



ARTICLE



<https://doi.org/10.1057/s41599-021-00994-5>

OPEN

Correlation of publication frequency of newspaper articles with environment and public health issues in fire-prone peatland regions of Riau in Sumatra, Indonesia

Masafumi Ohashi¹, Akihiro Kameda², Osamu Kozan^{3,4}, Masahiro Kawasaki^{4✉}, Windy Iriana^{5,9}, Kenichi Tonokura⁵, Daisuke Naito^{6,7} & Kayo Ueda⁸

Forest fires in peatlands emit pollutants to the atmosphere, affecting public health. Though air quality data and epidemiology information are helpful in the management of the environment and public health, they are not always available. We aimed to reveal the utility of newspaper articles for estimating the public health concerns posed by air pollution. Using the database of a local newspaper, *Riau Pos*, in Riau Province of Sumatra, Indonesia, we have studied spatiotemporal correlations between publication frequency of newspaper articles mentioning search terms relevant to health issues and Fire Radiation Power (FRP) of peatland fires. FRP from one of the NASA satellite databases has been used as an index of air pollution caused by peatland fires. Visibility data for airport operation were also used as an index of particulate matter concentrations. The study regions are primarily the fire-prone Riau Province and nearby areas in the central area of Sumatra, Indonesia, from 2009 to 2018. Newspaper articles related to public health are found to be associated with FRP and visibility, or regional air pollution.

¹ Graduate School of Science and Engineering, Kagoshima University, Kagoshima, Japan. ² National Museum of Ethnology, Sakura, Japan. ³ Center for Southeast Asian Studies, Kyoto University, Kyoto, Japan. ⁴ Research Institute for Humanity and Nature, Kyoto, Japan. ⁵ Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan. ⁶ Graduate School of Agriculture, Kyoto University, Kyoto, Japan. ⁷ Center for International Forestry Research, Sidang Barang, Bogor, Indonesia. ⁸ Graduate School of Global Environmental Studies, Kyoto University, Kyoto, Japan. ⁹ Present address: Center for Environmental Studies, Bandung Institute of Technology, Bandung, Indonesia. ✉email: kawasaki@moleng.kyoto-u.ac.jp

Introduction

Atropical peatland is an ecosystem that stores a large amount of terrestrial carbon over millennia. Indonesian peatlands store ~7% of the global peat carbon pool (Page et al., 2011). In recent decades, the frequency of forest fires across the tropics has increased (Global fire emissions database, 2020). Forest fires in Indonesia tend to occur annually in the dry season, even in years of non-El Niño Southern Oscillation, primarily by anthropogenic activities for developing agriculture and plantations (Murdiyarso and Adiningsih, 2007). Peatlands are concentrated in the lowland regions of Sumatra and Kalimantan in Indonesia. Smoke developed there is a major cause of transboundary haze pollution. Air quality during burning periods exceeds the maximum level on the Pollutant Standard Index. This air quality impacts on human health, especially on respiratory illnesses because fire smoke contains hazardous air pollutants by the emission of carbon monoxide, nitrogen oxides, sulfur oxide and organic compounds from underground burning of peat (Naeher et al., 2007). Frankenberg et al. (2004) reported statistical evidence linking the 1997 forest fire to increased difficulty in daily living activities and negative impacts on respiratory and general health in Sumatra and Kalimantan. Kim et al. (2017) utilized the satellite-based aerosol index from a wide region of Indonesia, finding that the episodic shock of air pollution was significantly linked to clinical depression in women in that region. Agustian et al. (2020) coupled global air quality proxies with the Indonesia Family Life Survey data for 2014, showing that the prevalence of health outcomes with the denominator of adults was higher in the southern area of Sumatra (3.3%) and Central Kalimantan (4.3%) than the national average (2.9%). All studies provided statistical evidence of the negative impacts of air pollution on chronic diseases and human welfare.

The impact of fires on air quality is not limited to the haze source regions alone as pollutants can be transported to the atmosphere over a hundred km away. Population exposure to haze pollution depends on the geospatial location of fire hotspots and the extent of burning of peatlands (Heil et al., 2007). The GEOS-Chem chemical transport model has shown that the southeastern region in Sumatra is heavily contaminated for the population-weighted regions (Kim et al., 2015). With the recent progress in informatics, contents acquired from social and mass media have become one of the new tools of disseminating information about air pollution. Sociological information is complementary for describing the extent and range of health impacts of haze dispersion in regions where environmental monitoring and health surveillance data are unavailable. Using posts to Sina Weibo on smoke and haze, the air quality index was estimated (Jiang et al., 2015; Tao et al., 2016). In California wildfire cases, social media data were used to map burned areas (Sachdeva and McCaffrey, 2018). Crowdsourced data are useful for estimating the haze pollution from wildfire smoke affecting particulate matter (PM) to predict air quality impacts on a statewide level (Kent and Capello, 2013). Social media are data sources for disease surveillance, such as infectious diseases and asthma (Broniatowski et al., 2013; Ram et al., 2015). It was also supplementary and complementary information for haze disaster management in Riau, Indonesia, according to a correlation calculation with tweets and hotspots (Kibanov et al., 2017). The Google Trends data had a linear time-series pattern with the official dengue fever disease report for Indonesia during 2012–2016 (Husnayain et al., 2019). For monitoring outbreaks of dengue fever in India, newspaper articles were useful in providing possible proxy data (Zhang et al., 2019). Newspaper information on social issues also helps in the development of public concern. Newspaper information is reliable because accuracy, quality and newsworthiness of articles are ensured by editorial review processes.

To quantify its disease burdens and develop mitigation strategies, there is an urgent need for exposure and health outcomes data on a local scale. The exposure estimates from scientific measurements can then be used to study the health effects of air pollution in urban populations. To investigate this relationship and develop air quality management strategies, small-scale spatial variations in air pollution levels require small-scale health outcome evaluation. The planning and implementation of the air quality management policy need to be executed accordingly. According to Agustian et al. (2020), this is even more important in Indonesia where, with the existence of local autonomy, the authority to develop such policies is placed with local/city governments. Hence, extracting health outcome data fit for the aforementioned purpose by analyzing local newspapers is also challenging.

This study reveals that the health impact of peatland fires can be estimated by choosing appropriate search terms in the analysis of spatiotemporal distributions. The easy availability of mass media data makes it appealing for peatland-fire responders and public health professionals. Since newspaper articles on agricultural economy of peatland can be an indicator of public concern about commercial logging and plantations, such search terms have also been examined in this study.

