

Spatial estimation of chronic respiratory diseases based on machine learning procedures—an approach using remote sensing data and environmental variables in quito, Ecuador



Cesar I. Alvarez-Mendoza ^{a,b,*}, Ana Teodoro ^{b,c}, Alberto Freitas ^{d,e}, Joao Fonseca ^{d,e}

^a Universidad Politécnica Salesiana, Grupo de Investigación Ambiental en El Desarrollo Sustentable GIADES, Carrera de Ingeniería Ambiental, Quito, 170702, Ecuador

^b University of Porto, Department of Geosciences, Environment and Land Planning, Faculty of Sciences, Rua Campo Alegre 687, Porto, 4169-007, Portugal

^c Earth Sciences Institute (ICT), Pole of the FCUP, University of Porto, Porto, 4169-007, Portugal

^d Department of Community Medicine, Information and Health Decision Sciences (MEDCIDS), Faculty of Medicine, University of Porto, Rua Dr. Plácido da Costa, 4200-450, Porto, Portugal

^e CINTESIS - Center for Health Technology and Services Research, Faculty of Medicine, University of Porto, Rua Dr. Plácido da Costa, 4200-450, Porto, Portugal

ARTICLE INFO

Keywords:

Remote sensing
Machine learning
Respiratory disease
Spatial models
Quito

ABSTRACT

Over the last few years, the use of remote sensing data in different applications such as estimation of air pollution concentration and health applications has become very popular and new. Thus, some studies have established a possible relationship between environmental variables and respiratory health parameters. This study proposes to estimate the prevalence of Chronic Respiratory Diseases, where there is a relationship between remote sensing data (Landsat 8) and environmental variables (air pollution and meteorological data) to determine the number of hospital discharges of patients with chronic respiratory diseases in Quito, Ecuador, between 2013 and 2017. The main objective of this study is to establish and evaluate an alternative LUR model that is capable of estimate the prevalence of chronic respiratory diseases, in contrast with traditional LUR models, which typically assess air pollutants. Moreover, this study also evaluates different analytic techniques (multiple linear regression, multi-layer perceptron, support vector regression, and random forest regression) that often form the basis of spatial models. The results show that machine learning techniques, such as support vector machine, are the most effective in computing such models, presenting the lowest root-mean-square error (RMSE). Additionally, in this study, we show that the most significant remote sensing predictors are the blue and infrared bands. Our proposed model is a spatial modeling approach that is capable of determining the prevalence of chronic respiratory diseases in the city of Quito, which can serve as a useful tool for health authorities in policy- and decision-making.

1. Introduction

During the last few years, remote sensing data have increasing been used in monitoring, spatial predictive modeling, surveillance, and risk assessment with respect to human health (Viana et al., 2017). These human health studies have also been associated with air pollution spatial modeling, which is connected to some vector-borne (Estallo, Benítez, Lanfri, Scavuzzo, & Almirón, 2016) and respiratory diseases (Ayres-Sampaio et al., 2014). In this context, spatial models relying on remote sensing data have been developed to identify different air pollutants, the most common of which are particulate matter less than 2.5

and 10 µm (PM2.5 and PM10, respectively) (Alvarez-Mendoza, Teodoro, Torres, et al., 2019; Alvarez et al., 2016), nitrogen dioxide (NO₂) (C. Zheng et al., 2018), tropospheric ozone (O₃) (Alvarez-Mendoza, Teodoro, & Ramirez-Cando, 2019), sulfur dioxide (SO₂) (Ibrahim Sameen et al., 2014), and carbon dioxide (CO₂) (Chejlarla, Maheshuni, & Mandla, 2016). The aforementioned air pollutants are greenhouse gases and precursors of global warming (EPA, 2017). Moreover, evidence of the adverse effects of exposure to air pollutants (PM2.5, PM10, O₃) on health has been collected in several countries around the world (Di et al., 2017). Specifically, air pollution is a threat to respiratory health, and several chronic respiratory diseases (CRDs), such as asthma, chronic

* Corresponding author. Universidad Politécnica Salesiana, Grupo de Investigación Ambiental en el Desarrollo Sustentable GIADES, Carrera de Ingeniería Ambiental, Quito, 170702, Ecuador

E-mail addresses: calvarezm@ups.edu.ec, up201510599@fc.up.pt (C.I. Alvarez-Mendoza), amteodor@fc.up.pt (A. Teodoro), alberto@med.up.pt (A. Freitas), jfonseca@med.up.pt (J. Fonseca).

obstructive pulmonary disease, and others, represent nearly 6% of global annual deaths (O'Connor et al., 2008; WHO, 2019). According to the World Health Organization (WHO), 92% of people around the world live in places with poor outdoor air quality, where the main risk factors of developing a CRD are related to the climate and the environment (Barry & Annesi-Maesano, 2017; WHO, 2019).

One of the most famous missions in satellite remote sensing is the Landsat program, launched in 1972. The most recent program satellite is the Landsat 8, which has the Operational Land Imager (OLI) and Thermal Infrared Sensor (Wulder et al., 2019) on board. This satellite provides a wide spatial-temporal perspective of the Earth, enabling a variety of applications and retrieving several variables, such as vegetation, land use, aerosol particles, and environmental and meteorological information, which can be retrieved and analyzed (Department of the Interior U.S. Geological Survey, 2016). Due to the potential of the variables collected by remote sensing of the Earth's environment (Chuvieco, 2016), it is possible to develop models to analyze air pollutants. Such models typically use air quality monitoring network (AQMN) data and remote sensing variables to conduct spatial modeling of air pollutants, using remote sensing-derived parameters in the form of environmental indexes, such as the normalized difference vegetation index (NDVI) (Zhai, Sang, Zhang, & An, 2015), and measures, such as aerosol optical thickness or depth (AOT or AOD) (Liu, Franklin, Kahn, & Koutrakis, 2007; Sicard, Anav, De Marco, & Paoletti, 2017). It is important to note that sensors, such as the MODIS instruments on Terra and Aqua and Landsat 8's OLI, allow us to obtain (directly or indirectly) this information. The Terra/Aqua MODIS instruments have a new AOT product (MODIS Multi-Angle Implementation of Atmospheric Correction - MAIAC AOD) with a low spatial resolution (1×1 km), which is ideal for regional studies (Lyapustin et al., 2012). Landsat 8's OLI is capable of retrieving AOT in fine spatial resolution (30 m); however, AOT information from ground stations is also needed (Bilal, Nichol, Bleiweiss, & Dubois, 2013). AOT measurements are retrieved by the blue and red bands of Landsat 8's OLI (Vermote, Justice, Claverie, & Franch, 2016); its infrared bands are also used to retrieve O_3 measurements (Famoso et al., 2017; S.; Zheng et al., 2017).

With respect to health studies based on remote sensing data, predictive models have been used to analyze air pollutants by combining geographic variables (traffic, land use, population, etc.) with remote sensing data, in the form of land use regression (LUR) models (Habermann, Billger, & Haeger-Eugensson, 2015a; Saucy et al., 2018). Thus, a LUR model could potentially be used to relate the environmental variables with the hospital discharge rates (Anderson et al., 2012; Ayres-Sampaio et al., 2014). A hospital discharge is defined as the release of a patient who has stayed at least one night in the hospital, including people who die in hospital care (Health at a Glance 2017, 2017). However, most LUR models do not consider the dynamics of geographical variables, because such variables are sometimes out of date or obsolete (Kreis, Nieuwenhuijsen, Kreis, & Nieuwenhuijsen, 2017). Some health studies have related hospital discharge with exposure to different traffic-related pollution, in which the NDVI and MODIS AOT are the most commonly used predictors. These studies aim to find a possible relationship between air pollution exposure and hospital discharge (Chang et al., 2019; Sorek-Hamer, Just, & Kloog, 2016).

LUR models use analytic techniques, such as multiple linear regression (MLR), stepwise regression, and multiple logistic regression (Cesar Ivan Alvarez et al., 2019; Chan et al., 2008; Heo & Bell, 2019). However, these techniques do not analyze the correlation between predictors, and it is well known that remote sensing variables have a high correlation or multi-collinearity (Jin, Wang, Jin, & Wang, 2019). An alternative to MLR is the use of more complex models, such as machine learning techniques (MLTs) in order to avoid multi-collinearity and other limitations presented in the linear LUR models. Examples of non-linear MLTs are multilayer perceptron (MLP), support vector regression (SVR) and random forest regression (RFR). Regarding MLP, several studies use this technique to build air pollutants models in order to map air quality in

cities, with a clear relationship with public health (Azeez et al., 2019; Li, Zhou, & Tong, 2019). SVR are often used to represent distribution of epidemiology spatial models using environmental variables (Ahangarcani, Farnaghi, Shirzadi, Pilesjö, & Mansourian, 2019). RFR are applied to assess new LUR model alternatives to PM associating with health effects (Brokamp, Jandarov, Rao, LeMasters, & Ryan, 2017).

