**CHAPTER 1**

**INTRODUCTION**

* 1. **Introduction to Climate Change**

Climate change is change in the climate observed in any region over the course of history, be it a small local region or a region as large as a continent or a planet. It is usually noticed as a change in temperature and precipitation. Thus, studies focusing on climate change usually analyse changes in climate variables for the region of interest over a certain period of time. Climate variables are not only important factors in determining the extent of change that has occurred to the climate in a region, but in planning for management of natural calamities such as floods, droughts and extreme heat/cold as well.

The IPCC calls evidences for climate change “unequivocal”. Ice sheets in Greenland and Antarctica have been found to be shrinking, with an average ice loss of 279 billion tons and 148 billion tons respectively, per year between 1993 and 2019. Glaciers in the Alps, Himalayas, Andes, Rockies, Alaska and Africa have also witnessed retreats. Sea levels have also risen about 20 centimetres in the last century, with the rate in the last two decades being nearly double that of the last century. Ocean waters have also seen an increased acidification of around 30%, due to the uptick in use of chemicals and fossil fuels, brought forth by the industrial revolution. Oceans are also responsible for storing 90% of the earth’s extra energy, and a consequence of climate change has been that the top 100 metres of oceans are showing a warming of more than 0.33°C since 1969. Increases in extreme weather events globally, especially intense rainfall events, are also a testament for climate change.

The IPCC also notes some drastic effects that climate change is expected to have in the future, such as the lengthening of frost-free seasons which are important for crops and plants to grow, which can in turn have adverse effects on entire ecosystems. Precipitation patterns are also expected to see a drastic change globally, and droughts and heat waves are projected to occur more frequently, from once-in-20-year extreme heat days, to every two or three years. Sea levels are also expected to rise by an even higher margin, increasing from the currently witnessed increase of 20 centimetres, to an increase of 30-240 centimetres. The Arctic Ocean is also expected to become ice-free in summer before 2050.

The United Nations Organisation calls climate change a “defining issue of our time” and the current times being a “defining moment” (United Nations Organization, 2021) i.e., immediate action needs to be taken to analyse, prevent and tackle climate change in as many regions as possible. While previous UN IPCC (Intergovernmental Panel for Climate Change) reports (Climate Change - IPCC 2007) focused on damages that may be caused due to a 2°C increase in the global temperatures, a recent report from 2018 switched that focus to an increase of 1.5°C, showing that many of the adverse impacts that were thought to be occurring at an increase of 2°C would occur at only 1.5°C (IPCC Special Report, 2015). Another IPCC report, from 2021, has called the current climate scenario witnessed globally as a “code red for humanity” (United Nations Organization, 2021), stressing the need for attention that needs to be put towards climate change ever more than before.

* 1. **Objectives**

**1.2.1 Analysis of Past Climate Trends**

One of the objectives of the study is to analyse the climate trends over the past years observed in the valley of Kashmir, for parameters such as mean temperature, precipitation and cloud cover, to visualise and understand the changes that have occurred.

**1.2.2 Providing Ready-to-Use Datasets**

The climate datasets pertaining to the region that are readily available for direct use only provide data for a few decades, while as the CRU dataset provides data for over a century. The problem with the CRU dataset is that it can be daunting to use without pre-processing it ideally. Pre-processed, ready-to-use datasets have thus been made available in the public domain for research purposes.

**1.2.3 Forecasting Climate Variables**

The major aim of the study is to forecast different climate variables such as mean temperature, average minimum temperature, average maximum temperature, cloud cover and precipitation, till the end of the current century, using the ARIMA model, to allow for better understanding of the projected changes and thus, in turn, structuring of better strategies to tackle climate change in the region.

* 1. **Motivation**

The valley of Kashmir is located 1600 metres above sea-level in the North-Western corner of the Himalayas, with an area ranging from (33.25°N,73.75°E) to (34.5°N,75.25°E), with some parts controlled by India and some by Pakistan. The region is considered to have a sub-tropical climate, which is sometimes also classified as sub-Mediterranean due to the rainfall distribution pattern (Meher-Homji, 1971). Mild summers and rigorous, severe winters are considered characteristic features of the climate in the region. The area holds great environmental, geographical and economical importance, with multiple rivers such as Jhelum, Bringhi, Lidder etc., originating from the glaciers located in the upper reaches of the valley, and then flowing through the valley, as well as towards other regions, such as Pakistan and Punjab, acting as a source of irrigation. The high-altitude origins of the rivers allow for generation of hydroelectricity, having an estimated potential of a whopping 20,000 MW (Ahmed U., 2021), and an estimated annual export revenue of $13 billion, although only about 20% of the total potential has been realised till date (Smith G., 2010; Ahmed M., 2019). Any drastic changes in climate can be disastrous to the hydroelectric potential of the region, since it is directly dependant on the glaciers in the region. One industry that has borne the brunt of climate change over the past few decades is the saffron industry, a sector with which the region has been involved for more than 2500 years, saffron has seen a steep decline in production, with the harvest in 2018 being only half of that in 1998 (Bukhari, 2020). Multiple studies and reports have been published regarding the scenario of climate change in the region with some key findings; such as the 27-38% decrease in glacier sizes (Romshoo, Fayaz, et al. 2020), 19.44% decrease in the daily rainfall (Wani, Baba, Seha, Bazaz, 2015), a predicted 6.93℃ increase in temperature by the year 2100 (Romshoo, Bashir, Rashid, 2020).

However, there are various shortcomings in the studies such as entirely leaving out major parts of Kashmir such as north Kashmir from the studies (Zaz, Rumshoo, et al., 2019), using small amounts of data (Zahoor ul, Khan, 2013), focusing only on individual variables without looking for correlations between them, and only using classical methods which aren’t able to lend as good results as possible using machine learning & deep learning methods such as ARIMA. Recent events that have occurred such as the unusually hot and humid summers of 2021 and 2020 with studies showing that there may be a further rise of 6.9°C in the temperature (Romshoo, Bashir, Rashid, 2020), the devastating floods of 2014 (Romshoo, 2015) and severe random upticks in rainfall for short periods of time that are unpredictable according to Sonam Lotus, the MET director of J&K, all indicate a change in the overall climate of the region (Das, 2021). This change requires to be analysed, studied and forecasts need to be made for the same, to understand the extent of climate change, the ways in which further deterioration can be prevented and strategies to minimise its effects can be developed.

* 1. **Report Outline**

The report starts by discussing previous studies and researches pertaining to the subject matter in Chapter 2 – Related Works, it also explores avenues of improvement in the previous studies.

In Chapter 3 – Datasets and Pre-Processing, the dataset used is discussed, as well as steps involved in its sourcing and the pre-processing performed on it before using it.

