A Multi-Prefecture Study Applying Multivariate Approaches for Predicting and Demystifying Weather Data Variations Affect COVID-19 Spread

Naveed Anwer Butt[1] , Huda Gull[1], Zulfqar Ali[2], Ghulam Muhammad[3], Salman A. AlQahtani[3]

[1]Department of Computer Science, Faculty of Computing and InformationTechnology, University of Gujrat, Gujrat,Pakistan.

[2]School of Computer Science and Electronic Engineering, University of Essex, United Kindgdom.

[3]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia.

Abstract

Since the first cases were reported in Wuhan, China in December 2019, the massive SARS virus, known as COVID-19, is spreading at an alarming rate and is endangering the world. The whole world is now affected by this terrible epidemic including Pakistan. Because it provides a rich resource for studying the variants affecting the COVID-19 contagion, the spread of COVID-19 in Pakistan makes it necessary to assess the link between COVID-19 cases and weather parameters for management and policymakers. The proposed study uses a multivariate approach to demystify the impact of meteorological conditions on COVID-19 contagion, where as a case study, different regions of Pakistan are chosen. The multivariate techniques used in the proposed approach comprises linear multiple regression, linear stepwise regression, multiple adaptive regression splines, and loess regression. We also implement spline regression models for deeper analysis. A machine learning regressor was used to validate the results of the spline curve regression model. The performance of machine learning models and splines is measured using different performance metrics. Experiments show that weather parameters such as temperature and humidity are prominent features in predicting the spread of COVID-19. The spline curve results show that, except for Baluchistan, during the first wave, the temperature was positively correlated with daily confirmed cases, and during the second wave, the temperature was negatively correlated with daily confirmed cases in all other provinces. Another finding of the study is that when data from March 2020 to February 2021 were included, the humidity was inversely associated with mortality and confirmed cases in all provinces. Thus, increased humidity inhibits the spread of COVID-19 and reduces mortality.

Keywords: COVID-19; Multivariate Approaches; Spline curve; Machine Learning; Weather variants

1. Introduction

These days pandemics are a worldwide concern. In Wuhan, China, at the end of December 2019, a large number of cases of pneumonia, later dubbed COVID-19 (Corona Virus Disease-19), arise for unknown reasons [1]. China, Hubei, Spain, America, and Italy were among the major areas afflicted by COVID-19. This worrisome pandemic now affects the entire world. From January to February 2020, the majority of the area saw the same average temperature range of 5°C to 11°C, with humidity levels ranging from 47 percent to 79 percent. The belief is that the high temperature will inhibit the spread of COVID-19 [2]. Some of the case studies revealed a link between the COVID-19 instances and different types of meteorological data. Several of the researchers were skeptical of the concept. The understanding was necessary to identify the reality underlying the variation of meteorological data that influenced COVID-19. This will directly affect the management people, policymakers, and stakeholders [3].

The COVID-19 arrived in Pakistan at the end of February, along with the sick individuals who had returned from Iran. Thousands of Pakistanis visit Iran's sacred sites each year. On February 23, 2020, Pakistan's authorities blocked the border with Iran to prevent further spread [4]. In Pakistan, COVID-19 is continuously spreading, with 581K confirmed cases and 22K active cases as of February 28, 2021. As a result, most past research has concentrated on identifying the link between COVID-19 and other factors such as control, prevention, causes, and diagnosis [5] [6]. Various examinations of the COVID-19 incidents focused on determining the pattern of climatic conditions. The bulk of the studies were attempting to determine if high temperatures influenced the spread of COVID-19 infections.

1.1 The role of the environment to the spread of COVID-19

Environmental variables play a role in the epidemiology of many infectious diseases [7]. COVID-19 instances and the temperature had a positive association, according to the correlation study. According to [8], many parameters such as humidity, temperature, and population density affect viral transmission. Environmental factors have also been implicated in the SARS outbreak [9]. An association between environmental variables and COVID-19 was found through epidemiological studies, but the conclusions were not clear [10]. According to the literature, climate change is associated with the emergence and spread of infectious diseases [11] [12]. Viral infections like the flu rage in dry or cold climates. Due to the mild season, the SARS pandemic was brought under control and ended in July 2003. COVID-19 occurs mainly in countries located in low-temperature regions [13]. Low average annual temperatures (37-63°F or 3-17°C) are typical of the spread of COVID-19. Chinese researchers have linked humidity, temperature, and the COVID-19 outbreak.

1.1.1. The effect of temperature and humidity

In the warm and humid regions of China, the transmission rate of COVID-19 is low. According to researchers in Finland and Spain, 95% of global infections occur in temperatures between 2 and 10 degrees Celsius and in dry regions. Outbreaks of COVID-19 have also been observed in hot and humid Malaysia, Indonesia, and Singapore [14]. Environmental stability is influenced by meteorological variables, which in turn control the viability of the virus. The spread of COVID-19 is strongly influenced by humidity and temperature [15]. The spread of COVID-19 in China is affected by relative humidity and temperature. According to [16], temperature, humidity, and respiratory viruses are related. Death rates from COVID-19 are also affected by changes in temperature and humidity [17]. According to [18], fluctuations in average temperatures can have a significant impact on COVID-19. However, according to [19], the ambient temperature may be a key factor in the spread of COVID-19.

Using the case study of Pakistan, several studies have found a correlation between climate variables such as temperature and humidity and the spread of COVID-19. Humera et. al. used time-series models to assess the impact of humidity, temperature, and other weather conditions on the spread of COVID-19 by determining correlations between the total number of cases, deaths, and weather variables in a specific area [20]. The results showed that meteorological conditions had a greater impact on the total number of cases and deaths than other characteristics such as community, age, and total population. Therefore, temperature and humidity are important criteria for predicting cases affected by COVID-19. Additionally, higher temperatures are associated with lower mortality from COVID-19 infection [21].

1.1.2. The effect of wind speed and air pollution

The impact of wind speed and air pollution on COVID-19 was investigated in Pakistan from March 10, 2020, to October 4, 2020. In Pakistan and its provinces, the wind speed was positively correlated with COVID-19, and PM2.5 was also positively correlated with COVID-19. However, there was a significant negative association in Sindh. Another study using meteorological variables (temperature, precipitation, and humidity) made recommendations for the total number of COVID-19 cases in Pakistan [22]. According to correlation studies, COVID-19 cases are positively correlated with temperature. It means that a spike in COVID-19 cases was observed in Pakistan, its provinces, and administrative units, because of rising temperatures from March 10, 2020, to August 25, 2020 [23].

Some of these studies found a significant positive correlation between daily temperature range (DTR) and COVID-19 mortality, while a significant negative correlation between COVID-19 mortality and temperature and absolute humidity relationship. A study found that only the average temperature was associated with the spread of COVID-19 [24]. However, other studies showed that temperature and humidity are not associated with the spread of COVID-19 [25]. This logical difference underscores the need for further investigation of the situation.

1.2. Motivation

The premise of the proposed study is that weather variations influence the dissemination of COVID-19. To determine the affiliation of the epidemiological data of multi prefecture of Pakistan combined with weather variants. For the trials, multivariate techniques are employed to determine the impact of weather variations on the spread of COVID-19. Multivariate techniques are statistical models that predict outcomes using many factors. Therefore, five ecologically distinct provinces are selected to investigate whether there is a link between climate change and the spread of COVID-19. Islamabad has a muggy subtropical climate, but Karachi's summers are hot and humid, influenced by cool breezes from the Arabian Sea. The climate of Punjab is semi-arid. Khyber Pakhtunkhwa (KPK) has a semi-arid climate with little rainfall throughout the year.

1.3. Contribution

The average temperature, humidity, wind speed, and dew point are the weather variables. The previous scholars' work was done by establishing a link between the two variables. The multivariate techniques employed in the proposed work utilized numerous independent factors to predict the one dependent variable. Linear Regression and Non-Linear Regression are two multivariate techniques. Both regressions are employed in the proposed study to determine which strategy best fits and has a minimal error for the COVID-19 and weather parameters datasets. The results of the proposed research are drawn using Multiple Linear Regression, Linear Forward stepwise Regression, Linear Backward stepwise Regression, Linear Both Regression, Spline Regression, MARS (Multivariate Adaptive Regression Splines), Loess Regression, and Loess Regression with direct surface models. The outcomes of the proposed model prediction are validated using several machine-learning models such as regression models and Spline

Regression model. Root mean square error (RMSE) assessment measures are used to compare the results of models to determine which model performs better.

The paper is organized as follows: Section 1 contains the introduction, section 2 describes methods and materials, section 3 provides the experimental results, and section 4 contains the discussion. The conclusion and future work for the planned study are included in the last section.

2. Methods and Materials

This section sheds a light on the architecture of the proposed scheme as shown in Figure 1. The details of the predictive model are discussed as follows.

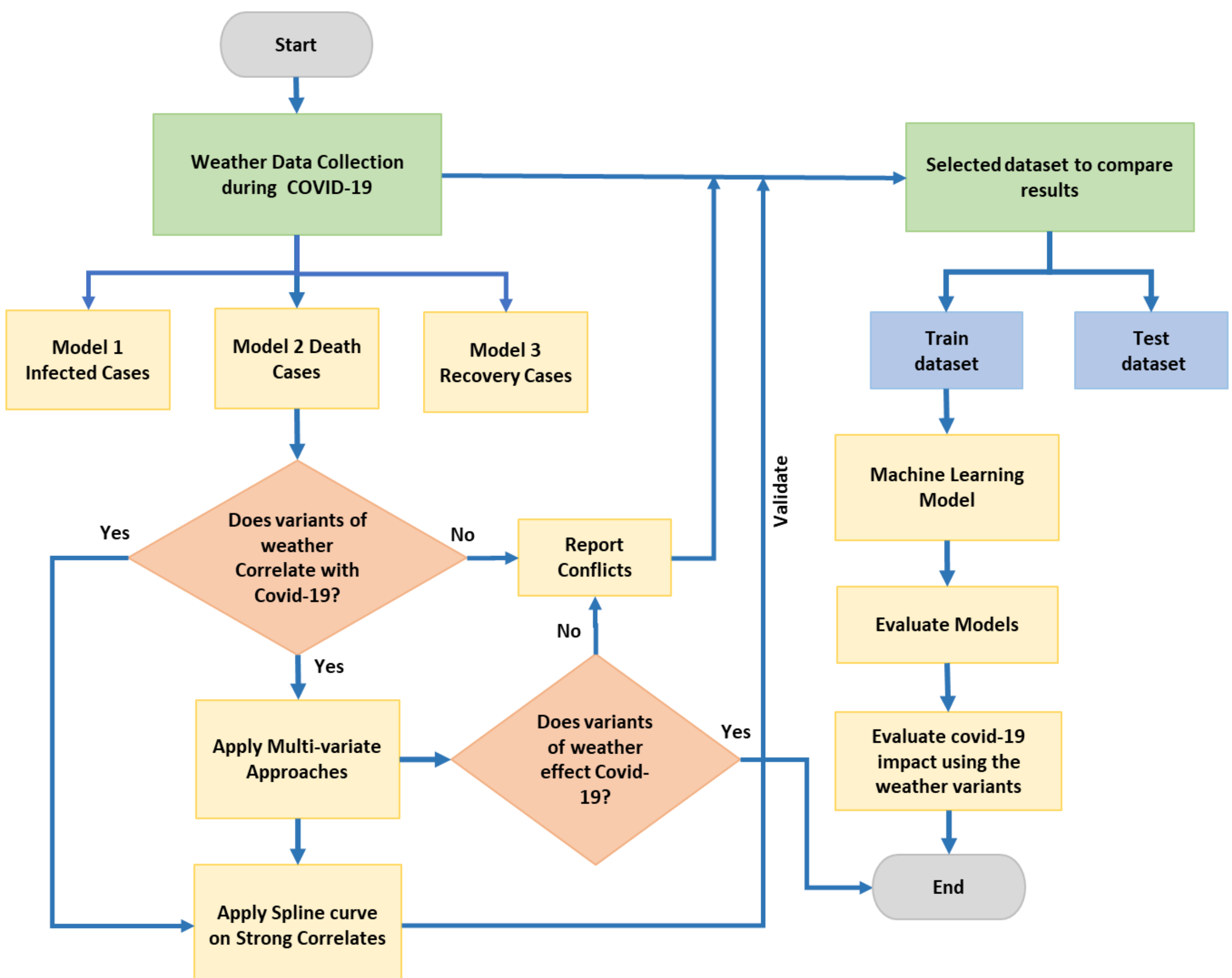


Figure 1: Schematic presentation of the proposed methodology for Predicting and Demystifying Weather Data Variations Affect COVID-19 Spread

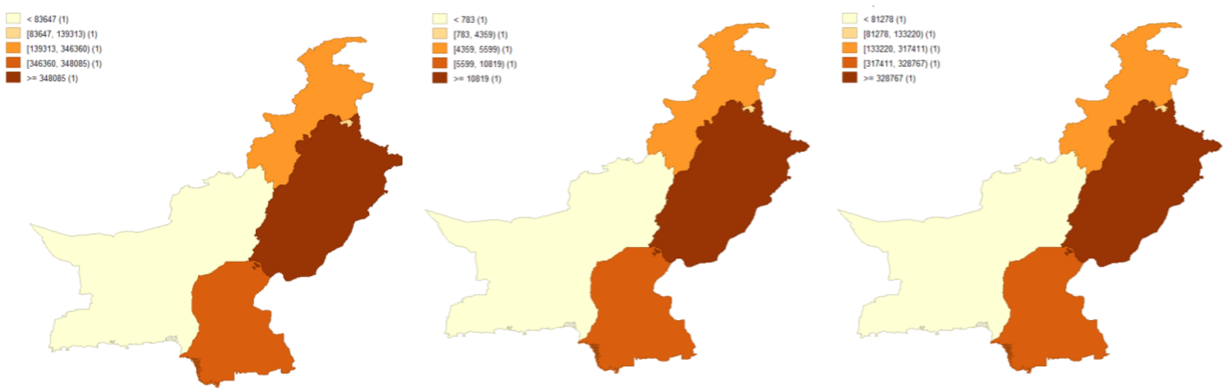
2.1 Data collection

A data analysis regarding a particular system serves as the foundation for a data-driven model. Finding correlations between system state variables (input and output) without knowing the system's physical phenomena explicitly is the basic idea behind data-driven models [26].

