e-Journal of Analysis and Applied Mathematics **2020(1)** (2020), 98 – 117
DOI 10.2478/ejaam-2020-0008

Time series analysis of radiant heat using 75 hours VIIRS satellite day and night band nightfire data

Jyoti U. Devkota

Communicated by Evren Hincal

Abstract. The nightfires illuminated on the earth surface are caught by the satellite. These are emitted by various sources such as gas flares, biomass burning, volcanoes, and industrial sites such as steel mills. Amount of nightfires in an area is a proxy indicator of fuel consumption and CO2 emission. In this paper the behavior of radiant heat (RH) data produced by nightfire is minutely analyzed over a period of 75 hour; the geographical coordinates of energy sources generating these values are not considered. Visible Infrared Imaging Radiometer Suite Day/Night Band (VIIRS DNB) satellite earth observation nightfire data were used. These 75 hours and 28252 observations time series RH (unit W) data is from 2 September 2018 to 6 September 2018. The dynamics of change in the overall behavior these data and with respect to time and irrespective of its geographical occurrence is studied and presented here. Different statistical methodologies are also used to identify hidden groups and patterns which are not obvious by remote sensing. Underlying groups and clusters are formed using Cluster Analysis and Discriminant Analysis. The behavior of RH for three consecutive days is studied with the technique Analysis of Variance. Cubic Spline Interpolation and merging has been done to create a time series data occurring at equal minute time interval. The time series data is decomposed to study the effect of various components. The behavior of this data is also analyzed in frequency domain by study of period, amplitude and the spectrum.

Keywords. Cluster analysis, discriminant analysis, principle components analysis, spectrum, VIIRS satellite DNB data.

2020 Mathematics Subject Classification. 62H25, 62H30, 62J10, 62M15, 47N50, 97K80.

1 Introduction

Energy is the source of all vital inputs. This energy is also needed for survival and economic development of a society. Night time activities play a very important role in this dynamics. Renewable and non renewable energies power the radiance created by night lights. These night lights illuminate bridges, streets, ship fleets, buildings etc. Nightfire are produced by burning gas and biomass, volcanoes and

steel plants. These are indications of fuel consumption and co2 emission. Several satellites capture the radiant heat produced by nightfire and hence this night radiance can be measured by free and easily available satellite data. The data from two satellite systems can be used for this purpose. The Defence Meteorological Satellite Program is the older and NASA-NOAA Suomi National Polar Partnership (SNPP) Program is the newer [1]. The SNPP satellite caries VIIRS sensors have higher spatial resolution and better lowlight detection and are hence more sensitive and thus provide good quality day /night band (DNB) radiance values. Thus they are better suited for measurement of radiant heat. This paper is based on data products from the Visible Infrared Imaging Radiometer Suite (VIIRS) night time sensors. They are also called Day/Night Band, or DNB. Elvidge, Zhinzhin, Hsu and Baugh discussed satellite pyrometry at night and also gave background information about these VIIRS data [2]. Koel and Simon modeled crude birth rate and maternal mortality ratio of India using night time satellite images [3]. Mann, Melaas and Mallick used VIIRS/DNB to measure electricity supply reliability in Maharastra India [4]. Yu et al. claim that VIIRS data is a useful tool for evaluating poverty at county level in China [5]. Sharma et al. combined Moderate Resolution Imaging Spectroradiometer (MODIS)-based multispectral data with the Visible Infrared Imager Radiometer Suite (VIIRS)-based night-time light (NTL) data for robust extraction and mapping of urban built-up areas [6]. Dou, Liu, He and Hue used VIIRS data for urban land extraction [7].

In this paper a detailed time series analysis of dynamics of change in radiant heat data RH is done. Real time RH data is monitored from 2 September 2018, 21:16 hours to 6 September 2018, 2:27 hours. This data is day and night band radiance satellite nightfire data. The entire RH data in the given time period is taken without any considerations to its geographical location. Only time of its measurement is considered. The results obtained here help us to identify underlying trends and patterns in the behavior of this data irrespective of the geographical location of its occurrence. Underlying categories and hidden relationships existing in the multivariate data created from this nightfire satellite data are found. The main contributions of this study are as follows:

(i) Exploration and cleaning of time series of DNB radiance 75 hour data,

(ii) Studying the behavior of mean RH for three consecutive days using Analysis of Variance,

(iii) Understanding the hidden categories and groups in the data,

(iv) Creating a interpolated mean RH multivariate data for 4 September 2018 by merging the data of 3 Sept. and 5 Sept. 2018 and thus creating a time series data of 1427 minutes,

(v) Decomposition of this data into different components of a time series and testing of stationarity,

(vi) Computation of auto-correlation for different lag,

(vii) Understanding the relationship between auto-correlations of different lag with multiple linear regression and principle components analysis,

(viii) Classifying the data in frequency domain by understanding the value of spectrum for different frequencies.

The research questions of this study are as follows:

(i) How does the radiant heat time series data observed from 2 September 2018 to 6 September 2018 behave irrespective of geographical coordinates of its energy source?

(ii) What are the underlying groups and categories that can be used in discriminating this data?

(iii) How does the time series of DNB radiant heat RH behave over a near real time interval of 1472 minutes in 24 hours?

(iv) What is the behavior of this time series in frequency domain?

(v) What underlying trends and patterns can be inferred from this RH time series data?

This paper is arranged in the following manner. This section is followed by Research Method section highlighting the statistical tools and the data used. This is followed by section on Result and Analysis. The last section in this article is the Conclusion.

