

Estimating the impact of daily weather on the temporal pattern of COVID-19 outbreak in India

Amitesh Gupta[1], Biswajeet Pradhan[2,3]* and Khairul Nizam Abdul Maulud[3]

[1]Remote Sensing & GIS Department, JIS University, Agarpara, Kolkata, India

[2]Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Information, Systems and Modelling, Faculty of Engineering & IT, University of Technology Sydney, Australia

[3]Earth Observation Center, Institute of Climate Change, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor, Malaysia

*Correspondence: Biswajeet.Pradhan@uts.edu.au or biswajeet24@gmail.com

Abstract

The COVID-19 pandemic has spread obstreperously in India. The increase in daily confirmed cases accelerated significantly from ~5 additional new cases (ANC)/day during early March up to ~249 ANC/day during early June. An abrupt change in this temporal pattern was noticed during mid-April, from which can be inferred a much reduced impact of the nationwide lockdown in India. Daily Maximum (T_{Max}), Minimum (T_{Min}), Mean (T_{Mean}) and Dew Point Temperature (T_{Dew}), Wind Speed (WS), Relative Humidity, and Diurnal range in Temperature and Relative Humidity during March 01 to June 04, 2020 over 9 major affected cities are analyzed to look into the impact of daily weather on COVID-19 infections on that day and 7, 10, 12, 14, 16 days before those cases were detected (i.e., on the likely transmission days). Spearman's correlation exhibits significantly lower association with WS, T_{Max}, T_{Min}, T_{Mean}, T_{Dew}, but is comparatively better with a lag of 14 days. Support Vector regression successfully estimated the count of confirmed cases (R^2>0.8) at a lag of 12-16 days, thus reflecting a probable incubation period of 14 ± 02 days in India. Approximately 75% of total cases were registered when T_{Max}, T_{Mean}, T_{Min}, T_{Dew}, and WS at 12-16 days previously were varying within the range of 33.6-41.3° C, 29.8-36.5° C, 24.8-30.4° C, 18.7-23.6° C, and 4.2-5.75 m/s respectively. Thus, we conclude that coronavirus transmission is not well correlated (linearly) with any individual weather

parameter; rather, transmission is susceptible to a certain weather pattern. Hence multivariate non-linear approach must be employed instead.

Keywords: COVID-19; weather; temporal trend; India

1. Introduction

In human history, it is apparent that pathogens have caused devastating consequences in social wellbeing and economies (Briz-Redón and Serrano-Aroca 2020). The recent novel coronavirus disease (COVID-19) is one prominent example of such a disastrous event that has grasped the world. The earliest outbreak of COVID-19 caused by Severe Acute Respiratory Syndrome CoronaVirus-2 (SARS-CoV-2) happened in Wuhan, Hubei Province, China during the late December, 2019, (Guan et al. 2020; Wu and McGoogan 2020; Zhu et al. 2020; Zu et al. 2020). Because of human-to-human transmissibility of the virus by contact, droplets and fomites (Wang et al. 2020a, Wang et al. 2020b), the transmission of this disease has become progressively more unpredictable and populations have become more vulnerable. Considering the rapid spread of the virus, the World Health Organization (WHO) declared an international public health emergency on January 30, 2020, and later on March 11, 2020, WHO declared this disease to be a global pandemic, due to the exponential surge in the total number of infections. Up to June 04, 2020, a total of 6,697,418 cases have been affirmed with 5.85% of these resulting in deaths worldwide (https://www.worldometers.info/coronavirus). Despite the fact that India registered its first case on January 29, 2020, the real outbreak occurred from March 02, 2020 onwards, and as of June 04, 2020, a total of 226,722 cases have been confirmed; however, the death rate (2.81%) is much lower than in the rest of the world.

Clinical investigations of COVID-19 identified respiratory droplets as the most common agent of infection (Ge et al. 2013; Huang et al. 2020) and the symptoms are also quite analogous to other coronavirus diseases such as MERS and SARS (Holshue et al. 2020; Perlman 2020; Tan et al. 2005; Wang et al. 2020a). WHO also reported that the SARS-CoV-2 virus initially causes respiratory disease, presents as a wide range of illness from asymptomatic or mild through to severe disease and death. Thus, the COVID-19 disease has close similarities in its presentation to influenza (https://www.who.int/westernpacific/news/q-a-detail/).

Environmental factors, such as daily weather and long term climatic conditions, may affect the epidemiological dynamics of this type of infectious disease (Dalziel et al. 2018;Yuan et al. 2006). Daily air temperature and relative humidity may impact on the transmission of coronavirus by affecting the persistence of the viral infections within its transmission routes (Casanova et al. 2010). A few studies accounting for climate and weather conditions found that these factors considerably affect the spatial distribution of the disease, along with its incubation period (Bedford et al. 2015; Lemaitre et al. 2019; Sooryanarain and Elankumaran 2015). Many years ago, Bull (1980) was the first to report that the mortality rate of pneumonia is intimately associated with changes in weather conditions. Other studies have revealed that among different climatic variables, air temperature affects influenza epidemics mostly in tropical regions (Tamerius et al. 2013), whereas the mid-latitude temperate regions experience influenza epidemics mostly during winter months (Bedford et al. 2015; Sooryanarain and Elankumaran 2015). Nevertheless, the response of COVID-19 transmission to weather patterns remains debatable, since studies carried out in different countries suggested an existing correlation between weather and the COVID-19 pandemic (Ficetola and Rubolini 2020; Liu et al. 2020; Ma et al. 2020; Oliveiros et al. 2020; Qi et al. 2020; Tosepu et al. 2020). Contradictorily, a few studies have reported that meteorological observations are not correlated with the outbreak pattern (Jamil et al. 2020; Mollalo et al. 2020; Shi et al. 2020; Xie and Zhu 2020). Studies carried out by (Wang et al. 2020a, Wang et al. 2020b) suggested that the spread of disease would decrease with an increase in temperature. Based on the USA model, a reduction of transmission in warmer conditions had been predicted for India (Gupta et al. 2020a) . However, in view of the long term climate record, it was found that comparatively hot areas in India are possibly going to be more affected by this disease (Gupta et al. 2020b). On the basis of regional data for several provinces in India, Goswami et al. (2020) reported on the inconsistency of the weather-infection interrelationship in India. Besides, the incubation period of COVID-19 may also vary spatially. The WHO reported an incubation period of 2-10 days for COVID-19 based on worldwide observation (World Health Organization 2020) while the National Health Commission in China had initially estimated an incubation period of 10-14 days for China (https://www.aljazeera.com/news/2020/01/chinas-national-health-commission-news-conference-coronavirus-200126105935024.html). The Centres for Disease Control and Prevention in United States of America estimate an incubation period of 2-14 days (https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html). On the other hand, some studies reported an incubation period of around 20 days (Bai et al. 2020; Guan et al. 2020).

COVID-19 has already made a significant indirect impact through reduction in anthropogenic activities on several environmental aspects in the Indian context (Gupta et al. 2020c), however, only a few studies have investigated the impact of daily weather on COVID-19 transmission nationwide, and since the incubation period of this disease in India is also not mentioned anywhere to date, there is a need for a comprehensive study about the impact of weather patterns on COVID-19 transmission in the Indian scenario. Thus, the present study is aimed at understanding the temporal patterns of the outbreak, any abrupt changes and the influence of daily weather conditions on the daily count of infected cases in India. We have also attempted to estimate the incubation period of COVID-19 based on five different time-frames: precisely on the day of the case detected, and with leads of 7, 10, 12, 14, and 16 days prior to the case detection.

2. Data and Methodology

2.1. Data collection

India, the largest country in South Asia, extends from 6° N to 38° N, and from 68° E to 98° E, comprising a land area of 3.287 million sq. km. with a total population of more than 1.2 billion (Census of India Website, 2011). The data of daily COVID-19 cases were collected from the official website of the Ministry of Health of India (https://www.mohfw.gov.in). Among a total of 725 districts in India, 618 districts have reported multiple confirmed cases. Several studies have reported that the disease spreads faster in the cities where population density is very high (Casanova et al. 2010; Ahmadi et al. 2020; Bonasera and Zhang 2020; Kang et al. 2020; Rocklöv and Sjödin 2020). Thus, among 53 'million cities' (where the total population is more than one million) in India, 9 cities have been selected for this study, from where more than 79% of the total cases in India have been reported up to June 4, 2020 (Figure 1). The daily weather data were collected from https://www.wunderground.com. Figure 2 shows the prevailing daily weather conditions in terms of Maximum, Minimum and Mean Temperature of Air, Diurnal Range in Air Temperature, Dew Point Temperature, Average Relative Humidity, Diurnal Range in Relative Humidity, and Wind Speed, in those cities. Since all the selected cities are located in different bio-climatic zones having different temperature characteristics (Gupta 2017), the variations in meteorological observations will also help to identify how spatially varying weather conditions influence the pattern of COVID-19 transmission in India.