Data quality is a crucial aspect when working with ML models. The process of optimizing data-driven approaches is based on patterns that algorithms can "learn" from data. Ensuring the high quality of data is one of the biggest issues in data collection. The quality of data can be affected by issues such as duplication, inaccuracy, inconsistent formatting, and incompleteness. The limited extrapolate by of data-driven models is a serious problem. Data-driven models can only be valid in regions where they have enough dense coverage of training data points, excepting significant assumptions on the learned function [27]. However, there could be chance of little variation in the dataset, so we can assume it will not skew our results.

According to the World Health Organization (WHO), a variety of environmental factors can influence the spread of COVID-19. Food, water, and weather are just some of the factors that play a role in the spread of COVID-19. Meteorological data (temperature, humidity, wind speed, dew point) have an impact on the number of infected cases [2]. Pakistan's Punjab, Sindh, Baluchistan, KPK, and Islamabad provinces are included, which encompass the whole mainland. Epidemiological and weather change data for March 2020-February 2021 data in Pakistani provinces are collected. The dataset of COVID-19 cases is obtained from the official website of the Government of Pakistan (https://COVID.gov.pk/stats/Pakistan, accessed February 2021). The COVID-19 cases dataset is arranged here by province. We collect three variables of "Total Confirmed Cases", "Total Deaths" and "Total Recovery" for a given date and calculate “daily new cases”, "daily new deaths" and "daily recoveries" based on these attributes.

Weather data is sourced from the Weather Underground database (https://www.wunderground.com, accessed March 2021), which contains historical weather data such as temperature, wind speed, humidity, and dew point, as well as minimum, maximum, and average values. To create the weather datasets for each province, the datasets for the major available stations were taken from the above databases daily. The stations are Badin, Karachi, Islamabad, Sargodha, Murree, Lahore, Multan, Sialkot, Jhelum, DIR, and Peshawar. Assigned each station under the designated province. In the collected dataset, each weather parameter has three parameters (min, max, and mean). For the proposed method, the mean value is considered. Average station data is calculated to approximate provincial weather. To handle null values, we take the average of adjacent values and assign it to the given value. Figure 2 shows a heat map of COVID-19 total confirmed, death, and recovered cases for all the provinces of Pakistan on Feb 28, 2021. The graphical representation of COVID-19 daily confirms, death, and recovery cases with weather parameters temperature and humidity are shown in Figure 3.



a) COVID-19 Confirmed cases in Pakistan by provinces till Feb 28, 2021

b) COVID-19 Death cases in Pakistan by provinces till Feb 28, 2021

c) COVID-19 Recovery cases in Pakistan by provinces till Feb 28, 2021

Figure 2: Heat map of COVID-19 total confirmed, death, and recovery cases for all the provinces of Pakistan till Feb 28, 2021.

2.2 Methodology

2.2.1. Machine Learning: Multivariate Approaches

A variety of models can be used on a fitted line plot to determine the pattern between two variables, including asymptotic regression, exponential, power (convex), concave 1, cubic, spline curve, nonlinear regression, and linear regression [28]. The proposed method uses a fitted model or multivariate approach to find the relationship between COVID-19 cases and weather parameters. Additionally, COVID-19 contagion is predicted based on variations in weather data. Most models are based on parametric fitting using local data to each specific region. Linear regression and nonlinear regression are two multivariate techniques. Two regressions are used in the proposed model to determine which strategy works best with the least error for the COVID-19 and weather parameter datasets.

The presented findings were constructed using multiple linear regression, linear forward stepwise regression, and linear backward stepwise regression Linear Both Regression, Spline Regression, MARS, Loess Regression, and Loess Regression with direct surface models.

Multiple Linear Regression

Multiple Linear Regression is a statistical model that uses two or more independent variables to predict the value of a dependent variable. This model is used to determine the variance in a data collection, and each independent variable contributes to the model's variance (total variance). There are two types of multiple regression: linear regression and non-linear regression. The following is a representation of the multiple linear regression equation as per the proposed analysis:

$$COVID\text{-}19\ cases = a + T \times m_1 + H \times m_2 + WS \times m_3 + DP \times m_4 \qquad (1)$$

In the above equation, the COVID-19 (confirm, death and recovery) cases are dependent variables and weather variants are independent variables. *T*, *H*, *WS*, and *DP* stand for temperature, humidity, wind speed, and dew point, respectively; 'a' represents the intercept, and m_1, m_2, m_3, and m_4 are the slopes of regression.

Linear Backward and Forward Stepwise Regression

The model iterates to find the most significant independent variable in stepwise regression. The linear forward regression model starts with an empty model and takes the most significant variable first, then the next, and so on until all variables have been considered. The least P-value, the drop-in residual sum of the square, and a big increase in R2 define the variable's importance. The threshold can be set with a preset value (P-value) to halt the model, however, this is rarely done.

2.2.2 Spline Regression Model

A segmented polynomial function is used in the spline regression model (SRM). This is utilized to keep the smoothness of the subsequent connecting points [29]. SRMs are important segmented polynomial functions that are employed for their ease of implementation, versatility, and assessing accuracy. With the basic fitting model, SRM helps to smooth the complicated curve. The polynomial regression is used for estimating to decide the degree of the better fitting model for the complex model. Also, evaluate which model is fit for the available dataset. Estimating the degree of a polynomial for the fitting model is extremely important; if the degree of a polynomial is large, there is a risk of overfitting; if the degree of a polynomial is little, there is a risk of information loss. In short, the SRM is an important approach for fitting data smoothly, although it can sometimes overfit or become wiggly. But the fact is that each polynomial function can be written as a simple linear function in the regression splines [30]. The independent variable range can be divided into portions known as knots, which divide the spline curve into parts that specify the end of the first spline curve and the start of the next curve part. In the end, the splines curve (SC) achieves the smoothen and fit on the given data. Its limitation is to make the edge parts to be linear.

The SRM transforms the 'd' degree-polynomial function, $f(x)$, into a simple linear basis function (spline), $b_i(x_i)$.

$$f(x) = y_i = \alpha + \beta_1 b_1(x_1) + \beta_2 b_2(x_2) + \ldots + \beta_d b_d(x_d) + \varepsilon_i \quad (2)$$

To create the spline model there is a need to define the knots to make the curve as smooth as it can be. There are two available options to define the knots in SC. One is to position and number of knots and the other is to define the degree of the polynomial to fit the curve on the data nodes. For the proposed methodology the implementation of the Spline Curve is used as the Splines2 package in R, where the knots and degree of polynomial parameters are tuned to get the desired results. The degree of freedom can be assigned for tracing the smooth curve that can be [1, nx], where nx specifies the number of x unique values. Lambda value can be defined, the λ can be computed as

$$\lambda = r * 256^{3*(spar-1)} \quad (4)$$

Where

$$r = tr\frac{(X'\,W\,X)}{tr(\Sigma)} \quad (5)$$

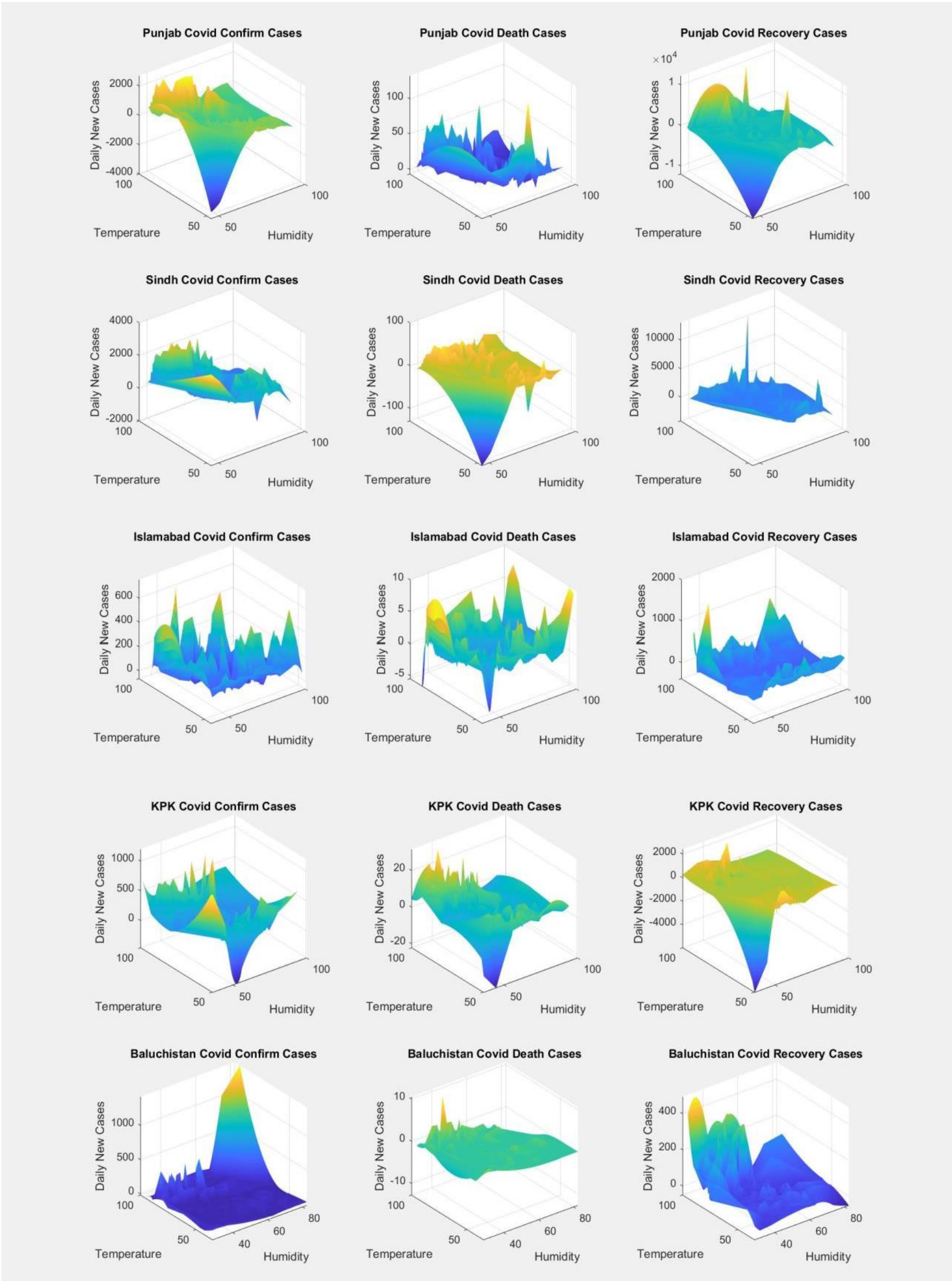


Figure 3: Graphical representation of COVID-19 daily confirm, death, and recovery cases with weather parameters temperature and humidity of Pakistan all Provinces.

$\sum$ = defined as a Matrix

$$\sum[i\,,j] = \ Integeral \frac{B''[i](t)}{B''[j](t)} \ dt \ \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6}$$

X is given as

$$X[i,j] = \ b_j(x_i) \qquad \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{7}$$

W is the diagonal matrix that assigned the weights. Spar is the function in the splines that is used to smoothen the curve which value is in [0,1] when its value is not defined then λ coefficient integral of the second square derivates in the fit criterion. All the above-mentioned equations are taken from (R: Fit a Smoothing Spline, n.d.)

2.2.3 Validation

Extra tree regressor, Decision Tree, SVM, Lasso Regression, Gradient Boosting regressor, Linear Regression, K Neighbor Regressor, Random Forest, Huber Regression, Least Angle Regression, Ridge Regression, Elastic Net, and AdaBoost Regressor, among others, were used for validation. Each of them received training on a different type of meteorological data as well as the COVID-19 scenarios. The proposed method's performance is assessed using standard performance measures such as MAE (Mean Absolute Error), MSE (Mean Square Error), R2, and RMSE (Root Mean Square Error).

A) Mean Square Error

Root Mean Square is the performance metric used for regression models [31]. MSE took the difference of points to its regression line and square the value. Here taking the square of value is important because it removes the negative sign to give the larger weight differences. The small value of MSE denotes the regression is the best fit for the data points [31]. The MSE is calculated by the equation given below:

$$MSE \ = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{8}$$

B) Root Mean Square Error

Root Mean Square Error is also a performance metric used for regressions models by evaluating the prediction error using standard deviation. The prediction error is the difference of data points to its fitted regression line. The residual is the other name of the prediction error. RMSE defined the data points concentration around the fitted regression line [31]. RMSE is the root square of MSE. RMSE is calculated by the equation given below:

$$RMSE = \ \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{9}$$

C) Mean Absolute Error

The Mean Absolute Error is measured by taking the magnitude of error for the prediction results [32]. It took the difference of actual data with results of model prediction, here the weights for each difference are equal. The value of matrix in between 0 to infinity, the smaller score showed the goodness of model that known as negatively oriented scores [31]. MAE is calculated by the equation given below:

$$MAE \ = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j| \tag{10}$$

D) R-Square Score

R-Square (R2) error is also a performance metric that evaluates the regression models [33]. Here the goodness of fit of the model is checked by using the R2 score. R2 is used to find the scatteredness of data points around the regression line. The score value must be between 0 % to 100%. The 0% showed that there is no variability and 100% showed the highest variability of data points by its means determined by the model. R2 score is the linear function that explains the variability of the independent variable [31]. R2 is calculated by the equation given below:

$$R^2 = \frac{VarianceOfModel}{TotalVariance} \quad (11)$$

3. Results

In the proposed method, different multivariate models are used to determine the impact of weather parameters (temperature, humidity, wind speed, dew point) on COVID-19 cases. In a multivariate model, one variable is the dependent variable, and the model must predict it based on using multiple independent variables and find out the joint effects of the independent variables on the dependent variable. In the proposed method, the COVID-19 case is the dependent variable and the weather parameter is the independent variable because, in the proposed study, we need to find the effect of the weather parameter on the COVID-19 case. This study correlates three relationships, one is the daily new confirmed cases and weather parameters, the other is the daily new deaths and weather parameters, and the third is the correlation between recovered cases and weather parameters in each province of Pakistan (Punjab, Sindh, Baluchistan, KPK, and Islamabad). The results of the model are categorized according to Pakistan's Punjab, Sindh, KPK, Baluchistan, and Islamabad provinces for detailed analysis.

