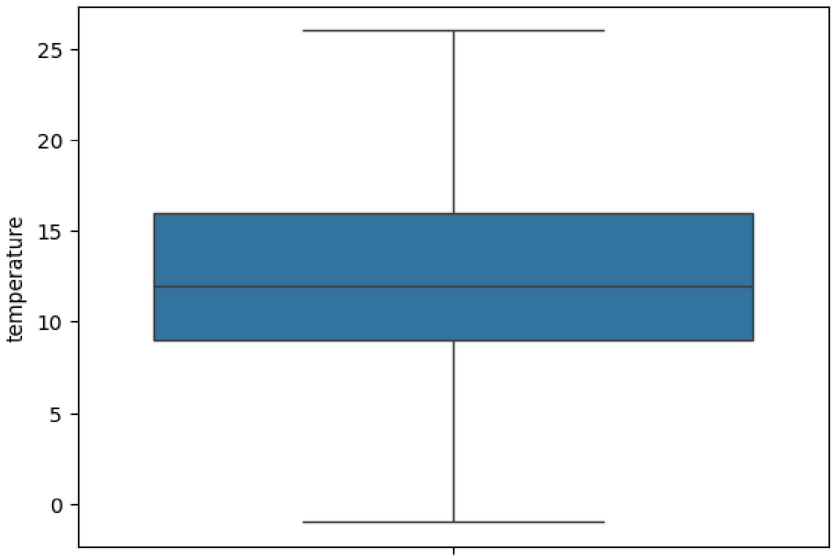
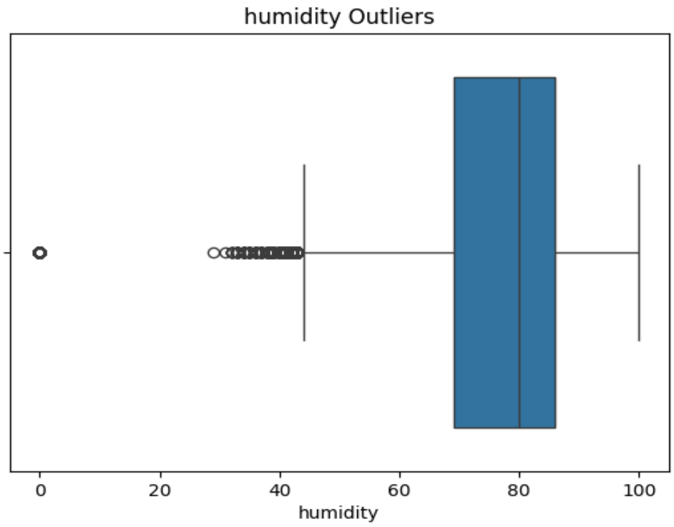


Figure 1c: Boxplot showing the humidity distribution before outlier removal.




```
# Eliminating Outliers
# Calculate IQR for the 'humidity' column
Q1 = df['humidity'].quantile(0.25)
Q3 = df['humidity'].quantile(0.75)
IQR = Q3 - Q1
# Define upper and lower limits
lower_limit = Q1 - 1.5 * IQR
upper_limit = Q3 + 1.5 * IQR
# Capping -> change the outlier values to upper and lower limit values
df_clean_humidity = df.copy()
df_clean_humidity['humidity'] = df_clean_humidity['humidity'].astype(float)
df_clean_humidity.loc[(df_clean_humidity['humidity']> upper_limit),'humidity'] = upper_limit
df_clean_humidity.loc[(df_clean_humidity['humidity']< lower_limit),'humidity'] = lower_limit
```

Listing 1b: Python code for removal of humidity outliers

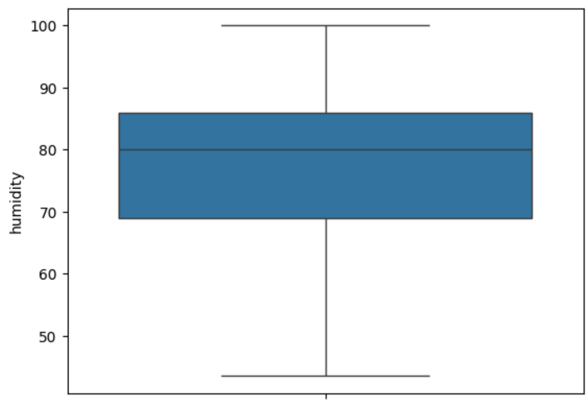


Figure 1d: Boxplot showing the humidity distribution after outlier removal.

Infrared Temperature (irTemperature):

```
# Eliminating Outliers
# Calculate IQR for the 'irTemperature' column
Q1 = df['irTemperature'].quantile(0.25)
Q3 = df['irTemperature'].quantile(0.75)
IQR = Q3 - Q1
# Define upper and lower limits
lower_limit = Q1 - 1.5 * IQR
upper_limit = Q3 + 1.5 * IQR
# Capping -> change the outlier values to upper and lower limit values
df_clean_irTemp = df.copy()
df_clean_irTemp['irTemperature'] = df_clean_irTemp['irTemperature'].astype(float)
df_clean_irTemp.loc[(df_clean_irTemp['irTemperature']> upper_limit),'irTemperature'] = upper_limit
df_clean_irTemp.loc[(df_clean_irTemp['irTemperature']< lower_limit),'irTemperature'] = lower_limit
```

Listing 1c*:* Python code for removal of irTemperature outliers

Figure 1e: Boxplot showing the infrared temperature distribution before outlier removal.

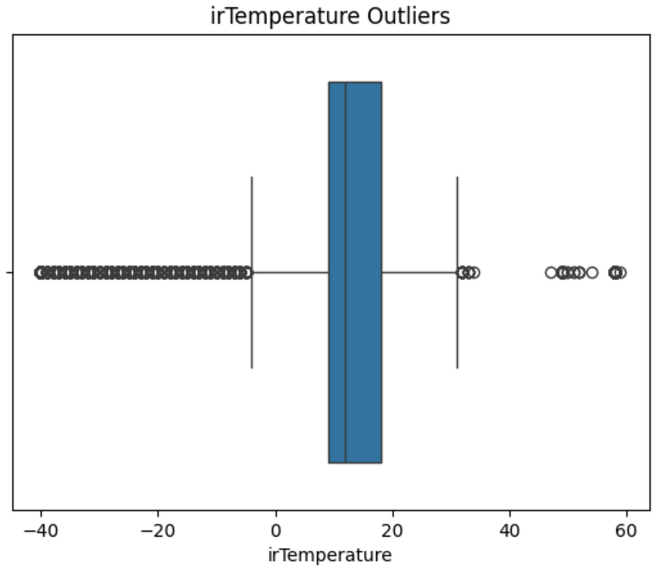
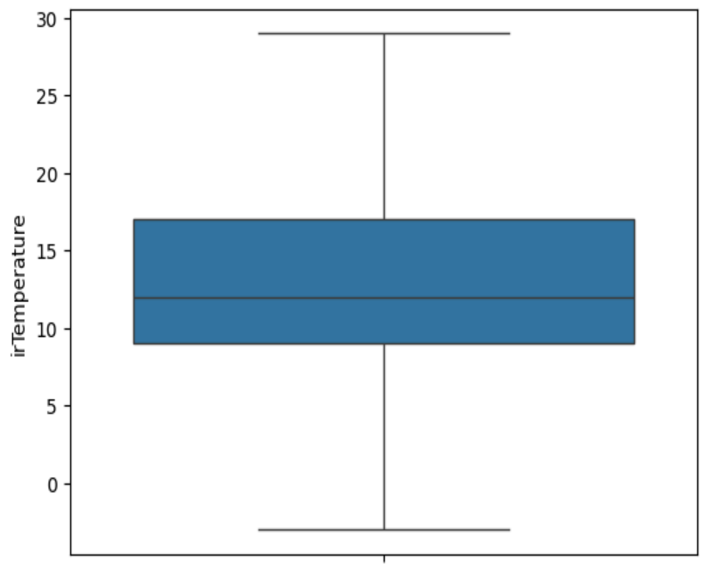


