Search for evidence of climate change in UK weather data

Author: Ethan Kosak-Hine

Supervisor: Professor Veronique Boisvert

March 2022

Abstract

This study uses traditional statistical analyses and a machine learning (ML) classification approach on historic Met Office weather station data to search for evidence of climate change. The study demonstrates the successful use of ML models to categorise UK weather data into before and after 1990, which serves as an inflexion point for climate change. Z-scores at the 5% confidence level were calculated to quantify the significance of the trend-over-time and ML analyses. The trend-over-time analysis provides evidence of climate change with a higher significance level than the ML approach. Rates of change calculated using trend-over-time analysis had a significance greater than 10σ. The ML approaches separated the classes with a significance of between 2σ and 4.2σ. Additionally, this study assesses the use of an ML approach in gaining insight into the physics of the climate and climate change.

Contents

1 Introduction

Climate change refers to the large-scale shifts in weather patterns resulting from the increased concentration of greenhouse gases in our atmosphere. This global challenge poses significant threats to human well-being and wildlife. These threats are manifested in a range of adverse impacts, including more frequent and severe natural disasters, rising food and water insecurity, and widespread species extinction and habitat loss [1, 2]. As a result, understanding the magnitude, direction, and causes of climate change is critical for developing effective mitigation and adaptation strategies.

Traditionally, researchers have used statistical analysis to detect evidence of climate change. This approach involved analysing long-term trends in weather variables. While traditional statistical analysis is useful, it has its limitations, such as difficulty in accounting for complex, non-linear relationships.

In recent years, machine learning (ML) has emerged as a promising approach for climate change research. ML algorithms can identify complex patterns between variables in large data sets, as well as help to develop predictive models that can forecast future climate trends.

This study explores the potential of combining traditional statistical analysis and ML classification approaches to detect evidence of climate change. First, a simple statistical analysis will be used to try and detect evidence of climate change. Next, a ML approach will be used whereby various models will be trained to classify data from different periods based on the degree of climate change that has occurred. Finally, the results of the two approaches will be compared. This study uses historic weather station data from the UK spanning over 150 years, allowing for long-term analysis.

The findings from this study are expected to contribute to the growing body of knowledge on climate change and demonstrate the potential of combining traditional statistical analysis and ML classification techniques. Ultimately, this research aims to test novel methods for analysing climate data and assess their viability in future research.

2 Physics of the climate and climate change

The physics of Earth's climate system is a complex and interdisciplinary field encompassing various aspects of atmospheric science, oceanography and geophysics. Earth's climate system has many components, including the atmosphere, oceans, land and ice, that all interact with each other to impact the Earth's weather and overall climate. One of the key components of the Earth's climate is its energy balance, the balance between incoming solar radiation and outgoing thermal radiation.

2.1 The Earth's energy balance

The Earth's climate is influenced by the balance between incoming solar radiation and outgoing thermal radiation [3]. An Energy Balance Model (EBM) is used to model this balance and understand the Earth's temperature changes. There are three main types of EBM: zero, one, and two-dimensional [4]. A zero-dimensional EBM represents the Earth as a single point, while one-dimensional models introduce a latitudinal dependence, breaking the Earth up into nine latitude regions. Two-dimensional models take this further, splitting the Earth up into a two-dimensional grid, useful for studying spatial and temporal variability [5, 6].

It is important to distinguish between effective temperature and surface temperature. The effective temperature is the temperature Earth would have if it were a perfect black body, whereas surface temperature takes into account the atmosphere, oceans, and land surface [3]. Although the effective temperature is a legitimate temperature measurement in that it measures the temperature in relation to Earth's outgoing longwave radiation, climate scientists are primarily concerned with surface temperature. This section will outline a simple zero-dimensional model to illustrate the basics of climate modelling. However, additional types of more complex EBMs are often used, such as three-dimensional models [5] and Earth System Models (ESMs) [6].

To build this model, first note that for the Earth's (effective) temperature to remain constant (or balanced), the incoming and outgoing energy (per unit time) must be equal, i.e. $P_{\text{in}} = P_{\text{out}}$. If the heat flows are not balanced, the overall temperature of the system will change. Now let's highlight some important assumptions and variables.

To simplify calculations, the Earth is modelled as a flat disk perpendicular to the Sun's rays due to the plane-parallel assumption [7]. This assumes that all points on the Earth's surface facing the Sun receive the same amount of solar radiation, as the curvature of the Earth's surface is relatively small compared to the distance from the Sun. Additionally, the Earth is assumed to be a black body that perfectly absorbs and radiates energy without reflecting any radiation.

Variables:

- R = Earth's radius $\approx$ 6371 km,
- $S_\odot$ = The energy flux density of the Sun, i.e. the energy flowing through a flat surface of 1 m^2. This is measured to be $S_\odot = 1368$ Wm^{-2} [8],
- α = The Albedo of the Earth, this is the fraction of the Sun's radiation which is reflected into space. This is measured to be around 0.3 [9],
- σ_{SB} = Stefan-Boltzmann constant = 5.67×10^{-8} $\text{Wm}^{-2}\text{K}^{-4}$,
- T_0 = conversion constant from Kelvin to degrees Celsius = 273.15,
- T_C = Earth's effective temperature in degrees Celsius.

The albedo α of the Earth is the fraction of the radiation from the Sun that is reflected into space, thus the total amount of solar radiation incident on the Earth surface per unit area is $S_\odot(1-\alpha)$. Modelling the area of the Earth that is exposed to the Sun's radiation as a disk of radius R, the total power absorbed by the Earth is therefore given by

$$P_{\text{in}} = \pi R^2 S_\odot (1-\alpha). \tag{1}$$

Assuming that the Earth is a black body, the outgoing power is given by the Stefan-Boltzmann law

$$P_{\text{out}} = A\sigma_{\text{SB}} T_K^4, \tag{2}$$

where A is the surface area of the Earth. Here we can assume the Earth is a sphere because it is the entire Earth that radiates energy, not just the surface facing the Sun (that we previously assumed to be a disk). This is given by $4\pi R^2$ and T_K is the Earth's effective temperature in Kelvin, where $T_K = T_0 + T_C$.

We are assuming the incoming and outgoing power is approximately balanced, so setting $P_{\text{in}} = P_{\text{out}}$ and solving for T_K, we get

$$T_K = \left(\frac{S_\odot(1-\alpha)}{4\sigma_{SB}} \right)^{\frac{1}{4}}. \tag{3}$$

After substituting in the quantities previously stated, we arrive at $T_K = 254.9$ K, or -18°C. This cannot be the whole picture, as this model predicts a global average temperature far below the measured global average temperature of around 15°C [10]. This is because this simple EBM

calculates Earth's effective temperature instead of its surface temperature. To find Earth's surface temperature, higher-order effects must be considered, such as the contribution of the atmosphere (and the Greenhouse Effect), the oceans, and land surface effects.

2.2 The Greenhouse Effect

If the simple energy balance formula is applied to Mars, the predicted average temperature agrees with the measured average temperature. On the other hand, for Venus, with its thick CO_2 atmosphere, the simple model we proposed does not work well at all. The explanation for this is the Greenhouse Effect [11].

The Greenhouse Effect warms Earth from what would otherwise be a frozen wasteland to a warm and habitable landscape. It occurs when gases in a planet's atmosphere trap heat that would have otherwise escaped into space, resulting in an overall higher temperature than it would otherwise have. Note that the presence of the Greenhouse Effect does not necessarily mean a planet's temperature will increase over time. If the concentration of GHGs remains constant, we would still expect the average temperature to be warmer than it would be in the absence of GHGs, but the temperature would remain in equilibrium [12].

The molecular structure of the components of Earth's atmosphere determines whether or not they are a GHG. Why is it, for example, that carbon dioxide is a GHG, but oxygen is not? At its core, it all comes down to the specific structure of the gas molecules [13]. Oxygen molecules are symmetric and lack an electric dipole moment, so they do not interact with atmospheric radiation. In contrast, carbon dioxide molecules are linear and symmetrical but possess asymmetric vibrations that can change the charge separation of the carbon and oxygen atoms, leading to a dipole moment[1]. The presence of a dipole moment means that the molecule has a positive and negative end, creating a non-uniform distribution of charge, thereby causing the molecule to behave like a magnet. Incident photons from the Sun can interact with these dipoles by exciting molecular vibrations or rotations, which change the strength of the dipole moment. Consequently, the molecule undergoes a fluctuation in its electromagnetic field, which results in the emission of another photon in a random direction [13]. On the scale of the entire atmosphere, this leads to some of this infrared radiation being emitted back towards Earth. At certain wavelengths, radiation from the Sun arrives almost unattenuated. However, at other wavelengths, it is removed almost entirely by absorption in the spectral bands of GHGs. Carbon dioxide, for example, has several strong absorption bands in the near-infrared region, as well as two in the mid-infrared region at 4.3 and 15 microns. The 15-micron band, in particular,

[1]Some non-linear molecules, such as water vapour and ozone, also have dipole moments due to their bent shapes [14]

falls near the peak of the Planck function for typical atmospheric temperatures. This is a function used to describe the spectral distribution of outgoing longwave radiation emitted by the Earth's surface. This function has the following form

$$R(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1}, \tag{4}$$

where R is the radiance, h is Planck's constant, c is the speed of light, k is Boltzmann's constant, λ is the wavelength, and T is the temperature in Kelvin. Given that the 15-micron band for carbon dioxide falls near the peak of this distribution, it is especially effective at absorbing outgoing infrared radiation emitted by Earth's surface. Figure 1 shows the absorption spectra of various atmospheric gases, including carbon dioxide.

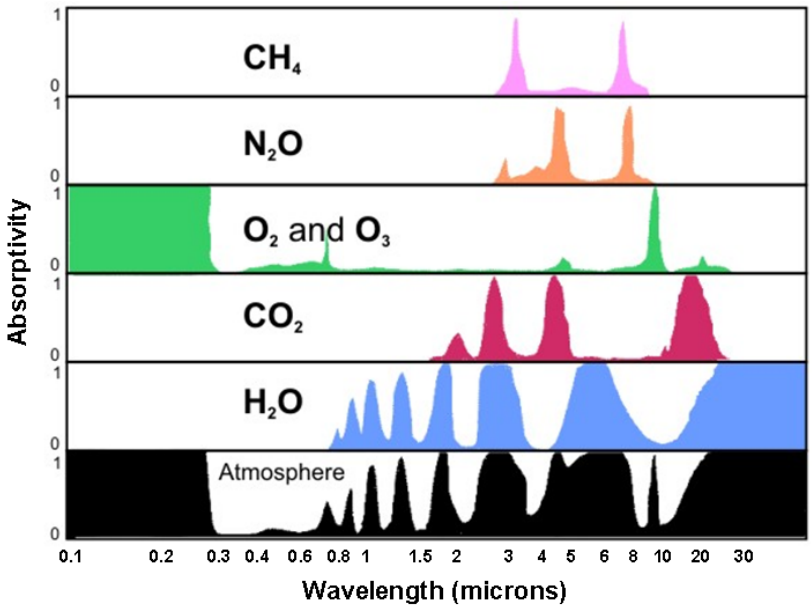


Figure 1: The absorption spectra of various gases in Earth's atmosphere, and the atmosphere as a whole [15].

The presence of GHGs in our atmosphere causes a warming of Earth's surface and troposphere (the lowest layer of the atmosphere) [16]. However, not all GHGs have the same impact on the warming of the Earth. The impact of a specific GHG on the environment can be quantified by defining the Global Warming Potential (GWP) of a GHG. The GWP is a measure of a GHGs ability to trap heat in the atmosphere relative to carbon dioxide. The GWP is calculated over a specific time horizon (usually 100 years) and is used to compare the warming impacts of different GHGs. The formula for the GWP is given by

$$\mathrm{GWP}(x) = \frac{\int_0^{TH} a_x \cdot [x(t)]dt}{\int_0^{TH} a_r \cdot [r(t)]dt}, \tag{5}$$

where TH is the time horizon over which the calculation is made, a_x is the radiative efficiency due to a unit increase in the atmospheric abundance of the substance in question (represents the amount of heat absorbed per unit of atmospheric mass), $[x(t)]$ is the time-dependent decay in the abundance of the instantaneous release of the substance (represents the change in concentration of the GHG over time). The corresponding quantities for the reference GHG are in the denominator [17].

A bar chart of the GWP for different GHGs is shown in Figure 2.

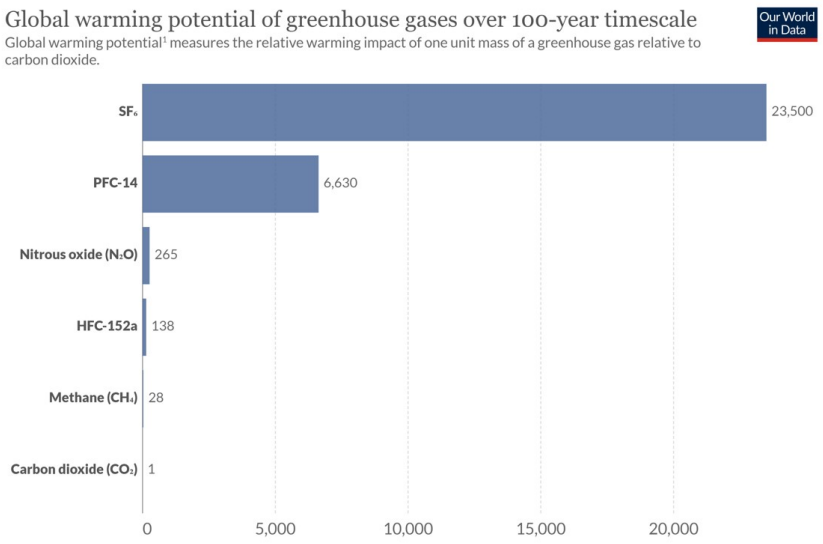
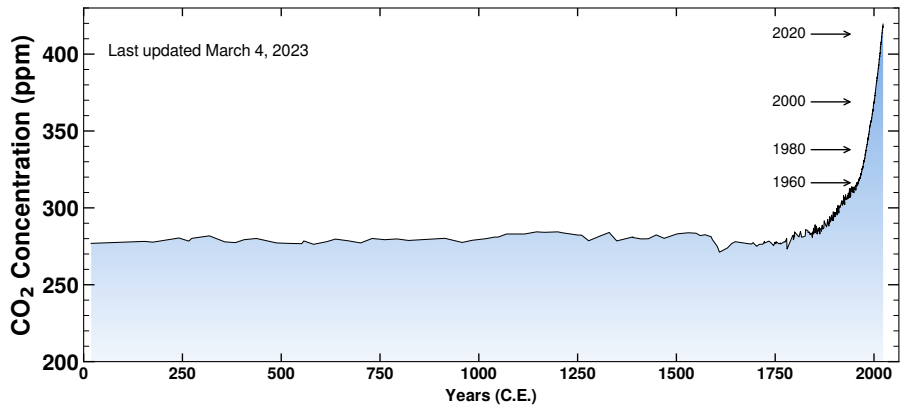
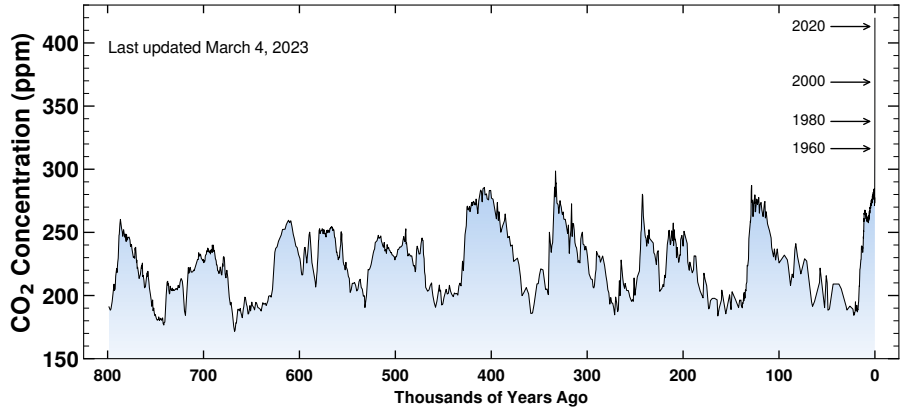


Figure 2: The Global Warming potential (GWP) of different Greenhouse Gases (GHGs) over a 100-year time scale, under the assumption that the reference GHG is carbon dioxide and that the TH is 100 years. The GWP is a measure of the amount of heat absorbed by a GHG relative to the same mass of carbon dioxide [18, 19].

As the amount of GHGs in the atmosphere increase, so does the warming effect induced by them. Figure 3a shows the global carbon dioxide concentration record for the past 2000 years. A sharp increase in global carbon dioxide concentration can be seen to begin around the 1800s, i.e. the industrial revolution [20]. Figure 3b shows this plot for the past 800,000 years. The periodic changes in carbon dioxide concentration can be observed and have been confirmed to coincide with the historical global warming and cooling of the Earth [21]. The current sharp increase in carbon dioxide concentration in Figure 3b is far larger and occurring faster than any of the previous cycles, indicating that the reason for the change is human activities.



(a)



(b)

Figure 3: The Keeling curve: a record of global atmospheric carbon dioxide concentration over a 2000 year time frame (a) and a 800,000 year time frame (b), measured in parts per million (ppm) [22].

2.3 Climate feedback mechanisms

Climate feedback mechanisms are a critical aspect of the Earth's climate system. Feedback mechanisms can amplify or dampen the effects of external forces on Earth's climate, such as changes in greenhouse gas concentrations, solar radiation, and other factors [23].

One of the most impactful feedback mechanisms is the positive feedback loop between GHG concentrations and temperature. As mentioned in Section 2.2, GHGs, such as carbon dioxide, methane, and water vapour, trap heat in the atmosphere and lead to a warming effect. This warming effect, in turn, causes more evaporation, leading to an increase in the concentration of water vapour, a potent GHG, in the atmosphere. This increase in water vapour amplifies the warming effect, leading to further warming, leading to even more evaporation and a continuous positive feedback loop.

Another powerful feedback mechanism is ice-albedo feedback. This mechanism involves the melting of polar ice caps, which exposes darker surfaces (such as liquid water) that absorb more solar radiation, leading to further warming and more melting. As more ice melts, more dark surfaces are exposed, amplifying the warming effect and leading to a positive feedback loop.

The carbon cycle feedback is another critical feedback mechanism that involves the exchange of carbon between the atmosphere, oceans, and land [24]. As the Earth warms, carbon stored in soils and permafrost is released into the atmosphere, amplifying the greenhouse effect and leading to further warming. This feedback mechanism can have significant consequences for the global carbon cycle, as it can either amplify or dampen the quantity of carbon stored in the Earth's carbon sinks.

