Spatial Knowledge and Information Canada, 2019, 7(2), 3

OLS and GWR LUR models of wildfire smoke using remote sensing and spatiotemporal data in Alberta

MOJGAN MIRZAEI
Department of Geography,
University of Calgary
mojgan.mirzaei@ucalgary.ca

STEFANIA BERTAZZON
Department of Geography,
University of Calgary;
Department SAGAS,
University of Florence
bertazzs@ucalgary.ca

ISABELLE COULOIGNER
Department of Geography,
University of Calgary
icouloig@ucalgary.ca

ABSTRACT

Wildfire smoke from forest fire is a major source of air pollution in Canadian cities. Wildfire smoke includes different types of gases and particles that adversely affect human health. Particulate Matter (PM), as the predominant pollutant in wildfire smoke, poses the greatest risk to human health. Accurate investigation of wildfire-related PM2.5 is critical to understand health-related effects. This research investigated PM2.5 concentration of wildfire smoke drifting over parts of Alberta in August 2017 from British Columbia, Montana, Idaho, and as far away as Washington State. We developed OLS and GWR land use regression models, which integrate the use of MODIS Aerosol Optical Depth data and temporal indicators to model PM2.5 concentration. The results provide estimates of PM2.5 at finer spatial resolution than ground-based records; these estimates could aid epidemiological studies to assess the health effects of wildfire smoke.

1. Introduction

Intensity and frequency of wildfire events have increased in recent decades and is likely to be aggravated by climate change. In the past decade, wildfires have come to the attention of public health and ecosystem studies (Youssouf et al. 2014). Wildfire PM and gaseous products can lead to acute and long term health impacts on exposed populations. Among these pollutants, fine particles are the most harmful (WHO 2000)

During the summer of 2017, Alberta experienced a severe smoke episode associated with different wildfires. In mid-August, smoke from wildfires in British Columbia, Montana, Idaho and Washington State has drifted over parts of Alberta, making the air quality (AQ) so poor that it made the headlines (e.g. CBC 2017).

AQ ground stations provide the most accurate data on PM2.5 concentration near the ground. However, due to their high operational cost, they have sparse distribution and limited spatial coverage, especially in remote rural area.

Land Use Regression (LUR) models and satellite observation based models can address these limitations. A variety of studies have used LUR and remote sensing based models to estimate PM2.5 concentration (Van Donkelaar et al. 2006; Liu et al. 2007; Van Donkelaar et al. 2015; Li et al. 2016); however, few studies reported smoke-based PM2.5 estimation during fire periods (Mirzaei et al. 2018; Hodzic et al. 2007).

Spatial data tend to exhibit spatial non-stationarity, defined as inconstant spatial variability (Anselin 1988). This spatial property can lead to spatial instability of regression coefficients (Fotheringham et al. 1998). Spatial non-stationarity can be addressed by geographically weighted regression (GWR) (Fotheringham et al. 1998).

The present study aimed to assess the performance of local GWR LUR to estimate PM2.5 concentration associated with wildfire in Alberta in August 2017 compared to the linear method.

2. Methods and Data

2.1 Study area and ground-based PM2.5 measurements

The study area (Figure 1) includes all Alberta Airshed Zones (AAZ). Twenty-four-hour $PM_{2.5}$ concentration were collected at 49 continuous AQ stations located in AAZ[1] over an extended period (Aug 7 to 22), centered on the fire event (August 13-18) and including 6 days before and 4 days after the event. Daily $PM_{2.5}$ concentrations (Figure 2) were averaged for the fire event period in each station.

