# FINAL REPORT

Title: Estimating fire smoke related health burden and novel tools to manage impacts on urban populations

JFSP PROJECT ID: 14-1-04-9

**February 2019**

Brian J Reich, **North Carolina State University**

Ana G Rappold, **US Environmental Protection Agency**

Fay H Johnston, **University of Tasmania**

Geoffrey G Morgan, **University of Sydney**

Neal L Fann, **US Environmental Protection Agency**

Martin E Cope, **Commonwealth Scientific and Industrial Research Organisation**

Richard A Broome, **Health Protection NSW**

FIRESCIENCE.GOV
Research Supporting Sound Decisions

DOI Office of Wildland Fire

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the U.S. Government. Mention of trade names or commercial products does not constitute their endorsement by the U.S. Government.

# FRONT MATTER

Table of Contents



List of Figures



List of Tables

# ABSTRACT

Fire smoke is a major contributor to both particulate matter (PM) and ozone exposure in urban centers. Epidemiological, clinical, and toxicological studies have demonstrated a casual relationship between these pollutants and cardiovascular and respiratory related deaths and illnesses. Given the expected increase in fire events due to climate and landscape changes, quantifying health effects of wildfire smoke and developing real-time tools to help air quality managers mitigate the effects of exposure are sorely needed. We comprehensively assessed the health risks posed by smoke exposure, and constructed new tools to estimate and forecast smoke concentration levels and associated health effects. We accomplished these goals with four specific aims. In Aim 1, we compared the chemical composition of fine PM emanating from fire smoke with typical urban PM in the US. This information fills an important gap in the literature, and enabled us to distinguish between smoke and non-smoke related exposures in air quality indices and epidemiological studies. In Aim 2, we conducted a systematic review and meta analysis of the risk estimates to evaluate the risks of smoke exposure for all relevant health outcomes. Using these pooled risk estimates of the health effects of smoke exposure, in Aim 3 we utilized the BenMAP-CE air pollution tool to characterize the health and economic value of the effects of fire smoke, such as the number of asthma exacerbations or hospital admissions and the associated economic cost. Finally, in Aim 4 we combined model-based predicted smoke exposure with health and economic assessment tools to provide real-time forecasts of health risk over space and time. This new tool provides easily interpretable information to assist officials in their decision processes to mitigate health effects, such as issuing hourly targeted advisories. To accomplish these tasks, we assembled a diverse team from US and Australian research institutions and government with extensive experience in air quality modeling, biostatistics, economics, and epidemiology.  In this report, we describe the results of this work and their impact on the field.

# BACKGROUND

Wildland fires are a major contributor to poor air quality, a leading risk factor for premature death and poor health globally (Ezzati, 2002). In its synthesis of the health literature, the U.S. EPA Integrated Science Assessments (EPA, 2013) found that short and long term exposure to fine particulate matter (PM) is causally associated with cardiovascular mortality and morbidity, while short and long term ozone exposure is causally associated with respiratory morbidity and mortality. Wildland fires can also release large quantities of stored carbon into the atmosphere further contributing to climatic changes that make some landscapes more vulnerable. Both the frequency and intensity of these fires are expected to increase globally due to dryer conditions, changes in landscape cover, and population growth (Bowman, 2009). It is therefore increasingly important to better characterize the contribution of fire smoke to the overall health burden and provide tools for managing impacts.

The association between short-term and longer-term exposure to PM from urban and industrial sources is very well established (Pope and Dockery, 2006). There is now solid evidence that extreme pollution from severe, episodic wild fires similarly affects a range of health outcomes including short-term mortality (Johnston et al, 2011; Sastry, 2002; Morgan et al, 2010) and hospitalization (Morgan et al, 2010; Delfino et al, 2009; Martin et al, 2013).

In the last several decades the fraction of population impacted by smoke has been steadily growing. The most recent 2011 U.S. EPA National Emissions Inventory estimated that wildfires were responsible for 20% of the 6.31 million tons of the total PM2.5 emitted that year. While these estimates vary from year to year depending on the intensity of a wildfire season, such events continue to degrade air quality at local and regional scales. In Australia, fires are integral to the landscapes and a characteristic feature of the environment. Changes in biota have resulted in frequent large fires near urban centers. The use of planned fire to limit the extent and occurrence of wildfire is now a central strategy used by land managers to reduce the risk from severe fires. While the direct damages from fires are well characterized, less research has been conducted toward understanding the indirect impacts on health and toward developing smoke management and exposure mitigation tools in affected populations.

In this project, we provide a comprehensive review of the literature on the effects of fire smoke on human health and estimates of smoke attributable health and economic burden. The burden analysis combines a meta-analysis of the epidemiological literature with dynamic model runs over the Continental United States with and without fire emissions. We also developed a real-time forecasting tool that combines station data with numerical model output using state-of-the-art data fusion methods. To achieve these objectives, we also develop several novel statistical method in the area of spatiotemporal analysis, causal inference, and machine learning. The expected benefit of the project is interpretable and actionable information on the likelihood of health impacts for public health officials and smoke managers during smoke events. Such information can facilitate decision processes involved in mitigating health effects during smoke events in the communities and justify future investments in fire prevention and public health protection programs.

# OBJECTIVES

The project was framed around the following four aims (text taken directly from the proposal)

**Aim 1: Characterize emissions from fire smoke over the last decade**

The objective of this aim is to investigate the chemical composition and spatial variability of PM2.5 from fire smoke events and contrast it with background PM2.5 and PM2.5 in urban centers. Fire smoke is a potent mix of chemical compounds including toxic gases and PM. Emerging epidemiologic literature suggests that specific chemical components may have stronger health effects than others. We hypothesize that the composition will differ from urban areas and have a greater fraction of organic material and secondary contribution from semi-volatile gases. The results of this comparison will provide important evidence regarding the generalizability of risk estimates from urban PM2.5 to fire smoke events, and the use of these risk estimates in health impacts assessment of fire smoke exposure (Aim 3).

**Aim 2: Develop fire smoke and health risk estimates based on a systematic review**

We will develop robust risk estimates for a range of health outcomes based on the systematic review of epidemiological studies of fire smoke exposure. For those outcomes with a sufficient number of studies we will use meta-analyses to provide a pooled risk estimate for use in the health impact assessment (Aim 3). Meta-regression techniques help identify characteristics that may explain differences in risk across studies, including population susceptibility and type of fire, the role of duration of exposure, and existence of safety threshold levels. Finally, we will use these results, combined with the results of Aim 1, to assess the generalizability of health risk estimates from urban air pollution compared to fire smoke.

There is some evidence that short-term mortality impacts are similar between fire smoke and urban PM. Studies have shown short-term impacts of fire smoke PM on respiratory morbidity to be generally of greater magnitude with equivalent concentrations of fire smoke derived PM compared to urban PM, while observed short-term cardiovascular morbidity is far less conclusive for fire smoke compared to urban PM. Additionally, there are no fire smoke studies of chronic exposure. Estimates of the long term impacts of exposure to fire smoke have used health risk estimates from studies of the long -term exposure to urban PM. The appropriateness of this approach will be informed by the results of Aim 1 and the similarities and differences in the concentration response functions from epidemiological and toxicological studies identified in Aim 2. The observed similarities and differences between urban and smoke PM effects have not been rigorously tested, yet are essential for robust assessments of community level health impacts (Aim 3).

**Aim 3: Perform a health economic impact assessment**

Using robust health risk estimates developed in Aim 2 and previously published estimates determined appropriate in Aim 2 for fire smoke events, we will perform health impacts assessment for each year of the study period. The assessment will potentially include cases of premature deaths, non-fatal heart attacks, respiratory and cardiovascular hospital visits, emergency department visits, acute respiratory symptoms, aggravated asthma, loss of productivity, and other outcomes. The health impact assessment will quantify the cumulative number of smoke-attributable premature deaths and illnesses, and the economic value of these outcomes. This approach will also focus on case studies of the impacts of specific events in North Carolina and Sydney, and nationally in both the US and Australia.

**Aim 4: Develop tools for predicting health impacts in real time**

As the population impacted by smoke grows, it is becoming increasingly important to develop tools for both prescribed and wild fires which can be used to inform public health officials and smoke managers on specific actions they should take during these events in real time. While a number of countries already have daily forecasts of air quality accessible from mobile devices, there are no programs that link forecasts to public health programs. Chemical transport model (CTM) forecasts of the spread of plumes already play a crucial role in control and intervention. As with any numerical model, CTM forecasts include error and systematic bias. In this aim, we will develop novel statistical models that combine several sources of data, including PM2.5 monitor observations, satellite measurements, and CTM output to estimate and adjust for biases and quantify forecast error. This will provide managers more reliable forecasts and measures of uncertainty, which are both critical to the decision-making process.