Figure 1. Location of the selected cities in India along with the total population of those cities.

2.2. Mann-Kendall Test

The nonparametric Mann-Kendall (MK) method (Kendall 1975;Mann 1945) was applied to the daily data of COVID-19 confirmed cases during March 01 to June 04, 2020 to detect statistically significant trends. The MK test takes as the null hypothesis (H_0) that there is no trend in the count of confirmed cases of infections; while the alternate hypothesis (H_1) is that there is a trend (increasing or decreasing) over time. The mathematical expressions for calculating MK Statistics S, $V\,(S)$ and standardized test statistics Z are as follows –

$$\mathrm{S} = \sum_{i=1}^{T-1} \sum_{j=i+1}^{T} \mathrm{sgn}\,(X_j - X_i), \tag{1}$$

$$\mathrm{sgn}\,(X_j - X_i) = \begin{Bmatrix} +1 & \text{if } (X_j - X_i) > 0 \\ 0 & \text{if } (X_j - X_i) = 0 \\ -1 & \text{if } (X_j - X_i) < 0 \end{Bmatrix} \tag{2}$$

$$\mathrm{V}\,(\mathrm{S}) = \frac{1}{18}\left[T(T-1)(2T+5) - \sum_{p=1}^{q} t_p\, p(p-1)(2p+5)\right] \tag{3}$$

$$Z = \begin{Bmatrix} \frac{S-1}{\sqrt{\mathrm{VAR}\,(S)}} & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ \frac{S+1}{\sqrt{\mathrm{VAR}\,(S)}} & \text{if } S < 0 \end{Bmatrix} \tag{4}$$

where, X_i and X_j are the daily observations, T is the length of the time series, t_p is the number of ties for the p^{th} value. Positive Z values designate an increasing trend and negative Z values signpost a negative trend. For $|Z| > Z_{1-\alpha/2}$, H_1 is accepted with rejection of H_0, considering the critical value of $Z_{1-\alpha/2}$ to be 1.96 for a p value of 0.05.

The statistic S is closely related to the Kendall's τ which is given by –

$$\tau = \frac{S}{D}$$

where

$$D = \left[\frac{1}{2}T(T-1) - \frac{1}{2}\sum_{p=1}^{q} t_p\, p(p-1)\right]^{\frac{1}{2}} \left[\frac{1}{2}p(p-1)\right]^{\frac{1}{2}} \tag{5}$$

2.3. Sen's Slope Estimator

Sen's slope (Sen 1968) is widely employed to estimate the magnitude of trends.

$$d_i = \text{Median}\left[\frac{x_j - x_k}{j-k}\right] \text{ for all } j > k \quad (6)$$

where d is the slope, x_j and x_k represent the corresponding data values at time j and k, ($1 \le k < j \le n$), n is the number of the variables.

$$Q_i = \begin{Bmatrix} d_{(n+1)/2} & \text{if n is odd} \\ \frac{1}{2}\left(d_{n/2} + d_{(n+2)/2}\right) & \text{ifn is even} \end{Bmatrix} \quad (7)$$

A positive Q_i value denotes an increasing trend; a negative Q_i value signifies a decreasing trend.

In this study, the MK test and Sen's Slope Estimator were implemented to investigate the trend of daily transmission over selected cities as well as all over the whole country. This helped to establish whether the temporal pattern of transmission varied in different cities with respect to the countrywide pattern or not.

Figure 2. Pattern of Daily Weather over the selected cities in India.

2.4. Pettitt test

Originally developed by (Pettitt 1979), the non-parametric Pettitt test is an effective method of identifying the change in the temporal trend in any time-series, because of its sensitivity to breaks in the middle of temporal records (Gao et al. 2011; Hänsel et al. 2016; Jaiswal et al. 2015; Mallakpour and Villarini 2016; Wijngaard et al. 2003). In this method, S is evaluated for all random variables from 1 to T; then the most prominent change point is determined as that where the value of $|S|$ found to be largest:

$$K_T = \max_{1 \le t < T}|S| \quad (8)$$

At a particular time t, the change point is detected when K_T is clearly different from zero at any particular level, where the significant level is estimated by:

$$p = 2 * \exp\left(\frac{-6K_T^2}{T^2 + T^3}\right) \tag{9}$$

The change point can be evaluated as statistically significant only when the estimated p-value becomes less than the pre-assigned significance level i.e. α.

2.5. Growth Rate

Growth Rate denotes the magnitude of alteration of any particular variable within a definite time period. Here, growth rate between March 01 and June 04, 2020 for the overall country and for each selected city was calculated using following formula –

$$\text{Growth rate (\%)} = \left\{\left(\frac{\text{NF}}{\text{NE}}\right)^{1/n} - 1\right\} \times 100 \tag{10}$$

Here, NF refers to the number of COVID-19 cases recorded on the 1st day of record, NE refers to the number of COVID-19 cases recorded on the last day of the study period (June 04, 2020), and n refers to the number of days between the first day of COVID-19 case detection and the last day of the study period.

2.6. Doubling time

The doubling time denotes the time taken for a count to be doubled. Here doubling time for the overall country and for each selected city was calculated using following formula –

$$\text{Doubling Time} = \frac{\text{n}\times\ln(2)}{\ln(\text{NE}/\text{NF})} \tag{11}$$

2.7. Spearman's correlation test

Spearman's rank correlation coefficient (r_s) calculates the association between the number of daily new cases and other input parameters. It summarizes how well the association between daily transmission and weather parameters can be quantified. The coefficient can be calculated via following equation –

$$r_s = 1 - 6\frac{\sum {d_i}^2}{n(n^2-1)} \tag{12}$$

where, n represents the number of alternatives, and d_i is the difference between the ranks of two parameters.

All the above mentioned statistics were based on a 95% confidence level.

2.8. Support Vector Machine

Support Vector Machine (SVM) is an extensively utilized machine learning technique. It is performed on the basis of statistical auto-adaptation and the structural risk minimization principle (Tien Bui et al. 2012). By creating a hyper-plane, the nonlinearity in the input dataset is reshaped into a linear entity (Jebur et al. 2014). The key factor behind this data transformation is a kernel function. Using the assigned training dataset, SVM puts the original input into a higher dimensional feature space, then finds the supreme fringe of separation among the observations, and constructs a hyper-plane at the centre of that extreme margin (Marjanović et al. 2011). Support vectors are nothing but the nearest training points to the produced hyper plane. Thus, this model adapts itself with new input observations, creates a hyper-plane and identifies the support vectors and thereafter acts on the input variables of the testing dataset to estimate the predicted variable. Further insights about how the mathematical computations and procedures work in SVM can be found in several papers such as Tien Bui et al. (2012), Pradhan (2013);Tehrany et al. (2014); Tehrany et al. (2015). However, the accuracy of estimation depends on the kernel type selected during the training of the model (Yao et al. 2008). The Radial basis function (RBF) kernel produces more exact results and is preferred over the linear, polynomial and sigmoid kernels, due to its higher capability in interpolation (Song et al. 2011).

In the present study, the log-transformed values of daily COVID-19 cases were estimated using several daily weather parameters, along with the elevation and population of those cities (Eq. 13).

$$\ln(\mathrm{NC}) = \mathrm{SVM}(\mathrm{T_{Max}} + \mathrm{T_{Min}} + \mathrm{T_{Mean}} + \mathrm{T_{range}} + \mathrm{T_{Dew}} + \mathrm{H_{Avg}} + \mathrm{H_{range}} + \mathrm{WS} + \mathrm{Ele} + \mathrm{Pop}) \quad (13)$$

where, NC is the number of New Confirmed Case, $\mathrm{T_{Max}}$ is Maximum Air Temperature (° C), $\mathrm{T_{Min}}$ is Minimum Air Temperature (° C), $\mathrm{T_{Mean}}$ is Mean Air Temperature (° C), T_{range} is Temperature Range (° C), $\mathrm{T_{Dew}}$ is Dew point Temperature (° C), $\mathrm{H_{Avg}}$ is Average Relative Humidity (%), $\mathrm{H_{Range}}$ is Range of Relative Humidity (%), WS is Wind Speed, Ele is Elevation (m), Pop is total Population.