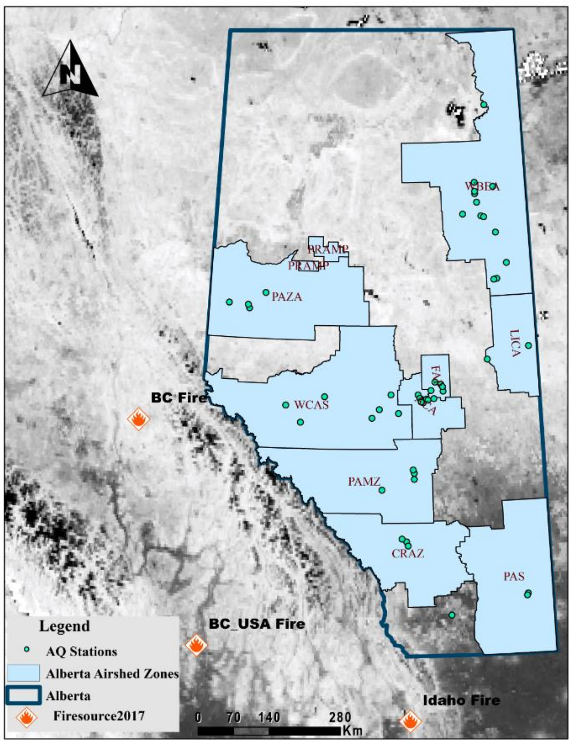

Figure 1. Distribution of the 49 Air Quality monitoring stations within the study area.

2.2 Predictors

The LUR model relied on the following predictor categories.

2.2.1 Temporal and spatial predictors

Temporal variables were wind speed, temperature, and humidity collected hourly at each monitoring station and averaged over the study period.

Spatial variables were industrial PM2.5 emission sources, road length, vegetation index, elevation, and distance from sources of fires. Industrial sources and road length were calculated over circular buffers of 5 and 10 km for industrial and 1 km for the road around each station. The Alberta road network was acquired from the National Road Network (NRN 2015), and industrial emission sources from the National Pollutant Release Inventory (NPRI 2016).

Normalized Difference Vegetation Index (NDVI) images (MOD13C2) were used as an indicator of vegetation cover. The 3 km spatial resolution (SR) images for August 2017 were collected from the NASA Giovanni website (Acker & Leptoukh 2007).

Elevation data were acquired from DMTI Spatial (DMTI 2010).

As the 2017 wildfire originated in different locations, three points were considered as sources of fire: one in BC, one in southwestern Alberta on the Canada-USA border, and one in Idaho (USA) (Figure 1). The Euclidean distance between each AQ station in the study area and each source of the fire was calculated and included in the model as a predictor.

The predictor variables pertaining to each AQ station location are presented in Table 1.

2.3 AOD images

Daily AOD images at 10x10 km SR, derived from MODIS terra collection data (NEO 2017) for the period of interest, were used as the AOD predictor. Further, averaged AOD product of MODIS at 1 degree, about 100 km, SR were collected from the NASA Giovanni website (Acker & Leptoukh 2007). They were used to fill some of the gaps in the 10x10 AOD images: 5x5 mean filter was applied, wherever possible, to calculate the missing values of the finer resolution images from their surrounding pixels; in areas where no surrounding pixels existed, the coarser resolution images were used to

---

[1]http://airdata.alberta.ca/RelatedLinks.aspx

simply fill gaps of the finer resolution image with its values.

2.4 Prediction Models

Traditional LUR models are described by standard regression equations (Eq.1), where the response variable y_i, *i.e.* observed PM2.5 concentration at location i is expressed as a function of k land use predictors, i.e., x_{i1} through x_{ik}, such as those detailed in Table 1. The β_0 through β_k coefficients are estimated using ordinary least squares (OLS).

$$y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i \quad (1)$$

Since global Moran's I spatial statistical test (Florax et al. 2003; Getis and Aldstadt 2004) of the PM2.5 concentration (Table 1) indicated that there was significant spatial autocorrelation, showing a likelihood of a clustered pattern, GWR was applied.

GWR applies a spatial weighting function on the spatial coordinates of each data point, *i.e.* (u_i, v_i), to subdivide the study area into local neighbourhoods, where local regressions are calculated (Eq. 2). Consequently, GWR produces n local regressions, each of them linear, and each one over a neighbourhood defined by the kernel function. A fixed bandwidth with a Gaussian kernel was selected. The bandwidth was determined automatically by minimizing a leave-one-out cross-validation (CV) score (Fortheringham et al. 2002).