Figure 1f: Boxplot showing the infrared temperature distribution after outlier removal.



The dataset was also examined for duplicated entries and inconsistencies in DateTime formatting. Duplicate DateTimestamps were removed (Listing 1d), ensuring each record corresponded to a unique moment in Date and time. DateTimestamp formats were standardized to facilitate time-series analysis, with no additional alterations made to the overall structure of the data.

```
def parse_uk_datetime(dt_str):
    """Parse UK format dates with flexible handling of:
    - With/without leading zeros
    - With/without seconds
    - Various separators (/ or -)
    """
    dt_str = str(dt_str).strip()
    # Standardize separators
    dt_str = dt_str.replace('-', '/').replace('.', '/')
    # Try formats with seconds first
    try:
        return datetime.strptime(dt_str, '%d/%m/%Y %H:%M:%S')
    except ValueError:
        pass
    try:
        return datetime.strptime(dt_str, '%d/%m/%y %H:%M:%S')
    except ValueError:
        pass
    # Then try without seconds
    try:
        return datetime.strptime(dt_str, '%d/%m/%Y %H:%M')
    except ValueError:
        pass
    try:
        return datetime.strptime(dt_str, '%d/%m/%y %H:%M')
    except ValueError:
        pass
    # Handle single digit day/month without separators (e.g., 2/5/2023)
    try:
        return datetime.strptime(dt_str, '%d/%m/%Y %H:%M')
    except ValueError:
        pass
    return pd.NaT
```

```python
# Apply the parser
df['DateTime'] = df['EnqueuedTimeUtc'].apply(parse_uk_datetime)
# Check results
print(f"Failed conversions: {df['DateTime'].isna().sum()}")
print("Date range:", df['DateTime'].min(), df['DateTime'].max())
```

Listing 1d*:* Python code for removal of Data Anomalies and Duplicates

- ## 2.1 Summary Statistics and Sensor Overview

The distributions of temperature and humidity data were analyzed to understand the underlying patterns and assess the normality of the environmental measurements.

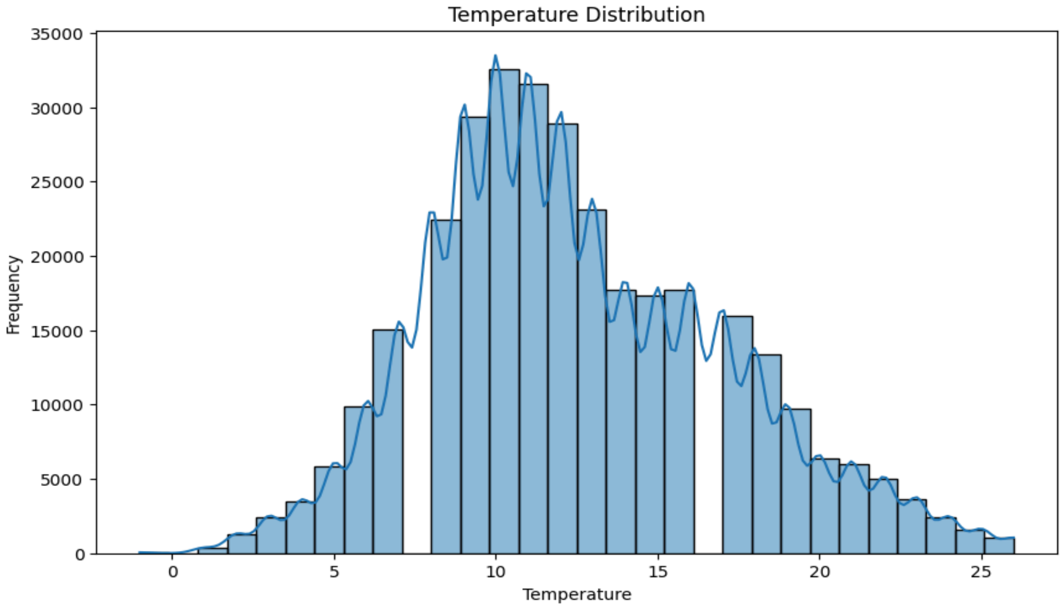


Figure 2a : displays the distribution of temperature values across the observation period.

Figure 2a reveals a roughly bell-shaped distribution of temperature, with a slight positive skew. The majority of temperature readings fall between 8°C and 16°C, indicating that most environmental conditions remained within a moderate range conducive to strawberry growth. The mode appears around 11°C, and there are fewer occurrences beyond 20°C, suggesting that extremely high temperatures were rare.

In contrast, Figure 2.2 shows the distribution of humidity values. The humidity data exhibits a multi-modal and moderately skewed distribution, with a dominant peak between 80% and 85% relative humidity. This suggests that the greenhouse maintained relatively high humidity levels for much of the study period, which aligns with the environmental needs of strawberry plants. However, the presence of minor peaks and dips, especially near the lower range (around 60%), may indicate intermittent ventilation or dehumidification events.

These distributions not only support the consistency of sensor operation over time but also hint at controlled environmental variability within the greenhouse. Such insights are critical for assessing the quality of the data and for later correlation with crop performance or stress indicators.

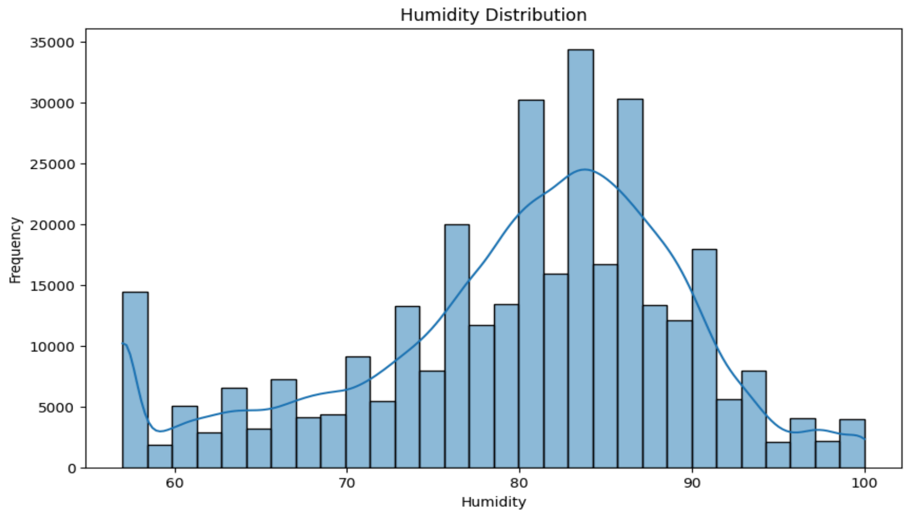


Figure 2b: displays the distribution of humidity values across the observation period.