We will combine these calibrated smoke forecasts with the risk estimates identified in Aim 2 and population incidence rates in BenMAP to provide new tools to map predicted air quality and health impacts in real time. These predictive tools will estimate the fire smoke related health impacts based on hourly forecast smoke exposures while properly characterizing uncertainty associated with these estimates. Such metrics will form the basis for determining an optimal methodology for a) calculating the forecast exposure and associated likely health effect using an approach which can cater for positional errors or magnitude errors of the smoke plume; b) for developing a communication protocol for issuing the forecast to health officials in a form which can be readily interpreted and effectively used in the health mitigation decision processes. To demonstrate the utility of this new approach, we will compare forecasted health risk in North Carolina and Sydney with the risk assessment performed in Aim 3.

## MATERIALS AND METHODS

We focus the material and methods section primarily around the work directly related to the four aims of the study. Several additional projects were initiated during the project period, but those not directly respondent to the specific aims are relegated to the publication list in the Appendix.

### Aim 1: Characterize emissions from fire smoke over the last decade

We define daily regional air pollution using monitoring sites for ozone and total PM2.5 collected by Federal Reference Method (FRM), and PM2.5 constituents from the Interagency Monitoring of PROtected Visual Environment (IMPROVE) network. Data were collected for 2006-2013 and the PM2.5 constituents of interest include sulfate, nitrate, potassium, potassium ion, elemental carbon and organic carbon. Satellite image analysis from the NOAA Hazard Mapping System (HMS) is used to determine days on which visible smoke plumes are detected in the vertical column of the monitoring site. We found HMS visible smoke plumes were detected on 6% of ozone (number of stations: n=1595), 8% of FRM (n=1058) and 6% of IMPROVE (n=264) network monitoring days. The plot below shows the monitoring locations and the plume for June 14, 2008.

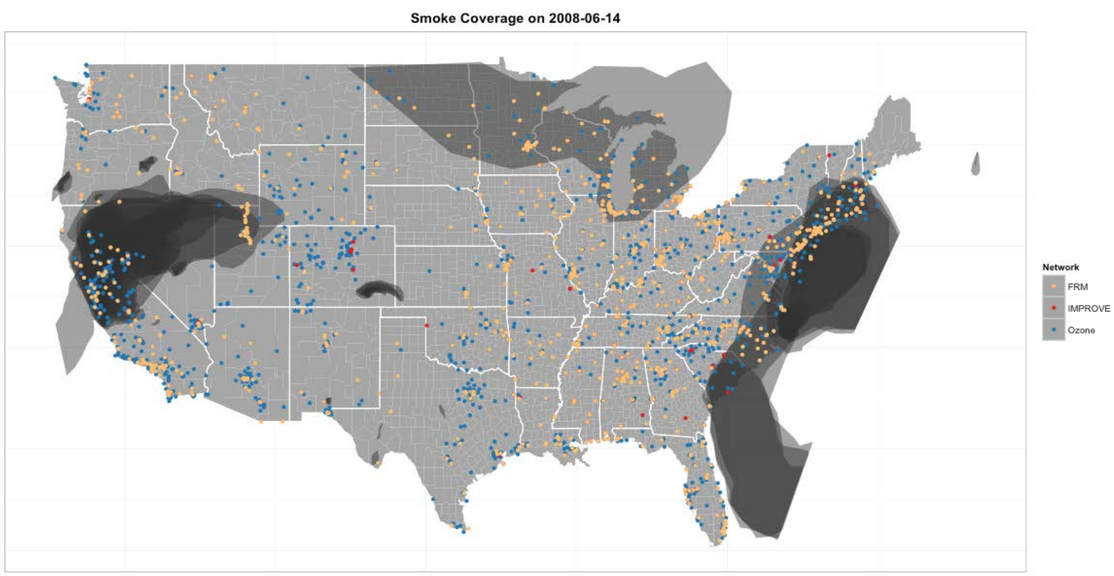


**Figure 1**: PM, ozone and smoke plume data on June 14, 2008.

The statistical analysis of these data has two stages: in the first stage we analyze data from each site separately using a linear regression to determine the effect of plume density on each pollutant; in the second stage we combine these site-specific estimates to determine the overall effect of plume density and spatial variation in this effect. From this final model output we map estimated by contribution of plume density at each site by multiplying the estimate effect by the proportion of days the site is affected by the plume. We also estimate the proportion of extremely high air pollution days that are attributable to smoke plumes.

In a follow up analysis, we use numerical model output and new tools from causal inference to refine estimates of the effect of wildland fire smoke on PM2.5 concentrations. We use the same station-data sources, but replace the HMS data with CMAQ models (https://www.epa.gov/cmaq) run with and without emissions from fires. Figure 2 below maps the percentage PM2.5 attributable to fires based only on these two CMAQ runs (i.e., 100(1-CMAQ without fires/CMAQ with fires)). These CMAQ runs and their difference are statistically calibrated using the station data in a hierarchical Bayesian analysis. The output from this analysis are largely the same as the initial analysis, but the addition of the numerical model runs without the fire emission forcing allows us to make stronger causal claims about the estimated contribution of fire emission to PM and resulting health burden.

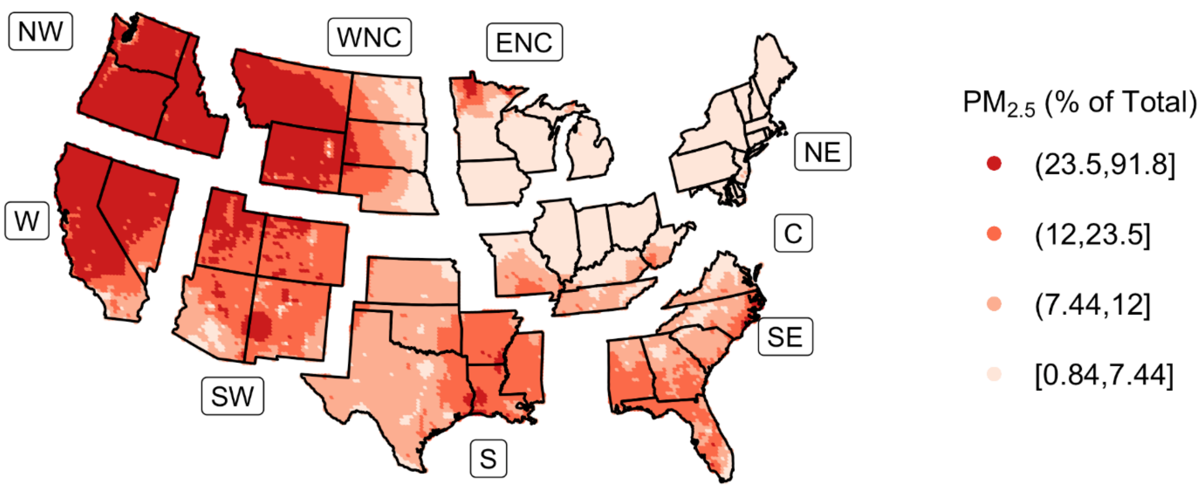


**Figure 2**: Percentage of PM2.5 attributed to fires based only on CMAQ model output

## Aim 2: Develop fire smoke and health risk estimates based on a systematic review

Our goal was to synthesize the state of the wildland fire epidemiological literature and to identify concentration-response parameters appropriate to use in a wildland fire air pollution risk assessment. In 2016, the team performed this systematic review and performed a quantitative meta-analysis. The team found that the number of relevant studies published in the United States and Australia was insufficient to inform a quantitative meta-analysis.

We selected effect coefficients in two stages. First, we identified the effect coefficients previously employed by the U.S. EPA to quantify the avoided premature deaths and illnesses associated with reduced levels of PM2.5 resulting from newly implemented air quality policies (EPA 2012). The Agency preferentially selects those coefficients that correspond to health endpoints identified by the Integrated Science Assessment (ISA) as exhibiting a causal, or likely-to-be-causal, relationship with short- or long-term exposure to fine particles (US EPA 2009). The ISA synthesizes the epidemiological, toxicological and controlled human exposure studies published to that point, and is peer reviewed by the independent Clean Air Scientific Advisory Committee (US EPA, 2009). The human health endpoints include premature death, respiratory and cardiovascular hospital admissions, respiratory emergency department visits, exacerbated asthma, upper and lower respiratory symptoms, acute respiratory symptoms, and work loss days (Table 1).

**Table 1.** Epidemiological Studies Used to Quantify $PM_{2.5}$-Attributable Risks in Primary Analysis

| Endpoint | Study | Study Population | Risk Estimate (95th Percentile Confidence Interval)[A] |
|---|---|---|---|
| **Hospital Admissions** | | | |
| Respiratory hospital admission | Pooled: Henderson et al. (2011), Johnston et al. (2007), Morgan et al. (2010) and Tham et al. (2009) | 0-99 | β=0.00137 (0.00039) |
| | Delfino et al. (2009) | 0-99 | β=0.00276 (0.00067) |
| | Zanobetti et al. (2009) | 0-99 | β=0.00207 (0.000446) |
| Cardiovascular hospital admission | Delfino et al. (2009) | 0-99 | β=0.000768 (0.00048) |
| **Mortality** | | | |
| Time series, all-cause | Zanobetti and Schwartz (2009) | 0-99 | β=0.000975 (0.000119) |

These epidemiological studies may not be well suited to this analysis for three reasons: (1) they do not consider PM2.5 particles originating from wildland fire events specifically; (2) they do not account for the episodic nature of wildland fires; and, (3) they generally consider PM2.5 concentrations at levels significantly below those observed during wildland fire episodes. To overcome these limitations, we employ meta-analysis of epidemiological studies addressing smoke events.