The total dataset was divided into a 70:30 ratio, where 70% of observations were used as a training dataset and the rest were used for testing. The accuracy of estimation was evaluated in terms of R^2, Root Mean Square Error (RMSE) and Mean Bias (MB).

All the analyses were done using R programs.

3. Results and Discussion

Table 1 presents the results of the MK test, Sen's slope and change point through the Pettitt test. All the results are found significant at α=0.05 level, thus, there is significant change in the transmission pattern. The calculated Sen's slope shows the rate of change in the count of additional new cases (ANC) in each day. For the whole country, an acceleration in confirmed cases of ~76 ANC/day was recorded from March 01 to June 04, 2020, whereas among the selected cities, Mumbai registered the highest rate of acceleration (~19 ANC/day). Analyses also revealed that the increasing rate of daily new cases accelerated abruptly all over the country after April 17, 2020, i.e. at the beginning of the 2nd phase of lockdown (April 15 to May 03, 2020). During the middle of April, a majority among the cities, namely, Ahmedabad, Chennai, Delhi, Indore, Kolkata, Pune and Mumbai registered the abrupt increase between April 12 and April 22, 2020. Only Hyderabad and Jaipur reported the abrupt increase during the 1st lockdown period (March 25 to April 14, 2020). Note too that all the major affected cities registered a 3 to 34-fold increase in transmission rate after the evaluated change point; explicitly, this incrementing rate was more than 20 times in Pune, Chennai and Ahmedabad after the estimated change point. The nationwide increase in transmission rate (case acceleration) was ~22 ANC/day before April 17 and afterward it amplified to ~174 ANC/day. Thus the entire country, and specifically the major affected cities, experienced an alarming rate of acceleration in the spread of COVID-19 during the lockdown period itself. Basically, the lockdown was implemented with a brief guideline of social distancing with an aim to reduce the occurrence of human-to-human transmission by avoiding gatherings at workplaces and at any other public places. Nationwide lockdowns resulted very effectively in reducing the growth rate of transmission in different countries across the world, such as China, Italy, France, Germany, United Kingdom (Gatto et al. 2020; Leung et al. 2020; Wurtzer et al. 2020). However, in India, the initial growth rate (acceleration) in infection up to March was approximately 5 ANC/day, and after that, the growth rate amplified multi-fold. Explicitly, the growth rate acceleration was ~49 ANC/day during April; it reached up to ~113 ANC/day during May 1 to May 20, and during May 21 to June 04, it was ~249 ANC/day (Figure 3). One of major reasons behind the initial slow growth rate might be surmised to be that the original virus had been transmitted through an infected immigrant; moreover, very few tests conducted throughout the country during March (fewer than 10,000 tests/day). Analysis also reveals that within this 96 days study period, the percentage growth rate for the overall country was 10.79%, whereas among the selected cities, Mumbai had the highest growth rate (9.98%) while Jaipur had the smallest growth rate (5.67%). Basically, the growth rate was higher in the cities, which had a higher rate of acceleration in COVID-19 cases. On the other hand, the doubling time of COVID-19 cases for Mumbai (7.31 days) and Chennai (7.57 days) was very close to the countrywide situation (7.85 days). Hyderabad registered the slowest doubling time of 12.58 days. This shows that the daily count of COVID-19 cases was doubling in less than 8 days throughout the country, which is also a measure of the drastic adverse situation in India. From Figure 3, it can be seen that an average of 53/1000 tests results were confirmed for infection during the entire study period. However, this positive rate was 35/1000 during the month of March; later, it rose to 44/1000 and 57/1000 during April 01-30, and May 01 - June 04, 2020, respectively. This shows that the probability of detecting confirmed cases also increased each week, which may be evidence of community transmission. The trend of daily new cases in the major affected cities (Figure 4) also indicates the large increase in daily transmission from May onwards. Figure 4 also shows that cities located at a lower elevation and having higher population registered a higher growth rate of transmission, thus agreeing with an early observation by (Gupta et al. 2020d). Of the 5 megacities in India, just three (Delhi, Mumbai, and Chennai) are the only cities where the count of daily infected cases exceeded 1200. One of the probable reasons behind such spikes in transmission rate might be the allowance to migrants to return to their native places, which instigated large crowds in various cities and gathering in transport hubs, as reported in many local and national newspapers, thus resulting in such an unforeseen increasing rate of transmission all over the country.

Table 1. Result of Mann-Kendall test, Sen's Slope, Pettit test, Grow rate and Doubling time.

Figure 3. Trend of daily confirmed cases in India. The weekly trend of number of confirmed cases per 1000 tests are shown inset.

The Spearman correlation analysis (Table 2) shows that there were mostly significant but still predominantly low correlations between the number of daily new cases and the various weather conditions. Among the 8 weather parameters, the correlation for T_{range} is non-significant over all time spans. Hence, the diurnal range of temperature is not significantly associated with the spread of COVID-19 cases in India. H_{avg} is associated significantly positively on the day of detection up to 10 days lag (i.e., when transmission presumably occurred). However, H_{range} is significantly negatively associated for 12-16 days prior to detection. Following the observations over the selected cities located in different geographical parts of the country, it is uncertain whether the higher humidity could reduce the infectivity of the coronavirus by reducing the suspension time of virus. This suggests that the role of humidity is quite complex and needs to be investigated further. On the other hand, all the temperature parameters (T_{max}, T_{min}, T_{mean}, T_{Dew}) are proportionately associated with COVID-19 transmission. The analysis also indicates that T_{max}, T_{min}, T_{mean}, T_{Dew} and WS on the day of the detection have the lowest correlations, which improves up to its peak at a time lag of 14 days. In other words, the maximum, minimum, mean and dew point temperature along with wind speed at 14 days prior to detection are closely allied with the number of infections. This suggests the interesting inference that weather conditions 14 days prior to the detection of infections had provided favorable conditions for virus transmutability. Surprisingly, perhaps, T_{min} is found to be better related than T_{mean} T_{max} T_{Dew}. Therefore, places with higher minimum temperature are more susceptible to COVID-19 transmission in India. A significant positive correlation between WS and daily transmission at a lag of 14 days infers that the virus might be able to transmigrate in high winds. Since most of the weather parameters are better correlated with the daily confirmed cases at a time lag of 14 days, this indicates an approximate incubation period of around 14 days for this disease in the Indian scenario. Therefore, considering the lag period of 14 days, the correlation analysis for each selected city (Table 3) shows that cities located away from the coast such as Delhi, Indore and Jaipur have better association with temperature parameters (T_{max}, T_{min}, T_{mean}, T_{Dew}) than the coastal cities such as Mumbai, Chennai and Kolkata. However, WS is relatively better correlated with COVID-19 cases in coastal cities than in the other cities. Interestingly, RH_{Avg} and RH_{Range} are also significantly related with COVID-19 cases in coastal cities only, while cities located in the interior didn't exhibit any significant correlation. That's why on the country-wide scale, correlations between COVID-19 cases and RH parameters were reporting as non-significant. Hence, a higher humidity with a higher wind speed could be favourable for virus transmissibility; while a higher temperature might favor virus transmission in semi-arid and interior areas. This also suggests that the geographical location of the cities plays a crucial role in the association of weather parameters with COVID transmission, which makes this interrelationship even more complex.

Table 2. Result of Spearman's correlation test in different time frames for all over India.

Table 3. Result of Spearman's correlation test in each cities considering a lag of 14 days only.

Figure 4. The daily trend of confirmed case in selected cities are shown. Inset is a scatter graph depicting the growth rate of transmission with respect to the population and elevation of those cities.