3.1 Model 1

Model 1 calculates linkages based on the number of confirmed cases and weather variations in Pakistani provinces. Where COVID-19 confirmed cases are the dependent variable and the weather parameter is the independent variable. As the correlation results showed there is a negative relationship between the humidity and COVID-19 cases, and for temperature, and wind speed there is a positive relationship with the confirmed cases. During 1-wave, confirmed COVID-19 cases were positively correlated with temperature and negatively correlated with humidity, as shown in Figure 4. 2-wave, on the other hand, showed a negative correlation between confirmed COVID-19 cases and temperatures in all designated areas except Baluchistan. Humidity has been shown to have a positive correlation. In order to estimate the relationship, linear regression requires only one dependent and independent variable. Multiple regression was used to examine the relationship between COVID-19 instances and weather variables in the proposed hypothesis. The RMSE performance metric is used to measure the results of multivariate techniques. The results of the multivariate methods for model 1, 1-wave and 2-wave are shown in Table 1. According to linear multivariate model analysis increase in 1oF temperature in Punjab, 42 new cases significantly increase. The linear multivariate regression showed an increase in 1-unit dew points the daily cases significantly decreased by 52.64. The linear multivariate regression results for Sindh showed the individual weather parameters are not significant for COVID-19 daily cases estimation rather than wind speed, but as jointly, weather variants are significant. In KPK, the linear multivariate regression results showed the 3% variation in the COVID-19 daily confirm cases explained by the weather variants. In the case of Baluchistan. The linear multivariate regression results showed an increase in 1° F temperature, wind speed the 5.50, and 6.11 daily confirm cases significantly increase respectively. With an increase in 1-unit dew points the daily confirms cases significantly decrease by 4.52. The 23% total variations in COVID-19 daily confirm cases explained by weather variants for Baluchistan. The multivariate models' results showed that the Loess Regression performed better in terms of RMSE in all province data.

3.2 Model 2

Model 2 calculates associations based on deaths and weather variations in Pakistani provinces, where COVID-19 death cases are the dependent variable and the weather parameter is the independent variable. During 1-wave, death COVID-19 cases were negatively correlated with temperature and positively correlated with humidity, as shown in Figure 5. On the other hand, 2-wave shows a positive correlation between confirmed COVID-19 cases and temperature in all regions. Humidity is positively correlated with Punjab, KPK, Baluchistan, and Islamabad. The results of the multivariate approach to Model 2 are shown in Table 2**.** The results showed that 4% of COVID-19 daily death cases in Punjab are explained by the weather parameters. The linear multivariate regression results showed the individual weather parameters are not significant for the estimation of daily death cases for Punjab but jointly it is significant. The linear multivariate regression results showed the 14% total variation in COVID-19 daily death cases explained by weather parameters in Sindh. As the correlation results for KPK showed a negative relationship for humidity, dew points with COVID-19 death cases. For temperature, and wind speed the correlation is positive. The linear multivariate regression results showed as individual only wind Speed can be used for estimation of COVID-19 death for KPK. And it showed 10% total variations in COVID-19 daily death cases explained by weather parameters. As the correlation results for Baluchistan showed a negative relationship between humidity with COVID-19 death cases. For temperature, wind speed, and dew points the correlation is positive. The Temperature is chosen for splines because it has a strong correlation with death cases. The multivariate Models' results showed the Loess Regression performed well with two weather parameters dew points and temperature with a minimum error rate but in terms of all-weather parameters, the Linear Both-stepwise model performed better. The linear multivariate regression results showed as individual weather parameters are not significant but jointly weather parameters are significant for the estimation of COVID-19 daily death for Islamabad. The multivariate models' results for Model 2 showed that the Loess Regression performed better in terms of RMSE in all province data.

3.3 Model 3

For Model 3, where COVID-19 recovery is the dependent variable and weather parameters are independently defined variables. Figure 6 shows the correlation results for each province. The results of the multivariate approach are shown in Table 3. The linear multivariate regression results showed the individual, as well as jointly weather parameters, are not significant for the estimation of COVID-19 recovery cases for Punjab. As the correlation of Sindh results shown in Figure 6, there is a positive relationship between the humidity, wind speed, dew points, and COVID-19 recovery cases, and a negative correlation with temperature. The linear multivariate regression results showed the individual as well as jointly the weather parameters are not significant for estimation of daily recovery cases. The multivariate models' results showed the Loess Regression with two independent variables Humidity and Temperature performed well with a minimum error rate. As the correlation, results of KPK showed in Figure 6, there is a positive relationship between the temperature, dew points, and COVID-19 recovery cases and negative for humidity. The linear multivariate regression results showed weather parameters are not significant for the estimation of COVID-19 daily death. As the correlation of Baluchistan results showed, there is a positive correlation for the temperature, dew points, and wind speed with COVID-19 recovery cases and negative for humidity. The linear multivariate regression results showed 4% of total variations in daily recovery cases explained by weather variants. The multivariate models' results for Model 3 showed that the Loess Regression performed better in terms of RMSE in all province data.

Table 1: Multivariate analysis using RMSE of COVID-19 daily confirm cases with weather parameters of 1-wave and 2-wave

	Multivariate analysis using RMSE of COVID-19 daily confirm cases with weather parameters of 1-wave					Multivariate analysis using RMSE of COVID-19 daily confirm cases with weather parameters of 2-wave				
Machine Learning (Multivariate Models)	**Punjab**	**Sindh**	**KPK**	**Baluchistan**	**Islamabad**	**Punjab**	**Sindh**	**KPK**	**Baluchistan**	**Islamabad**
Linear Multiple Regression	458.36	559.451	146.221	71.361	135.224	99.735	507.321	80.443	16.818	112.687
Linear Forward stepwise Regression	458.47	556.900	146.221	71.361	135.224	99.340	504.316	79.894	16.675	113.346
Linear Backward stepwise Regression	458.47	556.900	146.221	71.361	135.224	99.340	504.316	79.894	16.646	113.346
Linear Both stepwise Regression	458.47	556.900	152.621	71.077	135.224	99.340	504.316	79.894	16.646	116.539
MARS	386.39	437.346	158.717	68.046	121.519	100.542	584.646	87.435	15.760	106.092
Loess Regression	392.18	419.946	153.763	72.101	136.171	170.811	1419.38	108.870	24.577	160.010

Table 2: Multivariate analysis using RMSE of COVID-19 daily death cases with weather parameters of 1-wave and 2 -wave

	Multivariate analysis using RMSE of COVID-19 daily death cases with weather parameters of 1-wave					Multivariate analysis using RMSE of COVID-19 daily death cases with weather parameters of 2-wave				
Machine Learning (Multivariate Models)	**Punjab**	**Sindh**	**KPK**	**Baluchistan**	**Islamabad**	**Punjab**	**Sindh**	**KPK**	**Baluchistan**	**Islamabad**
Linear Multiple Regression	9.2751	10.525	4.6687	0.9380	1.16699	7.85639	7.00106	2.34439	0.5704	1.96686
Linear Forward stepwise Regression	9.1666	10.525	4.7775	1.0362	1.16699	8.02259	7.08314	2.38171	0.5752	1.96686
Linear Backward stepwise Regression	9.1666	10.528	4.7094	0.9193	1.16699	8.02259	7.01074	2.38171	0.5724	1.96686
Linear Both stepwise Regression	9.1666	11.098	4.7775	1.0215	1.15617	8.02259	7.08314	2.42906	0.5752	1.96686
MARS	8.0576	9.0707	4.2568	1.0389	1.34692	8.97662	7.14419	2.85432	0.5650	1.72889
Loess Regression	10.072	9.8481	3.9666	5.8227	1.31345	8.57131	9.05037	2.45949	0.6613	1.77838

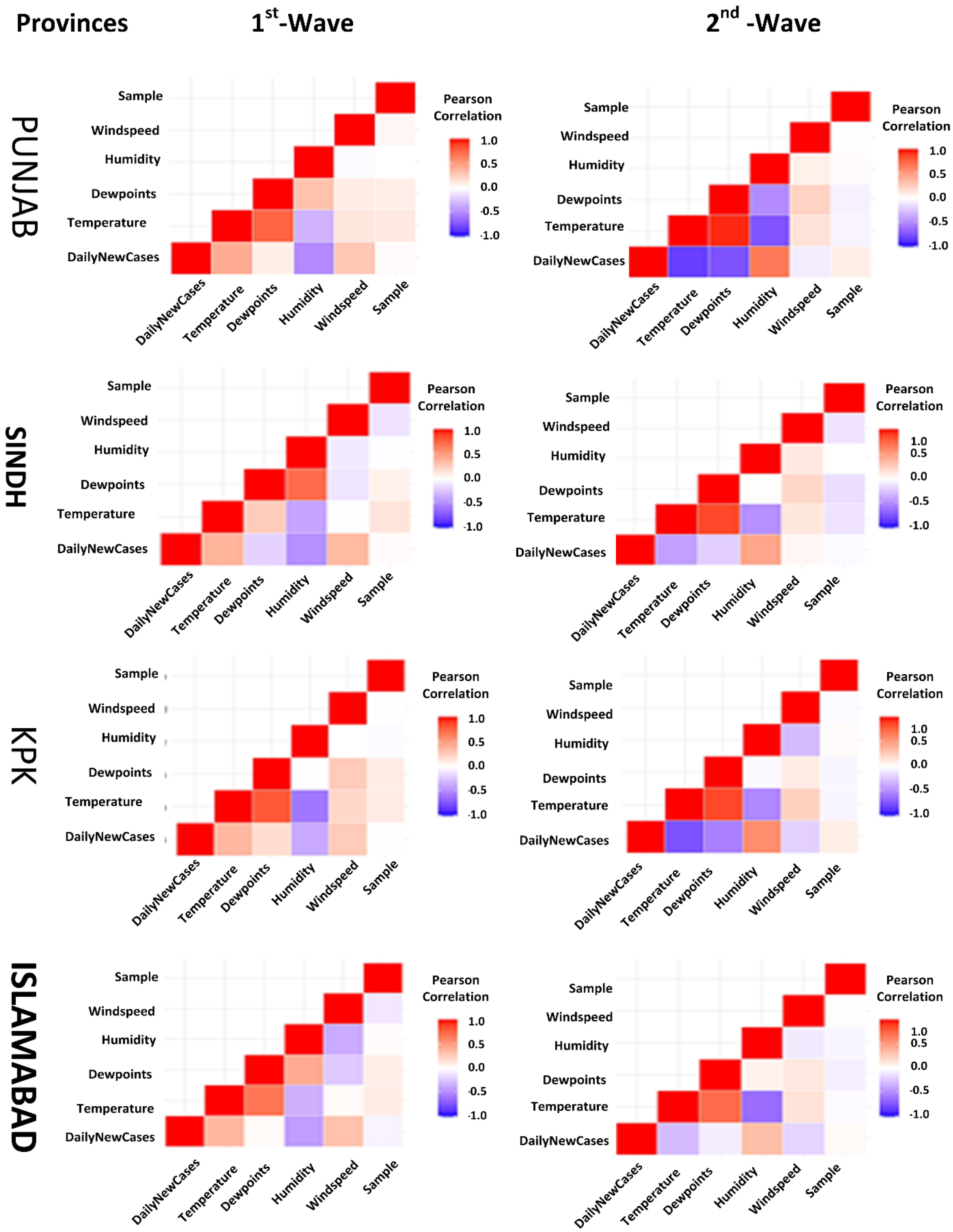


Figure 4: Model 1: Pearson Correlation results of COVID-19 daily confirm cases with the weather parameters.