2 Research method

2.1 Data

There are 28252 observations on time series of RH data measured by variable Radiant Heat unit W. This variable is key to estimating fuel consumption and CO2 emission. On different times ranging from 21:16 hours on 2 September 2018 to 2:27 hours on 6 September 2018, there are 539 intervals of values taken by factor time. This data is recorded over the interval of 75 hours. This satellite data is a product of National Aeronautics and Space Administration (NASA). They provide real-time imagery of our night time world. Their primary purpose is to support short term weather predictions and disaster response community [8]. A number of sources contribute to the DNB signal, including city lights, lightning, fishing fleet

navigation lights, gas flares, lava flows, and even auroras. When partial to full illumination from the moon is available, reflection of this lunar illumination off of ice, snow, and other highly reflective surfaces enable the study of ocean and terrestrial features. VIIRS is a scanning radiometer onboard the Suomi National Polar Partnership (SNPP) Satellite. The VIIRS collects visible and infrared imagery and radiometric measurements of land, atmosphere and oceans. It is sensitive to 22 wavelength bands, including a DNB with 750-m resolution. The DNB is sensitive to visible and near-infrared wavelengths ranging from daylight down to low levels of night time radiance. The ability of the DNB to detect the low levels of visible light present at night makes it well suited to studying night lights [4] [8]. VIIRS sensors have higher resolution and detect low light better than older Defence Meteorological Satellite Program Operational Linescan System (DMSP-OLS) system [10] [9]. Letu, Nakajima and Nishio estimated the co2 emission by power plants using DMSP-OLS of VIIRS data [11]. Rybnikova and Portnov claim that VIIRS data gives more accurate results than DMSP-OLS data when finding correlation between incidence of breast cancer and artificial lights at night [12]. Similarly Shi et al. claim that VIIRS data are better in forecasting total freight traffic for China than DMSP-OLS data [13]. Satellite data can be used to monitor various parameters related to earth's environment. Near real time monitoring of ecosystem using satellite remote sensing was proposed by Verbesselt, Zeileis and Herold [14].

2.2 Analysis of 75 hours near real time data

Cubic Spline: In spline interpolation a smooth curve is drawn through the given points. Here the connecting polynomial is of degree three. To create a time series where the values are classified according to equally spaced time intervals, cubic spline interpolation is done [15]. Let the given data points be $(x_i,\ y_i)$, $i = 0, 1, 2, ., n$ where, $a = x_0 < x_1 < x_2 < \ldots < x_n = b$. Let, $h_i = x_i - x_{i-1}$, $i = 1, 2, n$. Let $s_i(x)$ be the cubic spline or cubic polynomial defined in the interval $[x_{i-1},\ x_i]$. The conditions for the natural cubic spline are the following:

(i) $s_i(x)$ is atmost a cubic in each subinterval $[x_{i-1},\ x_i]$, $i = 1, 2, , n$,

(ii) $s_i(x_i) = y_i$, $i = 0, 1, 2, , n$,

(iii) $s_i(x)$, $s_i'(x)$ and $s_i''(x)$ are continuous in $[x_0, x_n]$,

(iv) $s_i''(x_0) = s_i''(x_n) = 0$. Then,

$$s_i(x) = \frac{1}{h_i}[\frac{(x_i - x)^3}{6}M_{i-1} + \frac{(x_i - x)^3}{6}M_i + (y_{i-1} - \frac{h_i^2}{6}M_{i-1})(x_i - x) + (y_i - \frac{h_i^2}{6}M_i)(x - x_{i-1})]$$

,

(v) $s_i''(x_i) = M_i$, for all i.

Time Series Analysis: A time series is a sequence of observations taken sequentially in time. An intrinsic feature of time series is that, adjacent observations are dependent. This behavior is called auto correlation. The nature of this dependence among observations of a time series is of considerable practical interest [17]. Further a statistical phenomenon that evolves in time according to probabilistic laws is called stochastic process. The analysis of a time series is the analysis of a particular realization of this stochastic process. The time series analysis is concerned with techniques for the analysis of this dependence. Stationary processes are important class of stochastic models. These processes assume that the process remains in statistical equilibrium with probabilistic properties that don't change over time, in particular varying about a fixed constant mean level and with constant variance. But some time series can be represented as non stationary processes, in particular having no constant natural mean level.

Auto Regressive (AR) Process: These processes are very useful in describing certain practically occurring series. These processes can be stationary or non stationary [16]. Here the current value of the process is expressed as a finite, linear aggregate of previous values of the process and a random shock a_t. Let us denote the values of a process at equally spaced time $t, t-1, t-2, ...$ by $z_t, z_{t-1}, z_{t-2} \ldots$ Also let $\tilde{z_t} = z_t - \mu$ be the series of the deviations from μ. Then,

$$\tilde{z_t} = \phi_1 \tilde{z_{t-1}} + \phi_2 \tilde{z_{t-2}} + \ldots + \phi_p \tilde{z_{t-p}} + a_t$$

is called autoregressive process of order p.

Auto Correlation Function: The stationarity assumption implies that the covariance between values z_t and z_{t+k} separated by k intervals of time or lag k must be same for all values of t under stationarity assumption [16]. This covariance will be called the auto covariance at lag k and is defined by

$$\gamma_k = cov[z_t, z_{t+k}].$$

Similarly, the autocorrelation at lag k is

$$\rho_k = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sqrt{E(z_t - \mu)^2 E(z_{t+k} - \mu)^2}}.$$

Spectrum: A time series can be represented in a time domain or a frequency domain. It can be analyzed based on the assumption that it is made up of sine and cosine waves with different frequencies [17]. The graphical representation

of spectrum of a time series for different values of frequencies is called a periodogram. This name originated from a device developed in 1898 for detecting and estimating the amplitude of sine components of known frequency, buried in noise. First a Fourier series model is fitted to the data. This gives the frequency and the amplitude which is plotted to get the periodogram.

