Statistical analysis of active fire remote sensing data: examples from South Asia

Jyoti U. Devkota ⬤

Abstract   Active fires emit aerosols and greenhouse gases in the atmosphere. In this paper, the behavior of active fires over a period of 1 year in Nepal, Bhutan, and Sri Lanka is studied using spatial statistics. In these countries, these fires are mainly forest and vegetation fires; they wreak havoc to the environment by damaging flora and fauna and emitting toxic gases. This study is based on data acquired through remote sensing of data acquisition platform, NASA's MODIS. Spatial statistics is used here to study the incidence of such fires with respect to geographical location. The behaviors of parameters of various autoregressive models like Spatial Durban Model, Spatial Lag Model, Spatial Error Model, Manski Model, and Kelegian Prucha Model are minutely analyzed. The best model with the highest pseudo R^2 is selected. The spatial behavior of the fire radiative power (FRP) for the three countries is also predicted using spatial interpolation and kriging. The burning potential of vegetations in unsampled areas is envisaged by thus predicting FRP. This study gives a country-wise perspective to the behavior of fire; this is with reference to South Asia. It holds a great significance for countries of the developing world which lack a strong backbone of good-quality official records. Through the statistical analyses of data collected by such platforms, important information on impact of forest fires can be indirectly assessed.

Introduction

Spatial data play an important role in today's world. These data are collected with reference to geographical locations. The spatial data of active fire are collected by various satellites. The sensors in these satellites are used to map area burned and assess characteristics of active fire. The effects on the ecology after these fires subside are also characterized with these sensors. Different space and airborne sensors that have been used to assess fire behavior are discussed in detail by Lentile et al. Changes in the environment before and during these fires can be detected by these sensors. Post-fire spectral response can also be assessed by these devices (Lentile et al., 2006).

Active fires in South Asia are primarily vegetation fires. There are many reasons for these fires. Agriculture residues such as straw, stalks, and husks are burnt by farmers. This is also done to clear the land for agriculture for next season. Although this has been one of the major contributors to air pollution of South Asia, it is thought to be the best and least expensive method for land clearing. This method has been thought to promote the growth of grasses in the

J. U. Devkota (✉)
Department of Mathematics, Kathmandu University,
Dhulikhel, Nepal
e-mail: drjdevkota@ku.edu.np

farms. It is also thought to boost agriculture and timber produce. But some fires are also caused by negligence and ignorance of the smokers and passersby. Hunters, poachers, grazers, and non-timber product collectors also deliberately set these forests on fire (Kunwar & Khaling, 2006).

Forest fires are one of the major reasons for forest degradation (Matin & Chitale, 2017). Reddy and Bird et al. claim that in recent years in South Asia: "51% of forest grid cells were affected by fires." Tropical moist deciduous forest and tropical dry deciduous forest have the highest incidences of forest fires (Reddy et al., 2019). Vadrevu et al. concluded that in South Asia, India had the highest number of annual fires followed by Pakistan and others. They also found that Nepal (82.84%) and Bhutan (75.56%) had the highest percentage of human-initiated forest fires (Vadrevu et al., 2019).

Many studies on fire hazards have been conducted. For example, Sakellariou et al. explored the variability of fire hazard in the Greek island of Skiathos and performed a spatiotemporal analysis (Sakellariou et al., 2020). Marin et al. studied the behavior of these forest fires in Mexico by using "georeferenced fire records" for the period of 2005–2015. The spatial and temporal relationships were examined with a "multiscalar drought index, the Standardized Precipitation-Evapotranspiration Index (SPEI)" (Marin et al., 2018). Su et al. used a geostatistical approach integrated with machine learning. It was used to improve the mapping accuracies of aboveground biomass in the northern Guangdong Province of China (Su et al., 2020). Eskandari et al. assessed quantitative temporal relationships using correlation and regression analyses. Statistically significant relationships are identified and described using climatic data from the Golestan Meteorological Administration and Fire Statistics, Iran. Spatial relationships between climatic variables and fire occurrence are also determined (Eskandari et al., 2020).

Statistics is an evidence-based and data-based approach to handling an issue. These issues can come from diverse fields. Today's world is the world of big data. Satellite data are also a source of big data. Statistical methods have a scope of vast interdisciplinary applications. For example, Johanna et al. have demonstrated benefits of integrating geostatistical covariance structure and ANOVA structure to linear mixed modeling framework. Examples from soil sciences are taken to demonstrate their findings (Johanna et al., 2020). Similarly, Bhunia et al. have used geostatistical models to study the spatial variability of lateritic soils. Soil nitrogen, pH, electrical conductivity, phosphorus, potassium, and organic carbon were measured. Surface maps of soil properties were prepared using the semivariogram model through Kriging techniques (Bhunia et al., 2018). Similarly, multivariate geostatistics has been used at the bus stop level on public transportation demand modeling (Marques & Pitombo, 2021). And non-linear geostatistical models are seen to be ideal in identification of geological and geometrical complexity of gold deposits (Afonseca & Costa, 2021).

This paper tries to fill the knowledge gap through a detailed statistical study of intensity of vegetation fire. Intensity of fire, also called fire energy output, is measured by fire radiative power (FRP). This intensity is also represented by brightness. Geostatistical analysis using variograms is also conducted here. Here, spatial correlation is estimated and modeled. In addition to this, this paper also aims to understand and predict the behavior of active fire in South Asia with the help of multivariate statistics. It also delves into the structural relationship between the variables. Bhutan, Nepal, and Sri Lanka are taken as examples from this region. Using active fire 1-year satellite data, it analyzes the country-wise behavior with respect to FRP and brightness. It also tries to explain the variogram of FRP with a mathematical model. Then, it predicts FRP of unsampled areas of vegetation using this model. This helps in projection of FRP of vegetation in unsampled areas if they caught fire.

The arrangement of this paper is in the following manner. This section is followed by the "Materials and methods" section, then by the section "Results and discussion." This is followed by the "Conclusions" section.

Materials and methods

Data

This study is based on active fire satellite data for three countries. These three countries are Bhutan, Nepal, and Sri Lanka. These are observations made by NASA's Moderate Resolution Imaging

Spectroradiometer (MODIS) satellite, from 18 September 2019 to 17 September 2020. This satellite noted 189, 982, and 441 cases of active fires in Bhutan, Nepal, and Sri Lanka, respectively, during this 1-year period.

Aboard Terra and Aqua satellites, MODIS is the main instrument. Terra orbits the earth around morning by passing north to south across the equator. Aqua orbits in the afternoon. "Terra MODIS and Aqua MODIS are viewing entire earth surface every 1 to 2 days. They acquire data in 36 spectral bands or groups of wave lengths" (NASA, 2021).