Methods

Newspaper database. This study investigated the spatiotemporal characteristics of newspaper articles relevant to public health. Respiratory illness is the most important health issue in fire-prone regions of Sumatra. Figure 1 shows a map of the fire radiative power (FRP) that accumulated in the central region of Sumatra for a decade from January 2009 to December 2018. FRP in units of power, W , is the rate at which the actively burning fire is emitting radiative energy at the time of observation. FRP is a measure of a fire release rate of thermal radiation, and is strongly related to the rate of fuel consumption and trace gas and aerosol emission (Wooster et al., 2005; Freeborn et al., 2008). FRP is, therefore, an estimator for the amounts of material being emitted to the atmosphere from that fire. The integrated FRP is high in the eastern and southern regions of Sumatra in Fig. 1. FRP within a 100 km radius around Pekanbaru peaked yearly in 2012–2014, and monthly around February and June, as shown in Fig. S1. As Riau Province has been heavily affected by the haze caused by peatland fires, an archive dataset in the PDF form for 2009–2018 was purchased from a local newspaper company, *Riau Pos* (<https://epaper.riaupos.co/hubungi-kami.html>). The documents are missing three time-frames; January 2009, October–December 2015, and December 2018. *Riau Pos* typically publishes 40 pages daily with mentioning international, domestic and local news in the categories of society, business, politics, sports, food and editorials. It covers the Riau Province and nearby regencies, such as Siak, Kampar, Bengkalis, Rokan Hulu, Indragiri Hulu, Kepulauan Meranti, Pelalawan, Indragiri Hilir, Rokan Hilir and Kuantan Singingi. Big cities covered by it are Pekanbaru (state capital), Padang, Bengkalis, Siak, Pelalawan, and Dumai. The total number of PDF pages examined was ~140,000, corresponding to 700,000 articles over the decade.

Selection of search terms for analysis. Using the software described below, newspaper articles on public health, air pollution, peatland fires and agricultural economy were selected. The searched terms are *penyakit akut* (acute illness), *penyakit pernapasan akut* (acute respiratory illness), *kerusakan asap* (smoke damage), *polusi udara* (air pollution), *kebakaran gambut* (peatland fire), *penggunaan lahan* (land use) and *perkebunan kelapa*

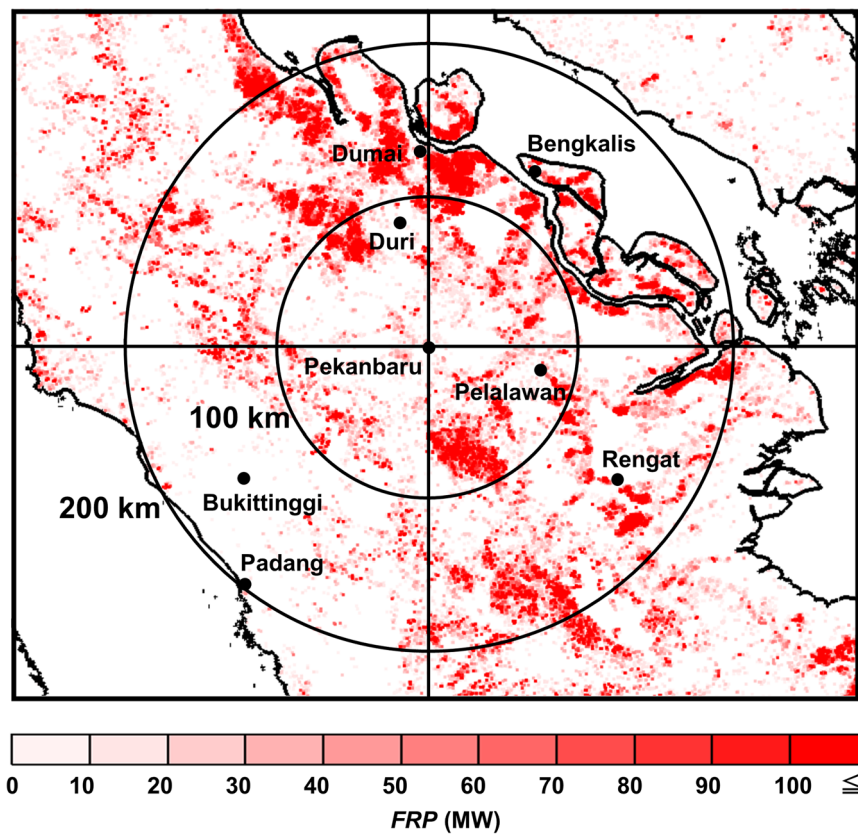


Fig. 1 Map of Fire Radiative Power (FRP) in the central region of Sumatra, Indonesia. Accumulated FRP for January 2009–December 2018. Circles represent radii centered at Pekanbaru, capital of Riau province.

sawit (oil palm plantation). Although some articles containing these search terms are not directly associated with fire events, they were not filtered based on their framing. This procedure for a popular search term reduced the temporal correlation between FRP and the article number because it appears even during a fire-less season, *e.g.*, on editorial opinions and review articles unrelated to fire events. This situation is specific to newspaper articles and is different from social media analysis because tweets usually synchronize with fire events. The number of other search terms listed in the Supplementary Information (SI) Section 1 was too low or too high for analysis. In the latter case, these articles appear throughout the year and are not associated with fire events.

Fire Radiation Power and airport visibility data. FRP is related to the rates of fuel consumption and emission of trace gases and aerosols, which is an appropriate index for surface fire and peat fire (ground fire) (Wooster et al., 2005; Freeborn et al., 2008). Since the correlation between monthly FRP and $PM_{2.5}$ is strong with a coefficient of determination $R^2 = 0.98$ for fires in 2015 around Sumatra and Borneo (Yin et al., 2020), FRP is an appropriate index to estimate the haze exposure to residents. The FRP data were obtained from the Fire Information for Resource Management System of the NASA Database (<https://firms.modaps.eosdis.nasa.gov/download/>). Since the newspaper articles were unavailable for the aforementioned time-frames, FRP data during the same time-frames were excluded from the analysis. Three-day FRP data of June 18–20, 2013, were also excluded for being highly irregular. The visibility for airport operation is reciprocally proportional to aerosol density that increases by fires (Griffin, 1980; Field et al., 2009). Ground-based monitoring data of airport visibility were obtained from Badan

Meteorologi, Klimatologi, dan Geofisika (BMKG; Meteorological, Climatological and Geophysical Agency of Indonesia, <http://aviation.bmkg.go.id/web/station.php>).