Figure 5 and Table 4 show the validation of estimated daily confirmed cases for all time spans using the non-linear multivariate Support Vector Regression Model with RBF kernel. The model performance in terms of R^2, RMSE, and MB are presented in Table 4. This shows that the SVM-based regression model is very efficient in establishing the complex relationship between the different weather parameters and the daily transmission of COVID-

19. However, it underestimates the prevalence of very high values (>1200 cases). Hence, we may conclude that no single weather parameter is enough to linearly correlate with the daily transmission. Instead, a non-linear multivariate approach is needed to estimate the daily transmission in India with high accuracy. Correlation analysis has evidently revealed a relatively higher degree of association with daily new cases for most of the parameters only when the time lag of 14 days is taken into consideration. The SVM based regression model also performs remarkably well with a time lag of 12 to 16 days. Together, these suggest a typical incubation period of 12 - 16 (14 ± 02) days for COVID-19 transmission in India. In order to better understand the influence of varying weather conditions, the response curves of each significant parameter to the cumulative confirmed cases were calculated and are shown in Figure 6, which reveals that cumulative COVID cases are very high only during a certain range of temperature and wind speed. Approximately 75% of the total confirmed cases were registered when the T_{Max}, T_{Mean}, T_{Min}, T_{Dew}, and *WS* parameters 12 - 16 days previously vary within a range of 33.6 - 41.3° C, 29.8 - 36.5° C, 24.8 - 30.4° C, 18.7 - 23.6° C, and 4.2 - 5.75 m/s respectively. Briefly, over the selected cities, there is quite large variation in daily weather conditions, hence, there is a broad range in the values of each weather parameter, among which a narrow range is highly favorable for virus transmission, as revealed by this study. This also suggets why the linear univariate correlation was so low for each of the parameters. Together, these give an idea of the favorable weather conditions for such transmission in India. In other words, the areas in India that experienced such weather patterns are those most likely to have been affected by this disease.

Figure 5. Validation of SVM based regression model for estimating daily transmission.

Table 4. Result of Validation of SVM based regression for estimating daily transmission.

Figure 6. Influence of weather parameters on count of confirmed cases with a lag of 12-16 days.

4. Conclusions

Unlike most studies, the present study investigated the impact of various weather parameters which include maximum, minimum, mean, and dew point temperature, temperature range, average humidity, humidity range and wind speed on the same day, as well as with time-lags of 7, 10, 12, 14, and 16 days prior to detection of the confirmed cases of COVID-19 in the Indian context. Additionally, the daily trends of confirmed cases in 9 of the most affected cities in India, along with a comparison of the entire country, have also been inspected in this study. The analyses revealed that the count of confirmed cases is not well correlated with any individual meteorological parameter because simple correlation depicts a linear relationship only. Rather than that, COVID-19 cases are significantly associated with a very certain range of temperature parameters and wind speed. Thus, much better than linear correlation, the non-linear SVM-based regression approach efficiently resolved this complex association and was able to estimate the daily cases of infection quite accurately with the help of the daily weather inputs. However, the positive correlation between daily transmission and air temperature, as well as wind speed, indicates that the daily transmission in highly populated areas in India has consequently increased during the current summer days of 2020. An approximate incubation period of 14 ± 02 days can also be identified from the data, which is a little longer than what WHO had estimated early in March. Therefore, in the prevailing weather conditions in India, the SARS-CoV-2 can be disseminated into the surrounding environment for around two weeks after being ingested from any other infected source.

The COVID-19 pandemic has resulted in a state of recrudescence in India. The daily confirmed cases have been rising at an acceleration rate of ~ 76 ANC/day since March 2, 2020 with a doubling rate of 7.85 days. This rate of acceleration all over the country reached approximately 249 ANC/day during the starting of June. Initially, 14 out of each 1000 tests revealed positive results during the first week of March, but the positive test rate escalated to 71/1000 tests in the first week of June. On the other hand, reduced strictness in subsequent phases of lockdowns, along with the allowing of interstate migration, had inevitably caused an easy pathway for transmission, hence resulting in an intractable circumstance all over the country. The cities with larger populations are cataloguing a higher rate of increase in daily cases. Moreover, a step-change in the rising trend over all the major affected cities has also been noted during mid-April, i.e., at the boundary between the first and second lockdowns. This signifies that the imposed lockdown was unsuccessful in reducing the COVID-19 transmission in India, unlike in e.g., South Korea, Japan, and Iran. Nonetheless, this study has limitations, since we were unable to include many other major affected cities due to lack of meteorological data availability. Moreover, the number of immigrants from abroad or other cities who were quarantined was not available; these might have enhanced the exactitude of the current analysis.

Funding: This research was supported by the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and Information Technology, in the University of Technology Sydney (UTS).

Conflicts of Interest: The authors declare no conflict of interest.

References

Ahmadi M, Sharifi A, Dorosti S, Ghoushchi SJ, Ghanbari N (2020) Investigation of effective climatology parameters on COVID-19 outbreak in Iran, Science of the Total Environment:138705

Bai Y, Yao L, Wei T, Tian F, Jin D-Y, Chen L, Wang M (2020) Presumed asymptomatic carrier transmission of COVID-19, Jama 323:1406-1407

Bedford T, Riley S, Barr IG, Broor S, Chadha M, Cox NJ, Daniels RS, Gunasekaran CP, Hurt AC, Kelso A, Klimov A, Lewis NS, Li X, McCauley JW, Odagiri T, Potdar V, Rambaut A, Shu Y, Skepner E, Smith DJ, Suchard MA, Tashiro M, Wang D, Xu X, Lemey P, Russell CA (2015) Global circulation patterns of seasonal influenza viruses vary with antigenic drift, Nature 523, 217–220.

Bonasera A, Zhang S (2020) Chaos, Percolation and the Coronavirus Spread, Frontiers in Physics 8:171

Briz-Redón Á, Serrano-Aroca Á (2020) A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain, Science of the Total Environment:138811

Bull G (1980) The weather and deaths from pneumonia, The Lancet 315:1405-1408

Casanova LM, Jeon S, Rutala WA, Weber DJ, Sobsey MD (2010) Effects of air temperature and relative humidity on coronavirus survival on surfaces, Applied and environmental microbiology 76:2712-2717

Census of India Website : Office of the Registrar General & Census Commissioner, India Available online: https://censusindia.gov.in/2011-common/census_2011.html (accessed on Jun 6, 2020).

Dalziel BD, Kissler S, Gog JR, Viboud C, Bjørnstad ON, Metcalf CJE, Grenfell BT (2018) Urbanization and humidity shape the intensity of influenza epidemics in US cities, Science 362:75-79

Ficetola GF, Rubolini D (2020) Climate affects global patterns of COVID-19 early outbreak dynamics, medRxiv

Gao P, Mu X-M, Wang F, Li R (2011) Changes in streamflow and sediment discharge and the response to human activities in the middle reaches of the Yellow River, Hydrology and Earth System Sciences 15:1-10

Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, Rinaldo A (2020) Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures, Proceedings of the National Academy of Sciences 117:10484-10491

Ge X-Y et al. (2013) Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor, Nature 503:535-538

Guan W-j et al. (2020) Clinical characteristics of coronavirus disease 2019 in China, New England journal of medicine 382:1708-1720

Gupta A (2017) An update of Temperature trends at major cities of India and their zonal analysis, Unpublished thesis, Savitribai Phule Pune University.

Gupta S, Raghuwanshi GS, Chanda A (2020a) Effect of weather on COVID-19 spread in the US: a prediction model for India in 2020, Science of The Total Environment:138860

Gupta A, Banerjee S, Das S (2020b) Significance of geographical factors to the COVID-19 outbreak in India, Modeling Earth Systems and Environment:1-9

Gupta A, Bhatt CM, Roy A, Chauhan P (2020c) COVID-19 lockdown a window of opportunity to understand the role of human activity on forest fire incidences in the Western Himalaya, India, Current Science 119 (2): 390-398

Gupta A, Banerjee S, Das S (2020d) Significance of geographical factors (climatic, topographic and social) to the COVID-19 outbreak in India. OSF Preprint.