"NASA's FIRMS give the global fire locations (hotspots)." NASA's FIRMS is an active fire locations data which represent the midpoint pixel measuring 1 km×1 km. It extracts from the MODIS Image using the thermal anomalies algorithms. FIRMS is part of NASA's Earth Observing System Data and Information System (EOSDIS). EOSDIS and twelve other Distributed Active Archive Centers (DAACs) provide access to data from NASA's Earth Science Missions (Fithria A & Ani A, 2017).

A comparison between data collected through a field survey of the Korea Forest Service (KFS) and satellite active fire data of MODIS was done by Lim et al. Examination of the spatial autocorrelation and related factors by fire source indicated that MODIS data had higher spatial autocorrelation. These results were found to be highly significant with respect to climate factors. KFS data were collected from post-fire surveys; they resulted in low spatial autocorrelation and reduced model accuracy owing to the wide distribution of data (Lim et al., 2019).

In this paper, the following statistical methods are used.

Spatial statistics

An important concern in here is to examine spatial patterning and spatial dependence among variables of interest. This means that values close together in space tend to be more familiar than those that are further apart (Lloyd, 2010).

Spatial autocorrelation

A key tool for the analysis of spatial autocorrelation is Moran's I coefficient. It measures spatial autocorrelation with neighbors of observations that are classified into various contiguity schemes.

Regression

An assumption of standard ordinary least squares regression is the independence of observations. But this assumption rarely holds true for spatial data. Spatial autoregressive models provide a means for accounting the spatial structure of the data. The different autoregressive models are namely (a) Spatially Lagged X model, (b) Spatial Error model, (c) Spatial Durban Error model, (d) Spatially Lagged Y model, (e) Spatial Durban model, (f) Manski All Inclusive model, and (g) Kelegian Prucha model.

Variogram

It is used to measure the spatial correlation. Here, the word variogram is synonymous with semivariogram. It plots a semivariogram as a function of distance. Here, a semivariogram is defined mathematically as follows (Bivand et al., 2013):

$$\gamma(\mathbf{h}) = \frac{1}{2}\mathbf{E}(\mathbf{Z(s)} - \mathbf{Z(s+h)})^2 \qquad (1)$$

where, under the assumption of intrinsic stationarity,

$$Z(s) = m + e(s) \qquad (2)$$

$$E(Z(s)) = m \qquad (3)$$

Here, $Z(s)$ is the observation of a variable at location s.

Under the assumption that a semivariogram can be estimated from N_h sample data pairs $Z(s_i)$, $Z(s_i + h)$ for a number of distances (or distance intervals) h_j by

$$\widehat{\gamma}\left(\tilde{\mathbf{h_j}}\right) = \frac{1}{2\mathbf{N_h}}\sum_{\mathbf{i=1}}^{\mathbf{N_h}} \mathbf{E}(\mathbf{Z(s)} - \mathbf{Z(s+h)})^2 \forall \mathbf{h_j} \in \tilde{\mathbf{h_j}} \qquad (4)$$

Interpolation and ordinary kriging

Geostatistics deals with the analysis of random fields $Z(s)$, with Z random, and s the non-random spatial index. Typically, at a limited number of sometimes
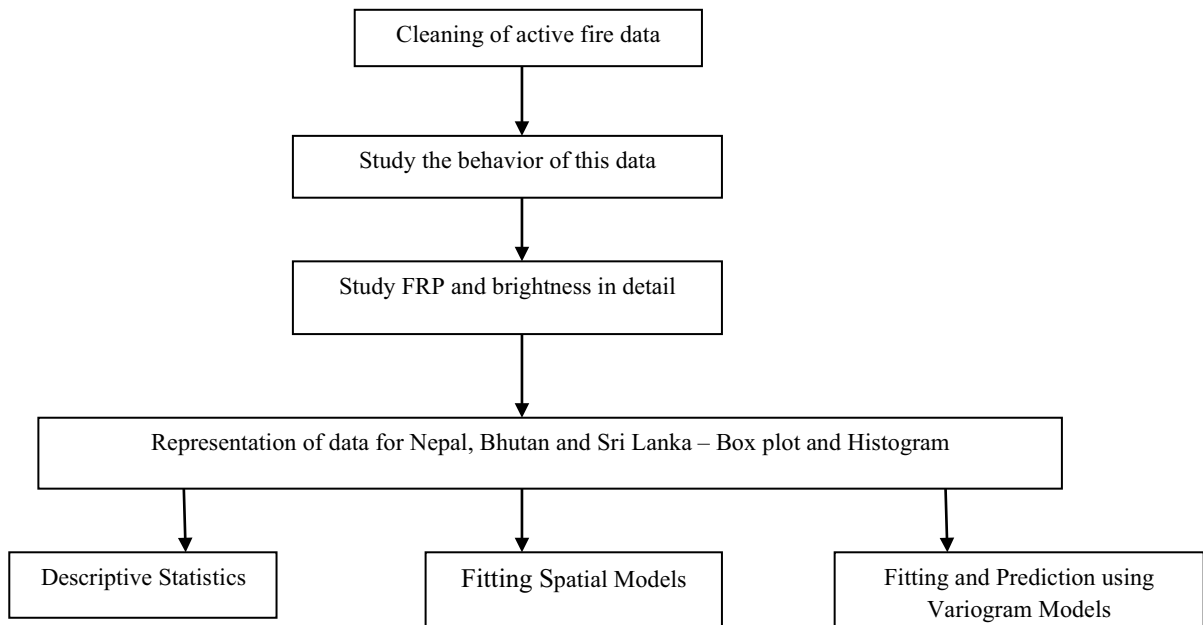
Fig. 1 Statistical methodologies used in the analysis of active fire data

arbitrarily chosen sample locations, measurements on Z are available and prediction (interpolation) is required at non-observed locations s_0. So, here in variogram modeling, the variogram is often used for spatial prediction (interpolation) or simulation of the observed process based on point observations.

Inverse distance weighted (IDW) interpolation is a non-geostatistical method of spatial prediction. The local influence of each measured point diminishing with distance is the main assumption of this method. Here, exponent (p) controls the distance. The lower the exponent, the more uniformly are the neighbor values incorporated in this interpolation. "If p = 0, the weights do not decrease with the distance and the estimated values at unsampled locations are equal to the mean of all the measured values; the value p = 2 is typically set by default in most applications, meaning that the importance of each measured location in determining a predicted value diminishes as a function of squared distance" (Gomez-Losada et al., 2019). Interpolation can also be done using a geostatistical model such as ordinary kriging (OK). In OK interpolation, the function determining the weights is called a variogram model. This model is a function fitted to the (empirical) variogram. The autocorrelation structure of the observed pattern is described

by this variogram. OK plays a critical role in spatial estimation. The interpolated estimates from IDW are always within the range of the observed values at sample locations. This differentiates IDW from OK (Bivand et al., 2013).

The methodology used in this paper is represented by the flowchart given in Fig. 1.

