

A model for rapid wildfire smoke exposure estimates using routinely-available data

Sean Raffuse¹, Susan O'Neill², and Rebecca Schmidt³

¹Air Quality Research Center, University of California Davis, Davis, CA, United States

²Pacific Northwest Research Station, USDA Forest Service, Seattle, WA, United States

³Department of Public Health Sciences, MIND Institute, University of California Davis School of Medicine, Davis, CA, United States

Correspondence: Sean Raffuse (sraffuse@ucdavis.edu)

Abstract. Urban smoke exposure events from large wildfires have become increasingly common in California and through the western United States. The ability to study the impacts of high smoke aerosol exposures from these events on the public is limited by the availability of high-quality, spatially-resolved estimates of aerosol concentrations. Methods for the assigning aerosol exposure often employ multiple data sets that are time consuming and expensive to create and difficult to reproduce.

5 As these events have gone from occasional to nearly annual in frequency, the need for rapid smoke exposure assessments has increased. The rapidfire R package provides a suite of tools for developing exposure assignments using data sets that are routinely generated and publicly available within a month of the event. Specifically, rapidfire harvests official air quality monitoring, satellite observations, meteorological modeling, operational predictive smoke modeling, and low-cost sensor networks. A machine learning approach (random forests regression) is used to fuse the different data sets. Using rapidfire, we produced
10 estimates of ground-level 24-hour average PM_{2.5} over for several large wildfire smoke events in California from 2017-2021. These estimates show excellent agreement with independent measures of PM_{2.5} from filter-based networks.

Copyright statement. The author's copyright for this publication is transferred to institution/company.

1 Introduction

- California smoke impact on the rise (and throughout the west)
- 15 – Expected to continue
- Health impact of wildfire smoke and need for research
- Need for rapid, inexpensive exposure modeling (increasing frequency and intensity of events)
- pedigree (NASA HAQAST and Yufei's work)
- growth of low-cost sensor networks and their fidelity for fires
- 20 – Compared with other methods (land use regression), this may be more suitable for fires because of the regional nature of the source

2 Methods

2.1 Input Data Sets

2.1.1 Official and Temporary Monitoring

25 Hourly $\text{PM}_{2.5}$ observations are available from monitoring stations across the United States via the AirNow network [ref]. Within California, about [number] of monitors were operating during the study period. These permanent monitors are a mixture of federal reference method or federal equivalent method instruments, meaning that they are approved by the US EPA to calculate and report air quality to the public.

During wildfires, temporary monitors are also deployed by several government agencies, such as the California Air Resources Board (CARB), and the USDA Forest Service (USFS). These are mostly [what]. Though they are not as accurate as the AirNow monitors [ref], they are deployed in regions where smoke impacts are significant and permanent monitoring is sparse or absent. Hourly $\text{PM}_{2.5}$ concentrations from both the permanent and temporary monitors were acquired using the `rapidfire::get_airnow_daterange` and `rapidfire::get_airstis_daterange` functions. These wrap the `monitor_subset` function from the `PWFSLSmoke` R package [Mazama Science]. `rapidfire::recast_monitors` was then used to calculate daily 24-hr averages from the hourly data. At least 16 hours are required to produce an average.

The daily average data from both the permanent and temporary monitors were combined into a single data set. The spatial extent of the monitors used in this analysis are shown in Figure [xx]. Portions of this monitor data set were withheld for development and validation of the model. $\text{PM}_{2.5}$ observations were interpolated to estimate concentrations at locations away from the monitors using ordinary kriging. 30% of the monitoring data were withheld as test data to develop model variograms using `rapidfire::create_airnow_variograms`.

– figure of Permanent and Temporary monitors.

2.1.2 Smoke Modeling

Air quality models provide ground-level estimates of $\text{PM}_{2.5}$ on an output grid. We processed daily average values acquired from the BlueSky Daily Run Viewer (Websky), developed by the USFS AirFire Team. Depending on the event year, different model runs were available. Modeling from Websky was chosen because it is available operationally, is high spatial resolution, and is focused specifically on modeling smoke aerosols from wildland fires. [Susan, can you help me here to describe which models were used and their references?] On some days, the model did not run successfully. For those days, data were backfilled by using the second or third day of a previous day's 72-hr model run.