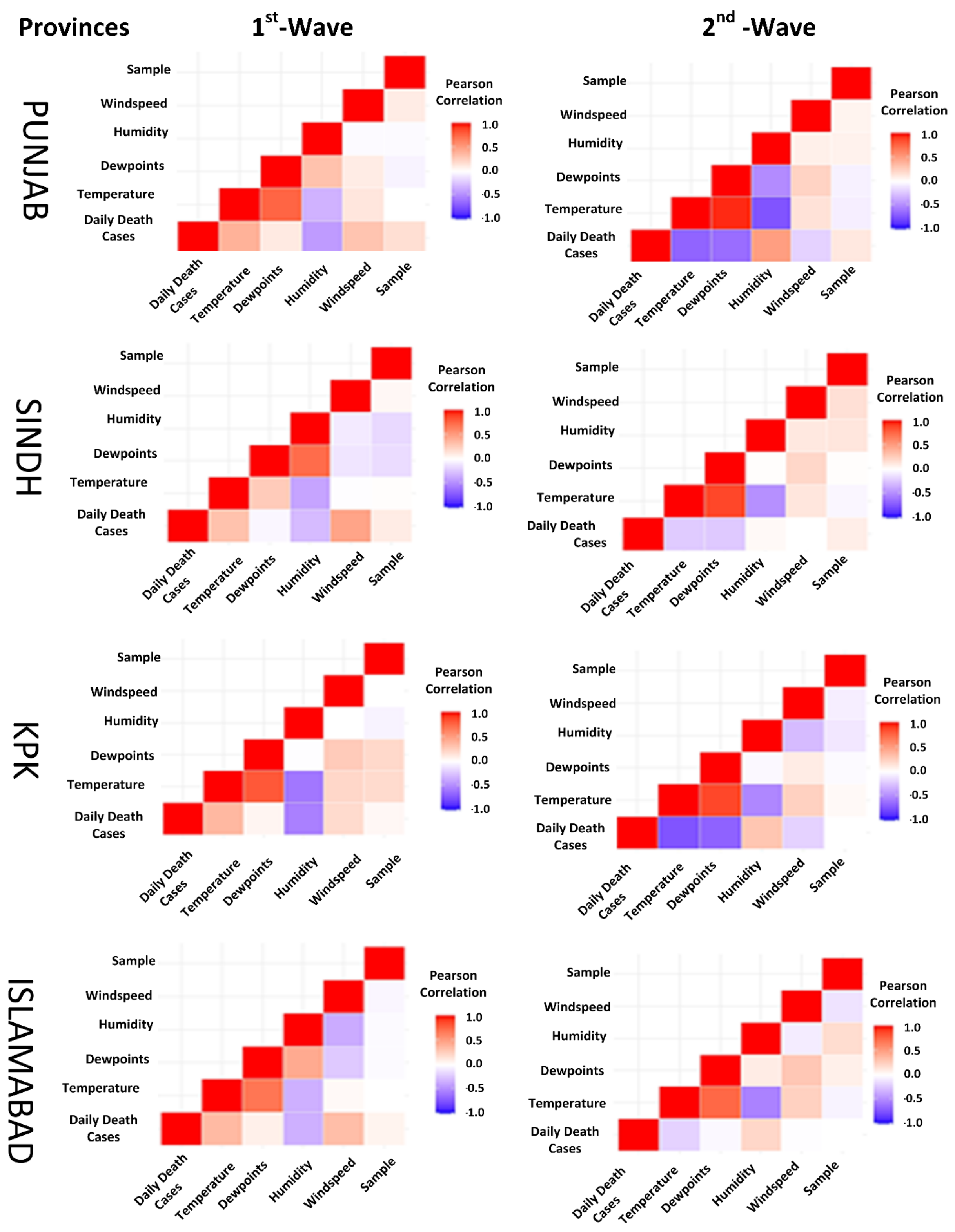


Figure 5: Model 2: Pearson Correlation results of COVID-19 daily death cases with the weather parameters.

3.4 Spline Regression

For detailed analysis, spline regression models were used to examine whether COVID-19 spread was affected by weather parameters. In the spline regression analysis, the dataset for each province was divided into two sub-datasets named 1-wave and 2-wave. The dataset contains data on confirmed COVID-19 cases and weather parameters (temperature, humidity, wind speed, dew point) from March 2020 to February 2021, divided into the first wave from March 2020 to September 24, 2020, and the Second wave on September 25, 2020, to February 2021. Spline regression analysis was performed on temperature, humidity, wind speed, and dew points for each province with daily confirmed COVID-19 cases as shown in Figures 7, 8, 9, and 10. Because spline regression is a fitted line model that finds patterns between variables. The graphical results show that in 1-wave, as the temperature increased, COVID-19 cases also increased in each province, but for 2-wave and not Baluchistan, and all provinces showed the same pattern i.e. COVID-19 cases increase as temperatures drop.

The Prediction Results of the Spline regression for COVID-19 daily confirmed cases with temperature, humidity, wind speed, and dew points are shown in Figure 10.

Table 3: Multivariate analysis using RMSE of COVID-19 daily recovery cases with weather parameters of 1-wave and 2 -wave

	Multivariate analysis using RMSE of COVID-19 daily recovery cases with weather parameters of 1-wave					Multivariate analysis using RMSE of COVID-19 daily recovery cases with weather parameters of 2-wave				
Datasets	**Punjab**	**Sindh**	**KPK**	**Baluchistan**	**Islamabad**	**Punjab**	**Sindh**	**KPK**	**Baluchistan**	**Islamabad**
Linear Multiple Regression	1906.71	757.719	469.806	49.1313	134.721	435.683	1108.87	157.310	30.925	125.101
Linear Forward stepwise Regression	1902.27	777.156	468.497	48.9560	107.199	435.710	1152.23	153.019	31.601	125.008
Linear Backward stepwise Regression	1904.14	757.674	469.446	48.9560	134.721	433.021	1144.23	155.544	32.034	125.008
Linear Both stepwise Regression	1902.27	777.156	460.546	48.9560	107.199	435.710	1177.52	153.019	31.601	125.008
MARS	1933.21	910.857	492.123	44.582	75.1374	471.785	1224.18	150.680	44.379	124.606
Loess Regression	1897.72	1008.48	485.515	54.4680	113.880	2964.81	1343.08	168.563	64.905	125.883

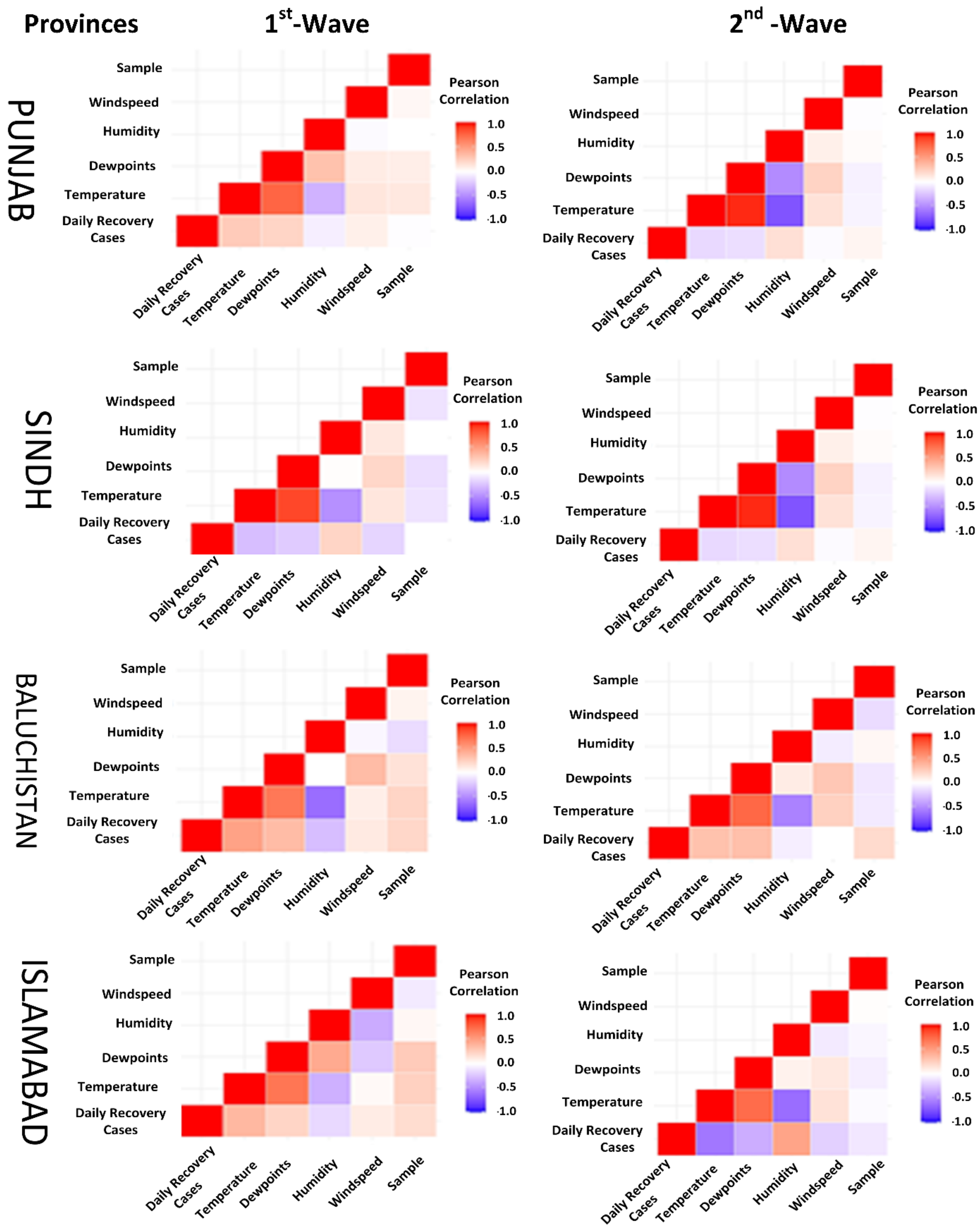


Figure 6: Model 3: Pearson Correlation results of COVID-19 daily recovery cases with the weather parameters.

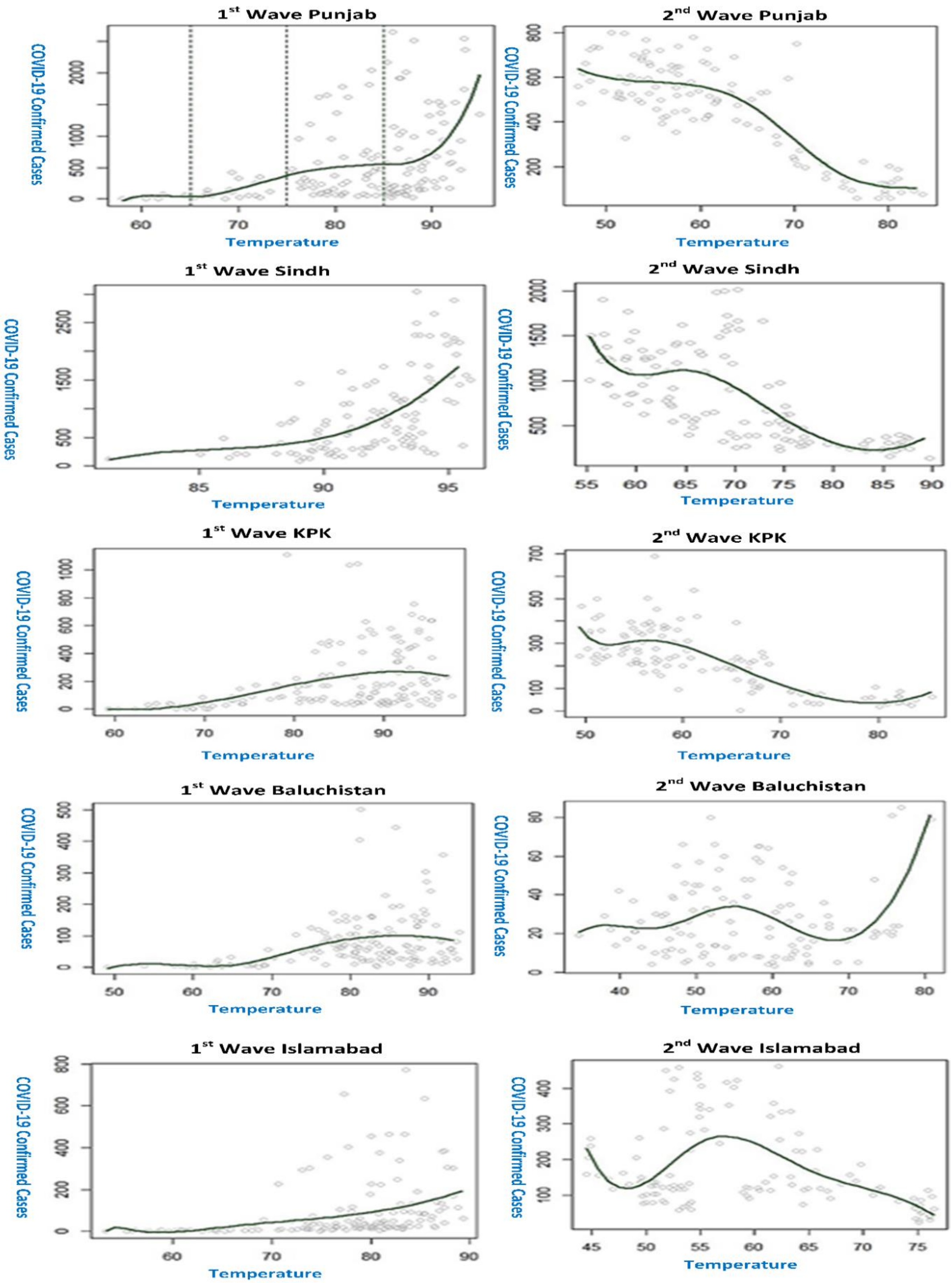


Figure 7: Spline Regression results of COVID-19 confirmed cases and temperature.