2.4 Climate change and its impacts

Climate change is the long-term shift in temperatures and weather patterns caused by human activities, including burning fossil fuels, deforestation, and industrial processes, resulting in significant and lasting changes in weather patterns [2]. England can expect to experience hotter, drier summers, milder, wetter winters, more extreme weather events, and rising sea levels [25].

Climate change also significantly impacts wildlife and biodiversity, such as direct habitat loss, population reductions, decoupling of coevolved interaction, increased spread of wildlife diseases, and extinction [26]. Furthermore, climate change poses less obvious but dangerous threats from ancient microorganisms released by thawing ancient ice [27]. Additionally, climate change leads to melting ice, releasing heavy metals, additional GHGs, and potentially even buried radioactive and biological waste [28].

For the UK, a significant concern is drought risk, particularly in the southeast and east of England, where water resources are already under pressure. The CCRA suggests that the frequency and severity of droughts are likely to increase, leading to water shortages and restrictions on water usage, which could have implications for food production and economic growth [29]. Climate change also significantly impacts human health, particularly through extreme weather events, which can cause deaths, injuries, and illnesses, especially among vulnerable populations [29].

Climate change is a pressing global issue, and the UK is not immune to its impacts. It is crucial to adapt to these changes and reduce emissions to limit the extent of the impacts on future generations [2].

3 Machine learning

3.1 Classification

Classification is a widely used supervised ML technique that categorises data into two classes based on specific features or attributes [30]. Its goal is to accurately predict the class of a new observation by identifying correlations and relationships within the data. To train a model, the data set is split into training and testing sets, and the model is trained on the former by adjusting its parameters to minimise prediction error [31]. The testing set is used to evaluate the model's performance. Overfitting, where a model learns the training data too well and fails to generalise to new data, is a common problem in classification [32]. A graphical representation of overfitting is shown in Figure 4.

Popular algorithms used in classification include logistic regression, decision trees, support vector machines, and neural networks, with the choice of the algorithm depending on the problem's nature and data quality [30].

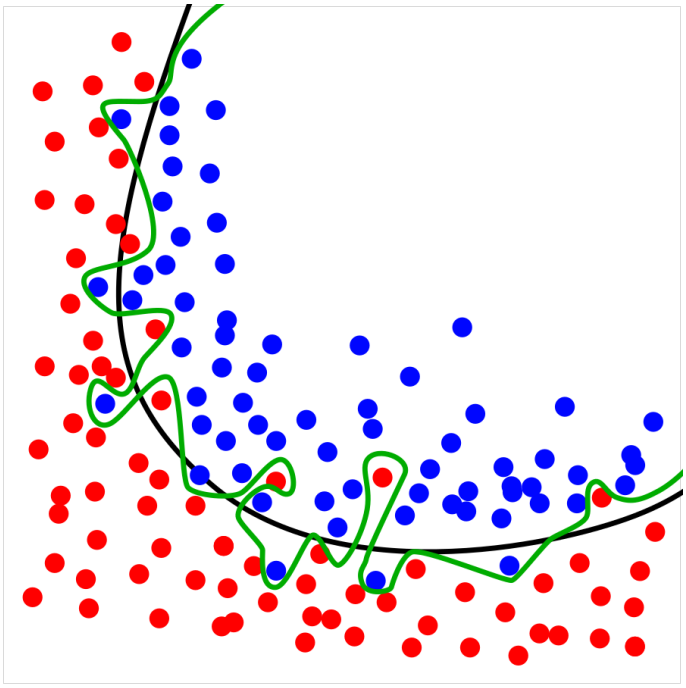


Figure 4: A graphical representation of overfitting in a classification algorithm. The red and blue dots represent instances of each class in a two-dimensional feature space. The black line indicates the optimal model, whereas the green line shows an overfit model [32].

3.2 Boosted Decision Trees

Decision trees are a classification algorithm that creates a tree-like model, where each branch represents a decision rule based on a feature of the data. Boosted decision trees (BDTs) improve upon this method by combining multiple decision trees to create a more accurate and robust model. Sequentially adding decision trees to the model corrects errors made by the previous tree, and the process continues until a stopping criterion is met. A BDT was chosen for this problem due to its robustness to noisy data and ability to handle complex non-linear relationships between features and the target variable. Additionally, BDTs are fast to train and predict, making them suitable for large data sets [33]. Figure 5 shows a graphical representation of a basic decision tree.

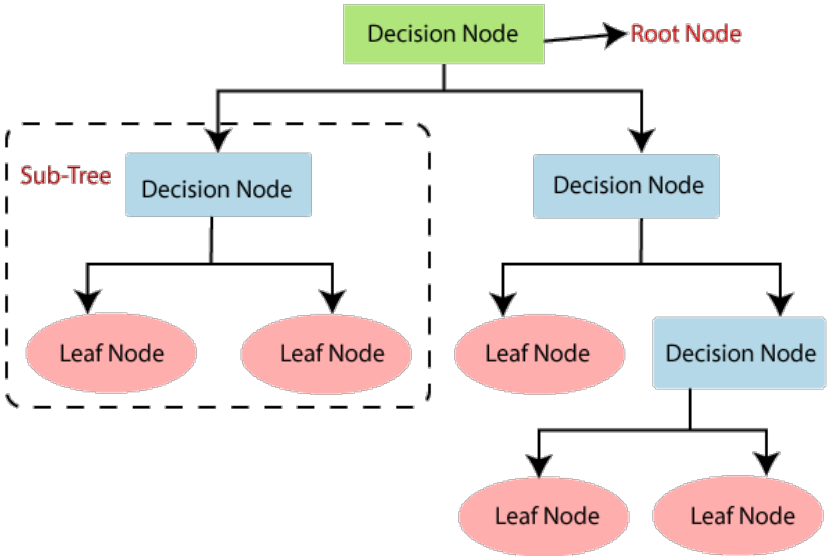


Figure 5: A graphical representation of a basic decision tree with all of the main features labelled [33].

4 Methodology

4.1 Data source and collection

The primary data set used in this study comes from the Met Office, the national meteorological service for the UK [34]. The data set consists of historical weather data from 37 weather stations located throughout the UK, spanning between 50 to 100 years of measurements. Monthly measurements of five weather variables were recorded, including the mean daily maximum temperature $t_{\max}$, the mean daily minimum temperature $t_{\min}$, the total number of air frost days per month, total rainfall per month and total sunshine duration per month [34]. Temperature measurements for each station are recorded at one-minute intervals, far smaller than the time intervals of the data set used here. The

data are then transmitted to a central collecting system where they are passed through numerous quality control checks [35].

To more precisely define the temperature variables $t_{\max}$ and $t_{\min}$, the process for arriving at a specific $t_{\max}$ measurement (for example) for any given month and weather station is as follows. First, the maximum temperature value for each day is identified. For any given month, the average of these daily maximum values is then calculated, giving the final $t_{\max}$ for that month. The same is done for $t_{\min}$ but the minimum daily temperature is used instead. Air frost days were defined as days when the air temperature fell to or below the freezing point of water at a height of at least one metre above the ground [36].

4.1.1 Temperature measurements

Temperature is a measure of the hotness or coldness of an object compared to a reference point. The kinetic energy of particles within a system determines an object's temperature, where a higher temperature indicates a higher kinetic energy band and vice versa. Measuring air temperature is essential in understanding weather patterns, assessing weather impacts on crops and vegetation, and forecasting climate change.

To ensure the accuracy and consistency of temperature measurements, the Met Office uses standardised methods when collecting temperature data for the historic weather stations used in this study. In the past, mercury thermometers were used, but today, platinum resistance thermometers (PRTs) or alcohol thermometers are used due to their increased accuracy and reduced risk over mercury thermometers.

At all their synoptic and supplementary weather stations employing an automatic system, the Met Office uses digital PRTs to measure air temperature [37]. A PRT is a piece of platinum wire whose electrical resistance increases approximately linearly with absolute temperature, making it useful as a temperature sensor. The resistance of the platinum wire is measured by passing a current through it, and a calibration equation converts the reading to a temperature measurement [38]. PRTs have a high resolution (0.01°C), repeatability, and stability across a wide range of temperatures (-200°C to 500°C) [39].

The PRTs are exposed in a Stevenson screen at the Met Office weather stations. The shelters are designed to protect the thermometers from direct sunlight, rain, snow, wind, leaves, and animals while still allowing air to circulate freely around them. The shelters are elevated 1.25 m above the ground and are white to reflect solar radiation [37].

4.1.2 Rainfall measurements

Since Met Office weather stations began taking records, the primary method for measuring rainfall was the storage rain-gauge [40]. Daily rainfall was measured at 0900 UTC by an observer from a 5 inch storage rain-gauge. The gauge has a sharp brass or steel rim of diameter 127 mm, at 30 cm above ground level, with a funnel that collects the rain in a narrow-necked bottle that is placed in a removable can. To make the rainfall measurements, the observer empties the rain into a graduated glass rain measure.

In the 1980s and 1990s, automated instruments were introduced across the network of Met Office weather stations. The tipping bucket rain-gauge became the standard instrument for automatically measuring rainfall. The collecting funnel has a sampling area of 750 cm^2, and the rim is raised 450 mm above the surrounding ground level. A mechanism records an event each time a rainfall increment of 0.2 mm has been detected.

4.1.3 Sunshine measurements

Historically, the Campbell-Stokes (CS) recorder has been the established instrument for measuring sunshine for over 100 years. It is a glass sphere that focuses sunlight onto a strip of cardboard, burning a trace whenever the sun is shining [41].

More recently, specifically in the automatic weather stations, the Met Office introduced the Kipp and Zonen (KZ) CSD-1 sensor. It is a pyrheliometer, a device that measures solar irradiance coming directly from the sun (in units of watts per square metre). It has a silicon photodiode sensor that converts the sunlight's direct radiation into an electrical signal which is transmitted to a data logger. It is recorded and processed to determine the duration of sunshine.

The CS recorder has many well documented weaknesses [42]. For example, on days with intense and intermittent sunshine, the burn pattern can spread, leading to an overestimation of the duration of direct sunshine. In addition, colder and damper weather conditions require more energy to burn a trace on the card than warmer and drier ones. During the dawn period, dew or frost may cover the sphere, which decreases the sunlight intensity passing through it. Finally, the measurement must be taken manually by an observer, resulting in a high degree of subjectivity affecting the sunshine measurements. Still, despite these limitations being well-known since the instrument's invention, the deficiencies are not easily overcome and result in a high degree of uncertainty. Regardless, many Met Office weather stations still use the CS recorder widely, with the KZ recorder being used to a lesser degree, especially within the historic Met Office stations.

4.2 Data preprocessing

Data preprocessing is an essential step in ML as it involves transforming raw data into a format suitable for the ML algorithm. It may also involve techniques such as filtering out less relevant data (reducing noise), focusing on the most relevant features [43], and handling missing values in the data sets [44].

This study divides data preprocessing into two categories: necessary preprocessing and optimisation preprocessing. Necessary preprocessing involves transformations and edits made to the data sets, which are necessary for any ML algorithm to parse them. These include changing data types, removing non-numerical characters, and handling missing values.

Optimisation preprocessing steps are applied once the necessary preprocessing steps have been completed. They involve transforming the data to improve the accuracy and efficiency of the ML model. These transformations include filtering out less relevant data or features (noise reduction), defining new variables as transformations of existing variables, and scaling the data to have the same units.

In summary, data preprocessing is a critical step in ML. Dividing it into necessary and optimisation preprocessing categories helps ensure that the data is transformed and optimised appropriately for use by the ML algorithm. Optimisation preprocessing is performed to improve the performance of the ML models, and hence will be discussed in the ML section of the study, in Section 5.2.

4.2.1 Necessary preprocessing

The raw data obtained from the MET office requires preprocessing before detailed analysis. Several non-numerical values are present in the data sets, making any in-depth analysis impossible. This section discusses the processes used to remove these values. Table 1 lists these non-numerical characters and their purposes in the MET office data sets. These characters include symbols for missing data, estimated or provisional data, and sunshine data recorded using a Kipp and Zonen sensor.

The first necessary preprocessing step is to remove these non-numerical characters. If the data is missing, the non-numerical characters are replaced by NaN values. However, for the initial study, any rows with missing data had to be removed entirely from the data set because the ML algorithms require complete data. Unfortunately, dropping these rows results in the loss of a significant amount of data, even though there may be only a small amount of missing data. This issue is particularly problematic as the data sets often span hundreds of years, with weather stations needing complete data from their inception. For example, the data from the Northern Ireland weather station, Ballypatrick Forest, had to be excluded from the study since it contained no complete data for any month.

Table 1: A list of the non-numerical characters found in the MET office data sets and their purposes [34].

Non-numerical character	**Purpose**
- - -	Indicates missing data (applies when there are more than two days missing per month).
*	Indicates estimated data (where available).
#	Applied after sunshine data recorded using a Kipp and Zonen sensor.
Provisional	Indicates provisional data, i.e. data that has not completed the quality control process.

5 Results

This investigation employs various techniques to analyse the weather data. The first method, a traditional analysis, looks at the distributions of the variables over time. Traditional data analysis techniques were used, including error analysis and a study into how exceptional the year 2022 was regarding the weather variables compared to the historical data.

Next, we apply various classification models to the data. The aim is to pick a year as an inflexion point for the climate whereby a classification model can identify a difference between any data recorded before the given year and after. If the models perform strongly in this exercise, this indicates that the climate has noticeably changed in some way. Further analysis can be done into the relative strengths of each variable to gain insights into the physics of climate change.

5.1 Traditional analysis

5.1.1 Trend over time

Plotting the data and fitting a simple linear model is an easy way to determine whether the temperature in the UK is following the global trend of increasing over time. This method can be applied to each of the five weather variables, and from this, the rate of change can be identified as the gradient of the fit. The uncertainty on this rate of change is just the calculated uncertainty on the gradient.

The raw data set, as produced by the Met Office, has a measurement for every month and station (unless missing values are present). For our purposes, a yearly value for each variable is needed to produce trend-over-time plots. First, we define a new variable t_{avg}. A new t_{avg} column is added to the original data set, which is the mean of t_{max} and t_{min} for each month and year of each weather station.

To determine a yearly value for each station, the average of all monthly $t_{\max}$, t_{avg}, $t_{\min}$, is calculated. The sum of all monthly measurements is taken to calculate the yearly total for the number of air frost days, rain and sun. The data set now consists of yearly data for 37 weather stations. This must be combined, so the average is calculated for all 37 stations. Finally, the data set comprises six variable columns (the three temperature variables, air frost, rain, and sun), with 170 rows (a row for each year from 1853 to 2022).

For this study, we consider two primary sources of uncertainty. The first is the intrinsic measurement uncertainty. Another, more prominent source of uncertainty is the number of stations taking data for any given year.

The data comprises 37 different weather stations from around the UK, with the earliest station taking data being in the Northern Ireland county of Armagh in 1853. In the traditional analysis, trend-over-time plots for the different variables are produced, and straight lines are fitted to the data to try and estimate a gradient. However, the fit depends on the uncertainties of the original measurements.

For the temperature measurements, the Met Office historically used mercury thermometers and then shifted to alcohol and digital thermometers sometime in the mid to late 20th century, when digital thermometers became commercially available [45, 46]. The resolution of all the historic weather station data temperature measurements is given to one decimal place, so we can set the uncertainty at 0.05°C based on this. Given the best evidence for the equipment quality used before introducing digital thermometers, the measurement uncertainty for pre-1950 is set to double the post-1950 value, 0.1°C. The measurement uncertainties for the other variables are set using the same assumption (that the uncertainty before 1950 is the resolution of the data and is half the resolution after 1950). For the number of air frost days, the measurement uncertainty is set to 1 day before 1950 and 0.5 days after 1950. For rain, 1 mm before and 0.5 mm after 1950, and for sun, 1 hour before and 0.5 hours after 1950.

There are, however, other factors that need to be considered. The number of stations taking data has not always been (and is currently not) 37. This is just the total number of historic weather stations. When records began at Armagh in 1853, it was the only station taking data. Given that the traditional trend-over-time analysis on the temperature in the UK is done by taking the mean temperature measurements of all stations in the UK, we would expect that for years where fewer stations were taking data, there would be a more significant uncertainty on the calculated mean temperature. A bar chart showing the number of historic MET office weather stations taking data for each year since records began is shown in Figure 6. We would also expect this relationship to apply to the other variables because they are all averages of the total value for each year and each station.

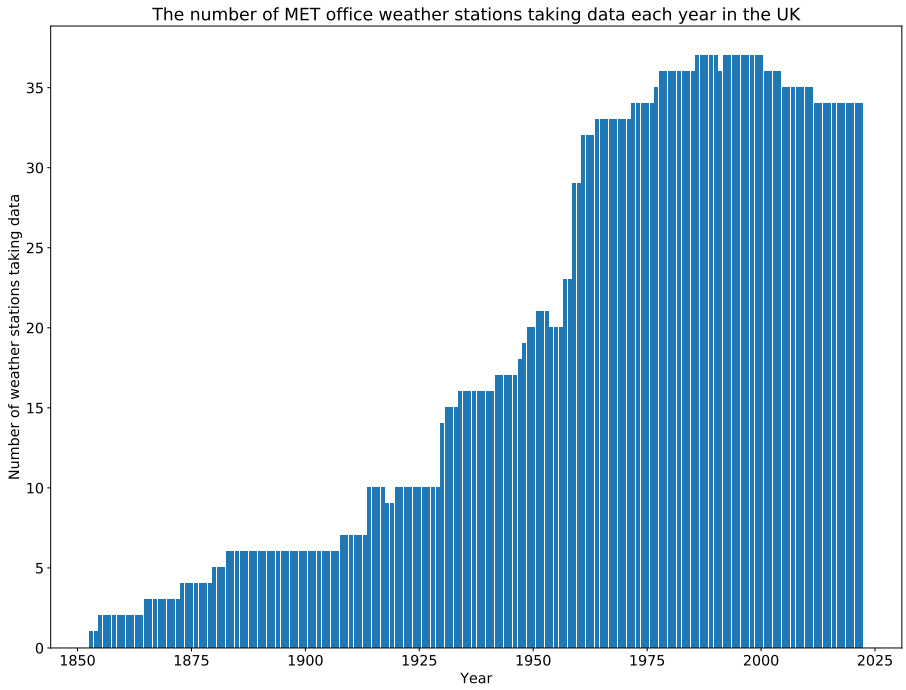


Figure 6: A bar chart showing the number of weather stations taking data for each year since records began in 1853.

Due to the drastic differences in regional weather in the UK [47], the years where only a few weather stations recorded data can be expected to have a very poor representation of the total average temperature for the UK as a whole. Given this, it makes sense to include a contribution to the uncertainty on each yearly temperature value from the number of stations taking data. The final source of uncertainty is the standard deviation of the data, which has a non-negligible contribution, especially for earlier data.