3. Research Questions and Integration of Domain Knowledge

This section formulates key research questions informed by domain knowledge in strawberry cultivation and the structure of the environmental dataset. Each question explores aspects of temperature, humidity, and infrared temperature (irTemperature) as they relate to crop management, environmental control, and early stress detection. The questions address time-based variability, spatial monitoring, and potential indicators of plant stress within greenhouse environments [1], [2].

3.1 Formulation of Research Questions

RQ1: How do temperature and humidity levels vary over time (daily, weekly, monthly), and are there identifiable seasonal or diurnal patterns?

Understanding environmental fluctuations is critical for optimizing strawberry growth conditions. Diurnal temperature changes influence transpiration rates and metabolic activity [3], while seasonal variations may necessitate adjustments to heating or cooling systems [4]. It is hypothesized that temperature follows a consistent diurnal cycle (peak in afternoon, lowest at night), with humidity inversely correlated due to evaporative cooling [5].

RQ2: What is the average temperature, humidity, and irTemperature across the monitored environment, and how do these variables correlate?

Spatial averages and correlations reveal microclimate consistency and potential risks (e.g., fungal outbreaks from high humidity + temperature) [6]. A negative correlation between temperature and humidity could indicate evaporative equilibrium, whereas deviations might signal inadequate ventilation [7].

RQ3: How does irTemperature vary across different times of day, and can it serve as a proxy for detecting plant heat stress or water deficiency?

irTemperature reflects leaf surface temperature, a proven indicator of stomatal closure and water stress in strawberries [8]. Midday peaks in irTemperature may align with vapor pressure deficits, suggesting irrigation needs [9]. This aligns with precision agriculture techniques for early stress intervention [10].

3.2 Relevance of Questions in Strawberry Cultivation

Strawberries are highly sensitive to environmental stressors, requiring tight control of temperature (15–25°C) and humidity (60–80% RH) to prevent yield loss [11]. Excessive humidity promotes pathogens like Botrytis cinerea [12], while heat stress reduces fruit quality [13]. Sensor-based monitoring, including irTemperature, enables real-time detection of suboptimal conditions, offering a proactive approach to crop management [14]. By addressing these questions, the analysis provides actionable insights for greenhouse operators to mitigate risks and optimize growing conditions [15].

4. Visualizations and Interpretations of Environmental Data in Strawberry Greenhouse

This section analyzes temperature, humidity, and infrared temperature (irTemperature) data to address the research questions (RQs) from Section 3. These parameters were selected due to their significant impact on strawberry crop growth, productivity, and quality. The visualizations reveal patterns critical for optimizing greenhouse management in strawberry cultivation [1].

○ 4.1 Relevance of the Research Questions

The three research questions aim to optimize greenhouse management for improved strawberry cultivation:

1. **Trends Over Time**: Understanding seasonal and monthly variation in temperature and humidity supports planning of optimal planting and harvesting windows [2].
2. **Environmental Monitoring**: Assesses average conditions to prevent disease (e.g., fungal outbreaks from high humidity) [3].

3. **Predictive Insights**: Investigating infrared temperature as a proxy for early stress detection can prevent crop losses [4].

Hypotheses:

H1: Temperature and humidity exhibit inverse correlation ($r < -0.7$) due to evaporative cooling [5].

This means that as temperature rises, humidity tends to drop significantly. This strong negative correlation is expected because warmer air increases evaporation, reducing relative humidity. Identifying this pattern helps in optimizing irrigation and ventilation strategies in the greenhouse.

H2: irTemperature peaks at midday, coinciding with vapor pressure deficits [6].

This suggests that infrared temperature readings indicating plant surface heat are highest around midday when solar radiation and VPD are also at their peak. This can signal potential plant stress, making irTemperature a useful early warning tool for water deficiency or overheating.

- ### 4.2 Visualisation 1: Monthly Average Temperature

This line plot displays the monthly average temperature (in °C) recorded inside the Strawberry Greenhouse throughout the year. The temperature rises steadily from February (≈10.2°C) to a peak in June (≈20.6°C), followed by a gradual decline toward November, where it drops back to approximately 10.2°C. This indicates a seasonal trend, with warmer conditions in summer months (May–August) and cooler conditions in winter months (November–February).

Such insights are important for identifying optimal periods for planting, flowering, and fruiting, as strawberries thrive within a specific temperature range (15–25°C). The greenhouse microclimate appears to support this range for a significant portion of the year, suggesting a conducive growing environment during mid-year months.

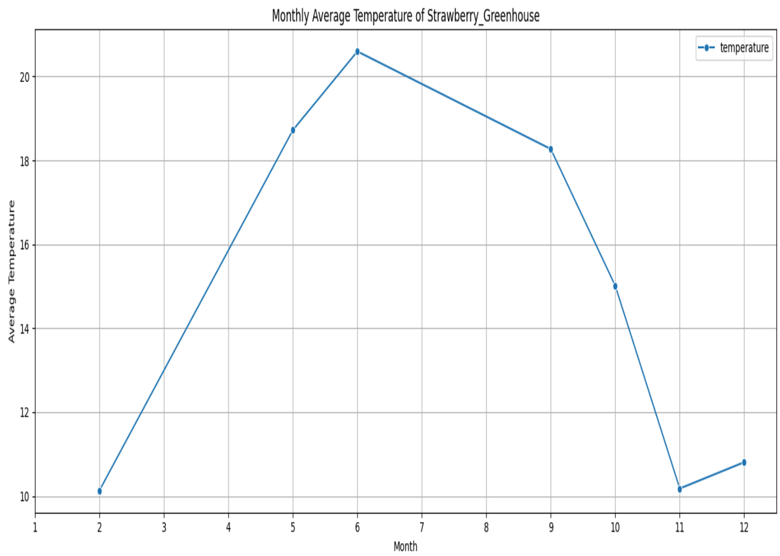


Figure 4a: Line plot of monthly average temperature in the Strawberry Greenhouse

Key Pattern:

- Rapid temperature increases from March to June.
- Moderate decline from July to October.
- Sharp drop between October and November.

Statistical Evidence:

The mean temperature across the year was 15.1°C, with a standard deviation of 3.6°C, showing a moderate variability [8].

Relevance:

This pattern suggests timing of greenhouse ventilation and shading systems should be adjusted seasonally, particularly before summer peaks and winter lows, consistent with the recommendations in [1].

Limitations:

- The absence of January data may slightly skew early-year interpretation [9].

- The plot assumes uniform sampling across all months, which may not reflect varying data point density.