Selection of study regions. The primary study region is Riau Province in the central part of the east coast of Sumatra Island. This province has the highest hotspot occurrence in Sumatra, as shown in Fig. 1. Approximately 5500 peat fire hotspots were detected in Riau Province during 2014, while 7892 across Sumatra according to Kibanov et al. (2017). Zaini et al. (2020) reported that among residents exposed to peatland fires in Riau province in 2015, respiratory symptoms were reported in 71.4% subjects and lung function was impaired in 72.6%, mostly with mild obstruction and restriction. Non-respiratory symptoms were reported by 84.7% subjects. In the years of 2015–2017 the acute infection respiratory illness was the top cause for outpatient visits to clinics/hospitals in Riau province. In 2015, the high FRP year, over 40% of outpatients to hospitals suffer from this illness, while about 20% in 2017. (Dinas Kesehatan Provinsi Riau, 2020) To determine search terms in newspapers for appropriately expressing a spatiotemporal variation of public health condition affected by fires and air pollution, we connect in time N_{page} containing the search terms with the FRP data of the region within a 100 km radius centered at Pekanbaru, the capital city of Riau Province. Kibanov et al. (2017) investigated the spatial characteristics of tweets using a mapping process based on the distance between the position of a tweet and a hotspot. People living closer to a hotspot are more affected by haze than those living farther away. The popularity of tweets on haze and related health distributed at distances of 0–200 km. Notable spikes were present around 100 km. Based on this report, a 100 km radius area around the study city is chosen for FRP data. As typical wind speed is 2–3 m/s (90–130 km/half-day) during the

daytime, a 100 km radius is a reasonable distance. Since the advection of air mass from the southern regions of Sumatra occurs, the air mass mixing might affect the correlation between N_{page} and FRP that is limited within 100 km radius.

Search engine for articles of newspaper database in the form of PDF files. A search engine for terms in PDF files was used. Details of the software are described in SI Section 2. Briefly, the search engine functions per page, i.e., the search tool counts “one” when one article with a certain search term appears on a page. However, when two or more articles with the same term appear on the same page, the engine does not count more than “one”. The article page number, N_{page} , containing a high frequency search term could level-off with the increase of the article number, N_{art} . The restricted number of articles due to space limitations in a copy of the daily newspaper was also considered. This level-off effect will be discussed below.

Results

Correlation between FRP and the number of article page containing search terms. Using the search engine described above, the total numbers of article pages containing search terms, “acute illness”, “smoke damage”, “air pollution” and “peatland fire” are 575, 984, 857, and 1618 pages during the decade, respectively. The annual N_{page} is shown in Fig. 2, along with the annual FRP distribution. Among these search terms, Fig. 3a–c show the correlations between the annual N_{page} of “acute illness”/“smoke damage”/“air pollution” and FRP on a linear scale without time-lag, resulting in correlation coefficients, $R = 0.82/0.64/0.79$, respectively. With a 1-year lag in the N_{page} distributions, the correlation coefficients are $R = 0.70/0.78/0.73$, respectively. The changes with and without a lag are small. These search terms can be used as potential indices of public health concern posed by air pollutants and fire haze.

The total number of article pages containing “peatland fire” is 1618 pages. The R value for this search term is -0.09 without time-lag (Fig. 3d). Since this result appears unreasonable, this search term is reanalyzed as follows. The total page number 1618 is much greater than the numbers of the other search terms, suggesting saturation in the analysis. Then, the analysis is transformed from annual to monthly because fire-prone periods are mainly in February and June–August, as will be described below and in SI Section 3. Figure 4a shows the level-off effect in the distribution of N_{page} , yielding a linear correlation with $R = 0.43$ for zero time-lag. To explain the saturation effect, we hypothesize a logarithmic function for the page number distribution. This hypothesis, specific to this analysis, is verified by the model calculation described in SI Section 3. The calculation suggests that when N_{art} is large, the level-off effect in N_{page} results in a log count:

$$N_{page} \propto \log_{10} \left(1 + \frac{N_{art}}{m} \right) \tag{1}$$

where m is the saturation parameter that depends on the total page number per daily copy and the article number limit per page. The level-off effect is stronger with a smaller m , as discussed in the model calculation. The correlation between the logarithmically scaled FRP and the linearly scaled N_{page} is shown in Fig. 4b, in which the solid line is a best-fit for $R = 0.43$ with zero-month lag and is essentially the same as the R value in Fig. 4a. R decreases to 0.15 with one-month lag in N_{page} . These results indicate that the saturation due to the present methodology reduces the R value but this search term correlates with fire events.

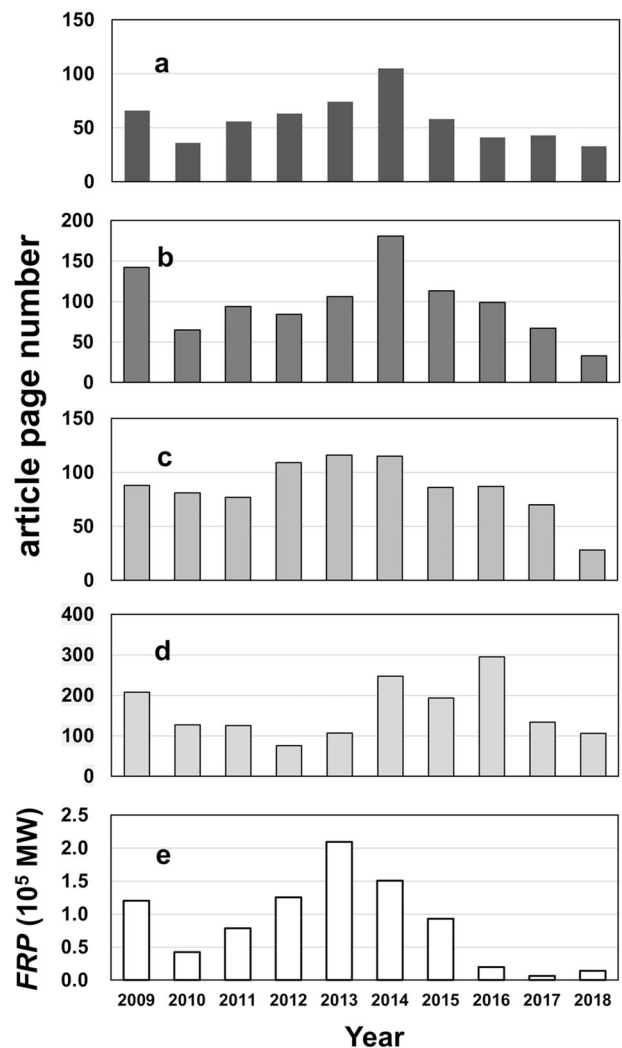


Fig. 2 Annual page numbers of articles relevant to search terms. a “acute illness”, **b** “smoke damage”, **c** “air pollution”, and **d** “peatland fire”. **e** annual Fire Radiation Power (FRP) around Pekanbaru within a radius of 100 km.

Next, we have analyzed the monthly correlation between the decade-integrated FRP and N_{page} (total 187 pages) of “acute respiratory illness”, as shown in Fig. 5. In Riau Province, the tropical equator weather is characterized by two fire-prone periods, February and June–August. Two seasonal peaks appear in both the figures for N_{page} and FRP . The R values with lags of FRP ranging of zero, one and two months are 0.62, 0.74 and 0.28, respectively. A 1-month lag yields the best temporal correlation. One-month lag observed here suggests that newsworthiness of public health increases with a lag of a few weeks after fire events (Xu et al., 2016).