The machine learning models and steps followed in using them are then discussed in Chapter 4 – ML Models and Methodology. It focuses on the ARIMA model, its usage and the fine-tuning involved.

The results and findings from the analysis phase and forecasting phase are discussed in Chapter 5 – Results, with proper visualisations and tables and key findings with regards to the climate trends in the region.

Finally, in Chapter 6 – Conclusion and Future Scope, the findings from the results and their real-world impacts are discussed briefly, and some improvements that can potentially be made for the studies are also explored.

**CHAPTER 2**

**RELATED WORK**

**2.1 Review of Literature**

Some studies focusing on the region of Kashmir have been conducted, with varying findings and results. Liner regression analysis was used to examine the rate of change of climatic indices using data from 1961-2005 and an overall increasing trend was found in the seasonal and annual average temperatures. (Islam, Khan, Rao, 2013). Another study, using observational data from six stations within the Kashmir valley, used WRF (Weather and Research Forecasting) model and ERA-Interim data. The main findings from the study were that the higher altitude stations exhibited a steep increase of 1.04°C to 1.13°C in the annual mean temperatures. (Zaz, Romshoo, Krishnamoorthy, Viswanadhapalli, 2019). Investigation of future climate change trends for the 21st century was done under 3 emission schemes (AIB, RCP4.5, RCP8.5) with the baseline period of the data being 1961-1990. The study used GFDL CM2.1 model and the conclusion made from the results was that the annual mean temperature was projected to increase by 4.5°C, 3.98°C and 6.93°C respectively under the 3 emission schemes. The study also found the different climatic zones would experience significant changes. (Romshoo, Bashir, Rashid, 2020). Changes in the glacier sizes have also been studied, by comparing satellite imagery from 1980 to that in 2018, showing a decrease of 27-38% in the sizes of different glaciers, suggesting the reason to be the increasing temperatures and decline in winter solid precipitation in the region, which if continued in the future, would adversely affect the economy in the region. (Romshoo, Fayaz, Meraj, Bahuguna, 2020).

Precipitation and temperature changes in India’s Bhagirathi River basin have been studied and forecasted using ARIMA, with the results showing an increasing trend for temperature in one station and a decreasing trend for another, while as the precipitation is found to be over-predicted in case of extreme rainfall events. (Dimri, Ahmad, Sharif, 2020). Air pollution modelling and forecasting has also been done using ARIMA model, with findings showing that meteorological variables have definite influence on the life cycle persistence of air pollutants. (Naseem, Rashid, Izhar, 2017). Another study has used ARIMA model for analysing of trends and modelling of pre-monsoon rainfall data using ARIMA model, with results indicating a significant rise in the pre-monsoon rainfall over the northwest part of the country (Narayanan, Basistha et al., 2013). Climate change has also been assessed and monthly rainfall forecasted over Khordha district in Odisha, India, with outstanding accuracy using ARIMA model (Swain, Patel, Nandi, 2018). In another paper, GDP forecasts have been made using ARIMA model for Portugal and Germany until the year 2031, showing a steady growth in both the countries. (Lhano, Francisco et al., 2021). The Nigerian economy has also been studied using the ARIMA model, with a study showing that the living standards in Nigeria would most likely worsen over the next decade unless the economic policy stance is not reviewed. (Nyoni, Thabhani, 2019).

More recently, ARIMA model has been put to use in various studies analysing and forecasting the COVID-19 pandemic, in terms of the spread and infection rates in different regions, and the effects thereof. One study focused on the top five affected countries, with the resulting forecasts being found to be within acceptable agreement with the observed data, and predictions of exponential curves in certain countries such as India, US and Brazil turning out to be true. (Sahai, Kumar et al., 2020). COVID-19 cases in India were also forecast in another study, using data from the Ministry of Health and Family Welfare (MoHFW), with results showing an increasing trend in actual and forecasted numbers of COVID-19 cases with approximately 1500 cases per day.

**2.2 Research Gap**

Although some studies have focused on the region of Kashmir, there are various shortcomings in each of them; none of the studies have taken machine learning approaches to forecasting the future climate trends and values for different variables such as temperature and precipitation, even though machine learning approaches can provide for better and more stable results than traditional mathematical models (Rolnick, Donti, Kaack, et al., 2019; Jebeile, J., Lam, V. & Räz, T., 2020). Another avenue of improvement is the data used by the studies, since almost all of them have taken only small datasets with records only available for a few decades (1981-2016) (1961-1990) (Khan, Islam, 2017). Further, some studies have restricted their region of study to a smaller region within the valley, as is evident from the coordinates mentioned in the studies, excluding statistically and environmentally important regions (Zaz, Romshoo et al., 2019), which can cast doubts on their findings and conclusions.

**CHAPTER 3**

**DATASETS**

**AND**

**PRE-PROCESSING**

**3.1 Dataset Description**

The study uses a timeseries data sourced from the Climate Research Unit (CRU) TS4.04 dataset (CRU TS4.04, 2020[[1]](#footnote-1)), which is in the public domain of the Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia in Norwich, UK. The main motive of the division is to provide data to the general public for research purposes. The data is an interpolated high resolution gridded dataset (0.5 x 0.5 grids) consisting of global monthly mean temperature, precipitation, cloud cover, potential evapotranspiration, etc. The data ranges from 1901 to 2019. The ‘Subsetter’ tool from the web processing service (WPS) available on the CEDA web archive portal (The Subsetter Tool, Web Processing Service, 2021[[2]](#footnote-2)) was used to extract the monthly precipitation, average cloud cover and mean temperature data for our region of study. While other datasets were available as well, such as one from the Indian Meteorological Department (IMD) Pune available on the Open Governmental Data Platform (Open Government Data Portal) and one from Berkeley Earth (Berkeley Earth Climate Data), these datasets cannot be refined to only include data from specific coordinates, have significantly lower number of entries than the CRU dataset and in some cases, the datasets provided being derivatives of the CRU timeseries itself (Background on the Indian Water Portal). The CRU data is also well documented and researched upon, leaving little to no doubt about its credibility and accuracy (Harris, I., Osborn, T.J., Jones, P. et al., 2020; H. Shi, T. Li, J. Wei, 2017; Salvacion, A.R., Magcale-Macandog, D.B., Cruz, P.C.S. et al., 2018; Dimri, Ahmad, Sharif, 2020). Another reason for the selection of the CRU timeseries in this study is previous studies have used sparsely populated datasets, with data only available for the past 3-4 decades, while as the CRU timeseries is populated with entries for well over a century, making this the first study in the region to be using such a large and extensive dataset, making for potentially better understanding of the past trends as well as more accurate forecasting.

|  |
| --- |
|  |
| **Fig 1. Data Sourcing Process** |

|  |
| --- |
|  |
| **Fig 2. Area of Study** |

**3.2 Pre-Processing**

The CRU data is provided in a gridded form, and requires some pre-processing before being used for analysis and forecasting future trends from it. The dataset comes included with metadata, contained in 44 header lines, which have no practical use and are thus removed completely.

