Multivariate Statistical Approach and Correlation Analysis for Climate Variable Modelling in Regional Weather Prediction

Keshav Kumar[1*], Madan Kumar[2], Mani Bhushan[3], Niraj Kumar Singh[4], Saurabh Kumar[5], Arvind Kumar[6], Nandan Kumar Raju[7], Vivek Kumar[7], Ashok Kumar Chaudhary[8]

[1] Assistant Professor, Dept. of Civil Engineering, Nalanda College of Engineering Chandi, DSTTE, Gov. of Bihar (India)
[2] Lecturer, Dept. of Civil Engineering, Government Polytechnic Muzaffarpur, Bihar, India-842001
[3] Principal in Charge, Government Engineering College Khagaria, DSTTE, Gov. of Bihar (India)
[4] Principal, Government Polytechnic Jamui, DSTTE, Gov. of Bihar (India)
[5] Lecturer, Dept. of Civil Engineering, Government Polytechnic Asthawan, Nalanda
[6] Assistant Professor, Dept. of Civil Engineering, Government Engineering College Sheikhpura
[7] Assistant Professor, Dept. of Civil Engineering, Supaul College of Engineering Supaul
[8] Assistant Professor, Dept. of EEE Engineering, Nalanda College of Engineering Chandi

Abstract
This study investigates the selection of predictor variables for temperature estimation using various statistical techniques across five districts in Bihar, namely Nalanda, Saran, Bhagalpur, Kaimur, and West Champaran. Bihar, located in the northern plains of India, has an agrarian economy that is highly sensitive to climatic and weather conditions. Understanding the variability and influence of different predictor variables is essential for effective management of agriculture, irrigation, hydrology, and sustainable development initiatives. The analysis is based on ten climatic predictor variables: specific humidity, maximum temperature, minimum temperature, mean sea level pressure, geopotential height, divergence, u-wind at 500 mBar, v-wind at 500 mBar, u-wind at 1000 mBar, and v-wind at 1000 mBar. These variables were sourced from the NCEP/NCAR Reanalysis dataset and subsequently converted from netCDF to tabular format using ArcMap (ArcGIS). Key statistical parameters such as mean, variance, and covariance matrices were computed, followed by the construction of correlation matrices for each district. Principal Component Analysis (PCA) was employed to examine the interrelationships among the variables and to identify dominant patterns influencing temperature variations. Visual tools including scatter plot matrices and histograms were used to further analyze variable interactions. The correlation monoplots helped determine the nature and strength of associations based on vector magnitudes and orientations. The outcomes of this study offer valuable insights for meteorologists and climate scientists in forecasting temperature and rainfall patterns. These findings can also support data-driven decision-making in climate-resilient agricultural planning and water resource management in the selected regions of Bihar.

Keywords: Predictor Selection, PCA, ArcGIS, Statistical Technique, Climate Variable Correlation, Predictor Variables.

Introduction

Precipitation plays a critical role in the global water cycle, and the effects of anthropogenic climate change on precipitation patterns have far-reaching implications, particularly for agriculture. India, with its diverse climatic zones, experiences complex precipitation dynamics influenced by the Indian Ocean, monsoonal winds, and regional topography. The country predominantly experiences three distinct rainfall seasons: the Southwest Monsoon (June to September
), the Northeast Monsoon (October to December), and the Pre-monsoon period (March to May). While regions such as the Western Ghats and northeastern states receive high rainfall, areas like Rajasthan and Gujarat are characterized by arid conditions. The Southwest Monsoon alone contributes over 75% of India's annual rainfall, making it vital for agricultural productivity and economic stability.

This study focuses on analysing the spatial and temporal variation in precipitation across the state of Bihar. Occupying a geographical area of 94,163 km², Bihar is the 12th largest state in India and is landlocked by Nepal to the north, West Bengal to the east, Jharkhand to the south, and Uttar Pradesh to the west. The state is endowed with fertile alluvial soil, an extensive river network, rich biodiversity, and significant agricultural potential. Bihar ranks third in vegetable production, fourth in fruit production, and is the largest producer of litchi, guava, and makhana in the country. The state's predominantly semi-humid climate supports diverse crop cultivation. Its soils vary from fertile alluvial plains to mountainous and laterite soils in the foothills of the Himalayas and southern regions, respectively.

Climate change refers to long-term shifts in temperature and weather patterns. While natural causes such as solar activity and volcanic eruptions can influence climate, the primary driver of recent global warming is the increased concentration of greenhouse gases (GHGs) from human activities. Burning fossil fuels, deforestation, and industrial processes release carbon dioxide (CO_2) and methane (CH_4), which trap heat in the atmosphere. Key sectors contributing to GHG emissions include energy, transport, industry, buildings, agriculture, and land use.
Global average surface temperatures have risen by approximately 1.1°C since the late 19th century, marking the Earth's warmest period in over 100,000 years. The decade from 2011 to 2020 was the hottest on record. The consequences of climate change are increasingly evident in the form of extreme droughts, water scarcity, wildfires, sea-level rise, floods, melting glaciers, severe storms, and biodiversity loss. United Nations assessments warn that global temperatures are on track to rise by 2.8°C by the end of the century if current policies continue. To avoid the most severe impacts, it is critical to limit the global temperature rise to 1.5°C.
India's National Communication (NATCOM) to the United Nations Framework Convention on Climate Change (UNFCCC) documents several observed climate trends. These include a rise of approximately 0.4°C in surface air temperatures over the past century. Regional warming trends have been noted along the west coast, central India, the interior peninsula, and northeastern regions, while cooling trends have been observed in northwestern and parts of southern India. Although all-India monsoon rainfall trends remain statistically insignificant, significant regional variations exist. The Himalayan region, a vital water source due to its extensive glacial reserves, faces uncertain long-term impacts from glacial melt, with implications for river flows, water security, and hydropower generation.
Bihar is particularly vulnerable to climate change due to its geographical location, socio-economic challenges, and reliance on agriculture. The state frequently experiences extreme weather events such as droughts and floods, which adversely impact food security, water availability, public health, and infrastructure. Key areas of concern includes Droughts and Floods, Agricultural Disruptions and Food Insecurity, Water Scarcity and Health Risks and Damage to Infrastructure.
This study aims to explore the interdependence of various climatic predictor variables influencing precipitation and temperature across Bihar. The insights derived from this analysis will support meteorologists and climate scientists in enhancing the accuracy of climate projections and formulating strategies for climate-resilient development and agricultural planning in the region.

STUDY AREA

For this study, five districts of Bihar namely, **Nalanda, Saran, Bhagalpur, Kaimur**, and **West Champaran** were selected based on their distinct topographical characteristics and geographic diversity. These districts represent different physiographic zones of the state, providing a comprehensive framework for analyzing climatic variables and precipitation patterns.

- **Nalanda** is situated in the southern part of Bihar at an elevation of approximately 67 meters above sea level. It features a rugged topography and is traversed by the Phalgu River. Nalanda lies at a latitude of **25.2622° N** and a longitude of **85.4788° E**, covering an area of approximately **2,641 km²**.
- **Saran**, located in the north-central part of Bihar, has a relatively flat terrain and an average elevation of **36 meters** above sea level. It spans an area of about **2,641 km²**, with coordinates at **25.9963° N** latitude and **84.9431° E** longitude. The district is intersected by the Ghaghara and Gandak rivers.
- **Bhagalpur**, positioned in eastern Bihar, is situated at an elevation of **52 meters** above sea level and covers an area of approximately **2,569 km²**. The geographical coordinates of Bhagalpur headquarters are **25.2372° N** latitude and **86.9746° E** longitude. The Ganges River flows prominently through this district, making it significant for hydrological studies.
- **Kaimur**, located in the southwestern region of Bihar, lies at an elevation of **48 meters** above sea level. It covers an area of around **3,363 km²** and is geographically positioned at **25.0426° N** latitude and **83.6056° E** longitude. The district includes hilly terrain and forms part of the Kaimur plateau region.
- **West Champaran**, situated in the northwestern part of the state, has an elevation of approximately **70 meters** above sea level and spans the largest area among the selected districts, covering about **5,228 km²**. Its coordinates are **27.1543° N** latitude and **84.3542° E** longitude. The district is drained by important rivers such as the **Gandak (Narayani)** and **Sikrahana (Little Gandak)**.