Results and discussion

An overview of incidence of active fire in the three countries of South Asia is given in Fig. 2. Here, the entire region is divided into grids. The incidences of active fire in these three countries are shown against this backdrop. The FRP of these fires is also given in figures on the right side of Fig. 2. The shades are from light to dark. Here, the darkest color indicates FRP of the lowest intensity. And the lightest color indicates FRP of the highest intensity. FRP can be used to quantify burned biomass, as it measures radiant energy released per unit time by burning vegetation (Costa & Fonesca, 2017). It can be seen from the map of Bhutan that the incidence of active fire is in the border areas, which adjoin the Indian states of West Bengal and Assam. The incidences of active fire in
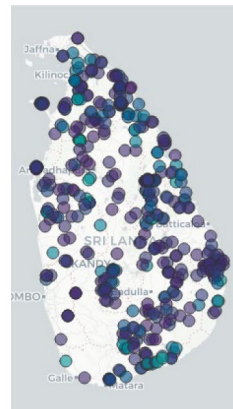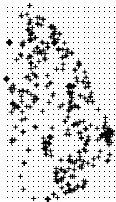
Bhutan
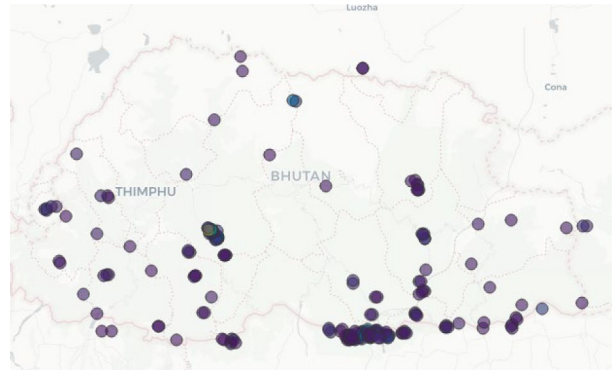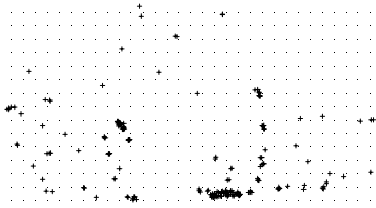
Nepal
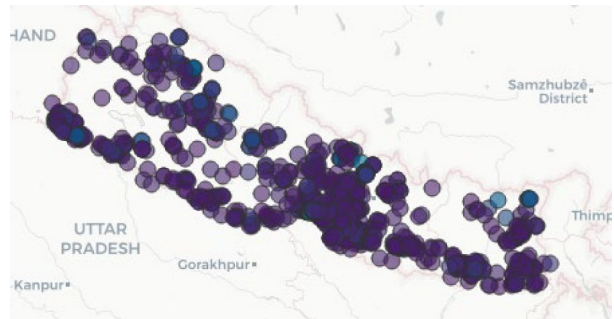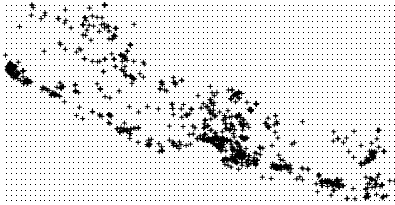
Sri Lanka

Fig. 2 Active fire and its FRP in Bhutan, Nepal, and Sri Lanka from 18 Sept. 2019 to 17 Sept. 2020

Table 1 Details of active fire from 18 September 2019 to 17 September 2020

Country	Total number	Types of fire				
		Time		Place		
		Daytime	Nighttime	Presumed vegetation fires	Static land source	Offshore
Bhutan	189	176	13	188	1	0
Sri Lanka	441	422	19	420	4	17
Nepal	982	913	69	981	1	0

Nepal are in the border areas, which adjoin the Indian states of Uttar Pradesh and Bihar. There are also incidences of fire in the hilly regions of Nepal. As seen from Fig. 2, the incidences of active fire are higher in the eastern and southern coasts of Sri Lanka.

The behavior of these fires is described in detail in Table 1 and Table 2. As seen from Table 1, in Bhutan, out of 189 observations of active fire, 176 took place in daytime and 13 during nighttime. Also, 188 of these were presumed vegetation fires, and 1 was due to other static land sources. In Sri Lanka, out of 441 incidences, 422 incidences of active fire took place during the daytime and 19 during the nighttime. Here, 420 were presumed vegetation fires, 4 were due to other static land sources, and 17 were offshore. In Nepal, out of 982 incidences of active fire, 913 occurred during daytime and 69 during nighttime. Presumed vegetation fires were sources of 981 fires and other static land sources generated 1 fire.

The variables closely related to the behavior of active fire are brightness and FRP. The behavior of these variables in three countries is summarized in Table 2, in terms of descriptive statistics. It is seen that although Bhutan had 189 incidences of fire in 1 year, it had the highest mean FRP among the three countries. The statistics describing FRP such as the quartiles and coefficient of variation (CV) are the highest for Bhutan.

This implies that although the incidence of fire is the least, in comparison to that of Nepal and Sri Lanka, the quantity of biomass burning in these fires is the highest. Nepal has the highest incidence of active fires with 982. But the average FRP is much lower than that of Bhutan. The median FRP for Nepal is the lowest. Sri Lanka has the most consistent type of active fire. The coefficient of variation is the lowest among these three countries. The spread of the variable brightness of active fire is only 2.7% of the mean. The spread of FRP is 68% of the mean. This indicates that Sri Lanka is most consistent with respect to the intensity of fire. This pattern is also reflected in box plots shown in Fig. 3 and Fig. 4. It can be seen from these figures that Bhutan has many outlier values on the higher range of the data. This means that although the number of such fires was the least, the intensity of these fires in terms of brightness and FRP was the highest. The box plot of Sri Lanka is the most consistent with the minimum number of outliers. The histogram of the behavior of brightness and FRP is given in Fig. 5 and Fig. 6. The brightness of active fire takes a nearly symmetrical shape in Sri Lanka. The FRP of most of active fires is between 0 and 20 MW for

Table 2 Descriptive statistics for active fire data for Bhutan, Nepal, and Sri Lanka

Sr. No	Country	Min	Max	Mean	Median	SD	Q1	Q3	CV
Brightness in Kelvin									
1	Bhutan	300.5	398.7	318.2	315.8	14.691	309.1	323.5	4.617
2	Nepal	300	370.3	317.7	317.7	9.874	310.6	323.4	3.108
3	Sri Lanka	3000	354.8	320.6	320.1	7.279	316.1	320.6	2.27
Fire radiative power (FRP) in MW									
1	Bhutan	3.50	417.60	26.75	14.30	44.277	9.20	27.30	165.496
2	Nepal	2.4	287.10	18.42	12.2	22.152	8.1	19.98	120.23
3	Sri Lanka	3.3	88.1	16.12	12.4	10.915	9.0	19.8	67.715