**3.2.1 Conversion of Gridded Data to Usable Data**

The data for each climate parameter is provided in the form of 0.5 x 0.5 grids, and needs to be converted to numeric values corresponding to each month, which is done by calculating the mean for cloud cover (%) and for all three of the temperature (°C) parameters (average, maximum and minimum). In a similar manner, to convert the gridded values to numeric values, the sum is taken for precipitation (mm/month), instead of its mean, since the total precipitation for the region is the required parameter, not the mean precipitation by area.

**3.2.2 Detection and Removal of Outliers**

Detection and removal of outliers allows machine learning models to better understand the underlying trends in the data and to more accurately model them. The ‘z-score’ of the distribution was calculated for this purpose, and instances with values not pertaining to the usual z-score range for Gaussian distributions i.e., to , were replaced by the mean of their immediately neighbouring values.

‘’ denoting the mean of the distribution and ‘’, the standard deviation.

**CHAPTER 4**

**ML MODELS**

**AND**

**METHODOLOGY**

**4.1 Model Description**

Auto-Regressive Integrated Moving Average (ARIMA) model (P. Whittle, 1951) is used for time series analysis to better understand the underlying trends in the data and to make forecasts for the future (Ding, Guorong et al., 2020; Sahai, Kumar et al., 2020; Khan, Mohammad, Gupta, 2020; Nyoni, Thabhani, 2019; Lhano, Francisco et al., 2021; Youssef, Noor, 2021) ARIMA tries to re-create newer values of the data based on the mean, deviation and differences of the past values. The usage of ARIMA model basically involves three main hyperparameters, p – which denotes the order of the AR (auto-regressive) term i.e., the number of prior values the current value in a timeseries is regressed upon, d – which denotes the numbers of differencing required for the timeseries to make it stationary and q – which denotes the order of the MA (moving average) term i.e., the number of error values occurring at various time intervals in the past that the regression error is a linear combination of. These AR, MA and I values need to be selected in such a way that the model fits the data in the best way possible. If the timeseries to be used is non-stationary, its trend is removed by differencing to obtain a stationary timeseries. Non-seasonal ARIMA models are denoted as ARIMA (p, d, q) and seasonal ARIMA models as ARIMA (p, d, q) (P, D, Q) m where “m” denotes the number of periods in each season. ARIMA is considered to be one of the best models to use, when dealing with long and stable timeseries data, especially for approximating historical patterns for the future (Jebeile, J., Lam, V. & Räz, T., 2020; Rolnick, Donti, Kaack, et al., 2019; Doost, Sadeghian, Farahani, Rasekhi, 2017).

An ARIMA model can be mathematically described as:

The equation follows the Box-Jenkins convention wherein the MA parameters (θ) are defined so that their signs are negative. Identifying the appropriate ARIMA model for some data begins by determining the order of differencing (d) required to make the timeseries stationary and remove any seasonality. Stationary series can still have autocorrelated errors, which is suggestive of the need for some number of AR terms (p≥1) and possibly some MA terms (q≥1) in the forecasting equation.

Separate ARIMA models have been fitted to mean temperature, maximum temperature and minimum temperature and cloud cover. Precipitation achieved a relatively higher RMSE value, due to the inability of the ARIMA model to adapt to extreme changes in the data, with values ranging from 60mm/month, all the way up to around 19000 mm/month.

**4.2 Methodology**

The data considered for the study is for the years 1901-2019, for each of the following climate variables; temperature (mean, maximum and minimum), precipitation and cloud cover. The data are provided in gridded form and are prepared as monthly mean values, except for precipitation, which is taken as the sum total of precipitation observed in the entire region. Upon preparation of the data files, the data is checked to ensure stationarity, after which the models that produce the best results in terms of forecast need to be identified. The identification process mainly focuses on determining the hyperparameter values for the ARIMA model, followed by testing the models using the evaluation metrics, as discussed below in section 4.2.1, and shown in fig 3.

**4.2.1 Ensuring Stationarity**

The first step while dealing with any timeseries data is to determine whether it is stationary or not, which can be done using multiple tests. In this study, we used one such test known as the Augmented Dickey-Fuller Test (ADF). The p-value from the test is used to test the null hypothesis and any values higher than the significant threshold of 0.05 (5%) deem the rejection of the null hypothesis false. The conclusion from the results thus reached was that only the cloud cover timeseries was non-stationary, while the rest of the data were stationary. To make the cloud cover data stationary, first-order differencing was performed. As auto-correlated errors could still exist in the differenced timeseries, AR (p≥1) and MA (q≥1) terms could be added to compensate for any mild under-differencing or over-differencing respectively. The ADF test results are shown in Table 1 and the differencing plots in fig 4.

|  |
| --- |
|  |
| **Fig 3. Methodology followed in the study** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Mean Temperature | Avg. Min. Temperature | Avg. Max. Temperature | Precipitation | Cloud Cover |
| Test Statistics | -4.519 | -4.186 | -4.938 | -5.494 | -0.629 |
| p-value | 0.000 | 0.001 | 0.000 | 0.000 | 0.864 |
| Lags | 23 | 23 | 23 | 24 | 23 |
| Observations | 1404 | 1404 | 1404 | 1403 | 1404 |
| Reject | True | True | True | True | False |
| **Table 1. Augmented Dickey-Fuller Test results for the dataset.** | | | | | |

|  |
| --- |
| 1. **(b)** |
| **Fig 4. Differencing plots for cloud cover:**  **(a). Before Differencing and (b). After Differencing** |

**4.2.2 Determination of ARIMA Hyperparameters**

Auto-correlation function (ACF) and partial auto-correlation function (PACF) plots were used for determining the values for the AR (p) and MA (q) terms. The ACF and PACF plots showing large values decomposing very slowly over time shows the need for differencing. The plots also show a confidence level (95%), towards which, if the values show a sharp cut-off in ACF, denotes some over differencing in the series and the need for an MA(q) term. In case of PACF, a sharp cut-off represents under differencing and the need for an AR (p) term. The ACF-PACF plots are shown in fig. 5 - 9. It was found that in case of temperature, maximum and minimum temperature, the ACF plot showed a sharp cut-off after the 3rd lag, thus providing us the value for the MA(q) term. The PACF plot showed a sharp cut-off after the 12th lag and a dip after the 4th lag, providing us with a value of either 12 or 4 for the AR(q) term, both of which were then used in the ARIMA model and the better performing of the two selected. So, the AR(p) term for temperature was 12 and that for minimum and maximum temperature was either 12 or 4, and the MA(q) term for all three of the temperature variables was 3, resulting in the selection of ARIMA (12,1,3) for temperature, ARIMA (12,1,3) for maximum temperature, ARIMA (4,1,3) for minimum temperature. Similar steps were followed to determine the ‘p’ and ‘q’ terms for both cloud cover and precipitation, with ACF plots for cloud cover showing a cut-off after the 12th lag and the PACF plot also shows a cut-off after the 12th lag, resulting in the selection of ARIMA (12, 1, 12). The ACF plot for precipitation shows a repeating pattern, indicative of one of the pattern cut-offs being the ideal value for the ‘q’ term, while its PACF plot shows a cut-off after the 11th lag, and ARIMA (11, 1, 12). The ‘m’ term is used to remove seasonality in the timeseries. Since the data is a monthly dataset, with each year having 12 months, the same value was chosen, i.e., m=12. Appropriate values for P, D, Q are then chosen for m=12 in a similar manner as the values for p, d and q.