Discriminant Analysis: It is a technique of multivariate statistics. The objective here is to classify the objects into mutually exclusive classes on the basis of some a priori information. Here the prediction or identification of group membership is done on the basis of one or more predictor or explanatory variables along with one criterion variable [17].

Cluster Analysis: In this multivariate analysis technique, relatively homogeneous groups are searched and put into clusters. Since similar objects form a cluster, all the sample points in any cluster will provide similar information about the population characteristic. Thus, for further analysis one may include one object from each cluster. In that sense the cluster analysis is a data reduction technique in rows of the data matrix [17].

3 Result and analysis

The pattern of change in day and night band nightfire data for radiant heat RH is minutely analyzed. The entire RH data for five days is considered in the analysis and no consideration is given to its geographical coordinates. Only times of their occurrences are considered. Hidden features that are overlooked by conventional means are statistically identified. Unlike this paper, in other works mainly these data are used in remote sensing and thus geographical coordinates are taken into consideration. The areas of night fires are identified. But the radiant heat RH generated by these burning sources are not statistically analyzed. For example, these nightfire data have been used in remote sensing active volcanoes [18]. It has been used to map gas flaring and policy compliance in Yasuni biosphere reserve, in a study of burnable and unleakable carbon in western amazon [19]. It has been used to estimate thermal emmission and emmissivity through estimation of CH4 and release of CO2 by using a formula [20].

Here the data are 28252 time series observations occurring at unequal time intervals. On different times ranging from 21:16 hours on 2 September 2018 to 2:27 hours on 6 September 2018, there are 539 values taken by the factor time. After removing the extreme values also called outliers 517 observations are left. This data is categorized as a multivariate data with variables RH, Time Of Day and Day. The objective of this analysis was to find underlying categories in the multivariate data classified according to RH, Time of Day, Day and Color. Here, RH is a ratio data

Table 1. Descriptive statistics

Colour		Mean	Standard Deviation
Dark	RH	3.291	1.894
	Time of Day	2.253	1.855
	Day	3.406	0.9830
Light	RH	3.970	2.197
	Time of Day	3.070	0.809
	Day	3.146	0.824
Total	RH	3.609	2.036
	Time of Day	2.635	1.518
	Day	3.284	0.921

Table 2. The eigenvalue

Function	Eigenvalue	% of Variance	Cumulative%	Canonical Correlation
1	.121	100.0	100.0	0.328

Table 3. Wilks Lambda

Chi-Square	df	Sig.
3216.148	3	0.000

and Time of Day is categorical data classified in 5 categories 1 - Night (9:00 PM to 3:00 A.M), 2 - Dawn (3:00 A.M. to 5:30A.M.), 3 - Morning (5:30 A.M. to 11:00 A.M), 4 - Noon (11:00A.M to 6:30PM), 5 - Evening (6:30 P.M to 9:00 P.M). Day is a categorical data ranging from 1 to 5 with Day 1 being Sept. 2, Day 2 being Sept. 3, Day 3 being Sept. 4, Day 4 being Sept.5 and Day 5 being Sept. 6, 2018. Color is a categorical ordinal data, with 1 being dark from 7:00 PM to 3: 00 A.M and 2 being light from 3:00 A.M. to 7:00 PM. The behavior of this data 5 time of days and 5 days is minutely analyzed. For the time series analysis and spectral analysis following steps are taken. The data is merged and interpolated for every minute of the day and 1427 minutely data of mean RH data for 4 September 2019 is obtained. The results obtained are the following.

(i) **Discriminant analysis** shows that Colour (dark and light) is the attribute that discriminates this 75 hour and five days RH data of 28252 observations. The light and dark are the colours outside the satellite station during the time of

Table 4. Canonical discriminant function coefficients

	Function
	1
RH	0.260
Time of the Day	0.521
Day	-0.361
(Constant)	-1.126

Table 5. Structure matrix

	Function
	1
Time of the Day	0.803
RH	0.486
Day	-0.408

Table 6. Functions at group centroids

	Function
Colour	1
Dark	-0.326
Light	0.370

Table 7. Classification results

	Colour	Predicted Group Membership		Total
		Night	Light	Total
Count	Night	9699	5334	15033
	Light	4019	9200	13219
%	Night	64.5	35.7	100.0
	Light	30.4	69.6	100.0
	a. 66.9% of original grouped cases correctly classified.			

recording of the data. Here light signifies time from 3:00 A.M. to 7:00 P.M that is day time; dark signifies time from 7:00 P.M. to 3:00 A.M. that is night time. The accuracy of predicted group membership of the attribute Colour is 66.9%. Here RH, Time of Day and Day are taken as independent variables. Table 1 gives the descriptive statistics of the multivariate data according to the discriminating group Colour. Table 2 gives the eigen value or the latent variable. This value helps separate the group. The significant value of Wilk's lambda given in Table 3 shows that this discriminating model is good fit for data. Table 4 gives the Discriminant Function coefficients. These coefficients behave as multiple regression coefficients and are used in making prediction equation. The prediction equation or Discrimination Function (DF) based on Table 4 is the following.

$$\text{DF} = -1.126 + 0.26 * \text{RH} + 0.521 * \text{Time of Day} - 0.361 * \text{Day} \quad (1)$$