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) \times x_{ik} + \varepsilon_i \quad (2)$$

Forward stepwise multiple linear regression (SMLR) was employed as a variable selection procedure to identify the significant predictors in the regression model.

LUR models were calculated in R (R Core Team 2018) using mainly the 'spdep' (Bivand & Piras 2015; Bivand et al. 2013), 'GWmodel' (Gollini et al. 2013), 'car' (Fox & Weisberg 2011), and 'lmtest' packages (Zeileis & Hothorn 2002).

3. Results and Discussion

Figure 2 shows the daily variability of $PM_{2.5}$ concentration recorded at the 49 AQ stations. The figure shows that PM2.5 concentration is under background level (20 µg/m^3) almost for all stations before and after the smoke event. It can also be seen that the daily averaged $PM_{2.5}$ concentration raised dramatically on August 13 and remained elevated until August 18.

Descriptive statistics of $PM_{2.5}$ concentration over the study period are presented in Table 1.

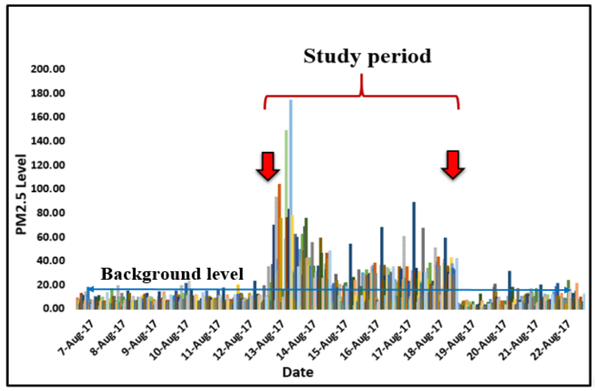

Figure 2. Daily variability of PM2.5 concentration at the 49 stations located in the study area.

Table 1. Descriptive statistics of response (PM2.5) and predictor variables

Response Variable	***Min***	***Max***	***Moran'I***	***P(I)***
PM2.5	9.85	57.9	0.6	0.00
Predictors	**Name/ Description**	**Unit/ Resolution**	**Range**	
AOD	Merged Aerosol Optical Depth	10x10 km	0.57	
NDVI	Vegetation Index	3x3 km	0.59	
TEMP	Temperature	Celcius	20.31	
RH	Relative humidity	Percentage	75.30	
WSP	Wind speed	km/hr at 10 m height	17.62	
Ind_5km/ 10km	Industrial points around each station within 5 and 10 km	Points in buffer	29 / 41	
Road 1km	Road length around each station	Length in buffer (meter)	46,711	
ELV	Elevation	meter	1171.28	
BC-Dis	Distance from fire in BC	kilometre	628,725	
Idaho-Dis	Distance from fire in Idaho	kilometre	997,808	
BC-US-dis	Distance from fire in BC_USA border	kilometre	749,683	

The variables identified by SMLR included AOD, wind speed, temperature, elevation, and BC-distance. AOD, followed by wind speed and temperature, were the three most significant variables in both OLS and GWR models.

Table 2 and Table 3 present the statistical results of OLS and GWR models respectively. The OLS LUR yielded a relatively high goodness-of-fit, with R^2 of 0.74 and adjusted R^2 of 0.71. However, the model performance was improved substantially by the use of GWR, with higher R^2, and lower AIC and RSS values compared to the OLS model.

Table 2: OLS results

	Coef	***Std.Error***	***P-Value***
Intercept	34.4	8.52	0.00
AOD	14.9	7.33	0.04
WSP	-1.23	0.28	0.00
TEMP	0.67	0.23	0.00
ELV	9.8E-03	5.3E-03	0.07
BC-Dis	-3.8E-05	7.1E-06	0.00
R^2	0.74	***Adj.R^2***	0.71
AIC	313	***RSS***	1306
LMerr Res	0.005	***Res Moran's I***	-0.006

Roads and industries are known as two important sources of PM2.5 in cities; however, these two variables were not significant in these models and were removed on the SLRM variable selection procedure. This result indicates that the presence/absence of wildfire smoke affects the model's predictors, as meteorological variables dominate the model, extruding those variables normally associated with PM. Similar results were obtained by our recent study of LUR models before, during, and after wildfire events (Mirzaei et al. 2018).