|  |
| --- |
| **(a) (b)**  **Fig 5.****ACF and PACF plots for Mean Temperature** |
|  |
| **(a) (b)**  **Fig 6. ACF and PACF plots for Average Minimum Temperature** |
|  |
| 1. **(b)**   **Fig 7. ACF and PACF plots for Average Maximum Temperature** |

|  |
| --- |
| **(a) (b)**  **Fig 8. ACF and PACF plots for Cloud Cover** |
| **(a) (b)**  **Fig 9. ACF and PACF plots for total precipitation.** |

**4.2.3 Evaluation Metrics used**

The results are graded using multiple evaluation metrics, such as RMSE, AIC, BIC as well as distribution metrics such as skew, kurtosis and stationarity R­2, all of which are inherent to the ARIMA model.

(i). Akaike Information Criterion (AIC):

K being the number of parameters.

AIC is an estimator of prediction error. When a statistical model is used to represent the process that generated the provided data, the representation is never exact, some information is always lost. AIC simply estimates this relative amount of information lost by our model.

(ii). Bayesian Information Criterion (BIC):

K being the number of parameters and N being the size of the dataset.

BIC tries to find the “True” model amongst a finite set of models. BIC introduces penalty terms into a model for each parameter that may increase the likelihood, thus preventing overfitting.

(iii). Root Mean Square Error (RMSE):

The RMSE denotes how close the values predicted by the trained model are to the actual observed values. The differences between the values, such as temperature or precipitation, known as residuals, help in the determination of a model’s quality.

(iv). R-squared:

The R-squared value denotes how well the model has fit on the training data, thus providing a measure for the model’s evaluation instead of the forecasts.

(v). Skew:

Skew is a statistical metric that shows how much the distribution of the data is concentrated on either side (left/negative or right/positive). It helps in determining if the predictions made are overtly biased towards either ends of the scale, a value close to 0 is preferred.

(vi). Kurtosis:

Kurtosis is a statistical measure that is used to determine if the predictions made contain any outliers, thus, helping in training a consistent model that doesn’t make extreme predictions.

**4.2.4 Selection of the Best ARIMA Model**

Once the appropriate values for the hyperparameters (p, d, q, P, D, and Q) are calculated, the most well-fitting model is determined based on the residual values of the ARIMA models. Data from the years 1901-2000 is used for training the models and forecasting is performed from 2001-2100, with the data from 2001-2019 acting as validation sets for the fitted models.

The most appropriate model for mean temperature is found to be ARIMA (12,0,3) (0,0,3)12, ARIMA (12,1,3) (0,1,3)12 for maximum temperature and for minimum temperature, ARIMA (4,1,3) (0,1,3)12. The model selected for precipitation based on the ACF/PACF plots, turns out to be ARIMA (11,1,12) with no seasonal part. In a similar manner ARIMA (12, 1, 12) was selected for modelling the cloud cover data.

**CHAPTER 5**

**RESULTS**

**5.1 Workstation Environment and Setup**

A computer system with an AMD Ryzen 5 4500U APU, 16 GB RAM and AMD Vega 8 GPU was used for training and visualising the data, as well as making forecasts. The training process was relatively fast, owing to the simplicity of the model, as well as the processing power made available, taking an accumulative time of less than a day, for all variables.

**5.1.1 Software Used**

The visualisation, modelling and forecasting process was undertaken using Python3 and an Anaconda environment consisting of various libraries such as Matplotlib, Seaborn, Statsmodels, Plotly, Pandas, and SciKit Learn.

Python is an interpreted high-level general-purpose programming language. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented and functional programming. Pandas is a software library written for Python, for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series and is used to pre-process, clean and manage datasets. Matplotlib is a plotting library for Python and its numerical mathematics extension NumPy. It is used to generate different kinds of plots with a high level of customisability. Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. Plotly is an interactive, open-source, and browser-based graphing library for Python. It is a high-level, declarative charting library that ships with over 30 chart types, including scientific charts, 3D graphs, statistical charts, SVG maps, financial charts, and more.

**5.2 Data Analysis and Key Observations**

A key part of our study was to analyse and visualise the extent of climate that has occurred over the years up until now, especially since it has been a topic that hasn’t received much attention over the years when it comes to the valley of Kashmir. The timeseries data was plotted using specialised software, like Matplotlib, Plotly and Seaborn, and trend lines fitted upon the graphs to better visualise the direction and change in the temperatures, precipitation and cloud cover over the period of study. The analysis resulted in some key observations;

**5.2.1 Drastic changes in temperature**

The change in mean temperature over the period of 119 years in the region can be observed in the temperature graphs (Fig. 10-11), showing about 10-25% change in mean temperature in springs, summers and autumns. The more worrying part is the gradual but extreme increase of about 200% in the mean temperatures over the past winters, as shown in Table 2 and Table 3. This is a worrying sign for the region, considering its ecologically sensitive nature. Such drastic and extreme increase in the mean temperature, especially for winters, can result in the accelerated melting of glaciers which may result in flooding and even affect the perennial nature of the rivers originating from them. The rise in mean temperature is also a problem for most of the fruit-bearing trees in the region such as apple, apricot, walnut, all of which have their own “Chilling hour” requirements, insatiability of which can cause abnormalities in the yield. (Stafne, 2017; Patel, N. R., et al, 2019; Das, Biswajit, et al., 2011; Salama, Abdel-Moety, et al, 2021).

**5.2.2 Precipitation spikes**

Upon analysis of the precipitation data, extreme spikes in precipitation were noticed in summers and autumns, causing flash floods in localised areas such as that in 2010 (The Guardian, 2010) and major floods like that of 2014 (Greater Kashmir, 2015). The precipitation data when separately plotted for 1901-1990 and 1991-2019 shows a lower mean and deviation in the former time period and a higher mean and deviation, pointing to more erratic rainfalls. The precipitation keeps peaking in late summers and early autumns, as shown in Table 4 and Fig. 12.