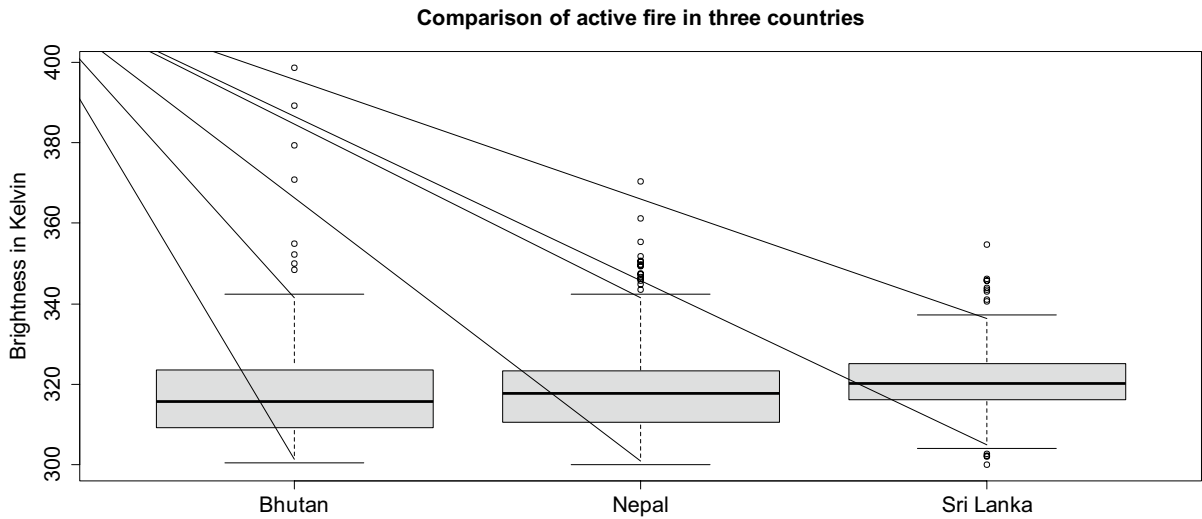**Comparison of active fire in three countries**

Fig. 3 Box plot of brightness of active fire in the three countries

Nepal and Sri Lanka. This is also validated by the Q3 values of Table 2. But for Bhutan, although the incidences of such fires are very low in comparison with those of the Nepal and Sri Lanka, the intensity in terms of FRP is the highest.

The residuals obtained from simple linear regression are tested for spatial dependence using Moran's *I*. It is a test under the null hypothesis that the data are not spatially correlated. A significant value of Moran's *I* standard deviation indicates that the

regression residuals are spatially correlated. This is based on a simple non-spatial regression model of brightness on scan, FRP, and brightness_T31 that is constructed. The dependent variable brightness is channel 21/22 brightness temperature of the fire pixel measured in Kelvin. The independent variable scan represents 1-km fire pixel and is representative of actual pixel size. The independent variable FRP represents fire radiative power measured in megawatts. And brightness_T31 is channel 31 brightness
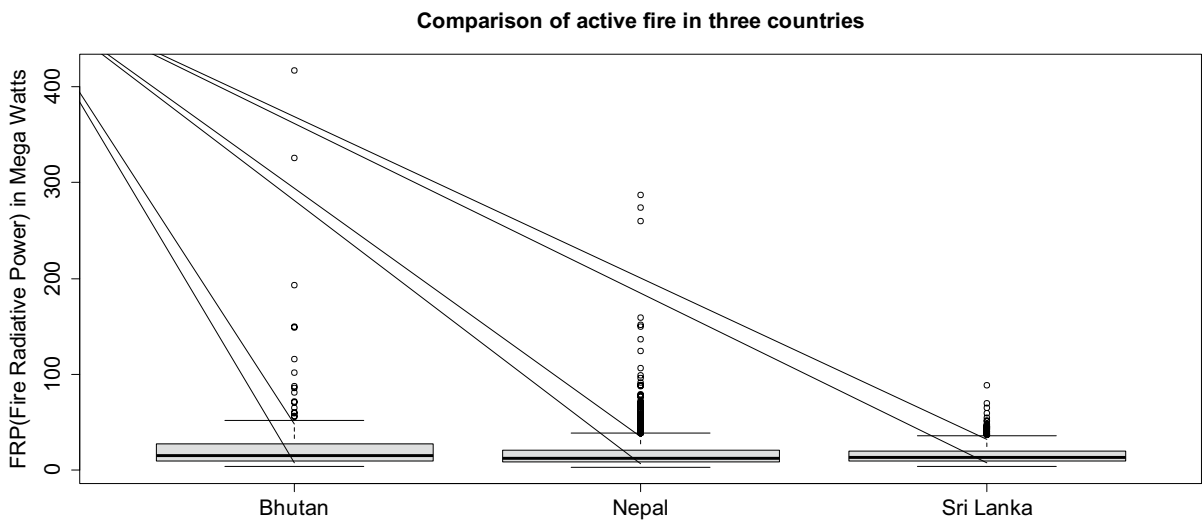
Comparison of active fire in three countries

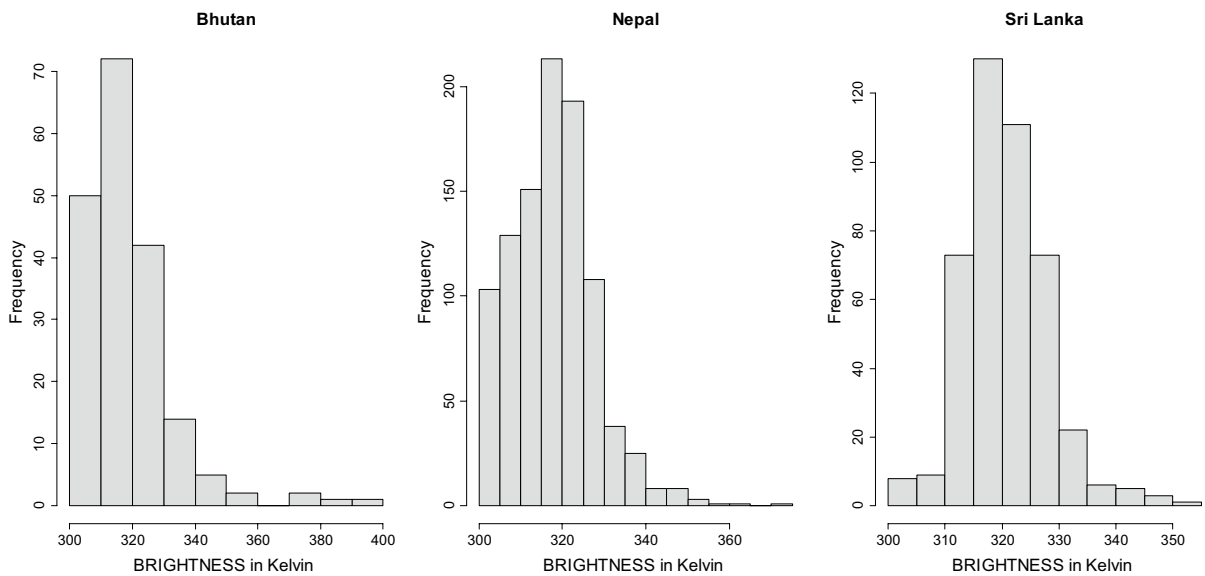**Fig. 4** Box plot of FRP of active fire in the three countries

Fig. 5 Histogram of brightness of active fire in the three countries

temperature of the fire pixel measured in Kelvin. For autoregressive models, the neighbor file is based on 5 nearest neighbors contiguity scheme. These 5 nearest neighbors are identified using great circle distances. In these autoregressive models, brightness is regressed on scan, FRP, and brightness_T31.