Hänsel S, Medeiros DM, Matschullat J, Petta RA, de Mendonça Silva I (2016) Assessing homogeneity and climate variability of temperature and precipitation series in the capitals of north-eastern Brazil, Frontiers in Earth Science 4:29

Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, Spitters C, Ericson K, Wilkerson S, Tural A, Diaz G, Cohn A, Fox L, Patel A, Gerber SI, Kim L, Tong S, Lu X, Lindstrom S, Pallansch MA, Weldon WC, Biggs HM, Uyeki TM, Pillai SK (2020) First case of 2019 novel coronavirus in the United States, New England Journal of Medicine 382:929-936

Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, The lancet 395:497-506

Jaiswal R, Lohani A, Tiwari H (2015) Statistical analysis for change detection and trend assessment in climatological parameters, Environmental Processes 2:729-749

Jamil T, Alam I, Gojobori T, Duarte CM (2020) No evidence for temperature-dependence of the covid-19 epidemic, Cold Spring Harbor Laboratory 1-11.

Jebur MN, Pradhan B, Tehrany MS (2014) Manifestation of LiDAR-derived parameters in the spatial prediction of landslides using novel ensemble evidential belief functions and support vector machine models in GIS IEEE, Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8:674-690

Kang D, Choi H, Kim J-H, Choi J (2020) Spatial epidemic dynamics of the COVID-19 outbreak in China, International Journal of Infectious Diseases 94:96-102

Kendall MG (1975) Rank Correlation Methods. London: Charles Griffin & Co. Ltd,

Lemaitre J, Pasetto D, Perez-Saez J, Sciarra C, Wamala JF, Rinaldo A (2019) Rainfall as a driver of epidemic cholera: comparative model assessments of the effect of intra-seasonal precipitation events, Acta tropica 190:235-243

Leung K, Wu JT, Liu D, Leung GM (2020) First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment, The Lancet 395: 1382-1393

Liu J, Zhou J, Yao J, Zhang X, Li L, Xu X, He X, Wang B, Fu S, Niu T, Yan J, Shi Y, Ren X, Niu J, Zhu W, Li S, Luo B, Zhang K (2020) Impact of meteorological factors on the COVID-19 transmission: A multi-city study in China, Science of the Total Environment:138513

Ma Y, Zhao Y, Liu J, He X, Wang B, Fu S, Yan J, Niu J, Zhou J, Luo B (2020) Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China, Science of The Total Environment:138226

Mallakpour I, Villarini G (2016) A simulation study to examine the sensitivity of the Pettitt test to detect abrupt changes in mean, Hydrological Sciences Journal 61:245-254

Mann HB (1945) Nonparametric tests against trend Econometrica, Journal of the econometric society 245-259

Marjanović M, Kovačević M, Bajat B, Voženílek V (2011) Landslide susceptibility assessment using SVM machine learning algorithm, Engineering Geology 123:225-234

Mollalo A, Vahedi B, Rivera KM (2020) GIS-based spatial modeling of COVID-19 incidence rate in the continental United States, Science of The Total Environment:138884

Oliveiros B, Caramelo L, Ferreira NC, Caramelo F (2020) Role of temperature and humidity in the modulation of the doubling time of COVID-19 cases, medRxiv

Organization WH (2020) Coronavirus disease 2019 (COVID-19): situation report, 72

Perlman S (2020) Another decade, another coronavirus. The New England Journal of Medicine, Mass Medical Society 382:760-762

Pettitt A (1979) A non-parametric approach to the change point problem, Applied Statistic 28:126-135

Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS, Computers & Geosciences 51:350-365

Qi H, Xiao S, Shi R, Ward MP, Chen Y, Tu W, Su Q, Wang W, Wang X, Zhang Z (2020) COVID-19 transmission in Mainland China is associated with temperature and humidity: A time-series analysis, Science of the Total Environment:138778

Rocklöv J, Sjödin H (2020) High population densities catalyse the spread of COVID-19, Journal of travel medicine 27:taaa038

Sen P (1968) Estimates of the regression coefficient based on Kendall's tau, Journal of the American Statistical Association 63:1379-1389

Shi P, Dong Y, Yan H, Zhao C, Li X, Liu W, He M, Tang S, Xi S (2020) Impact of temperature on the dynamics of the COVID-19 outbreak in China, Science of The Total Environment:138890

Song S, Zhan Z, Long Z, Zhang J, Yao L (2011) Comparative study of SVM methods combined with voxel selection for object category classification on fMRI data, PloS one 6:e17191

Sooryanarain H, Elankumaran S (2015) Environmental role in influenza virus outbreaks, Annual Review of Animal Biosciences 3:347-373

Tamerius JD, Shaman J, Alonso WJ, Bloom-Feshbach K, Uejio CK, Comrie A, Viboud C (2013) Environmental predictors of seasonal influenza epidemics across temperate and tropical climates, PLoS Pathog 9:e1003194

Tan J, Mu L, Huang J, Yu S, Chen B, Yin J (2005) 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 59:186-192

Tehrany MS, Pradhan B, Jebur MN (2014) Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS, Journal of hydrology 512:332-343

Tehrany MS, Pradhan B, Mansor S, Ahmad N (2015) Flood susceptibility assessment using GIS-based support vector machine model with different kernel types, Catena 125:91-101

Tien Bui D, Pradhan B, Lofman O, Revhaug I (2012) Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and Naive Bayes Models, Mathematical problems in Engineering 2012: 1-26

Tosepu R, Gunawan J, Effendy DS, Lestari H, Bahar H, Asfian P (2020) Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia, Science of The Total Environment:138436

Wang C, Horby PW, Hayden FG, Gao GF (2020a) A novel coronavirus outbreak of global health concern, The Lancet 395:470-473

Wang M, Jiang A, Gong L, Luo L, Guo W, Li Chuyi, Zheng J, Li Chaoyong, Yang B, Zeng J, Chen Y, Zheng K, Li H (2020b) Temperature significant change COVID-19 Transmission in 429 cities. MedRxiv

Wang Y, Wang Y, Chen Y, Qin Q (2020c) Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID- 19) implicate special control measures, Journal of medical virology 92:568-576

Wijngaard J, Klein Tank A, Können G (2003) Homogeneity of 20th century European daily temperature and precipitation series International Journal of Climatology: A Journal of the Royal Meteorological Society 23:679-692

Wu Z, McGoogan JM (2020) Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention Jama 323:1239-1242

Wurtzer S, Marechal V, Mouchel J-M, Maday Y, Teyssou R, Richard E, Almayrac JL, Moulin L (2020) Evaluation of lockdown impact on SARS-CoV-2 dynamics through viral genome quantification in Paris wastewaters, medRxiv

Xie J, Zhu Y (2020) Association between ambient temperature and COVID-19 infection in 122 cities from China, Science of The Total Environment 724:138201

Yao X, Tham L, Dai F (2008) Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China Geomorphology 101:572-582

Yuan J, Yun H, Lan W, Wang W, Sullivan SG, Jia S, Bittles AH (2006) A climatologic investigation of the SARS-CoV outbreak in Beijing, China American journal of infection control 34:234-236

Zhu N, Zhang D, Wang W China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*, *382*(8), pp.727-733.

Zu ZY, Jiang MD, Xu PP, Chen W, Ni QQ, Lu GM, Zhang LJ (2020) Coronavirus disease 2019 (COVID-19): a perspective from China Radiology:200490

Table 1. Result of Mann-Kendall test, Sen's Slope, Pettit test, Grow rate and Doubling time.

Cities	M-K Tau	Sen's Slope	Change Point	Slope before Change Point	Slope after Change Point	Growth Rate (%)	Doubling Time (days)
Ahmedabad	0.90*	7.77*	17-04-2020	0.46*	12.75*	8.85	8.31
Chennai	0.89*	8.73*	18-04-2020	0.91*	31.28*	8.89	7.57
Delhi	0.87*	10.85*	22-04-2020	1.78*	34.86*	8.67	8.53
Hyderabad	0.63*	1.47*	31-03-2020	0.51*	1.77*	5.87	12.68
Indore	0.70*	1.5*	14-04-2020	0.43*	1.49*	6.43	11.04
Jaipur	0.70*	1*	09-04-2020	0.19*	1.08*	5.67	12.25
Kolkata	0.82*	1.3*	17-04-2020	0.25*	1.43*	6.38	11.34
Mumbai	0.88*	19.08*	18-04-2020	2.79*	38.2*	9.98	7.31
Pune	0.83*	3.51*	12-04-2020	0.52*	10.63*	7.95	9.75
All over India	0.95*	76.11*	17-04-2020	21.55*	173.54*	10.79	7.85

Table 2. Result of Spearman's correlation test in different time frames for allover India.