The structure matrix in Table 5 shows that Time of Day has highest loading on the DF followed by RH and Day. Table 6 gives the group centroids and are mean discriminant score for each group. After substituting the values of independent variables in Equation 1, if the discriminant score is closer to -0.326, then the group predicted is dark; and if it is closer to 0.370, then the group predicted is light. Cut Score $= (-0.326 + .370)/2 = 0.022$. If an individual person's score on the DF (calculated by plugging in their scores on RH, Time Of Day and Day to the equation (1)) is above 0.022, then they were probably recorded during light time. If their DF score is below 0.022, then they were probably the recorded during dark time. As seen from Table 7, category that is actually in dark and predicted dark is 64.5% . On the other hand categories that are in light and are predicted light are 69.6%. The overall accuracy of this model given in equation (1) is 66.9% which is between 64.5% and 69.6%. Here other variables namely RH, Time of Day and Day are taken as independent variables. Thus the discriminating factor for the RH data collected for five days irrespective of its geographical coordinates can be discriminated by characteristic dark and light.

(ii) **Cluster analysis** of the same five days dataset for different time ranges after removing outliers has 517 RH observations. This analysis shows that there are three hidden groups. The dendogram shown in the Figure 1, shows the organization of these clusters. These clusters are homogeneous within and heterogenous between. The one way analysis of variance (ANOVA) of RH with respect to the clusters given in Table 8. It shows that the difference between the clusters is highly significant with a p value of 0. The vertical axis denoted by height represents the degree of dissimilarity between clusters.

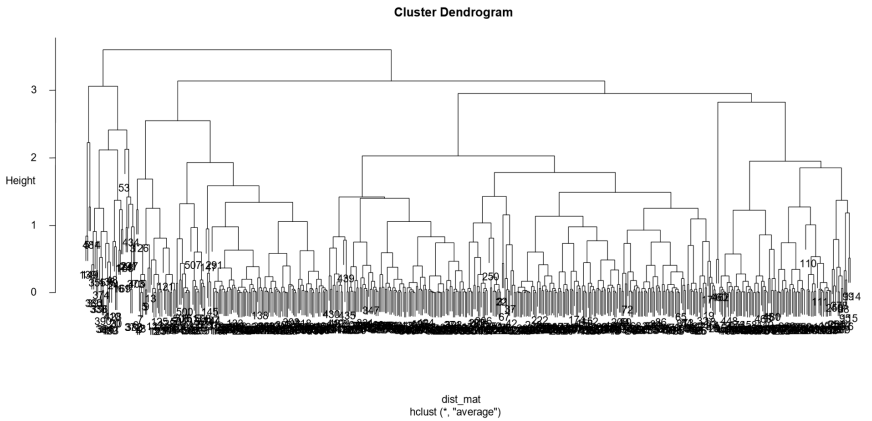

Figure 1. Cluster Dendogram of distribution of RH across 75 hours

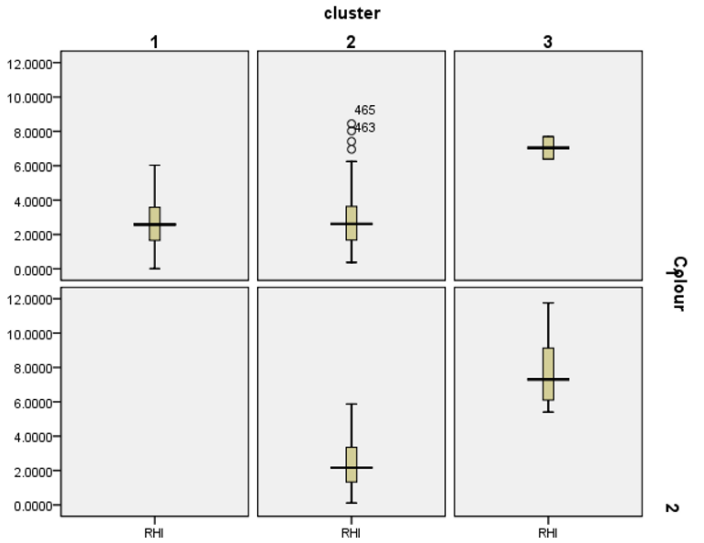

Figure 2. Boxplot representing clusters and colour

Cluster 1, cluster 2 and cluster 3 comprises of 91, 232 and 26 observations respectively. The box plot given in Figure 2 helps understand the organization

of clusters. As seen from Figure 2, cluster 1 comprises of lower values of RH recorded only during the 1, implying 7:00 P.M. to 3:00 A.M. Cluster 3 comprises of only higher values of RH recorded during 1 and during 2. Here color 2 represents data recorded from 3:00 A.M. to 7:00 P.M. Cluster 2 comprises of lower values of RH recorded during 1 and 2.