In this context, the aim of this study is to estimate and compare spatial empirical models, based on LUR models, that are capable of determining the number of hospital discharges of patients with CRDs (HCRD) in Quito, Ecuador, between 2013 and 2017, using remote sensing data, air pollution field measurements, and meteorological data as predictors and considering three different complex machine learning techniques: MLP, SVR, and RFR. The spatial model selected will allow us to map the prevalence of HCRD. This approach will provide insight into and an understanding of the most significant spatial predictors and the spatial distribution of HCRD in the city of Quito. Furthermore, the present study is an innovative approach to the use of remote sensing data in human health studies.

2. Materials and methods

2.1. Study area

The study area is the most populated zone of Quito, Ecuador. The area is divided into 45 administrative urban districts. Its latitude is $0^{\circ}30'S$ to $0^{\circ}10'N$, its longitude is $78^{\circ}10'W$ to $78^{\circ}40'W$, and the equatorial line crosses through it (Fig. 1). Quito's annual median temperature is $17^{\circ}C$, and its elevation is about 2800 m above sea level. This specific study area was chosen for the following reasons: (i) it is covered by nine AQMN stations; (ii) its road traffic is relatively high; and (iii) the urban downtown is located in this area. The influence zones of each AQMN station were established through Thyssen polygons from AQMN dot location and their respective intersection with the urban districts. Thus, each AQMN influence area contains several districts in order to merge the data, representing each area as a unit in this study.

2.2. Data collection

2.2.1. Remote sensing data

Between 2013 and 2017, 46 Landsat 8 level 2 images were acquired over the study area. The on-demand images were obtained from the Land Satellite Data Systems Science Research and Development website (<https://espa.cr.usgs.gov/>). The main advantage of level 2 images is that they use the Landsat 8 Surface Reflectance Code to generate products with geometrical, radiometric, and atmospheric corrections (USGS, 2019). These products have a spatial resolution of 30 m. The products used in this study as predictors are the surface reflectance OLI bands, the top of the atmosphere, brightness temperature (BT), and some pre-processed indexes, such as the NDVI (Tucker, 1979), the soil-adjusted vegetation index (SAVI) (A. . Huete, 1988), and the enhanced vegetation index (EVI) (A. Huete et al., 2002). Moreover, considering the high cloud density in the Andean Region (César I Alvarez, Teodoro, & Tierra, 2017), the images were filtered, and only images with a maximum of 10% cloud density over the study area were considered.

BT was converted to land surface temperature (LST) using the emissivity equation according to (Sobrino, Jiménez-Muñoz, & Paolini, 2004, 2008) and the inversion of Planck's function, as shown in Equation (1):

$$LST = \frac{BT}{\left(1 + \left(\frac{\lambda * BT}{\rho}\right) \ln e\right)} - 273.15 \quad (1)$$

where BT is obtained from Landsat 8 level 2 images in kelvin degree (K), λ is the center wavelength of the Landsat 8 TIR 1 band ($10.8 \mu m$)

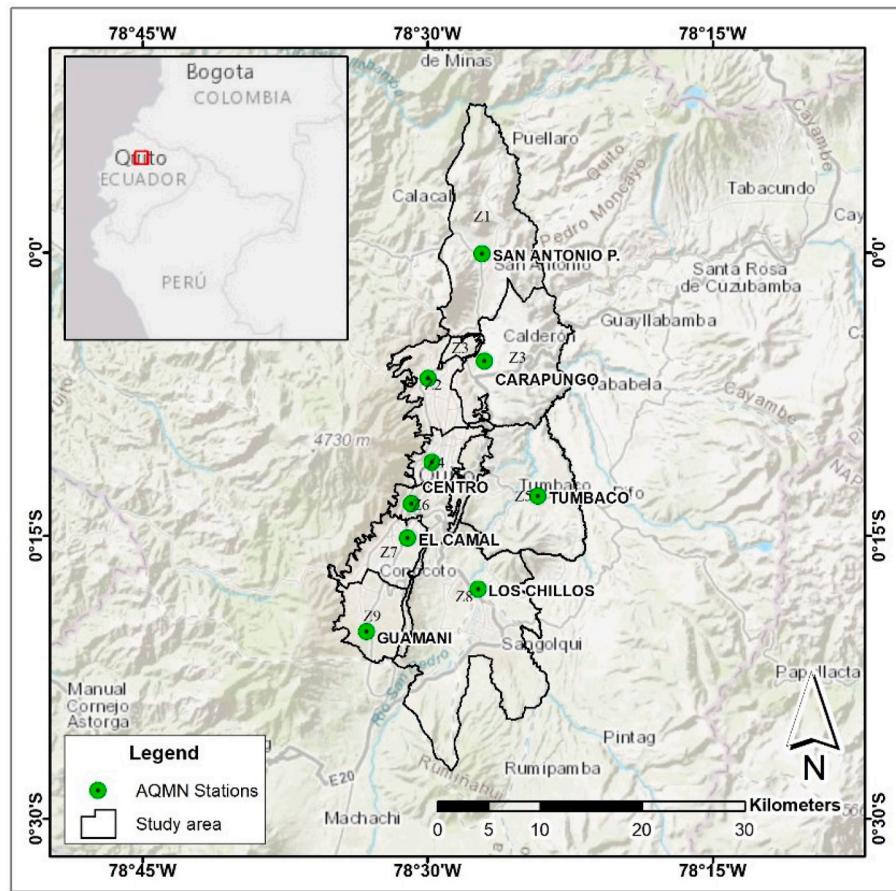


Fig. 1. The study area (Quito, Ecuador). The green dots represent the air quality monitoring network (AQMN) stations and their influence areas in black outline polygons.

(Ghaleb, Mario, & Sandra, 2015), ρ is expressed in Equation (2), and ε is the emissivity derived from Equation (3), which has to be selected according to the NDVI evaluation in the study area. The result is the LST in degrees Celsius ($^{\circ}\text{C}$).

$$\rho = \frac{h^*c}{s} \quad (2)$$

where h represents Planck's constant ($6.63e-34 \text{ Js}$), c is the speed of light ($2.99e-8 \text{ ms}^{-1}$), and s is the Boltzmann constant ($1.38e-23 \text{ m}^2\text{kgs}^{-2}\text{K}^{-1}$).

$$\varepsilon = \begin{cases} \varepsilon_s, & \text{NDVI} < \text{NDVI}_s \\ \varepsilon_s + (\varepsilon_v - \varepsilon_s)P_V, & \text{NDVI}_s \leq \text{NDVI} \leq \text{NDVI}_v \\ \varepsilon_v, & \text{NDVI} > \text{NDVI}_v \end{cases} \quad (3)$$

where ε_s is the emissivity for the soil (0.973) and ε_v is the emissivity for the vegetation (0.985) (Sobrino et al., 2008). NDVI_v is the NDVI for the vegetation (0.2), and NDVI_s is the NDVI for the soil (0.5) (Vieira, Teodoro, & Gomes, 2016). P_V represents the proportion of vegetation in the study area according to Equation (4).

$$P_V = \left(\frac{\text{NDVI} - \text{NDVI}_s}{\text{NDVI}_v - \text{NDVI}_s} \right)^2 \quad (4)$$

Moreover, the cloud pixels are removed from each satellite image considering the information available in the level 2 pixel quality band (Band QA), considering a pixel filter. Before aggregate the data, a filter in RStudio was computed to every pixel in each image. The null and cloud values were removed using the quality band of Landsat-8, in order to avoid the noise pixels. This filter was an additional alternative to clean the pixel data, because this study only considers images with a maximum of 10 percent of cloud density. All the processes were

computed on RStudio with the raster v2.9-5 package.

2.2.2. Field measurement data

Most of the models that calculate air pollutants require airfield measurements. In this work, field data were obtained from the Quito AQMN, known as "Red Metropolitana de Monitoreo Atmosférico de Quito" (REMMAQ) (Secretaría del Ambiente de Quito, 2018a; Secretaría del Ambiente de Quito, 2018b). This AQMN has been in operation since 2002, providing hourly field measurements of air pollutants and meteorological variables. REMMAQ has nine georeferenced stations (Fig. 1), which collect the following air pollution variables of interest to this study: carbon oxide (CO), PM2.5, PM10, SO₂, O₃, and NO₂. The following meteorological variables were considered in this study: pressure, wind direction, relative humidity, precipitation, wind speed, air temperature, and solar irradiance. The Environmental Secretary of Quito manages the REMMAQ, and the data are available to download for free on her website (<http://www.quitoambiente.gob.ec/ambiente/index.php/datos-horarios-historicos>). Spatial air pollutant raster for each trimester of every year were computed using the inverse distance weight (IDW) algorithm with a 30 m spatial resolution and a power of 2.0 (Jumaah, Ameen, Kalantar, Rizeei, & Jumaah, 2019; Lee et al., 2018; Rivera-González et al., 2015). All the information was processed with the R packages *rgdal* v1.4-4 and *gstat* v2.0-2. On the other hand, the AQMN influences areas were computed using the Thyssen polygon from the AQMN points shapefile and intersecting with the district shapefile.