We performed a systematic review and quantitative meta-analysis of epidemiological studies addressing smoke events in the U.S., Australia, South America and Asia. The systematic review employed a machine learning technique to identify relevant literature. The machine learning literature review identified a total of 276 epidemiological studies, of which we judged 21 to be suitable to be included in the quantitative meta-analysis (Figure 3). These 21 studies reported a total of 917 relative risk and odds ratios, the majority of which were for respiratory (n=455) or cardiovascular (n=308) hospital or emergency department visits. These studies reported a smaller number of effect estimates for all-cause or non-accidental deaths (n=139).

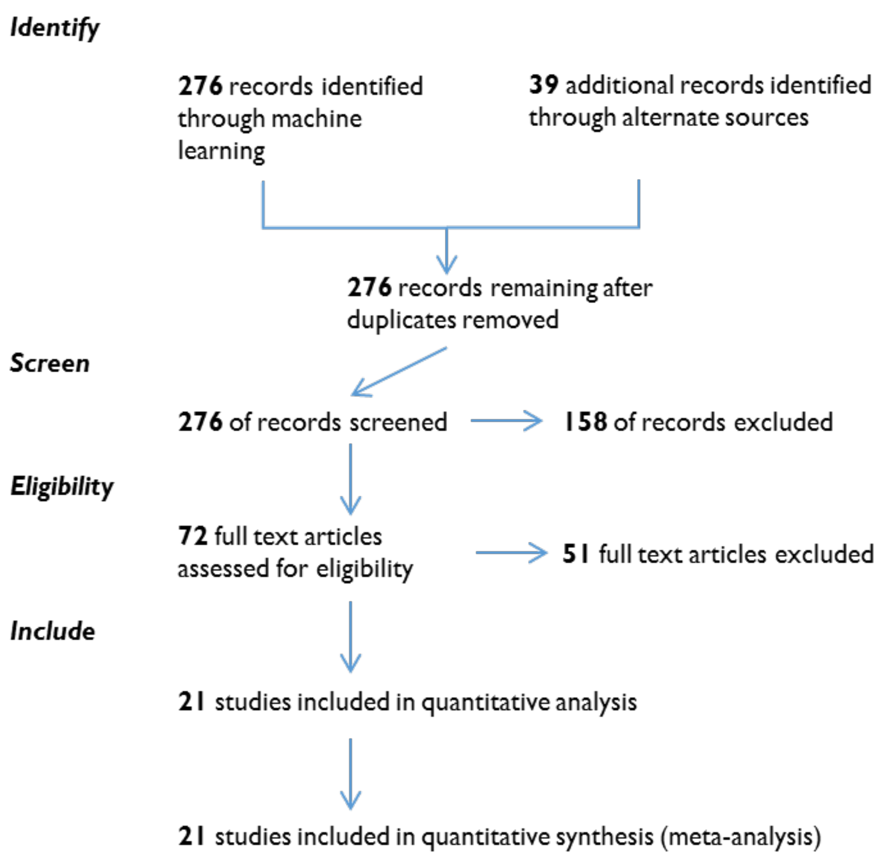


**Figure 3**: PRISMA Diagram

### Aim 3: Perform a health economic impact assessment

In Fann et al. (2018) we characterized the PM2.5-related premature deaths and illnesses in the contiguous United States using chemical transported modeled PM2.5 from wildland fire episodes over a 5-year period beginning in 2008. We apply concentration-response relationships developed both from a systematic review and quantitative meta-analysis of wildland fire epidemiological studies as well as those from epidemiology studies not specifically considering wildland fire episodes. We also report the economic value of these adverse health impacts. The final analysis characterized both the overall magnitude, as well as the distribution, of adverse health impacts among various population subgroups classified by age and race.

We employed a health impact function to quantify the number of excess wildland fire-attributable premature deaths and illnesses in each of five years as follows:

$$\Delta y = y_o\,(e^{\beta \cdot \Delta x}-1) \cdot Pop$$

In this function, β is the risk coefficient reported in an air pollution epidemiological studies or in the meta analysis performed as part of this manuscript. The term $y_0$ is the baseline incidence rate for the health outcome of interest and Δx is annual mean concentration of PM2.5. Finally, Pop is the population for which we estimate health impacts. We use this health impact function in the environmental Benefits Mapping and Analysis Program—Community Edition (BenMAP-CE, v1.1) to estimate counts of PM2.5 attributable deaths in each of five years from 2008 to 2012.

**Aim 4: Develop tools for predicting health impacts in real time**

Working towards this aim we realized that an obstacle was real-time assessment of the current location of smoke plumes emanating from a fire. We propose using a deep learning algorithm for near real-time assessment of smoke in satellite images, trained on output of currently operational methods such as HMS product or analytically derived classification of aerosol. The goal is to supplement current approaches to be able to take advantage of new generation satellite data for public health management and enhancement of forecasting applications. To this end, we implement a Fully Convolutional Network (FCN), a class of deep learning algorithms that are state-of-the-art for speed and accuracy in performing semantic segmentation or pixel-wise classification. Our FCN takes arbitrarily-sized satellite images, learns local features indicative of smoke plumes and conducts pixel-wise binary classification (presence or absence of smoke). Our algorithm is trained to automatically detect smoke plumes in satellite imagery using Himawari-8 products. The algorithm we develop is applicable to any geographic region where some form of ground truth can be obtained. It also has the potential to operational in near-real time, allowing the algorithm to flag the presence of fire smoke for use in public health messaging. This new information can be used by fire managers, health and environmental agencies and the general public to better manage the health risks for fire smoke.

Substantial progress has been made towards performing real-time bias-correction of chemical transport model (CTM) smoke predictions. We are currently making use of fixed-site monitors and the addition of AOD from the geostationary Himawari-8 satellite. We are combining these data using a hierarchical Bayesian spatiotemporal model fit using Markov Chain Monte Carlo sampling. The model includes CTM and AOD as predictors of the monitor data, and exploits spatiotemporal correlation in the residuals to improve forecast. Prediction improvement is currently being evaluated using two Sydney case studies. The model is sufficiently fast to be run in real time and R code is publically available.

Using these forecasts, an online Health Impact Assessment (HIA) tool for Australia is currently under development. The tool is preconfigured with Australian health incidence, population, and economic data and recommended source-specific risk estimates. Users are able to upload pollutant scenarios and calculate attributable health impacts and associated economic costs by small spatial units. The tool will allow a user to download the results as a spreadsheet or a dynamic report. In addition to the graphical interface, the tool can also be accessed programmatically, this will facilitate integration into existing infrastructure.

## RESULTS AND DISCUSSIONS

### Aim 1: Characterize emissions from fire smoke over the last decade

Our analysis of station data and HMS smoke plume data indicate unhealthy levels of O3 and PM2.5 were, respectively, 3.3 and 2.5 times more likely to occur on plume days than on clear days. Plumes were associated with an average increase of 2.6 p.p.b. (2.5, 2.7) in O3 and 2.9 µg/m3 (2.8, 3.0) in PM2.5 nationwide, but the magnitude of effects varied by location. The largest impacts were observed across the southeast. High impacts on O3 were also observed in densely populated urban areas at large distance from the fires throughout the southeast. Fire smoke substantially affects regional air quality and accounts for a disproportionate number of unhealthy days. This work appears in Larsen et al (2018).

Next we turn to the follow up analysis that replaced HMS plume indicators with CMAQ model runs with and without the fire emissions forcing. Regressing the station data onto the difference between these CMAQ runs we found that the CMAQ runs tended to overestimate the amount of PM2.5 that resulted from fires (e.g., comparing Figure 2 with the figure below). After adjusting for this bias, the estimated percentage of PM2.5 that is attributable to fire smoke exceeds 30% in large areas of the West and Northwest.