(iii) **Time series Analysis** of 75 hours time series data on radiant heat RH is done. The data are at unequal time interval. To create time series with equal time intervals, first the data for these three days is merged and then cubic spline interpolation is done. Thus 1427 minutely equally spaced Mean RH data for 4 September 2019 is generated. This is done under the assumption that there is no difference in the mean RH for three consecutive days holds. The behavior of the time series data is analyzed with decomposition of the time series data and detailed analysis of autocorrelation function using multiple linear regression and principle components analysis. The results are obtained applying given statistical methodologies in the following steps.

a. **Analysis of Variance (ANOVA):** To test the assumption that there is no difference in the mean RH with respect to three consecutive days, ANOVA is conducted. The ANOVA of RH with respect to day showed that there is no difference 3 Sept., 4 Sept., and 5 Sept. 2018 with respect to mean RH. As shown in Table 9, the p value is 0.528. This assumption is true as between three consecutive days there will be no significant difference between the amount of night lights. Then a time series data occurring at every minute is created for 4 Sept. 2018.

b. **Cubic Spline Interpolation:** To create an equally spaced time series occuring at every minute, the 487 values of RH for 3 September, 4 September and 5 September 2018 are merged. And then cubic spline interpolation is conducted. Thus a time series for 4 September 2018, occurring at every minute is generated. The length of this time series thus generated is 1427.

c. **Outliers:** To clean the data, outliers are removed. The length of the time series reduces from 1427 to 1381, as RH higher than 20.06 are removed. The outliers are defined here as the mean RH values lying outside the outer fence computed in the following manner.

$$Q_3 - Q_1 = 6.046776 - 1.394829 = 4.651947$$

Outer Fence

$$Q_1 - (Q_3 - Q_1) * 3 = 1.394829 - 4.651947 * 3 = -12.56101\ 0$$

(RH values are not negative)

$$Q_3 + (Q_3 - Q_1) * 3 = 6.046776 + 4.651947 * 3 = 20.00262$$

RH values higher than 20.06 are treated as outliers.

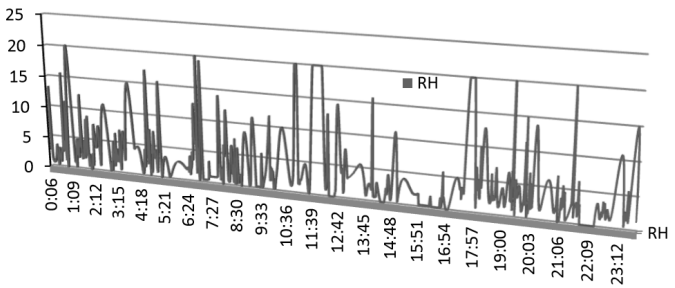

Figure 3. Time series data of radiant heat RH on 4 Sept 2018

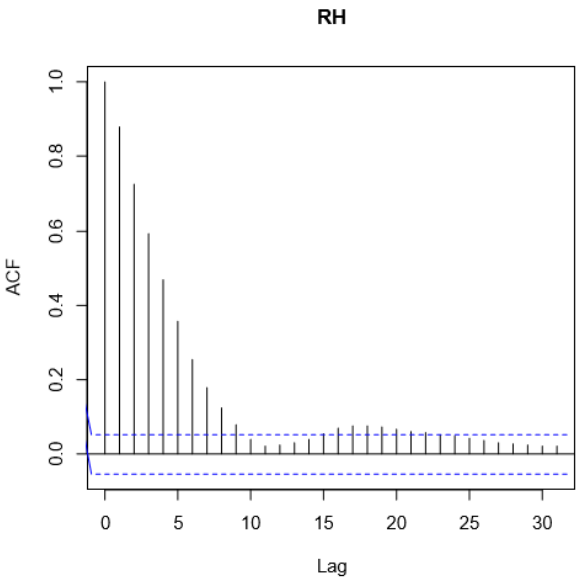

Figure 4. The autocorrelation function of RH

d. **Auto Correlation Function:** To explore the time series, its autocorrelation function (acf) and Dickey Fuller's test for stationarity is conducted. The minute wise time series data of RH generated is shown in Figure 3. The series seems to be stationary. This is validated by Dickey Fuller's test of stationarity, shown in Table 10. The Dickey Fuller's test for stationarity rejects the null hypothesis that the series is not stationary. As seen from Table 10, the time series and its first order differences are stationary. The autocorrelation function of this time series is shown in Figure 4. It can be seen that there is a significant autocorrelation even till the lag of 22 minutes. It also shows a positive association between successive values of RH. The autocorrelation values are not normally distributed. The Shapiro wilks statistics W takes the value 0.8558 with a p value of 0.000556. So this validates the fact that acf is not normally distributed. The mean of acf is 0.251364 and the standard deviation is 0.28872. Weibull, Gamma and Lognormal probability distributions are fitted to acf values. Weibull distribution best explains the probability distribution with smallest AIC of -6.713. The parameters of Weibull distribution are given by scale = 0.299 and shape = 0.976

e. **Multiple Linear Regression:** The value of Mean RH recorded at the current minute is significantly dependent on the mean RH values recorded in the lag of upto 22 minutes. The acf values have reflected this. The linear regression of the value of Mean RH (without intercept) recorded in the current minute on the RH values in lag upto 22 minutes is highly significant. The p value of this regression is 2.2e-16 with coefficient of determination $R^2 = 0.9031$.

f. **Principle Components Analysis:** Principal components analysis is conducted on the RH data of lag of upto 22 minutes. The autocorrelation function shown in Figure 4, revealed significant auto correlation up to lag of last 22 minutes. These are classified as 22 variables and the aim is to reduce these correlated variables into fewer orthogonal variables. Four principle components obtained after principle components analysis explained 83% of the total variance contributed by 22 lag variables. So the data is reduced from 22 correlated variables to 4 independent variables. The RH value for the current time is regressed on four predicted RH values. These predicted are obtained from the components matrix of these four principle components. The regression is significant with an R^2 of 0.527. The first principle component has highest factor loading on RH with lag 9, lag 10, lag 11, lag 12 and lag 13. Here the values of factor loadings are higher than 0.70 but less than 0.75. The

second principle component has highest factor loading on RH with lag 17, 18 and 19. Here the factor loadings are higher than 0.6 but less than 0.65. The third component has high loadings on RH with lag 20, 21 and 22 with factor loadings higher than 0.5 and less than 0.55. The fourth principle component has highest factor loading RH with lag 1 and lag 2 with factor loadings higher than 0.40 but less than 0.5.