The contribution from the measurement uncertainty was defined previously as ± 0.1°C for pre-1950 and ± 0.05°C for post-1950. Let's define an array of values for the measurement uncertainty called σ_m as

$$\sigma_m = \begin{bmatrix} 0.1 \\ 0.1 \\ \vdots \\ 0.05 \end{bmatrix}. \tag{6}$$

Using $t_{\max}$ as an example, the total uncertainty of each measurement would be

$$\sigma_i = \sqrt{\sigma_m^2 + \left(\frac{\sigma_{t_{\max}}}{\sqrt{n_i}}\right)^2}, \tag{7}$$

where $\sigma_{t_{\max}}$ is the standard deviation of the $t_{\max}$ data, and n_i is an element of the array composed

of the number of weather stations taking data for each year. The total uncertainty σ is then parsed to the curve fitting function, which determines the uncertainty on the fit parameters.

The plots for $t_{\max}$, t_{avg} and $t_{\min}$ over time are shown in Figure 7. Linear fits were made and show a positive rate of change. The shaded regions are the fit values minus the weighted standard deviations. The error bars indicate that the data for earlier years tend to have a larger uncertainty than in recent years.

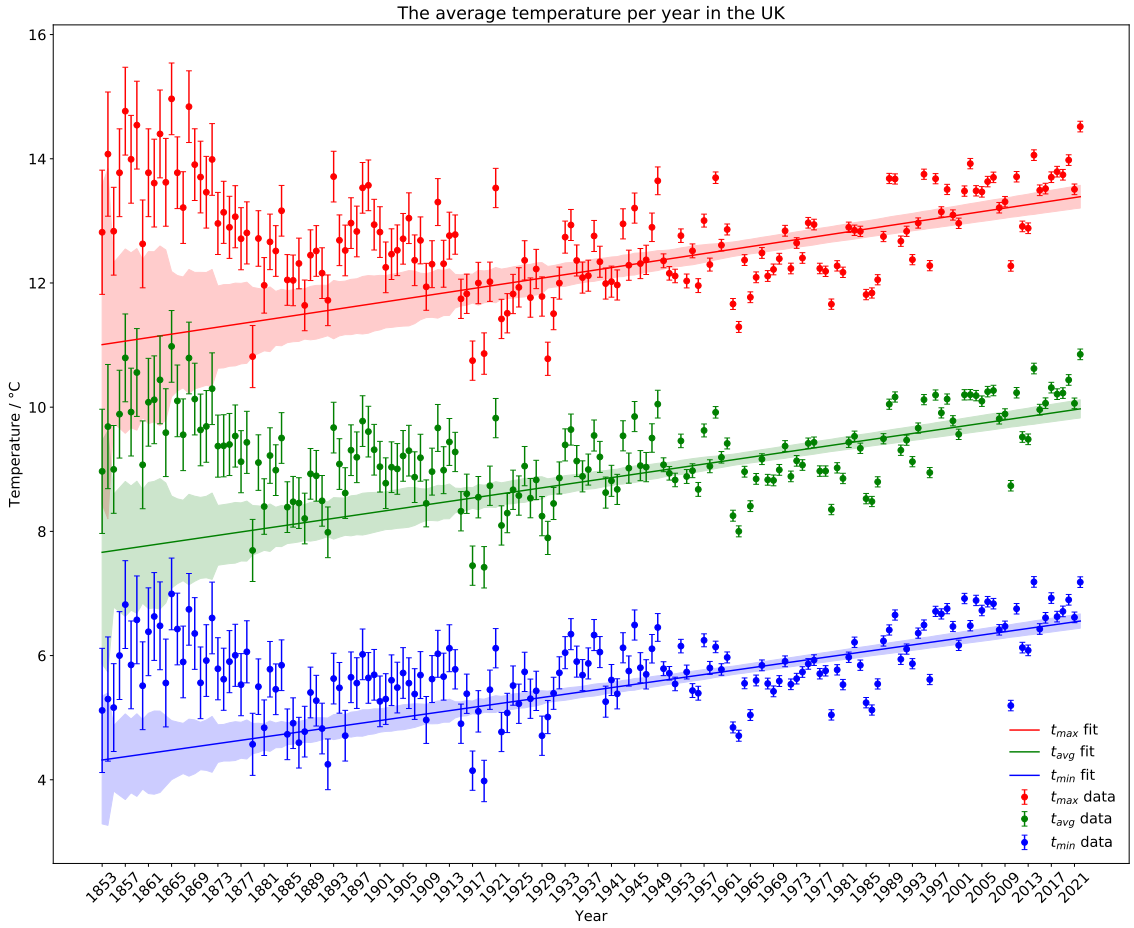


Figure 7: A plot of the average monthly maximum temperature per year $t_{\max}$, average monthly temperature per year t_{avg} and average monthly minimum temperature per year $t_{\min}$ with a linear fit to historical MET office weather data ranging from 1853 to 2022.

The trend-over-time plots for the total number of air frost days, rainfall and sunshine are shown in Figure 8, 9, and 10.

The rate of change, their respective uncertainties and the significance of each slope were determined for each weather variable and are shown in Table 2.

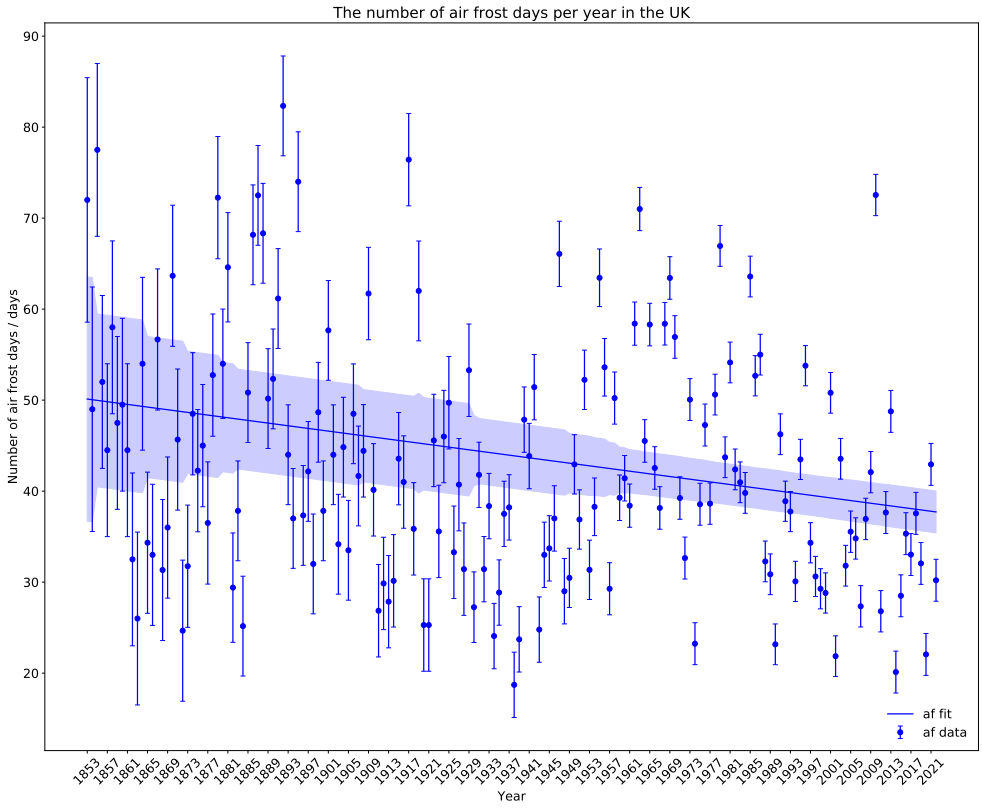


Figure 8: A plot of the average total yearly number of air frost days in the UK with a linear fit to historical MET office weather data ranging from 1853 to 2022.

Table 2: A table showing the rate of change of the average monthly maximum temperature per year $t_{\max}$, average monthly temperature per year t_{avg} and average monthly minimum temperature per year $t_{\min}$, average total yearly number of air frost days, average total yearly rainfall and average total yearly sunshine. The Z-scores for the slope of a linear regression model at 5% significance level are also given.

Variable	**Rate of change/year**	**Z-score (significance) of slope (5% confidence level)**
$t_{\max}$	$(8.16 \pm 0.41) \times 10^{-2}$°C	19.9
t_{avg}	$(8.65 \pm 0.35) \times 10^{-2}$°C	24.5
$t_{\min}$	$(9.22 \pm 0.33) \times 10^{-2}$°C	28.1
af	$-(7.34 \pm 0.66) \times 10^{-2}$ days	-11.1
rain	0.90 ± 0.05 mm	19.4
sun	1.19 ± 0.09 hours	13.6

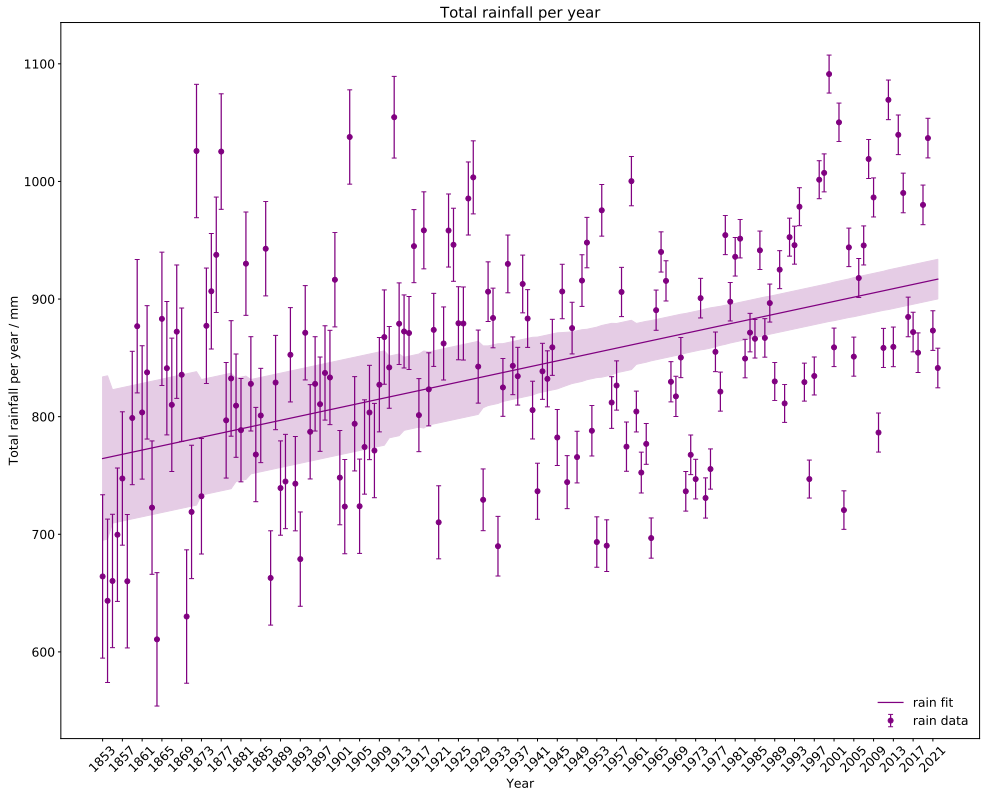


Figure 9: A plot of the average total yearly rainfall in the UK with a linear fit to historical MET office weather data ranging from 1853 to 2022.

To assess the significance of the slope in this study, a statistical hypothesis test was performed using a significance level of 0.05. The null hypothesis was that the slope was equal to zero, implying no significant trend in the data over time. The alternative hypothesis was that the slope was not equal to zero, indicating a significant trend in the data over time. The null hypothesis was rejected for all variables, implying a significant trend over time for all of the variables.

A significant rate of change was determined for each of the six variables, with t_{min} found to have the largest significance of 28.1 and the number of air frost days found to have the smallest significance of -11.1. This supports the results of a 1997 study that found that the minimum temperature increased with a larger rate of change (+1.79°C per 100 years) than the maximum temperature (+0.82°C) [48] (globally). However, the rate of change here was found to be significantly higher than in the cited study.

There are some consequential problems with this method and the reliability of its results. The significance is calculated assuming that the errors in the model are normally distributed, that the observations are independent and that the most appropriate model is linear. These assumptions are made to simplify the analysis, but a more thorough investigation using traditional statistical methods

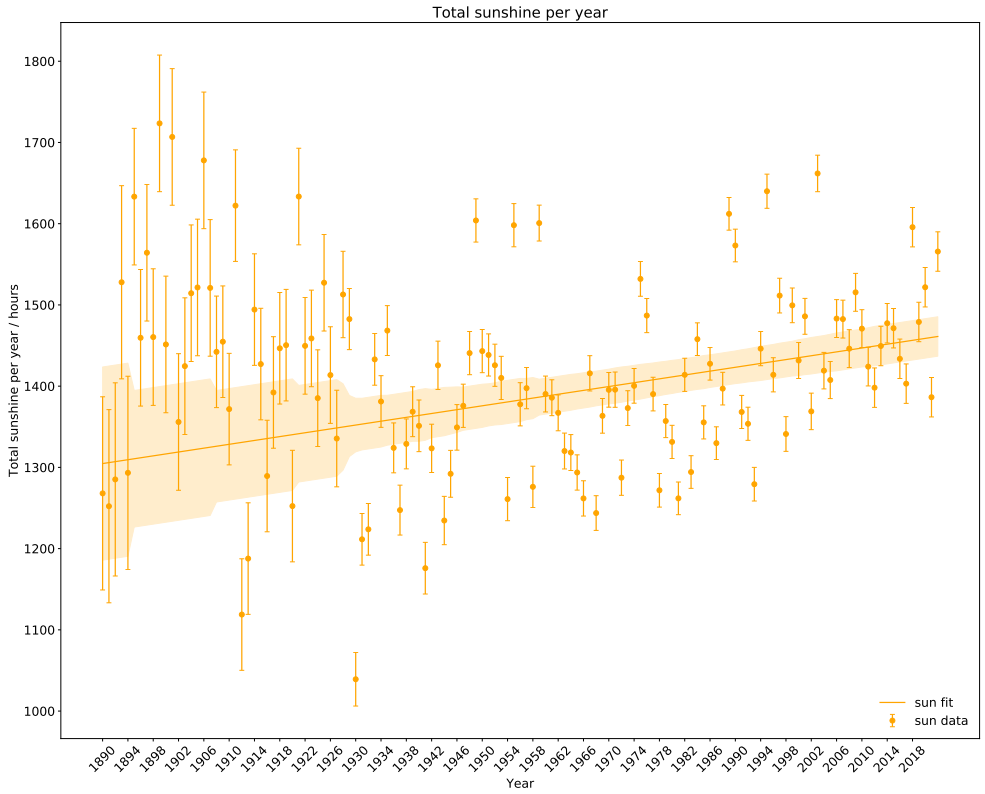


Figure 10: A plot of the average total yearly number of sunshine hours in the UK with a linear fit to historical MET office weather data ranging from 1853 to 2022.

should carefully evaluate these assumptions.

A variable of interest is the change in temperature since pre-industrial times. The average global temperature has increased by at least 1.1°C since 1880[2] , with most of the warming occurring since 1975 [50]. However, this does not mean that every region is experiencing the same amount of warming. The UK is one of the regions that, according to the Met Office, has seen a higher-than-average increase in temperature [51]. The Met Office reports that the increase in temperature since pre-industrial times is estimated to be 1.2°C.

The simple model in this study could be used to make the same predictions for each temperature variable. The fit to the data was used to estimate $t_{\max}$, t_{avg}, and $t_{\min}$ for 1880, 1900, 2022 and 2100. The 1880 values were used as a benchmark to measure the change in temperature since pre-industrial times. These values are shown in Table 3, and the changes in the other weather variables, the total number of air frost days per year, total rainfall per year and total sunshine per year, are given in Table 4.

[2]The global average temperature in 1880, though not historically pre-industrial, is used as a benchmark for the pre-industrial temperature. This is because global temperature records began around 1880, and the change in temperature in that time would be sufficiently small to permit the use of 1880 as the reference value [49].

Table 3: Predictions for $\Delta t_{\max}$, Δt_{avg}, $\Delta t_{\min}$, where each variable is the change in the respective temperature measurement since 1880. These changes are given for 1900, 2022 and 2100.

Year	$\Delta t_{\mathbf{max}}$ **(± 0.8) / °C**	$\Delta t_{\mathbf{avg}}$ **(± 0.7) / °C**	$\Delta t_{\mathbf{min}}$ **(± 0.6) / °C**
1900	0.2	0.2	0.2
2022	1.1	1.2	1.3
2100	1.8	1.9	2.0

Table 4: Predictions for the total number of air frost days per year, total rainfall per year and total sunshine per year, where each variable is the change in the respective temperature measurement since 1880. These changes are given for 1900, 2022 and 2100.

Year	**Δaf (± 13) / days**	**Δrain (± 91.6) / mm**	**Δsun (± 172.1) / hours**
1900	-4	45.1	59.3
2022	-13	155.3	203.9
2100	-18	225.7	296.3

According to this simple model, the average monthly temperature in the UK is predicted to increase by (1.2±0.7)°C in 2022 relative to pre-industrial times. This value agrees with the Met Office reported value of (1.2±0.12)°C, implying that, at least within the limit it is used, it can provide an accurate estimate of UK temperatures.

5.1.2 2022 study

Evaluating the significance of a particular year can provide valuable insights into the impacts of climate change on the environment. A 2020 Met Office study described that in 2020, summers which see days above 40°C somewhere in the UK have a return time of 100-300 years, and without mitigating GHG emission, this can decrease to 3.5 years by 2100 [52]. Records began being broken sooner than expected. Two significant milestones were passed in 2022 - a daily maximum temperature of more than 40°C was reached, and a national average temperature over the year of more than 10°C was passed [53].

These record-breaking milestones meant 2022 is the warmest year on record for the UK. Figure 11 shows the daily UK average temperatures through 2022. The orange peaks indicate periods above the average for the time of the year and the blue troughs show below-average periods. Note that the

plot shows the average temperature for the UK as a whole, so the peaks and troughs are less extreme than those seen in individual locations.