Among the six autoregressive spatial models tested for all the three countries, the one with maximum pseudo R^2 is selected. These autoregressive models are namely (a) Spatially Lagged X model, (b) Spatial Error model, (c) Spatial Durban Error model, (d) Spatially Lagged Y model, (e) Spatial
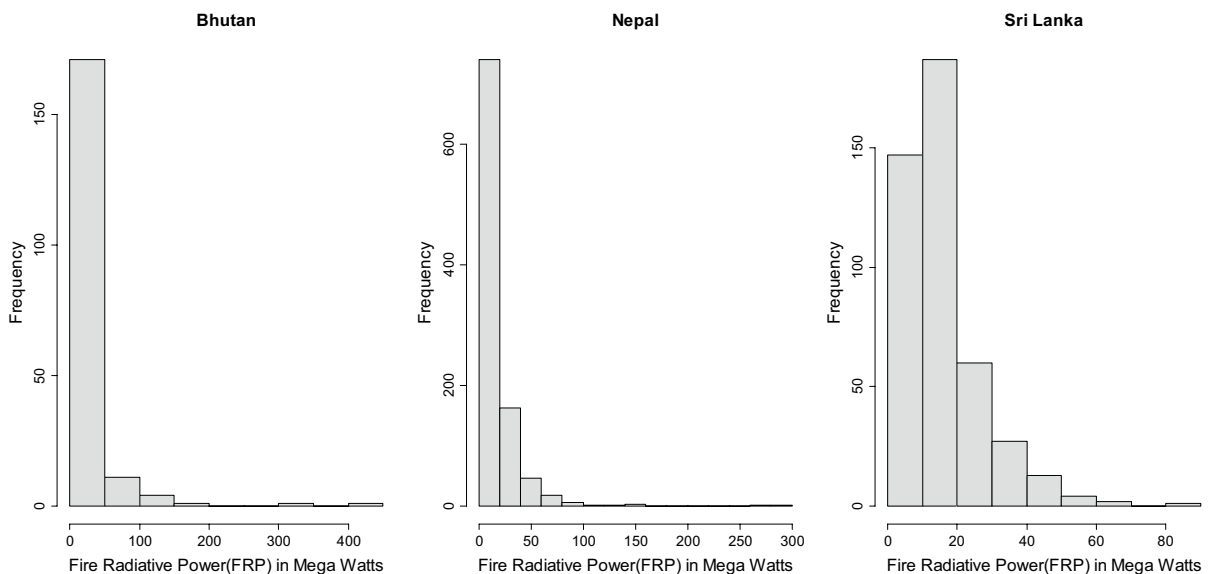
Fig. 6 Histogram of FRP of active fire in the three countries

Durban model, (f) Manski All Inclusive model, and (g) Kelegian Prucha model.

Spatially Lagged X model (Local Spatial Model) given below in Eq. (5) is the most suitable model for Bhutan:

$$y = X\beta + WXT + e \tag{5}$$

This is also seen in Table 3. This model explains how neighboring explanatory variables behave. The behavior of incidence of active fire

Table 3 Parameters of the spatial models for three countries

Sr. No	Country	Model	Model parameters	Accuracy (Pseudo R^2)	Remarks
1	Bhutan	Spatially Lagged X model (Local) $Y = X\beta + WXT + e$	$\beta_0 = 86.006\ *,$ $\beta_1 = -5.631\ *$ $\beta_2 = 0.636\ *$ $\beta_3 = 0.255\ *$ (lag coefficients) $W_1 = 3.269$ $W_2 = 0.14$ $W_3 = 0.004$	0.842	Best model
		SEM Spatial Error (Global), Y = X $\beta + u, u = \lambda Wu + e$	$\beta_0 = 113.231\ *,$ $\beta_1 = -5.255\ *$ $\beta_2 = 0.697\ *$ $\beta_3 = 0.258\ *$ (spatially lagged error multiplier) $\lambda = 0.144$	0.839	Second-best model
2	Sri Lanka	Spatial Durban Error model (Local) $Y = X\beta + WX\theta + u,$ $u = \lambda Wu + e$	$\beta_0 = 174.398\ *,$ $\beta_1 = -8.712\ *$ $\beta_2 = 0.429\ *$ $\beta_3 = 0.581\ *$ (lag coefficients) $W_1 = -0.395$ $W_2 = 0.067$ $W_3 = 0.093*$ (spatially lagged error multiplier) $\lambda = 0.312*$	0.808	Best model
		SEM Spatial Error (Global), Y = X $\beta + u, u = \lambda Wu + e$	$\beta_0 = 197.073\ *,$ $\beta_1 = -8.621\ *$ $\beta_2 = 0.423\ *$ $\beta_3 = 0.571\ *$ (spatially lagged error multiplier) $\lambda = 0.342*$	0.806	Second-best model
3	Nepal	Spatial Durban Error model (Local) $y = X\beta + WX\theta + u,$ $u = \lambda Wu + e$	$\beta_0 = 91.7403\ *,$ $\beta_1 = -6.038\ *$ $\beta_2 = 0.897\ *$ $\beta_3 = 0.315\ *$ (lag coefficients) $W_1 = -0.479$ $W_2 = -0.13\ *$ $W_3 = 0.01$ (spatially lagged error multiplier) $\lambda = 0.407*$	0.771	Best model
		SEM Spatial Error (Global), Y = X $\beta + u, u = \lambda Wu + e$	$\beta_0 = 59.168\ *,$ $\beta_1 = -6.02\ *$ $\beta_2 = 0.874\ *$ $\beta_3 = 0.316\ *$ (spatially lagged error multiplier) $\lambda = 0.411*$	0.769	Second-best model

Table 4 Variogram modeling of 1-year FRP data

Sr. No	Country	Variogram model	Parameters	Sum of squares between variogram model and sample variance	RMSE (IDW-cross validation)	RMSE (OK)
1	Bhutan	Matern	Sill = 1680.89, range = 772.077, kappa = 0.05	211,896,984	37.678	39.639
2	Nepal	Matern	Sill = 64.868, range = 407.119, kappa = 0.05	58,591,690	19.099	20.829
3	Sri Lanka	Matern	Sill = 29.243, range = 200.762, kappa = 0.05	179,552	11.094	10.656

in neighboring area affects the behavior of active fire in that area. Here, the X values of neighboring areas affect the incidence of active fire in that area. But there is no global spillover effect. Active fire data of Bhutan shows just significant spatial correlation at $\alpha = 0.05$, and Moran's I standard deviation of the residual of the non-spatial simple linear model is 1.6883. This implies that the incidence of active fire in the neighboring area significantly affects active fire in that area with a p value of 0.04568. But here, the sign of the independent variables, namely scan, brightness_T31, and FRP, does not change when the coefficients of the lag variables are considered. These lag variables represent neighboring areas.