**5.2.3 Potential for using temperature and cloud cover forecasts to predict heavy rainfall**

Cloud cover is one of the most overlooked climate factors when it comes to analyses of the valley’s climate. A trendline was fitted upon the data points to analyse the co-dependence of the variables. The trendline is a locally weighted polynomial regression line fit using weighted least squares giving more weight to points near the point whose response is being estimated and less weight to the points further away. Upon analysis of the data and the trendline, it was observed that the cloud cover usually sits around a small range of 41-47% and a relatively larger range of 23-34% in summers and autumns respectively, correlating to precipitations of 3000-8000 mm in summers and 1000-3000 mm in autumns. The deviances and spikes in precipitations can be correlated to cloud covers greater than 55% in summers and 40% in autumns, as shown in Fig. 13-14. A multi-variate approach, using the precipitation and cloud cover data, can thus be taken in early prediction of such spikes, which can thus aid in prevention of catastrophic loss of property and life, such as the ones caused in 2010 and 2014.

|  |
| --- |
| **Fig 10. Mean temperature over 1901-2019 in the autumn season (Sep, Oct, Nov)** |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Fig 11. Mean temperature over 1901-2019 in the winter season (Dec, Jan, Feb)** | | | | | | | |
| **Years** | **Mean** | **Standard Deviation** | **Min** | **25%** | **50%** | **75%** | **Max** |
| 1901-1990 | 0.870 | 1.509 | -3.28 | -0.27 | 1.003 | 1.942 | 4.488 |
| 1991-2019 | 1.506 | 1.420 | -1.21 | 0.36 | 1.669 | 2.709 | 5.000 |
| **Table 2. Changes in mean temperature for winter (1991-2019) and (1901-1990)** | | | | | | | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Years** | **Mean** | **Standard Deviation** | **Min** | **25%** | **50%** | **75%** | **Max** |
| 1901-1990 | 11.685 | 0.555 | 10.57 | 11.29 | 11.68 | 12.04 | 13.21 |
| 1991-2019 | 12.231 | 0.391 | 11.61 | 11.88 | 11.88 | 12.52 | 12.93 |
| **Table 3. Changes in mean temperature for autumn (1901-2000) and (2001-2019)** | | | | | | | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Years** | **Mean** | **Standard Deviation** | **Min** | **25%** | **50%** | **75%** | **Max** |
| 1901-1990 | 11960.7 | 2278.3 | 7491 | 10227 | 11651 | 13350 | 17319 |
| 1991-2019 | 13295.7 | 2427.7 | 10421 | 11490 | 12620 | 14761 | 18997 |
| **Table 4. Total yearly precipitation compared over the years, 1901-1990 vs. 1991-2019** | | | | | | | |

|  |
| --- |
| **Fig 12. Comparison of mean monthly precipitation for 1901-1990 and 1991-2019** |

|  |
| --- |
| **Fig 13. Correlation between cloud cover & precipitation spikes during autumns** |

|  |
| --- |
|  |
| **Fig 14. Correlation between cloud cover & precipitation spikes during summers** |

**5.3 Modelling and Future Forecasts**

The time-series data for each of the variables, viz. mean temperature, average maximum temperature, average minimum temperature, precipitation and cloud cover was separately trained upon & forecasted using different ARIMA models. The training data used was from the years 1901-1999 and forecasts were made for the next century i.e., 2000-2100 with data from 2000-2019 being used for validation of the forecasts. Although the ADF tests showed the data for precipitation, temperature, minimum and maximum temperature to be stationary, some seasonality was found using ACF/PACF plots in minimum and maximum temperature, which showed the timeseries having some sequentially recurring trends in the lags and slow decay over time.

**5.3.1 Temperature Forecasts**

The results for the mean temperature, average minimum temperature and average maximum temperature show a very low RMSE value, resulting in very good results in terms of accuracy. The mean temperature shows the most prominent change, doubling by the end of the century for the month of January and showing an increase of about 50% on average for the other two months of winter (December and February), when compared to that in 2020. Springs are also predicted to get hotter, which further worsens the issue regarding the lack of the “chilling period” required by fruit trees in the region, as discussed earlier in the analysis phase. The summers and autumns are seen to be getting cooler when looking at the mean temperature for Jun-Nov, whilst also seeing an increase in the minimum and maximum temperatures, pointing to an increase in the erratic behaviour seen in the said seasons over recent years.

The models fit for each of the temperature variables (mean, maximum and minimum) have been shown in detail in Fig. 15-17 and their evaluation results in Table 5.

Yearly forecast results have been, and shown as values in 20-year intervals between 2020 and 2100 in Table 6-8, and to better visualise the change predicted in the temperature, plots showing the same are provided in Fig. 18-20.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Climate Variable** | **AIC** | **BIC** | **RMSE** | **Skew** | **Kurtosis** | **R2** |
| **Mean Temperature** | 3442.07 | 3537.94 | 1.1296 | -0.06 | 3.92 | 0.973 |
| **Minimum Temperature** | 3288.46 | 3343.84 | 1.1504 | 0.17 | 3.70 | 0.970 |
| **Maximum Temperature** | 3432.28 | 3527.93 | 1.1514 | 0.09 | 3.84 | 0.974 |
| **Table 5. Model evaluation results for mean, minimum and maximum temperature** | | | | | | |

|  |
| --- |
| **Fig 15. ARIMA (12,0,3) (0,0,3)12 model validation plot for mean temperature (°C)** |

|  |
| --- |
| **Fig 16. ARIMA (12,1,3) (0,1,3) 12 model validation plot for minimum temperature (°C)** |