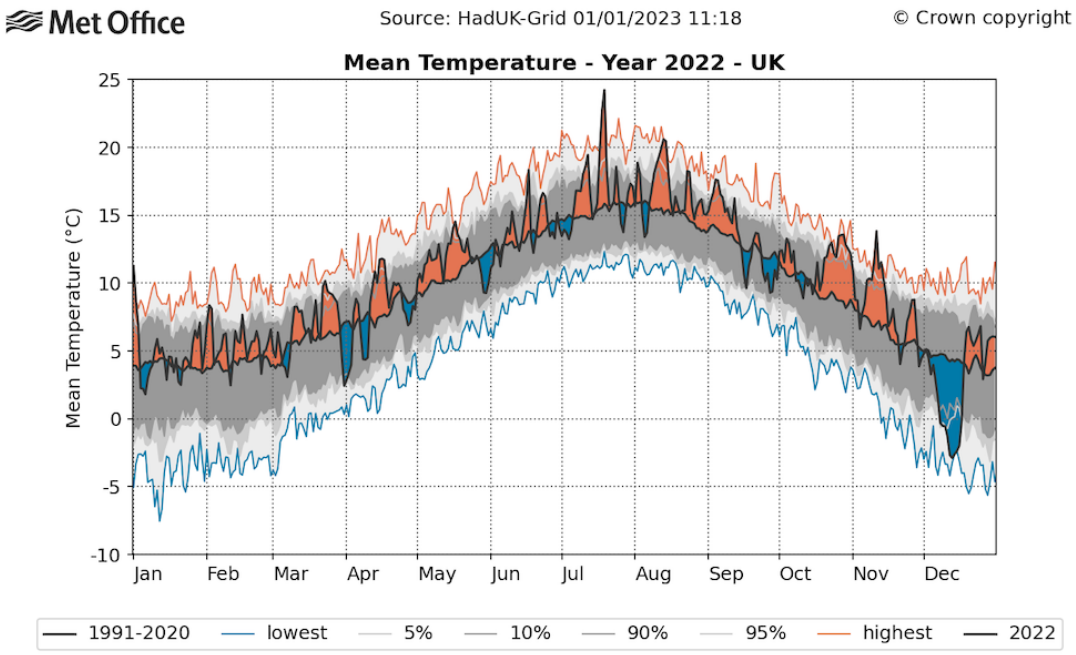


Figure 11: A plot of daily UK average temperature for 2022. The orange shaded areas are periods of above average temperature, blue shading is below average, and the solid black line is the average temperature for 1991-2020. The grey shading indicates the 5th, 10th, 90th and 95th percentiles of the temperature distribution and the red and blue lines are the highest and lowest values for each day of the year based on daily data from 1960 to 2022 [53].

In this section, we analyse the significance of 2022 compared to all historical UK climate data and try to confirm results from other studies. Two approaches will be employed to test the significance of 2022. First, the most exceptional years for the temperature variables will be ranked and compared with 2022. Second, distributions of each climate variable are shown, and the statistical significance of the 2022 averages with respect to the total mean is presented.

As shown in Figure 12, 2022 had the second-highest average temperature in the UK, behind 1865. For $t_{\max}$ and $t_{\min}$, 2022 was fifth and second, respectively, providing further evidence that 2022 was an exceptional year for temperature records in the UK. The bar charts for the top five years with the highest $t_{\max}$ and $t_{\min}$ are provided in Appendix A.

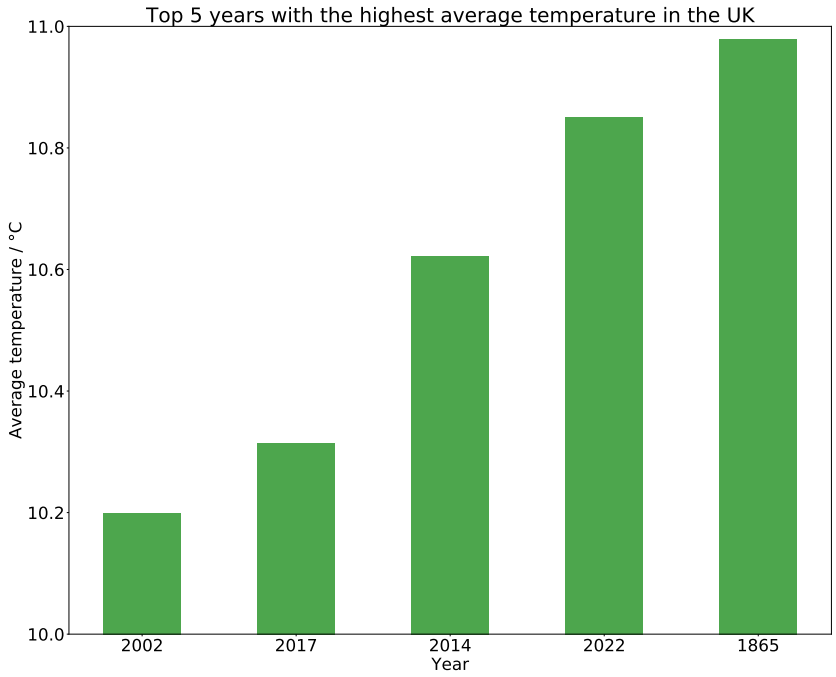


Figure 12: A bar chart of the top five years with the highest average temperature in the UK.

Another way of analysing the significance of 2022 is to produce distributions for each variable and determine the significance of the maximum 2022 value with respect to the mean of the distribution. Distributions of each temperature variable, t_{max}, t_{avg} and t_{min}, are shown in Figure 13. Note that the values being distributed are the yearly average values for each station, i.e. the mean has not been taken for each year over all stations. If this were the case, there would be too little data to fit a Gaussian curve, resulting in a poor estimation of the significance of the maximum 2022 values.

For each distribution, the significance of the maximum 2022 temperature value is greater than 2σ. This indicates that the maximum 2022 temperatures are unlikely to be caused by statistical fluctuations in the average temperatures of the UK. Furthermore, the statistical significance increases for the 2022 maximum of all three variables when taken with respect to pre-1990 data, providing further evidence of climate change.

The distributions for the yearly average total number of air frost days, rainfall, and sunshine are given in Appendix A.

5.1.3 Conclusions

The results of the basic trend-over-time model indicate that the UK's average monthly temperature is expected to change by (1.2±0.7)°C, which aligns with the reported value from the Met Office of (1.2±0.12)°C. This suggests that the model may offer a reasonably accurate estimate of UK temper-

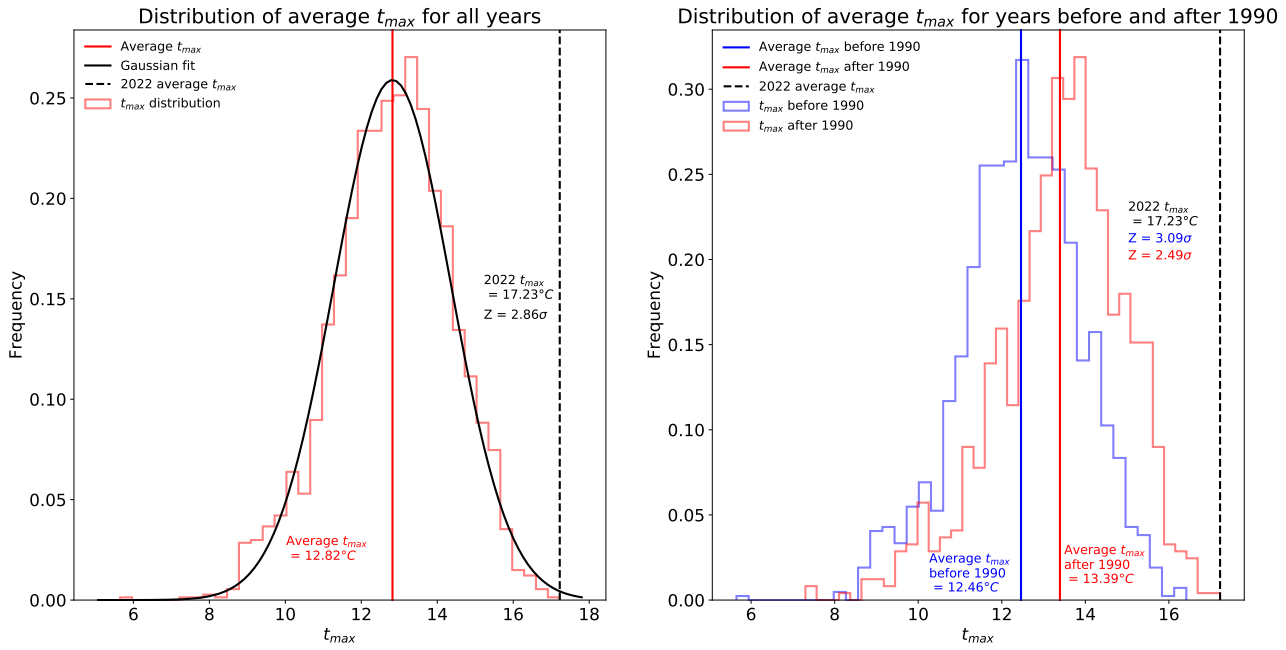


Figure 13: Histograms of the the yearly average $t_{\max}$ values for each weather station. The left panel shows the distribution for all years with a Gaussian fit. The average $t_{\max}$ value is 12.82°C. The maximum 2022 $t_{\max}$ value is 17.23°C, and its significance with respect to the mean is $Z = 2.86\sigma$. The right panel shows distributions of $t_{\max}$ for before and after 1990. The average $t_{\max}$ for before 1990 is 12.46°C, and 13.39°C after. The significance of the maximum 2022 $t_{\max}$ value with respect to before 1990 is $Z = 3.09\sigma$ and $Z = 2.49\sigma$ after.

atures, at least within its designated limits. It should be noted, however, that the model relies on several assumptions and may not be reliable for long-term forecasts. Furthermore, the high uncertainties associated with all projected changes suggest that the trends and values should be cautiously interpreted. Despite its limitations, the model can serve as a straightforward tool for predicting and visualising changes in the UK's climate. It is similar in utility to the simple zero or one-dimensional energy balance model described in Section 2.1.

The 2022 study has shown that 2022 was significant in terms of weather in the UK. It adds to the growing trend of a rapidly changing climate in the UK and the rest of the world.

5.2 Machine Learning

In this investigation, classification algorithms are used to try and categorise data into either before or after 1990. If the model successfully classifies data based on the climate variables, it would provide additional proof of climate change (assuming that the measurement equipment and standards remain constant). The year 1990 holds significance in the context of climate change in the UK, as it serves

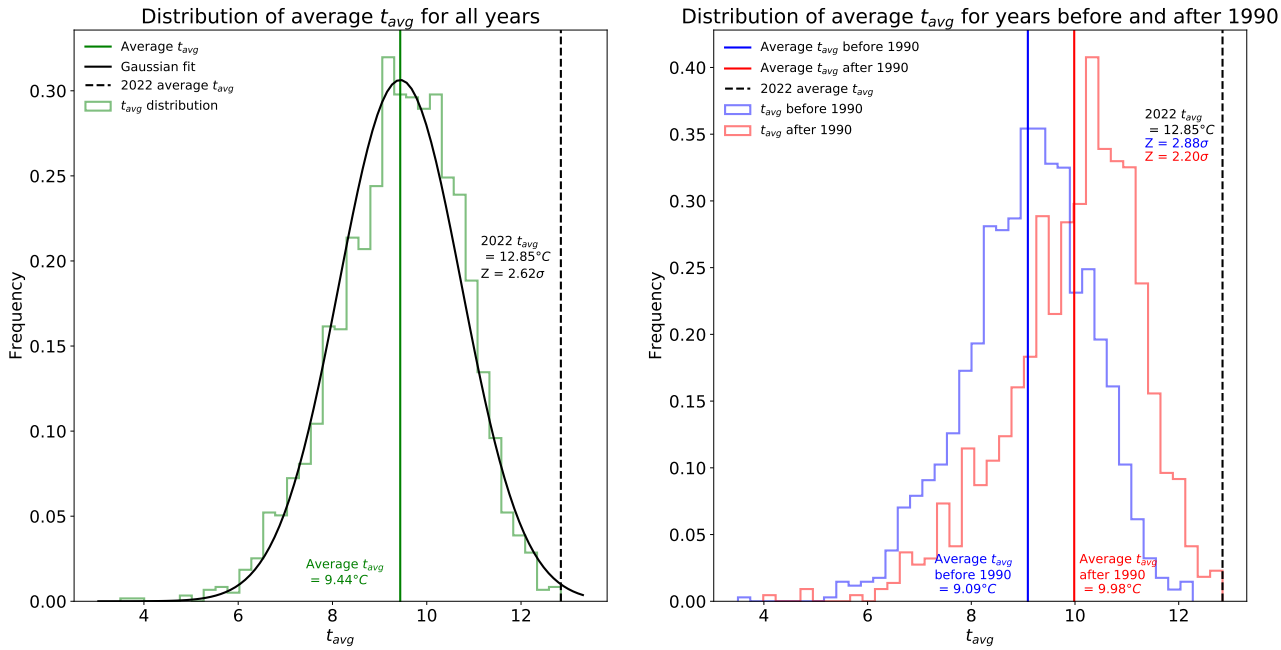


Figure 14: Histograms of the the yearly average $t_{\max}$ values for each weather station. The left panel shows the distribution for all years with a Gaussian fit. The average $t_{\max}$ value is 12.82°C. The maximum 2022 $t_{\max}$ value is 17.23°C, and its significance with respect to the mean is $Z = 2.86\sigma$. The right panel shows distributions of $t_{\max}$ for before and after 1990. The average $t_{\max}$ for before 1990 is 12.46°C, and 13.39°C after. The significance of the maximum 2022 $t_{\max}$ value with respect to before 1990 is $Z = 3.09\sigma$ and $Z = 2.49\sigma$ after.

as a benchmark for evaluating recent changes in temperature and weather. According to a report by the Environmental Agency, the ten warmest years in the UK have all occurred since 1990, and the 21st century has been warmer in central England than the previous three centuries [25]. Using 1990 as a reference point allows for a comparison of the current state of the climate with a period before substantial human-induced alterations to the atmosphere. This approach is used to try and provide a unique method of attaining evidence of climate change, along with insights into climate change that traditional statistical methods cannot provide, thereby providing a better understanding of the underlying physics of the climate system.

There are various ways to transform or redefine variables to improve the accuracy of a ML model. This study takes an iterative approach, beginning with the simplest case and iteratively introducing new transformations to the data to test for performance improvement. First, the nominal case is presented, which is the five variables from the raw data (with some simple redefinitions). Next, new variables are introduced that use data that would have otherwise been discarded in the nominal case. Finally, models are trained on the raw data (with some transformations to allow the data to be parsed

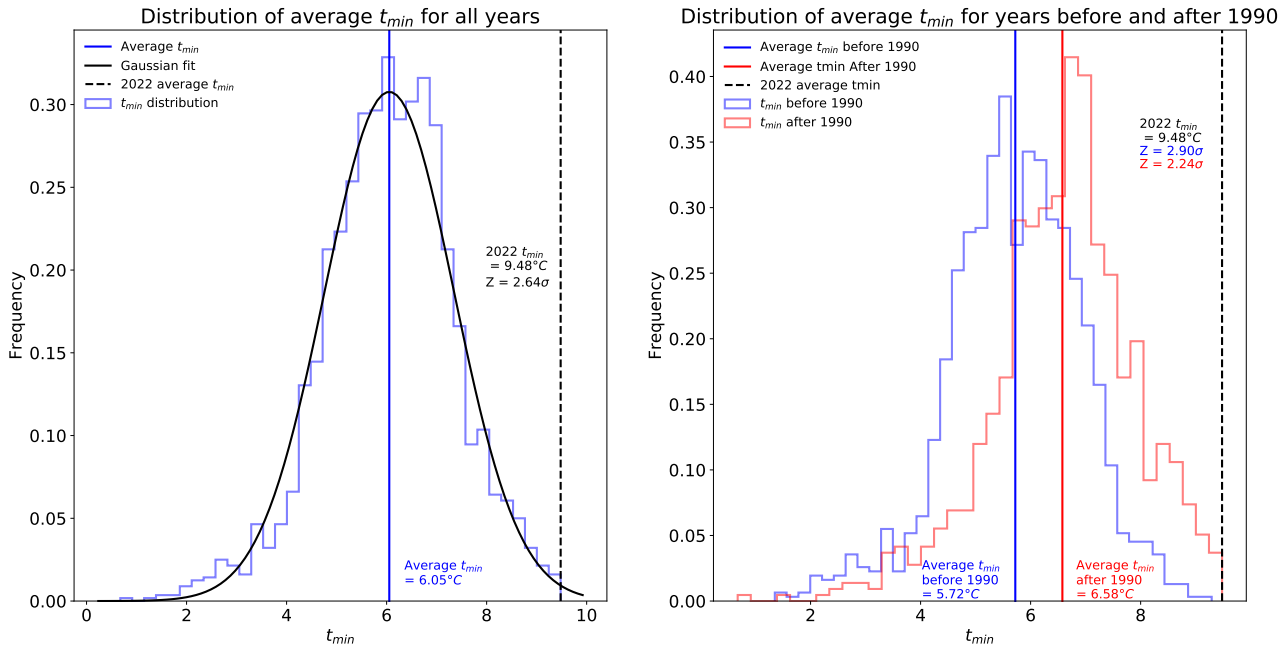


Figure 15: Histograms of the the yearly average $t_{\max}$ values for each weather station. The left panel shows the distribution for all years with a Gaussian fit. The average $t_{\max}$ value is 12.82°C. The maximum 2022 $t_{\max}$ value is 17.23°C, and its significance with respect to the mean is $Z = 2.86\sigma$. The right panel shows distributions of $t_{\max}$ for before and after 1990. The average $t_{\max}$ for before 1990 is 12.46°C, and 13.39°C after. The significance of the maximum 2022 $t_{\max}$ value with respect to before 1990 is $Z = 3.09\sigma$ and $Z = 2.49\sigma$ after.

to the model).

5.2.1 Nominal case

The nominal case is the simplest set of variables that will be used to train a model. To successfully perform a classification, the data set must have one set of measurements (each of the five weather variables) for a given year and station. The classification becomes impossible if the model is trained on the raw data, as it contains many conflicting measurements for a given year (i.e. measurements for each month but with the same year label). To solve this problem, the nominal five variables are redefined. The $t_{\max}$ for a given year is redefined from the raw data set as the average of the original $t_{\max}$ for the summer months (June to August) and the same for $t_{\min}$ but for the winter months (December to February). The air frost days, sun exposure and rainfall variables are redefined as the sum of the original variables for all months of a given year.

As mentioned in Section 4.2.1, during the necessary preprocessing step, much of the data is discarded if it is incomplete. Due to the nature of the data set, this is common, meaning that for the

nominal case, a considerable fraction of useful data is discarded.

The correlation matrix for the nominal data set is shown in Figure 16. Here, only the nominal five variables were used. As expected, $t_{\max}$ is correlated with sun exposure, implying that cloud coverage strongly impacts the maximum temperature. We also find that $t_{\min}$ is negatively correlated with the number of air frost days, indicating that the higher the minimum temperature, the less likely an air frost day is to occur. These are all apparent insights, but this method will become more useful once we expand past the nominal data set and introduce new variables. However, one unique insight the correlation can provide in the nominal case is the correlation of a weather variable with the target variable, i.e. how much a given weather variable influences the BDT model. The weather variable most strongly correlated with the target variable is $t_{\min}$, indicating that the minimum temperature is potentially a variable of interest concerning the physics of climate change. Additionally, $t_{\max}$ shows a strong negative correlation with the rainfall variable. This insight is not necessarily obvious, and so provides further evidence of the benefits of this approach compared to the traditional statistical analysis.