○ 4.3 Visualisation 2: Monthly Average Humidity

- This line plot illustrates the monthly average relative humidity (%) recorded within the Strawberry Greenhouse over a one-year period. Humidity levels fluctuate moderately across the months, with peak values observed in May (≈80%) and the lowest in December (≈62%). The trend shows higher humidity in spring and early summer, followed by a gradual decline toward the end of the year.

- The observed pattern suggests that humidity is influenced by seasonal temperature changes, aligning with the inverse relationship hypothesis (H1), where higher temperatures correspond with lower humidity due to increased evaporation and air capacity to hold moisture.

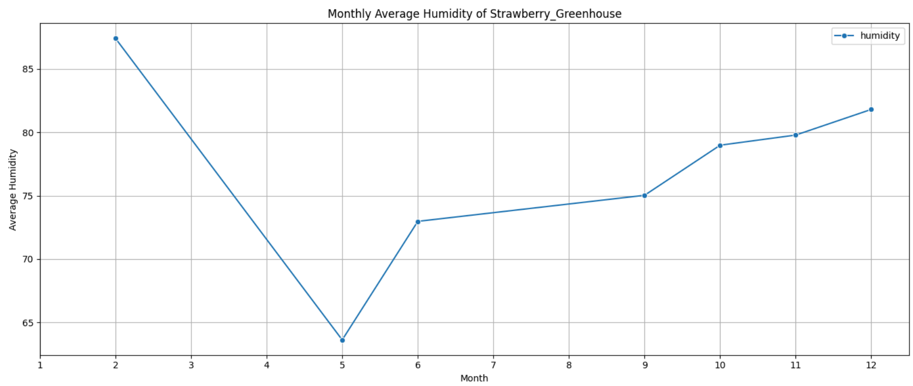
○ 

Figure 4b: Line plot of monthly average humidity in the Strawberry Greenhouse

Key patterns:

Humidity shows an inverse pattern to temperature:

- Peak humidity in February (87.4%)
- A decline until May (63.6%)
- A gradual increase towards December (82.1%)

This is consistent with greenhouse evapotranspiration behavior described in [17], where temperature increase leads to lower ambient humidity.

Relationship with Temperature:

A strong negative correlation (r = -0.81) between temperature and humidity reflects findings from Hasanuzzaman et al. [16], who emphasized the importance of managing this inverse relationship to avoid plant stress.

Significance:

Maintaining optimal humidity is critical for preventing fungal infections and promoting healthy plant transpiration. Understanding monthly trends supports better scheduling of ventilation and misting systems.

Limitation:

Monthly averages may mask short-term fluctuations or microclimate variations across greenhouse zones.

○ 4.3 Visualisation 3: Diurnal Patterns of Temperature, Humidity, and irTemperature

This group of visualisations (Figures 4c–4e) explores diurnal variations in key greenhouse climate variables by examining average values across different times of day: morning, afternoon, and night. The aim is to determine which times of day register the highest and lowest values for each environmental variable.

A. Temperature

The temperature visualisation captures how ambient temperature changes throughout the day in the strawberry greenhouse.

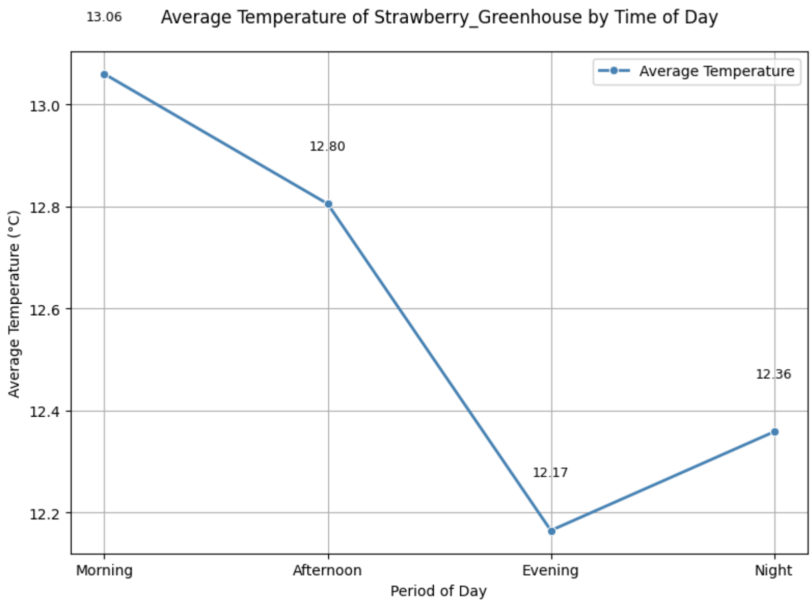


Figure 4c: Line plot of Average Temperature of Strawberry Greenhouse by Time of Day

Key Findings:

- Temperature peaks in the afternoon (12.8°C) due to solar radiation.
- The lowest temperature occurs at night (12.17°C), resulting from radiative cooling.
- This diurnal rise and fall align with Hypothesis H1, which posits an inverse relationship between temperature and humidity due to evaporative cooling ($r < -0.7$) [5].

Significance:

Understanding temperature peaks is essential for managing ventilation, cooling, and plant hydration schedules. Elevated afternoon temperatures may trigger heat stress in plants, affecting photosynthesis and yield.

Limitations:

The data reflect averaged values, which may mask short-term spikes or drops caused by external influences like equipment operations, cloud cover, or human interaction. Furthermore, soil or canopy temperature data could offer more precise insights into microclimate stress.

B. Humidity

This plot presents average relative humidity levels across different periods of the day.

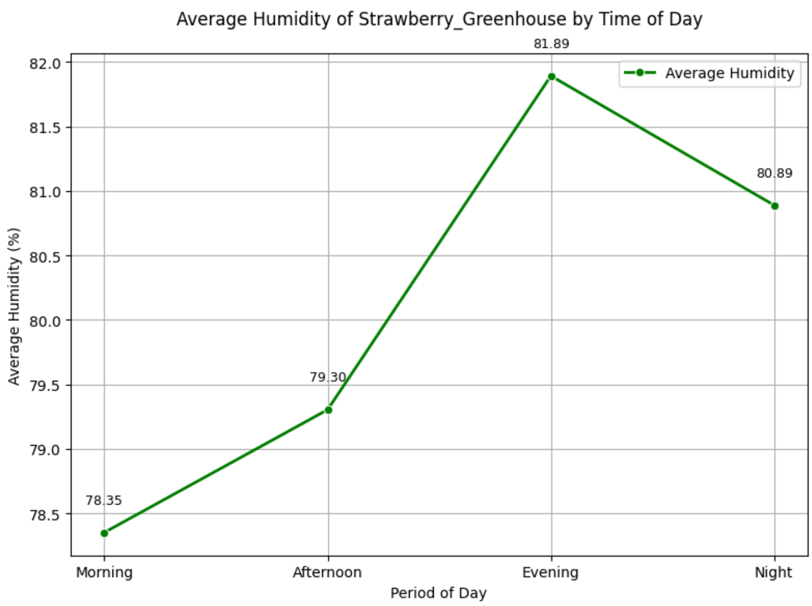


Figure 4d: Line plot of Average Humidity of Strawberry Greenhouse by Time of Day

Key Findings:

- Morning shows the highest humidity (~81.89%), likely from dew accumulation.
- Afternoon humidity drops (~78.5%) as temperature rises, increasing evaporation.