Table 1: Latitude, longitude, area and % area distribution of the districts in terms of area of Bihar.

District	Latitude	Longitude	Area (km²)	% of State Area
Nalanda	25.2622° N	85.4788° E	2,367	2.51%
Saran	25.8560° N	84.8568° E	2,641	2.80%
Bhagalpur	25.2372° N	86.9746° E	2,569	2.72%
Kaimur	25.0426° N	83.6056° E	3,363	3.57%
West Champaran	27.1543° N	84.3542° E	5,228	5.55%

Figure 1: Map of Bihar showing the 5 districts with colour

Table 1 depicts the details about the latitude, longitude and others of all selected districts of Bihar used in this study. These districts were chosen to provide a spatially diverse representation of Bihar's climatic and geographical conditions. Their variation in elevation, river systems, and geographic location contributes to differing precipitation patterns and climate responses, making them ideal for this analysis. Fig. 1 represents the map of all selected district of Bihar state.

Methodology

The methodology employed in this research integrates a comprehensive approach to analysing climatic variables and their interrelationships using advanced statistical and data visualization techniques. This section delineates the systematic procedures adopted for data collection, preprocessing, variable selection, correlation analysis, and dimensionality reduction through Principal Component Analysis (PCA). The overarching goal is to identify key climate predictors and understand their interactions to facilitate accurate weather forecasting and climate modelling.

Selection of Potential Predictors

Identifying relevant predictors for temperature modelling is a complex task due to the multitude of dynamic and interdependent variables influencing atmospheric behaviour. In real-world scenarios, especially in climatology, large volumes of data are generated from various sources, necessitating the use of robust multivariate statistical techniques for efficient analysis.

Incorporating numerous variables into predictive models often leads to increased computational complexity, reduced model interpretability, and potential degradation in predictive performance. Therefore, effective variable selection and dimensionality reduction techniques are essential to enhance model efficiency, reduce noise, and improve the accuracy of predictions.

In this study, statistical methods are employed to identify and select the most relevant predictor variables for temperature downscaling. These variables are Specific humidity, Maximum temperature, Minimum temperature, Mean sea-level pressure, Divergence, Geopotential height, Horizontal Wind, Vertical Wind, Horizontal Wind, and Vertical Wind.

Two primary statistical techniques are utilized:

1. **Pearson Product Moment Correlation (PPMC):**

This method quantifies the linear relationship between individual predictor variables and the target variable (temperature). It helps in identifying predictors with strong positive or negative correlations, thereby facilitating the initial screening of relevant variables.

Pearson correlation assesses the degree of association between two datasets, each represented by a variable. These variables may reflect deterministic or stochastic data. In this context, the variables correspond to the percentage values of two types of errors: complex function errors (e_x) and measured CT errors (e_m). The correlation coefficient, denoted as r, is computed using Equation. This coefficient quantifies the linear relationship between two variables, ranging from +1 (indicating perfect positive correlation) to –1 (indicating perfect negative correlation). The value is derived by evaluating how two variables interact relative to their combined variability. Commonly known as the Pearson correlation coefficient, it provides a standardized measure of dependence between datasets.

$$r = \frac{\sum_{t=1}^{n}(X_t - \bar{X})(Y_t - \bar{Y})}{\sqrt{\sum_{t=1}^{n}(X_t - \bar{X})^2 + \sum_{t=1}^{n}(Y_t - \bar{Y})^2}} \quad \ldots(1)$$

While the correlation coefficient is a useful tool for evaluating measurement accuracy, it is important to apply it with certain limitations in mind. In this study, the focus is on examining the relationship between function points, which are treated as stochastic in nature. The correlation between these declared stochastic values is determined based on Equation.

2. **Principal Component Analysis (PCA):**
PCA is employed as a dimensionality reduction tool to transform a set of possibly correlated variables into a smaller number of uncorrelated components. This approach captures the most significant variance in the data while minimizing redundancy, enabling the selection of principal predictors that contribute meaningfully to temperature variability.
By applying these statistical techniques, the study aims to identify a robust subset of predictors that can effectively capture the temperature dynamics for use in statistical downscaling models. This process not only enhances computational efficiency but also improves the reliability and interpretability of model outputs.
In this study, bilinear interpolation was employed as a resampling technique to estimate unknown values within a two-dimensional grid based on the known values of four surrounding points. This method was selected due to its ability to produce smoother and more accurate results than nearest-neighbour interpolation while maintaining computational efficiency. The interpolation process involved determining the value at an intermediate point by performing linear interpolation first in one direction (either x or y), and then again in the other direction. Specifically, for a point located within a rectangular cell defined by four adjacent grid points, the function values at those four corners were used to compute the value at the desired location. This was done by interpolating the values along one axis to get two intermediate values, which were then interpolated along the second axis to obtain the final result. Prior to applying bilinear interpolation, the dataset was pre-processed to ensure that the grid structure was regular and complete, with any missing values addressed through appropriate cleaning techniques. The method was implemented using ArcGIS and was applied to improve the resolution of spatial data and support further analysis by generating a continuous surface from discrete input values.

Result Analysis And Discussion

NALANDA

The scatter plot matrix illustrates the pairwise relationships among key meteorological variables, including Specific Humidity, Maximum and Minimum Temperatures, Mean Sea Level Pressure (mslp), Divergence (div), Geopotential Height (hgt), and wind components at both 1000 mb and 500 mb levels (uwnd, vwnd) as shown in Fig. 2. Diagonal histograms show the distribution of each variable, while the lower triangle scatter plots depict their bivariate associations, overlaid with red ellipses indicating correlation strength and direction.
Notable positive correlations are observed between maximum and minimum temperatures, as well as between temperature variables and specific humidity—consistent with the thermodynamic principle that warmer air can hold more moisture. Conversely, mslp demonstrates a negative correlation with temperature and humidity, aligning with meteorological expectations that low-pressure systems are typically associated with warmer, moister conditions. Geopotential height also shows moderate correlation with temperature and pressure variables.