Figure 16: The correlation matrix for the nominal data set, the standard 5 variable case.

An AdaBoost classifier using a decision tree base estimator was trained on the nominal data set. A parameter search was performed using a grid search. The optimal hyperparameters and classifier performance is shown in Table 5.

Table 5: The training and test set score, 95% confidence interval, and optimal hyperparameters for BDTs trained on the nominal data set, with and without feature selection.

	Nominal (without feature selection)	**Nominal (with feature selection)**
Training set score	0.82	0.79
Test set score	0.65	0.65
Confidence interval (95%)	[0.60, 0.70]	[0.61, 0.70]
Max depth	2	2
Minimum samples per leaf	20	20

Next, the features were ranked using the F-test, a statistical test that measures the variability between two sets of observations. The F-value is found between each feature and the target variable and is then ranked based on their relative 'importance' to the target variable. This is analogous to the correlation of the variables to the target variables in the correlation matrix. Figure 17 shows the relative feature importance for each feature. This feature ranking method can also be used to select the 'K' best features, i.e. only select the most important features for training the model. Decreasing the number of features is a dimensionality-reducing technique and can help mitigate overtraining. Using this ranking, the top four features were selected, and a BDT was trained on the reduced data [54].

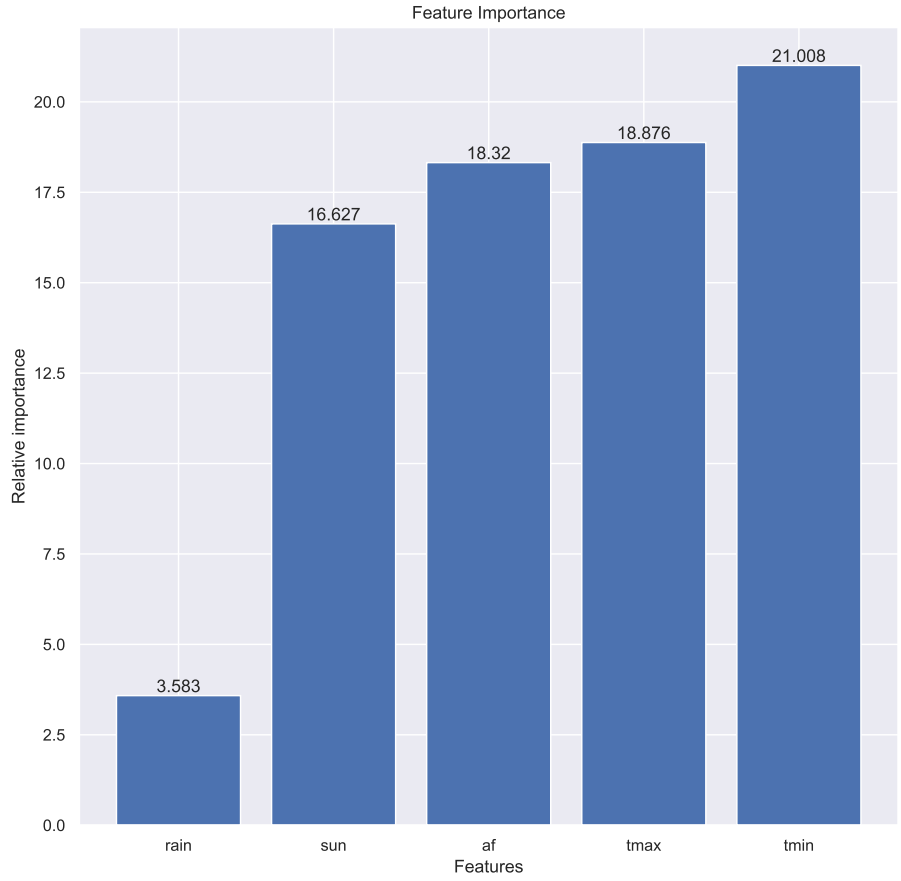
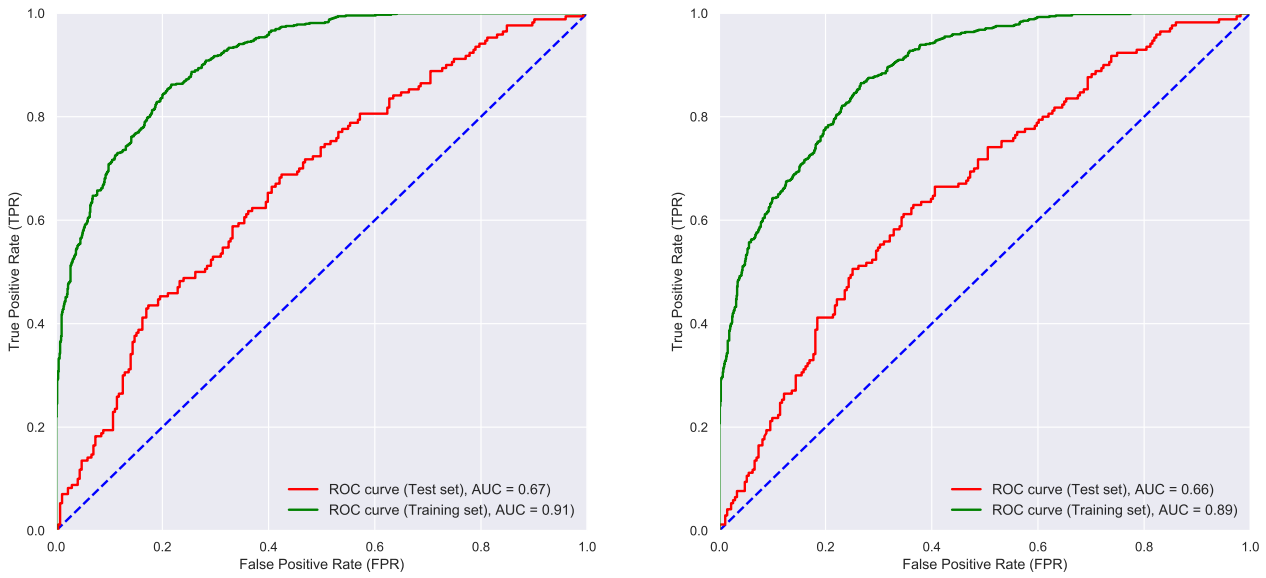


Figure 17: A bar chart of the relative feature importance for the nominal case, ranked using an F-test.

The feature importances here broadly confirm the findings of the correlation matrix. However, a naive look at the correlation matrix concludes that t_{min} is as important as the air frost variable. In contrast, the F-test feature rankings indicate the air frost days variable is slightly less important than t_{max}.

The Receiver Operating Characteristic (ROC) curve for the performance of the BDT with and without feature selection is shown in Figure 18. The ROC curve is a graphical representation of the performance of a binary classifier system as the discrimination threshold is varied. It visually represents the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR). TPR is the proportion of positive instances that are correctly classified as positive, and the FPR is the proportion of negative instances that are incorrectly classified as positive. A good classifier will have a high TPR and a low FPR, meaning it correctly identifies positive instances as positive and correctly identifies negative instances as negative.



(a) (b)

Figure 18: The Receiver Operating Characteristic (ROC) curve for a BDT trained on the nominal data set for: (a) no feature selection, and (b) with four features selected out of the nominal five.

The Area Under the Curve (AUC) is the area under a ROC curve and is a scalar value that summarises the performance of a binary classifier. A perfect classifier will have an AUC value of 1, meaning that the ROC curve passes through the upper left corner of the graph (100% true positive rate and 0% negative rate). A random classifier will have an AUC value of 0.5, providing no better performance than random guessing.

The AUC value is a particularly useful metric in this case, as our classes are imbalanced. Given that the historic data spans over 100 years and the threshold year is chosen to be 1990, meaning there will be more instances of data from before 1990 than after 1990, the standard accuracy metric for the classifier may not provide an adequate measure of the classifier's performance [55]. This is because a classifier that predicts the majority class for all instances will achieve a greater-than-chance accuracy, even if it fails to classify any instances of the minority class correctly. In cases such as this, the AUC value of the ROC curve provides a more meaningful and nuanced measure of the performance of the classifier.

The ROC curves in Figure 18 show a large separation between the training and test sets, indicating a large degree of overfitting. The ROC curve for the BDT with feature selection is shown in Figure 18b. It displays worse performance on the training and testing set, indicating that, at least for the nominal case, feature selection is insufficient to improve the performance and deal with the low quantity of data.

The BDT output is a score assigned to each input event, representing the likelihood that the data point belongs to before or after 1990. The output score is obtained by combining the predictions of multiple decision trees, each of which splits the data based on a selected feature and assigns a score to each leaf node.

The BDT output for the model with feature selection is visualised in Figure 19 as a plot that shows the distribution of scores for the entire data set. Plots with a high degree of separation between the two classes indicate good performance. The plot shows a significant overlap between the two classes, indicating poor performance with this data set.

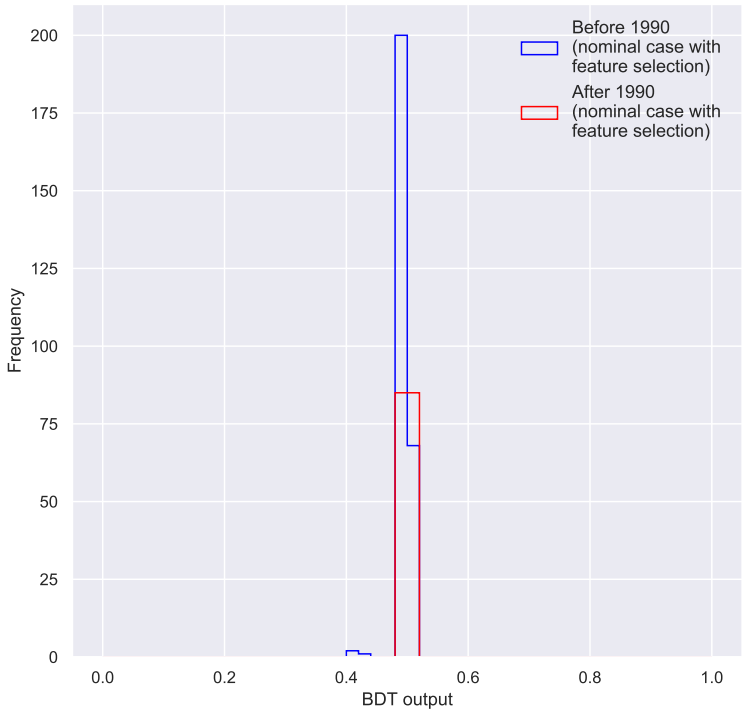


Figure 19: Distributions of the BDT output for the two classes: before and after 1990 (nominal case).

5.2.2 Introducing new variables

The raw data provided by the MET office only contains five weather variables, which, paired with the relatively small data set, imposes some limitations on the potential strength of the ML model that can be trained. It is possible, however, to create new variables that provide some new information to the model. As explained in Section 4.2.1, much of the data in the nominal case is removed due to necessary preprocessing steps. It is possible to find novel ways of transforming the data such that the model can draw new helpful information from it. In this section, we propose the introduction of new variables, which can be provided alongside the initial nominal five variables to improve the model's accuracy and reduce overfitting.

Here we define several new variables. As discussed in Section 5.2.1, for the nominal case, $t_{\max}$ for a given year is redefined from the raw data set as the average of the original $t_{\max}$ for the summer months (June to August), and the same for $t_{\min}$ but for the winter months (December to February). The air frost days, sunshine and rainfall variables are redefined as the sum of the original variables for all months of a given year.

The new variable $(t_{\max})_{\max}$ is defined as the maximum value of the original $t_{\max}$ for a given year - so essentially, the average monthly maximum temperature for the warmest month of the year. $(t_{\min})_{\min}$ is defined as the minimum value of the original $t_{\min}$ for a given year - so essentially, the average monthly minimum temperature for the coldest month of the year. For air frost days, sun exposure and rainfall we also define $\text{af}_{\max}$, $\text{af}_{\min}$, $\text{rain}_{\max}$, $\text{rain}_{\min}$, $\text{sun}_{\max}$ and $\text{sun}_{\min}$. These are each variable's maximum and minimum values for a given year. Introducing these new variables provides the model with information that would have been lost in the nominal case.

The correlation matrix for the data set with additional variables is shown in Figure 20.

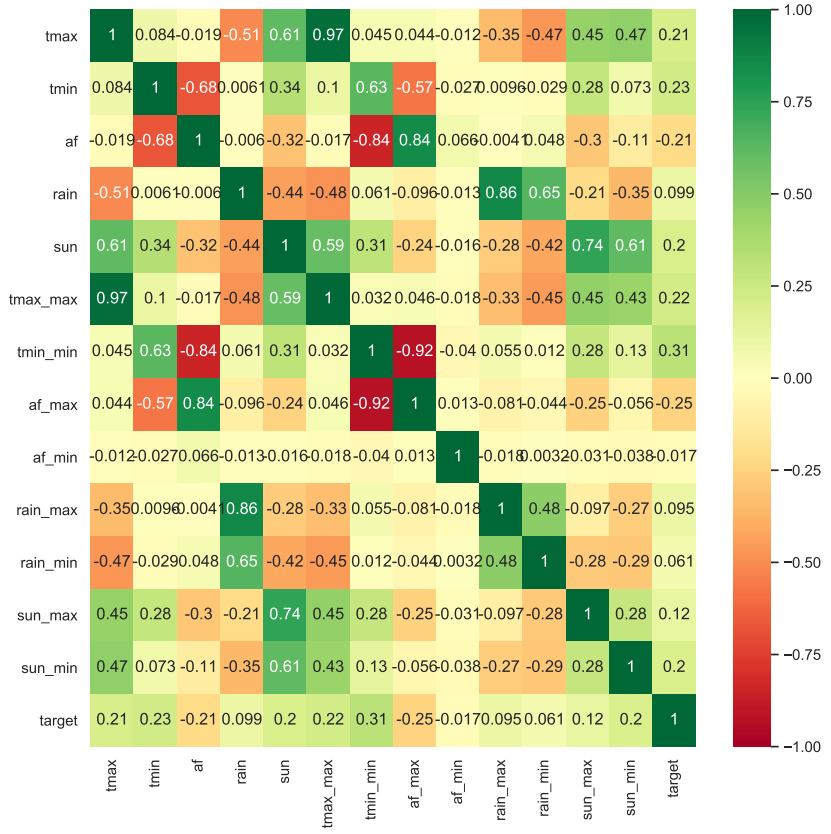


Figure 20: The correlation matrix for the case of the additional variable, composed of the nominal five variables and eight additional variables.

Introducing new variables revealed new strong correlations. For example, $(t_{\min})_{\min}$ is 92% nega-

Table 6: The training and test set score, 95% confidence interval, and optimal hyperparameters for BDTs trained on the additional features data set, with and without feature selection.

	Additional variables (without feature selection)	**Additional variables (with feature selection)**
Training set score	1	0.76
Test set score	0.71	0.65
Confidence interval (95%)	[0.70, 0.78]	[0.61, 0.70]
Maximum depth	3	2
Minimum samples per leaf	50	20
Maximum features	3	7
Number of estimators	50	100

tively correlated with af_{max}. Interestingly, none of the variables have a particularly high correlation with the target variable, indicating that this data set may still not be enough to perform well in the classification.

An AdaBoost classifier was trained on a data set including the new variables. A grid search was performed over a broader range of possible hyperparameters with 5-fold cross-validation. The performance of the model and the optimal hyperparameters are shown in Table 6, and Figure 21 shows the relative feature importance for each feature.

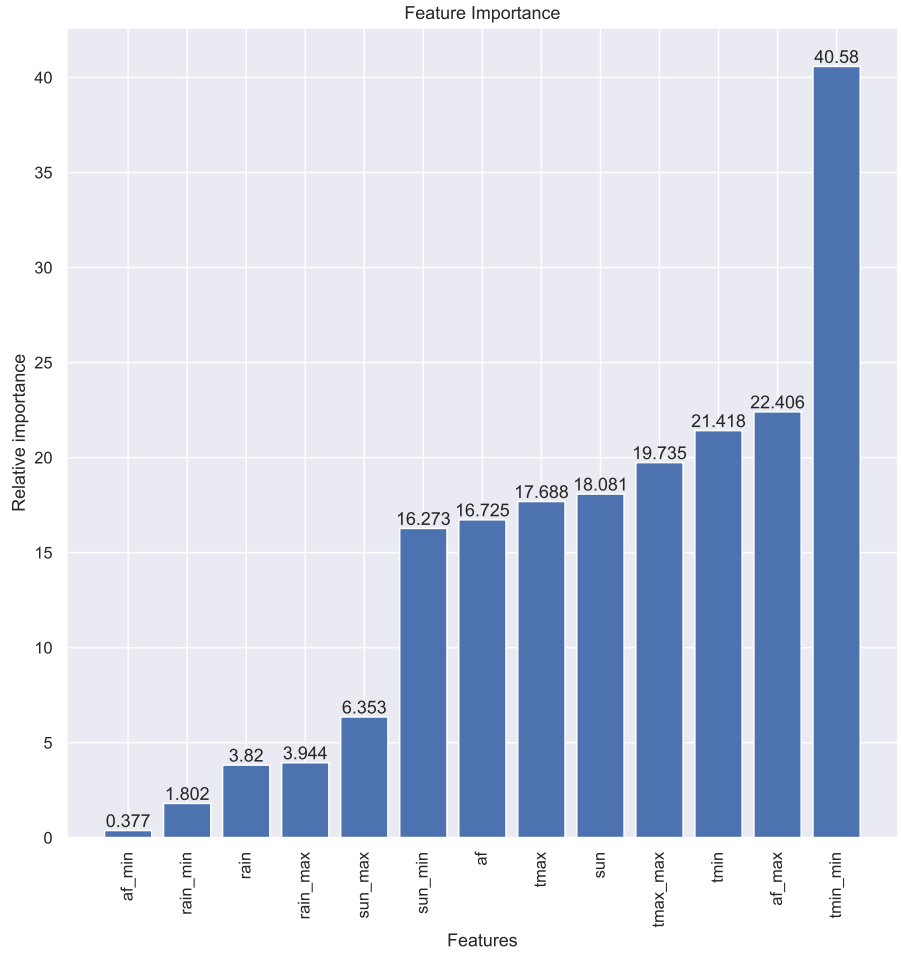
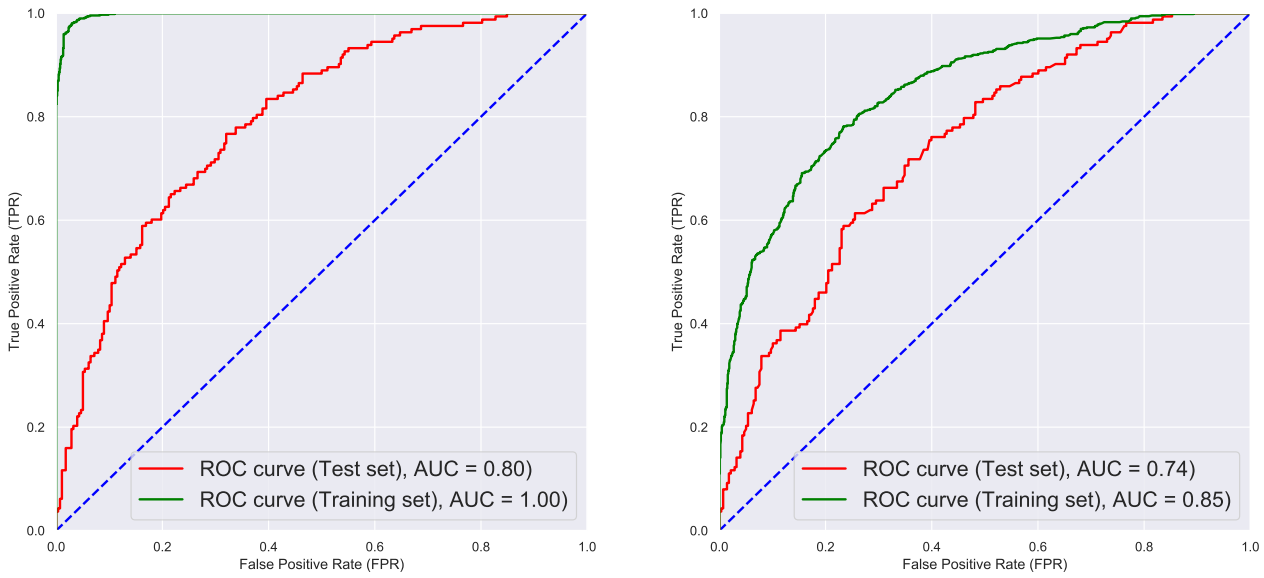


Figure 21: The ranked features for the additional variables case.