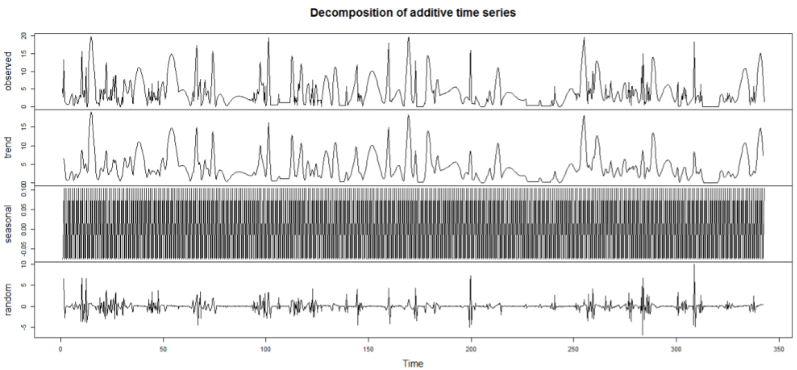

Figure 5. Decomposition of additive time series

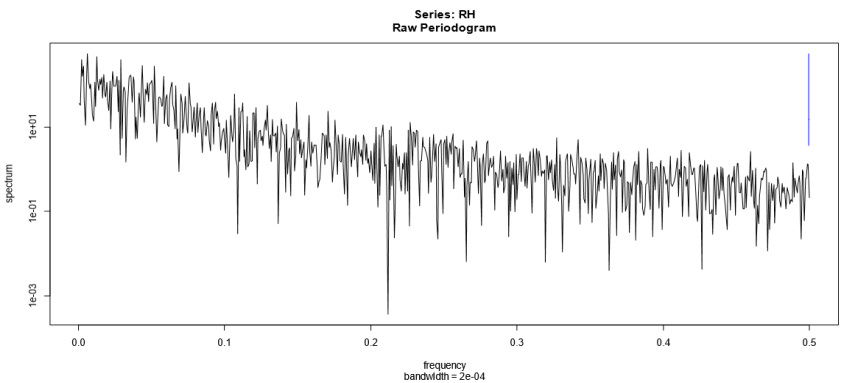

Figure 6. Periodogram of RH time series logrithmic scale

g. **Decomposition of Time Series:** To find out whether the three components of the time series are additive or multiplicative in nature, the residual sum of squares from the above mentioned two decompositions were found. The residual sum of square when additive model is fitted

to the 1381 equally spaced time series was 5.712423e-29. After fitting a multiplicative model to this time series, the residual sum of squares was 4.627641e-28. Lower value of residual sum of squares indicates that additive model is suitable. The decomposition of the time series using additive model is shown in Figure 5. So we can conclude that RH=Trend+Seasonal+Random. This decomposition is done with a frequency of 4 reflecting four quarters of the day.

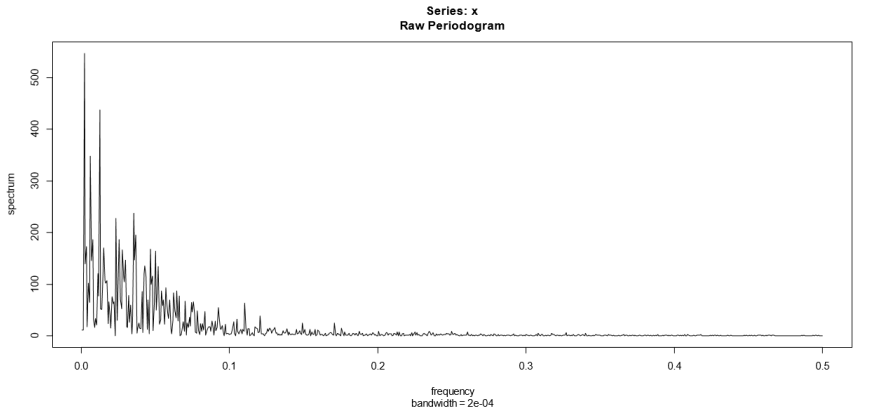

Figure 7. Periodogram of RH time series on original scale

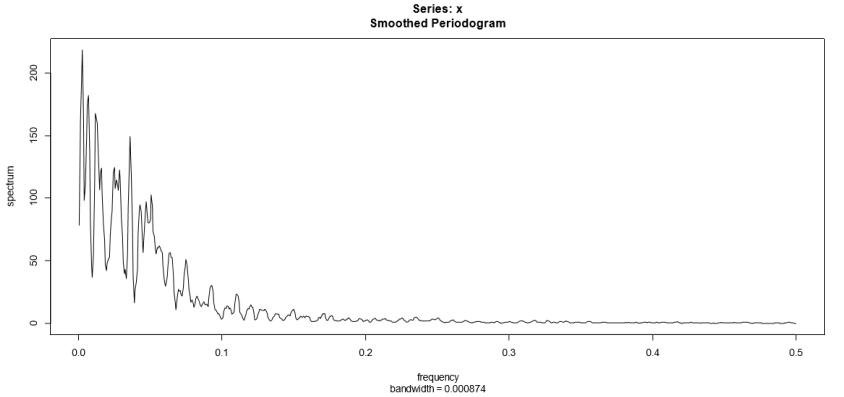

Figure 8. Smoothed periodogram of RH time series on original scale

(iv) **Spectrum** of 75 hours time series data on radiant heat RH is analyzed here. The spectrum of the data has a period is 360 minutes. This result seems plau-