Parameters	On that Day	7 Days ago	10 Days ago	12 Days ago	14 Days ago	16 Days ago
Maximum Temperature	0.161*	0.198*	0.231*	0.336*	0.347*	0.272*
Minimum Temperature	0.244*	0.285*	0.319*	0.417*	0.436*	0.351*
Mean Temperature	0.199*	0.248*	0.287*	0.409*	0.430*	0.337*
Temperature Range	-0.032	-0.043	-0.066	-0.064	-0.075	-0.054
Dew Point Temperature	0.222*	0.235*	0.261*	0.238*	0.269*	0.238*
Relative Humidity	0.128*	0.10*	0.10*	0.002	0.008	0.034
Humidity Range	-0.016	-0.024	-0.056	-0.13*	-0.149*	-0.095
Wind Speed	0.108*	0.133*	0.163*	0.221*	0.255*	0.193*

Table 3. Result of Spearman's correlation test in each cities considering a lag of 14 days only.

Parameters	Ahmedabad	Chennai	Delhi	Hyderabad	Indore	Jaipur	Kolkata	Mumbai	Pune
Maximum Temperature	0.46*	0.23*	0.51*	0.31*	0.53*	0.51*	0.15*	0.22*	0.27*
Minimum Temperature	0.51*	0.38*	0.56*	0.34*	0.58*	0.55*	0.17*	0.29*	0.29*
Mean Temperature	0.45*	0.29*	0.50*	0.33*	0.51*	0.54*	0.14*	0.28*	0.23*
Temperature Range	-0.06	-0.19*	-0.07	-0.08	-0.07	-0.13	-0.26*	-0.23*	-0.12
Dew Point Temperature	0.27*	0.25*	0.30*	0.28*	0.29*	0.17*	0.24*	0.26*	0.25*
Relative Humidity	-0.18	0.21*	-0.39*	-0.50*	-0.13	-0.20	0.45*	0.46*	-0.17
Humidity Range	-0.07	-0.18*	-0.11	-0.08	-0.11	-0.12	-0.24*	-0.22*	-0.09
Wind Speed	0.32*	0.49*	0.21*	0.20*	0.13*	0.17*	0.32*	0.37*	0.24*

Table 4. Result of Validation of SVM based regression for estimating daily transmission.

	R^2	RMSE	MB
On that Day	0.6414	199.2929	-48.1804
7 Days ago	0.7015	202.1743	-40.0377
10 Days ago	0.8286	223.1949	-42.0560
12 Days ago	0.8503	186.0126	-66.9880
14 Days ago	0.8680	178.3891	-43.6459
16 Days ago	0.8714	202.2428	-60.0658

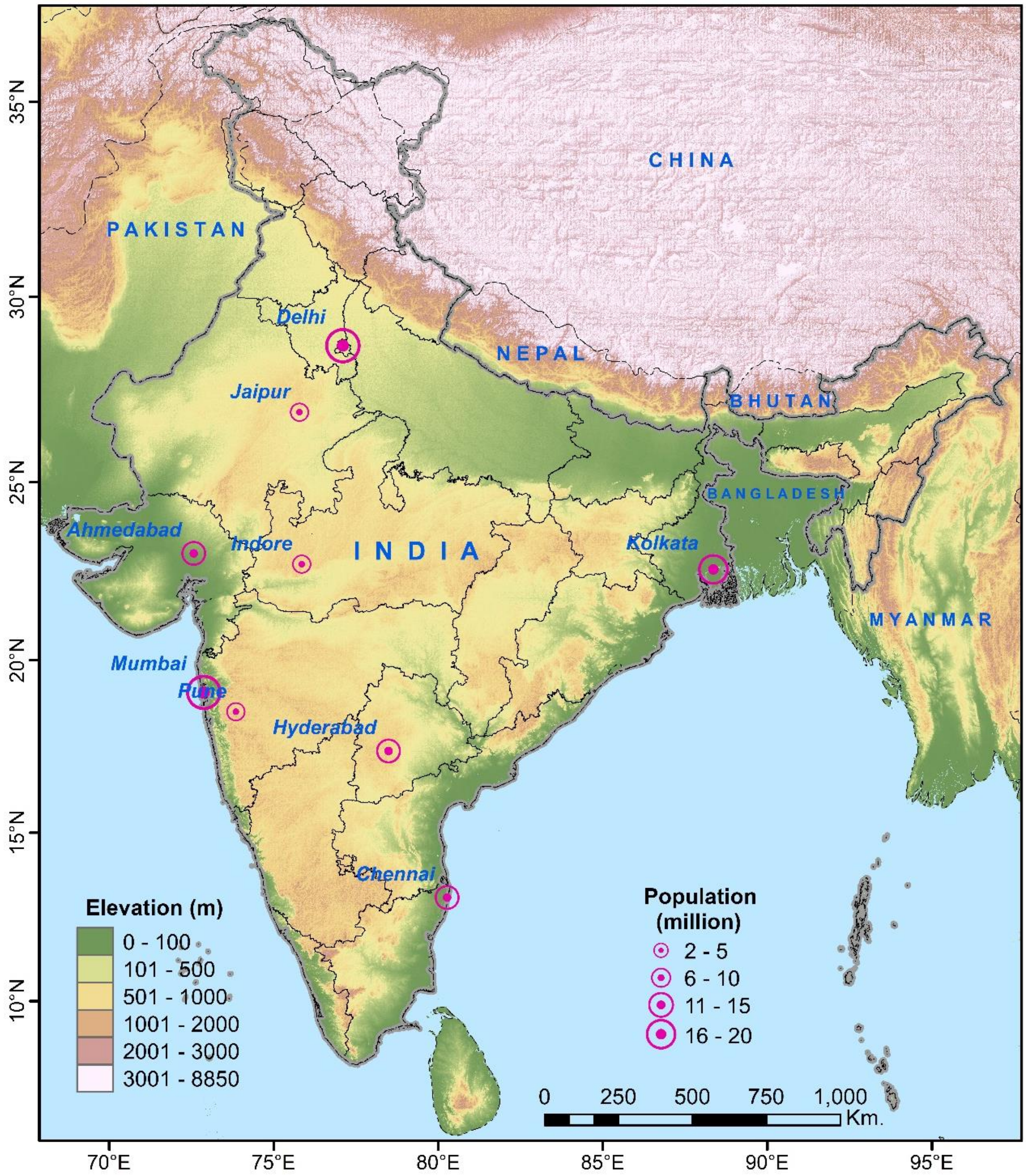

Figure 1. Location of the selected cities in India along with the total population of those cities.