The Spatial Error Model (Global) is also studied as the second-best model. It is explained by Eqs. (6) and (7):

$$Y = X\beta + u \tag{6}$$

$$u = \lambda Wu + e \tag{7}$$

Here, λ is a spatially lagged error multiplier. In a global model, the impact of one region spills over to the other, even when they are not specified as neighbors. So the behavior of global model in explaining the incidence of active fire is also studied. As seen from Table 3, the accuracy of these models, explained by pseudo R^2, takes the values 0.842 and 0.839 respectively.
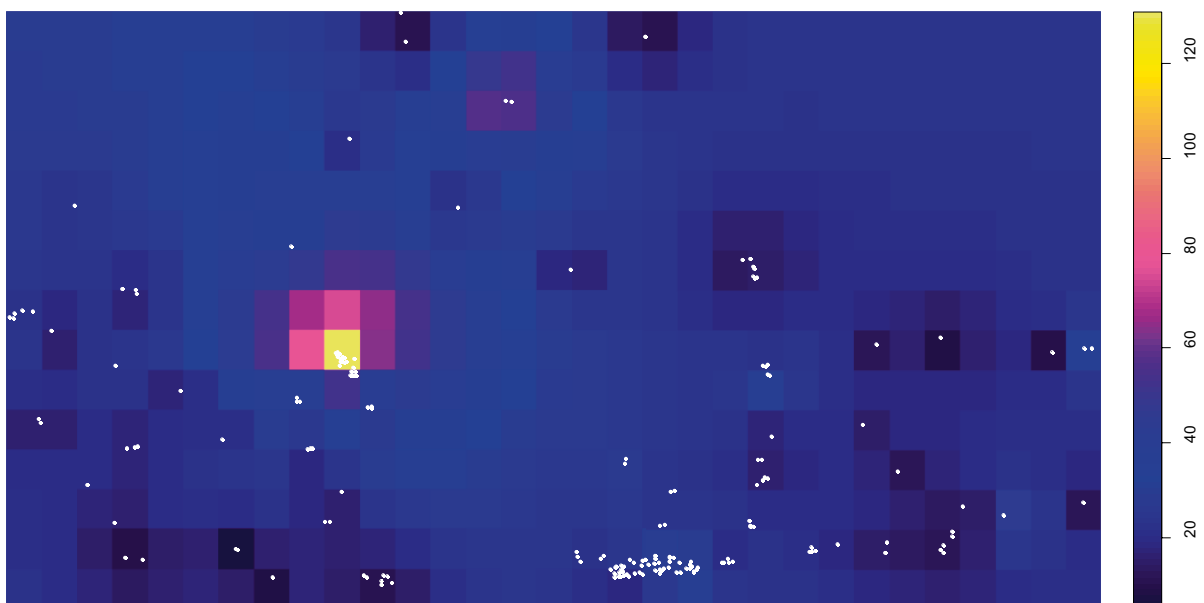

Fig. 7 Spatially interpolated FRP for Bhutan with white dots as observed values of 1-year FRP data

Fig. 8 Spatially interpolated FRP for Nepal with white dots as observed values of 1-year FRP data

As seen from Table 3, for Sri Lanka, the Spatial Durban Error model represented by Eqs. (8) and (9) is the most suitable model:

$$Y = X\beta + WX\theta + u \tag{8}$$

$$u = \lambda Wu + e \tag{9}$$

It is a local model. It has lag coefficients that study the impact of the neighboring area on the explanatory variables. But here, the sign of the independent variables, namely brightness_T31 and FRP, does not change when the coefficients of the lag variables are considered. Moran's I statistics standard deviation takes a value of 6.2236 and shows a high significant spatial correlation with a p value of 2.43E–10. This implies that the incidence of active fire in Sri Lanka shows a very high spatial correlation.

But Spatial Error Model (Global) is also studied as the second best-model. It is explained by Eqs. 6 and 7 given below. Here, λ is a spatially lagged error multiplier. In a global model, the impact of one region spills over to the other, even when they are not specified as neighbors. The pseudo R^2 values for the first and the second models are 0.808 and 0.806, respectively.

For Nepal, active fire data is highly spatially correlated. Here, Moran's I statistics standard deviation takes a value of 11.239 with a p value < 2.2E–16. Among the three countries, Nepal's data shows

the highest spatial correlation. Bhutan's active fire data shows the least spatial correlation. At $\alpha = 0.01$,
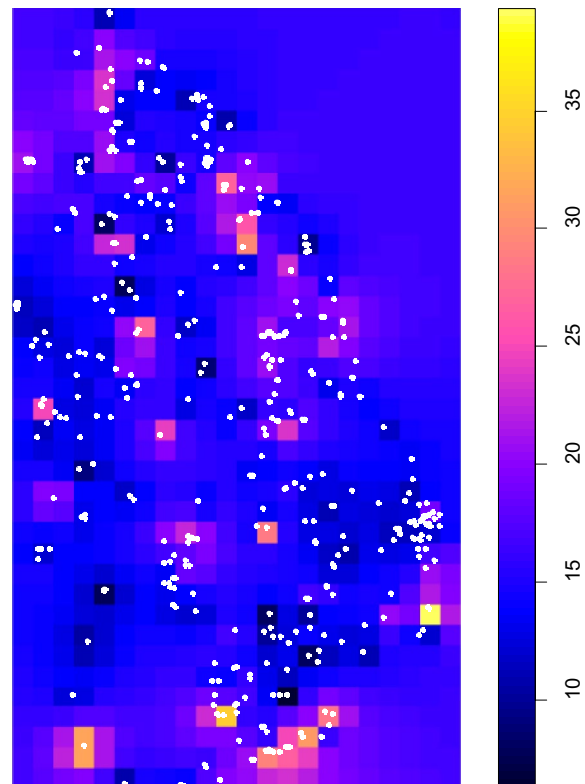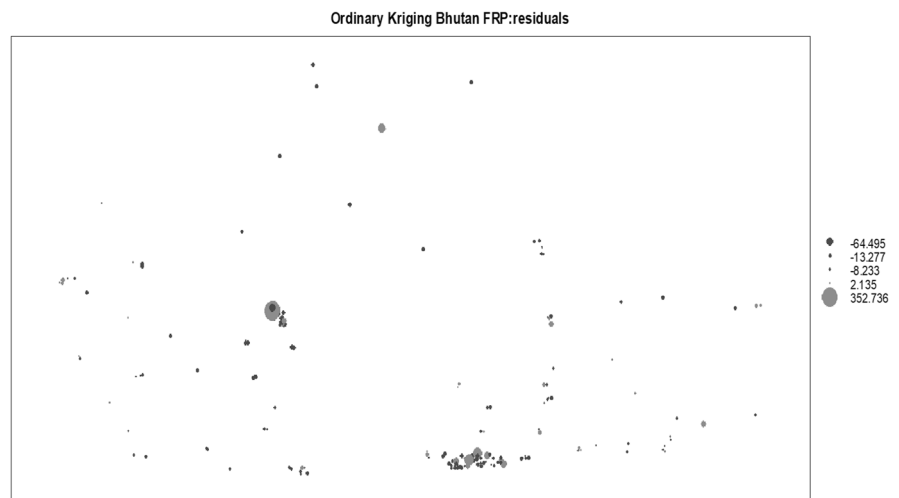