Correlation between FRP and number of article pages containing “acute respiratory illness” geo-tagged with a city name: spatial patterns inferred from newspaper articles. Figure 6 shows the annual FRP and N_{page} of “acute respiratory illness” + “Pekanbaru”, where an appreciable correlation is observed. The correlation coefficients between FRP and N_{page} are $R = 0.63$ and 0.78 with a time-lag of 0 and 1 year in FRP , respectively. The slightly higher R value with the 1-year lag suggests that the newsworthiness of risk perception is partly developed from public experience in the previous year. Considering district tendencies, Pekanbaru, Bengkalis, Siak, Pelalawan, Dumai and Duri in Riau

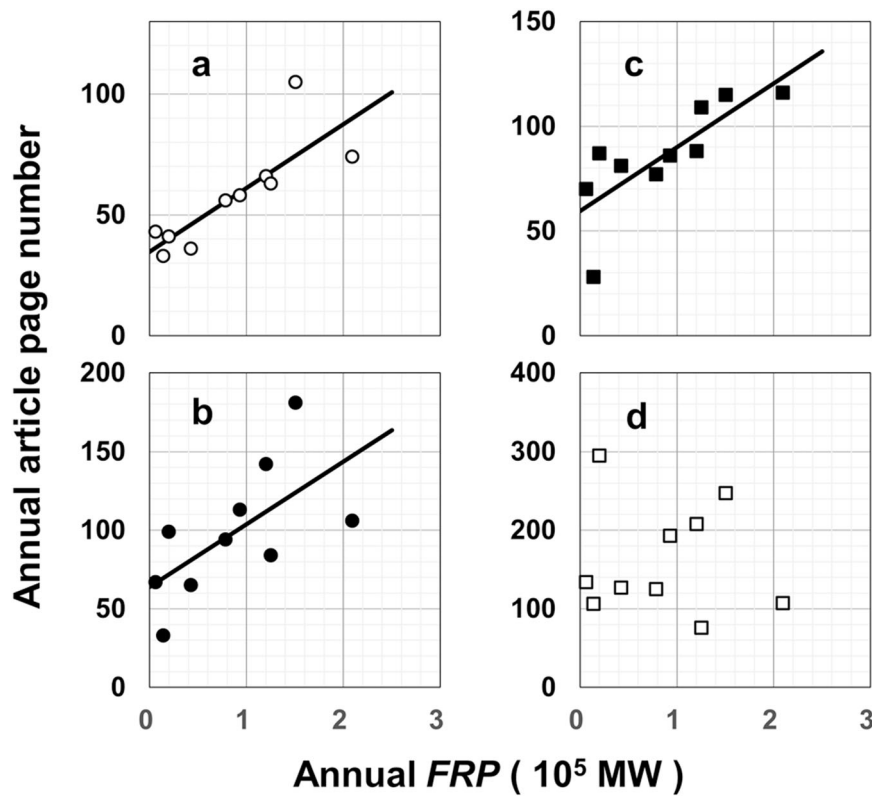


Fig. 3 Correlation diagrams of annual data during 2009-2018. **a** Article page number of “acute illness” vs. *FRP*. $R = 0.82$, **b** “smoke damage” $R = 0.64$, **c** “air pollution” $R = 0.79$, **d** “peatland fire” $R = -0.09$.

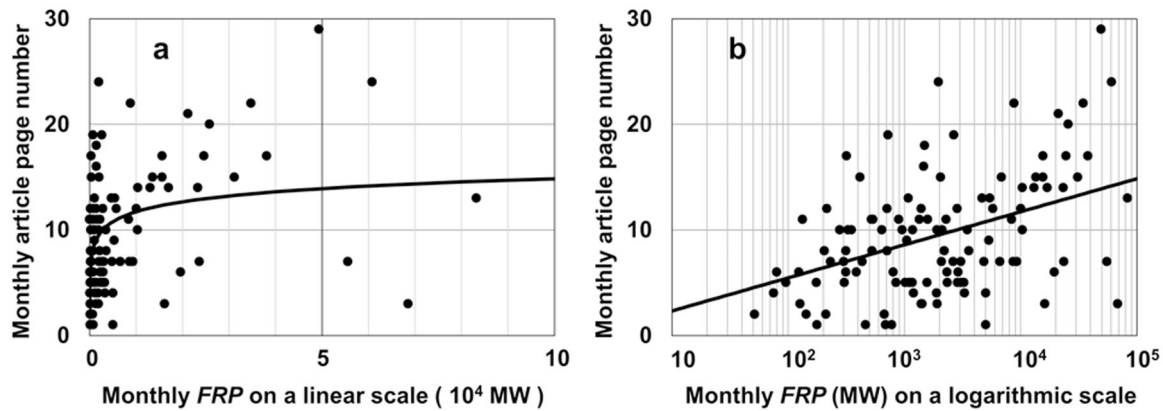


Fig. 4 Correlation diagrams of monthly data during 2009-2018. Article page number of “peatland fire” vs. *FRP*. The article number is on a linear scale while *FRP* is on **a** a linear scale or **b** a logarithmic scale. The solid line in **b** is a best-fit with $R = 0.43$, which corresponds to the solid curve in **a**.

Province, and Padang and Bukittinggi in West Sumatra Province are searched. These cities have over 100,000 inhabitants. Since the BMKG stations are located only in three cities, newspaper information is useful for public health management in other cities. Table 1 shows the decade-integrated N_{page} for “acute respiratory illness” + “city name” and their ratios over city population. The six cities located in Riau Province have higher ratios than those of the two cities in West Sumatra. The statistics on pneumonia cases for children under five in 2018 and 2019 reported that the percentages are Siak > Dumai, Pekanbaru, Pelalawan > Bengkalis. (Dinas Kesehatan Provinsi Riau, 2020) This order is in fair agreement with the ratios in Table 1.

To confirm the correlation between those ratios and air quality data, airport visibility data for 2012–2019 at three BMKG stations (Pekanbaru, Padang, and Rengat) are examined to provide

spatiotemporal information on the haze contamination of air. Figure 7 shows the reversal visibility data at the three stations, corresponding to the optical density of polluted air or haze level (Griffin, 1980). The temporal profile at the Pekanbaru BMKG station peaks in 2014 and 2015, synchronizing with the peaks in the annual N_{page} for “acute respiratory illness” as seen in Fig. 6a. Regarding visibility, the eastern side of the central region of Sumatra, Pekanbaru and Rengat, suffered heavily from haze compared to the western side region, Padang. These observations support the results of newspaper analysis in Table 1, i.e., the N_{page} of Pekanbaru is seven times higher than that of Padang.