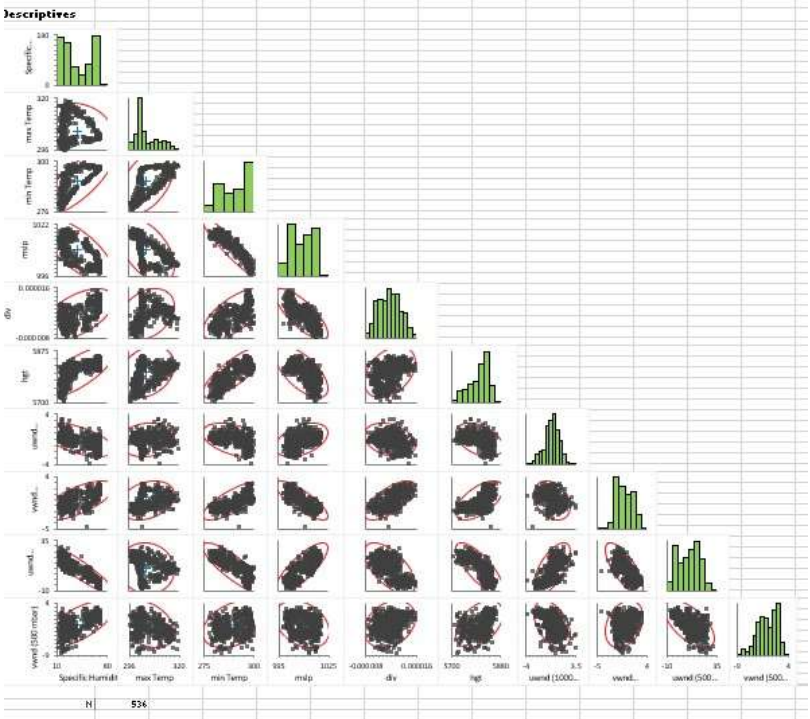

Figure 2: Scatter plots and histograms of Nalanda district

Table 2: Correlation between various predictor variables of Nalanda district

Pearson's r	Sp.Hum.	Tmax	Tmin	mslp	div	hgt	uwnd	vwnd	uwnd	vwnd
Sp.Hum.	-	-0.150	0.776	-0.710	0.554	0.744	-0.588	0.685	-0.920	0.684
Tmax	-0.150	-	0.498	-0.501	0.317	0.274	0.116	0.295	-0.068	-0.345

Tmin	0.776	0.498	-	-0.933	0.674	0.814	-0.433	0.776	-0.847	0.374
Mslp	-0.710	-0.501	-0.933	-	-0.775	-0.648	0.434	-0.816	0.808	-0.317
Div	0.554	0.317	0.674	-0.775	-	0.378	-0.296	0.686	-0.621	0.280
Hgt	0.744	0.274	0.814	-0.648	0.378	-	-0.418	0.599	-0.771	0.446
Uwnd	-0.588	0.116	-0.433	0.434	-0.296	-0.418	-	-0.311	0.636	-0.455
Vwnd	0.685	0.295	0.776	-0.816	0.686	0.599	-0.311	-	-0.738	0.310
Uwnd	-0.920	-0.068	-0.847	0.808	-0.621	-0.771	0.636	-0.738	-	-0.613
Vwnd	0.684	-0.345	0.374	-0.317	0.280	0.446	-0.455	0.310	-0.613	-

From Table 2, after correlating maximum temperature with nine predictor variables (specific humidity, minimum temperature, mean sea level pressure, divergence, geopotential height, u wind500, vwnd500, uwnd1000, vwnd1000), it is clear that maximum temperature is highly correlated (affected) with mean sea level pressure and minimum temperature and least correlated with u wind at 1000 mBar pressure. After correlating minimum temperature with nine predictor variables (specific humidity, maximum temperature, mean sea level pressure, divergence, geopotential height, u wind500, vwnd500, uwnd1000, vwnd1000), it is clear that minimum temperature is highly correlated (affected) with mean sea level pressure and least correlated with v wind at 1000 mBar pressure. The correlation matrix (Table 2) shows a strong positive relationship between Specific Humidity and Tmin ($r = 0.776$) and between humidity and hgt ($r = 0.744$), indicating that higher humidity is often associated with increased minimum temperature and geopotential height. Tmax has a moderate positive correlation with Tmin ($r = 0.498$), while mslp demonstrates a strong negative correlation with Tmin ($r = -0.933$), confirming the inverse pressure-temperature relationship. Wind components at both pressure levels exhibit negative correlations with mslp and Tmin, suggesting the role of dynamic atmospheric processes.

Table 3 presents the results of Principal Component Analysis (PCA) for Nalanda, highlighting the variance explained by each component. The first principal component accounts for 60.5% of the total variance, indicating it captures the most significant patterns in the dataset. Combined with the second component, they explain 78.8% of the total variance. The cumulative proportion reaches 91.5% by the fourth component, suggesting that the first four components sufficiently represent the underlying data structure. Components beyond the sixth contribute minimally, implying that dimensionality can be effectively reduced without significant information loss. This supports efficient data simplification and further multivariate analysis.

Table 3: Principal Component of Nalanda district

Principal Component			
Component	**Variance**	**Proportion**	**Cumulative proportion**
1	6.053	0.605	0.605
2	1.824	0.182	0.788
3	0.6774	0.068	0.855
4	0.5953	0.060	0.915
5	0.3714	0.037	0.952
6	0.2219	0.022	0.974
7	0.1550	0.015	0.990
8	0.06396	0.006	0.996
9	0.03311	0.003	1.000
10	0.004553	0.000	1.000

Table 4: Coefficients of principal component of Nalanda district

	Component									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Sp. Humidity	**0.364**	**0.266**	**0.034**	**-0.147**	**0.222**	**-0.236**	**0.274**	**-0.359**	**0.227**	**-0.643**
max Temp	**0.104**	**-0.661**	**-0.302**	**0.095**	**-0.397**	**0.194**	**-0.033**	**0.093**	**0.188**	**-0.460**
min Temp	**0.382**	**-0.186**	**-0.166**	**-0.072**	**-0.052**	**-0.090**	**0.303**	**-0.225**	**0.522**	**0.601**
mslp	**-0.372**	**0.230**	**-0.098**	**-0.106**	**0.051**	**-0.081**	**-0.449**	**0.181**	**0.734**	**-0.082**

div	**0.299**	**-0.192**	**0.641**	**0.181**	**-0.237**	**-0.464**	**-0.395**	**0.003**	**0.049**	**0.018**
hgt	**0.330**	**0.002**	**-0.549**	**-0.336**	**0.084**	**-0.294**	**-0.535**	**-0.056**	**-0.299**	**0.071**
uwnd	**-0.239**	**-0.310**	**0.266**	**-0.838**	**-0.015**	**-0.145**	**0.190**	**0.120**	**-0.039**	**-0.010**
vwnd	**0.339**	**-0.155**	**0.283**	**-0.152**	**0.467**	**0.656**	**-0.318**	**-0.027**	**0.088**	**0.009**
uwnd	**-0.386**	**-0.140**	**0.039**	**-0.011**	**-0.107**	**0.089**	**-0.217**	**-0.872**	**-0.045**	**0.037**
vwnd	**0.229**	**0.478**	**0.081**	**-0.285**	**-0.703**	**0.363**	**-0.063**	**-0.031**	**-0.006**	**0.009**

Table 4 shows the component loadings of ten meteorological variables across ten principal components for Nalanda. The first component, explaining the highest variance (60.5%), is heavily influenced by minimum temperature (0.382), specific humidity (0.364), and geopotential height (0.330), indicating these variables drive the primary variability in the dataset. The second component is shaped mainly by maximum temperature (–0.661) and v-wind at 500 mb (0.478), reflecting temperature and upper-air wind influence. Divergence strongly loads on the third component (0.641), and u-wind at 1000 mb dominates the fourth (–0.838). This loading matrix helps identify key drivers behind each principal component for further interpretation and modeling.