2.2.3. Hospital discharges of patients with chronic respiratory diseases

The National Institute of Statistics and Census (INEC) is the official government institution in Ecuador in charge of collecting and

disseminating information about population and other socioeconomic statistics and variables. This information is public and available on a district scale (<http://www.ecuadorencifras.gob.ec/camas-y-egresos-hospitarios/>). One of the variables included in this information is the number of hospital discharges (the number of released patients who stayed at least one night in the hospital, including people who died in hospital care) organized by their home district. This variable is classified according to the International Classification of Diseases 10th version (ICD-10) from the WHO (WHO, 2016). Considering the aim of this study, only hospital discharges of patients with CRDs were considered—those with ICD-10 classification codes of J40–J47, between 2013 and 2017. This filter includes the most significant CRDs, such as asthma and bronchitis. A summary of hospital discharges in each AQMN area of influence was computed for each trimester of each year. The main reason to group the dataset by trimester was the availability of matched data. Furthermore, population data are necessary to compute HCRD (the number of patients per 10,000 people who are admitted to the hospital with a CRD) to compare the different urban districts. This variable is a continuous dependent variable. Finally, we group the district variable considering the intersection with the AQMN area of influence.

2.3. Input dataset

In order to compile a unique dataset encompassing the remote sensing data, environmental variables (air pollution and meteorological field data), and HCRD, all the variables were correlated by trimester, year, and AQMN area of influence between 2013 and 2017.

We compute a mask of each AQMN area with each 30-m raster variable (See section 2.2). Then, the median pixel value to each AQMN area was estimated. The null or black pixel values are filtered before entry in the median calculation. On the other hand, the CRDs variable is obtained from the merging of districts. Each AQMN area of influence contains several districts. Table 1 shows the variables used in this study and their respective statistics. Moreover, Fig. 2 shows the correlation matrix in order to define the collinearity between variables (see Table 2).

2.4. Models establishment

A LUR model is an empirical model that considers some geographical predictors as independent variables and a dependent variable. The first

Table 2
Mean of RMSE and R² for all the models tested.

Model	RMSE Test Data	R ² Test Data
Multiple Linear Regression (MLR)	3.619	0.327
Multilayer Perceptron (MLP)	3.052	0.254
Support Vector Regression (SVR)	2.043	0.506
Random Forest Regression (RFR)	3.046	0.119

step in establishing such a model is the selection of the input predictors. The simplest model with the least number of independent variables should be found in order to avoid overfitting. Here, the Bayesian information criterion (BIC) was considered to conduct backward elimination, by which the lowest BIC values were used to choose the predictors (Chen et al., 2014; Zhang, 2016). Moreover, the input dataset was tested to define the existence of spatial autocorrelation and heterogeneity using the Moran's Index (Bertazzon, Johnson, Eccles, & Kaplan, 2015).

Then, the models were computed, considering different MLTs in order to compare linear (MLR) and non-linear regression models (MLP, SVR, and RFR). In each model, 90% of the dataset was used as training data, and 10% of the dataset was used as test data in a 10-fold cross validation.

MLR is probably the simplest and most common analytic technique used in building a predictive model. It computes a linear relationship between the independent (predictors) and the dependent variables (Olive, 2017). However, MLR does not analyze the correlation between predictors—a major limiting factor when considering remote sensing variables (G. Chen & Meentemeyer, 2016), which are highly correlated. In contrast, MLP with a back-propagation learning process is classified as an artificial neural network (ANN) model, and it can be used in the classification of remote sensing data. MLP uses a series of neuronal activities where the ideal is to have interconnection weights in a multi-layer perceptron (Mather & Tso, 2003). In this study, a non-linear MLP with an architecture defined by a hidden layer and four hidden nodes was computed according to (Sheela & Deepa, 2013) and evaluated. Moreover, the maximum step used was 1e+08 in order to find the convergence of the model and a learning rate with a value of 0.01. The threshold is a non-linear function with a backpropagation algorithm. The R package *neuralnet v1.44.2* was used to compute the MLP. SVR is a non-linear transformation of an MLT, which works as a support vector

Table 1
Descriptive statistics of the input variables.

No.	Variable	Min.	Max.	Mean	Median	Units/scale
0	HCRD	0.334	23.433	4.463	3.689	Hospital discharges per 10,000 people with chronic respiratory disease
1	Coastal aerosol band (B1)	0.029	0.077	0.056	0.060	Surface reflectance (0–1)
2	Blue band (B2)	0.034	0.095	0.068	0.072	Surface reflectance (0–1)
3	Green band (B3)	0.062	0.136	0.098	0.101	Surface reflectance (0–1)
4	Red band (B4)	0.050	0.149	0.105	0.111	Surface reflectance (0–1)
5	Near-infrared (NIR) (B5)	0.182	0.291	0.231	0.228	Surface reflectance (0–1)
6	Short-wave infrared 1 (B6)	0.170	0.268	0.208	0.206	Surface reflectance (0–1)
7	Short-wave infrared 2 (B7)	0.092	0.218	0.159	0.163	Surface reflectance (0–1)
8	Normalized Difference Vegetation Index (NDVI)	0.171	0.721	0.359	0.312	0–1
9	Soil-Adjusted Vegetation Index (SAVI)	0.101	0.408	0.209	0.184	0–1
10	Enhanced vegetation index (EVI)	0.106	0.428	0.217	0.190	0–1
11	Land Surface temperature (LST)	15.031	39.758	26.232	26.299	degrees Celsius
12	Pressure (P)	712.945	761.178	740.476	741.018	mb
13	Wind direction (WD)	58.155	273.426	142.357	146.345	degrees
14	Relative humidity (RH)	49.140	84.582	69.190	72.632	percentage (%)
15	Precipitation (PR)	0.000	4.443	0.406	0.000	mm
16	Wind speed (WS)	0.879	2.482	1.686	1.743	m/s
17	Air temperature (AT)	11.749	17.421	14.957	15.041	degrees Celsius
18	Solar irradiance (SR)	0.092	278.691	166.728	215.724	W/m ²
19	CO	0.435	0.852	0.622	0.598	µg/m ³
20	NO ₂	11.458	35.256	23.055	22.169	µg/m ³
21	O ₃	7.518	44.055	22.786	22.130	µg/m ³
22	PM2.5	10.441	23.504	16.490	16.316	µg/m ³
23	SO ₂	0.839	7.829	3.459	3.273	µg/m ³

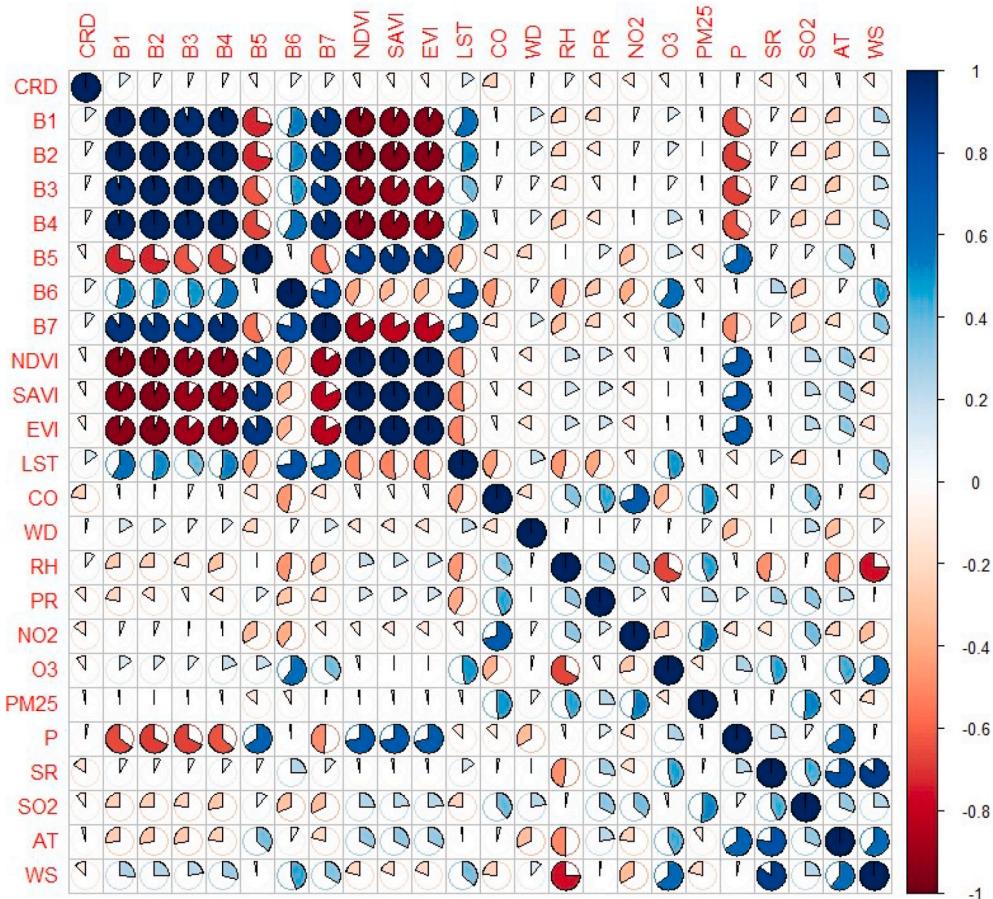


Fig. 2. Correlation graph between all the input variables.