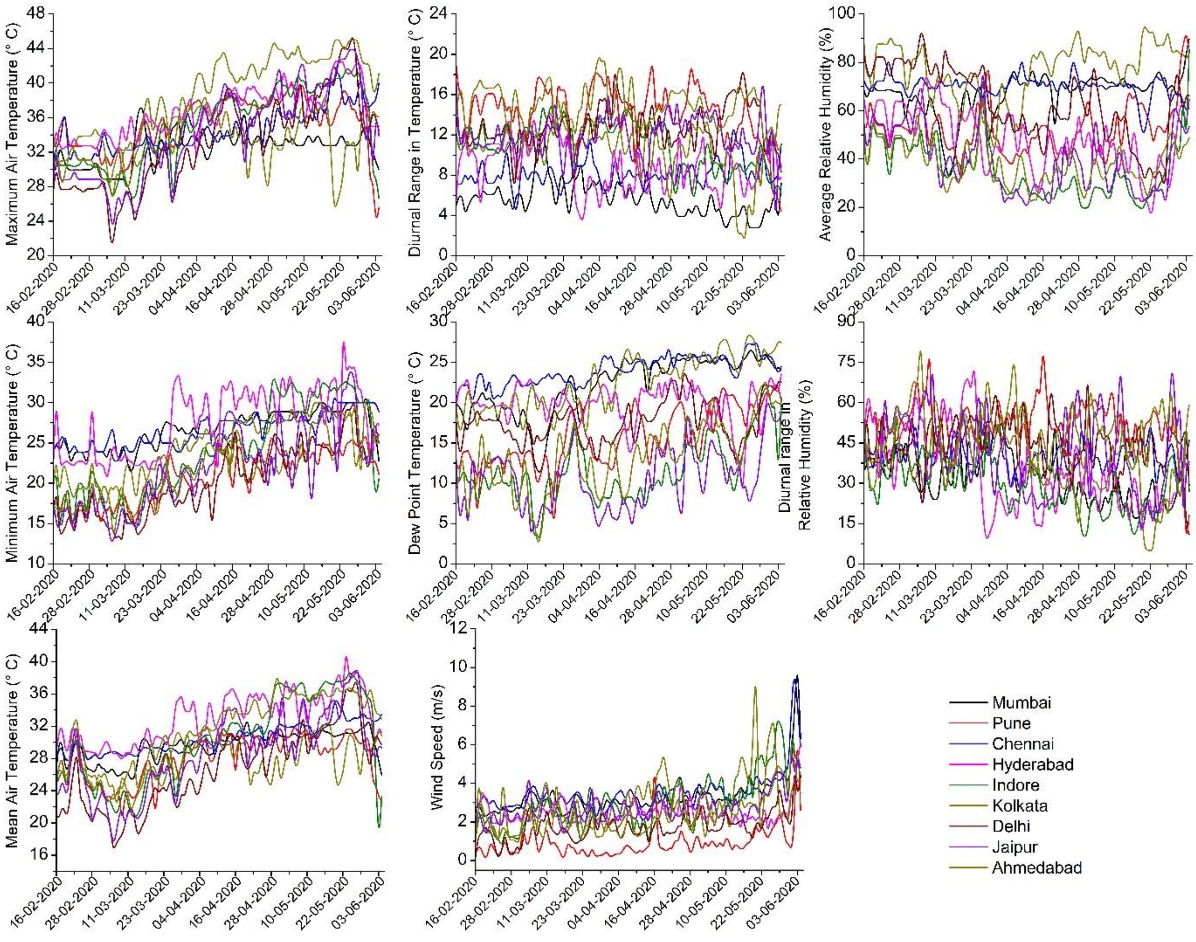

Figure 2. Pattern of daily weather over the selected cities in India.

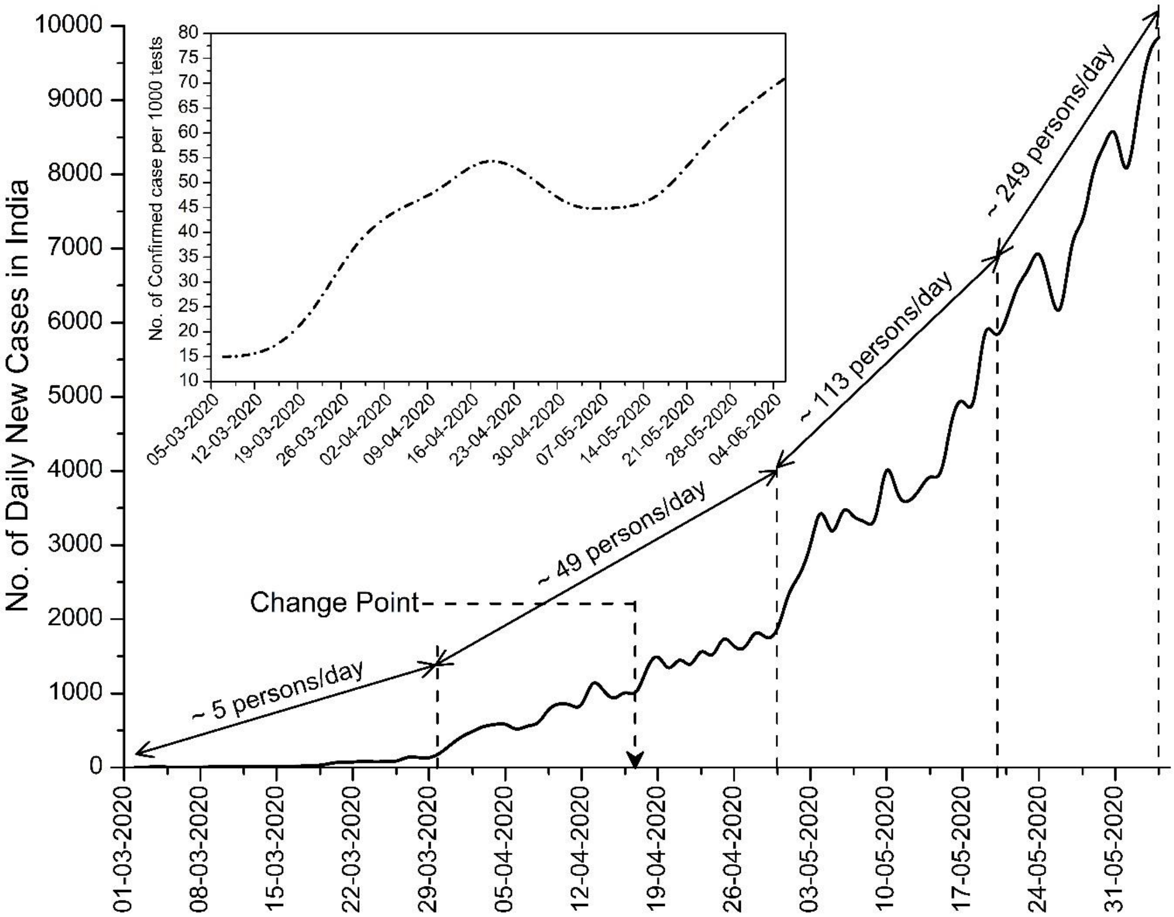

Figure 3. Trend of daily confirmed cases in India. The weekly trend of number of confirmed cases per 1000 tests are shown inset.

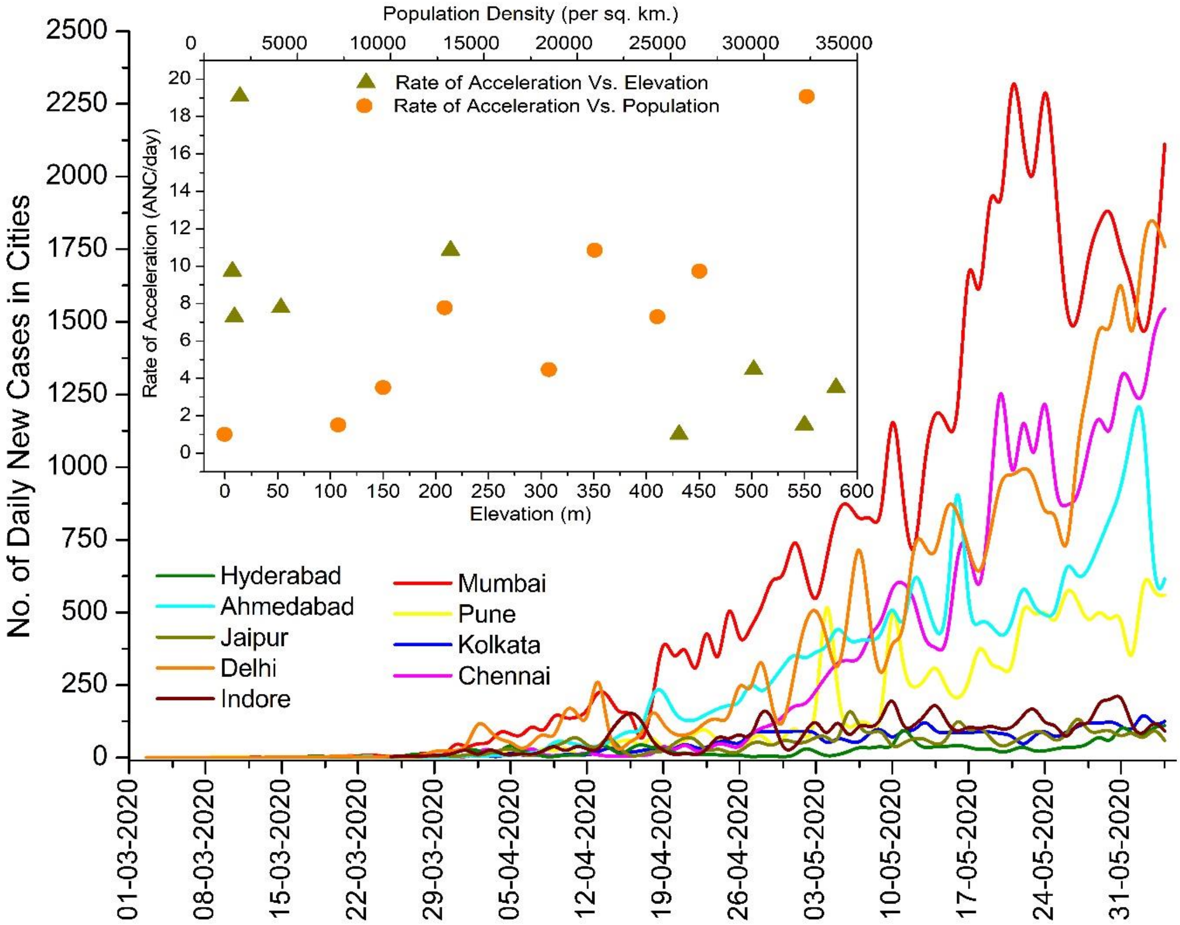

Figure 4. The daily trend of confirmed case in selected cities are shown. Inset is a scatter graph depicting the growth rate of transmission with respect to the population and elevation of those cities.