|  |
| --- |
| **Fig 17. ARIMA (4,1,3) (0,1,3)12 model validation plot for maximum temperature (°C)** |
| **Fig 18. Yearly mean temperature (°C) changes over the years 2020-2100** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Month** | **2020** | **2040** | **2060** | **2080** | **2100** |
| **JAN** | 1.264 | 1.837 | 2.246 | 2.574 | 2.874 |
| **FEB** | 2.243 | 2.723 | 3.165 | 3.544 | 3.880 |
| **MAR** | 5.773 | 5.951 | 6.251 | 6.565 | 6.856 |
| **APR** | 10.664 | 10.596 | 10.685 | 10.846 | 11.027 |
| **MAY** | 15.376 | 15.287 | 15.232 | 15.233 | 15.270 |
| **JUN** | 18.657 | 18.698 | 18.617 | 18.521 | 18.442 |
| **JUL** | 19.864 | 19.972 | 19.927 | 19.813 | 19.681 |
| **AUG** | 18.904 | 18.894 | 18.856 | 18.769 | 18.652 |
| **SEP** | 16.029 | 15.823 | 15.748 | 15.697 | 15.639 |
| **OCT** | 11.773 | 11.527 | 11.444 | 11.440 | 11.463 |
| **NOV** | 7.050 | 7.033 | 7.053 | 7.131 | 7.247 |
| **DEC** | 3.132 | 3.475 | 3.696 | 3.899 | 4.114 |
| **Table 6. Mean temperature (°C) forecasts for the years 2020-2100** | | | | | |
|  | | | | | |
| **Fig 19. Yearly average minimum temperature (°C) changes over the years 2020-2100** | | | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Month** | **2020** | **2040** | **2060** | **2080** | **2100** |
| **JAN** | -3.335 | -3.289 | -3.247 | -3.206 | -3.165 |
| **FEB** | -2.127 | -2.074 | -2.034 | -1.993 | -1.952 |
| **MAR** | 2.133 | 2.163 | 2.203 | 2.244 | 2.284 |
| **APR** | 6.710 | 6.746 | 6.788 | 6.829 | 6.870 |
| **MAY** | 10.222 | 10.277 | 10.318 | 10.359 | 10.400 |
| **JUN** | 13.827 | 13.864 | 13.904 | 13.945 | 13.985 |
| **JUL** | 15.638 | 15.668 | 15.709 | 15.750 | 15.791 |
| **AUG** | 15.081 | 15.131 | 15.173 | 15.214 | 15.255 |
| **SEP** | 11.638 | 11.684 | 11.724 | 11.765 | 11.806 |
| **OCT** | 6.584 | 6.612 | 6.652 | 6.693 | 6.734 |
| **NOV** | 1.629 | 1.672 | 1.714 | 1.755 | 1.796 |
| **DEC** | -1.579 | -1.527 | -1.486 | -1.445 | -1.404 |
| **Table 7. Average minimum temperature (°C) forecasts for the years 2020-2100** | | | | | |
|  | | | | | |
| **Fig 20. Yearly average maximum temperature (°C) changes over the years 2020-2100** | | | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Month** | **2020** | **2040** | **2060** | **2080** | **2100** |
| **JAN** | 5.799 | 5.814 | 5.829 | 5.844 | 5.858 |
| **FEB** | 7.324 | 7.339 | 7.354 | 7.369 | 7.383 |
| **MAR** | 11.890 | 11.905 | 11.920 | 11.934 | 11.949 |
| **APR** | 17.580 | 17.595 | 17.609 | 17.624 | 17.639 |
| **MAY** | 21.982 | 21.997 | 22.012 | 22.027 | 22.041 |
| **JUN** | 25.899 | 25.914 | 25.929 | 25.944 | 25.958 |
| **JUL** | 25.623 | 25.638 | 25.653 | 25.667 | 25.682 |
| **AUG** | 24.672 | 24.687 | 24.702 | 24.717 | 24.731 |
| **SEP** | 23.565 | 23.580 | 23.595 | 23.610 | 23.624 |
| **OCT** | 19.675 | 19.689 | 19.704 | 19.719 | 19.734 |
| **NOV** | 14.185 | 14.200 | 14.215 | 14.229 | 14.244 |
| **DEC** | 8.516 | 8.531 | 8.545 | 8.560 | 8.575 |
| **Table 8. Average maximum temperature (°C) forecasts for the years 2020-2100** | | | | | |

**5.3.2 Cloud Cover Forecasts**

The ARIMA model fitted for cloud cover achieved a relatively higher RMSE than what was achieved for the model fitted for temperature values. The model thus tends to underestimate the peaks of the cloud and makes forecasts that are around 4-8% lower than what the actual values may be.

The evaluation metrics for the model are shown in Table 9 and the model fit values are compared to actual observed values for validation in Fig. 21. The cloud cover values for different seasons have also been visualised in Fig. 22-25, with springs showing a very slight decrease, winters showing a relatively larger decline in values, summers and autumns showing slightly increasing values as the years go on.

The forecast results from the model are presented in Table 10, in 20-year intervals.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **AIC** | **BIC** | **RMSE** | **Skew** | **Kurtosis** | **R2** |
| **Cloud Cover** | 6262.51 | 6389.22 | 11.892 | -0.462 | -0.967 | 0.2483 |
| **Table 9. ARIMA model evaluation results for cloud cover** | | | | | | |
|  | | | | | | |
| **Fig 21 ARIMA (12,1,12) model validation plot for cloud cover (%)** | | | | | | |
|  | | | | | | |
| **Fig 22. Cloud cover (%) forecast changes for winter over the years 2020-2100** | | | | | | |
|  | | | | | | |
| **Fig 23. Cloud cover (%) forecast changes for spring over the years 2020-2100** | | | | | | |
|  | | | | | | |
| **Fig 24 Cloud cover (%) forecast changes for summer over the years 2020-2100** | | | | | | |

|  |
| --- |
|  |
| **Fig 25. Cloud cover (%) forecast changes for autumn over the years 2020-2100** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Month** | **2020** | **2040** | **2060** | **2080** | **2100** |
| **JAN** | 50.845 | 50.598 | 50.356 | 50.122 | 49.894 |
| **FEB** | 54.339 | 54.054 | 53.778 | 53.509 | 53.247 |
| **MAR** | 56.310 | 56.225 | 56.127 | 56.020 | 55.904 |
| **APR** | 48.436 | 48.314 | 48.201 | 48.094 | 47.992 |
| **MAY** | 41.633 | 41.708 | 41.776 | 41.841 | 41.902 |
| **JUN** | 36.039 | 36.447 | 36.850 | 37.243 | 37.628 |
| **JUL** | 46.171 | 46.194 | 46.212 | 46.225 | 46.236 |
| **AUG** | 46.365 | 46.330 | 46.297 | 46.264 | 46.231 |
| **SEP** | 32.266 | 32.273 | 32.288 | 32.314 | 32.351 |
| **OCT** | 24.373 | 24.717 | 25.054 | 25.382 | 25.701 |
| **NOV** | 27.485 | 27.730 | 27.970 | 28.207 | 28.441 |
| **DEC** | 45.190 | 44.861 | 44.541 | 44.228 | 43.923 |
| **Table 10. Forecasts for average yearly cloud cover (%)** | | | | | |