Public concern about fire causes. Since media attention and public concern cycles are linked, risk perception in relation to

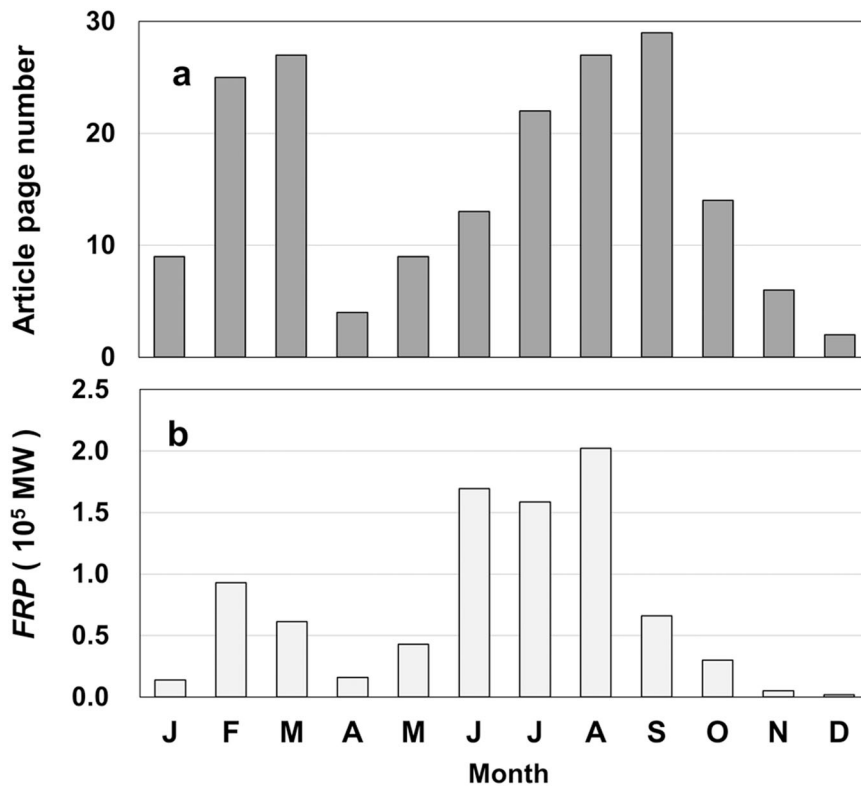


Fig. 5 Monthly data accumulated during 2009-2018. a Article page number of “acute respiratory illness”, **b** FRP. $R = 0.62, 0.74,$ and 0.28 with time-lag zero, one, and 2 months in the distribution of page number, respectively.

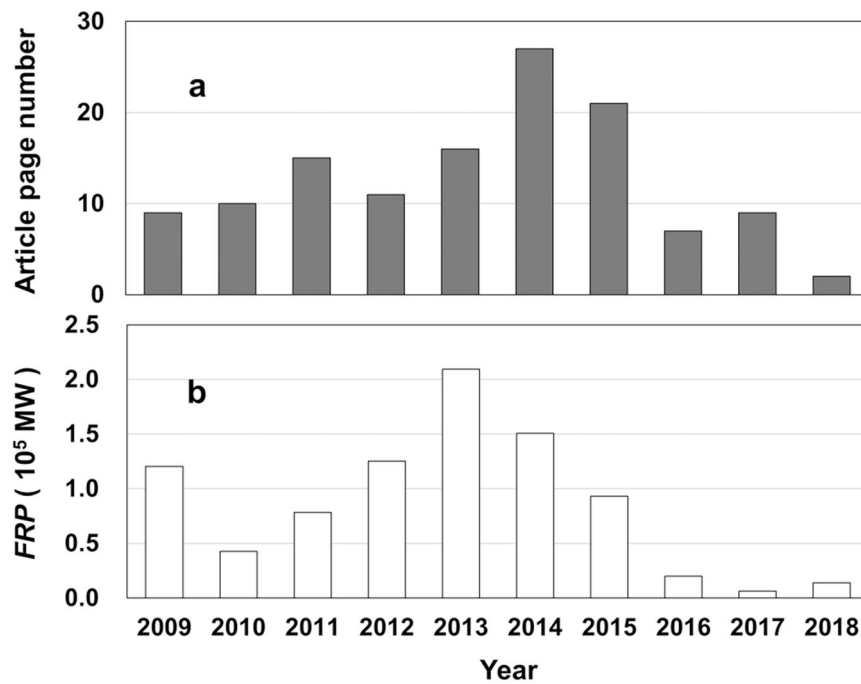


Fig. 6 Annual data for 2009-2018. a Article page number of “acute respiratory illness” + “Pekanbaru”, **b** FRP. $R = 0.63$ and 0.78 with time-lag zero and 1 year, respectively.

peatland fire is investigated with the search terms relevant to fire destruction as well as an agricultural economy, *i.e.*, “land use” and “oil palm plantation”. The total page numbers for the decade are 3446 and 4881 pages, respectively. The *R* values between the annual N_{page} and *FRP* are 0.59 and 0.71, respectively, indicating that the newspaper identifies well the fire cause, namely, the conversion of forests and peatland into palm oil plantations through slash-and-burn techniques induces severe fire effects. Because these articles cover various newspaper categories such as the economy, politics, living and environment, the level-off effect

is weak for a large N_{page} as is mentioned in the model calculation in SI Section 3, *i.g.*, the maximum page number of these search terms per daily copy is large. In this case, the saturation parameter *m* in Eq. (1) becomes large.

Discussion

Here, we have found appreciable correlations between the index of Fire Radiation Power and the annual article page numbers of “acute illness”/“smoke damage”/“air pollution”/“acute respiratory illness”. An appreciable correlation is also observed between the fire index and the article page number of “acute respiratory illness” + “Pekanbaru” (state capital name). The six cities located in Riau Province of the central region of Sumatra are characterized by the higher article ratios per population for “acute respiratory illness” than the two cities in West Sumatra province. Higher *FRP* or concentration of particulate matter are observed in Riau Province in visibility measurements. Concerning the risk perception in relation to peatland fire, the search terms for the fire destruction are strongly correlated with fire events.

Table 1 Regional dependence of page numbers of articles mentioning both “acute respiratory illness” and “city name”.

| City ^a | N_{page}^b | Population ^c 10 ⁵ | Ratio ^d | <i>FRP</i> ^e 10 ² GW | Province ^f |
|-------------------|--------------|--|--------------------|---|-----------------------|
| Pekanbaru | 127 | 9.0 | 14 | 9 | Riau |
| Siak* | 60 | 3.8 | 16 | 13 | Riau |
| Pelalawan* | 39 | 3.0 | 13 | 10 | Riau |
| Dumai* | 27 | 2.5 | 11 | 21 | Riau |
| Bengkalis* | 28 | 5.0 | 6 | 15 | Riau |
| Duri* | 7 | 1.4 | 5 | 18 | Riau |
| Rengat | 16 | 0.9 | — | 10 | Riau |
| Padang | 19 | 9.4 | 2 | 0.2 | West Sumatra |
| Bukittinggi* | 1 | 1.2 | 1 | 1.2 | West Sumatra |

^aThe asterisks indicate metropolis areas where BMKG stations are not available.
^bTotal number of article pages accumulated in the period February 2009–November 2018.
^cCensus in 2010 from <http://citypopulation.info/>.
^d N_{page} divided by population (units of 10⁵).
^eFire Radiation Power (*FRP*) in the area within a radius of 100 km centered at each city, which are accumulated for 2009–2018.
^fRiau Province is in the east side of the central region of Sumatra while West Sumatra Province in the west side.