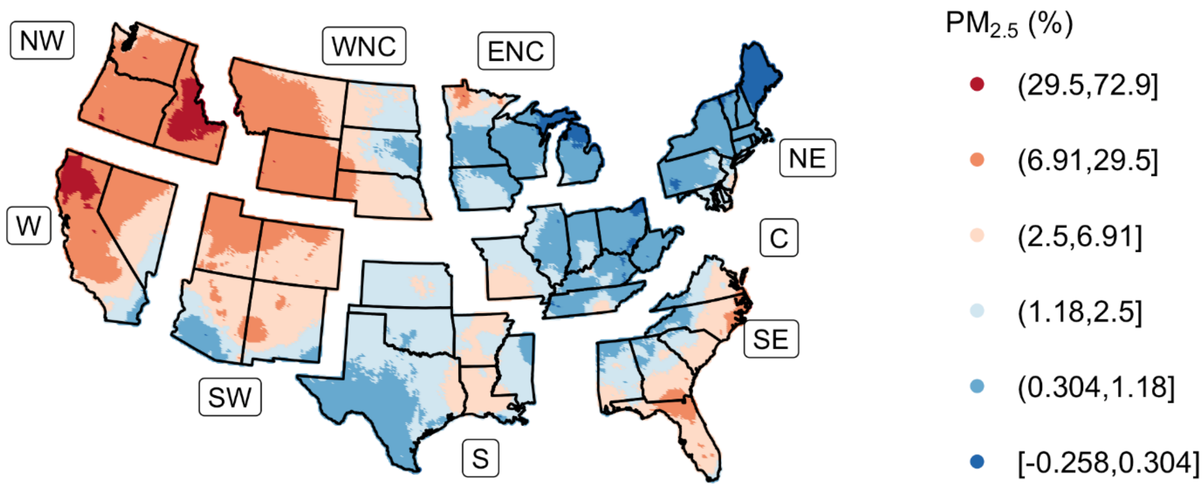


**Figure 4**: Bayesian estimate of fire-contributed PM2.5 as percentage of the total.

After estimating the causal effect of fire smoke onto PM2.5 concentrations, we applied health effect estimates to estimate the total health burden for fire smoke. The figure below shows the estimated cumulative number of excess hospital admissions our analysis attributed to fire smoke over the fire seasons between 2008 and 2012. Not surprisingly, the health burden is the highest in highly-populated areas in the Western US.

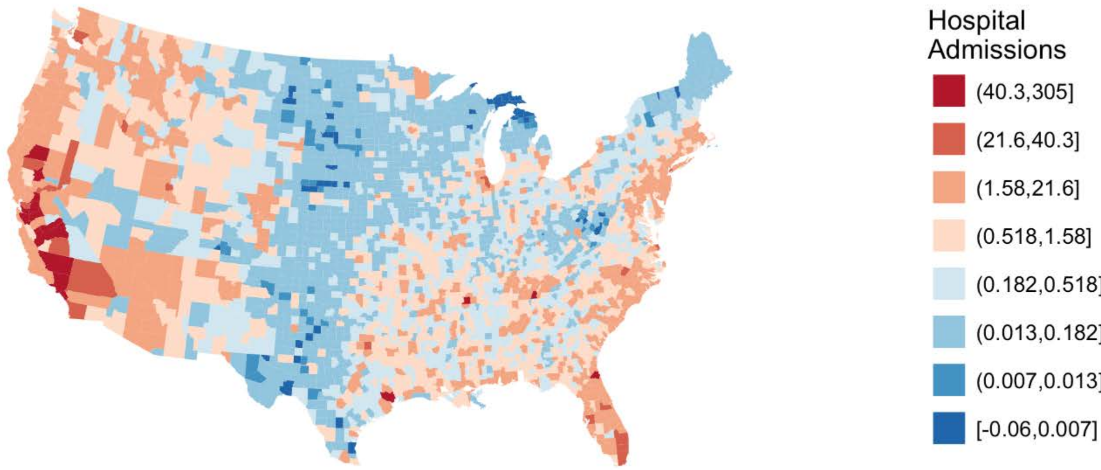


**Figure 5:** Distribution of cumulative health burden by county. For each county, we aggregated the number of hospitalizations for respiratory illness across all age groups related to fire-contributed PM2:5 (the Bayesian estimate). The map displays the number of hospital admissions estimated over the 2008-2012 wildfire seasons.

### Aim 2: Develop fire smoke and health risk estimates based on a systematic review

We determined that 4 studies reported risk estimates that were suitable for pooling in a quantitative meta-analysis. Using the Metafor library in the R statistical package, we perform a random effects meta-analysis. A limitation of this meta-analysis is that it incorporates studies of wildland fire events outside of the U.S.; using the pooled estimate may then introduce uncertainties that we discuss further below.

### Aim 3: Perform a health economic impact assessment

Nearly all states in the U.S. experienced elevated fine particle concentrations from wildland fire events over the 5-year period. Certain states were affected by severe wildland fire events occurring across two or more years, including Louisiana (some of which may be due to debris burning after an active hurricane season), California, Idaho, and Georgia (Figure 6). The maximum predicted annual mean wildland fire-attributable PM2.5 concentration is about 42 µg/m3 (Table 1). The median value ranges from between 0.3 µg/m3, in 2009 and 0.8 µg/m3 in 2012, while the population-weighted annual mean PM2.5 concentration ranges from between 0.6 µg/m3 in 2009 to 1.1 µg/m3 in 2008 (Table 2). In the western U.S, states including California, Oregon and Idaho and Montana see among the greatest impacts from wildland fires. Southeastern states including North Carolina, South Carolina, Georgia, Louisiana, Arkansas, Florida and eastern Texas are most affected.

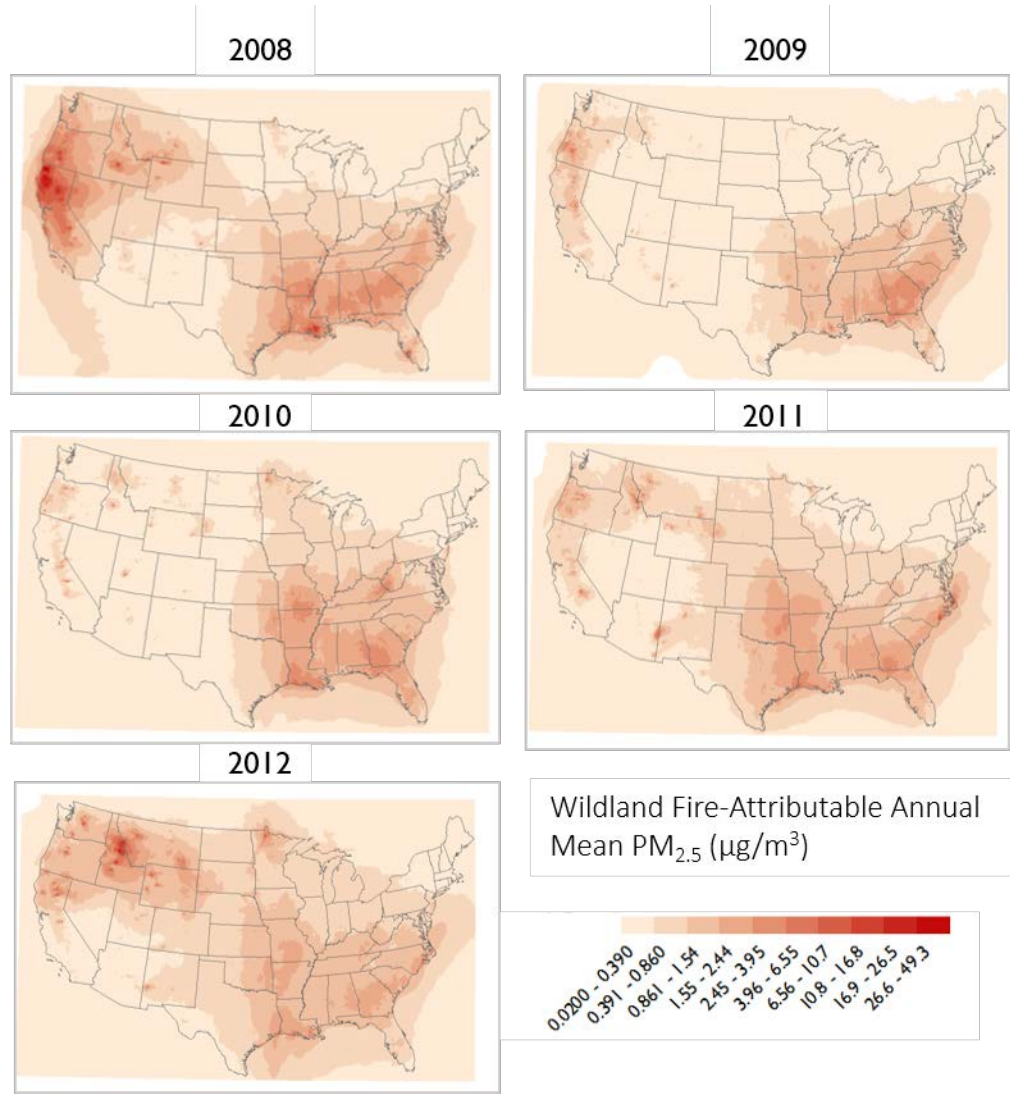


**Figure 6**. Annual mean wildfire-attributable PMR2.5 concentrations (2008-2012)

When quantifying wildland fire-related impacts using concentration-response relationships from epidemiological studies of wildland fire events, we find that the number of deaths and illnesses is significantly smaller. We estimate between about 1,500 and 8,000 excess deaths and between 5,000 and 6,500 hospital admissions per year (Table 3). We estimate the largest number of excess deaths, hospital admissions and other impacts for the year 2008, when wildland fire attributable PM2.5 concentrations are the greatest. Summing the economic value of the premature deaths and hospital admissions we estimate a total dollar value of between $12B and $20B per year (2016$). The present value of these economic values across the 5-year period is $65B (3% discount rate, 2016$).