From Figure 21, we can see that certain variables have low feature importance relative to the other features, namely rain, af_{min}, rain_{max}, rain_{min} and sun_{max}. This indicates that these features could be removed, reducing the model's complexity and overfitting. Given this, the top eight out of the possible 13 features were selected, and an AdaBoost classifier was retrained on the reduced data set. The results of the grid search are given in Table 6.

The ROC curves in Figure 22 still show a significant separation between the training and test sets, indicating a substantial degree of overfitting. However, the performance on the test set has somewhat improved from the nominal case. For no feature selection, the AUC score went from 0.67 for the test set to 0.80. Although there is still a high degree of overfitting, the improved test set performance indicates that the approach of introducing new variables is promising. The ROC curve for the BDT with feature selection is shown in Figure 22b. It displays worse performance on the training and testing set, indicating that feature selection is still insufficient for the additional variables case to improve the performance and deal with the low quantity of data.



(a) (b)

Figure 22: The Receiver Operating Characteristic (ROC) curve for a BDT trained on the additional variables data set for: (a) no feature selection, and (b) with eight features selected out of the possible 13.

Finally, the BDT output distributions for the additional variables case with feature selection are shown in Figure 23. A slightly higher degree of separation can be seen; however, there is still significant overlap, indicating poor performance overall.

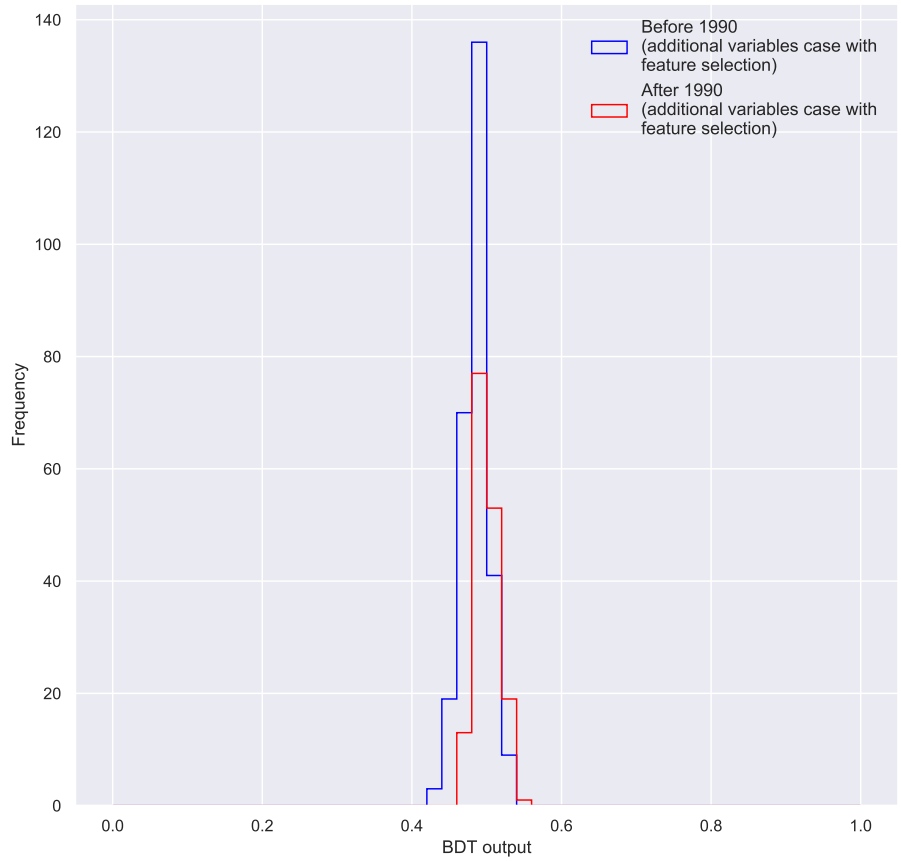


Figure 23: Distributions of the BDT output for the two classes: before and after 1990 (additional variables case).

5.2.3 Complete data

So far, BDTs have been trained on the nominal case (five variables, one value for each year and each station) and the added variable case. The next approach is to transform the data to be parsed to the BDT in its complete form without destroying any information. In this approach, a new variable is created for each weather variable for each month. We end up with a data set of 60 total variables (the five initial variables multiplied by the 12 months). For example, instead of having a $t_{\max}$ for each year and station, we now have $t_{\max 1}$, $t_{\max 2}$, ... $t_{\max 12}$ for each year and station. The nominal data set contained 2205 rows x 5 columns, equaling 11025 total data points. The added variable case contained 2205 rows x 13 columns, equaling 28665 total data points. The complete data case contains 2205 rows x 60 columns, totaling 132300 data points. The benefit of this approach is that it increases the size of the data set without adding too much unnecessary noise or complexity (which could lead to overfitting and a reduction in model performance) and prevents the destruction of potentially useful information by deletion or taking averages for certain months.

The correlation matrix for the complete data set is shown in Figure 24. As expected, many highly correlated variables mean more information for the BDT. However, many weakly correlated areas of the matrix also indicate that feature selection will be necessary to remove useless features that add noise to the data.

The correlation matrix for the complete data set is shown in Figure 24.

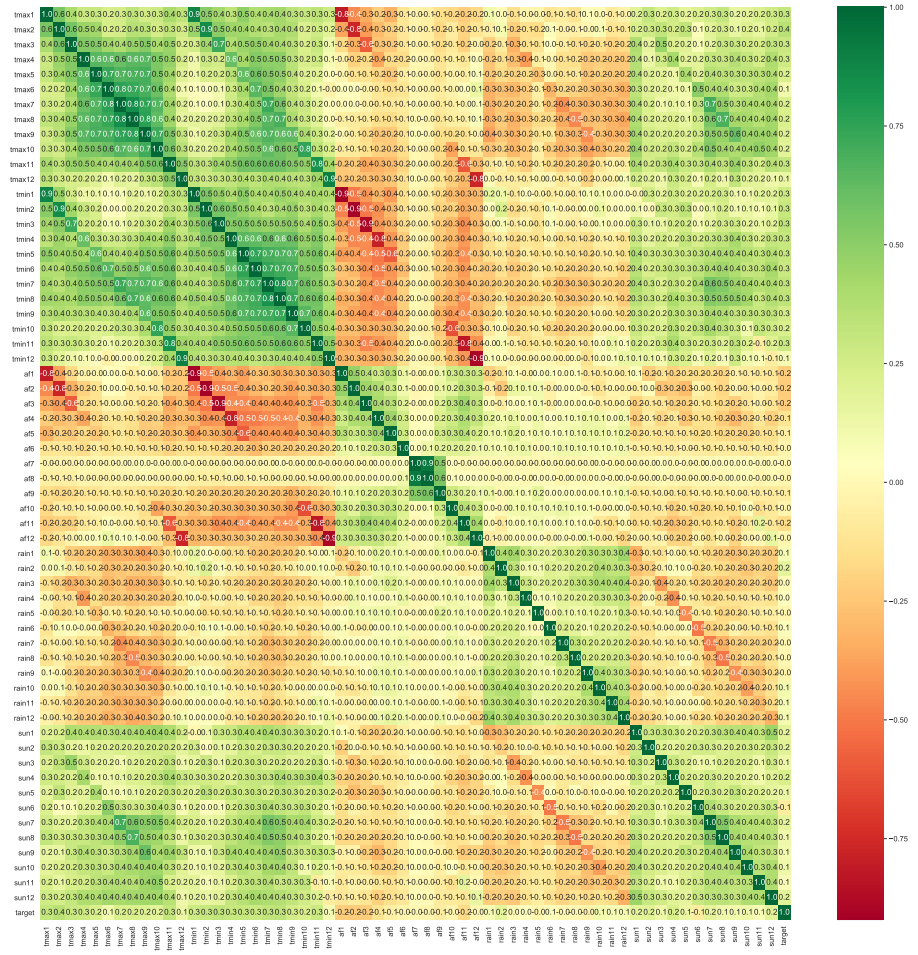


Figure 24: The correlation matrix for all 60 variables of the complete data set.

The correlation matrix shows that variables from the same month tend to be more strongly correlated. it shows that the number of air frost days in the winter months is strongly negatively correlated with the $t_{\max}$ for those months. Additionally, the target variable is still weakly correlated with most of the features, indicating that no one variable strongly impacts the classification performance. This is, however, the benefit of ML models over traditional statistical analysis. They extract information from the many weakly correlating variables to successfully classify new data.

Again, an AdaBoost classifier was trained on the complete data set, and a grid search was performed. The performance of the model and the optimal hyperparameters are included in Table 7.

Table 7: The training and test set score, 95% confidence interval, and optimal hyperparameters for BDTs trained on the complete data set, with and without feature selection.

	Complete data (without feature selection)	**Complete data (with feature selection)**
Training set score	1	0.95
Test set score	0.91	0.87
Confidence interval (95%)	[0.88, 0.93]	[0.83, 0.90]
Maximum depth	5	3
Minimum samples per leaf	10	10
Maximum features	7	20
Number of estimators	100	30

Figure 25 shows the relative importance of each of the 60 features.

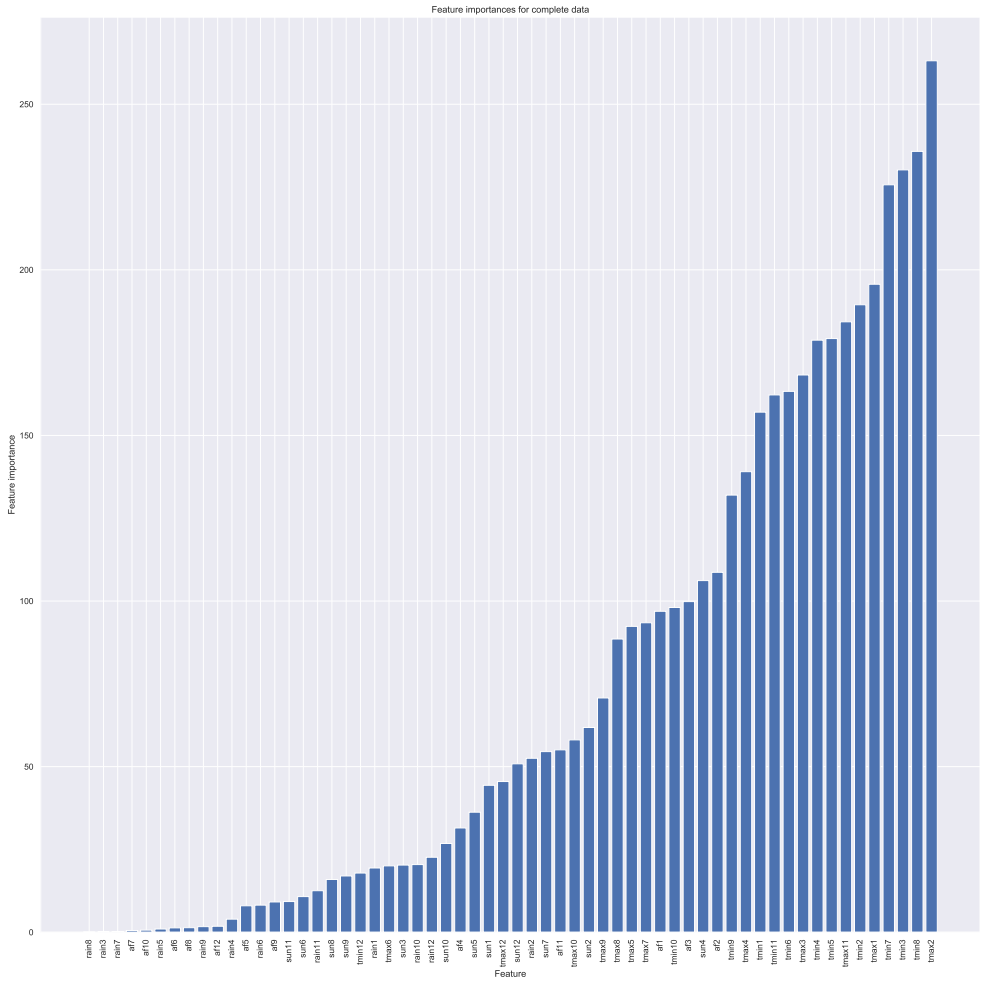
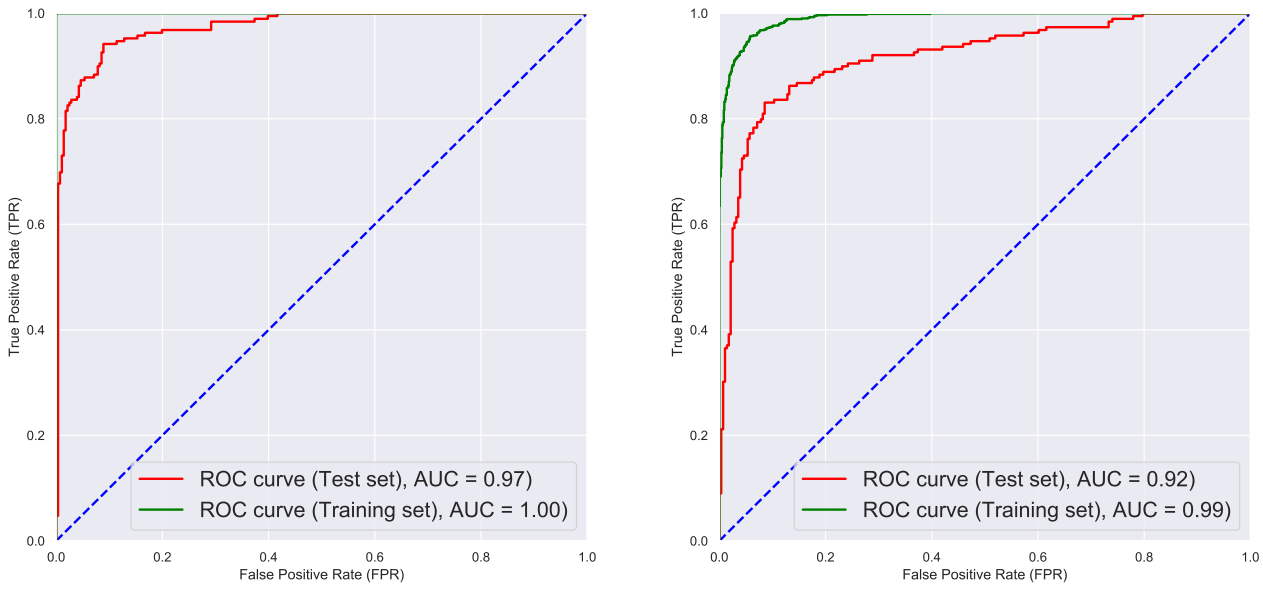


Figure 25: The ranked features for the complete data set.

As shown in Figure 25, many of the features show low relative feature importance. From this, the top 23 most important features were selected, and a new BDT model was trained on the reduced data set. The number 23 was selected after a trial and error approach to identify what reduced overfitting the most while maintaining a high performance level.

Figure 26 shows the ROC curve for the BDT trained on the complete data for the case with and without feature selection.

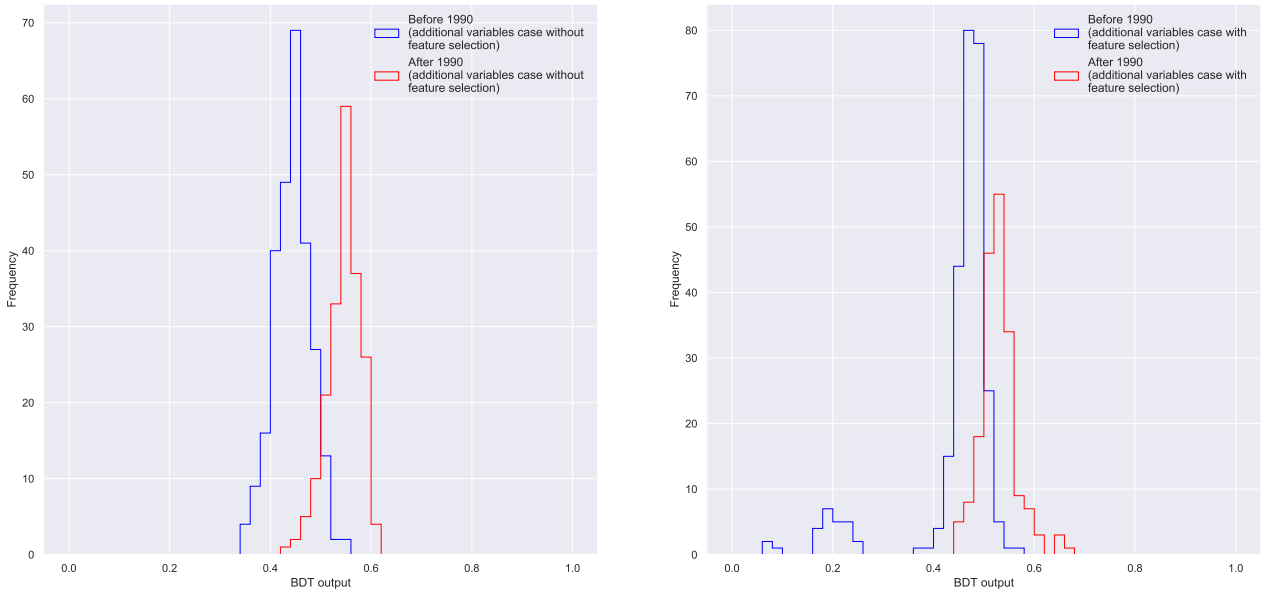


(a) (b)

Figure 26: The Receiver Operating Characteristic (ROC) curve for a BDT trained on the complete data set for: (a) no feature selection, and (b) with 23 features selected out of the possible 60.