Public health and Fire Radiation Power via mutual correlations with particulate matter. Peatland fires emit organic aerosols, which are the primary components of smoke-related surface PM concentrations. Peatland fires are classified as surface and peat fires. Surface fires are detected by satellite observations as hotspots, which typically burn during the daytime. Peat fires detected by thermal analysis of satellite images continue to burn for ~2 weeks, emitting a large amount of PM to air (Atwood et al., 2016). Studies conducted in air-polluted regions have found that acute increases in ambient air pollution led to increases in the hospital and doctor visits for respiratory infections (North et al., 2019). Review papers have reported consistent evidence of associations between wildfire smoke exposure with general respiratory morbidity and with exacerbations of asthma and chronic obstructive pulmonary disease (Reid et al., 2016; Uda et al., 2019). Epidemiological evidence

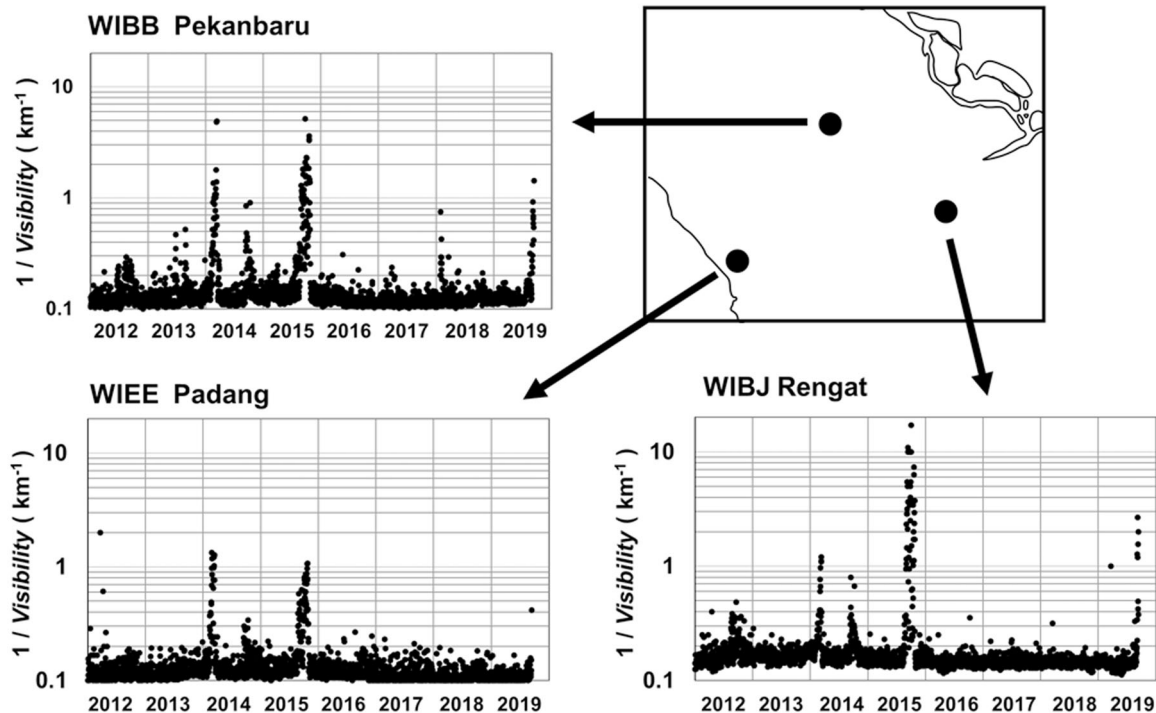


Fig. 7 Spatiotemporal variations of the reverse values of visibility on a logarithmic scale for three BMKG stations. Pekanbaru, Padang, and Rengat in the central region of Sumatra. The abbreviations are the station names of World Meteorological Organization.

and plausible toxicological mechanisms suggest an association between wildfire smoke exposure and respiratory illness. A multivariate analysis was performed on the severity of respiratory problems caused by PM from the 1997 haze disaster in Sumatra (Kunii et al., 2002). As the correlation between FRP and monthly average of $PM_{2.5}$ is strong for fires in Sumatra (Yin et al., 2020), FRP is considered an appropriate index for estimating haze exposure to residents and respiratory illness. In our study, the newspaper article volume on public health has a strong relationship with FRP and PM concentration.

MODIS hotspot measurement by NASA, has been adopted as an index of peatland fire intensity or air pollution in tropical regions. In this study, we use FRP as an index because the hotspot number tends to be lower than the real number. One of the causes is the thick smoke that blocks fire signals from reaching the MODIS sensor, in addition to cloud constraints, fire area, and the time factor of the satellite trajectory compared with the timing of the fire. Even so, a strong correlation was reported between the weekly numbers of hotspots and tweets on “haze-health” in Riau Province, 2014, i.e., between the number of detected hotspots on a linear scale and the weekly number of conversations in a logarithmic scale, confirming that people have exponentially increased conversations on “haze-health” after a fire (Kivanov et al., 2017). Regarding the publication frequency of newspaper articles, Figs. 3, 5, and 6 suggest an appreciable correlation between FRP and the search terms of “acute illness”, “smoke damage”, “air pollution” and “acute respiratory illness”.

Spatial variances in newspaper article numbers on public health. This section discusses the spatio-relationship between FRP and N_{page} in the context of health issues. Identification of unhealthy conditions of geospatial locations from mass media would be useful for the prioritization of disaster response and humanitarian action during haze disasters. The results described above suggest that “acute illness” and “acute respiratory illness” are useful search terms, reflecting the public health impact of air pollution and smoke damage from peatland fires. In the Tweeter content analysis, tweets about wildfire and smoke predicted public health impacts (Sachdeva and McCaffrey, 2018). The assessment of population-level exposure in the United States was explored using daily posts from Facebook (Ford et al., 2017). Comparing the $PM_{2.5}$ levels, the best correlations reported were for “air quality”, “wildfire” and “smoke” since wildfire smoke was the source of the highest variability in surface $PM_{2.5}$ concentrations. The search terms “haze” and “health” were relevant to residents across Sumatra, becoming popular terms in social media (Kibanov et al., 2017). A study of local newspapers in India showed a significant correlation between the number of dengue fever cases and the number of associated articles from local major newspapers (Zhang et al., 2019).