2.1.3 Satellite Aerosol Optical Depth

50 Satellite aerosol optical depth (AOD) is a measure of the total columnar aerosol light extinction from the satellite sensor to the ground. AOD is indirectly related to $\text{PM}_{2.5}$, with the relationship depending on aerosol type, humidity, and aerosol vertical profile [ref]. We used AOD from the Multi-Angle Implementation of Atmospheric Correction (MAIAC) project [ref]. MAIAC

is an algorithm that uses time series analysis and additional processing to improve aerosol retrievals, atmospheric correction, and, importantly, cloud detection from the MODerate-resolution Imaging Spectroradiometers (MODIS) onboard NASA's Terra and Aqua satellites [Lyapustin et al, 2011a,b; 2021; 2018]. Past work has shown that thick smoke is often mistaken for clouds in the standard MODIS algorithms [ref], which hampers their use in wildfire conditions.

The `rapidfire::maiac_download` function can be used to acquire the 1-km daily atmosphere product (MCD19A2) which contains AOD. Clouds prevent the retrieval of AOD, and there are sometimes clouds present even in the hot, dry conditions during California wildfires. The data fusion algorithm requires a complete data set, so a placeholder value must be used to gap-fill in locations under clouds. Previous work has used model-simulated AOD, along with meteorological variables in a data fusion approach to gap-fill satellite-observed AOD [Zou 2019]. For this work, where clouds cover less of the domain, we took a simpler approach. Missing AOD values were filled using a three-stage focal average, available in `rapidfire::maiac_fill_gaps_complete`. [more to describe the figure]

– figure illustrating the gap filling approach

65 **2.1.4 Low-cost Sensors**

2.1.5 Meteorology

2.2 Data Fusion

2.3 Model Training and Validation

2.3.1 IMPROVE, CSN, CA FRM

70 **3 Code?**

4 Results

4.1 Model development plots? (Test vs. Training)

4.2 Comparison to filter measurements

4.2.1 crossplots

75 **4.2.2 time series?**

4.3 Maps

4.4 Compare to other method (interpolation alone)

5 Discussion

- Importance plots
- 80 – application for health studies
- advantages over existing methods
- limitations

6 Content section with citations

 See the R Markdown docs for bibliographies and citations.

- 85 Copernicus supports biblatex and a sample bibliography is in file `sample.bib`. Read (Feynman and Vernon Jr., 1963), and (see Dirac, 1953).

7 Content section with R code chunks

You should always use `echo = FALSE` on R Markdown code blocks as they add formatting and styling not desired by Copernicus. The hidden workflow results in 42.

90 You can add verbatim code snippets without extra styles by using ````` without additional instructions.

```
sum <- 1 + 41
```

8 Content section with list

If you want to insert a list, you must

- leave

95 – empty lines

- between each list item

because the `\tightlist` format used by R Markdown is not supported in the Copernicus template. Example:

- leave

100 – empty lines

- between each list item

9 Examples from the official template

9.1 FIGURES

105 When figures and tables are placed at the end of the MS (article in one-column style), please add

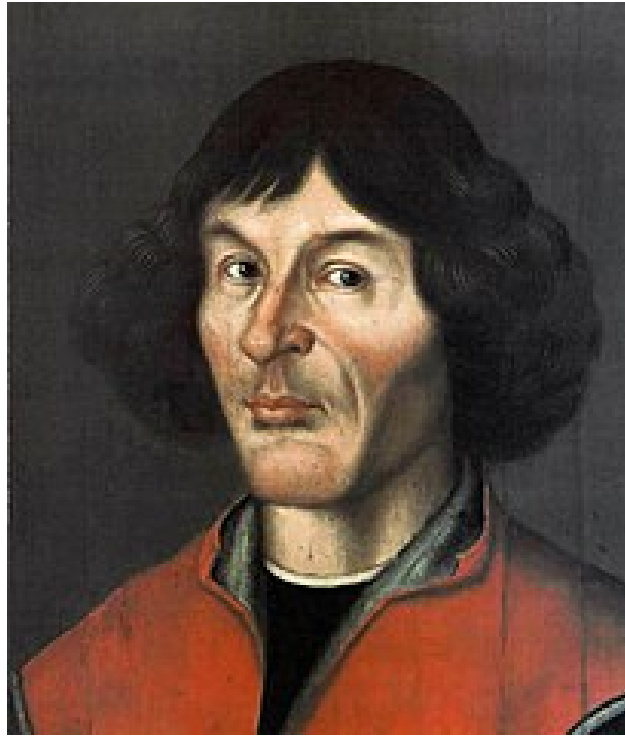


Figure 1. one column figure

between bibliography and first table and/or figure as well as between each table and/or figure.

9.1.1 ONE-COLUMN FIGURES

Include a 12cm width figure of Nikolaus Copernicus from Wikipedia with caption using R Markdown.