**5.3.3 Precipitation Forecasts**

The ARIMA model produces a high RMSE value for the precipitation dataset, mainly due to its inability to adapt to erratic changes in the data points which is essentially a high variance between the points. The model is unable to replicate the peaks that are expected when comparing the forecasted values with the observed values and thus creates a flatter, middle of the ground curve. The model evaluation results are discussed in Table 11. The validation fit for the model has been shown as a plot in Fig. 26, and the values forecast using the model are presented in Table 12, with visualisations for the same in Fig. 27, showing an overall decrease in the total precipitation in almost all seasons, especially the summers, pointing to drier climates.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **AIC** | **BIC** | **RMSE** | **Skew** | **Kurtosis** | **R2** |
| **Precipitation** | 18005.49 | 18127.13 | 504.923 | -0.073 | -1.106 | 0.457 |
| **Table 11. ARIMA model evaluation results for precipitation** | | | | | | |

|  |
| --- |
|  |
| **Fig 26. ARIMA (11,1,12) model validation plot for precipitation (mm/month)** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Month** | **2020** | **2040** | **2060** | **2080** | **2100** |
| **Jan** | 827.259 | 822.059 | 829.926 | 846.947 | 873.467 |
| **Feb** | 1292.45 | 1353.94 | 1408.87 | 1463.19 | 1515.35 |
| **Mar** | 1482.47 | 1513.28 | 1544.88 | 1571.42 | 1592.40 |
| **Apr** | 1135.92 | 1144.92 | 1153.38 | 1163.77 | 1179.93 |
| **May** | 850.69 | 877.804 | 909.056 | 949.669 | 997.282 |
| **Jun** | 1163.02 | 1196.69 | 1236.43 | 1271.64 | 1298.92 |
| **Jul** | 1659.21 | 1655.83 | 1634.69 | 1600.16 | 1555.79 |
| **Aug** | 1612.55 | 1547.77 | 1479.96 | 1415.04 | 1353.35 |
| **Sep** | 983.912 | 945.346 | 921.974 | 904.800 | 891.905 |
| **Oct** | 434.783 | 458.288 | 471.460 | 478.244 | 478.894 |
| **Nov** | 336.650 | 317.251 | 290.615 | 258.438 | 220.468 |
| **Dec** | 483.213 | 427.805 | 381.267 | 339.437 | 304.823 |
| **Table 12. Forecasts for total precipitation (mm/month) for the years 2020-2100** | | | | | |

|  |
| --- |
|  |
| **Fig 27. Forecasted average precipitation (mm/month) for the years 2020-2100** |

**CHAPTER 6**

**CONCLUSION  
AND  
FUTURE SCOPE**

**6.1 Conclusion**

The climate of any region is mainly dependant on the temperature and precipitation observed there.

In the study, these factors, in addition to cloud cover were studied using the ARIMA model and forecasts were made.

The temperature variables i.e., mean, maximum and minimum temperature were predicted by the model with very low error values and it was found that the mean temperatures for the region, especially in winters are going to shift further from the natural range, towards an overall hotter climate, with a forecasted increase of nearly 2°C for winters from that in 2019, in mean temperatures.

The temperatures show such an increase that the region is devoid of any months during the winter season with a mean temperature lower than 1°C, from the year 2048. This is a worrying sign for a region used to and needing of sub-zero temperatures for winters.

The minimum and maximum temperatures also see a rise, although less pronounced than the mean temperature. The forecasts also show the summers and autumns in the region to get ever so slightly cooler, while the springs show a relatively higher increase in temperature.

Although the model was found to be unable to predict extreme rainfall events with a reasonably small error value, the mean precipitation forecasts were found to be closer to actuality. The decrease in precipitation and increase in cloud cover during summers implies that an increasingly humid climate is abound for the valley, which can further prove to be disastrous for the valley, taking into account the sensitive nature of the region.

It is thus concluded from the study that the forecast results achieved using ARIMA modelling, will prove beneficial in providing assistance to scientists and stakeholders to develop strategies and plans to conserve the climate of the region.

**6.2 Future Scope**

Although strongly accurate results with a very low error values were achieved for mean, average minimum and average maximum temperature variables, the results for cloud cover and precipitation did not reflect similar characteristics, owing to the inability of the ARIMA model to fit to, and replicate peaks and drastic changes. It can, thus, be said that with use of different machine learning models, the results can be further improved upon, with a better prediction rate for extreme events.

Another area that can be improved upon, is the use of higher resolution datasets, which can allow for finer, cluster-wise analyses and forecasts. Such datasets can allow for separation of data for planes, and mountainous regions, providing a clearer picture of the impact on glaciers and snow-caps in the region.

**REFERENCES**

United Nations Organization, Global Issues – Climate Change. *United Nations*. Extracted on 02 June, 2021 https://www.un.org/en/global-issues/climate-change

IPCC, 2007: *Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Core Writing Team, Pachauri, R.K and Reisinger, A. (eds.)]. IPCC, Geneva, Switzerland, 104 pp.

Global Warming of 1.5°C, *Intergovernmental Panel on Climate Change (IPCC) Special Report 2015,* IPCC.

AR6 Climate Change 2021, *Intergovernmental Panel on Climate Change (IPCC) Special Report 2021,* IPCC.

Allen, M.R., O.P. 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, 2018: Framing and Context. In: *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*[Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. In Press.

United Nations Organization, IPCC report: ‘Code red’ for human driven global heating, warns UN chief. *United Nations*. https://news.un.org/en/story/2021/08/1097362

Meher-Homji, VM On the Mediterranean climatic regime of West Pakistan. *Arch. Met. Geoph. Biocl. B.***19,** 277-286 (1971).

Ahmed, Development of Hydropower Projects in Jammu and Kashmir, *International Journal of Trend in Research and Development*, Volume 6(1).

Ahmed. Are Kashmir’s New Hydro-Electric Projects Economically viable? *The Wire.* https://science.thewire.in/environment/kashmir-new-ratle-hydroelectric-power-projects-economically-viable/

Bukhari, Climate change ravages Kashmir's 'red gold' saffron crop, Phys.org (2020, December 22)

Romshoo, S.A., Fayaz, M., Meraj, G. *et al.* Satellite-observed glacier recession in the Kashmir Himalaya, India, from 1980 to 2018. *Environ Monit Assess* **192,** 597 (2020).

M. H. Wani, Climate change in Kashmir valley: Is it initiating transformation of mountain agriculture?.*Indian Journal of Economics & Development*. 2015.

Romshoo, S.A., Bashir, J. & Rashid, I. Twenty-first century-end climate scenario of Jammu and Kashmir Himalaya, India, using ensemble climate models. *Climatic Change* **162,** 1473–1491 (2020).

Zaz, S. N., S. A. Romshoo, R. T. Krishnamoorthy, and Y. Viswanadhapalli. ‘Analyses of Temperature and Precipitation in the Indian Jammu and Kashmir Region for the 1980–2016 Period: Implications for Remote Influence and Extreme Events’. *Atmospheric Chemistry and Physics* 19, no. 1 (2019): 15–37.

Islam, Zahoor & Rao, L.A.K. (2013). Climate Change Scenario in Kashmir Valley, India, based on Seasonal and Annual Average Temperature Trends. Disaster Advances. 6. 30-40.