Fig. 9 Spatially interpolated FRP for Sri Lanka with white dots as observed values of 1-year FRP data

Fig. 10 Residuals from the ordinary kriging of 1-year FRP data for Bhutan

Moran's *I* statistics is not significant for Bhutan, but they are very highly significant for Sri Lanka and Nepal.

Similar to Sri Lanka, Spatial Durban Error model, given in Eqs. (8) and (9), is the most suitable model.

Here, the independent variables scan and brightness_T31 change the sign for the lag variables. This is related to neighboring areas. But for the other independent variable (FRP), no change in sign is seen.

But the Spatial Error Model (Global) is also studied as the second-best model. It is explained by Eqs. (6) and (7). The pseudo R^2 values for the first and the second models are 0.771 and 0.769 respectively.

Values with asterisks (*) are statistically significant.

The results of variogram modeling are given in Table 4. It has been used in quantifying the variability of vegetation FRP. Four variogram models (Matern, Spherical, Exponential, and Gaussian) were tested. Matern function with the smallest root mean square error seems to be the most suitable model in explaining the variogram of sample FRP of the three countries. Matern function has been successful in variogram modeling in other situations. Mianasny and McBratney have highlighted the importance of (semi) variogram as keystone to geostatistics and the flexibility of Matern function in explaining the variogram of soil variability (Minasny & McBratney 2005). Matern function is a special case of exponential function and fits the soil parameters' variogram close to the origin (Shaheen & Iqbal 2018).

Fig. 11 Residuals from the ordinary kriging of 1-year FRP data for Nepal
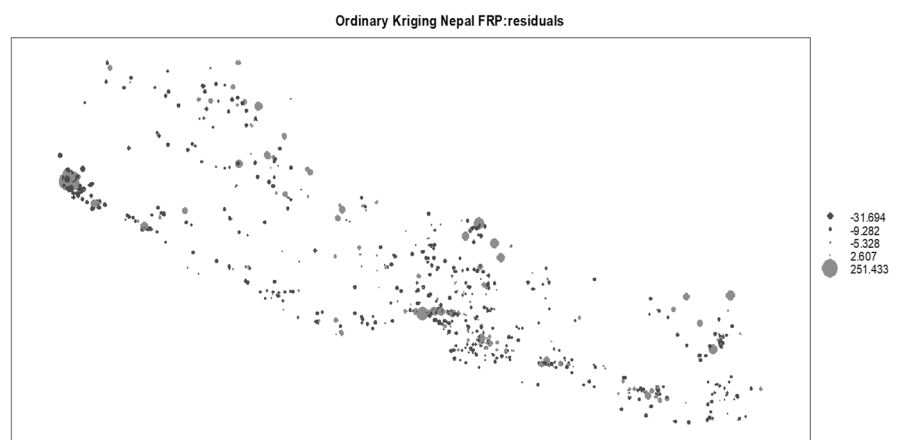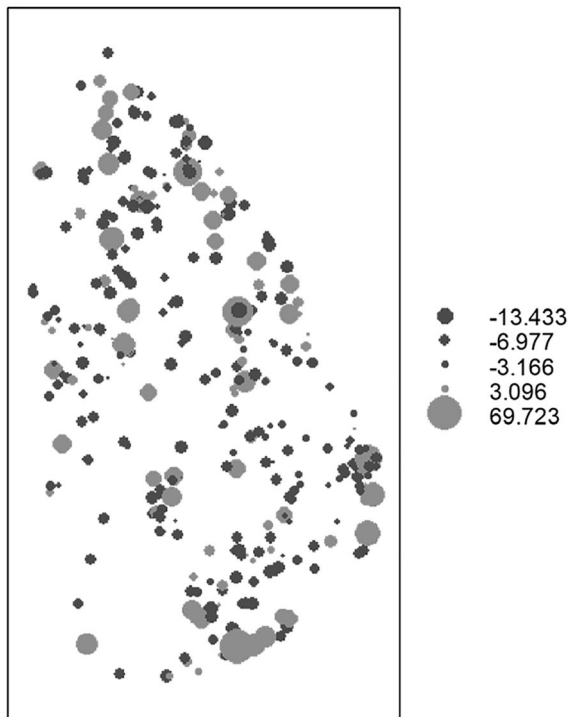
Ordinary Kriging Sri Lanka FRP:residuals

Fig. 12 Residuals from the ordinary kriging of 1-year FRP data for Sri Lanka

The FRP of the burning potential of the biomass at the unsampled area is predicted using IDW and OK. The results of interpolation FRP using IDW for the unsampled areas of Bhutan, Nepal, and Sri Lanka are given in Fig. 7, Fig. 8, and Fig. 9. The darker the color, the lower is the interpolated value of FRP for the unsampled areas. IDW is a non-geostatistical model, and here $p = 2$.

The residuals obtained from the OK of the Matern model given in Table 3 are given in Fig. 10, Fig. 11, and Fig. 12. We see that the residuals take the highest value for Bhutan and the lowest value for Sri Lanka. As seen from Table 4, the root mean square error (RMSE) for Sri Lanka is the lowest for predictions made by IDW and OK. This is because of the consistent nature of FRP data. Here, RMSE are the prediction errors from IDW and kriging. We see that predictions from OK have higher RMSE than those from IDW for Bhutan and Nepal.

Conclusions

Forest fires wreak havoc to the environment and damage the flora and fauna of that region. These fires in South Asia are either natural or manmade. In this paper, the intensity of these fires is statistically analyzed. FRP and brightness data are used. Three countries from South Asia namely Bhutan, Nepal, and Sri Lanka are taken as a model in this study of active fires.