9.1.2 TWO-COLUMN FIGURES

110 You can also include a larger figure.

9.2 TABLES

You can add `\LaTeXtable` in an R Markdown document to meet the template requirements.

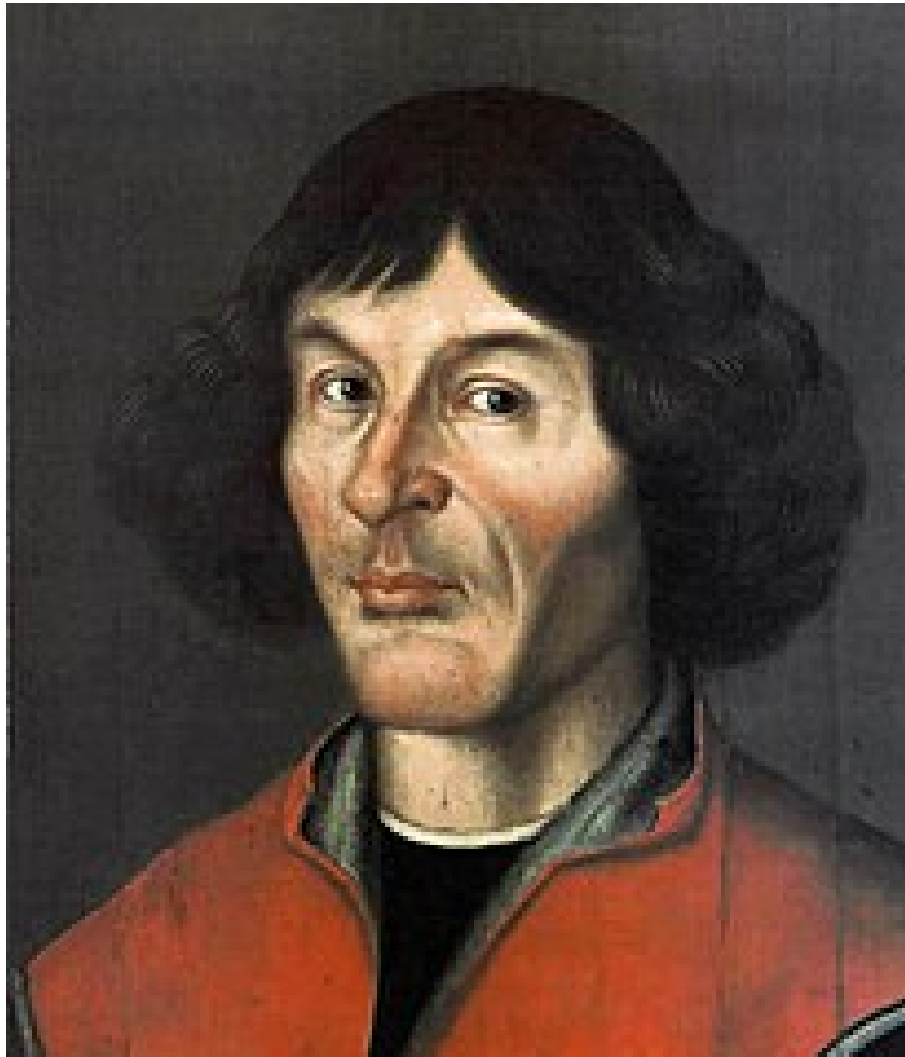


Figure 2. two column figure

Table 1. TEXT

a	b	c
1	2	3

Table Footnotes

Table 2. TEXT

a	b	c
1	2	3

Table footnotes

9.2.1 ONE-COLUMN TABLE

9.2.2 TWO-COLUMN TABLE

115 **9.3 MATHEMATICAL EXPRESSIONS**

All papers typeset by Copernicus Publications follow the math typesetting regulations given by the IUPAC Green Book (IUPAC: Quantities, Units and Symbols in Physical Chemistry, 2nd Edn., Blackwell Science, available at: <http://old.iupac.org/publications/book1993>).