Table 3: GWR results

	Coef median	***Coef range***	***t-Value mean***	***t-Value range***
Intercept	37.4	25.5	3.98	2.2
AOD	16.4	42.6	1.72	4.79
WSP	-1.14	1.28	-3.6	0.05
TEMP	0.67	0.48	2.8	2.91
ELV	9.0E-03	8.6E-03	1.63	1.17
BC-Dis	-4.4E-05	4.8E-05	-5.5	3.78
R^2	0.84	***Adj.R^2***		0.77
AIC	288	***RSS***		830
LMerr Res	1.92	***Res Moran's I***		-0.11

Observed versus GWR predicted PM2.5 concentration, as well as OLS and GWR residuals are shown in Fig. 3.

The observed PM2.5 concentration is higher in the western parts of Alberta mainly due to the longer distance to the fire(s) of interest. The GWR fitted concentration follows this pattern through its association with the selection of distance to BC wildfire among all three sources of fire.

It can be seen in the residuals maps that not only did the GWR model performed better than the OLS model, but also that this difference is greater for lower PM2.5 concentration (shown in grey), relative to higher concentration. Over- and under-estimates do not present any spatial pattern but demonstrate that more work needs to be done for a more accurate model.

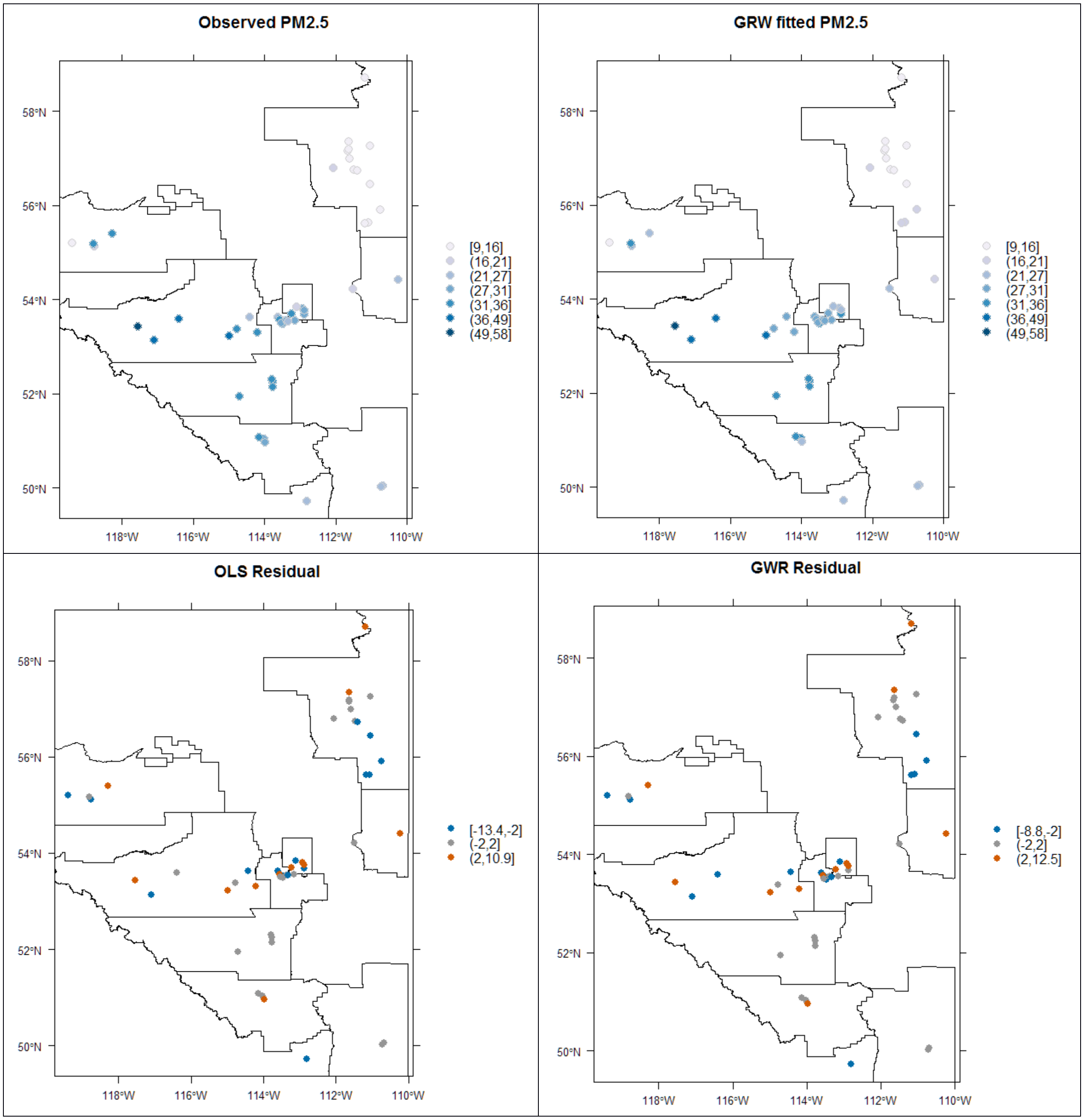

Figure 3 Observed and GWR predicted PM2.5 concentration, as well as OLS LUR and GWR LUR residuals (orange corresponds to underfitted values and blue to overfitted ones)

4. Conclusion

Due to wildfire events from BC and the USA, some parts of Alberta experienced a high level of smoke and PM2.5 concentration in August 2017.

In the present study, we modelled the spatial distribution of PM2.5 concentration due to wildfire smoke using OLS and GWR land use regression that integrated MODIS AOD, meteorological data, and spatial variables. The OLS results indicated that the global model performed relatively well; however, the GWR LUR has achieved a better model performance, as shown by higher R^2 and lower AIC and RSS.