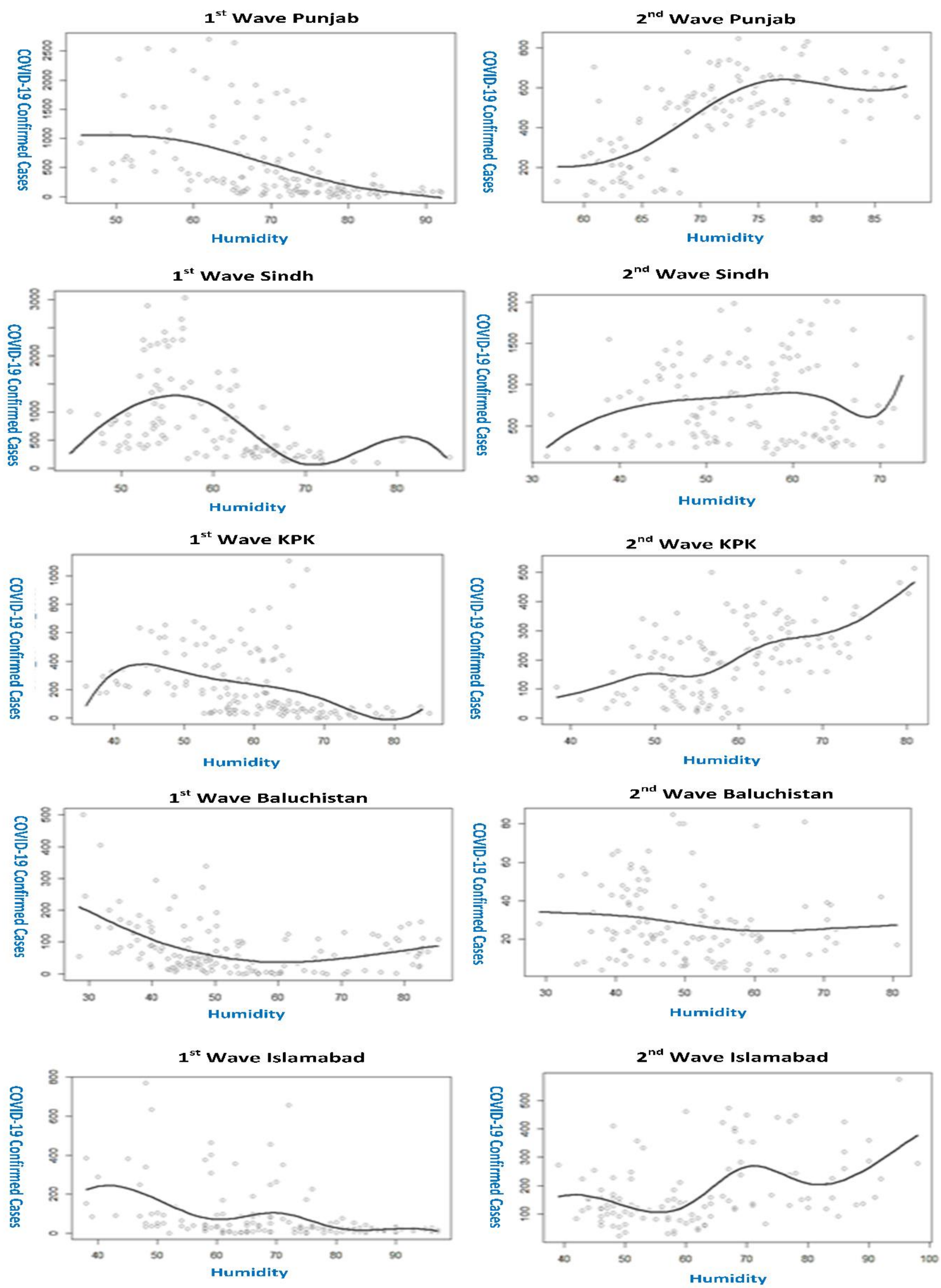


Figure 8: Spline Regression results of COVID-19 confirmed cases and humidity.

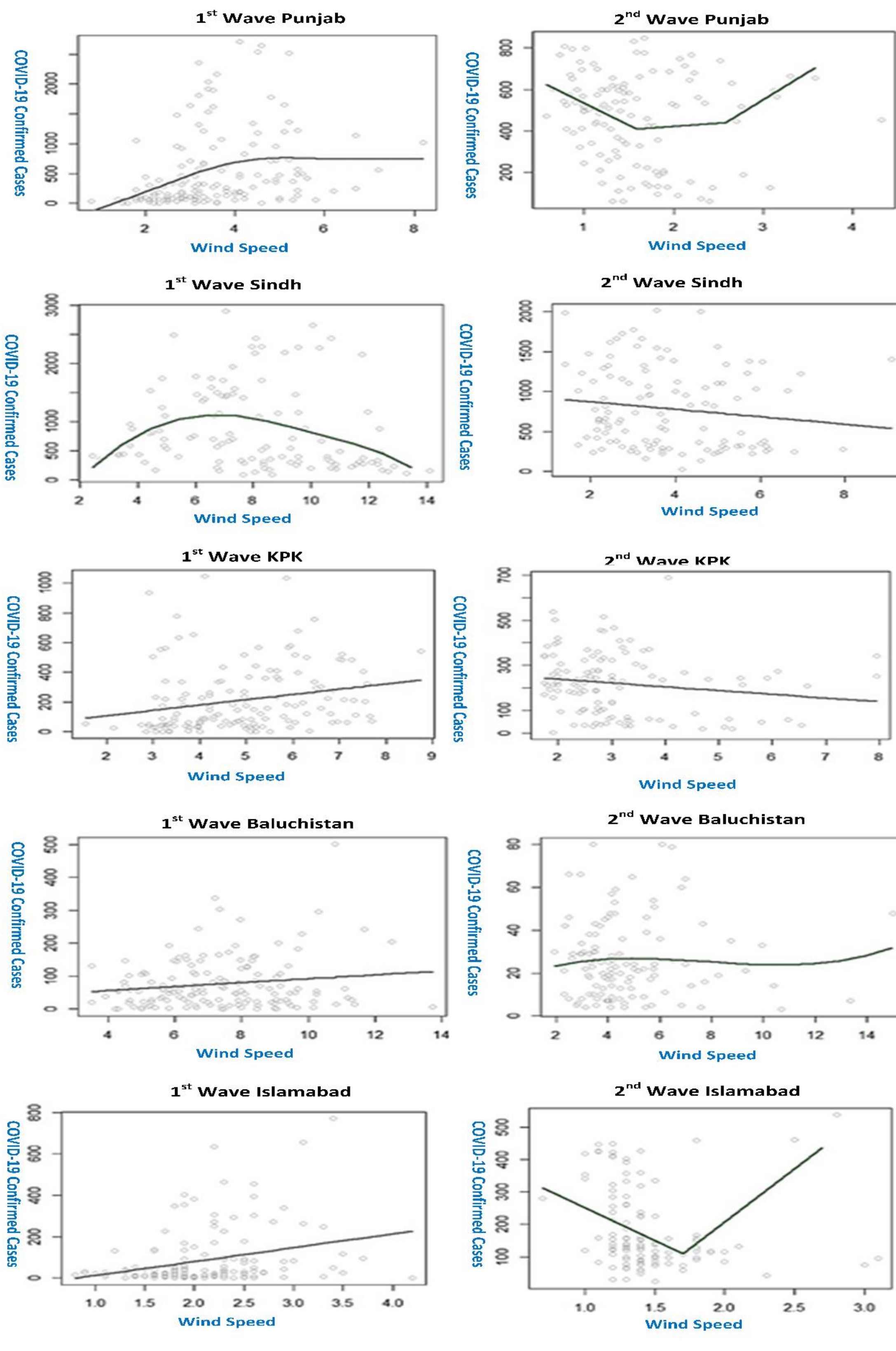


Figure 9: Spline Regression results of COVID-19 confirmed cases and wind speed.

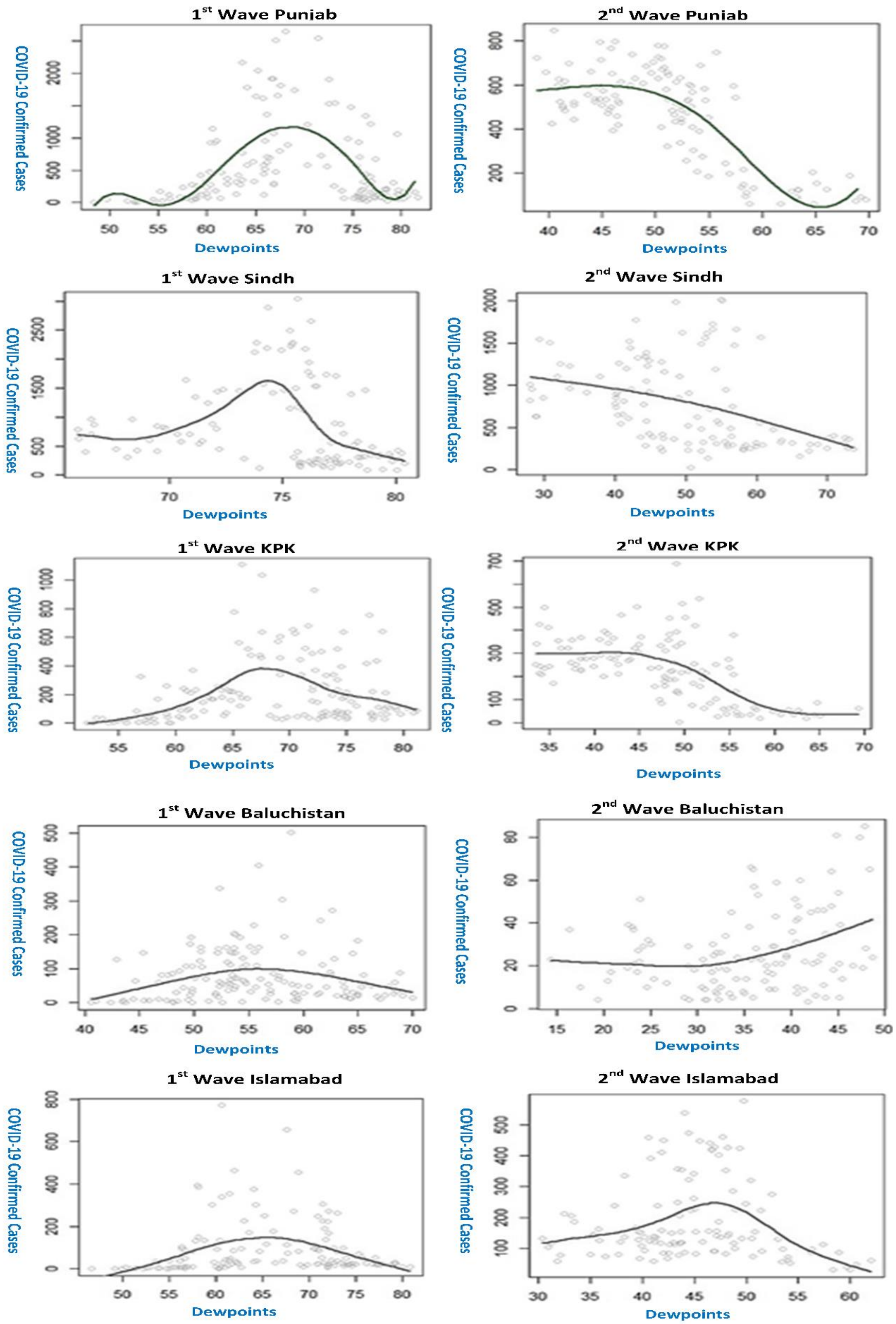
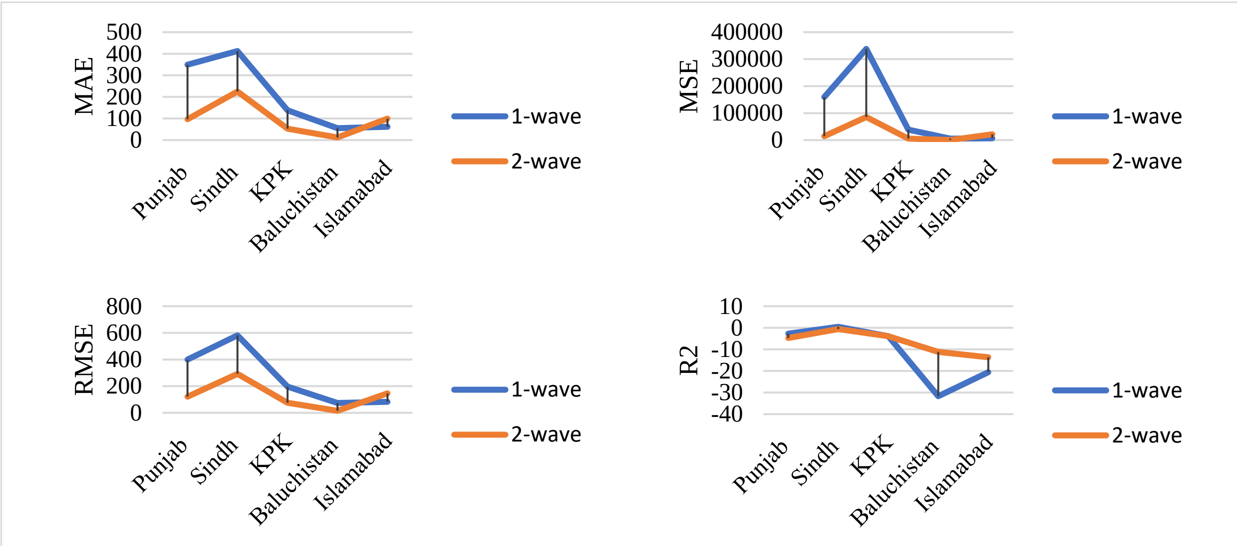
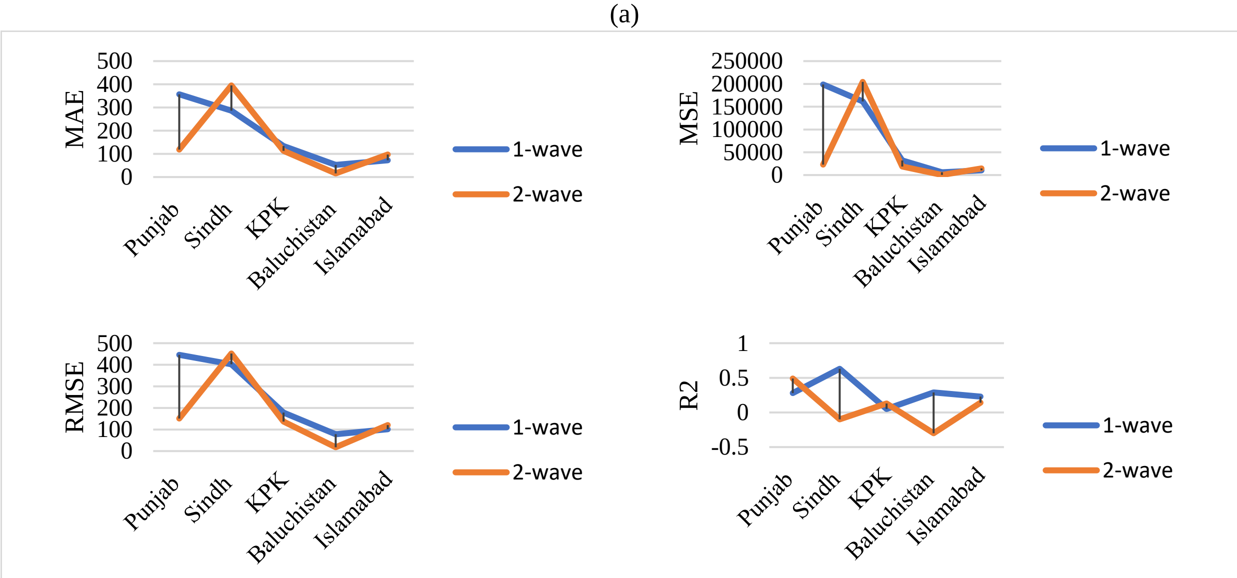