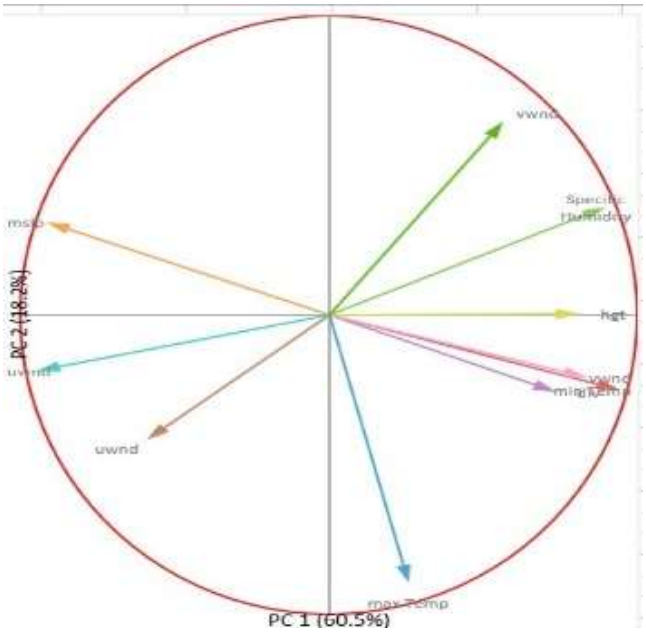

Figure 3: Correlation monoplot of Nalanda district

The Fig. 3 illustrates the contributions of meteorological variables to the first two principal components (PC1 and PC2), which together explain 78.7% of the total variance (60.5% by PC1 and 18.2% by PC2). Variables such as specific humidity, geopotential height (hgt), and minimum temperature align closely with PC1, indicating strong positive contributions to the primary component. In contrast, maximum temperature and u-wind components exhibit negative associations with PC1. The value of PC2 is influenced primarily by mean sea level pressure (mslp) and upper-air wind vectors. The vector lengths reflect the strength of contribution, while angles between them suggest correlations—smaller angles imply stronger positive relationships among variables.

SARAN

Correlation: Specific Humidity, max Temp, min Temp, mslp, div, hgt, uwnd1000, vwnd1000, uwnd500, vwnd500.

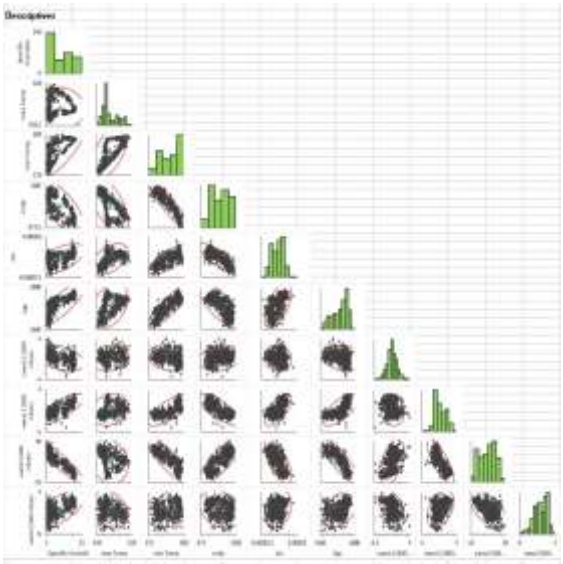

Figure 4: Scatter plots and histograms of Saran district

Table 5: Correlation between various predictor variables of Saran district

Pearson's r	Sp. Hum	max Temp	min Temp	mslp	Div	Hgt	uwnd	vwnd	uwnd	vwnd
Sp.Hum.		-0.040	0.781	-0.724	0.597	0.827	-0.419	0.791	-0.924	0.658
max Temp	-0.040	-	0.576	-0.564	0.469	0.345	0.166	0.202	-0.148	-0.293
min Temp	0.781	0.576	-	-0.930	0.778	0.874	-0.224	0.754	-0.843	0.351
Mslp	-0.724	-0.564	-0.930	-	-0.813	-0.730	0.211	-0.783	0.801	-0.292
Div	0.597	0.469	0.778	-0.813	-	0.592	-0.093	0.674	-0.672	0.248
Hgt	0.827	0.345	0.874	-0.730	0.592	-	-0.299	0.701	-0.840	0.466
Uwnd	-0.419	0.166	-0.224	0.211	-0.093	-0.299	-	-0.170	0.468	-0.376
Vwnd	0.791	0.202	0.754	-0.783	0.674	0.701	-0.170	-	-0.792	0.421
Uwnd	-0.924	-0.148	-0.843	0.801	-0.672	-0.840	0.468	-0.792	-	-0.592
Vwnd	0.658	-0.293	0.351	-0.292	0.248	0.466	-0.376	0.421	-0.592	-

From Table 5, after correlating maximum temperature with nine predictor variables (specific humidity, minimum temperature, mean sea level pressure, divergence, geopotential height, u wind500, vwnd500, uwnd1000, vwnd1000**),** it is clear that maximum temperature is highly correlated (affected) with minimum temperature and mean sea level pressure and least correlated with u wind at 1000 mBar pressure. After correlating minimum temperature with nine predictor variables (specific humidity, maximum temperature, mean sea level pressure, divergence, geopotential height, u wind500, vwnd500, uwnd1000, vwnd1000**),** it is clear that minimum temperature is highly correlated (affected) with mean sea level pressure and least correlated with u wind at 500 mBar pressure. In Saran district (Table 5), patterns are similar: Tmin and Specific Humidity correlate strongly (r = 0.781), while mslp exhibits a strong negative relationship with Tmin (r = -0.930). Geopotential height correlates positively with Tmin (r = 0.874), signifying its influence on temperature dynamics. Wind components show weaker associations compared to Nalanda, though significant negative correlations with uwnd at 1000 mb (r = -0.843) indicate possible influence of low-level circulation patterns. The principal component of Saran district is depicted in Table 6. The coefficients of principal component of Saran district is presented in Table 7. The correlation monoplot of Saran district is shown in Fig. 5.

Table 6: Principal Component of Saran district

Principal Component			
Component	Variance	Proportion	Cumulative proportion
1	6.144	0.614	0.614
2	1.854	0.185	0.800
3	0.7513	0.075	0.875
4	0.4263	0.043	0.918
5	0.3679	0.037	0.954
6	0.2199	0.022	0.976
7	0.1341	0.013	0.990
8	0.06378	0.006	0.996
9	0.03089	0.003	0.999
10	0.008152	0.001	1.000

Table 7: Coefficients of principal component of Saran district

Component										
	1	2	3	4	5	6	7	8	9	10
Sp.Hum	0.365	0.249	-0.136	0.043	-0.223	0.218	-0.268	0.360	-0.258	0.648
max Temp	0.141	-0.621	0.314	-0.356	0.172	-0.282	0.124	-0.109	-0.124	0.465
min Temp	0.382	-0.184	0.071	-0.184	-0.056	0.056	-0.214	0.143	-0.620	-0.570
mslp	-0.368	0.215	-0.055	-0.150	-0.094	0.233	0.552	-0.236	-0.593	0.136
div	0.321	-0.235	-0.103	0.437	0.547	0.481	0.328	0.036	0.010	0.005
hgt	0.360	0.005	0.037	-0.468	-0.389	0.322	0.439	0.113	0.417	-0.124
uwnd	-0.146	-0.410	-0.840	-0.211	-0.073	0.112	-0.166	-0.118	0.026	0.021
vwnd	0.346	-0.001	-0.300	0.377	-0.274	-0.617	0.429	-0.001	-0.088	-0.026
uwnd	-0.382	-0.151	-0.039	-0.047	0.101	-0.113	0.217	0.868	-0.020	-0.060
vwnd	0.217	0.475	-0.255	-0.466	0.606	-0.272	0.055	0.022	0.016	-0.010

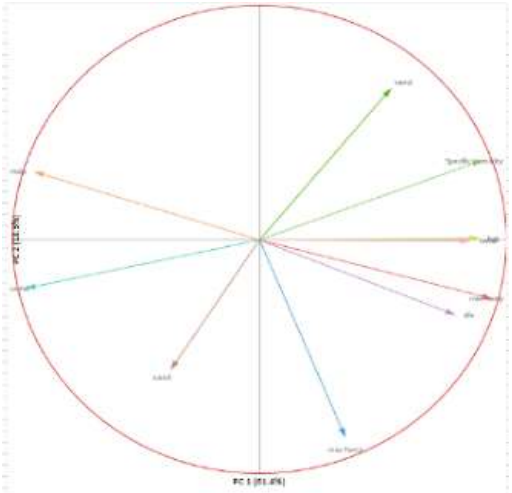

Figure 5: Correlation monoplot of Saran district

BHAGALPUR

Correlation: Specific Humidity, max Temp, min Temp, mslp, div, hgt, uwnd1000, vwnd1000, uwnd500, vwnd500.