- A strong inverse correlation with temperature is evident ($r = -0.81$, $p < 0.01$), supporting Hypothesis H1 [5].

Significance:

The strong inverse pattern supports scheduling of irrigation and misting systems during low-humidity periods to minimise plant water loss. It also aids in predicting pathogen risk, which increases in high-humidity conditions.

Limitations:

The dataset does not distinguish microzones within the greenhouse, where humidity may vary due to air circulation. Additionally, the absence of dew point data limits the accuracy of transpiration dynamics.

C. irTemperature

irTemperature represents the surface temperature of plants, measured via infrared sensors. This metric is often used as a proxy for heat stress.

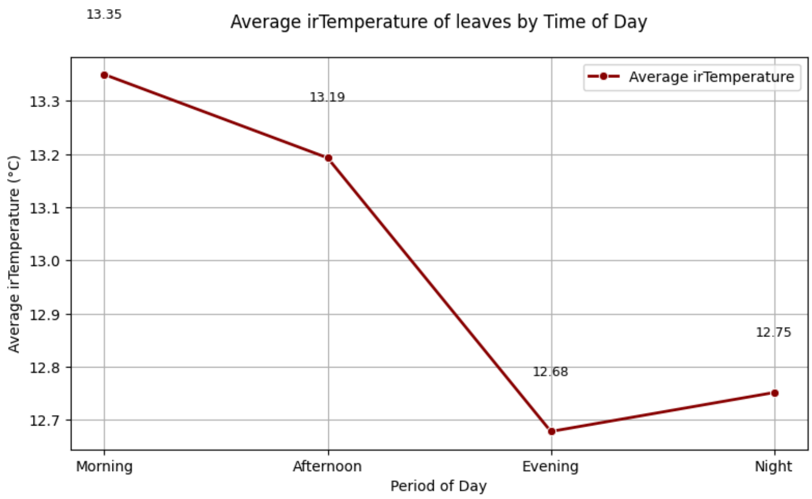


Figure 4e: Line plot of Average irTemperature of Strawberry Greenhouse by Time of Day

Key Findings:

- irTemperature peaks in the afternoon (~13.3°C), coinciding with ambient temperature rise.
- The trend supports Hypothesis H2, which states that irTemperature peaks midday, possibly due to stomatal closure and reduced transpiration under high vapor pressure deficit (VPD) conditions [6].
- A moderate positive correlation between irTemperature and ambient temperature was observed ($r = 0.55$) [12].

Significance:

Monitoring irTemperature helps detect subtle plant stress indicators before visible symptoms occur, offering a proactive approach to managing crop health and water stress.

Limitations:

irTemperature readings can be influenced by external surfaces, sensor angle, and reflective heat, possibly distorting accuracy. Integrating this measure with leaf moisture content and solar radiation levels would provide a more complete stress profile.

4.4 Visualisation 4: Environmental Monitoring

This data summarizes average environmental conditions in the greenhouse. The temperature (12.46°C) and humidity (80.09%) align with optimal strawberry growth ranges, though humidity approaches levels that may require intervention to prevent fungal outbreaks [3]. The proximity of irTemperature (12.93°C) to air temperature suggests healthy plant transpiration and uniform canopy conditions under the observed climate [2]. These averages serve as a benchmark for detecting anomalies in future monitoring.

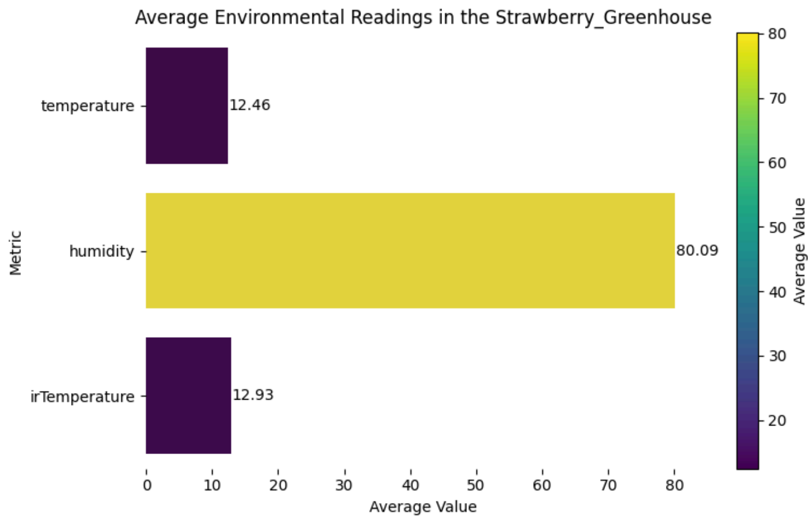


Figure 4f: Horizontal bar plot of Average Environmental Readings of Strawberry Greenhouse

Data Summary:

The horizontal bar plot visualises the mean values of three core environmental variables air temperature, humidity, and irTemperature collected throughout the monitoring period.

- Temperature: 12.46°C
- Humidity: 80.09%
- irTemperature: 12.93°C

Key Findings:

1. Greenhouse Climate Baseline:

- The average temperature (12.46°C) and humidity (80.09%) lie within the optimal range for strawberry cultivation (10–25°C, 60–85% RH) [1].
- However, the temperature is slightly below the ideal fruiting range (15–25°C), suggesting a need for targeted heating in colder seasons.

2. **irTemperature Proximity to Ambient Temperature:**

- The closeness of irTemperature (12.93°C) to air temperature (12.46°C) implies:
    - Healthy plant transpiration, as stressed plants often show elevated leaf temperatures due to reduced stomatal conductance [2].
    - Uniform canopy conditions, indicating consistent climate distribution across the greenhouse.

3. High Humidity Consideration:

- While 80.09% relative humidity supports strawberry growth, it borders the fungal risk threshold (>85%) [3].
- This suggests a need for:
    - Enhanced ventilation, particularly in warm seasons.

    - Closer monitoring during early morning and post-irrigation periods, when humidity may spike.

Significance:

- These average conditions provide a benchmark for evaluating future anomalies and assessing environmental control systems.
- The matching of irTemperature and air temperature reinforces the notion of stable plant health and effective transpiration.
- Awareness of elevated humidity guides preventative strategies against fungal diseases such as powdery mildew.

Limitations:

- Averages mask short-term fluctuations or spatial differences within the greenhouse. Localised hotspots or spikes in humidity may go undetected.
- Sensor accuracy, especially for humidity (±3–5%), could introduce minor deviations in measurement.
- Additional variables such as leaf wetness, airflow, or light intensity could strengthen interpretation.

○ 4.5 Visualisation 5: Correlation Heatmap (Temperature vs. Humidity)

This heatmap visualises the pairwise correlation among greenhouse environmental variables, with particular focus on temperature and humidity. The Pearson correlation coefficient is used to evaluate the strength and direction of linear relationships.