Romshoo. The 2014 Kashmir Flood: The Extreme of the Extremes. *Greater Kashmir* (2015, March 14).

Shafiq, M.u., Ramzan, S., Ahmed, P. *et al.* Assessment of present and future climate change over Kashmir Himalayas, India. *Theor Appl Climatol* **137,** 3183–3195 (2019).

Das. Beyond Hot Summer, Kashmir Facing Multiple Ecocide Offense. *Kashmir Observer Web Desk*. Kashmir Observer (2021, August 25).

Rolnick, David, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, *et al.* Tackling Climate Change with Machine Learning. 2019.

Jebeile, J., Lam, V. & Räz, T. Understanding climate change with statistical downscaling and machine learning. *Synthese* (2020).

Chris Huntingford et al 2019 Environ. Res. Lett. **14** 124007.

J&K Climate Data Github repository. *Fahad Farooq.* https://github.com/fahad-farooq/climatedatajk

University of East Anglia Climatic Research Unit; Harris, I.C.; Jones, P.D.; Osborn, T. (2020): CRU TS4.04: Climatic Research Unit (CRU) Time-Series (TS) version 4.04 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2019).

Data Subsetter. Web Processing Service (WPS). *Centre for Environmental Data Analysis Climate Research Unit*. https://www.ceda-wps-ui.ceda.ac.uk/processes/execute?wps=063fb257cfd94a3f89ba7f3753f7864e&process=SubsetCRUTimeSeries

Open Government Data (OGD) Platform, *Government of India*. https://data.gov.in/catalog/all-india-seasonal-and-annual-mean-temperature-series

Climate Data. *Berkeley Earth.* http://berkeleyearth.org/data/

Background on the Meteorological Datasets, *Indian Water Portal.* https://www.indiawaterportal.org/articles/background-meteorological-datasets

Harris, I., Osborn, T.J., Jones, P. *et al.* Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci Data* **7,** 109 (2020).

Shi, Haiyun, Tiejian Li, and Jiahua Wei. ‘Evaluation of the Gridded CRU TS Precipitation Dataset with the Point Raingauge Records over the Three-River Headwaters Region’. *Journal of Hydrology* 548 (2017): 322–32.

Salvacion, A.R., Magcale-Macandog, D.B., Cruz, P.C.S. *et al.* Evaluation and spatial downscaling of CRU TS precipitation data in the Philippines. *Model. Earth Syst. Environ.* **4,** 891–898 (2018).

Dimri, T., Ahmad, S. & Sharif, M. Time series analysis of climate variables using seasonal ARIMA approach. *J Earth Syst Sci* **129,** 149 (2020).

Kashmir flash floods leave dozens dead, *The Guardian* (2010, August 06).

Youssef, Jamile and Ishker, Nermeen and Fakhreddine, Nour (2021): GDP Forecast of the Biggest GCC Economies Using ARIMA

Lhano, Francisco and Gouveia, Jorge and Perestrello, Francisco and Melo, Alexandre and Gomes, Miguel, Using an ARIMA Model to Forecast GDP Until 2031 for Portugal and Germany (May 27, 2021).

NYONI, THABANI (2019): Is Nigeria's economy progressing or backsliding? Implications from ARIMA models

Khan, Farhan Mohammad, and Rajiv Gupta. ‘ARIMA and NAR Based Prediction Model for Time Series Analysis of COVID-19 Cases in India’. *Journal of Safety Science and Resilience* 1, no. 1 (2020): 12–18.

Ding, Guorong, Xinru Li, Fan Jiao, and Yang Shen. ‘Brief Analysis of the ARIMA Model on the COVID-19 in Italy’. *MedRxiv*, 2020.

Sahai, Alok Kumar, Namita Rath, Vishal Sood, and Manvendra Pratap Singh. ‘ARIMA Modelling & Forecasting of COVID-19 in Top Five Affected Countries’. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 14, no. 5 (2020): 1419–27.

Arash Yoosef Doost, Mohammad Sadegh Sadeghian, Mohammad Ali Node Farahani, Ana Rasekhi. Comparison between Performance of Statistical and Low Cost ARIMA Model with GFDL, CM2.1 and CGM 3 Atmosphere-Ocean General Circulation Models in Assessment of the Effects of Climate Change on Temperature and Precipitation in Taleghan Basin. *American Journal of Water Resources*. Vol. 5, No. 4, 2017, pp 92-99.

Patel, N. R., et al. "Geospatial Technology for Climate Change Impact Assessment of Mountain Agriculture." *Remote Sensing of Northwest Himalayan Ecosystems*. Springer, Singapore, 2019. 381-400.

Das, Biswajit, et al. "Harvest maturity standards and fruit quality of some apple cultivars under high altitudal conditions." *Indian Journal of Horticulture* 68.2 (2011): 170-179.

Salama, Abdel-Moety, et al. "Temperate Fruit Trees under Climate Change: Challenges for Dormancy and Chilling Requirements in Warm Winter Regions." *Horticulturae* 7.4 (2021): 86.

Stafne, Eric T. *Chilling-hour Requirement of Fruit Crops*. Mississippi State University Extension, 2017.

Whittle, Peter. *Hypothesis testing in time series analysis*. Vol. 4. Almqvist & Wiksells boktr., 1951.

Swain, S., S. Nandi, and P. Patel. "Development of an ARIMA model for monthly rainfall forecasting over Khordha district, Odisha, India." *Recent Findings in Intelligent Computing Techniques*. Springer, Singapore, 2018. 325-331.

Praveen, B, Sharma, P. Climate variability and its impacts on agriculture production and future prediction using autoregressive integrated moving average method (ARIMA). *J Public Affairs*. 2020; 20:e2016.

Priya Narayanan, Ashoke Basistha, Sumana Sarkar, Sachdeva Kamna, Trend analysis and ARIMA modelling of pre-monsoon rainfall data for western India, *Comptes Rendus Geoscience*, Volume 345, Issue 1, 2013, Pages 22-27, ISSN 1631-0713.

Naseem, Faiza, et al. "An integrated approach to air pollution modeling from climate change perspective using ARIMA forecasting." *Journal of Applied Agriculture and Biotechnology* 2.2 (2018): 37-44.

Smith G., “In Kashmir, water treaty means less power to the people.” *The Globe and Mail* (2010, 27th July). https://www.theglobeandmail.com/news/world/in-kashmir-water-treaty-means-less-power-to-the-people/article1389907/

1. https://data.ceda.ac.uk/badc/cru/data/cru\_ts/cru\_ts\_4.04 [↑](#footnote-ref-1)
2. https://ceda-wps-ui.ceda.ac.uk/processes [↑](#footnote-ref-2)