machine classifier. They work in a higher dimensional space. The main difference is that SRV uses a continuous number as a dependent variable (Vapnik, 1998). The SVR uses as hyperparameters a radial kernel with a cost 1, gamma 0.1 and epsilon 0.1 in a beginning. Then, a tuning of the

SVR hyperparameters are considering in base to sensibility analysis. The R package used to compute SVR was *e1071 v1.7-2*. Finally, the last MLT employed was RFR. It is based on ensemble learning, which uses the training dataset to generate multiple decision trees, making it less

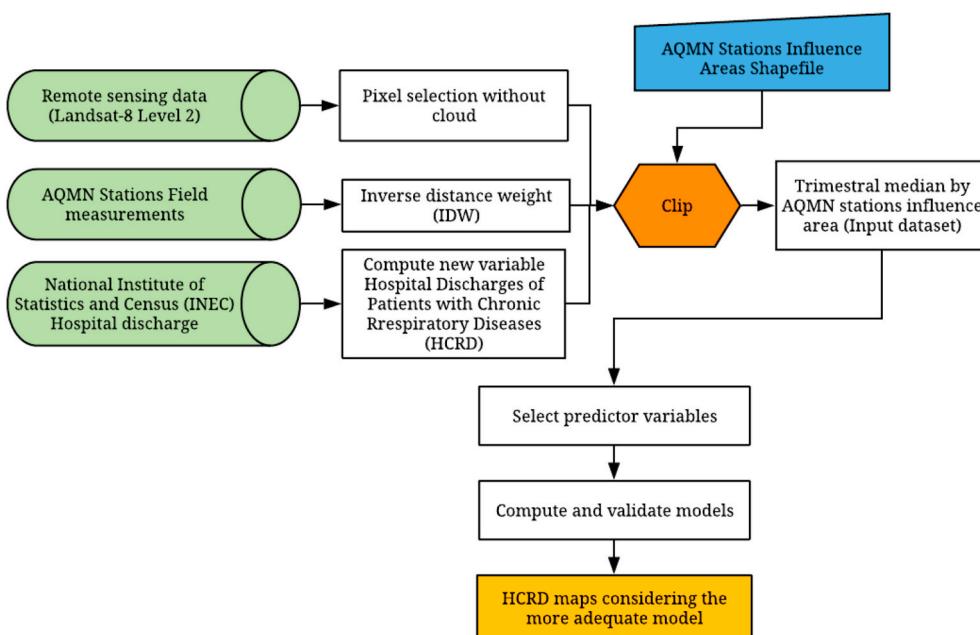


Fig. 3. Workflow of the methodology applied in this work.

sensitive to the overfitting problem. The decision trees are simply combined according to their weights. Moreover, RFR is considered to be one of the most effective non-parametric ensemble learning methods in image analysis (Ghorbanzadeh et al., 2019). Our RFM used 500 trees to set up the error. Moreover, the number of variables tried at each split are three. The R package *randomForest v4.6-14* was used to implement RFR in this study.

In the model evaluation, the coefficient of determination (R^2) between the observed values and the predicted values and the root-mean-square error (RMSE) were compared. Models (considering the test dataset) with a higher R^2 and lower RMSE were selected to develop a spatial map of HCRD for each trimester for each year. The final model developed a raster file with 30 m of spatial resolution. Fig. 3 shows the workflow of the methodology used in this study.

3. Results

3.1. Selected predictor variables

The final dataset considered 162 observations, which included all the variables (the remote sensing, environmental, and HCRD variables). The dataset consisted of 24 variables (one dependent variable and 23 predictors), including registers by trimester, year, and the AQMN area of influence. The lowest BIC values were chosen in order to consider only the most significant variables, avoiding multi-collinearity. Moreover, we tested the input dataset to define the existence of spatial autocorrelation and heterogeneity using the Moran's Index. The index results are values near to 0 in all the input variables, making a randomly dispersed. This could happen because we already considered clusters (AQMN area of influences) to group the input data (Bertazzon et al., 2015). A total of 10 predictors (B1, B2, B7, EVI, LST, RH, SR, AT, CO, and SO₂) were

considered as inputs in all the MLTs (p-value < 0.050). Equation (5) shows the MLR established with the 10 predictors considered:

$$\text{HCRD} = I + aB1 + bB2 + cB7 + dEVI + eLST + fRH + gSR + hAT + iCO + jSO_2 \quad (5)$$

where HCRD is the hospital discharges per 10,000 people with chronic respiratory disease; I is the intercept; a, b, c, d, and e are the coefficients in each predictor; and the other variables are described in Table 1. Thereby, the models can assess the use of other alternatives as remote sensing variables to predict epidemiology cases.

3.2. Comparison and evaluation of the models

The results of the MLP presented in Fig. 4 allow us to analyze the structure of the neural network used in the different k-fold cross validation. The diagram of the MLP is defined consider a hidden layer and four hidden nodes.

Fig. 5 shows the comparison between R^2 and RMSE for all the models established. The non-linear SVR model showed the best adjustment in the test data considering a 10-fold cross-validation. The SVR model presents the lowest RMSE (2.0439) and the highest R^2 (0.5066), considering the test data. Thus, this model, considering the test data, was used to map HCRD despite its limitations. Fig. 6 presents the SVR model by trimester (and year).

In function to the sensitivity, the relative importance variables is showed in Table 3 for the SVR model.

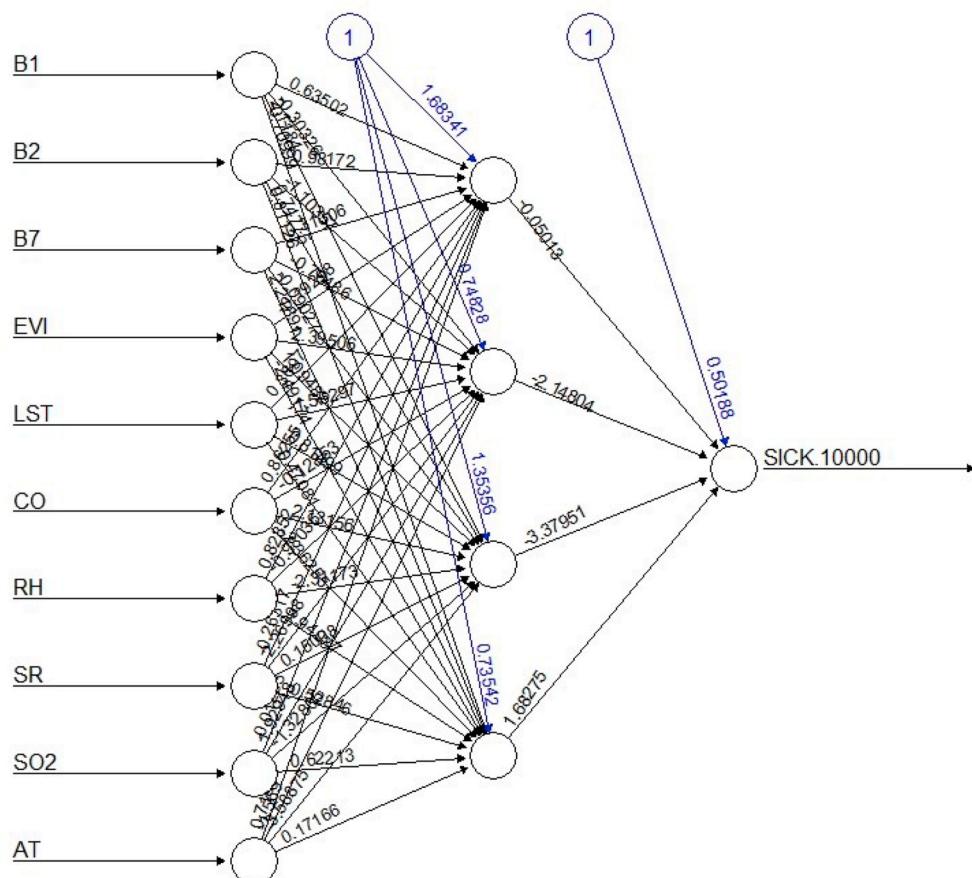


Fig. 4. MLP Diagram in a k-fold cross validation.