Bhutan had minimum occurrences of active fire in the 1-year period of Sept. 2019 to Sept. 2020. This was in contrast to Nepal, which had the highest incidence. But some of these fires in Bhutan were of very high intensity. These fires are mainly vegetation fires in all the three countries. The coefficient of variation of brightness and FRP was highest in Bhutan, in comparison to Nepal and Sri Lanka. This indicates that the highest variance is Bhutan in contrast to the lowest in Sri Lanka. The distribution of these values for Sri Lanka is also symmetrical, as reflected by the box plots of brightness and FRP. The p value of Moran's I standardized variate for the residuals of a non-spatial simple linear model is 0.04568 for Bhutan, in contrast to 2.43E–10 for Sri Lanka and < 2.2E–16 for Nepal. This implies that unlike Nepal and Sri Lanka, the behavior of active fire is universal throughout Bhutan and does not depend on its geographical coordinates. This is done at the 1% level of significance.

Among several autoregressive spatial models tested, Spatially Lagged X model (local), Spatial Durban Error model (local), and Spatial Durban Error model (local) were the best in explaining the brightness of these active fires for Bhutan, Nepal, and Sri Lanka respectively. The coefficients of determination pseudo R^2 are 0.842, 0.771, and 0.808.

The variogram of FRP was best explained by the Matern function for all the three countries. The spatial variability of FRP has been quantified with variogram analysis here. The FRP of burning potential of vegetation in unsampled areas was predicted by IDW and OK. These models have best explained the active fire for Sri Lanka, as the RMSE took minimum values here. Symmetric and consistent nature FRP values are the main reasons.

Here, statistics is used to offer critical insights in understanding the behavior of vegetation fires. Such studies can provide important guidelines for

strengthening management of such fires in South Asia. In addition to this, through this detailed statistical study of FRP and brightness, many variables closely related to the incidence of fire can be indirectly assessed. This is especially useful for countries in the developing world that have knowledge gap due to scarce data.

Author contribution   This paper is written by a sole author.

Availability of data and material   This research is based on open-access satellite data.

Declarations

Conflict of interest   The author declares no competing interests.

References

Afonseca B. D. & Costa J. F. (2021). Dynamic anisotropy and non linear geostatistics supporting short term modeling of structurally complex gold mineralization. *Geosciences*, 74(2), https://doi.org/10.1590/0370-44672020740034

Bhunia, G. S., et al. (2018). Assessment of spatial variability of soil properties using geostatistical approach of lateritic soil. *Annals of Agrarian Science, 16*(4), 436–443. https://doi.org/10.1016/j.aasci.2018.06.003

Bivand R. et al. (2013). Applied spatial data analysis with R, Second edition, Springer, NY. http://www.asdar-book.org/

Costa B. S. & Fonseca E. L. (2017). The use of fire radiative power to estimate the biomass consumption coefficient for temperate grasslands in the Atlantic forest biome. *Revista Brasileira de Meteorologia*, 32(2). https://doi.org/10.1590/0102-77863220004

Eskandari, S., Miesel, J. R., & Pourghasemi, H. R. (2020). The temporal and spatial relationships between climatic parameters and fire occurrence in northeastern Iran. *Ecological Indicators*, *118*. https://doi.org/10.1016/j.ecolind.2020.106720

Fithria, A., & Ani, A. (2017). Frequency and intensity of forest and land fire incidents in the Banjar district in 2015. *Journal of Biodiversity and Environmental Sciences, 11*(5), 12–19. https://doi.org/10.1016/j.ecolind.2020.106720

Johanna, S. F., et al. (2020). Linear mixed models and geostatistics for designed experiments in soil science: Two entirely different methods or two sided of same coin? *European Journal of Soil Science, 72*(1), 47–68. https://doi.org/10.1111/ejss.12976

Kunwar, R. M., & Khaling, S. (2006). Forest fire in terai Nepal - causes and community management interventions. *International Forest Fire News, 34*, 46–54.

Lentile L.B. et al. (2006). Remote sensing techniques to assess active fire characteristics and post fire effects. *International Journal of Wild land Fire* 15, 319–345. https://core.ac.uk/download/pdf/17270328.pdf

Lim, C. H., et al. (2019). Can satellite-based data substitute for surveyed data to predict the spatial probability of forest fire? A geostatistical approach to forest fire in the Republic of Korea. *Geomatics, Natural Hazards and Risk, 10*(1), 719–739. https://doi.org/10.1080/19475705.2018.1543210

Lloyd, C. D. (2010). Spatial data analysis – an introduction for GIS users. Oxford University Press.

Gómez-Losada, Á., Santos, F. M., Gibert, K., & Pires, J. C. (2019). A data science approach for spatiotemporal modelling of low and resident air pollution in Madrid (Spain): implications for epidemiological studies. *Computers, Environment and Urban Systems*,*75*, 1-11. https://doi.org/10.1016/j.compenvurbsys.2018.12.005

Marin, P. G., et al. (2018). Drought and spatiotemporal variability of forest fires across Mexico. *Chinese Geographical Science, 28*, 25–37. https://doi.org/10.1007/s11769-017-0928-0

Marques S. F. & Pitombo C. F. (2021) Applying multivariate geostatistics for transit ridership modeling at the bus stop level. *Boletim de Ciências Geodésicas*, 27 (2). https://doi.org/10.1590/1982-2170-2020-0069

Matin, M. A., et al. (2017). Understanding forest fire patterns and risk in Nepal using remote sensing, geographic information system and historical fire data. *International Journal of Wild Land Fire, 26*, 276–286. https://doi.org/10.1071/WF16056

Minasny, B., & McBratney, A. B. (2005). The matern function as a general model for soil variograms. *Geoderma, 128*, 192–207.

NASA. (2021). Moderate resolution imaging spectroradiometer. https://modis.gsfc.nasa.gov/about/

Reddy, C. S., et al. (2019). Identification and characterization of spatio-temporal hotspots of forest fires in South Asia. *Environmental Monitoring Assessment, 191*, 791. https://doi.org/10.1007/s10661-019-7695-6

Sakellariou, S., Cabral, P., Caetano, M., Pla, F., Painho, M., Christopoulou, O., & Vasilakos, C. (2020). Remotely sensed data fusion for spatiotemporal geostatistical analysis of forest fire hazard. *Sensors*, *20*(17), 5014.

Shaheen, A., & Iqbal, J. (2018). Spatial distribution and mobility assessment of carcinogenic heavy metals in soil profiles using geostatistics and random forest, boruta algorithm. *Sustainability*, *10*(3), 799.

Su, H., et al. (2020). Machine learning and geostatistical approaches for estimating aboveground biomass in Chinese subtropical forests. *Forest Ecosystem, 7*, 64. https://doi.org/10.1186/s40663-020-00276-7

Vadrevu, K. P., et al. (2019). Trends in vegetation fires in South and Southeast Asian countries. *Science Report, 9*(1), 7422. https://doi.org/10.1038/s41598-019-43940-x