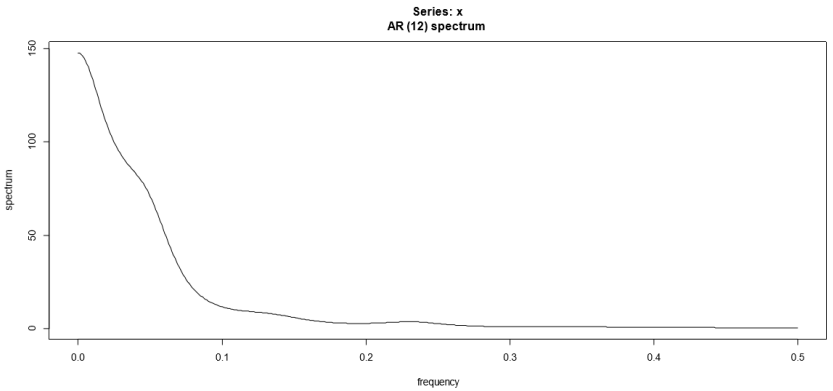

Figure 9. Spectrum of auto regressive process of RH data in original scale

sible as this implies that the lights dim and brighten after a period of 6 hours. These lights are caught by the satellite. At the source they are either the night lights or reflection of light from various surfaces during the day time. The spectrum of the logarithm of RH data is given in Figure 6 and original RH in Figure 7. From this figure we get an idea about the value of spectral coefficients that are significantly different from zero. At low frequency the spectrum values are much higher than zero. As the frequency increases the spectrum values becomes closer to zero. Here each frequency denotes number of cycles in 1381 minutes. One day comprises of 1368 minutes with outliers removed. So frequency 1 is $1/1381 = 0.000724$ and frequency 2 is 2/1381 and frequency k is k/1381 and so on. The angular frequency is $(2 * \pi)/1381 = 0.00455$ radians per day. Fundamental period is 1381 minutes. This periodogram is smoothened in Figure 8. The periodograms in Figure 6 and Figure 7 compare spectrum values in logarithmic and original scale. The total area under the curve given in Figure 7 gives the total variance of RH data over a period of one day. Higher values at low frequency in Figure 6 and 7 also indicate that low frequencies add the most to the variance of this process. The smoothed periodogram on original scale is given in Figure 8. The maximum value of the spectrum is 218.2879. This corresponds to the frequency 0.003 which 4 th minute of the day 4 September 2018. The corresponding period is 360 minutes, equivalent to 6 hours. The spectrum of original RH data as an autoregressive process is shown in Figure 9. We see as observed before that the spectral values are high for lower frequencies and low for higher frequencies.

Table 8. ANOVA of mean RH with respect to clusters

RH	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	998.601	2	499.301	209.754	.000
Within Groups	1223.531	514	2.380		
Total	2222.132	516			

Table 9. ANOVA of mean RH with respect to 3 days

RH	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5.622	2	2.811	.640	.528
Within Groups	2125.746	484	4.392		
Total	2131.368	486			

Table 10. Values of augmented Dickey Fueller's test unit root test

Variable	Dickey Fuller's Test Statistic	P value	Remarks
RH	-6.3236	3.454e-10	Stationary
First Order Difference of RH	-31.5316	< 2.2e-16	Stationary

4 Conclusion

Here dynamics of change in RH time series data is statistically analyzed. Unlike other papers, the nightfire data is not used here in remote sensing of volcanos, burning biomass or gas. Here the behavior of nightfire radiant heat variables RH is analyzed in detail in general and also with respect to time. The geographical location of the burning source of RH is not taken into consideration. Time Series Analysis, Principle Components Analysis, Discriminant Analysis, Cluster Analysis and Multiple Regression are statistical tools used here to analyze these data. These data are collected from 21:16 hours on 2 September 2018 to 2:27 hours on 6 September 2018. Discriminant Analysis shows that Colour (dark and light) is the attribute that discriminates this five days and 75 hour RH data of 28252 observations. The light and dark are the colours outside the satellite station during the time of recording of the data. Here light signifies time from 3:00 A.M. to 7:00 P.M that is day time; dark signifies time from 7:00 P.M. to 3:00 A.M. that is night. Discrimant Analysis revealed that the data can be discriminated as collected during light or dark period with 66.9% accuracy. Cluster analysis of this five days data set shows that the data can be divided into three clusters. These clusters are ho-

mogeneous within and heterogenous between. Cubic spline interpolation is used to create a time series of mean RH for every minute of 4 September 2018. Auto Correlation Function shows a significant auto correlation upto a lag of 22 minutes. These 22 lag variables have a very significant correlation with each other. Principle Component Analysis identified four components on the basis of 22 lag variables of mean RH data. These components explained 83% of the total variance contributed by 22 these lag variables. These highly correlated readings of mean RH data recorded in last 22 minutes are reduced to four orthogonal principal components. This shows that the mean RH data recorded at current minute is dependent on the RH readings of last 22 minutes. The current minute reading of mean RH data is also regressed on these 22 lag variables using Multiple Linear Regression. The p value of this regression is 2.2e-16 with coefficient of determination $R^2 = 0.9031$. This shows that this regression is highly significant and 90.31% variance of the mean RH data is explained by this regression. Principal components analysis reduced the independent variables from 22 to 4. The regression of current reading of RH on four principal components had $R^2 = 0.527$. The decomposition of the time series revealed that additive model is more suitable than multiplicative model. The spectrum of this data has a period is 360 minutes. This result seems plausible as this implies that the lights dim and brighten after a period of 6 hours. These lights are caught by the satellite. At the source they are either the night lights or reflection of light from various surfaces during the day time.