**Table 2.** Premature deaths and illnesses attributable to wildfire-related $PM_{2.5}$ concentrations in each year calculated using alternative concentration-response functions (95% confidence intervals)

| | | | Year | | |
|---|---|---|---|---|---|
| **Endpoint** | *2008* | *2009* | *2010* | *2011* | *2012* |

| | | | | | |
|---|---|---|---|---|---|
| Alternative Estimates of Respiratory Hospital Admissions | | | | | |
| Delfino et al. (2009) | 8,500 (4,400—12,000) | 5,200 (2,700—7,700) | 6,200 (3,200—9,100) | 6,300 (3,300—9,300) | 6,400 (3,300—9,400) |
| Zanobetti et al. (2009) | 6,300 (3,600—9,000) | 3,900 (2,300—5,500) | 4,600 (2,600—6,500) | 4,700 (2,700—6,700) | 4,800 (2,800—6,800) |
| Alternative Estimates of Long-Term PM2.5-Related Premature Deaths | | | | | |
| Krewski et al. (2009) | 14,000 (9,700—19,000) | 8,700 (5,800—11,000) | 10,000 (6,900—14,000) | 11,000 (7,300—14,000) | 11,000 (7,600—15,000) |
| Lepeule et al. (2012) | 32,000 (16,000—48,000) | 19,000 (9,800—29,000) | 23,000 (12,000—35,000) | 24,000 (12,000—36,000) | 25,000 (13,000—38,000) |

[A]*Values rounded to two significant figures; all functions estimated for populations ages 0-99*
[B]*Pooled estimate of 4 $PM_{10}$ studies*

**Table 3**. Estimated economic value of wildfire-attributable $PM_{2.5}$-related premature deaths and respiratory hospital admissions (2008 to 2012) (Billions of 2010$)[A].

| Year | | | | | |
|---|---|---|---|---|---|
| *2008* | *2009* | *2010* | *2011* | *2012* | **Present Value** |
| $122 ($11—$330) | $68 ($6.3—$190) | $82 ($7.7—$220) | $68 ($6.4—$190) | $70 ($6.6—$190) | $380 ($35—$1,000) |

[A] Sum of Delfino et al. (2009) hospital admission estimates and Krewski (2009) long-term mortality.

We estimate the cumulative number of wildland fire-attributable PM2.5-related deaths and respiratory hospital admissions in these highly affected areas as well as all other areas over the 5-year period, and then quantified the percentage of total deaths and hospital admissions occurring among populations of each race. We found that, when comparing the share of wildland fire-attributable deaths estimated to occur in the highly affected and less affected areas among population subgroups, that: (1) black populations represent a larger share of the fraction of deaths in the highly affected areas and a smaller share in the less affected areas; (2) white populations experience a smaller share of deaths in the highly affected areas and a larger share in the less affected areas.

**Aim 4: Develop tools for predicting health impacts in real time**

In Larsen (2019+) we train the model that uses satellite images to identify smoke plumes using stochastic gradient descent and a binary cross entropy loss function commonly used to optimize semantic segmentation algorithms. The learned parameters are used to make predictions on the out-of-sample test set (~30% of the data) and the model is evaluated by the average loss. We repeat the cycle of training and testing until the average loss converges (about 20 cycles). Upon convergence, we report the performance metrics, pixel accuracy and intersection over union (IoU), as shown for one day in Figure 7. Across all days, our classification accuracy is over 95%, therefore we believe this method will prove to be useful to generate data that can be feed into the short-term forecasting model.

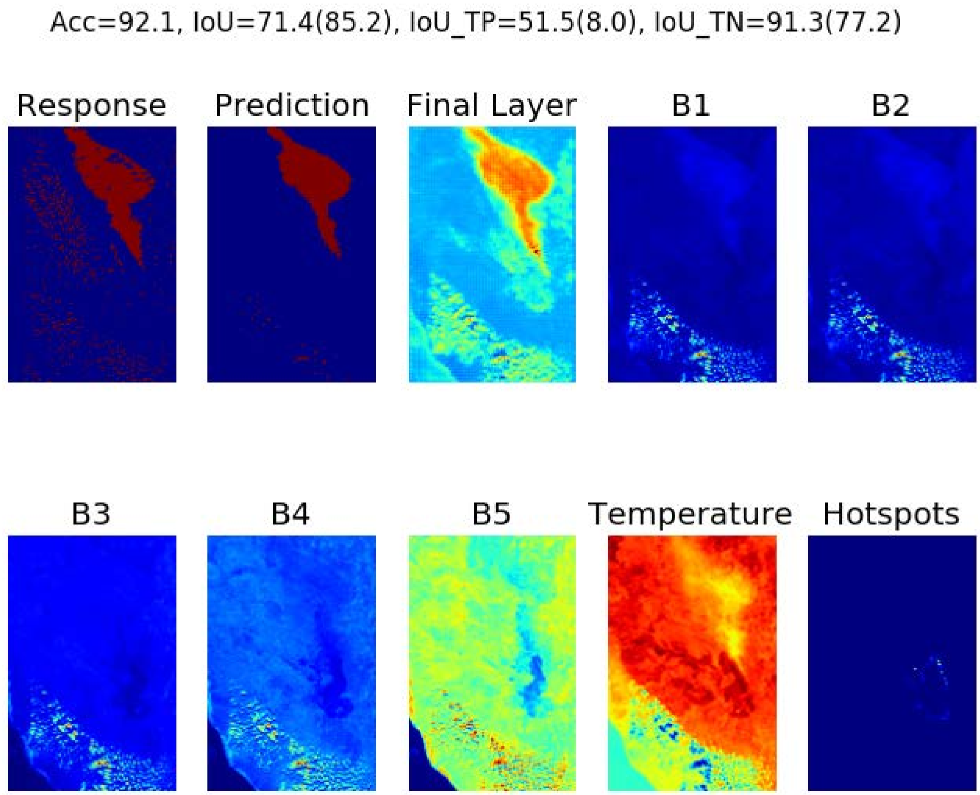


**Figure 7**. Performance of the smoke classifier algorithm. The top left plot is the true locations of the plume (red), the adjacent plot is the prediction from the machine learning algorithm, the ``final layer'' is the machine learning output before thresholding, and the remaining plots are images used as covariates in the predictive model. The metrics reported are as follows: Acc=percent of correctly classified pixels; IoU=Intersection over Union (%); IoU_TP=IoU for true positives; IoU_TN= IoU for true negatives. Values in parenthesis are the weighted versions of each metric.

We are nearing completion of the real-time forecasting tool. Figure 8 provides a screenshot of the current version. A journal paper describing this tool is under preparation for submission in 2019 and the tool will be made available as an R package and R Shiny web app. There has been significant interest from government in the development of this tool. Presentations and demonstrations have been given to state government departments from NSW and Victoria and to the national Bureau of Meteorology. We met with NSW government departments to demonstrate online Health Impact Assessment tool on December 17, 2018. Departments included NSW Health, NSW Office of Environment and Heritage, NSW EPA, and NSW Rural Fire Service (the agenda appears in the appendix). We also met Victorian government departments and the national Bureau of Meteorology to demonstrate the tool to on July 18, 2018. Local governments have also expressed interest. We also developed a related methodology and implementation for the NSW Office of Environment and Heritage (OEH).