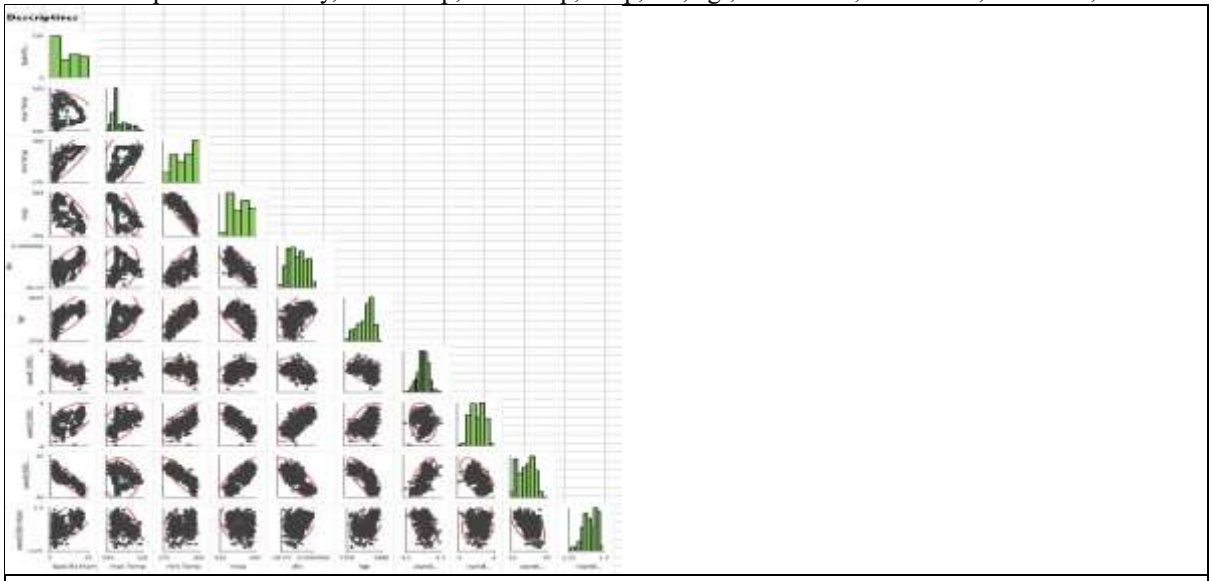

Figure 6: Scatter plots and histograms of Bhagalpur district

Bhagalpur exhibits the strongest correlation between Specific Humidity and Tmin (r = 0.857) among all districts (Table 8). The negative correlation between mslp and Tmin (r = -0.917) is consistent with earlier observations.

Divergence (div) shows a strong positive correlation with Tmin (r = 0.788) and humidity (r = 0.766), indicating that convergence or divergence significantly influences moisture distribution. The scatter plots and histograms of Bhagalpur district is presented in Fig. 6 and Correlation monoplot is presented in Fig. 7.

Table 8: Correlation between various predictor variables of Bhagalpur district

Pearson's r	Sp. Hum	max Temp	min Temp	mslp	div	hgt	uwnd	vwnd	uwnd	vwnd
Sp. Hum	-	-0.065	0.857	-0.761	0.766	0.780	-0.630	0.617	-0.927	0.632
max Temp	-0.065	-	0.431	-0.503	0.231	0.248	0.146	0.469	-0.076	-0.362
min Temp	0.857	0.431	-	-0.917	0.788	0.842	-0.479	0.768	-0.878	0.393
mslp	-0.761	-0.503	-0.917	-	-0.824	-0.647	0.448	-0.839	0.811	-0.288
div	0.766	0.231	0.788	-0.824	-	0.559	-0.420	0.740	-0.777	0.395
hgt	0.780	0.248	0.842	-0.647	0.559	-	-0.453	0.541	-0.782	0.446
uwnd	-0.630	0.146	-0.479	0.448	-0.420	-0.453	-	-0.197	0.667	-0.422
vwnd	0.617	0.469	0.768	-0.839	0.740	0.541	-0.197	-	-0.644	0.155
uwnd	-0.927	-0.076	-0.878	0.811	-0.777	-0.782	0.667	-0.644	-	-0.581
vwnd	0.632	-0.362	0.393	-0.288	0.395	0.446	-0.422	0.155	-0.581	-

From Table 8, after correlating maximum temperature with nine predictor variables (specific humidity, minimum temperature, mean sea level pressure, divergence, geopotential height, u wind500, vwnd500, uwnd1000, vwnd1000**),** it is clear that maximum temperature is highly correlated (affected) with mean sea level pressure and least correlated with specific humidity. After correlating minimum temperature with nine predictor variables (specific humidity, maximum temperature, mean sea level pressure, divergence, geopotential height, u wind500, vwnd500, uwnd1000, vwnd1000**),** it is clear that minimum temperature is highly correlated (affected) with mean sea level pressure and least correlated with v wind at 1000 mBar pressure. For Bhagalpur district the principal component and coefficient of principal component is presented in Table 9 and Table 10, respectively.

Table 9: Principal Component of Bhagalpur district

Principal Component			
Component	**Variance**	**Proportion**	**Cumulative proportion**
1	6.255	0.625	0.625
2	1.849	0.185	0.810
3	0.6141	0.061	0.872
4	0.5331	0.053	0.925
5	0.3187	0.032	0.957
6	0.1953	0.020	0.976
7	0.1250	0.012	0.989
8	0.06661	0.007	0.996
9	0.03350	0.003	0.999
10	0.01049	0.001	1.000

Table 10: Coefficients of principal component of Bhagalpur district

Component										
	1	2	3	4	5	6	7	8	9	10
Sp.Hum	0.371	0.210	-0.072	0.010	-0.281	0.016	-0.302	0.413	0.207	0.655
max Temp	0.105	-0.650	0.226	-0.262	0.499	0.046	0.060	-0.054	0.187	0.393
min Temp	0.384	-0.128	0.044	-0.183	-0.029	0.084	-0.289	0.208	0.535	-0.616
Mslp	-0.366	0.222	-0.017	-0.180	-0.182	0.089	0.436	-0.180	0.707	0.147
Div	0.344	-0.066	-0.254	0.435	0.083	0.685	0.383	-0.007	0.002	-0.010
Hgt	0.330	0.022	0.125	-0.673	-0.348	0.127	0.404	0.020	-0.346	-0.040
Uwnd	-0.240	-0.333	-0.766	-0.289	-0.193	0.167	-0.279	-0.115	-0.047	0.035

Vwnd	0.311	-0.315	-0.297	0.261	-0.215	-0.658	0.396	-0.023	0.102	-0.023
Uwnd	-0.380	-0.139	-0.063	-0.017	0.085	0.008	0.273	0.858	-0.065	-0.106
Vwnd	0.210	0.483	-0.427	-0.267	0.648	-0.188	0.113	0.036	0.011	-0.009