Bibliography

[1] World Bank, Night time lights revisited, *Policy research working paper*, 2015, 7496.

[2] C. D. Elvidge, M. Zhizhin, F-Ch. Hsu and K. E. Baugh, VIIRS nightfire: Satellite pyrometry at night, *Remote Sens.* **5** (2013) 4423-4449, DOI: 10.3390/rs5094423.

[3] K. Roychowdhury and S. Jones, Nexus of health and development: Modelling crude birth rate and maternal mortality ratio using night time satellite images, *ISPRS International Journal of Geo-Information* **3** (2014) 693-712, DOI: 10.3390/ijgi3020693.

[4] E. Melas , M. Mann and A. Malik, Using VIIRS day/night band to measure electricity supply reliability preliminary results from Maharashtra, *India, Remote Sens.* **8** (2016) 7-11, DOI: 10.3390/rs8090711.

[5] B. Yu, K. Shi, Y. Hu, C. Huang, Z. Chen and J. Wu, Poverty evaluation using NPP-VIIRS nighttime light composite data at the county level in China, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **8(3)** (2015) 1217-1229, DOI: 10.1109/JSTARS.2015.2399416.

[6] R. C. Sharma, R. Tateishi, K. Hara, S. Gharechelou and K. Iizuka, Global mapping of urban built-up areas of year 2014 by combining MODIS multispectral data with

VIIRS night time light data, *International Journal of Digital Earth* **9(10)** (2016) 1004-1020, DOI: 10.1080/17538947.2016.1168879.

[7] Y. Dou, Z. Liu, C. He and Huanbi Yue, Urban land extraction using VIIRS night time light data: An evaluation of three popular methods, *Remote Sens.* **9(2)** (2017) 13-25, DOI: 10.3390/rs9020175.

[8] C. Elvidge, M. Zhinzhin, F-Ch. Hsu and K. Baugh, What is so great about nighttime VIIRS data for the detection and characterization of combustion sources, *Proceedings of Asia Pacific Advanced Network* **35** (2013) 33-48, DOI: 10.7125/APAN.35.5.

[9] C. D. Elvidge, K. Baugh, M. Zhizhin, F. C. Hsu and T. Ghosh, VIIRS night-time lights, *International Journal of Remote Sensing* **38(21)** (2017) 5860-5879, DOI:10.1080/01431161.2017.1342050.

[10] W. Guo, D. Lu, Y. Wu and J. Zhang, Mapping impervious surface distribution with integration of SNPP VIIRS-DNB and MODIS NDVI data, *Remote Sens.* **7(9)** (2015) 12459-12477, DOI: 10.3390/rs70912459.

[11] H. Letu, T. Y. Nakajima and F. Nishio, Regional-scale estimation of electric power and power plant CO2 emissions using defense meteorological satellite program operational linescan system nighttime satellite data, *Environ. Sci. Technol. Letter.* **1(5)** (2014) 259-265, DOI: 10.1021/ez500093s.

[12] N. A. Rybnikova and B. A. Portnov, Outdoor light and breast cancer incidence: a comparative analysis of DMSP and VIIRS-DNB satellite data, *International Journal of Remote Sensing* **38(21)** (2017) 5952-5961, DOI: 10.1080/01431161.2016.1246778.

[13] K. Shi, B. Yu, Y. Hu, C. Huang, Y. Chen, Y. Huang, Z. Chen and J. Wu, Modelling and mapping total freight traffic in China using NPP-VIIRS night time light composite data, *GIScience and Remote Sensing* **52(3)** (2015) 274-289, DOI: 10.1080/15481603.2015.1022420.

[14] A. Zeileis, J. Verbesselt and M. Herold, Near real-time disturbance detection using satellite image time series, *Remote Sensing of Environment* **123** (2012) 98-108, DOI: 10.1016/j.rse.2012.02.022.

[15] S. S. Sastry, *Introductory Methods of Numerical Analysis*, PHI Learning Private Limited, 185-187, 2017.

[16] G. Jenkins, G.Box and G. Reinsel, *Time Series Analysis*, John Wiley and Sons, 7 - 41, 2008.

[17] K.C. Bhuyan, *Multivariate Analysis and its Applications*, New Central Book Agency (P) Ltd, 171-233, 2010.

[18] G. M. Trifonov, M. N. Zhizhin, D. V. Melnikov and A. A. Poyda, VIIRS Nightfire remote sensing volcanoes, *Procedia Computer Science* **119** (2017) 307-314.

[19] F. Facchinelli, S. E. Pappalardo, D. Codato, A. Diantini, G. D. Fera, E. Crescini and M. De Marchi, Unburnable and unleakable carbon in Western Amazon: Using VIIRS nightfire data to map gas flaring and policy compliance in the YasunÃbiosphere reserve, *Sustainability* **12(58)** (2020) 1-26.

[20] X. Zhang, B. Shavings, B. Shoghli, C. Zygarlicke and C. Wocken, Quantifying gas flaring and CH4 consumption using VIIRS, *Remote Sensing* **7(8)** (2015) 9529-9541.

Received March 15, 2020; revised November 12, 2020; accepted December 1, 2020.

Author information

Jyoti U. Devkota, Department of Mathematics, Kathmandu University, Nepal.
E-mail: drjdevkota@ku.edu.np