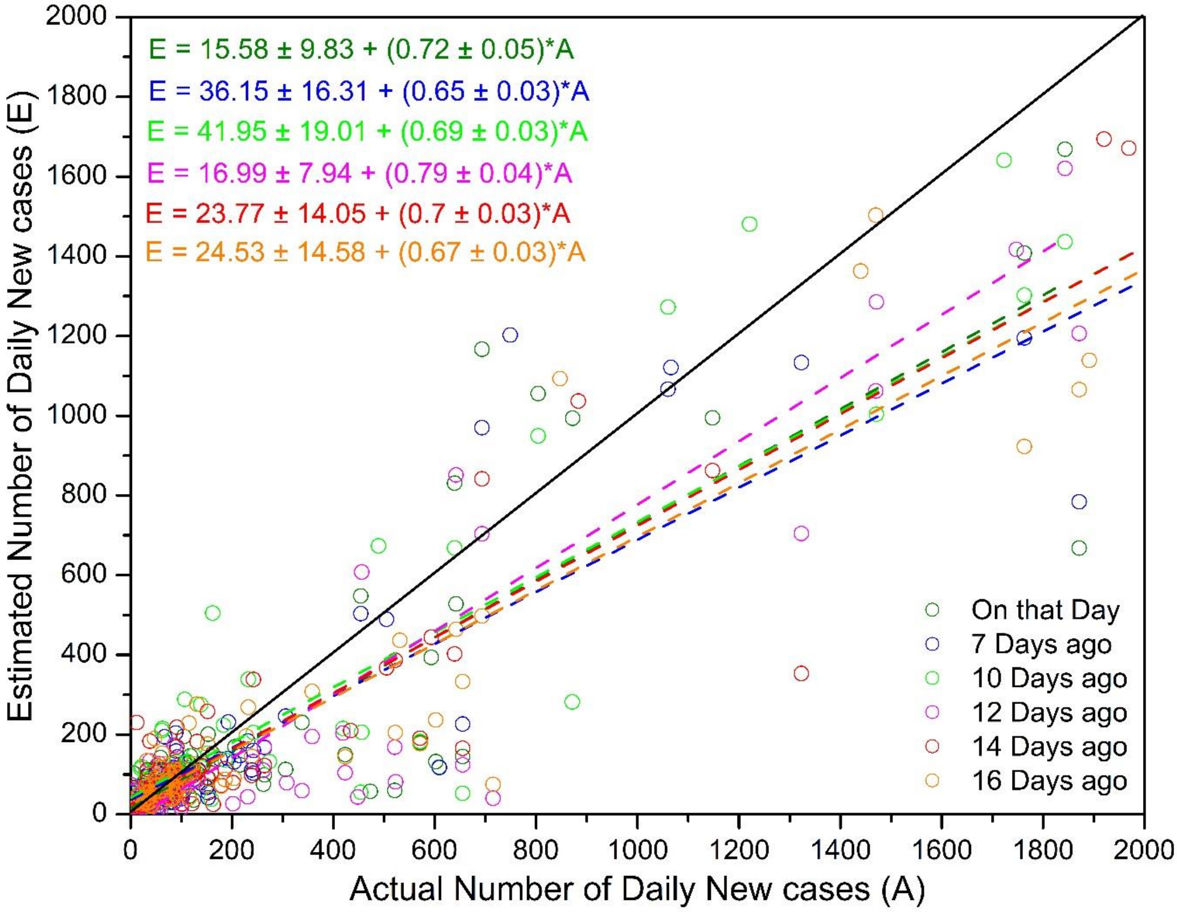

Figure 5. Validation of SVM based regression model for estimating daily transmission.

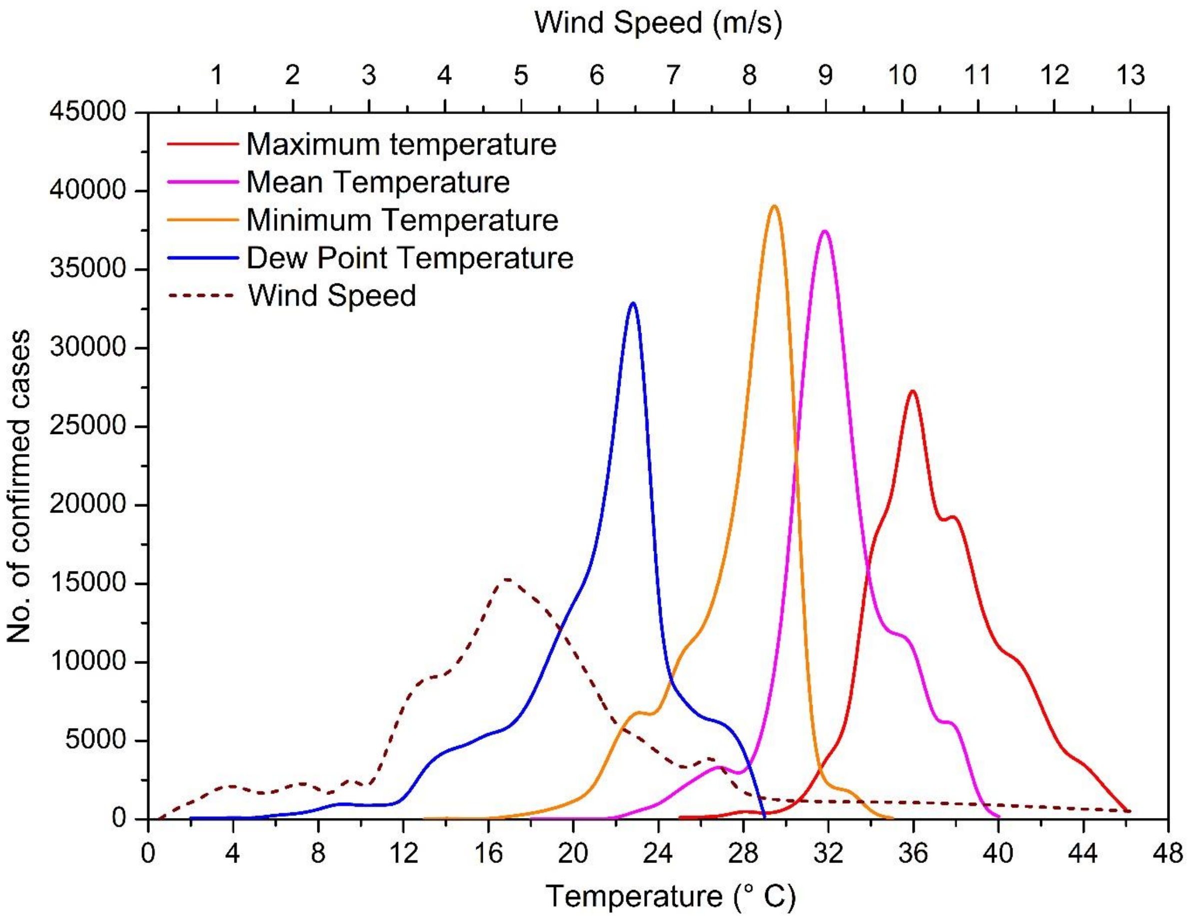

Figure 6. Influence of weather parameters on count of confirmed cases with a lag of 12-16 days.