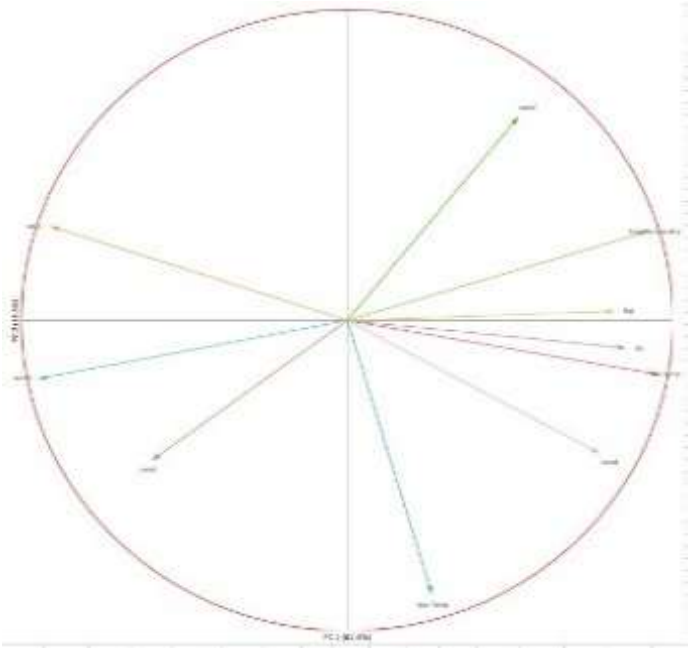

Figure 7: Correlation monoplot of Bhagalpur district

KAIMUR

Correlation: Specific Humidity, max Temp, min Temp, mslp, div, hgt, uwnd1000, vwnd1000, uwnd500, vwnd500. The scatter plots and histograms of Kaimur district is presented in Fig. 8 and Correlation monoplot is presented in Fig. 9.

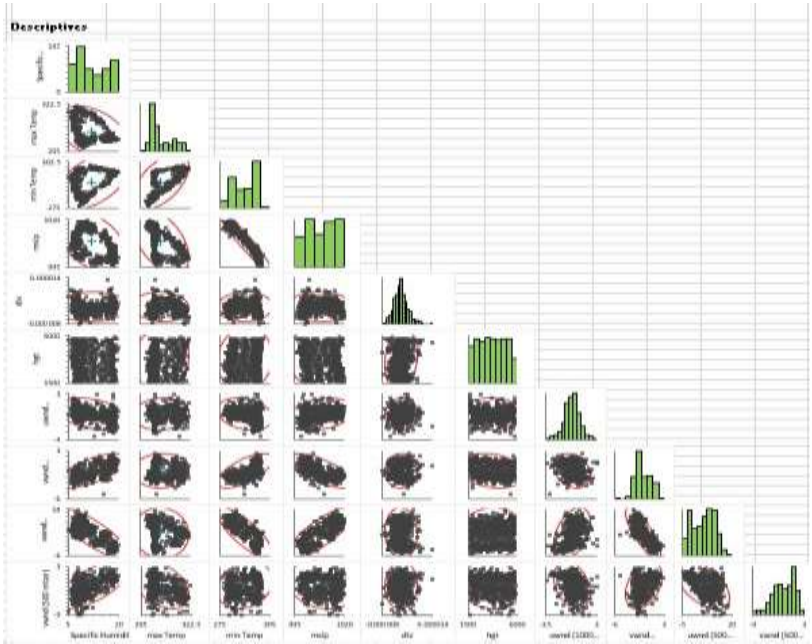

Figure 8: Scatter plots and histograms of Kaimur district

Table 11: Correlation between various predictor variables of Kaimur district

Pearson's r	sp.Hum	max Temp	min Temp	mslp	div	hgt	uwnd	vwnd	uwnd	vwnd
Sp.Hum	-	-0.417	0.488	-0.493	0.048	-0.036	-0.362	0.687	-0.822	0.664
max Temp	-0.417	-	0.555	-0.497	-0.081	0.083	0.073	0.035	-0.051	-0.391
min Temp	0.488	0.555	-	-0.956	0.032	0.080	-0.258	0.674	-0.803	0.221

Mslp	-0.493	-0.497	-0.956	-	-0.037	-0.122	0.277	-0.708	0.802	-0.221
Div	0.048	-0.081	0.032	-0.037	-	0.153	0.040	0.034	-0.059	0.053
Hgt	-0.036	0.083	0.080	-0.122	0.153	-	-0.040	-0.221	0.040	-0.095
Uwnd	-0.362	0.073	-0.258	0.277	0.040	-0.040	-	-0.263	0.420	-0.325
Vwnd	0.687	0.035	0.674	-0.708	0.034	-0.221	-0.263	-	-0.801	0.382
Uwnd	-0.822	-0.051	-0.803	0.802	-0.059	0.040	0.420	-0.801	-	-0.529
Vwnd	0.664	-0.391	0.221	-0.221	0.053	-0.095	-0.325	0.382	-0.529	-

From Table 11, after correlating maximum temperature with nine predictor variables (specific humidity, minimum temperature, mean sea level pressure, divergence, geopotential height, u wind500, vwnd500, uwnd1000, vwnd1000**),** it is clear that maximum temperature is highly correlated (affected) with minimum temperature and least correlated with v wind at 500 mBar pressure. After correlating minimum temperature with nine predictor variables (specific humidity, maximum temperature, mean sea level pressure, divergence, geopotential height, u wind500, vwnd500, uwnd1000, vwnd1000**),** it is clear that minimum temperature is highly correlated (affected) with mean sea level pressure and least correlated with divergence. Table 12 represent the Principal Component of Kaimur district. The coefficients of principal component of Kaimur district is presented in Table 13.

Table 12: Principal Component of Kaimur district

Principal Component			
Component	**Variance**	**Proportion**	**Cumulative proportion**
1	4.355	0.435	0.435
2	2.019	0.202	0.637
3	1.181	0.118	0.755
4	0.9496	0.095	0.850
5	0.7071	0.071	0.921
6	0.4523	0.045	0.966
7	0.1952	0.020	0.986
8	0.06784	0.007	0.993
9	0.05537	0.006	0.998
10	0.01758	0.002	1.000

Table 13: Coefficients of principal component of Kaimur district

Component										
	1	2	3	**4**	5	6	7	8	9	10
Sp.Hum	0.389	0.338	-0.052	0.013	-0.233	0.203	-0.391	0.231	0.497	-0.427
max Temp	0.056	-0.663	0.099	-0.020	0.162	-0.339	0.060	-0.209	0.380	-0.463
min Temp	0.412	-0.336	-0.009	0.027	-0.040	-0.107	-0.230	0.168	0.298	0.731
Mslp	-0.416	0.317	0.040	-0.008	0.060	0.013	0.020	-0.521	0.618	0.260
Div	0.026	0.041	-0.648	0.613	0.445	-0.028	-0.030	0.014	0.030	-0.031
Hgt	-0.014	-0.165	-0.727	-0.431	-0.427	0.094	0.230	-0.108	0.055	0.004
Uwnd	-0.222	-0.145	0.072	0.614	-0.724	-0.127	-0.050	-0.057	-0.042	-0.007
Vwnd	0.410	0.025	0.175	0.238	-0.032	0.369	0.759	-0.073	0.156	0.018
Uwnd	-0.462	-0.035	0.004	-0.024	0.041	-0.065	0.296	0.763	0.329	0.022
Vwnd	0.267	0.425	-0.046	-0.048	-0.089	-0.816	0.263	0.016	-0.007	0.004

WEST CHAMPARAN

Correlation: Specific Humidity, max Temp, min Temp, mslp, div, hgt, uwnd1000, vwnd1000, uwnd500, vwnd500. Table 14 reveals strong positive correlations between Specific Humidity and Tmin (r = 0.829) and between humidity and vwnd (r = 0.840). The negative correlation between mslp and Tmin (r = -0.909) is consistent with synoptic-scale pressure-temperature interactions. Unlike other districts, wind variables here display stronger interrelationships, implying higher sensitivity of this region to circulation patterns due to its northwestern location and proximity to Himalayan foothills.