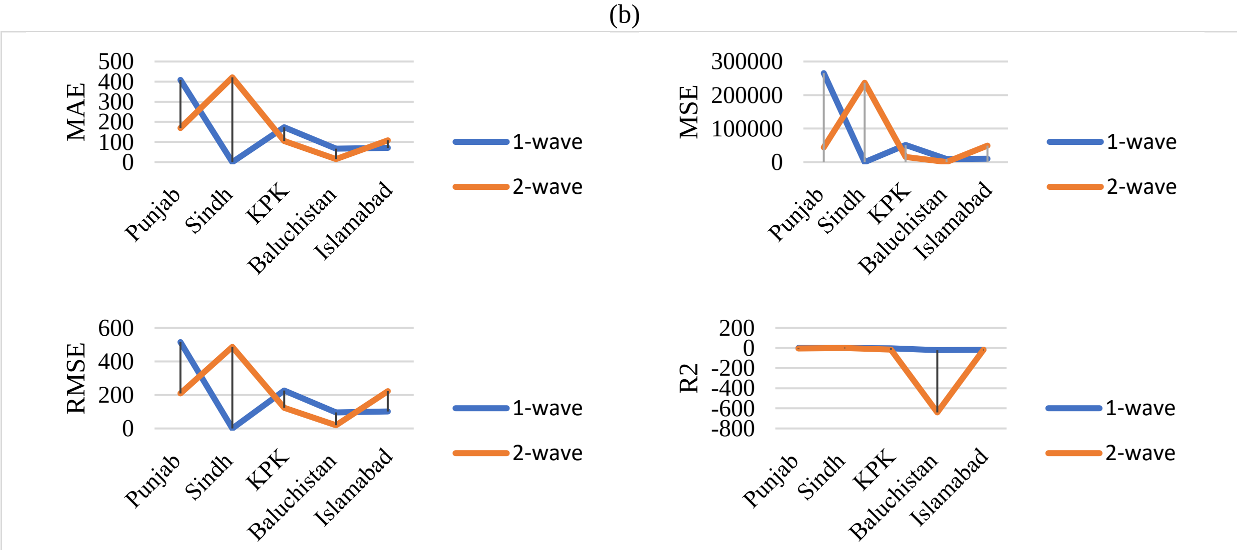
Figure 10: Spline Regression results of COVID-19 cases and dew points.


MAE
MSE
RMSE
R2
1-wave
2-wave
Punjab
Sindh
KPK
Baluchistan
Islamabad

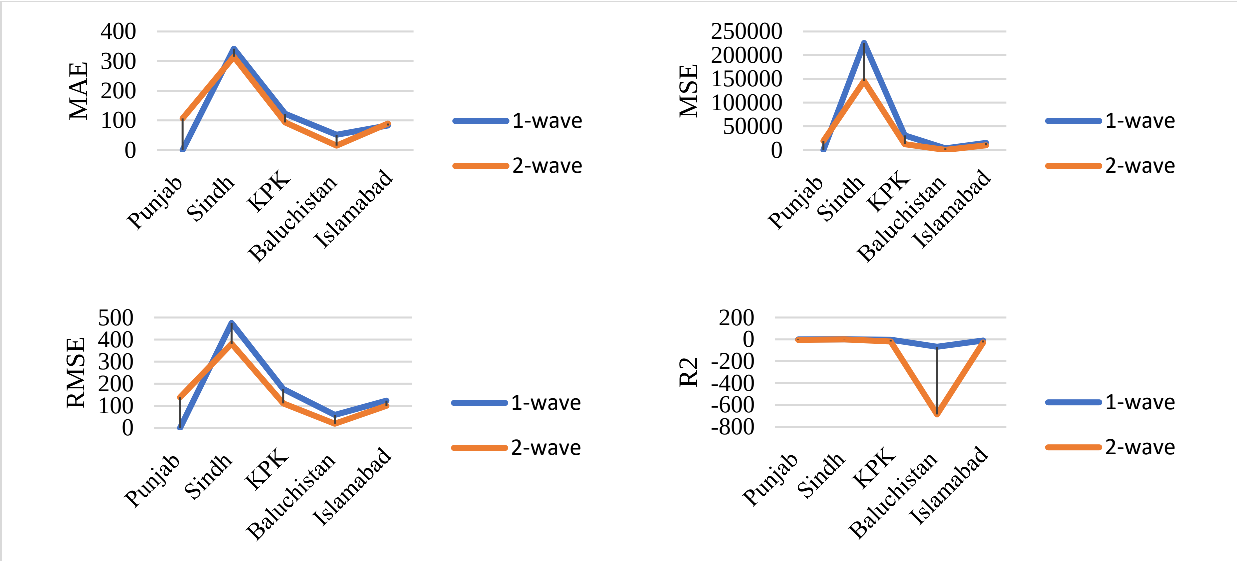
(a)


MAE
MSE
RMSE
R2
1-wave
2-wave
Punjab
Sindh
KPK
Baluchistan
Islamabad

(b)


MAE
MSE
RMSE
R2
1-wave
2-wave
Punjab
Sindh
KPK
Baluchistan
Islamabad

(c)



(d)

Figure 11: Prediction Results of Spline regression for COVID-19 daily confirmed cases with (a) temperature, (b) humidity, (c) wind speed and (d) dew points.

3.5 Validation

To validate the results of the spline model regression, we used a machine learning regression models and with non-linear regression methods i.e., linear, quadratic and cubic [34]. For model validation, the dataset contains daily confirmed cases in Punjab, Pakistan, from March 2020 to September 2020 as the dependent variable, and temperature as the independent variable, with a split of 70:30 and K-cross vaidation. As shown in Table 4, the MSE, RMSE, and R2 evaluation metrics show that the spline regression model works effectively. The K-neighbor regressor showed the best results using the MAE rating metric, while the R2 rating metric performed poorly. Another dataset used for validation is the Punjab 2-wave from September 2021 to February 2021. The spline regression model performed well on the MSE, RMSE, and R2 evaluation metrics, as shown in Table 5. Light Gradient Boosting Regressor shows the best results using the MAE evaluation metric. Passive-aggressive regressors perform poorly on MAE, MSE, and RMSE, while Huber regressors perform poorly on R2. Light Gradient Boosting Regressor shows the best results using the MAE evaluation metric. Passive-aggressive regressors perform poorly on MAE, MSE, and RMSE, and Huber regressors perform poorly on R2. From the detailed analysis of correlations in Figure 5, KPK humidity is strongly negatively correlated with COVID-19 deaths, so the KPK dataset of deaths takes from March 2020 to February 2021 with a split of 0.75. Table 6 shows the experimental results of the state-of-the-art prediction algorithm of death cases on the Pakistan-KPK COVID-19 dataset with humidity variables. Table 7 shows additional dataset results with daily confirmed COVID-19 cases and temperatures in Sindh from March 2020 to September 2021. For training and testing, the dataset is divided by 0.75. As the results show, splines do not perform well on this dataset. Cat Boost performs better in MAE, MSE, and RMSE, and for R2, Extra Tree Regressor outperforms other models.

Table 4 : State-of-the-art algorithm's prediction results of COVID-19 confirmed cases in Pakistan-Punjab for 1-wave with temperature

	70-30				**10-Fold**				**20-Fold**			
Models	**MAE**	**MSE**	**RMSE**	**R2**	**MAE**	**MSE**	**RMSE**	**R2**	**MAE**	**MSE**	**RMSE**	**R2**
KNN	323.1	203805.2	451.4	0.2	376.36	372848.28	610.61	0.45	360.6	293688.07	541.92	6.14
RF	442.13	461665	679.45	0.42	401.99	328056.36	572.76	0.03	412.73	334623.11	578.46	0.55
SVM	367.54	363817.7	603.17	0.12	425.07	479140.78	692.19	0.24	427.52	482907.63	694.91	0.36
RR	407.9	275677.4	525.04	0.15	418.76	330307.84	574.72	0.00	420.5	332391.56	576.53	0.71
LR	407.9	275679.2	525.05	0.15	418.77	330308.17	574.72	0.00	420.5	332391.82	576.53	0.71
ABR	479.07	354626.7	595.5	0.09	452.35	379755.25	616.24	0.39	446.69	370527.81	608.7	1.15
LASSO	407.9	275670.5	525.04	0.15	418.75	330308.01	574.72	0.00	420.48	332391.74	576.53	0.71
LLAR	407.9	275670.3	525.04	0.15	418.75	330308.14	574.72	0.00	420.49	332391.89	576.53	0.71
LGBM	378.76	277879.4	527.14	0.14	412.14	354362.93	595.28	0.05	415.6	358255.36	598.54	0.68
PAR	382.8	456066.5	675.32	0.4	657.95	879987.26	938.07	4.99	603.03	797667.72	893.12	2.1
SC	349	159875.5	399.844	2.7	348	159565.7	399.844	2.6	366	15999.5	401.354	2.9

KNN= K Neighbors Regressor; RF= Random Forest; SVM= Support Vector Machine; RR= Ridge Regressor; LR= Linear Regression; ABR= AdaBoost Regressor; LASSO=Lasso Regressor; LLAR= Lasso Least Angle Regression; LGBM= Light Gradiant Boosting Machine; PAR= Passive Aggressinve Regressor; SC= Spline Curve;

Table 5: State-of-the-art algorithm's prediction results of COVID-19 confirmed cases in Pakistan-Punjab for 2-wave with temperature.

	70-30				**10-Fold**				**20-Fold**			
Models	**MAE**	**MSE**	**RMSE**	**R2**	**MAE**	**MSE**	**RMSE**	**R2**	**MAE**	**MSE**	**RMSE**	**R2**
KNN	207.9	70709.6	265.91	-.66	95.53	17800.17	133.41	-0.45	90.45	14680.27	121.16	-.27
RF	93.8	17135.8	130.9	0.56	91.14	14235.87	119.31	0.65	90.51	14578.14	120.74	0.44
SVM	146.8	35717.5	188.9	0.08	68.23	44179.32	210.18	0.013	167.05	43577.55	208.75	0.13
RR	98	15434.9	124.23	0.61	97.03	14464.38	120.26	0.65	97.03	14464.8	120.26	0.42
LR	98.1	15435.6	124.2	0.6	97.03	14464.37	120.26	0.65	97.03	14464.81	120.26	0.42
ABR	106.89	18637.06	136.51	0.52	7.429	17659.14	132.88	0.576	95.3	17140.82	130.92	0.37
LASSO	98.04	15431.2	124.2	0.6	97.05	14464.49	120.26	0.65	97.05	14464.89	120.27	0.42
LLAR	98.04	15430.8	124.2	0.6	97.05	14464.47	120.268	0.65	97.05	14464.86	120.26	0.42
LGBM	92.7	15241.48	123.45	0.61	90.01	13843.44	117.65	0.665	88.09	13175.05	114.78	0.51
PAR	347.16	154333.6	392.8	-2.9	352.3	77326.61	421.1	-3.37	313.72	144749.97	380.46	-.01
SC	96.2	14532.97	120.5	-4.7	95.58	14800.37	118.91	-4.5	120.45	149887.97	124.09	-5.1

KNN= K Neighbors Regressor; RF= Random Forest; SVM= Support Vector Machine; RR= Ridge Regressor; LR= Linear Regression; ABR= AdaBoost Regressor; LASSO=Lasso Regressor; LLAR= Lasso Least Angle Regression; LGBM= Light Gradiant Boosting Machine; PAR= Passive Aggressinve Regressor; SC= Spline Curve;

Table 6: State-of-the-art algorithm's prediction results of COVID-19 death cases in Pakistan-KPK with Humidity

	70-30				10-Fold				20-Fold			
Models	**MAE**	**MSE**	**RMSE**	**R2**	**MAE**	**MSE**	**RMSE**	**R2**	**MAE**	**MSE**	**RMSE**	**R2**
KNN	4.9	37.39	6.11	-0.09	4.31	34.54	5.87	-0.13	4.35	35.34	5.94	-0.23
RF	4.93	41.4	6.43	-0.44	4.18	30.66	5.53	-.029	4.21	30.66	5.53	-0.03
SVM	4.32	27.19	5.21	0.05	4.06	31.46	5.6	0.033	4.05	31.41	5.6	0.005
RR	4.32	26.59	5.16	0.08	4.58	33.57	5.79	0.053	4.57	33.42	5.78	0.12
LR	4.32	26.59	5.16	0.08	4.58	33.57	5.79	0.05	4.57	33.42	5.78	0.12
ABR	4.71	31.98	5.66	-0.11	4.46	32.83	5.73	-0.07	4.46	32.86	5.73	-0.24
LASSO	4.32	26.56	5.15	0.08	4.58	33.54	5.79	-0.05	4.56	33.39	5.77	0.12
LLAR	4.31	26.55	5.15	0.08	4.56	33.51	5.78	0.04	4.55	33.38	5.77	0.11
LGBM	4.4	27.62	5.26	0.04	4.21	31.51	5.61	0.008	4.15	31.15	5.58	0.06
PAR	11.47	166.27	12.89	-0.78	5.9	57.97	7.61	-0.95	5.58	52.58	7.25	-0.95
SC	4.27	25.98	5.09	0.024	4.8	26.05	5.89	0.028	4.51	26.99	5.76	0.030

KNN= K Neighbors Regressor; RF= Random Forest; SVM= Support Vector Machine; RR= Ridge Regressor; LR= Linear Regression; ABR= AdaBoost Regressor; LASSO=Lasso Regressor; LLAR= Lasso Least Angle Regression; LGBM= Light Gradiant Boosting Machine; PAR= Passive Aggressinve Regressor; SC= Spline Curve;

Table 7: State-of-the-art algorithm's prediction results of COVID-19 daily confirmed cases in Pakistan-Sindh with Temperature.