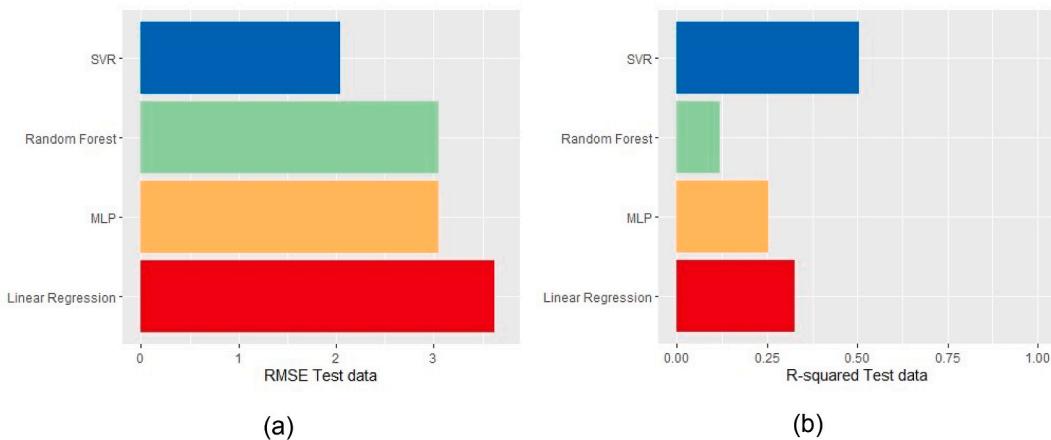


Fig. 5. Comparison between models considering the average of RMSE and R^2 using 10-fold cross validation: (a) RMSE test data; (b) R^2 test data.

4. Discussion

In most cases, LUR models are used to calculate air pollutants (Alvarez-Mendoza, Teodoro, & Ramirez-Cando, 2019; Alvarez-Mendoza, Teodoro, Torres, et al., 2019; Habermann, Billger, & Haeger-Eugensson, 2015b; Zou et al., 2016), and geographical variables, such as roads, traffic, land use, etc., are used to establish LUR MLR models. However, these models include geographic variables that are not always available or updated in a timely manner. Some studies have also compared the air pollution calculated by LUR models with health data (Ayres-Sampaio et al., 2014; Kallawicha et al., 2018). However, this comparison is only performed considering categorical variables (Fan, Li, Fan, Bai, & Yang, 2016) and not with numerical variables in order to quantify the value.

In this study, spatial models were developed to compute HCRD, considering remote sensing and environmental variables (air pollution and meteorological ground data) as predictors. The predictors were chosen considering their relationships with variables that may potentially affect respiratory health, such as vegetation, land use, climate, and air pollution. Air pollution is defined as the presence of one or more substances in the air (Humans, of, to, & Cancer, 2016). Some studies have shown that air pollution is a serious issue, posing a grave threat to respiratory health (Sweileh, Al-Jabi, Zyoud, & Sawalha, 2018). On the other hand, climate variables, specifically meteorological variables, such as temperature or humidity, have a direct influence on the potential to acquire CRDs (Arundel, Sterling, Biggin, & Sterling, 1986; D'Amato, Cecchi, D'Amato, & Annesi-Maesano, 2014). Moreover, using remote sensing data, it is possible to derive some of these environmental variables. One such example is the high correlation between the Landsat 8 blue and red bands with AOT (Vermote et al., 2016). As noted above, AOT influences the retrieval of air pollutants (Y. Chen et al., 2014; Gupta et al., 2006; Hadjimitsis, 2009). In this context, 24 predictors were considered as inputs in the original model with a matched dataset of 162 observations. With the use of BIC, 10 significant predictor variables were selected for use in the final spatial HCRD models using all the dataset. In some cases, if we choose in a separated way the feature selection, we can have a perturbation in the estimation of parameters (Hastie, 2009). The remote sensing variable predictors included the coastal aerosol, blue, and SWIR-2 bands (bands 1, 2, and 7, respectively). The blue band is more correlated with PM (Vermote et al., 2016) and the B7 with O₃ concentrations (Famoso et al., 2017; S.; Zheng et al., 2017). Additionally, EVI and LST were also selected. The environmental variable predictors were CO and SO₂, and the meteorological variables were HR, SR, and AT. Several studies have reported a correlation between these variables and the presence of CRDs (Kutlar Joss, Eeftens, Gintowt, Kappeler, & Künzli, 2017).

Four MLTs were selected to compute the model: (i) linear MLR; (ii)

non-linear MLP; (iii) SVR; and (iv) RFR. During the computations, 90% of the dataset was used as training data, and 10% of the dataset was used as test data, considering a 10-fold cross validation. The main advantage of non-linear MLTs is the avoidance of concerns regarding multicollinearity. Several studies have found that the use of MLP and SVR with remote sensing data improves the performance of regression models using ground data (Ayeju et al., 2019). RFRs are often implemented in prediction analyses, because they provide better accuracy (Hastie, Tibshirani, Friedman, & Franklin, 2005). The results show that the use of SVR created the most successful model. However, we applied a tuning in the SVR hyperparameters in order to increase the accuracy in the model. Thus, SVR had the highest R^2 (0.5066) and the lowest RMSE (2.0434) with the test data. This model helped to develop a spatial map of HCRD in different trimesters in the years between 2013 and 2017 (Fig. 6). In particular, the third trimester was selected for mapping, because there were more available images taken during that trimester during the five years of the study sample period. Additionally, there was more variation in the rates of hospital discharges in September according to the input data. It is also worth noting that there was a significant increase in reported air pollutant concentrations in some areas between 2013 and 2017 (Secretaría del Ambiente de Quito, 2018a). Indeed, in the north and east regions of the city, there were higher values of HCRD, which was likely due to the fact that these areas had higher rates of air pollutants, traffic, and population. Thus, these results allow us to identify possible trends in the growth patterns of CRDs in the next few years.

The main limitations of this study were as follows: (i) There were a limited number of satellite images available without high cloud density (César I Alvarez et al., 2017). In future research, a possible improvement could be the use of more remote sensors to combine data or to develop and apply new techniques to remove cloud interference (Alvarez-Mendoza, Teodoro, & Ramirez-Cando, 2018). (ii) The REEMAQ and INEC data were incomplete for some months and years. In some cases, the stations were unavailable or did not have complete quality data. On the other hand, some hospital discharge data were lacking information regarding location, or such information suffered from poor-quality codification or registration. We discarded these data in order to obtain a more accurate dataset; however, in future work, we will extend this study for a longer period of time in order to improve our models. (iii) The percentages of training and test data may not have been ideal, because the evaluation of such data can change the final results. Our future research could consider different cut-off points in the dataset. (iv) Despite the fact that the spatial HCRD maps gave us a general idea of the presence of CRDs and possible future trends, these maps must be improved with more input data.

Despite the limitations, this study could be applied in regions with few quantities of air quality and health studies, in order to have a more