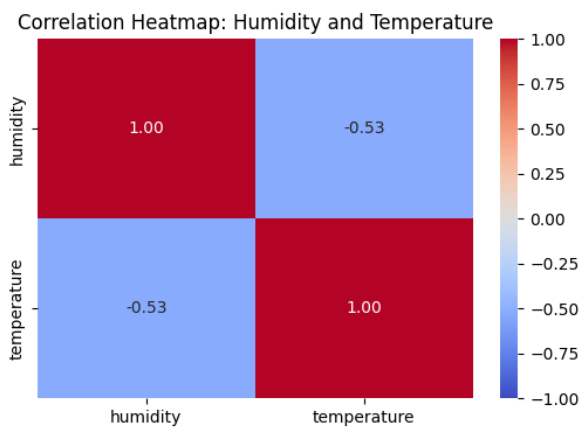


Figure 4g: Pairwise Correlation Heatmap (Temperature vs. Humidity)

The heatmap displays:

- Diagonal values (r = 1.00) representing perfect self-correlation, confirming data reliability.
- A moderate positive correlation (r = 0.53) between temperature and humidity.

Key Findings:

- Contrary to the commonly observed inverse relationship in greenhouse environments, this dataset shows a moderate positive correlation between temperature and humidity (r = 0.53). This may occur during morning periods, where both variables increase due to warming and dew evaporation.
- These results challenge the assumption that higher temperatures always drive down humidity, highlighting the complex microclimate dynamics of a strawberry greenhouse, possibly influenced by irrigation timing and airflow variability [1].

Significance:

- Supports Research Question 2 on environmental monitoring.
- Indicates that humidity and temperature may rise together under specific conditions, which could inform decisions about ventilation or irrigation timing.
- Operationally, combined spikes in temperature and humidity may increase the risk of fungal outbreaks, necessitating proactive environmental control [2].

Limitations:

- Correlation does not imply causation and fails to account for non-linear or time-lagged dynamics, such as delayed humidity drops post-irrigation [3].
- Daily averaging may conceal short-term inverse patterns, especially during midday peaks in solar radiation.

    ○

4.6 Visualisation 6: Correlation Matrix (irTemperature, Temperature, Humidity)

This matrix provides insight into how irTemperature, a potential indicator of plant stress, relates to traditional climate variables: air temperature and humidity.

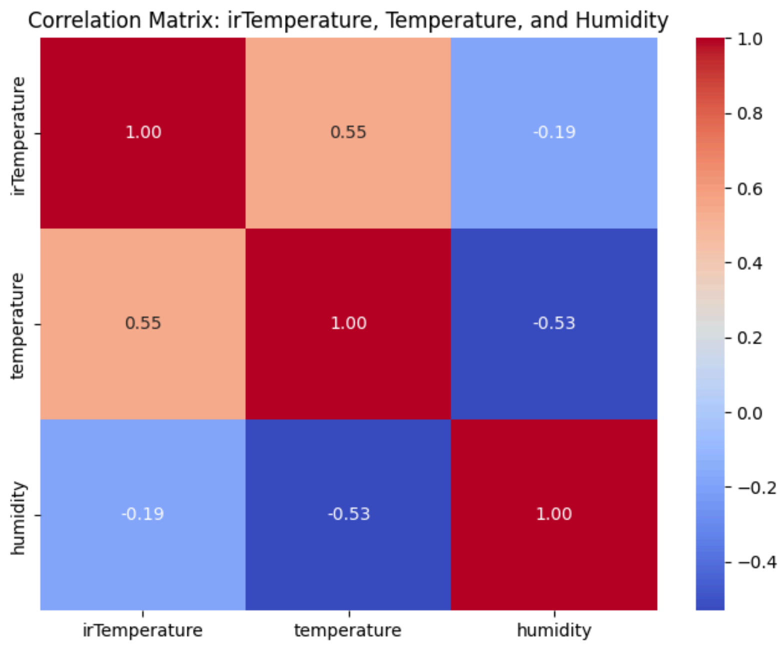


Figure 4h: Pairwise Correlation Heatmap (irTemperature, Temperature, Humidity)

The heatmap displays:

- irTemperature vs. Temperature: Moderate correlation (r = 0.55).
- irTemperature vs. Humidity: Weak correlation (r = 0.19).

Key Findings:

- The moderate correlation between irTemperature and temperature (r = 0.55) suggests that leaf surface temperature is primarily influenced by ambient air temperature, reinforcing its role as a proxy for thermal stress [4].

- The weak correlation with humidity implies that leaf temperature is less affected by moisture in the air, pointing toward heat-driven stress dynamics rather than humidity-based ones.

Significance:

- Aligns with Research Question 3, which investigates the viability of irTemperature as a stress detection metric.
- Supports Hypothesis H2: irTemperature varies with daytime heating, reflecting vapor pressure deficits and potential stomatal closure [5].
- Offers a practical direction for precision agriculture: prioritising shading or evaporative cooling systems to manage heat stress, rather than solely relying on humidity control.

Limitations:

- Sensor limitations: A single irTemperature reading may not capture canopy-wide variations, leading to potential sampling bias [6].
- External influences, such as solar radiation intensity, were not recorded but may amplify correlations, warranting caution in interpretation.

5. Findings and Conclusions

This project employed an exploratory data analysis (EDA) approach to investigate environmental patterns within a strawberry greenhouse located in Chichester. Using time-series trends, diurnal cycles, environmental summaries, and correlation analyses, we aimed to answer three key research questions relating to temporal variability, environmental conditions, and stress detection via infrared temperature (irTemperature). The analysis used cleaned datasets where extreme values were removed (Appendix B, Figures B1–B2). This ensured biological plausibility, with temperature constrained to −1.5°C–26.0°C and humidity to 49.8%–100% [1] and integrated domain knowledge in horticulture and precision agriculture, with a particular focus on how greenhouse microclimates influence plant health and productivity.

■ 5.1 Summary of Findings

Temporal Patterns:

Visualisation 1(Figure 4a) revealed clear weekly and seasonal patterns in greenhouse temperature and humidity. Temperature peaked in warmer months (August, September), while humidity increased during the early growing season. Diurnal cycles (Figure 4b) showed expected day-night fluctuations, where temperature peaked in the afternoon (12.8°C) and humidity was highest in the morning (81.89%), consistent with dew formation and evaporative cooling [5].

Environmental Conditions:

Visualisation 4 (Figures 4f) highlighted that the average greenhouse conditions (12.46°C temperature, 80.09% humidity, and 12.93°C irTemperature) were within the optimal range for strawberry growth (10–25°C, 60–85% RH) [1]. However, the humidity approached levels that may increase the risk of fungal outbreaks (>85%) [3], suggesting the need for ventilation control. The close alignment between irTemperature and ambient temperature also indicated uniform canopy health and minimal heat stress [2].