- Physical quantities/variables are typeset in italic font (t for time, T for Temperature)
- 120 Indices which are not defined are typeset in italic font (x, y, z, a, b, c)
- Items/objects which are defined are typeset in roman font (Car A, Car B)
- Descriptions/specifications which are defined by itself are typeset in roman font (abs, rel, ref, tot, net, ice)
- Abbreviations from 2 letters are typeset in roman font (RH, LAI)
- Vectors are identified in bold italic font using \boldsymbol{x}
- 125 Matrices are identified in bold roman font
- Multiplication signs are typeset using the LaTeX commands `\times` (for vector products, grids, and exponential notations) or `\cdot`
- The character `*` should not be applied as multiplication sign

9.4 EQUATIONS

130 9.4.1 Single-row equation

Unnumbered equations (i.e. using `$$` and getting inline preview in RStudio) are not supported by Copernicus.

$$1 \times 1 \cdot 1 = 42 \tag{1}$$

$$A = \pi r^2 \tag{2}$$

$$x = \frac{2b \pm \sqrt{b^2 - 4ac}}{2c}. \tag{3}$$

135 9.4.2 Multiline equation

$$3 + 5 = 8 \tag{4}$$

$$3 + 5 = 8 \tag{5}$$

$$3 + 5 = 8 \tag{6}$$

9.5 MATRICES

$$\begin{matrix} x & y & z \\ 140 & x & y & z \\ & x & y & z \end{matrix}$$

9.6 ALGORITHM/PROGRAMMING CODE

If you want to use algorithms, you need to make sure yourself that the \LaTeX packages `algorithms` and `algorithmicx` are installed so that `algorithm.sty` respectively `algorithmic.sty` can be loaded by the Copernicus template. Both need to be available through your preferred \LaTeX distribution. With TinyTeX (or TeX Live), you can do so by running

```
145 tinytex::tlmgr_install(c("algorithms", "algorithmicx"))
```

Copernicus staff will no accept any additional packages from your LaTeX source code, so please stick to these two acceptable packages. They are needed to use the example below

9.7 CHEMICAL FORMULAS AND REACTIONS

For formulas embedded in the text, please use `\chem{ }`, e.g. $A \rightarrow B$.

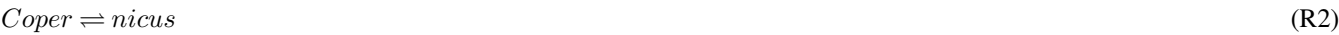
150 The reaction environment creates labels including the letter R, i.e. (R1), (R2), etc.

```
i ← 10
if i ≥ 5 then
  i ← i − 1
else
  if i ≤ 3 then
    i ← i + 2
  end if
end if
```

- `\rightarrow` should be used for normal (one-way) chemical reactions
- `\rightleftharpoons` should be used for equilibria
- `\leftrightarrow` should be used for resonance structures



155



9.8 PHYSICAL UNITS

160 Please use `\unit{}` (allows to save the `math/$` environment) and apply the exponential notation, for example 3.14 km h^{-1} (using LaTeX mode: `\(3.14\,, \unit{...} \)`) or 0.872 ms^{-1} (using only `\unit{0.872\,, m\,, s^{-1}}`).

10 Conclusions

The conclusion goes here. You can modify the section name with `\conclusions[modified heading if necessary]`.

Code and data availability. use this to add a statement when having data sets and software code available

165 *Sample availability.* use this section when having geoscientific samples available

Video supplement. use this section when having video supplements available

Appendix A: Figures and tables in appendices

Regarding figures and tables in appendices, the following two options are possible depending on your general handling of figures and tables in the manuscript environment:

170 **A1 Option 1**

If you sorted all figures and tables into the sections of the text, please also sort the appendix figures and appendix tables into the respective appendix sections. They will be correctly named automatically.

A2 Option 2

If you put all figures after the reference list, please insert appendix tables and figures after the normal tables and figures.

175 To rename them correctly to A1, A2, etc., please add the following commands in front of them: `\appendixfigures` needs to be added in front of appendix figures `\appendixtables` needs to be added in front of appendix tables

Please add `\clearpage` between each table and/or figure. Further guidelines on figures and tables can be found below.

Author contributions. Daniel wrote the package. Josiah thought about pottery. Markus filled in for a second author.

Competing interests. The authors declare no competing interests.

180 *Disclaimer.* We like Copernicus.

Acknowledgements. Thanks to the rticles contributors!

References

Dirac, P.: The lorentz transformation and absolute time, *Physica*, 19, 888–896, [https://doi.org/10.1016/S0031-8914\(53\)80099-6](https://doi.org/10.1016/S0031-8914(53)80099-6), 1953.

Feynman, R. and Vernon Jr., F.: The theory of a general quantum system interacting with a linear dissipative system, *Annals of Physics*, 24,
185 118–173, [https://doi.org/10.1016/0003-4916\(63\)90068-X](https://doi.org/10.1016/0003-4916(63)90068-X), 1963.