In this case, feature selection was not enough to significantly reduce overfitting. It again reduced the model's performance on the test set, as indicated by the reduced AUC score in the feature selection ROC curve in Figure 27b.

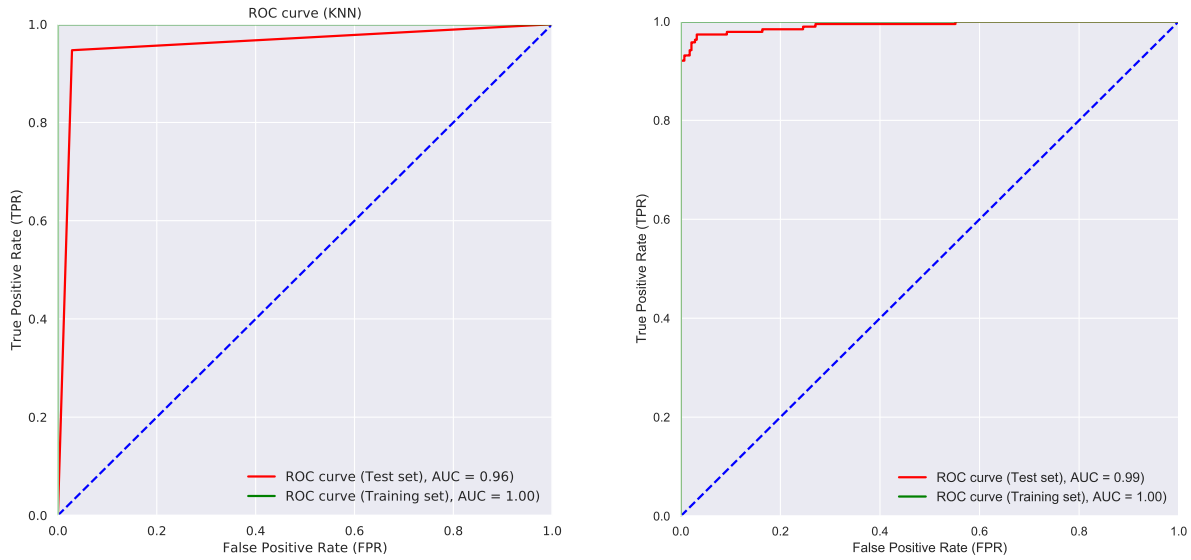
Figure 27 shows the BDT output distributions for the complete data, with and without selection.



(a) (b)

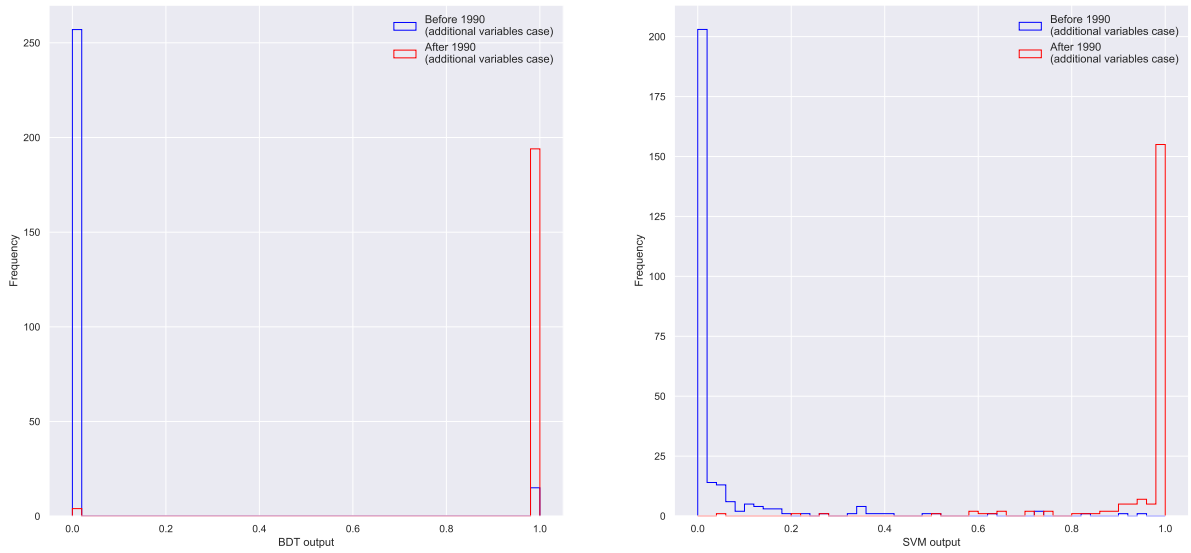
Figure 27: Distributions of the BDT output for the two classes: before and after 1990 (complete data case), for (a) no feature selection and (b) with feature selection.

Finally, two other ML algorithms were tried: a K-Nearest Neighbours (KNN) model and a Support Vector Machine (SVM). Figure 28 and Figure 29 show the ROC curves and outputs for each model. As they are not the focus of the study, hyperparameter optimisation was not performed. For the KNN, the training and test set scores were 0.97 and 0.96, respectively. For the SVM, the training and test set scores were 0.96 and 0.95, respectively.



(a) (b)

Figure 28: The Receiver Operating Characteristic (ROC) curve for (a) KNN and (b) SVM models, trained on the complete data set.



(a) (b)

Figure 29: Distributions of the output for before and after 1990 for (a) KNN and (b) SVM models, trained on the complete data set.

Compared to all of the BDT models, both the KNN and SVM show significant improvement in accuracy and AUC scores. The smaller separation between the training and test set accuracies indicates minimal overfitting.

6 Analysis

The primary goal of this study was to compare the use of traditional statistical methods to ML approaches in searching for evidence of climate change in UK weather data. Additionally, using a classification approach allows for novel ways to gain insight into the physics of the climate and climate change, primarily by analysing the correlation matrix of a data set. As discussed in Section 5.1.2, the effects of climate change are already being felt, so it is expected that both methods would be effective at finding evidence. In this section, we focus on how to quantify and compare the strength of each approach and discuss their possible utility in future research.

The method for quantifying the significance of the determined rate of change over time for each variable was already discussed at length in Section 5.1.1. However, this has not been the case for the ML study. For each of the ML models, the significance of the separation between the classes is found by calculating the significance between the means of the output distribution of each class. This is done because a common significance quantity is needed to compare the two approaches.

The formula for the Z-score (significance) of the separation between the two classes for a given model output is

$$Z = \frac{\bar{B} - \bar{A}}{\sqrt{\sigma_B^2 + \sigma_A^2}}, \tag{8}$$

where B is the before-1990 class, A is the after-1990 class. Consequently, $\bar{B}$ is the mean of the before-1990 distribution, $\bar{A}$ is the mean of the after-1990 distribution and σ_B and σ_A are the standard deviations of the before and after-1990 distributions, respectively. The Z-score indicates the degree of separation between the two distributions in units of standard deviations. A higher significance value means that the two distributions are more significantly separated and hence more easily distinguishable.

Table 8 shows the Z-scores for the traditional analysis results and the classification results.

As shown in Table 8, the traditional analysis significantly outperformed the ML approach for every variable. The ML models theoretically have access to more information than in the traditional analysis as they are trained on data from all features. This is as opposed to the approach used in the traditional analysis, which can only be analysed one feature at a time. Although this means that the ML models should be able to draw on more information and have a more statistically significant form of evidence, the likely explanation is the small data set. Training a model on a data set this small, especially given how the low correlation between the features and the target variables, means that the ML approach is likely not suitable for a data set this size. However, especially for the complete

Table 8: The significance of the calculated rate of change from the traditional analysis and the class separation from the ML model output. The significance was calculated for the BDT trained on complete data with feature selection. The significance was calculated for the KNN and SVM that were trained on complete data, but no feature selection.

Variable	**Z-score (significance) / σ**
Traditional analysis	
t_{max}	19.9
t_{avg}	24.5
t_{min}	28.1
af	-11.1
rain	19.4
sun	13.6
Machine learning	
BDT	2.0
KNN	3.6
SVM	4.2

data, the models showed a high level of accuracy even with the small data set. This indicates that this approach would likely be suitable for data with much a much smaller time resolution.

The feature importance plot from Figure 25 and the correlation matrix from Figure 24 can be used to provide insight into the physics of the climate and climate change. Although a detailed analysis is outside the scope of this report, features of interest will be noted here.

The feature ranking highlights which features are most important for a model to be successful in its classification. This implies that high-ranking features are likely to be the variables which are most affected by climate change. As in the complete data case, this also allows insight into which months and variable pairings are likely to be impacted the most by climate change. Figure 25 shows that the maximum monthly temperature in February has the highest feature importance, implying that warmer winters are likely to be a prominent effect of climate change in the UK. This is confirmed by other statistical analyses of UK weather data that see significant and positive trend parameters for winter months and report reduced significance in spring and summer months like May, June and July [56]. The rainfall and number of air frost days tend to have the lowest importance, indicating that they may be impacted less by climate change.

The correlation matrix, as it is independent of the classification, does not provide information on the physics of climate change but, rather, the physics of the climate itself. The correlation matrix shows that the air frost days and $t_{\max}$ for a given month tend to have a strong correlation, especially for the winter months. The rainfall in August also showed a higher than the typical correlation for that variable type with the $t_{\max}$ for August (50% negative correlation). This is not necessarily obvious, as warmer air can hold more water, leading to an increase in rainfall on average across the world. Some areas receive less rain due to changes in wind patterns, so this correlation may hint at a coupling between rainfall, $t_{\max}$, and changing wind patterns [1].

7 Conclusions

This study used traditional statistical analyses and a machine learning (ML) classification approach on historic Met Office weather station data to search for evidence of climate change. The significance of the year 2022 was evaluated using a trend-over-time approach. The nominal data set provided by the Met Office was preprocessed and transformed to optimise the performance of the classifiers for the ML approach. The study demonstrated the successful use of ML models to categorise UK weather data into before and after 1990, which serves as an inflexion point for climate change.

To quantify the significance of the trend-over-time and ML analyses, Z-scores (at the 5% confidence level) were calculated for the slope of the fit to the weather data and the separation between the mean

of each model output distribution, respectively. The trend-over-time analysis provided evidence of climate change with a higher level of significance than the ML approach. Rates of change calculated using trend-over-time analysis had a significance greater than 10σ.

In contrast, the significance of the results of the ML study ranged between 2σ and 4.2σ. This vast difference in the reliability of the models is likely due to the size of the data set, as ML models generally do not perform well on small data sets [57]. Despite this, the ML study still showed some success, indicating that it may be a promising approach when applied to larger data sets.

Additionally, this study aimed to assess the use of a ML approach in gaining insight into the physics of the climate and climate change. We showed that an analysis of the correlation matrix and F-test ranking of the weather variables could provide information on the potential effects of climate change and the coupling between weather variables. For example, an analysis of the feature ranking delivered evidence that the winter months in the UK may see a more significant increase in temperature in the summer months. Furthermore, the correlation matrix for the weather variables implied that changing wind and rainfall patterns may be linked to the maximum daily temperature.

Ultimately, the results of this study provide a test of novel methods for searching for evidence of climate change. It contributes to a growing body of evidence that our climate has changed significantly, and continues to do so, due to human activities.

References

[1] Met Office, "Effects of climate change." https://www.metoffice.gov.uk/weather/climate-change/effects-of-climate-change#:~:text=Warmer%20air%20can%20hold%20more,becoming%20more%20intense%20as%20well. (Accessed: 9 March 2023).

[2] United Nations, "What Is Climate Change?." https://www.un.org/en/climatechange/what-is-climate-change. (Accessed: 9 March 2023).

[3] F. W. Taylor, *Elementary climate physics*. Oxford: Oxford University Press, 2005.

[4] G. R. North, "Multiple solutions in energy balance climate models," *Global and Planetary Change*, vol. 2, no. 3, pp. 225–235, 1990.

[5] G. L. Russell, J. R. Miller, and D. Rind, "A coupled atmosphere-ocean model for transient climate change studies," *Atmosphere-Ocean*, vol. 33, no. 4, pp. 683–730, 1995.

[6] G. M. Flato, "Earth system models: an overview," *WIREs Climate Change*, vol. 2, no. 6, pp. 783–800, 2011.

[7] *Energy Balance Models*, ch. 3, pp. 81–116. John Wiley Sons, Ltd, 2005.

[8] D. E. Reichle, "Chapter 3 - energy relationships between organisms and their environment," in *The Global Carbon Cycle and Climate Change* (D. E. Reichle, ed.), pp. 15–41, Elsevier, 2020.

[9] S. Cohen and G. Stanhill, "Chapter 32 - changes in the sun's radiation: the role of widespread surface solar radiation trends in climate change: dimming and brightening," in *Climate Change (Third Edition)* (T. M. Letcher, ed.), pp. 687–709, Elsevier, third edition ed., 2021.

[10] NOAA National Centers for Environmental Information, "Monthly Global Climate Report for Annual 2022." https://www.ncei.noaa.gov/access/monitoring/monthly-report/global/202213, Jan 2023. (Accessed: 15 February 2023).

[11] S. B. Idso, "The co2 greenhouse effect on mars, earth, and venus," *Science of The Total Environment*, vol. 77, no. 2, pp. 291–294, 1988.

[12] W. R. Cline, "Scientific Basis for the Greenhouse Effect," *The Economic Journal*, vol. 101, pp. 904–919, 07 1991.

[13] D. Archer, *Greenhouse gases*, pp. 63–86. John Wiley Sons, 2nd ed., 2011.

[14] LibreTexts, "Dipole moments - chemistry libretexts." https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Physical_Properties_of_Matter/Atomic_and_Molecular_Properties/Dipole_Moments, nov 2020. (Accessed: 3 March 2023).

[15] S. Seman, "Lesson 2: The earth's energy budget." https://www.e-education.psu.edu/meteo3/l2_p7.html. Accessed on: 3 March 2023.

[16] Encyclopedia Britannica, "Greenhouse Effect." https://www.britannica.com/science/greenhouse-effect, Jan 2023. (Accessed: 15 February 2023).

[17] J. Houghton, Y. Ding, D. Griggs, M. Noguer, P. van der Linden, X. Dai, M. Maskell, and C. Johnson, *Climate Change 2001: The Scientific Basis*, vol. 881., p. 881. 01 2001.

[18] Our World in Data, "Global warming potentials of greenhouse gases over 100-year timescale." https://ourworldindata.org/grapher/global-warming-potential-of-greenhouse-gases-over-100-year-timescale-gwp. (Accessed: 15 February 2023).

[19] I. Adopted, "Climate change 2014 synthesis report," *IPCC: Geneva, Szwitzerland*, 2014.

[20] NOAA, "Carbon dioxide now more than 50% higher than pre-industrial levels." https://www.noaa.gov/news-release/carbon-dioxide-now-more-than-50-higher-than-pre-industrial-levels#:~:text=Carbon%20dioxide%20now%20more%20than,National%20Oceanic%20and%20Atmospheric%20Administration. (Accessed: 5 March 2023).

[21] K. D. Bennett, "Milankovitch cycles and their effects on species in ecological and evolutionary time," *Paleobiology*, vol. 16, no. 1, p. 11–21, 1990.

[22] Scripps Institution of Oceanography, UC San Diego, "The Keeling Curve." https://keelingcurve.ucsd.edu/. (Accessed: 5 March 2023).

[23] NASA, "The Study of Earth as an Integrated System." https://climate.nasa.gov/nasa_science/science/#:~:

text=Climate%20feedbacks%3A%20processes%20that%20can,Clouds. (Accessed: 6 March 2023).

[24] Z. Hausfather and R. Betts, "Analysis: How 'carbon-cycle feedbacks' could make global warming worse." https://www.carbonbrief.org/analysis-how-carbon-cycle-feedbacks-could-make-global-warming-worse/, Apr 2020. (Accessed: 6 March 2023).

[25] Environmental Agency, "Climate change impacts and adaptation." https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/758983/Climate_change_impacts_and_adaptation.pdf, Nov 2018. (Accessed: 19 February 2023).

[26] J. R. MAWDSLEY, R. O'MALLEY, and D. S. OJIMA, "A review of climate-change adaptation strategies for wildlife management and biodiversity conservation," *Conservation Biology*, vol. 23, no. 5, pp. 1080–1089, 2009.

[27] A. El-Sayed and M. Kamel, "Future threat from the past," *Environmental Science and Pollution Research*, vol. 28, pp. 1287–1291, Jan 2021.

[28] W. Colgan, H. Machguth, M. MacFerrin, J. D. Colgan, D. van As, and J. A. MacGregor, "The abandoned ice sheet base at camp century, greenland, in a warming climate," *Geophysical Research Letters*, vol. 43, no. 15, pp. 8091–8096, 2016.

[29] E. Tompkins, "3rd uk climate change risk assessment (ccra3) evidence report," project report, January 2022.

[30] S. Chowdhury and M. P. Schoen, "Research paper classification using supervised machine learning techniques," in *2020 Intermountain Engineering, Technology and Computing (IETC)*, pp. 1–6, 2020.

[31] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, p. 160, Mar 2021.

[32] B. Penkovsky, "Model Uncertainty Estimation." https://penkovsky.com/neural-networks/day8/, Apr 2022. (Accessed: 4 March 2023).

[33] Javatpoint, "Decision Tree Classification Algorithm." https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm. (Accessed: 4 March 2023).

[34] Met Office, "Historic station data." https://www.metoffice.gov.uk/research/climate/maps-and-data/historic-station-data. (Accessed: 27 January 2023).

[35] Met Office, "Weather stations." https://www.metoffice.gov.uk/weather/learn-about/how-forecasts-are-made/observations/weather-stations. (Accessed: 3 March 2023).

[36] Met Office, "What is frost?." https://www.metoffice.gov.uk/weather/learn-about/weather/types-of-weather/frost-and-ice/frost#:~:text=be%20aware%20of.-,Air%20frost,one%20metre%20above%20the%20ground. (Accessed: 27 January 2023).