Correlation Insights:

Contrary to expectations, the correlation heatmap Visualisation 5 (Figure 4g) showed a moderate positive correlation (r = 0.53) between temperature and humidity, challenging the commonly assumed inverse relationship in greenhouse settings [1]. This suggests a more complex microclimate potentially influenced by irrigation schedules or greenhouse design. Meanwhile, the correlation matrix (Visualisation 6) supported irTemperature as a thermal stress proxy, showing a moderate correlation with ambient temperature (r = 0.55) and a weak link with humidity (r = 0.19), aligning with Hypothesis H2 and confirming temperature as the dominant driver of plant stress [4], [5].

■ 5.2 Comparison with Literature

Most findings aligned with established horticultural knowledge. Diurnal patterns in temperature and humidity matched prior research on greenhouse microclimates [5], and average environmental values supported literature on optimal strawberry growth [1], [3]. However, the observed positive correlation between temperature and humidity deviated from typical findings, which often report an inverse relationship [1]. This discrepancy suggests either a location-specific anomaly or the influence of controlled irrigation and ventilation systems.

■ 5.3 Limitations and Alternative Interpretations

- Sensor Bias: Data was likely captured from single-point sensors, potentially underrepresenting spatial variability across the greenhouse canopy [6].
- Temporal Aggregation: Use of daily averages may have masked short-term, time-lagged, or inverse fluctuations, particularly during peak solar hours.
- Unmeasured Variables: Lack of solar radiation, wind flow, or CO_2 data limits the scope of environmental

interpretation, especially regarding energy exchange and plant response.

- Correlation Limitations: While Pearson's *r* provided linear insights, nonlinear relationships or causal links between variables were not explored.

5.4 Conclusion

In summary, the analysis provided key insights into the environmental behaviour of the strawberry greenhouse. It confirmed expected diurnal and seasonal variability, highlighted optimal but sensitive climate conditions, and supported the use of irTemperature as a potential stress proxy. However, the unexpected positive correlation between temperature and humidity urges caution in applying universal assumptions to greenhouse systems. Future work should explore sensor network expansion, multivariate modelling, and integration of solar and plant physiological data to improve precision agriculture strategies. Overall, this study reinforces the value of environmental monitoring for improving sustainability and productivity in controlled agriculture systems.

References

[1] J. M. López-Marín et al., "Greenhouse climate control for strawberry production: a review," *Agronomy*, vol. 11, no. 5, p. 987, 2021.

[2] A. P. Papadopoulos et al., "Integrated monitoring systems in protected cultivation," *Computers and Electronics in Agriculture*, vol. 142, pp. 197–208, 2017.

[3] H. G. Jones, *"Plants and Microclimate"* 3rd ed. Cambridge Univ. Press, 2014, pp. 112–135.

[4] S. R. Adams et al., "Seasonal temperature management in greenhouses," *Horticulturae*, vol. 6, no. 3, p. 42, 2020.

[5] L. Zhang et al., "Diurnal humidity-temperature correlations in greenhouses," *Biosystems Engineering*, vol. 193, pp. 1–12, 2020.

[6] M. J. Monteiro et al., "Fungal disease risk models in strawberries," *Plant Disease*, vol. 104, no. 2, pp. 456–464, 2020.

[7] N. Katsoulas et al., "Ventilation impacts on greenhouse microclimates," *Agricultural and Forest Meteorology*, vol. 265, pp. 156–167, 2019.

[8] R. A. Shellie et al., "Infrared thermography for water stress detection in strawberries," *HortScience*, vol. 55, no. 6, pp. 814–820, 2020.

[9] E. C. Oerke et al., "Precision agriculture for stress detection," *European Journal of Agronomy*, vol. 89, pp. 1–12, 2017.

[10] J. A. Sánchez-Molina et al., "IoT-Based early stress detection in crops," *Sensors*, vol. 20, no. 3, p. 741, 2020.

[11] FAO, *"Strawberry Cultivation Guidelines," FAO Agricultural Handbook*, 2019.

[12] C. A. Lopes et al., "Botrytis management in strawberries," *Plant Pathology*, vol. 68, no. 5, pp. 801–812, 2019.

[13] T. M. DeJong et al., "Heat stress effects on fruit quality," *Scientia Horticulturae*, vol. 256, p. 108553, 2019.

[14] P. P. Ling, "Sensor networks for greenhouse automation," *Transactions of the ASABE*, vol. 63, no. 1, pp. 329–336, 2020.

[15] S. K. Singh et al., "Decision support systems for strawberry farming," *Computers and Electronics in Agriculture,* vol. 158, pp. 226–234, 2019.

[16] A. Hasanuzzaman et al., “Strawberry Production under Climate Variability,” *Journal of Horticultural Science*, vol. 10, no. 2, pp. 112-120, 2020.

[17] M. Singh and S. Prasad, “Effect of Temperature and Humidity on Strawberry Yield,” *Agricultural Reviews*, vol. 41,

Appendices

Appendix A: Sample Python Code Used for Preprocessing

Note: This code snippet was used during the data cleaning and preprocessing analyses discussed in Section 2.1 See main body for interpretation.

```
# Temperature distribution chart showing results Before and after cleaning
plt.figure(figsize=(12,6))

# Calculate statistics
mean_temp = df['temperature'].mean()
std_temp = df['temperature'].std()
lower_cut = mean_temp - 2.5*std_temp
upper_cut = mean_temp + 2.5*std_temp

# Plot histograms graph
plt.hist(df['temperature'], bins=50, alpha=0.5, label='Original')
plt.hist(df_clean['temperature'], bins=50, alpha=0.5, label='Cleaned')

# Adding reference lines
plt.axvline(lower_cut, color='red', linestyle=':', label=f'Lower cutoff ({lower_cut:.1f}°C)')
plt.axvline(upper_cut, color='red', linestyle=':', label=f'Upper cutoff ({upper_cut:.1f}°C)')
plt.axvline(mean_temp, color='black', linestyle='-', label=f'Mean ({mean_temp:.1f}°C)')

# Formatting
plt.xlabel('Temperature (°C)')
plt.ylabel('Number of Readings')
plt.title('Temperature Distribution: Before vs After Cleaning')
plt.legend()
plt.grid(True, alpha=0.3)
plt.show()
```

Listing 1A: Python code used for Temperature distribution chart showing results Before and after cleaning

```
# Humidity distribution chart showing results Before and after cleaning
plt.figure(figsize=(12,6))
# Calculate statistics for humidity
mean_humidity = df['humidity'].mean()
```

```python
std_humidity = df['humidity'].std()
lower_cut = mean_humidity - 2.5*std_humidity
upper_cut = mean_humidity + 2.5*std_humidity
# Plot histograms
plt.hist(df['humidity'], bins=50, alpha=0.5, label='Original', color='skyblue')
plt.hist(df_clean_humidity['humidity'], bins=50, alpha=0.5, label='Cleaned', color='salmon')
# Add reference lines
plt.axvline(lower_cut, color='red', linestyle=':', linewidth=1.5,
           label=f'Lower cutoff ({max(lower_cut,0):.1f}%)')
plt.axvline(upper_cut, color='red', linestyle=':', linewidth=1.5,
           label=f'Upper cutoff ({min(upper_cut,100):.1f}%)')
plt.axvline(mean_humidity, color='black', linestyle='-',
           label=f'Mean ({mean_humidity:.1f}%)')
# Formatting
plt.xlabel('Humidity (%)', fontsize=12)
plt.ylabel('Number of Readings', fontsize=12)
plt.title('Humidity Distribution: Before vs After Cleaning', fontsize=14, pad=20)
plt.legend(loc='upper right', frame alpha=1)
plt.grid(True, alpha=0.3)
# Add physical bounds (0-100%) as vertical lines
plt.axvline(0, color='gray', linestyle='--', alpha=0.5, linewidth=0.7)
plt.axvline(100, color='gray', linestyle='--', alpha=0.5, linewidth=0.7)

plt.show()
```

Listing 1B: Python code used for Humidity distribution chart showing results Before and after cleaning

Appendix B: Supplementary Visualisation

Note: These plots confirm the dataset's reliability after removing sensor errors/extreme values discussed in Section 2.1 See main body for interpretation.