As for the regional characteristics of public concern regarding “acute respiratory illness”, the ratios of N_{page} divided by *population* are shown in Table 1, which are much higher in Riau Province than in West Sumatra. The FRP is also high in the eastern region (Fig. 1). Figure 7 indicates the association of the search terms with the PM. The large ratios, 11–18, are independent for five most populated eastern cities, Pekanbaru, Siak, Pelalawan, Dumai and Rengat, because haze spreads all over the eastern side of the central region of Sumatra (Kim et al., 2015). Close to the eastern coast, one-third of the land is underlain by peatlands (Sze et al., 2019). The eastern region is characterized by high FRP as shown in Fig. 1, and suffers from the haze from underground peat burns during the fire periods (Marlier et al.,

2015a). The air quality in the fire season was the worst in the eastern cities, Pekanbaru, Siak, Bengkalis and Dumai (Kibanov et al., 2017). Among these cities, Siak, Pelalawan and Dumai have no BMKG stations. Local information on acute illnesses can be obtained from newspaper articles, even if haze data from weather stations are unavailable.

Newspaper articles on the causes of peatland fires. Palm oil is an important vegetable oil in terms of production quantity and Indonesia is the world’s largest producer (Koh and Ghazoul, 2010). The demand for palm oil is driving deforestation to clear land for new plantations. The clearance and maintenance of lands are managed mainly through prescribed fire (Marlier et al., 2015b). In the evaluation of the social factors behind the 2015 extreme fire event in Sumatra, a variable of agricultural economy, such as the proportion of plantation land holdings, is one of the factors in explaining fire count at the regency-level (Sze et al., 2019). In combating peatland fires, high levels of acknowledgment and engagement are needed for changes toward more sustainable behaviors. In this aspect, newspaper media plays an important role in shaping public perceptions of fire causes. The newspaper search terms “land use” and “oil palm plantation” appear under various article categories. The appreciable *R* values suggest that newsworthiness of the fire causes, namely, the conversion of forests and peatlands into palm oil plantations through slash-and-burn techniques, enhances the changes in people’s behavior toward more sustainable actions. Newspaper media are influenced by public perception and support public involvement. Thus, government or planning authorities are expected to increase the dissemination of information on the prevention of fires through newspaper articles.

Conclusions

Newspaper epidemiology information is not generated with the primary purpose of doing epidemiology but a useful source in the analysis of public health and environmental contexts. (Salathé, 2018). This study has revealed the utility of articles from a local newspaper, *Riau Pos*, for filling in estimates of “acute illness (*penyakit akut*)” and “acute respiratory illness (*penyakit pernapasan akut*)” related to air pollution. Newspaper analysis can be a potential tool for public health sectors for estimating health impacts from deteriorated air quality due to smoke haze. To solve the problems caused by fire and particulate matter, the local government needs to find a way of managing peatlands. That will require strengthening governance and accountability. Newspapers play an important role in informing people about the initiatives taken to manage such social issues. The interdisciplinary scientific framework to evaluate public health implications is publicly available through a decision support tool of mass media. As the development of digital databases in Indonesia, interdisciplinary research like this study will become popular in media research. The present methodology will contribute to the issue of forest fires globally because forest fires are not only a problem in Indonesia but also occur in many other countries

Data availability

The datasets supporting the conclusions of this article are included in the article and Supplementary Information. If readers would like further information about data, please contact the authors.

Received: 24 June 2021; Accepted: 11 November 2021;