[37] Met Office, "How we measure temperature." https://www.metoffice.gov.uk/weather/guides/observations/how-we-measure-temperature. (Accessed: 28 February 2023).

[38] NPL, "What is a platinum resistance thermometer?." https://www.npl.co.uk/resources/q-a/platinum-resistance-thermometer. (Accessed: 28 February 2023).

[39] NIST, "Mercury Thermometer Alternatives: Platinum Resistance Thermometers (PRTs)." https://www.nist.gov/pml/mercury-thermometer-alternatives/mercury-thermometer-alternatives-platinum-resistance, Mar 2022.

[40] Met Office, "How we measure rainfall." https://www.metoffice.gov.uk/weather/guides/observations/how-we-measure-rainfall#:~:text=For%20many%20years%20the%20Met,0.2%20mm%20has%20been%20detected. (Accessed: 9 March 2023).

[41] A. Kerr and R. Tabony, "Comparison of sunshine recorded by campbell–stokes and automatic sensors," *Weather*, vol. 59, no. 4, pp. 90–95, 2004.

[42] P. HE, "The performance of a campbell-stokes sunshine recorder compared with a simultaneous record of the normal incidence irradiance," *METEOROL. MAG.; ISSN 0026-1149; GBR; DA. 1981; VOL. 110; NO 1305; PP. 102-109; BIBL. 2 REF.*, 1981.

[43] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence*, vol. 97, no. 1, pp. 245–271, 1997. Relevance.

[44] J. Huang, Y.-F. Li, and M. Xie, "An empirical analysis of data preprocessing for machine learning-based software cost estimation," *Information and Software Technology*, vol. 67, pp. 108–127, 2015.

[45] Met Office, "How do we measure the weather?." https://www.metoffice.gov.uk/weather/learn-about/met-office-for-schools/other-content/other-resources/how-to-measure-the-weather#:~:text=We%20used%20to%20use%20thermometers,our%20thermometer%20a%20Stevenson%20screens. (Accessed: 18 February 2023).

[46] F Mueller, B Martin, J Chandler, R Martin, "Electronic, digital thermometer patent." https://patents.google.com/patent/US3729998A/en. (Accessed: 18 February 2023).

[47] Met Office, "UK regional climates." https://www.metoffice.gov.uk/research/climate/maps-and-data/regional-climates/index#:~:text=In%20general%2C%20places%20in%20the,than%20in%20autumn%20and%20winter. (Accessed: 27 February 2023).

[48] D. R. Easterling, B. Horton, P. D. Jones, T. C. Peterson, T. R. Karl, D. E. Parker, M. Salinger, V. Razuvayev, N. Plummer, P. Jamason, and C. Folland, "Maximum and minimum temperature trends for the globe," *Science*, vol. 277, no. 5324, p. 364–367, 1997.

[49] M. Allen, O. Dube, W. Solecki, F. Aragón-Durand, W. Cramer, S. Humphreys, M. Kainuma, J. Kala, N. Mahowald, Y. Mulugetta, R. Perez, M. Wairiu, and K. Zickfeld, *Framing and Context*, pp. 49–92. Cambridge, UK and New York, NY, USA: Cambridge University Press, 2018.

[50] J. Hansen, R. Ruedy, M. Sato, and K. Lo, "Global surface temperature change," *Reviews of Geophysics*, vol. 48, no. 4, 2010.

[51] G. Madge, "2023 set to be tenth consecutive year at 1°C or above." https://www.metoffice.gov.uk/about-us/press-office/news/weather-and-climate/2022/2023-global-temperature-forecast, Dec 2022. (Accessed: 6 March 2023).

[52] N. Christidis, M. McCarthy, and P. A. Stott, "The increasing likelihood of temperatures above 30 to 40 °c in the united kingdom," *Nature Communications*, vol. 11, p. 3093, Jun 2020.

[53] M. McCarthy, N. Christidis, and P. Stott, "Met Office: A review of the UK's climate in 2022," *Carbon Brief*, Jan 2023.

[54] B. Venkatesh and J. Anuradha, "A review of feature selection and its methods," *Cybernetics and Information Technologies*, vol. 19, no. 1, pp. 3–26, 2019.

[55] C. X. Ling, J. Huang, and H. Zhang, "Auc: A better measure than accuracy in comparing learning algorithms," in *Advances in Artificial Intelligence* (Y. Xiang and B. Chaib-draa, eds.), (Berlin, Heidelberg), pp. 329–341, Springer Berlin Heidelberg, 2003.

[56] T. J. Vogelsang and P. H. Franses, "Are winters getting warmer?," *Environmental Modelling Software*, vol. 20, no. 11, pp. 1449–1455, 2005.

[57] A. Althnian, D. AlSaeed, H. Al-Baity, A. Samha, A. B. Dris, N. Alzakari, A. Abou Elwafa, and H. Kurdi, "Impact of dataset size on classification performance: An empirical evaluation in the medical domain," *Applied Sciences*, vol. 11, no. 2, 2021.

Appendix A

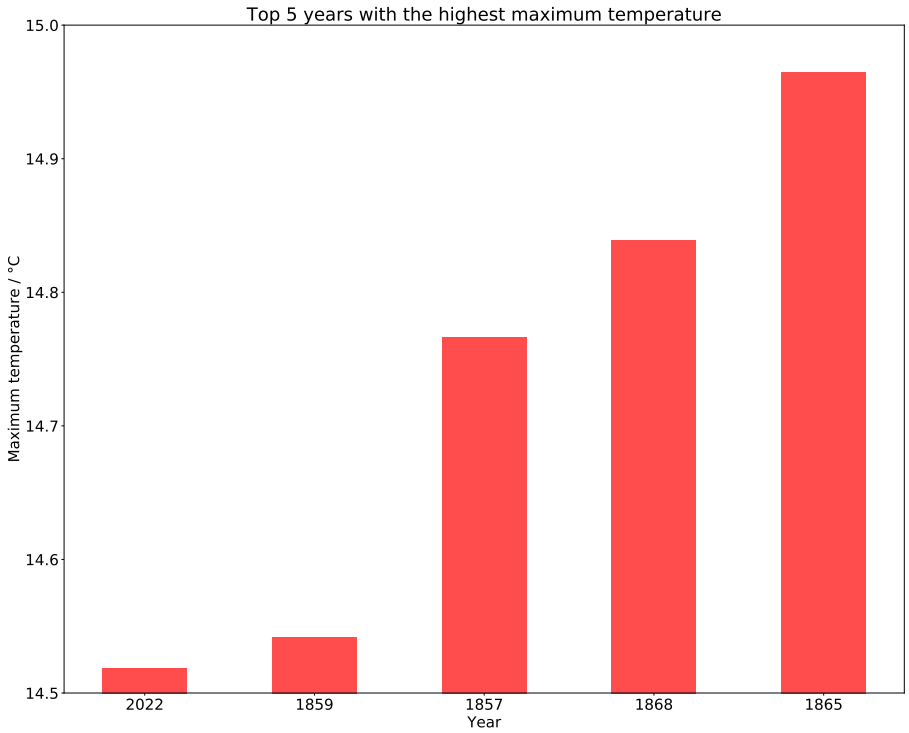


Figure 30: A bar chart of the top five years with the highest $t_{\max}$ in the UK.

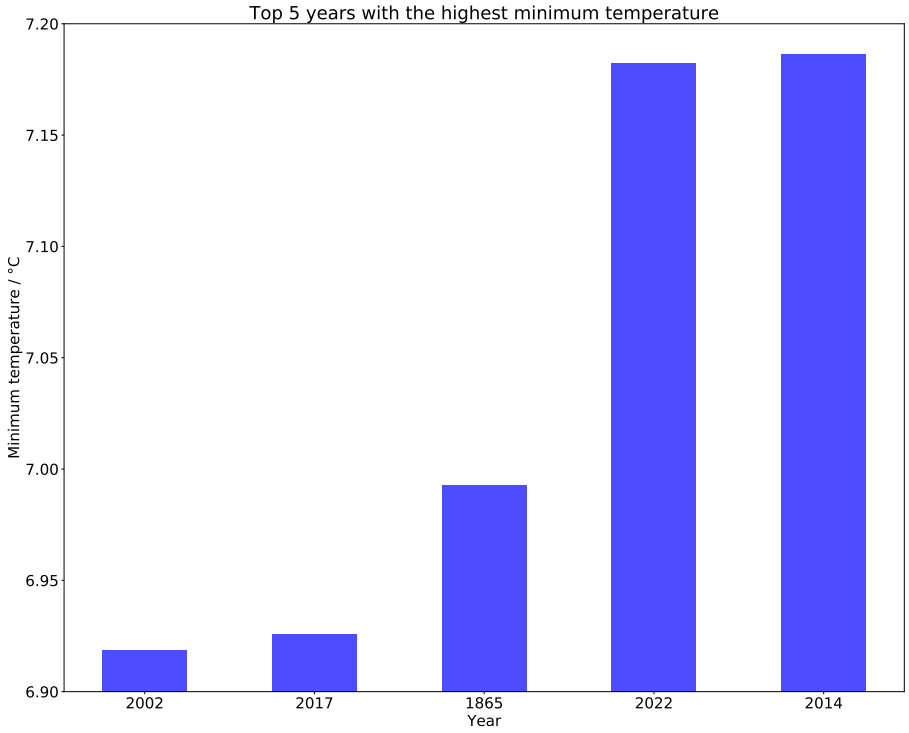


Figure 31: A bar chart of the top five years with the highest $t_{\min}$ in the UK.

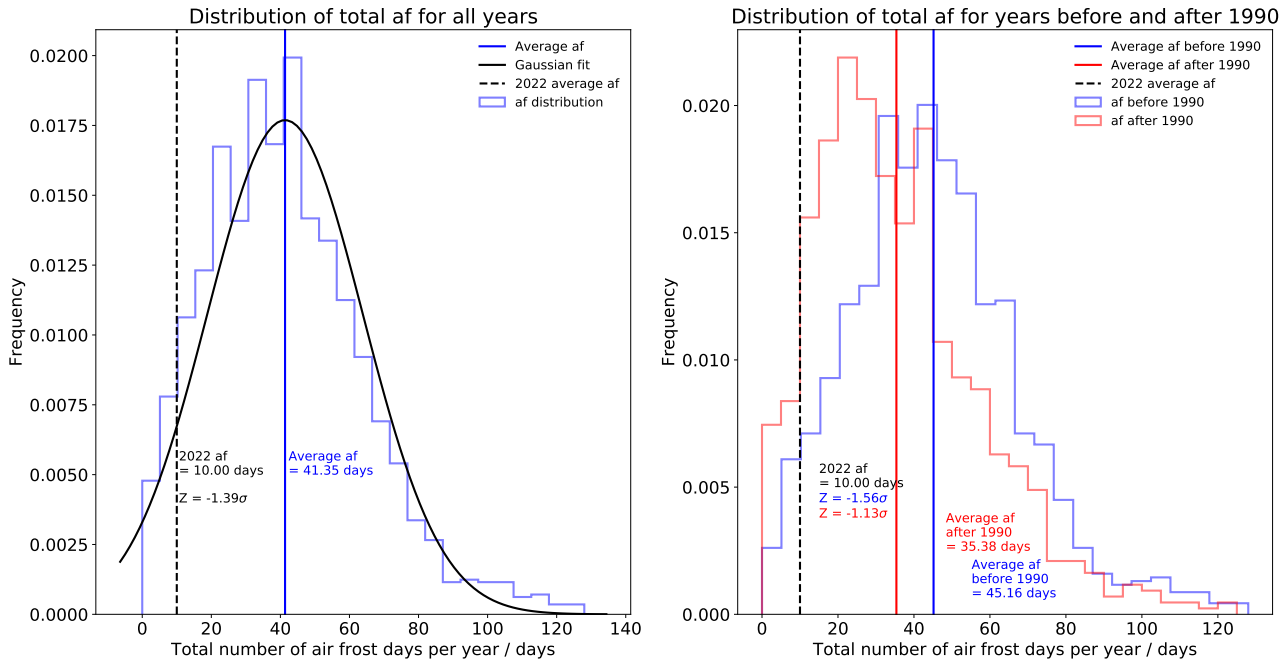


Figure 32: Histograms of the total yearly average number of air frost days for each weather station. The left panel shows the distribution for all years with a Gaussian fit. The average number of air frost days is 41. The *average* 2022 total number of air frost days is 10, and its significance with respect to the mean is $Z = -1.39\sigma$. The right panel shows distributions of the total yearly average number of air frost days before and after 1990. The average number of air frost days before 1990 is 45, and 35 after. The significance of the *average* 2022 total number of air frost days with respect to before 1990 is $Z = -1.56\sigma$ and $Z = -1.13\sigma$ after.

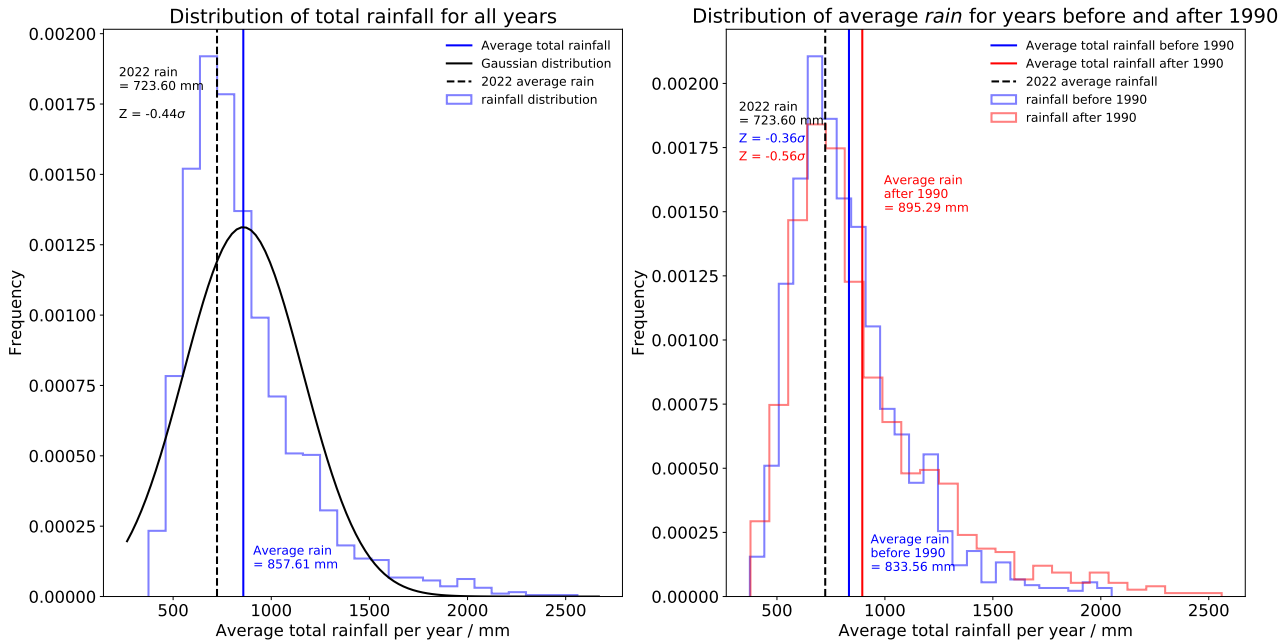


Figure 33: Histograms of the total yearly average rainfall for each weather station. The left panel shows the distribution for all years with a Gaussian fit. The average rainfall is 860 mm. The *average* 2022 total rainfall is 720 mm, and its significance with respect to the mean is $Z = -0.44\sigma$. The right panel shows distributions of the total yearly average rainfall before and after 1990. The average rainfall before 1990 is 830 mm, and 900 mm after. The significance of the *average* 2022 total rainfall with respect to before 1990 is $Z = -0.36\sigma$ and $Z = -0.56\sigma$ after.

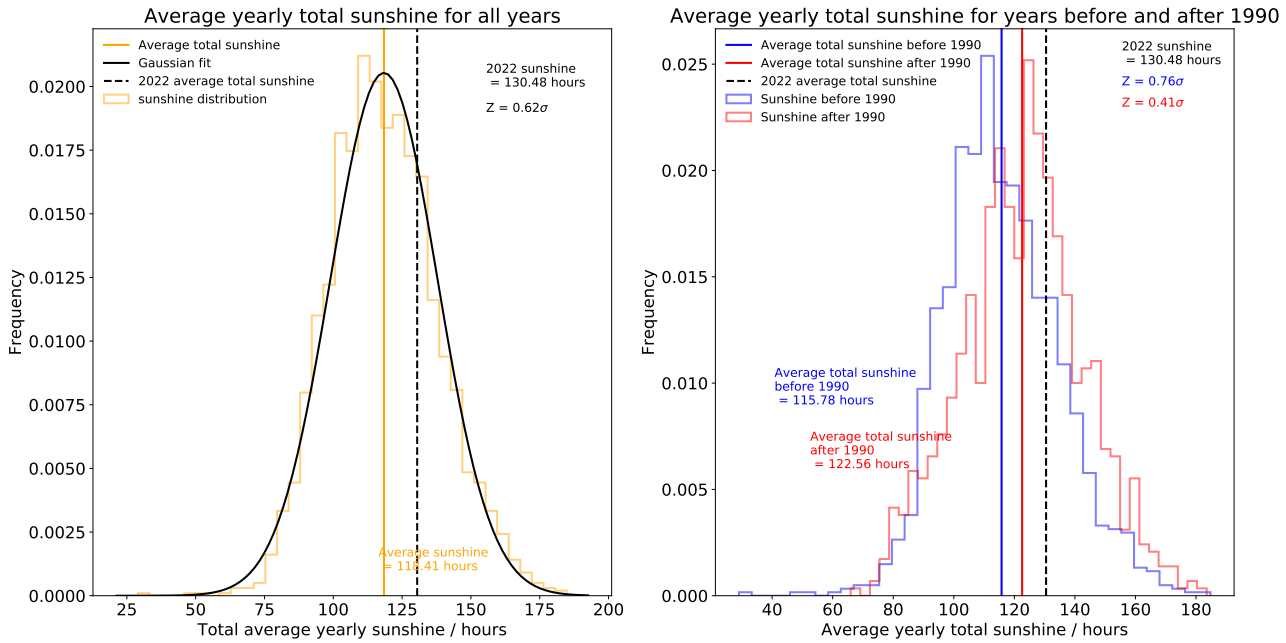


Figure 34: Histograms of the total yearly average sunshine for each weather station. The left panel shows the distribution for all years with a Gaussian fit. The average sunshine is 120 hours. The *average* 2022 total sunshine is 130 hours, and its significance with respect to the mean is $Z = 0.62\sigma$. The right panel shows distributions of the total yearly average sunshine before and after 1990. The average sunshine before 1990 is 116 hours, and 120 hours after. The significance of the *average* 2022 total sunshine with respect to before 1990 is $Z = 0.76\sigma$ and $Z = 0.41\sigma$ after.