	70-30				10-Fold				20-Fold			
Models	**MAE**	**MSE**	**RMSE**	**R2**	**MAE**	**MSE**	**RMSE**	**R2**	**MAE**	**MSE**	**RMSE**	**R2**
KNN	544.9	419605	647.7	0.27	494.826	335533.94	579.25	0.0009	483.71	386739.2	621.88	0.34
RF	680.07	722570.78	850.04	-.26	441.96	322589.87	567.96	-0.14	439.49	314753.23	561.02	-.08
SVM	582.01	631705.23	794.79	-0.1	505.79	402039.26	634.06	-0.057	506.65	402343.39	634.3	-.13
RR	577.65	501039.56	707.8	0.12	512.04	391613.42	625.79	0.039	512.38	391177.05	625.44	0.12
LR	577.67	501156.09	707.9	0.13	512.041	391613.48	625.79	0.039	512.38	391177.12	625.44	0.12
ABR	626.7	665140.55	815.6	-.16	455.97	330987.16	575.31	-0.11	449.66	324286.94	569.46	-.04
LASSO	577.66	501080.9	707.8	0.12	512.05	391613.35	625.79	0.039	512.4	391177.21	625.44	0.12
LLAR	577.67	501135.28	707.9	0.12	512.06	391613.32	625.79	0.039	512.41	391177.41	625.44	0.12
LGBM	600.58	604115.4	777.2	-.05	426.58	307964.36	554.94	-0.16	426.51	307332.7	554.37	-.08
PAR	582.4	613653.4	783.36	-.07	811.45	1023932.8	1011.8	-2.311	738.94	852030.75	923.05	-1.9
SC	412.5	337890	581.2	0.46	408.7	334478	591.2	0.44	442.5	348390	598.2	0.47

KNN= K Neighbors Regressor; RF= Random Forest; SVM= Support Vector Machine; RR= Ridge Regressor; LR= Linear Regression; ABR= AdaBoost Regressor; LASSO=Lasso Regressor; LLAR= Lasso Least Angle Regression; LGBM= Light Gradiant Boosting Machine; PAR= Passive Aggressinve Regressor; SC= Spline Curve;

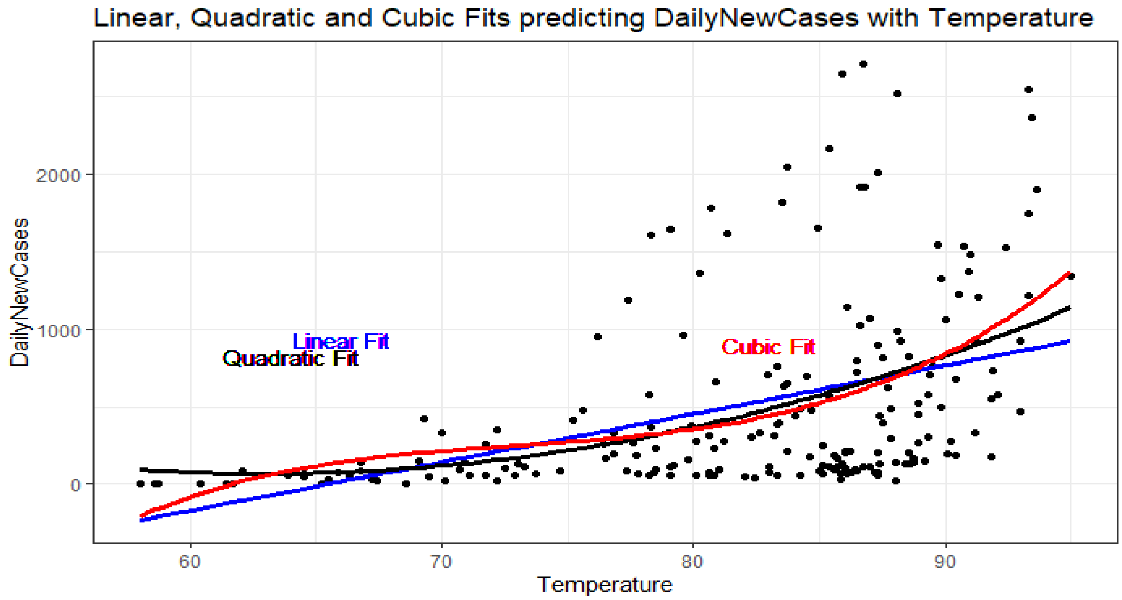


Figure 12: Prediction results of COVID-19 confirmed cases in Pakistan-KPK 1st wave with Temperature.

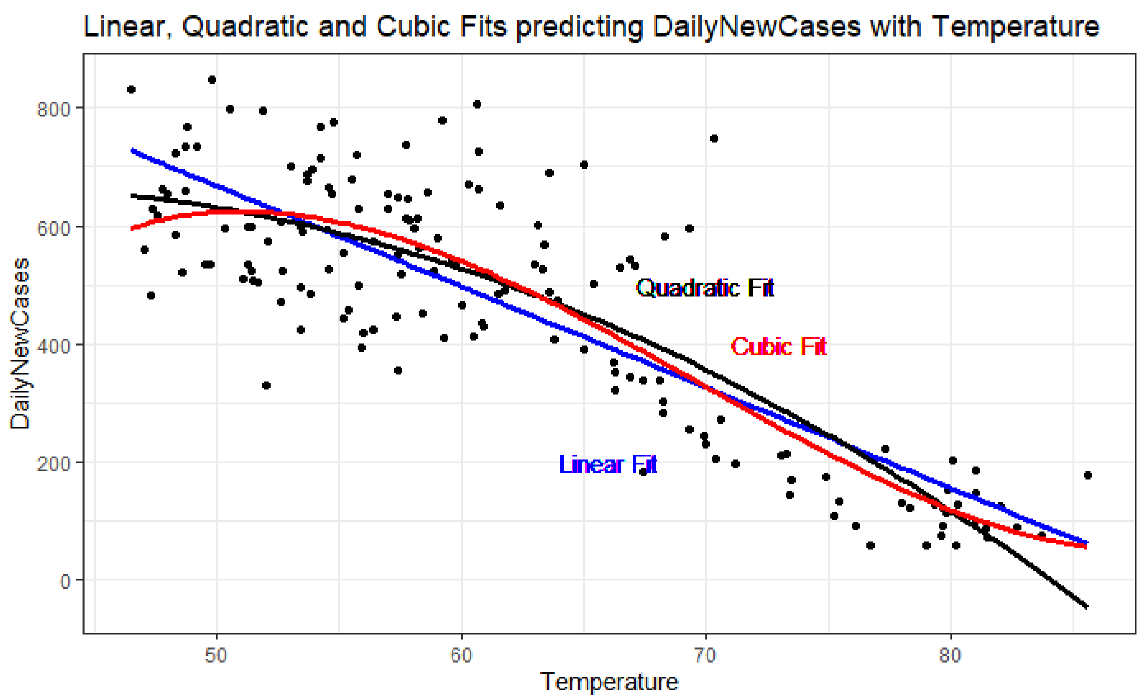


Figure 13: prediction results of COVID-19 confirmed cases in Pakistan- Punjab 2nd wave with Temperature

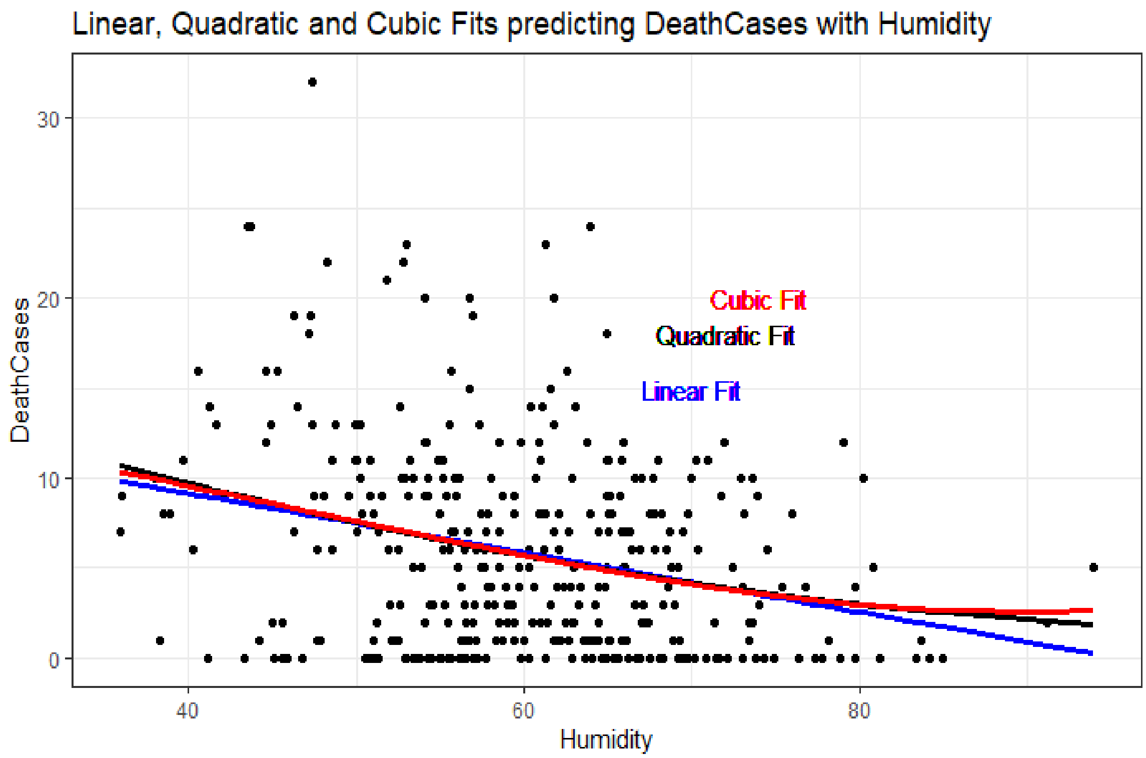


Figure 14: prediction results of COVID-19 death cases in Pakistan-KPK 1st wave with Humidity

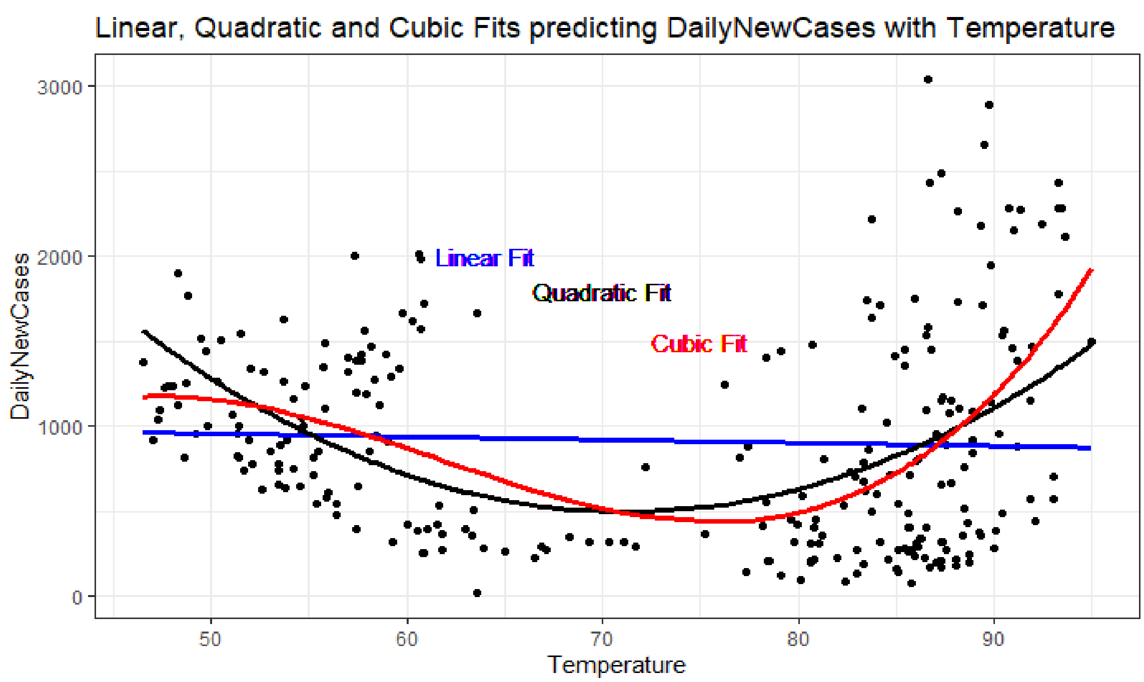


Figure 15: prediction results of COVID-19 daily confirmed cases in Pakistan-Sindh with temperature.

4. Discussion

Wu et al. state that for every 1-degree Celsius increase in temperature, the reproduction number of COVID-19 decreases by 0.0383 [10]. This means the virus spreads more slowly in areas with warmer temperatures. However, the trends examined in this study behaved somewhat differently. The virus spreads more slowly in areas with cooler temperatures and less population. Additionally, the number of COVID- 19 reproductions depends largely on the precautions a place takes to prevent the spread of the virus. For example, facilities with larger temperature measurements and no COVID-19 containment procedures will almost certainly record more cases. Conversely, locations with lower temperature readings that take all action will record a lower number of instances. The authors in [35] could not find a clear link between temperature and pandemic transmission, as the factors involved are numerous and varied. The trend analysis presented in this paper shows that there is a correlation between temperature and the speed of the spread of COVID-19. However, due to the considerable variability and uncertainty of the data provided, a perfect connection would not transmit relevant information.