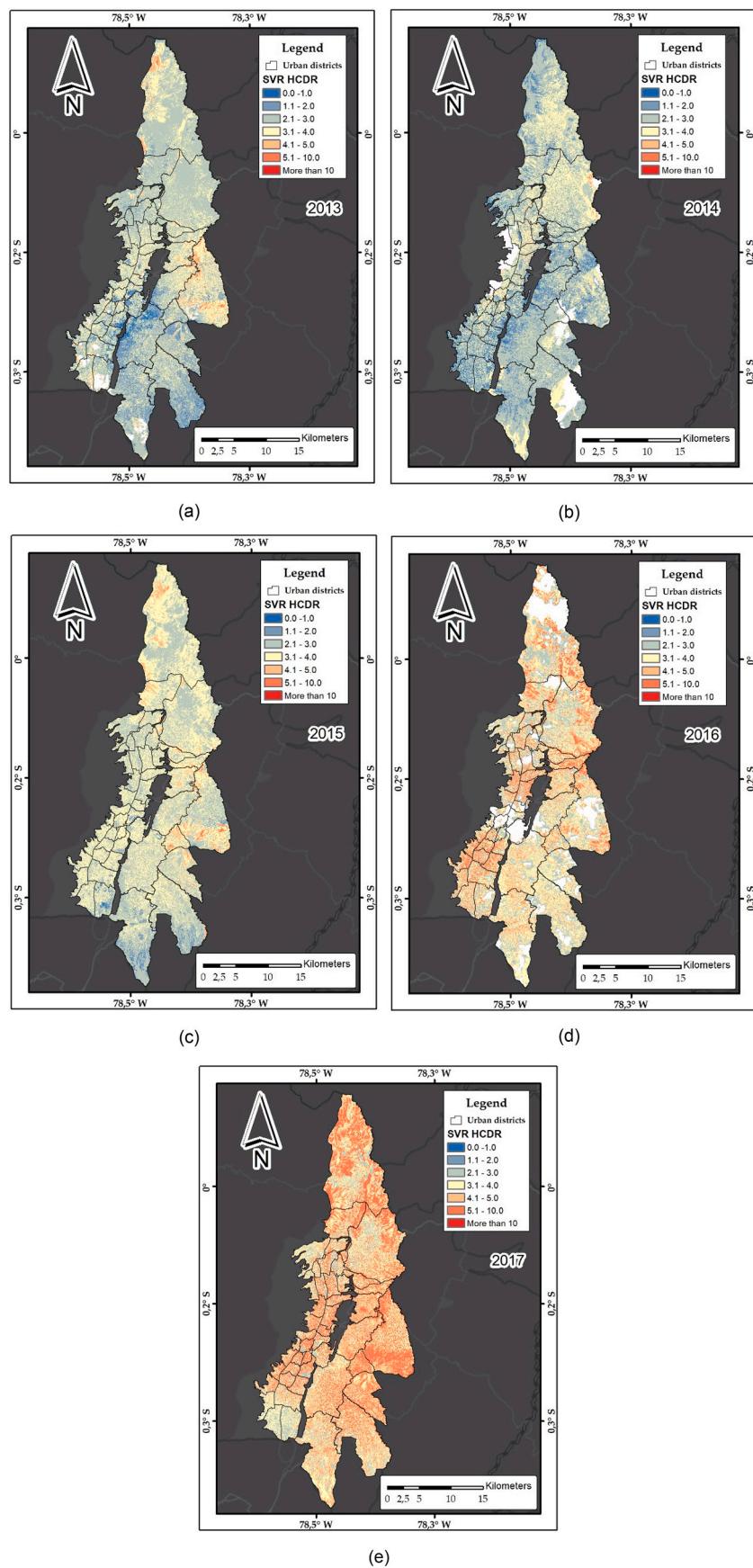


Fig. 6. HCRD maps considering the third trimester of the year (July–September) using the SVR model in (a) 2013, (b) 2014, (c) 2015, (d) 2016, and (e) 2017. The white areas show cloud presence.

Table 3

Relative importance of the input variables in the SVR model.

No.	Variable	Importance
1	Coastal aerosol band (B1)	0.085
2	Blue band (B2)	0.044
3	Short-wave infrared 2 (B7)	0.067
4	Enhanced vegetation index (EVI)	0.131
5	Land Surface temperature (LST)	0.047
6	Relative humidity (RH)	0.125
7	Air temperature (AT)	0.212
8	Solar irradiance (SR)	0.035
9	CO	0.204
10	SO ₂	0.049

accuracy public health planning, especially in countries where the quality of health data is questionable. Thereby, an alternative of modeling geographical variables is non-linear LUR models, adding more spatial predictors, as remote sensing data. Despite, MLP, SVR and RFR are black boxes in most of their different processes, they show better results than a traditional LUR linear model avoiding, for instance, situations of multicollinearity. On the other hand, remote sensing variables allow to get updated (temporal resolution) and high spatial resolution data to update the non-linear LUR models in real time. In the case of Ecuador, Quito is the unique city with an AQMN in all the country. The second largest city, Guayaquil with 3 million people, did not have this kind of information. Thereby, the application of these kind of models in cities without field data could provide a first approach of spatial distribution of air pollutants and its possible relationship with epidemiological data, being a feasible alternative to the governmental health institutions.

5. Conclusions

This study proposed an innovative, alternative use of LUR models to establish a spatial modeling approach to calculating the number of hospital discharges of patients with CRDs in Quito, Ecuador. The proposed model considered geographical predictors, specifically remote sensing data (Landsat 8) and environmental variables (air pollution and meteorological information) from 2013 to 2017. The most significant predictors were the red band, the short infra-red band (B7), CO, and SO₂. The first two are related with PM2.5 and O₃, while CO and SO₂ are related with air pollution. Different machine learning techniques were tested. RFR performed best considering the training dataset, and SVR performed best considering the test dataset. These models allowed us to generate spatial maps identifying areas with a high prevalence of chronic respiratory diseases, representing an effective approach to using remote sensing data in public health research. This work also provides more information about the spatial distribution of respiratory diseases, which can help in the identification and eradication of their possible causes. Moreover, the main objective of this approach is to improve traditional linear LUR models with more updated variables, as remote sensing data.

CRediT authorship contribution statement

Cesar I. Alvarez-Mendoza: Writing - original draft. **Ana Teodoro:** Supervision. **Alberto Freitas:** Formal analysis, Supervision. **Joao Fonseca:** Data curation, Formal analysis.

Acknowledgments

The study is part of a PhD thesis in surveying engineering at the University of Porto, Portugal, supported by the Salesian Polytechnic University, Ecuador. This work was supervised at the University of Porto by Prof. Ana Cláudia Teodoro.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.apgeog.2020.102273>.

References

- Ahangarcani, M., Farnaghi, M., Shirzadi, M. R., Pilesjö, P., & Mansourian, A. (2019). Predictive risk mapping of human leptospirosis using support vector machine classification and multilayer perceptron neural network. *Geospatial Health*, 14(1). <https://doi.org/10.4081/gh.2019.711>.
- Alvarez-Mendoza, C. I., Teodoro, A., & Ramirez-Cando, L. (2018). Improving NDVI by removing cirrus clouds with optical remote sensing data from Landsat-8 – a case study in Quito, Ecuador. *Remote Sensing Applications: Society and Environment*, 13, 257–274. <https://doi.org/10.1016/j.rsase.2018.11.008>, October 2018.
- Alvarez-Mendoza, C. I., Teodoro, A., & Ramirez-Cando, L. (2019). Spatial estimation of surface ozone concentrations in Quito Ecuador with remote sensing data, air pollution measurements and meteorological variables. *Environmental Monitoring and Assessment*, 191(3), 155. <https://doi.org/10.1007/s10661-019-7286-6>.
- Alvarez, C. I., Padilla Almeida, O., Álvarez Mendoza, C. I., & Padilla Almeida, O. (2016). Estimación de la contaminación del aire por PM10 en Quito a través de índices ambientales con imágenes LANDSAT ETM+. *Revista Cartográfica*, 92, 135–147. <http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso,url,uid&db=a9h&AN=122331600&lang=pt-br&site=eds-live&scope=site&authtype=sso>.
- Alvarez, C. I., Teodoro, A. C., Ordoñez, J., Benítez, A., Freitas, A., & Fonseca, J. (2019). Modeling the prevalence of respiratory chronic diseases risk using satellite images and environmental data. In N. Chrysoulakis, T. Erbertseder, Y. Zhang, & F. Baier (Eds.), *Remote sensing technologies and applications in urban environments IV* (Vol. 11157, p. 7). SPIE. <https://doi.org/10.1117/12.2532508>.
- Alvarez, C. I., Teodoro, A., & Tierra, A. (2017). Evaluation of automatic cloud removal method for high elevation areas in Landsat 8 OLI images to improve environmental indexes computation. *Proceedings of SPIE*. <https://doi.org/10.1117/12.2277844>.
- Alvarez-Mendoza, C. I., Teodoro, A. C., Torres, N., Vivanco, V., et al. (2019). Assessment of remote sensing data to model PM10 estimation in cities with a low number of air quality stations: A case of study in Quito, Ecuador. *Environments*, 6(7), 85. <https://doi.org/10.3390/ENVIRONMENTS6070085>, 6.
- Anderson, H. R., Butland, B. K., van Donkelaar, A., Brauer, M., Strachan, D. P., Clayton, T., et al. (2012). Satellite-based estimates of ambient air pollution and global variations in childhood asthma prevalence. *Environmental Health Perspectives*, 120(9), 1333–1339. <https://doi.org/10.1289/ehp.1104724>.
- Arundel, A. V., Sterling, E. M., Biggin, J. H., & Sterling, T. D. (1986). Indirect health effects of relative humidity in indoor environments. *Environmental Health Perspectives*, 65, 351. <https://doi.org/10.1289/EHP.8665351>.
- Ayehu, G., Tadesse, T., et al. Ayehu, G., Tadesse, T., Gessesse, B., & Yigrem, Y. (2019). Soil moisture monitoring using remote sensing data and a stepwise-cluster prediction model: The case of upper blue nile basin, Ethiopia. *Remote Sensing*, 11(2), 125. <https://doi.org/10.3390/rs11020125>.
- Ayres-Sampaio, D., Teodoro, A. C., Siller, N., Santos, C., Fonseca, J., & Freitas, A. (2014). An investigation of the environmental determinants of asthma hospitalizations: An applied spatial approach. *Applied Geography*, 47, 10–19. <https://doi.org/10.1016/j.apgeog.2013.11.011>.
- Azeez, O., Pradhan, B., Shafri, H., Shukla, N., Lee, C.-W., & Riziee, H. (2019). Modeling of CO emissions from traffic vehicles using artificial neural networks. *Applied Sciences*, 9(2), 313. <https://doi.org/10.3390/app9020313>.
- Barry, M., & Annese-Maesano, I. (2017). Ten principles for climate, environment and respiratory health. *European Respiratory Journal*, 50(6), 1701912. <https://doi.org/10.1183/13993003.01912-2017>.
- Bertazzon, S., Johnson, M., Eccles, K., & Kaplan, G. G. (2015). Accounting for spatial effects in land use regression for urban air pollution modeling. *Spatial and Spatio-Temporal Epidemiology*, 14–15, 9–21. <https://doi.org/10.1016/j.sste.2015.06.002>.
- Bilal, M., Nichol, J. E., Bleiweiss, M. P., & Dubois, D. (2013). A Simplified high resolution MODIS aerosol retrieval algorithm (SARA) for use over mixed surfaces. *Remote Sensing of Environment*, 136(September 2013), 135–145. <https://doi.org/10.1016/j.rse.2013.04.014>.
- Brokamp, C., Jandarov, R., Rao, M. B., LeMasters, G., & Ryan, P. (2017). Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. *Atmospheric Environment*, 151, 1–11. <https://doi.org/10.1016/j.atmosenv.2016.11.066>.
- Chan, C.-C., Chuang, K.-J., Chen, W.-J., Chang, W.-T., Lee, C.-T., & Peng, C.-M. (2008). Increasing cardiopulmonary emergency visits by long-range transported Asian dust storms in Taiwan. *Environmental Research*, 106(3), 393–400. <https://doi.org/10.1016/j.envres.2007.09.006>.
- Chang, H.-T., et al. Chang, H.-T., Wu, C.-D., Pan, W.-C., Lung, S.-C. C., & Su, H.-J. (2019). Association between surrounding greenness and schizophrenia: A Taiwanese cohort study. *International Journal of Environmental Research and Public Health*, 16(8), 1415. <https://doi.org/10.3390/ijerph16081415>.
- Chejarla, V. R., Maheshuni, P. K., & Mandla, V. R. (2016). Quantification of LST and CO₂ levels using Landsat-8 thermal bands on urban environment. *Geocarto International*, 31(8). <https://doi.org/10.1080/10106049.2015.1094522>.
- Chen, S., Goo, Y.-J. J., & Shen, Z.-D. (2014). A hybrid approach of stepwise regression, logistic regression, support vector machine, and decision tree for forecasting fraudulent financial statements. *TheScientificWorldJOURNAL*, 968712. <https://doi.org/10.1155/2014/968712>, 2014.