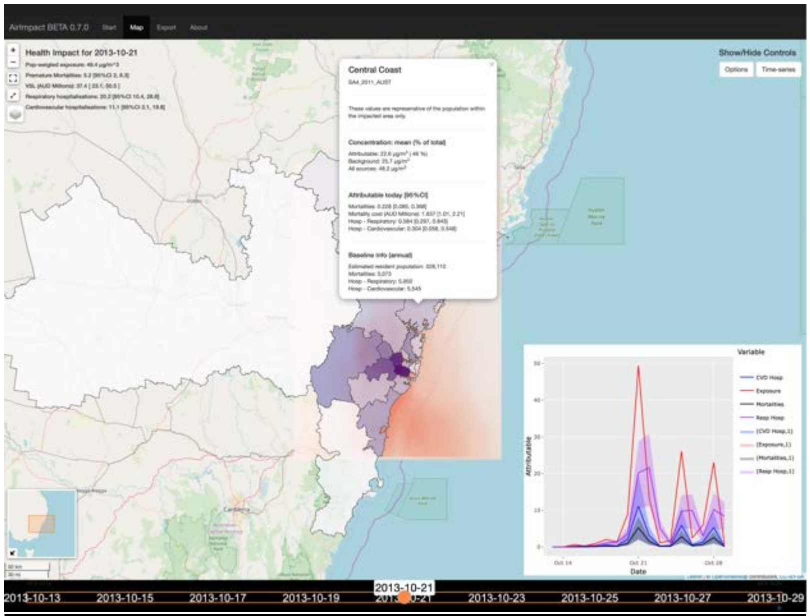


**Figure 8**. Screenshot from the online Health Impact Assessment (HIA) tool for Australia.

# CONCLUSIONS

Joint Fire Science Program proposal titled ``Estimating fire smoke related health burden and novel tools to manage impacts in urban populations'' has been an extraordinarily productive research effort to characterize air quality impacts of wildland fires across U.S. over an extended period of time, to characterize health burden related to those impacts and to develop two prototype tools for managing smoke related public health response. Among the air pollutants found in smoke (direct or long-range transport) fine particulate matter and ozone are of highest concern to public health due to the amount of exposure as well as the size of affected population. We estimated that unhealthy levels of ozone and fine particulate matter were 3.3 and 2.5 times more likely to occur on days when smoke plumes are detected. As such long-range fire smoke may be characterized as one of the largest sources of unhealthy air quality in U.S. today.

In working toward accomplishing proposed aims we identified several knowledge gaps. Whereas epidemiologic evidence has consistently demonstrated health impacts of smoke exposure we identified lack of consistent exposure metrics which would unable quantitative meta-analysis and comparison of health impacts relative to other sources of air pollution.  We also recognized that currently smoke forecasting tools lack the precision needed to accurately predict air quality surface and consequently health impacts in real time on the spatial scale relevant to public health. Low number of monitors and heterogeneity between monitoring instruments limits the ability to evaluate model performance even under best circumstances. Different monitors have different performance with respect to variation and bias.

The cross disciplinary approach and wide range of experience on the research team provided an opportunity to address aspects of these knowledge gaps. Most notably, we quantified health burden attributed to wildland fires in terms of excess hospitalizations and deaths as well as monetary value of the burden in both U.S. and AU. We developed a number of statistical downscaling techniques to learn the relationship between model-forecasted and observed air quality measurements and improve accuracy of short term forecasting. With these methods we also develop health burden forecasting tools.  Finally, during the granting period a new generation satellites became available which provide real-time data products but have not been linked to smoke detection and health burden. We developed an artificial intelligence framework for new generation satellite data products which can be used both to communicate health impacts and calibrate smoke models. The proposed tools have a real potential to be scaled into real time tools leading to a valuable tool for policy makers.

Engagement in JFSP grant related research provided a number of training opportunities for graduate students. Alexandra Larsen (PhD, 2018, NC State) quantified the long-range transport of wildland fire smoke, developed a novel statistical framework to estimate causal effect of fires on ambient air pollution and calculated related health burdens and she developed an artificial intelligence framework for identify smoke plumes directly from satellite-based true color imagery (i.e. the spectrum visible to the human eye).  This smoke plume data will be used in conjunction with near real time modelled and measured particulate data to provide an integrated picture of smoke exposure across an airshed. This new approach can be used by fire managers, health and environmental agencies and the public to better manage the health risks for fire smoke. Joshua Horsely (PhD student, University of Sydney) developed a data fusion model that combines numerical model output with station data to forecast air pollution, and created a public website for users to interact with his methods. Suman Majumder (PhD candidate, NC State) developed a novel statistical downscaling with spatial misalignment to accurately forecast air quality impacts and Arnab Hazra (PhD, 2018, NC State) developed an ensemble approach for statistical downscaling to calibrate forecasts of fine particulate matter in smoke against ground monitors and thus forecast the health burden in real time.

Our research efforts have been recognized in both the research community but also more broadly. Joshua Horsely a PhD student at University of Sydney, Australia who has been engaged in JFSP research won the First Place: Spark Award: Best Student Presentation, First Place at the 2nd International Smoke Symposium in Long Beach California for presentation titled ``Forecasting fire smoke exposure and health impacts in Australia''. Alexandra Larsen, a PhD student at North Carolina State University received the best poster of the day award at CMAS 2015 Conference and the best presentation award for the American Statistical Association's Conference on Environmental Statistics (2018). Larsen et al (2017) was selected by the Journal of Nature for press release and was picked up by 10 major news channels. Fann et al. was mentioned by 9 news outlets including The Washington Post, Herald Tribune and the Wired magazine. Broome et al. 2017 was picked up by 5 news outlets and a policy source. Finally, we anticipate that machine learning tool for smoke detection may be tested in operational setting at CISRO, AU.

# LITERATURE CITED


Bowman DMJS, Balch JK, Artaxo P, et al (2009). Fire in the Earth System. *Science,* **324,** 481 - 484.

Delfino RJ, Brummel S, Wu J, et al (2009). The relationship of respiratory and cardiovascular hospital admissions to the southern California wildfires of 2003. *Occupational and Environmental Medicine,* **66**, 189.

Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJ, Comparative Risk Assessment Collaborating G (2002). Selected major risk factors and global and regional burden of disease. *Lancet*, **360**, 1347-1360.

Fann N, Alman B, Broome RA, Morgan GG, Johnston FH, Pouliot G, Rappold AG (2018). The health impacts and economic value of wildland fire episodes in the US: 2008–2012. *Science of the Total Environment*, **610**, 802-9.

Henderson SB, Brauer M, Macnab YC, Kennedy SM (2011). Three measures of forest fire smoke exposure and their associations with respiratory and cardiovascular health outcomes in a population-based cohort. *Environmental Health Perspectives*, **119**, 1266-1271.

Johnston F, Hanigan I, Henderson S, Morgan G, Bowman D (2011). Extreme air pollution events from bushfires and dust storms and their association with mortality in Sydney, Australia 1994-2007. *Environmental Research*, **111**, 811-816.

Larsen AE, Reich BJ, Ruminski M, Rappold AG (2018). Impacts of fire smoke plumes on regional air quality, 2006-2013. *Journal of Exposure Science and Environmental Epidemiology*, **28**, 319-27.

Larsen AE, Yang S, Reich BJ, Rappold AG (2019+). A spatial causal analysis of wildland fire-contributed PM2:5 using numerical model output. Submitted to the *Journal of the American Statistical Association*.

Larsen AE, Reich BJ, Rappold AG (2019+). A deep learning approach to identify smoke plumes in satellite imagery in near real-time for public health management. In preparation.

Martin KL, Hanigan IC, Morgan GG, Henderson SB, Johnston FH (2013). Air pollution from bushfires and their association with hospital admissions in Sydney, Newcastle and Wollongong, *Australian and New Zealand Journal of Public Health*, **37**, 238-243.

Morgan G, Sheppeard V, Khalaj B, et al (2010). The effects of bushfire smoke on daily mortality and hospital admissions in Sydney, Australia, 1994 to 2002. *Epidemiology*, **21**, 47-55.

Pope CA, Dockery DW (2006). Health effects of fine particulate air pollution: lines that connect. *Journal of the Air & Waste Management Association*, **56**, 709-742.

Sastry N (2002). Forest fires, air pollution, and mortality in Southeast Asia. *Demography*, **39**,1-23.

Tham R, Erbas B, Akram M, Dennekamp M, Abramson MJ (2009). The impact of smoke on respiratory hospital outcomes during the 2002–2003 bushfire season, Victoria, Australia. *Respirology*, **14**, 69–75.

US EPA (2009) Integrated Science Assessment (ISA) for Particulate Matter (Final). U.S.

Environmental Protection Agency, Washington, DC.

US EPA (2012). Regulatory Impact Assessment for the PM NAAQS RIA. U.S. Environmental Protection Agency, Washington, DC.

US EPA. (2013) Integrated Science Assessment for Particulate Matter. U.S. Environmental Protection Agency, Washington, DC.

Zanobetti A, Schwartz J (2009). The effect of fine and coarse particulate air pollution on mortality: a national analysis. *Environmental Health Perspectives*, **117**, 898–903.

Zanobetti A, Franklin M, Koutrakis P, Schwartz J (2009). Fine particulate air pollution and its components in association with cause-specific emergency admissions. *Environmental Health*, 8, 58.