Every day, as new data points are added at an exponential rate, the correlations between variables show a narrative that changes over time. In [10], the authors analyzed data from January 21 to 23, 2020, and found an inverse relationship between temperature and COVID-19 reproduction numbers. The authors examined data from January 23 to February 10, 2020, and found no significant association between temperature and COVID-19 reproduction numbers. In the proposed study, COVID-19 transmission was significantly positively correlated with temperature during the first wave. Most studies found an inverse correlation between temperature and COVID-19 [7] [9] [36]. However, there is no evidence that the number of COVID-19 cases decreases when the weather warms [37]. While the proposed study found that the association between temperature and COVID-19 was generally positive for all provinces in Pakistan during the first wave, it was predominantly negative for all provinces except Baluchistan during the second wave. Temperature appears to have had a complex effect on the above studies. Most other studies have shown that temperature in most places had a complex impact on the spread of COVID-19 during this particular time. Our understanding is that under strict lockdown conditions, temperature and humidity can affect the spread of COVID-19. Our approach has the distinct advantage of using a multivariate approach rather than just linear regression to model the spread of COVID-19. After weather change confirmed a significant association with COVID-19 transmission, it can be concluded that weather change is a key factor in the regression, resulting in a valid multiple regression. The findings of the multivariate model revealed that weather variables might be important factors in predicting the COVID-19 spread.

However, during the spread of Covid-19, a significant humidity dependence was found, which is comparable to other findings in the US (NJ) and Japan [38] [39]. In the current study, humidity was also shown to have an effect on mortality, which is consistent with previous findings in Italy and Pakistan [22] [2] [23]. Current findings on the temperature-dependent transmission of COVID-19 are comparable to the effects of temperature on the transmission of other respiratory infectious diseases such as SARS and influenza [40]. Our results provide evidence that lower temperatures are associated with higher rates of COVID-19 transmission during 2-wave At the same time, the results are consistent with recent epidemiological and laboratory studies that have found an inverse correlation between the transmission rate of COVID-19 and temperature [41] [42]. During 1-wave, COVID-19 transmission showed positive results, with temperatures similar to previous surveys [43].

However, some limitations must be taken into account. First, we used the COVID-19 cases for the provinces, averaging weather data collected from several stations to represent province weather data. However, there could be subtle variation in the dataset. In addition, all data of confirmed cases were from the official website (https://covid.gov.pk/), there could be little delay while gathering data. From this, we can conclude that it does not affect the results. Second, the implemented model is completely data-dependent. Although governments have established first-rate disease surveillance systems, there may still be underreporting, which could affect our key findings, especially at the start of the COVID-19 epidemic.

5. Conclusion

Experimental results show that weather factors can affect the COVID-19 contagion and is an important aspect to consider when predicting COVID-19 cases. correlation results show that the temperature during the first wave is positively correlated with the daily confirmed cases in each province, while the temperature during the second wave is negatively correlated with the daily confirmed cases in Baluchistan. Spline regression gave similar results. During the first wave, the public was unaware of social distancing and other factors affecting COVID-19, so the temperature effect was reduced. Higher temperatures during the second wave resulted in fewer confirmed COVID-19 cases, but we cannot conclude that temperature alone is responsible for the spread of COVID-19. This by no means that the virus cannot survive certain temperature settings. On the other hand, governments can make the most of this information by taking proper precautions to prevent the spread of the virus. The dataset may be expanded in the future as data on COVID-19 cases and meteorological factors for the prospective study were collected from March 2020 to February 2021. 3rd and 4th-wave analyses can be performed by augmenting the dataset to better understand the connection between the two variants. Other features such as vaccinated people, pollution, and air quality can also improve model performance.

Conflict of Interest Statement: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou and L. Zhao, "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding," *Lancet,* vol. 395(10224), pp. 565-574, 2020.

[2] Z. Malki, E.-S. Atlam, A. E. Hassanien and e. al., "Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches," *Chaos, Solitons & Fractals,* vol. 138, no. 110137, 2020.

[3] G. Muhammad, F. Alshehri, F. Karray and e. al., "A comprehensive survey on multimodal medical signals fusion for smart healthcare systems," *Information Fusion,* vol. 76, pp. 355-375, 2021.

[4] S. Raza, M. Rasheed and M. Rashid, "Transmission Potential and Severity of COVID-19 in Pakistan," *Preprints,* vol. 2020040004, 2020.

[5] S. P. Adhikari, S. Meng, Y.-J. Wu, Y.-P. Mao, R.-X. Ye, Q.-Z. Wang, C. Sun, S. Sylvia, S. Rozelle, H. Raat and H. Zhou, "Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review," *Infectious Diseases of Poverty,* vol. 9, no. 1, p. 29, 2020.

[6] G. Muhammad, S. Alqahtani and A. Alelaiwi, "Pandemic Management for Diseases Similar to COVID-19 Using Deep Learning and 5G Communications," *IEEE Network,* vol. 35, no. 3, pp. 21-26, 2021.

[7] P. Shi, Y. Dong, H. Yan, C. Zhao, X. Li, W. Liu, M. He, S. Tang and S. Xi, "Impact of temperature on the dynamics of the COVID-19 outbreak in China," *Science of The Total Environment,* vol. 728, no. 138890, 2020.

[8] B. D. Dalziel, S. Kissler, J. R. Gog, C. Viboud, O. N. Bjørnstad, C. J. E. Metcalf and B. T. Grenfell, "Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities," *Science,* vol. 362, no. 6410, pp. 75-79, 2018.

[9] M. F. F. Sobral, G. B. Duarte, A. I. G. P. Sobral, M. L. M. Marinho and A. S. Melo, "Association between climate variables and global transmission oF SARS-CoV-2," *Science of The Total Environment,* vol. 729, no. 138997, 2020.

[10] Y. Wu, W. Jing, J. Liu, Q. Ma, J. Yuan, Y. Wang, M. Du and M. Liu, "Effects of temperature and humidity on the daily new cases and new deaths of COVID-19 in 166 countries," *Science of The Total Environment,* vol. 729, no. 139051, 2020.

[11] P. Gale, A. Brouwer, V. Ramnial, L. Kelly, R. Kosmider, A. R. Fooks and E. L. Snary, "Assessing the impact of climate change on vector-borne viruses in the EU through the elicitation of expert opinion," *Epidemiology & Infection,* vol. 138, no. 2, pp. 214-225, 2010.

[12] P. Stott, "How climate change affects extreme weather events," *Science,* vol. 352, no. 6293, 2016.

[13] L. Gaur, U. Bhatia, N. Z. Jhanjhi and e. al., "Medical image-based detection of COVID-19 using Deep Convolution Neural Networks," *Multimedia Systems,* pp. 1-22, 2021.

[14] M. Ahmadi, A. Sharifi, S. Dorosti, S. J. Ghoushchi and N. Ghanbari, "Investigation of effective climatology parameters on COVID-19 outbreak in Iran," *Science of The Total Environment,* vol. 729, no. 138705, 2020.

[15] Y. Ma, S. Pei, J. Shaman, R. Dubrow and K. Chen, "Roles of meteorological conditions in COVID-19 transmission on a worldwide scale," *Nature Communications,* vol. 12, no. 3602, 2021.

[16] M. M. Sajadi, P. Habibzadeh, A. Vintzileos, S. Shokouhi, F. Miralles-Wilhelm and A. Amoroso, "Temperature, humidity, and latitude analysis to estimate potential spread and seasonality of coronavirus disease 2019 (COVID-19)," *JAMA Network Open,* vol. 3, no. 6, p. e2011834, 2020.

[17] Y. Ma, Y. Zhao, J. Liu, X. He, B. Wang, S. Fu, J. Yan, J. Niu, J. Zhou and B. Luo, "Effects of temperature variation and humidity on the death of COVID-19 in Wuhan," *Science of The Total Environment,* vol. 724, no. 138226, 2020.

[18] D. N. Prata, W. Rodrigues and P. H.Bermejo, "Temperature significantly changes COVID-19 transmission in (sub) tropical cities of Brazil," *Science of The Total Environment,* vol. 729, no. 138862, 2020.

[19] M. Jahangiri, M. Jahangiri and M. Najafgholipour, "The sensitivity and specificity analyses of ambient temperature and population size on the transmission rate of the novel coronavirus (COVID-19) in different provinces of Iran," *Science of The Total Environment,* vol. 728, no. 138872, 2020.

[20] H. Batool, A. Karamat, K. Waheed, S. Anwar, S. A. Haider, S. M. A. Naqvi and M. Javed, "Clinical and laboratory characteristics of COVID-19 infection in patients presenting to a tertiary care hospital," *Biomedica,* vol. 37, no. 3, pp. 179-184, 2021.

[21] H. Batool and L. Tian, "Correlation Determination between COVID-19 and Weather Parameters Using Time Series Forecasting: A Case Study in Pakistan," *Mathematical Problems in Engineering,* vol. 2021, no. 9953283, 2021.

[22] R. Basray, A. Malik, W. Waqar, A. Chaudhry, M. W. Malik, M. A. Khan, J. A. Ansari and A. Ikram, "Impact of environmental factors on COVID-19 cases and mortalities in major cities of Pakistan," *Journal of Biosafety and Biosecurity,* vol. 3, no. 1, pp. 10-16, 2021.

[23] A. Raza, M. T. I. Khan, Q. Ali, T. Hussain and S. Narjis, "Association between meteorological indicators and COVID-19 pandemic in Pakistan," *Environmental Science and Pollution Research,* vol. 28, p. 40378–40393, 2021.

[24] S. Thangariyal, A. Rastogi, A. Tomar, A. S. Bhadoria and S. Baweja, "Impact of temperature and sunshine duration on daily new cases and death due to COVID-19," *Journal of Family Medicine and Primary Care,* vol. 9, no. 12, p. 6091–6101, 2020.

[25] R. Xu, H. Rahmandad, M. Gupta, C. DiGennaro, N. Ghaffarzadegan, H. Amini and M. S. Jalali, "Weather conditions and COVID-19 transmission: estimates and projections," *medRxiv,* 2020.

[26] A. M. Schweidtmann, J. M. Weber, C. Wende and e. al., "Obey validity limits of data-driven models through topological data analysis and one-class classification," *Optimization and Engineering,* vol. 23, p. pages855–876, 2022.

[27] D. P. Solomatine and A. Ostfeld, "Data-driven modelling: some past experiences and new approaches," *Journal of Hydroinformatics,* vol. 10, no. 1, pp. 3-22, 2008.

[28] A. Heidari, N. J. Navimipour, M. Unal and S. Toumaj, "Machine learning applications for COVID-19 outbreak management," *Neural Computing and Applications,* vol. 34, p. 15313–15348, 2022.

[29] C. d. Boor, A Practical Guide to Splines, New York: Springer-Verlag, 1978.

[30] K. Demertzis, L. Magafas and D. Tsiotas, "Flattening the COVID-19 Curve: The" Greek" case in the Global Pandemic," *arXiv:2010.12040 [stat.AP],* 2020.

[31] F. Rustam, A. A. Reshi, A. Mehmood and e. al., "COVID-19 Future Forecasting Using Supervised Machine Learning Models," *IEEE Access,* vol. 8, pp. 101489-101499, 2020.

[32] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research,* vol. 30, no. 1, pp. 79-82, 2005.

[33] H. Ji-Hyeong and C. Su-Young, "Consideration of manufacturing data to apply machine learning methods for predictive manufacturing," in *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*, Vienna, Austria, 2016.

[34] A. Heidari, N. J. Navimipour, M. Unal and ShivaToumaj, "The COVID-19 epidemic analysis and diagnosis using deep learning: A systematic literature review and future directions," *Computers in Biology and Medicine,* vol. 141, p. 105141, 2022.

[35] C. Poirier, W. Luo, M. S. Majumder and e. al., "The role of environmental factors on transmission rates of the COVID-19 outbreak: an initial assessment in two spatial scales," *Scientific Reports,* vol. 10, p. 17002, 2020.

[36] M. Sarmadi, N. Marufi and V. K. Moghaddam, "Association of COVID-19 global distribution and environmental and demographic factors: An updated three-month study," *Environmental Research,* vol. 188, no. 109748, 2020.

[37] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu and e. al., "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet,* vol. 395, no. 10223, pp. 497-506, 2020.

[38] B. Doğan, M. B. Jebli, K. Shahzad and e. al., "Investigating the Effects of Meteorological Parameters on COVID-19: Case Study of New Jersey, United States," *Environmental Research,* vol. 191, p. 110148, 2020.

[39] Y. Diao, S. Kodera, D. Anzai and e. al., "Influence of population density, temperature, and absolute humidity on spread and decay durations of COVID-19: A comparative study of scenarios in China, England, Germany, and Japan," *One Health,* vol. 12, p. 100203, 2021.

[40] J. Tan, L. Mu, J. Huang, S. Yu, B. Chen and J. Yin, "An initial investigation of the association between the SARS outbreak and weather: with the view of the environmental temperature and its variation," *Journal of Epidemiology & Community Health,* vol. 59, no. 3, p. 186–192, 2005.

[41] Q. Bukhari, J. M. Massaro, S. Ralph B. D'Agostino and S. Khan, "Effects of Weather on Coronavirus Pandemic," *International Journal of Environmental Research and Public Health,* vol. 17, no. 15, p. 5399, 2020.

[42] M. Ujiie, S. Tsuzuki and N. Ohmagari, "Effect of temperature on the infectivity of COVID-19," *International Journal of Infectious Diseases,* vol. 95, pp. 301-303, 2020.

[43] S. R. Babu, N. N. Rao, S. V. Kumar and e. al., "Plausible Role of Environmental Factors on COVID-19 Transmission in the Megacity Delhi, India," *Aerosol Air Quality Research,* vol. 20, no. 10, p. 2075–2084, 2020.

[44] C. Poirier, W. Luo, M. S. Majumder, D. Liu, K. D. Mandl, T. A. Mooring and M. Santillana, "The role of environmental factors on transmission rates of the COVID-19 outbreak: an initial assessment in two spatial scales," *Scientific Reports,* vol. 10, no. 17002, 2020.