Published online: 01 December 2021

References

- Agustian D, Rachmi CN, Indraswari N, Molter A, Carder M, Rinawan FR, van Tongeren M, Drijeana D (2020) Feasibility of Indonesia Family Life Survey Wave 5 (IFLS5) data for air pollution exposure–response study in Indonesia. *Internat. J Environ Res Public Health* 17:9508. <https://doi.org/10.3390/ijerph17249508>
- Atwood EC, Englhart S, Lorenz E, Halle W, Wiedemann W, Siegert F (2016) Detection and characterization of low temperature peat fires during the 2015 fire catastrophe in Indonesia using a new high-sensitivity fire monitoring satellite sensor (FireBird). *PLoS ONE* 11:e0159410. <https://doi.org/10.1371/journal.pone.0159410>
- Broniatowski D, Paul MJ, Dredze M (2013) National and Local Influenza Surveillance through Twitter: an analysis of the 2012–2013 Influenza Epidemic *PLoS ONE* 8:e83672. <https://doi.org/10.1371/journal.pone.0083672>
- Dinas Kesehatan Provinsi Riau (2020) Profil Kesehatan Provinsi Riau 2019 and other years. <http://www.dinkes.riau.go.id>
- Field RD, van der Werf GR, Shen SP (2009) Human amplification of drought-induced 190 biomass burning in Indonesia since 1960. *Nat Geosci* 2:185–188. <https://doi.org/10.1038/ngeo0443>
- Ford B, Burke M, Lassman W, Pfister G, Pierce JR (2017) Status update: is smoke on your mind? Using social media to assess smoke exposure. *Atmos Chem Phys* 17:7541–7554. <https://doi.org/10.5194/acp-17-7541-2017>
- Frankenberg E, McKee D, Thomas D (2004) Health Consequences of forest fires in Indonesia. *Demography* 42:109–129. <https://doi.org/10.1353/dem.2005.0004>
- Freeborn PH, Wooster MJ, Hao WM, Ryan CA, Nordgren BL, Baker SP, Ichoku C (2008) Relationships between energy release, fuel mass loss, and trace gas and aerosol emissions during laboratory biomass fires. *J Geophys Res Atmos* 113:D01301. <https://doi.org/10.1029/2007JD008679>
- Griffin GW (1980) Relations between the prevailing visibility, nephelometer scattering coefficient and sun photometer turbidity coefficient. *Atmos Environ* 14:577–584. [https://doi.org/10.1016/0004-6981\(80\)90089-X](https://doi.org/10.1016/0004-6981(80)90089-X)
- Global fire emissions database (2020). <http://www.globalfiredata.org/updates.html>
- Heil A, Langmann B, Aldrian E (2007) Indonesian peat and vegetation fire emissions: study on factors influencing large-scale smoke haze pollution using a regional atmospheric chemistry model. *Mitig Adapt Strat Glob Change* 12:113–133. <https://doi.org/10.1007/s11027-006-9045-6>
- Husnayain A, Fuad A, Lazuardi L (2019) Correlation between Google Trends on dengue fever and national surveillance report in Indonesia. *Global Health Action* 12:1552652. <https://doi.org/10.1080/16549716.2018.1552652>
- Jiang W, Wang Y, Tsou M-H, Fu X (2015) Using social media to detect outdoor air pollution and monitor air quality index (AQI): a geo-targeted spatiotemporal analysis framework with Sina Weibo (Chinese Twitter). *PLoS ONE* 10:e0141185. <https://doi.org/10.1371/journal.pone.0141185>
- Kent JD, Capello HT (2013) Spatial patterns and demographic indicators of effective social media content during the Horseshoe Canyon fire of 2012. *Cartogr Geogr Inf Sci* 40:78–89. <https://doi.org/10.1080/15230406.2013.776727>
- Kibanov M, Stumme G, Amin I, Lee JG (2017) Mining social media to inform peatland fire and haze disaster management. *Soc Netw Anal Min* 7:art. 30. <https://doi.org/10.1007/s13278-017-0446-1>
- Kim PS, Jacob DJ, Mickley LJ, Koplitz SN, Marlier ME, DeFries RS, Myers SS, Chew BN, Mao YH (2015) Sensitivity of population smoke exposure to fire locations in equatorial Asia. *Atmos Environ* 102:11–17. <https://doi.org/10.1016/j.atmosenv.2014.09.045>
- Kim Y, Knowles S, Manley J, Radoias V (2017) Long-run health consequences of air pollution: Evidence from Indonesia's forest fires of 1997. *Econ Hum Biol* 26:186–198. <https://doi.org/10.1016/j.ehb.2017.03.006>
- Koh LP, Ghazoul J (2010) Spatially explicit scenario analysis for reconciling agricultural expansion, forest protection, and carbon conservation in Indonesia. *Proc Natl Acad Sci USA* 107:11140–11144. <https://doi.org/10.1073/pnas.1000530107>
- Kunii O, Kanagawa S, Yajima I, Hisamatsu Y, Yamamura S, Amagai T, Ismail IT (2002) The 1997 haze disaster in Indonesia: its air quality and health effects. *Arch Environ Health* 57:16–22. <https://doi.org/10.1080/00039890209602912>
- Marlier ME, DeFries RS, Kim PS, Gaveau DL, Koplitz SN, Jacob DJ, Mickley LJ, Margono BA, Myers SS (2015a) Regional air quality impacts of future fire emissions in Sumatra and Kalimantan. *Environ Res Lett* 10:054010. <https://doi.org/10.1088/1748-9326/10/5/054010>
- Marlier ME, DeFries RS, Kim PS, Koplitz SN, Jacob DJ, Mickley LJ, Myers SS (2015b) Fire emissions and regional air quality impacts from fires in oil palm, timber, and logging concessions in Indonesia. *Environ Res Lett* 10:085005. <https://doi.org/10.1088/1748-9326/10/8/085005>
- Murdiyarso D, Adiningsih ES (2007) Climate anomalies; Indonesian vegetation fires and terrestrial carbon emissions. *Mitig Adapt Strat Glob Change* 12:101–112. <https://doi.org/10.1007/s11027-006-9047-4>
- Naeher LP, Brauer M, Lipsett M, Zelikoff JT, Simpson CD, Koenig JQ, Smith KR (2007) Wood smoke health effects: a review. *Inhal Toxicol* 19:67–106. <https://doi.org/10.1080/08958370600985875>
- North CM, Rice MB, Ferkol T, Gozal D, Hui C, Jung SH, Kuribayashi K, McCormack MC, Mishima M, Morimoto Y, Song Y, Wilson KC, Kim W, Fong KM (2019) Air pollution in the Asia-Pacific region. *Respirology* 24:484–91. <https://doi.org/10.1111/resp.13531>
- Page SE, Rieley JO, Banks CJ (2011) Global and regional importance of the tropical peatland carbon pool. *Glob Change Biol* 17:798–818. <https://doi.org/10.1111/j.1365-2486.2010.02279.x>
- Ram S, Zhang W, Williams M, Pengetnze Y (2015) Predicting asthma-related emergency department visits using big data. *IEEE J Biomed Health Inform* 19:1216–1223. <https://doi.org/10.1109/JBHI.2015.2404829>
- Reid CE, Brauer M, Johnston FH, Jerrett M, Balmes JR, Elliott CT (2016) Critical review of health impacts of wildfire smoke exposure. *Environ Health Perspect* 124:1334–1343. <https://doi.org/10.1289/ehp.1409277>
- Sachdeva S, McCaffrey S (2018). Using Social Media to Predict Air Pollution during California Wildfires. *Proceed Internat Conf Social Media & Society*, Copenhagen, Denmark. 365–369, ACM 978-1-4503-6334-1/18/07. 2018. <https://doi.org/10.1145/3217804.3217946>
- Salathé M (2018) Digital epidemiology: what is it, and where is it going? *Life Sci Soc Policy* 14:1. <https://doi.org/10.1186/s40504-017-0065-7>
- Sze JS, Jefferson, Lee JSH (2019) Evaluating the social and environmental factors behind the 2015 extreme fire event in Sumatra. *Indonesia Environ Res Lett* 14:015001. <https://doi.org/10.1088/1748-9326/aace1d>
- Tao Z, Kokas A, Zhang R, Cohan DS, Wallach D (2016) Inferring Atmospheric Particulate Matter Concentrations from Chinese Social Media Data. *PLoS ONE* 11:e0161389. <https://doi.org/10.1371/journal.pone.0161389>
- Uda SK, Hein L, Atmoko D (2019) Assessing the health impacts of peatland fires: a case study for Central Kalimantan, Indonesia. *Environ Sci Pollut Res* 26:31315–31327. <https://doi.org/10.1007/s11356-019-06264-x>
- Wooster MJ, Roberts G, Perry GLW, Kaufman YJ (2005) Retrieval of biomass combustion rates and totals from fire radiative power observations: FRP derivation and calibration relationships between biomass consumption and fire radiative energy release. *J Geophys Res Atmosphere* 110(D24):D24311. <https://doi.org/10.1029/2005JD006318>
- Xu Q, Li X, Wang S, Wang C, Huang F, Gao Q, Wu L, Tao L, Guo J, Wang W, Guo X (2016) Fine particulate air pollution and hospital emergency room visits for respiratory disease in urban areas in Beijing, China, in 2013. *PLoS ONE* 11:e0153099. <https://doi.org/10.1371/journal.pone.0153099>
- Yin S, Wang X-F, Guo M, Santoso H, Guan H (2020) The abnormal change of air quality and air pollutants induced by the forest fire in Sumatra and Borneo in 2015. *Atmos Res* 243:105027. <https://doi.org/10.1016/j.atmosres.2020.105027>
- Zaini J, Susanto AD, Samoedro E, Bionika VC, Antariksa B (2020) Health consequences of thick forest fire smoke to healthy residents in Riau, Indonesia: a cross-sectional study. *Med J Indonesia* 29:58–63. <https://doi.org/10.13181/mji.oa.204321>
- Zhang Y, Ibaraki M, Schwartz FWT (2019) Disease surveillance using online news: an extended study of dengue fever in India. *Trop Med Health* 47:58. <https://doi.org/10.1186/s41182-019-0189-y>

Acknowledgements

The authors thank BMKG (Indonesia) and NASA (U.S.A.) for providing us with the data of airport visibility and Fire Radiation Power in Indonesia. This work was supported by a research program of Consultative Group on International Agricultural Research from Center for International Forestry Research, a KAKENHI Grant (18KK0294) of Japan Society for Promotion of Sciences, and a project (No. 14200117) of Research Institute for Humanity and Nature.

Competing interests

The authors declare no competing interests.

Ethical approval

The paper does not involve work directly with human participants.

Informed consent

The paper does not involve work directly with human participants.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1057/s41599-021-00994-5>.

Correspondence and requests for materials should be addressed to Masahiro Kawasaki.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021