PC1 explains 62.5% variance (Table 15) and is heavily loaded on Tmin, humidity, and hgt (Table 16). PC2 (14%) captures Tmax and vwnd, while PC3 emphasizes divergence, reflecting complex interplay between thermodynamic and circulation patterns in this flood-prone region.

The scatter plots and histograms of West Champaran district is presented in Fig. 10 and Correlation monoplot is presented in Fig. 11.

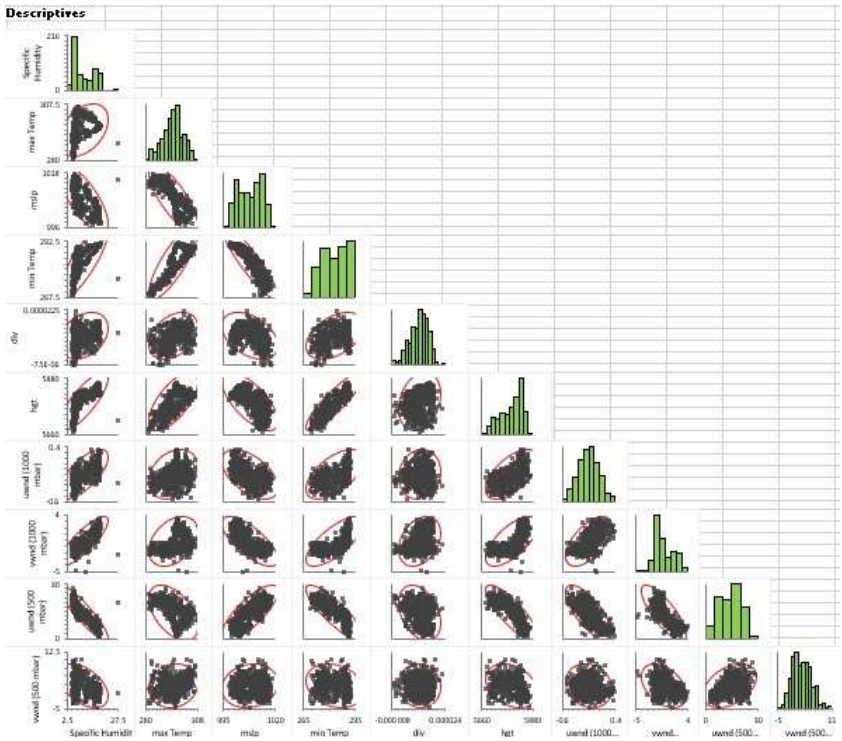

Figure 9: Scatter plots and histograms of West Champaran district

From Table 14, after correlating maximum temperature with nine predictor variables (specific humidity, minimum temperature, mean sea level pressure, divergence, geopotential height, u wind500, vwnd500, uwnd1000, vwnd1000**),** it is clear that maximum temperature is highly correlated (affected) with minimum temperature.

Table 14: Correlation between various predictor variables of West Champaran district

Pearson's r	Sp.Hum	max Temp	mslp	min Temp	div	hgt	uwnd	vwnd	uwnd	vwnd
Sp.Hum	-	0.399	-0.745	0.829	0.327	0.801	0.698	0.840	-0.908	-0.480
max Temp	0.399	-	-0.780	0.823	0.369	0.673	0.280	0.376	-0.508	0.120
Mslp	-0.745	-0.780	-	-0.909	-0.419	-0.686	-0.545	-0.733	0.784	0.139
min Temp	0.829	0.823	-0.909	-	0.446	0.885	0.589	0.737	-0.859	-0.229
Div	0.327	0.369	-0.419	0.446	-	0.260	0.113	0.326	-0.362	-0.069
Hgt	0.801	0.673	-0.686	0.885	0.260	-	0.604	0.673	-0.824	-0.307
Uwnd	0.698	0.280	-0.545	0.589	0.113	0.604	-	0.596	-0.744	-0.160
Vwnd	0.840	0.376	-0.733	0.737	0.326	0.673	0.596	-	-0.797	-0.419
Uwnd	-0.908	-0.508	0.784	-0.859	-0.362	-0.824	-0.744	-0.797	-	0.417
Vwnd	-0.480	0.120	0.139	-0.229	-0.069	-0.307	-0.160	-0.419	0.417	-

Table 15: Principal Component of West Champaran district

Principal Components			
Component	Variance	Proportion	Cumulative proportion
1	6.250	0.625	0.625
2	1.405	0.140	0.765
3	0.8918	0.089	0.855
4	0.6162	0.062	0.916
5	0.3507	0.035	0.951
6	0.2263	0.023	0.974
7	0.1389	0.014	0.988
8	0.07907	0.008	0.996
9	0.03067	0.003	0.999
10	0.01232	0.001	1.000

After correlating minimum temperature with nine predictor variables (specific humidity, maximum temperature, mean sea level pressure, divergence, geopotential height, u wind500, vwnd500, uwnd1000, vwnd1000**),** it is clear that minimum temperature is highly correlated (affected) with mean sea level pressure and least correlated with v wind at 1000 mBar pressure. Principal Component of West Champaran district is presented in Table 15 while the Coefficients of principal component is depicted in Table 16.

Table 16: Coefficients of principal component of West Champaran district

Component										
	1	2	3	4	5	6	7	8	9	10
Sp.hum	-0.368	0.234	0.004	-0.080	0.047	-0.206	0.531	-0.507	0.302	0.360
max Temp	-0.272	-0.539	-0.068	0.391	-0.077	0.216	-0.294	0.021	0.400	0.422
mslp	0.357	0.224	0.011	-0.049	-0.454	-0.430	-0.220	0.092	0.609	-0.035
min Temp	-0.385	-0.176	-0.007	0.162	-0.044	-0.001	0.090	-0.158	0.279	-0.826
div	-0.176	-0.289	0.779	-0.454	-0.242	-0.028	-0.085	-0.039	-0.064	0.037
hgt	-0.355	-0.012	-0.149	0.271	-0.458	-0.526	-0.137	-0.028	-0.518	0.063
uwnd	-0.284	0.181	-0.445	-0.555	-0.296	0.377	-0.350	-0.168	0.008	0.001
vwnd	-0.340	0.203	0.067	-0.124	0.627	-0.384	-0.506	0.121	0.109	0.016
uwnd	0.377	-0.140	0.038	0.098	0.118	-0.060	-0.359	-0.813	-0.126	-0.060
vwnd	0.143	-0.629	-0.404	-0.446	0.145	-0.394	0.203	0.061	0.013	-0.009

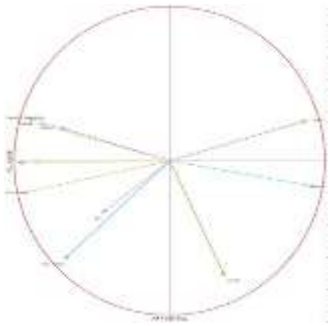

Figure 10: Correlation monoplot of West Champaran district

Conclusion

This Project deals with the study of "predictor selection for temperature using different statistical techniques". It comprises of basic knowledge about climate change, surface temperature, rainfall, extreme weather events, rise in sea level, impacts of climate change, downscaling- need, categories, advantages and disadvantages, GCMs. For better understanding of predictor variables and methodologies, references from various researchers have been taken which are being mentioned. The "study areas" selected for the study are "Nalanda, Saran, Bhagalpur, Kaimur and West Champaran".