- Chen, Y., Han, W., Chen, S., & Tong, L. (2014). Estimating ground-level PM2.5 concentration using Landsat 8 in Chengdu, China. *Proceedings of SPIE*, 9259. <https://doi.org/10.1117/12.2068886> (February), 925917–925931.
- Chen, G., & Meentemeyer, R. (2016). Remote sensing of forest damage by diseases and insects. In Q. Weng (Ed.), *Remote sensing for sustainability* (2017th ed., p. 357). CRC Press https://books.google.pt/books?id=7da_DQAAQBAJ.
- Chuvieco, E. (2016). *Fundamentals of satellite remote sensing: An environmental approach* (2nd ed.). CRC Press <https://books.google.pt/books?id=-nCmWAAQBAJ>.
- Department of the Interior U.S. Geological Survey. (2016). In *Landsat 8* (Vol. 8). L8 Data users handbook.
- Di, Q., Wang, Y. Y., Zanobetti, A., Wang, Y. Y., Koutrakis, P., Choirat, C., et al. (2017). Air pollution and mortality in the medicare population. *New England Journal of Medicine*, 376(26), 2513–2522. <https://doi.org/10.1056/NEJMoa1702747>.
- D'Amato, G., Cecchi, L., D'Amato, M., & Annese-Maesano, I. (2014). Climate change and respiratory diseases. *European Respiratory Review: An Official Journal of the European Respiratory Society*, 23(132), 161–169. <https://doi.org/10.1183/09059180.00001714>.
- Epa. (2017). Inventory of U.S. Greenhouse gas emissions and sinks: 1990–2017. <https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks>.
- Estallo, E. L., Benitez, E. M., Lanfri, M. A., Scavuzzo, C. M., & Almirón, W. R. (2016). MODIS environmental data to assess chikungunya, dengue, and zika diseases through Aedes (stegomyia) aegypti oviposition activity estimation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(12), 5461–5466. <https://doi.org/10.1109/JSTARS.2016.2604577>.
- Famoso, F., Wilson, J., Monforte, P., Lanzafame, R., Brusca, S., & Lulla, V. (2017). Measurement and modeling of ground-level ozone concentration in catania, Italy using biophysical remote sensing and GIS. In *International journal of applied engineering research* (Vol. 12). <http://www.ripublication.com>.
- Fan, J., Li, S., Fan, C., Bai, Z., & Yang, K. (2016). The impact of PM2.5 on asthma emergency department visits: A systematic review and meta-analysis. *Environmental Science and Pollution Research*, 23(1), 843–850. <https://doi.org/10.1007/s11356-015-5321-x>.
- Ghaleb, F., Mario, M., & Sandra, A. (2015). Regional landsat-based drought monitoring from 1982 to 2014. *Climate*, 3(3), 563–577. <https://doi.org/10.3390/cli3030563>.
- Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S., Tiede, D., Aryal, J., et al. (2019). Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sensing*, 11(2), 196. <https://doi.org/10.3390/rs11020196>.
- Gupta, P., Christopher, S. A., Wang, J., Gehrig, R., Lee, Y., & Kumar, N. (2006). Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmospheric Environment*, 40(30), 5880–5892. <https://doi.org/10.1016/j.atmosenv.2006.03.016>.
- Habermann, M., Billger, M., & Haeger-Eugensson, M. (2015a). Land use regression as method to model air pollution. Previous results for gothenburg/Sweden. *Procedia Engineering*, 115, 21–28. <https://doi.org/10.1016/j.proeng.2015.07.350>.
- Habermann, M., Billger, M., & Haeger-Eugensson, M. (2015b). Land use regression as method to model air pollution. Previous results for gothenburg/Sweden. *Procedia Engineering*, 115, 21–28. <https://doi.org/10.1016/j.proeng.2015.07.350>.
- Hadjimitsis, D. G. (2009). Aerosol optical thickness (AOT) retrieval over land using satellite image-based algorithm. *Air Quality, Atmosphere and Health*, 2(2), 89–97. <https://doi.org/10.1007/s11869-009-0036-0>.
- Hastie, T., et al. (2009). Springer series in statistics the elements of statistical learning. *The Mathematical Intelligencer*, 27(2), 83–85. <https://doi.org/10.1007/b94608>.
- Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of statistical learning: Data mining, inference and prediction. *The Mathematical Intelligencer*, 27 (2), 83–85.
- Health at a Glance 2017. (2017). Oecd. https://doi.org/10.1787/health_glance-2017-en.
- Heo, S., & Bell, M. L. (2019). The influence of green space on the short-term effects of particulate matter on hospitalization in the U.S. for 2000–2013. *Environmental Research*, 174, 61–68. <https://doi.org/10.1016/j.envres.2019.04.019>.
- Huete, A. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X).
- Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., & Ferreira, L. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1–2), 195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2).
- Humans, I. W. G., of to, E. C. R., & Cancer for R.on, I. A. (2016). *Outdoor air pollution*. International Agency for Research on Cancer, World Health Organization. <https://books.google.com.ec/books?id=bioFvgAACAAJ>.
- Ibrahim Sameen, M., Kubaisy, M. A. Al, Nahhas, F. H., Ali, A. A., Othman, N., & Hason, M. M. (2014). Sulfur dioxide (SO₂) monitoring over kirkuk city using remote sensing data. *Journal of Civil & Environmental Engineering*, 4(5), 1–6. <https://doi.org/10.4172/2165-784X.1000155>.
- Jin, J., Wang, Q., Jin, J., & Wang, Q. (2019). Evaluation of informative bands used in different PLS regressions for estimating leaf biochemical contents from hyperspectral reflectance. *Remote Sensing*, 11(2), 197. <https://doi.org/10.3390/rs11020197>.
- Jumaah, H. J., Ameen, M. H., Kalantar, B., Rizeei, H. M., & Jumaah, S. J. (2019). Air quality index prediction using IDW geostatistical technique and OLS-based GIS technique in Kuala Lumpur, Malaysia. *Geomatics, Natural Hazards and Risk*, 10(1), 2185–2199. <https://doi.org/10.1080/19475705.2019.1683084>.
- Kallawicha, K., Chuang, Y., Lung, S. C., Wu, C., Han, B., Ting, Y., et al. (2018). Outpatient visits for allergic diseases are associated with exposure to ambient fungal spores in the greater taipei area (pp. 2077–2085). <https://doi.org/10.4209/aaqr.2018.01.0028>.
- Khreis, H., Nieuwenhuijsen, M., Khreis, H., & Nieuwenhuijsen, M. J. (2017). Traffic-related air pollution and childhood asthma: Recent advances and remaining gaps in the exposure assessment methods. *International Journal of Environmental Research and Public Health*, 14(3), 312. <https://doi.org/10.3390/ijerph14030312>.
- Kutlar Joss, M., Eeftens, M., Gintowt, E., Kappeler, R., & Künzli, N. (2017). Time to harmonize national ambient air quality standards. *International Journal of Public Health*, 62(4), 453–462. <https://doi.org/10.1007/s00038-017-0952-y>.
- Lee, J. H., Ryu, J. E., Chung, H. I., Choi, Y. Y., Jeon, S. W., & Kim, S. H. (2018). Development of spatial scaling technique of forest health sample point information. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(3), 751–756. <https://doi.org/10.5194/isprs-archives-XLII-3-751-2018>.
- Liu, Y., Franklin, M., Kahn, R., & Koutrakis, P. (2007). Using aerosol optical thickness to predict ground-level PM2.5 concentrations in the st. Louis area: A comparison between misr and MODIS. *Remote Sensing of Environment*, 107(1–2), 33–44. <https://doi.org/10.1016/j.rse.2006.05.022>.
- Li, L., Zhou, X., & Tong, W. (2019). *Spatiotemporal analysis of air pollution and its application in public health*. Elsevier Science. <https://books.google.com.ec/books?id=vCa-DwAAQBAJ>.
- Lyapustin, A. I., Wang, Y., Laszlo, I., Hilker, T., Hall, G. F., Sellers, P. J., et al. (2012). Multi-angle implementation of atmospheric correction for MODIS (MAIAC): 3. Atmospheric correction. *Remote Sensing of Environment*, 127, 385–393. <https://doi.org/10.1016/j.rse.2012.09.002>.
- Mather, P., & Tso, B. (2003). *Classification methods for remotely sensed data*. CRC Press. <https://books.google.pt/books?id=u0l5HSj0pKEC>.
- Olive, D. J. (2017). *Linear regression*. Springer International Publishing. <https://books.google.pt/books?id=d9WyDgAAQBAJ>.
- O'Connor, G. T., Neas, L., Vaughn, B., Kattan, M., Mitchell, H., Crain, E. F., et al. (2008). Acute respiratory health effects of air pollution on children with asthma in US inner cities. *The Journal of Allergy and Clinical Immunology*, 121(5), 1133–1139. <https://doi.org/10.1016/j.jaci.2008.02.020>.
- Rivera-González, L. O., Zhang, Z., Sánchez, B. N., Zhang, K., Brown, D. G., Rojas-Bracho, L., et al. (2015). An assessment of air pollutant exposure methods in Mexico City, Mexico. *Journal of the Air and Waste Management Association*, 65(5), 581–591. <https://doi.org/10.1080/10962247.2015.1020974>.
- Saucy, A., Röösli, M., Künzli, N., Tsai, M. Y., Sieber, C., Olaniyan, T., et al. (2018). Land use regression modelling of outdoor NO₂ and PM2.5 concentrations in three low income areas in the western cape province, South Africa. *International Journal of Environmental Research and Public Health*, 15(7). <https://doi.org/10.3390/ijerph15071452>.
- Secretaria del Ambiente de Quito. (2018a). IAMQ/18. <https://doi.org/10.1590/s1809-98232013000400007>.
- Secretaria del Ambiente del Ecuador. (2018b). Red Metropolitana de Monitoreo Atmosférico de Quito. <http://www.quitoambiente.gob.ec/ambiente/index.php/generalidades>.
- Sheela, K. G., & Deepa, S. N. (2013). Review on methods to fix number of hidden neurons in neural networks. *Mathematical Problems in Engineering*, 1–11. <https://doi.org/10.1155/2013/425740>, 2013.
- Sicard, P., Navar, A., De Marco, A., & Paoletti, E. (2017). Projected global ground-level ozone impacts on vegetation under different emission and climate scenarios. *Atmospheric Chemistry and Physics*, 17, 12177–12196. <https://doi.org/10.5194/acp-17-12177-2017>.
- Sobrino, J. A., Jiménez-Muñoz, J. C., & Paolini, L. (2004). Land surface temperature retrieval from LANDSAT TM 5. *Remote Sensing of Environment*, 90(4), 434–440. <https://doi.org/10.1016/j.rse.2004.02.003>.
- Sobrino, J. A., Jiménez-Muñoz, J. C., Soria, G., Romaguera, M., Guanter, L., Moreno, J., et al. (2008). Land surface emissivity retrieval from different VNIR and TIR sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 46(2), 316–327. <https://doi.org/10.1109/TGRS.2007.904834>.
- Sorek-Hamer, M., Just, A. C., & Kloog, I. (2016). Satellite remote sensing in epidemiological studies. *Current Opinion in Pediatrics*, 28(2), 228–234. <https://doi.org/10.1097/MOP.0000000000000326>.
- Sweileh, W. M., Al-Jabi, S. W., Zyoud, S. H., & Sawalha, A. F. (2018). Outdoor air pollution and respiratory health: A bibliometric analysis of publications in peer-reviewed journals (1900 – 2017). *Multidisciplinary Respiratory Medicine*, 13(1), 1–12. <https://doi.org/10.1186/s40248-018-0128-5>.
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0).
- Usgs. (2019). Landsat 8 surface reflectance code (LaSRC) product guide. https://prd-wret.s3-us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1368_L8_Surface_Reflectance_Code_LaSRC_Product_Guide-v2.0.pdf.
- Vapnik, V. N. (1998). *Statistical learning theory*. Wiley. <https://books.google.pt/books?id=GwoAQAAQAAJ>.
- Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, 185, 46–56. <https://doi.org/10.1016/j.rse.2016.04.008>.
- Viana, J., Vasco Santos, J., Manuel Neiva, R., Souza, J., Duarte, L., Cláudia Teodoro, A., et al. (2017). Remote sensing in human health: A 10-year bibliometric analysis. *Remote Sensing*. <https://doi.org/10.3390/rs9121225>.
- Vieira, D., Teodoro, A., & Gomes, A. (2016). Analysing land surface temperature variations during fogo island (Cape Verde) 2014–2015 eruption with Landsat 8 images. *Proceedings of SPIE*, 10005(October), 1000508. <https://doi.org/10.1117/12.2241382>.
- WHO. (2016). *ICD-10 version:2016*. WHO. <https://icd.who.int/browse10/2016/en>.
- WHO. (2019). *Chronic respiratory diseases (CRDs)*. WHO; World Health Organization. <https://www.who.int/respiratory/en/>.

- Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., et al. (2019). Current status of Landsat program, science, and applications. *Remote Sensing of Environment*, 225, 127–147. <https://doi.org/10.1016/J.RSE.2019.02.015>.
- Zhai, L., Sang, H., Zhang, J., & An, F. (2015). Estimating the spatial distribution of PM_{2.5} concentration by integrating geographic data and field measurements. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. <https://doi.org/10.5194/isprarchives-XL-7-W4-209-2015>. XL-7/W4(July), 209–213.
- Zhang, Z. (2016). Variable selection with stepwise and best subset approaches. *Annals of Translational Medicine*, 4(7), 136. <https://doi.org/10.21037/atm.2016.03.35>.
- Zheng, C., Zhao, C., Li, Y., Wu, X., Zhang, K., Gao, J., et al. (2018). Spatial and temporal distribution of NO₂ and SO₂ in Inner Mongolia urban agglomeration obtained from satellite remote sensing and ground observations. *Atmospheric Environment*, 188, 50–59. <https://doi.org/10.1016/J.ATMOSENV.2018.06.029>.
- Zheng, S., Zhou, X., Singh, R., Wu, Y., Ye, Y., & Wu, C. (2017). The spatiotemporal distribution of air pollutants and their relationship with land-use patterns in hangzhou city, China. *Atmosphere*, 8(6), 110. <https://doi.org/10.3390/atmos8060110>.
- Zou, B., et al. Zou, B., Chen, J., Zhai, L., Fang, X., & Zheng, Z. (2016). Satellite based mapping of ground PM_{2.5} concentration using generalized additive modeling. *Remote Sensing*, 9(1), 1. <https://doi.org/10.3390/rs9010001>.