Overall, we have demonstrated the potential of integrating satellite AOD data with spatial and temporal variables to accurately predict PM2.5 concentration during wildfire smoke events. Building on these promising results, our models can be further improved by using more spatiotemporal variables and better methodology to fill AOD images' gaps, so that we can develop daily models of the PM2.5 plume.

Acknowledgment

Mojgan Mirzaei wishes to thank "Eyes High Doctoral Recruitment Scholarship" for supporting her doctoral work. Stefania Bertazzon wishes to thank the Canadian Institutes for Health Research (CIHR) Institute for Population and Public Health for funding the research on air pollution and public health. We are grateful to our colleagues and members of the Geography of Health research group of the O'Brien Institute for Population Health for their advice and insightful discussions.

References

Acker, J.G. & Leptoukh, G., 2007. Online Analysis Enhances Use of NASA Earth Science Data. *Eos, Transactions American Geophysical Union*, 88(2), p.14. Available at: http://doi.wiley.com/10.1029/2007EO020003.

Anselin, L., 1988. *Spatial econometrics: methods and models* Springer., Kluwer, New York. Available at: https://www.springer.com/gp/book/9789024737352.

Bivand R. et al., 2013. *Applied Spatial Data Analysis with R* Springer, ed., New York.

Bivand, R. & Piras, G., 2015. Comparing Implementations of Estimation Methods for Spatial Econometrics. *Journal of Statistical Software*, 63(18). Available at: http://www.jstatsoft.org/v63/i18/.

CBC, 2017. "High risk" air quality warning issued for Calgary as B.C. wildfire smoke returns. *CBC News*. Available at: https://www.cbc.ca/news/canada/calgary/calgary-air-quality-index-health-aqhi-bc-forest-fires-alberta-smoky-skies-august-31-1.4269864.