Ten types of predictor variables are selected which are: - "Specific humidity, maximum temperature, minimum temperature, mean sea level pressure, divergence, geopotential height, u wind at 500 mBar, v wind at 500 mBar, u wind at 1000 mBar, v wind at 1000 mBar". Data of various predictor variables are collected for each study area at four values of latitude and longitude. Data are collected from "NCEP/NCAR Reanalysis" website in netCDF format which is converted to table format using "ArcMap of ArcGIS" software. After that, data of all predictor variables are made precise to the exact latitude and longitude of each study area using "bilinear interpolation" in MS Excel. Interpolated data are used to find the "correlation matrix" and to draw the "scatter plots" and "histograms" between various predictor variables. Further, mean, variance of data are found and they are used to find the relationship between various predictor variables using principal component analysis (PCA). After PCA, correlation monoplot is drawn which helps to analyse whether the two variables are, positively correlated, negatively correlated or not correlated. The correlation between ten principal components and ten predictor variables is mentioned in the table of coefficients of PCA. The same procedure is adopted for each study area and analysis using PCA is done individually.

From the study, the conclusion obtained is that the predictor variables affect each other in a different way. No predictor variable affect all study area in a mutual way but the major predictor variable which affects temperature in all the districts is “mean sea level pressure”. The correlation of mean sea level pressure with maximum temperature of Nalanda, Saran, Bhagalpur, Kaimur and West Champaran are:- 0.501, 0.564, 0.503, 0.497, 0.780 respectively. The correlation of mean sea level pressure with minimum temperature of Nalanda, Saran, Bhagalpur, Kaimur and West Champaran are:- 0.933, 0.930, 0.917, 0.956, 0.909 respectively. These values are high and near about one which indicate high dependence of temperature with mean sea level pressure. Thus, due to mean sea level pressure variations, there is a high chance of climate change which may cause flood and drought in certain regions of Bihar. This study can help meteorologists to use the correlation between various predictor variables to predict the rainfall and climate change in the specific areas (study areas) taken in the project.

Reference

1. Anandhi, Aavudai, et al. "Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine." International Journal of Climatology 28.3 (2008): 401-420.
2. Vimont, Daniel J., David S. Battisti, and Rosamond L. Naylor. "Downscaling Indonesian precipitation using large-scale meteorological fields." International journal of climatology 30.11 (2010): 1706-1722.
3. Kannan, S., and Subimal Ghosh. "Prediction of daily rainfall state in a river basin using statistical downsclaing from GCM output." Stochastic Environmental Research and Risk Assessment 25 (2011): 457-474.

4. Timbal, Bertrand, et al. Understanding the anthropogenic nature of the observed rainfall decline across South Eastern Australia. Centre for Australian Weather and Climate Research, 2010.
5. Singh, Dharmaveer, Sanjay K. Jain, and R. D. Gupta. "Statistical downscaling and projection of future temperature and precipitation change in middle catchment of Sutlej River Basin, India." Journal of Earth System Science 124 (2015): 843-860.
6. Tisseuil, Clement, et al. "Statistical downscaling of river flows." Journal of Hydrology 385.1-4 (2010): 279-291.
7. Chen, Shien-Tsung, Pao-Shan Yu, and Yi-Hsuan Tang. "Statistical downscaling of daily precipitation using support vector machines and multivariate analysis." Journal of hydrology 385.1-4 (2010): 13-22.
8. Saraf, Vidya R., and Dattatray G. Regulwar. "Assessment of climate change for precipitation and temperature using statistical downscaling methods in Upper Godavari River Basin, India." Journal of Water Resource and Protection 8.1 (2016): 31-45
9. Pichuka, Subbarao, et al. "Development of a method to identify change in the pattern of extreme streamflow events in future climate: Application on the Bhadra reservoir inflow in India." Journal of Hydrology: Regional Studies 9 (2017): 236-246
10. Alizadeh, Mohamad Javad, et al. "A distributed wind downscaling technique for wave climate modeling under future scenarios." Ocean Modelling 145 (2020): 101513.
11. Misra, Saptarshi, Sudeshna Sarkar, and Pabitra Mitra. "Statistical downscaling of precipitation using long short-term memory recurrent neural networks." Theoretical and applied climatology 134 (2018): 1179-1196.
12. Misra, Saptarshi, Sudeshna Sarkar, and Pabitra Mitra. "Statistical downscaling of precipitation using long short-term memory recurrent neural networks." Theoretical and applied climatology 134 (2018): 1179-1196.
13. Sachindra, D. A., et al. "Statistical downscaling of precipitation using machine learning techniques." Atmospheric research 212 (2018): 240-258
14. Vyshnevskyi, Viktor I., and Olena A. Donich. "Climate change in the Ukrainian Carpathians and its possible impact on river runoff." Acta Hydrologica Slovaca 22.1 (2021): 3-14.
15. Zhang, Q., et al. "A novel statistical downscaling approach for analyzing daily precipitation and extremes under the impact of climate change: Application to an arid region." Journal of Hydrology 615 (2022): 128730.
16. N. Aronszajn, G. Gagliardo, Interpolation spaces and interpolation methods, Ann. Mat. Pura Appl. 68 (1965) 51–118.
17. J. Meng and Y. Yang, “Symmetrical Two-Dimensional PCA with Image Measures in Face Recognition,” Int J Adv Robotic Sy, Vol. 9, 2012.
18. Z. Haiyang, “A Comparison of PCA and 2DPCA in Face Recognition,” in Electrical Power Systems and Computers: Springer, Vol. 99, 2011, pp. 445-449.
19. SDSM (Statistical DownScaling Model) by University of Loughborough http://copublic.lboro.ac.uk/cocwd/SDSM/sdsmmain.html.
20. https://nihroorkee.gov.in/sites/default/files/uploadfiles/Downscaling-Climate-ChangeStudies.pdf.
21. Fowler, H. J., and Wilby, R. L. (2007). Beyond the downscaling comparison study. International Journal of Climatology, 27: 1543–1545 (2007).
22. Teutschbein, Claudia, Fredrik Wetterhall, Jan Seibert (2011). Evaluation of different downscaling techniques for hydrological climate-change impact studies at the catchment scale. Climate Dynamics, DOI 10.1007/s00382-010-0979-8.
23. Abbasi, Daniel R. (2005) “Americans and Climate Change, A Summary of Insights and Recommendations from the 2005 Yale University Forestry and Environmental Science Conference on Climate Change.”
24. Singer, S. Fred, and Avery, Dennis T. (2007) Unstoppable Global Warming-Every 1,500 Years Lanham, Maryland: Rowman and Littlefield Publishing Group, Inc.
25. Sneider, C., Golden, R., Gaylen, F. (2007) Global Systems Science- Climate Change. Berkeley: The Regents of the University of California.
26. Abbass, K., Qasim, M.Z., Song, H. et al. A review of the global climate change impacts, adaptation, and sustainable mitigation measures. Environ Sci Pollut Res **29**, 42539–42559 (2022). https://doi.org/10.1007/s11356-022-19718-6