## APPENDIX A: CONTACT INFORMATION FOR KEY PERSONNEL

Brian J Reich
North Carolina State University
Raleigh, NC USA
brian_reich@ncsu.edu

Ana G Rappold
US Environmental Protection Agency
Research Triangle Park, NC USA
Rappold.Ana@epa.gov

Fay H Johnston
University of Tasmania
Tasmania, Australia
Fay.Johnston@utas.edu.au

Geoffrey G Morgan
University of Sydney
New South Wales, Australia
geoffrey.morgan@sydney.edu.au

Neal L Fann
US Environmental Protection Agency
Research Triangle Park, NC USA
Fann.Neal@epa.gov

Martin E Cope
Commonwealth Scientific and Industrial Research Organisation
Aspendale, VIC, Australia
Martin.Cope@csiro.au

Richard A Broome
Sydney South West Area Health Service
New South Wales, Australia
rbroo@doh.health.nsw.gov.au

# APPENDIX B: COMPLETED/PLANNED SCIENTIFIC/TECHNICAL PUBLICATIONS/SCIENCE DELIVERY PRODUCTS

## Articles in peer-reviewed journals


Berrocal VJ, Guan Y, Muyskens A, Wang H, Reich BJ, Chang HH (2019+). A comparison of statistical and machine learning methods for creating national daily maps of ambient PM concentration. Submitted to *Environmental Health Perspectives*.

Broome RA, Johnston FH, Horsley J, Morgan GG (2016). A rapid assessment of the impact of hazard reduction burning around Sydney, May 2016. *The Medical Journal of Australia*, **205**, 407-8.

Fann N, Alman B, Broome RA, Morgan GG, Johnston FH, Pouliot G, Rappold AG (2018). The health impacts and economic value of wildland fire episodes in the US: 2008–2012. *Science of the Total Environment*, **610**, 802-9.

Hanigan IC, Morgan GG, Williamson GJ, Salimi F, Henderson SB, Turner MR, Bowman DMJS, Johnston F (2018). Extensible database of validated biomass smoke events for health research. *Fire*, **1**, 50.

Hazra A, Reich BJ, Shaby BA, Staicu AM (2019+). A semiparametric Bayesian model for spatiotemporal extremes. In revision, *Journal of the American Statistical Association*.

Horsley JA, Broome RA, Johnston FH, Cope M, Morgan GG (2018). Health burden associated with fire smoke in Sydney, 2001-2013. *Medical Journal of Australia*, **208**, 309-10.

Huang YN, Reich BJ, Fuentes M, Sankarasubramanian A (2019+). Complete spatial model calibration. Accepted, *Annals of Applied Statistics*.

Johnston FH (2017). Bushfires and planned burns: Tips for your patients in managing smoke. *Respiratory Medicine Today*, **2**, 2.

Johnston FH (2017). Understanding and managing the health impacts of poor air quality from landscape fires. *Medical Journal of Australia*, 207, 229-230.

Larsen AE, Reich BJ, Ruminski M, Rappold AG (2018). Impacts of fire smoke plumes on regional air quality, 2006-2013. *Journal of Exposure Science and Environmental Epidemiology*, **28**, 319-27.

Larsen AE, Yang S, Reich BJ, Rappold AG (2019+). A spatial causal analysis of wildland fire-contributed PM2:5 using numerical model output. Submitted to the *Journal of the American Statistical Association*.

Larsen AE, Reich BJ, Rappold AG (2019+). A deep learning approach to identify smoke plumes in satellite imagery in near real-time for public health management. In preparation.

Majumder S, Reich BJ, Rappold AG (2019+). Statistical downscaling with spatial misalignment: Application to wildland fire PM emissions forecasting. To be submitted to the *Annals of Applied Statistics*.

Morris SA, Reich BJ, Thibaud E, Cooley DA (2017). A space-time skew-t model for

threshold exceedances. *Biometrics*, **73**, 749-778.

Salimi F, Henderson SB, Morgan GG, Jalaludin B, Johnston JH (2017). Ambient particulate matter, landscape fire smoke, and emergency ambulance dispatches in Sydney, Australia. *Environment International*, **99**, 208-212

## Graduate theses (masters or doctoral)

Alexandria Larsen, 2018, PhD, Department of Statistics, North Carolina State University

Arnab Hazra, 2018, PhD, Department of Statistics, North Carolina State University

Joshua Horsely, 2019+, PhD, Department of Epidemiology, University of Sydney

Suman Mujumder, 2019+, PhD, Department of Statistics, North Carolina State University

## Conference or symposium presentations

Larsen (2018), ASA/ENVR Workshop, Asheville, NC.

Larsen (2018), Joint Annual Meeting of the International Society of Exposure Science (ISES) and the International Society for Environmental Epidemiology (ISEE), Ottowa, CA.

Larsen (2018), Joint Statistical Meetings, Vancouver, CA.

Larsen (2018), International Chinese Statistical Association Meeting, New Brunswick, NJ.

Horsley (2018), Research Meeting of the Centre for Air pollution, Energy, and Health Research, Hobart, Tasmania.

Horsley (2018), Scientific Meeting of the Australasian College of Toxicology & Risk Assessment, Perth, AY.

Horsley (2018), Joint Annual Meeting of the International Society of Exposure Science and the International Society for Environmental Epidemiology, Ottawa, CA.

Reich (2018), Joint Statistical Meetings, Vancouver, BC.

Reich (2017), NASA Jet Propulsion Lab, Pasadena, CA, 2017.

Horsley (2017), Clean Air Society of Australia and New Zealand (CASANZ) Conference, Brisbane, AU.

Morgan (2017), Scientific Meeting of the Australasian College of Toxicology & Risk Assessment, Canberra, AU.

Horsley (2017), International Society for Environmental Epidemiology Conference, Sydney, AU.

Horsley (2017), University of Sydney School of Public Health Research Showcase, Sydney, AU.

Cope (2017), Interdisciplinary Biomass Burning Initiative Workshop, Boulder, CO.

Broome (2017), NSW Environment Protection Authority Clean Air Summit, Sydney, AU.

Morgan (2017), University of Sydney, Sydney, AU.

Horsley (2016), International Smoke Symposium, Long Beach, CA.

Morgan (2015), Clean Air Society of Australia and New Zealand Conference, Melbourne, AU.

## Posters

Larsen (2018), NC State Graduate Research Symposium, Raleigh, NC

Larsen (2017), Joint Statistical Meetings, Baltimore, MD

Larsen (2017), Atlantic Causal Inference Conference, Chapel Hill, NC

Larsen (2016), International Fire Behavior & Fuels Conference, Portland, OR

Larsen (2015), Community Modeling & Analysis System Conference, Chapel Hill, NC

## Workshop materials and outcome reports

Morgan, Broome and Horsley presented a 3-day workshop on Health Impact Assessment of Haze for the Malaysian Ministry of Health, 26-29 July 2017, Malacca, Malaysia. Also, as part of the development of the short-term forecasting tools we organized two workshop to foster collaboration between academic and government partners. The agendas are below.

### Collaboration to Enhance Emissions Inventory Development

**Event:** 3 day emissions inventory workshop including half day session on exposure modelling and health
**Date:** Tues 10 to Thurs 12 April 2018
**Location:** Sydney CBD
**Participants:** State and territory staff and academic science partners
**Workshop objectives and outcomes:**

- Establish a shared understanding of the priorities, status and future plans for jurisdictional emission inventories, including: end user needs and priorities; current status and critical gaps; frameworks for inventory development; current projects and future plans (funded and proposed)
- Identify shared priorities, opportunities for collaboration and key components of the framework needed to support work towards nationally consistent, up-to-date and comprehensive emissions inventories.
- Establish working arrangements and agree next steps to progress collaborative actions for coordinated emissions inventory development.
- Identify population exposure assessment and health related priorities and opportunities for collaboration

**Agenda:**

| Day 1 | Emissions Inventory | Who |
|---|---|---|
| 10am | Welcome & introductions<br>Workshop objectives and outcomes | Jurisdiction heads and Facilitator<br>CAR representative |
| 10-12.30pm | Information sharing by jurisdictions and research partners on emissions inventories:<br>• end user needs and priorities including exposure assessment and health<br>• current status and critical gaps<br>• frameworks for inventory development<br>• current projects and future plans (funded and proposed) | Jurisdictional rep presentations; including invited policy and health rep presentations |
| 12.30-1.30pm | Lunch | |

| | | |
|---|---|---|
| **1.30-3pm** | Towards national consistency in emissions inventories<br>• Primary components of consistent inventories<br>• Priorities for exposure assessment and health<br>• Framework needed to support delivery | Facilitated discussion |
| 3-5pm | Focus sessions on major sources<br>• Residential wood heating<br>• Vehicles | Jurisdictional staff & invited science partner presentations |
| 5-5.30pm | Day 1 wrap up:<br>• end user priorities, critical gaps and opportunities for collaboration<br>• elements of a framework to support national consistency<br>• exposure assessment and health perspective | |
| **Day 2** | **Emissions Inventory** | Who |
| 9.30am | Day 2 objectives and outcomes | |
| 9.30-11.30am | Focus sessions on major sources, continued<br>• Major industrial, including power generation<br>• Off-road engines, shipping, rail, aircraft | Jurisdictional staff & science partner presentations |
| 11.30 – 12am | Summary of priorities and opportunities identified for collaboration during Day 2 morning session | Facilitator |
| **12-1pm** | Lunch | |
| **1 – 3.00pm** | Priorities, conceptual framework and collaborative actions going forward:<br>• Major priorities (including exposure assessment and health)<br>• Governance and work arrangements<br>• Framework to support nationally consistent, up-to-date inventories, and data and method development and sharing<br>• Source category building blocks/modules for initial focus<br>• Inventory years – base case and future projection years<br>• Collaborative approaches for focus source sector<br>• Funding opportunities | Facilitated discussion |
| **3.00-4.30pm** | **Proposal for funding under NCAA to support a national emissions inventory**<br><br>Agreeing the business case (based on earlier discussions):<br>• What are the critical knowledge, data and information gaps and how do they affect our ability to manage air quality and reduce impacts?<br>Purpose, scope, method and deliverables:<br>• What is the purpose, scope and outputs of the project?<br>• What resources (people, funding) and timeframes will be required?<br>• How should the project be implemented?<br>Do we seek to develop (a) and (b) or just (b) through the NCAA priority setting process?<br>a) nationally consistent emissions inventory approach for jurisdictional inventories<br>b) national emissions inventory (NEI) | Facilitated discussion |
| **4.30-5pm** | Day 2 Wrap up and Next Steps | Facilitator |