DMTI, 2010. The Gold Standard Canada's Most Complete and Accurate Mapping Data. *DMTI Spatial*. Available at: https://www.dmtispatial.com/ [Accessed August 22, 2018].

Fotheringham, A.S., Charlton, M.E. & Brunsdon, C., 1998. Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis. *Environment and Planning A: Economy and Space*, 30(11), pp.1905–1927.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. *Chichester*: *Wiley*

Fox, J. & Weisberg, S., 2011. An R Companion to Applied Regression, Second Edition, Sage Publications. Available at: https://socialsciences.mcmaster.ca/jfox/Books/Companion-2E/index.html.

Florax RJGM, Folmer H, Rey SJ, 2003. Specification searches in spatial econometrics: the relevance of Hendry's methodology. *Reg Sci Urban Econ*, 33(5), 557–79.

Getis A. and Aldstadt J., 2004. Constructing the spatial weights matrix using a local statistic. *Geogr Anal*, 40, 297–309

Gollini, I. et al., 2013. GWmodel: an R Package for Exploring Spatial Heterogeneity using Geographically Weighted Models. Available at: http://arxiv.org/abs/1306.0413.

Hodzic, A. et al., 2007. Wildfire particulate matter in Europe during summer 2003: Meso-scale modeling of smoke emissions, transport and radiative effects. *Atmospheric Chemistry and Physics*, 7(15), pp.4043–4064.

Li, S., Joseph, E. & Min, Q., 2016. Remote sensing of ground-level PM2.5 combining AOD and backscattering profile. *Remote Sensing of Environment*, 183, pp.120–128. Available at: http://dx.doi.org/10.1016/j.rse.2016.05.025.

Liu, Y. et al., 2007. Using aerosol optical thickness to predict ground-level PM2.5concentrations in the St. Louis area: A comparison between MISR and MODIS.

Remote Sensing of Environment, 107(1–2), pp.33–44.

Mirzaei, M., Bertazzon, S. & Couloigner, I., 2018. Modeling Wildfire Smoke Pollution by Integrating Land Use Regression and Remote Sensing Data: Regional Multi-Temporal Estimates for Public Health and Exposure Models. *Atmosphere*, 9(9), p.335. Available at: http://www.mdpi.com/2073-4433/9/9/335.

NEO, 2017. NEO, Nasa Earth Observations. *NASA*. Available at: https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MODAL2_D_AER_OD&date=2018-08-01.

NPRI, 2016. National Pollutant Release Inventory. *Environment and Climate Change Canada*. Available at: http://www.ec.gc.ca/inrp-npri/default.asp?lang%BCEn&;n%BC4A577BB9-1 [Accessed August 22, 2018].

NRN, 2015. National Road Network - NRN. *Natural Resources Canada*. Available at: https://open.canada.ca/data/en/dataset/3d282116-e556-400c-9306-ca1a3cada77f.

Van Donkelaar, A. et al., 2015. High-Resolution Satellite-Derived PM 2.5 from Optimal Estimation and Geographically Weighted Regression over North America. *Environmental Science & Technology*, 49(17), pp.10482–10491. Available at: http://pubs.acs.org/doi/10.1021/acs.est.5b02076.

Van Donkelaar, A., Martin, R. V. & Park, R.J., 2006. Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing. *Journal of Geophysical Research: Atmospheres*, 111(21), pp.1–10.

WHO, 2000. Air quality guidelines for Europe. *Environmental Science and Pollution Research*, 3(1), pp.23–23. Available at: http://link.springer.com/10.1007/BF02986808%5Cnhttp://www.springerlink.com/index/10.1007/BF02986808.

Youssouf, H. et al., 2014. Quantifying wildfires exposure for investigating health-related effects. *Atmospheric Environment*, 97, pp.239–251. Available at: http://dx.doi.org/10.1016/j.atmosenv.2014.07.041.

Zeileis, A. & Hothorn, T., 2002. Diagnostic Checking in Regression Relationships. *R News*, 2(3), pp.7–10. Available at: https://cran.r-project.org/doc/Rnews/.