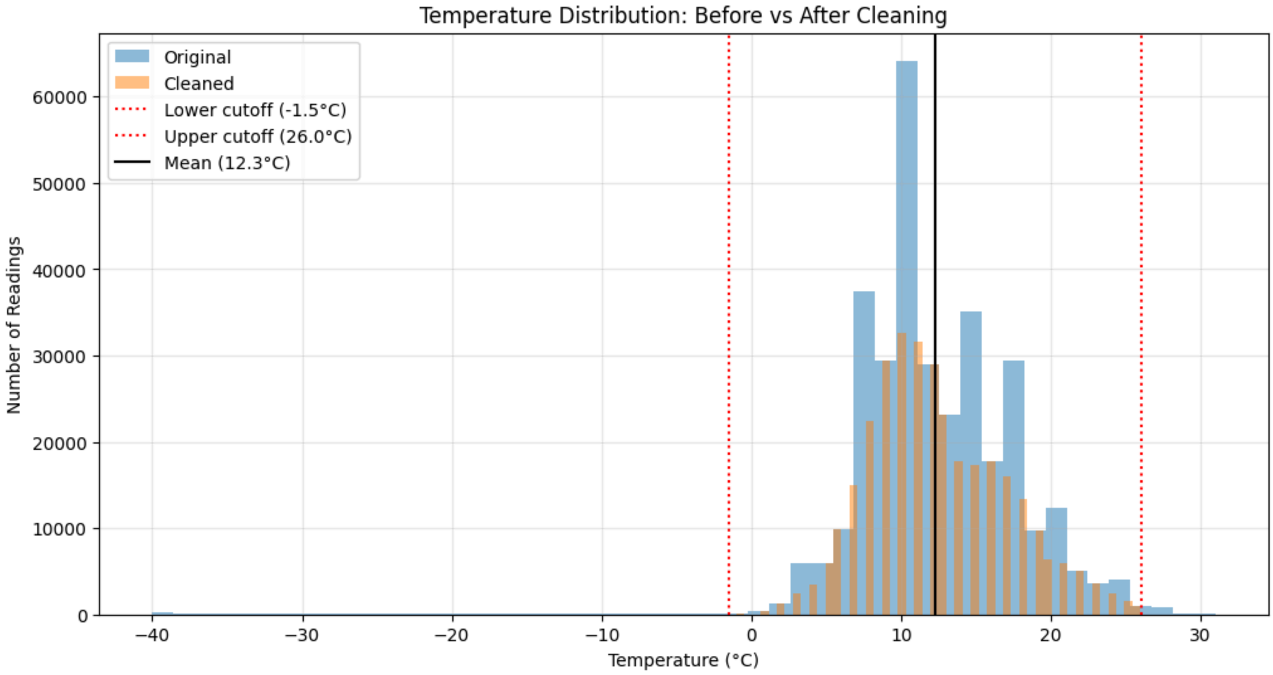


Figure B1: Temperature Distribution Before vs After Data Cleaning (valid range: −1.5°C to 26.0°C).

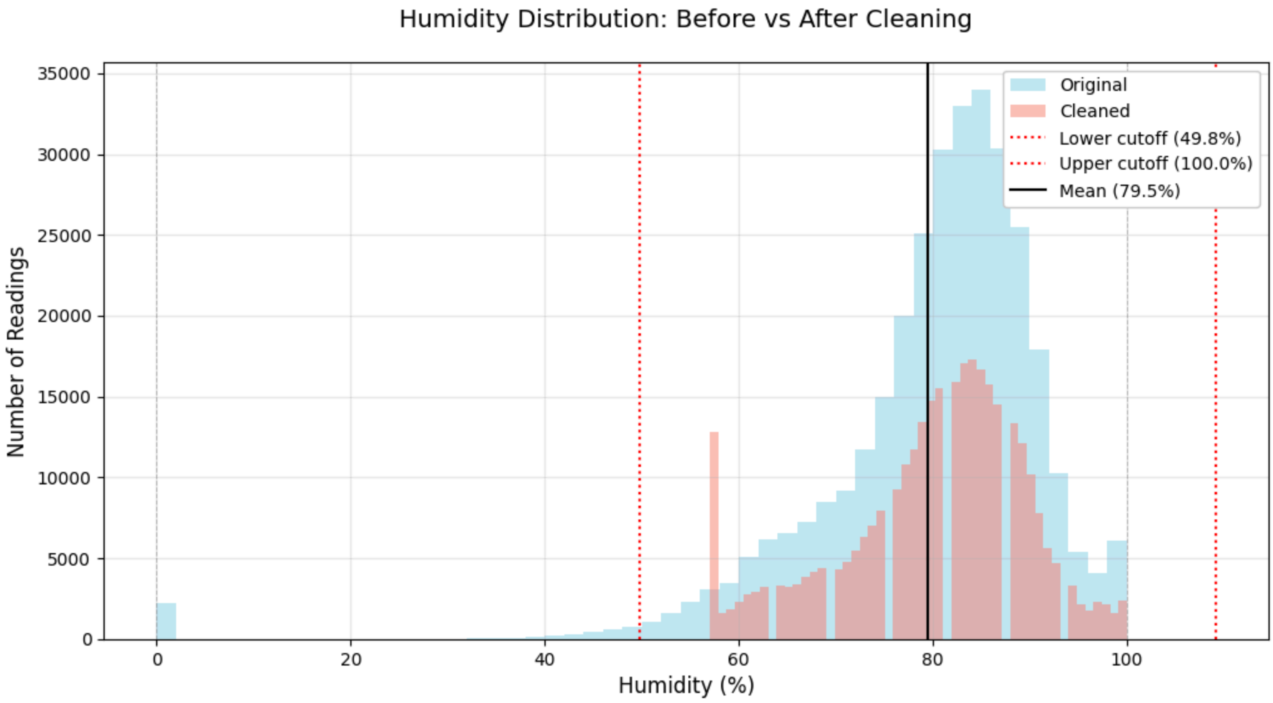


Figure B2: Humidity Distribution Before vs After Data Cleaning (valid range: 49.8% to 100%).