| Day 3 (morning) | Emissions Inventory | Who |
|---|---|---|
| 9.30am | Day 3 objectives and outcomes | |
| 9-11.30am | Focus sessions on other sources<br>• Landscape fires (planned burns, wildfires, agricultural)<br>• Domestic/commercial<br>• Small industrial (non-NPI)<br>• Agricultural (other)<br>• Wind blow dust (regional)<br>• Biogenic<br>• Sea salt<br>• Future sources including potential exposure and health concerns | Jurisdictional staff & invited science partner presentations |
| 11.30 – 12am | Summary of priorities and opportunities identified for collaboration during Day 3 morning session<br>CAR – health and exposure assessment perspective | Facilitator |
| **12-1pm** | Lunch | |

## Air Pollution Exposure Assessment Approaches for Australia

**Event:** 0.5 day session 12 April 2018, 1-5pm
**Participants:** State and territory staff and science partners - Co-hosted by Centre for Air pollution, energy and health Research (CAR)
**Workshop objectives and outcomes:**

- Share information on air pollution exposure assessment approaches which have been developed and applied in Australia, including data input requirements,
- Identify opportunities for collaboration to improve exposure modelling and assessment for health.

| Day 3 (afternoon) | Exposure Modelling and Assessment | Who |
|---|---|---|
| 1-1.15pm | Introductions<br>Workshop objectives and outcomes | Facilitator |
| **1.15 – 4.15pm** | Exposure Modelling Approaches:<br>Speakers to be determined – but could include:<br>• Richard Broome / Martin Cope –Health gains from reductions in air pollution using scenario based CTM models<br>• Luke Knibbs / Chris Cowie – National and local Land Use Regression Models<br>• Ivan Hanigan / Martin Cope – combining data and models for improved population exposure assessment<br>• Prof Fabio Ramos (Univ Sydney Centre for Translational data science) – Spatio-temporal exposure modelling<br>• John Innis – Population exposure assessment in Tasmania<br>• Yvonne Scorgie – Population exposure assessment methods applied in NSW to support policy scenario analysis | Invited presentations |
| **4:15-4.45** | Opportunities for collaboration | Facilitated discussion |
| 4.45-5pm | Wrap up | Facilitator |

# APPENDIX C: METADATA

We have made the main datasets from the project publically available.

(1) The datasets and code from Alex Larsen's papers

Larsen et al (2018). Impacts of fire smoke plumes on regional air quality, 2006-2013. JESEE.

Larsen et al (2019+). A spatial causal analysis of wildland fire-contributed PM2.5 using numerical model output. In revision, JASA.

are posted for the public at

https://drive.google.com/drive/folders/1KIa99dCV6OhHhnD8bwoDjhjaiC0IQP17

These files are described in detail in section (i)-(viii) below.

(2) The team has also published a paper on a dataset that includes extreme air pollution events attributable to fire events. The paper is

Hanigan, Ivan C., et al. (2018). Extensible database of validated biomass smoke events for health research. Fire 1.3: 50.

and the abstract is given in section (ix) below. The data are freely available on

https://github.com/swish-climate-impact-assessment/biomass_smoke_events_db

and an R package was written to support the air pollution data processing and is available at Github:

https://github.com/swish-climate-impactassessment/BiosmokeValidatedEvents.

Directory: Impact_Fire_Smoke_Plumes

(i) Data Files

/Data/AQS_SPECData/ contains data on speciated PM2.5 in the U.S. for 2000-2013.

/Data/dailyOzone/ contains data on ozone in the U.S. for 2006-2013.

/Data/dailyPM2.5FRM/ contains data on total PM2.5 in the U.S. for 2006-2013.

/Data/HMS_Smoke/ contains spatial data on smoke plumes in the U.S. for 2004-2014.

/Data/GSOD_Temp/ contains data on temperature in the U.S. for 2005-2014.

/Data/ParameterCodes.csv contains the relevant pollutant codes to the analysis.

(ii) Data Preprocessing Code

/FRM_data compilation.R combines all years of total PM2.5 data into one file.

/IMPR_Data Compilation.R combines all years of total PM2.5 data into one file.

/Ozone_Data Compilation.R combines all years of ozone data into one file.

/addGSOD.R adds the temperature data to the pollution data.

/addHMS.R add the smoke plume data to the pollution data.

(iii) Analysis Code

/AQI Analysis.R runs the AQI analysis for Ozone and total PM2.5.

/FRM_Analysis.R runs the analysis for the total PM2.5 data and creates Figures 2b and 3b.

/IMPR_Analysis.R runs the analysis for the speciated PM2.5 data.

/Ozone_Analysis.R runs the analysis for the Ozone data and creates Figures 2a and 3a.

/aqiToColor.R converts AQI numeric values to color codes.

/calcPropPlume.R calculates the proportion of plume days.

/disc_colorMap.R creates a discretized map of the results.

/Model Comparison.R contains the log-likelihood functions for the analysis.

/National Estimates.R computes confidence intervals for the national estimates.

/pmTOaqi.R converts pollutant values to AQI numeric values.

/stage1Regression.R runs the stage 1 regression analysis.

(iv) Figures Code

/Figure1.R creates Figure 1.

Directory: Spatial_Causal_Wildfire_PM

(v) Data Files

/Data/CMAQ_2008_2012/ contains data on modeled PM2.5 w/ and w/out fire emissions in the U.S. for 2008-2012.

/Data/FRM_pm25.Rdata contains data on monitored total PM2.5 in the U.S. for 2008-2012.

/Output/ contains the MCMC output for each region

(vi) Data Preprocessing Code

/Data_08_12.R reformats the monitored and modeled PM2.5 data for the analysis, and adds HMS plume data.

/map_state_data.R visualizes the modeled and monitored data.

/addHMSPlume.R adds the smoke plume data to the modeled and monitored data.

(vii) Analysis Code

/glm.R runs a glm to provides the starting values for the MCMC parameters.

/MCMC_v2_2.R runs the MCMC for a given region.

/get_WFData_regionGrid.R subsets the PM2.5 data into a given region of the U.S.

/regions9.R groups the U.S. into 9 regions.

/regions18.R groups the U.S. into 18 regions.

/thinCMAQ.R provides lats/lons for a prediction grid that algins with CMAQ.

/calc_N.R calculates the MCMC sample size given trim and burn-in values.

/mu_start.Rdata starting values for the lower level mean parameters.

/makeA.R produces a fire indicator matrix.

/stateMap_wRings.R maps estimates and standard errors.

/gdd.R is an older version of get_WFData_regrionGrid.

(viii) Figures Code

/figures.R creates the figures.

/getWFData_regionGrid.R

/make.plotdf.R

/post.summaries.R

/calcN.R

/get_ts_data.R

/calc_corr.R

/map.puzzle.R

/scatterplot.compare.R

/timeSeries.R

/plotCorr.R

(ix) Extensible database of validated biomass smoke events for health research

The extensible Biomass Smoke Validated Events Database is an ongoing, community driven, collection of air pollution events which are known to be caused by vegetation fires such as bushfires (also known as wildfire and wildland fires), or prescribed fuel reduction burns, and wood heaters. This is useful for researchers of health impacts who need to distinguish smoke from vegetation versus other sources. The overarching aim is to study statistical associations between biomass smoke pollution and health. Extreme pollution events may also be caused by dust storms or fossil fuel smog events and so validation is necessary to ensure the events being studied are from biomass. This database can be extended by contribution from other researchers outside the original team. There are several available protocols for adding validated smoke events to the database, to ensure standardization across datasets. Air pollution data can be included, and free software was created for identification of extreme values. Protocols are described for reference material needed as supporting evidence for event days. The utility of this database has previously been demonstrated in analyses of hospitalization and mortality. The database was created using open source software that works across operating systems. The prospect for future extensions to the database is enhanced by the description in this paper, and the availability of these data on the open access Github repository enables easy addition to the database with new data by the research community.