Challenges in Studying Wildfire Health Impacts – Exposure Assessment

National Academies of Sciences – Implications of the California Wildfires for Health, Communities, and Preparedness

Colleen Reid, PhD MPH
Assistant Professor, Department of Geography
University of Colorado Boulder
Email: Colleen.Reid@Colorado.edu

Image Credit: NASA
When investigating a wildfire retrospectively, how do you assess which people were exposed?
Exposure Assessment Methods Used in Wildfire Epidemiological Studies

- Temporal Comparisons
  - Relatively easy
  - Can’t estimate dose-response
  - Associations may not be solely due to smoke exposure

Holstius et al. 2012, EHP
When investigating a wildfire retrospectively, how do you assess how much exposure those people experienced?
Exposure Assessment Difficulties with Wildfires

• Monitoring data
  • Lots of data – relatively easy to access
• The monitors are not always in all of the locations that you want
• Many EPA PM$_{2.5}$ monitors only measure every sixth or third day
• Leads to spatial and temporal averaging of exposure measurements
  • But, smoke plumes migrate quickly, changing exposures over smaller spatial and temporal scales
Exposure Assessment Methods Used in Wildfire Epidemiological Studies

- Nearest air pollution monitor or average of regional monitors

https://www.epa.gov/outdoor-air-quality-data/interactive-map-air-quality-monitors
Exposure Assessment Methods Used in Wildfire Epidemiological Studies

• Satellite Data – Aerosol Optical Depth
  • Benefits
    • Full spatial coverage (except for cloud masking)
    • Actual measurement
  • Drawbacks
    • Full column measurement and not just at ground level
    • Missing data

MODIS AOD August 2018

https://earthobservatory.nasa.gov/global-maps/MODAL2_M_AER_OD
Exposure Assessment Methods Used in Wildfire Epidemiological Studies

• Air pollution models
  • Dispersion models
    • HYSPLIT (Thelen et al., 2013)
    • CalPuff (Henderson et al. 2011)
  • Chemical Transport Models (CTMs)
    • GEOS Chem (Liu et al. 2016; Liu et al. 2017)
    • WRF-Chem (Gan et al. 2017)
    • CMAQ (DeFlorio-Barker et al., 2019)

• Advantages
  • Complete spatial coverage
  • Can get just ground-level concentration estimates
  • Allow estimation of counterfactual

• Disadvantages
  • Dependent on the inputs
  • Uncertainties in emissions estimates from fires
Exposure Assessment Methods Used in Wildfire Epidemiological Studies

• Blended Models
  • Statistically combine CTMs, satellite data, and monitoring data
  • Sometimes also auxiliary data

• GWR method – Gan et al. 2017 and Lassman et al. 2017

• Machine learning method – Reid et al. 2015 and Yao et al. 2018
Adapt Land Use Regression Modeling with Machine Learning and Adding Temporal Component

- Include novel spatiotemporal datasets
- Apply machine learning methods to
  - Select from a long list of predictor variables
  - Select from a variety of statistical algorithms
<table>
<thead>
<tr>
<th>Model</th>
<th>CV-RMSE (μg/m²)</th>
<th>CV-R²</th>
<th>no. of variables selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>random forest</td>
<td>1.513</td>
<td>0.796</td>
<td>20</td>
</tr>
<tr>
<td>bagged trees</td>
<td>1.687</td>
<td>0.672</td>
<td>27</td>
</tr>
<tr>
<td><strong>generalized boosting model</strong></td>
<td><strong>1.489</strong></td>
<td><strong>0.803</strong></td>
<td><strong>29</strong></td>
</tr>
<tr>
<td>elastic net regression</td>
<td>1.848</td>
<td>0.558</td>
<td>28</td>
</tr>
<tr>
<td>multivariate adaptive regression splines</td>
<td>1.642</td>
<td>0.701</td>
<td>28</td>
</tr>
<tr>
<td>lasso regression</td>
<td>1.821</td>
<td>0.558</td>
<td>28</td>
</tr>
<tr>
<td>support vector machines</td>
<td>1.556</td>
<td>0.761</td>
<td>16</td>
</tr>
<tr>
<td>gaussian processes</td>
<td>1.580</td>
<td>0.746</td>
<td>16</td>
</tr>
<tr>
<td>generalized linear model</td>
<td>1.821</td>
<td>0.558</td>
<td>29</td>
</tr>
<tr>
<td>K-nearest neighbors</td>
<td>2.030</td>
<td>0.387</td>
<td>2</td>
</tr>
<tr>
<td>generalized additive model</td>
<td>1.607</td>
<td>0.725</td>
<td>26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>CV-RMSE (μg/m²)</th>
<th>CV-R²</th>
<th>no. of variables selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>model with smallest CV-RMSE for subsets of variables</td>
<td>1.495</td>
<td>0.799</td>
<td>13</td>
</tr>
<tr>
<td>model with fewer variables whose CV-RMSE was within 1.5% of the smallest CV-RMSE</td>
<td>1.521</td>
<td>0.790</td>
<td>14</td>
</tr>
</tbody>
</table>

Large circles are observed values at monitors, small circles are predicted values.
Figure 3: The mean leave-one-location-out (LOLO) cross-validated estimates of RMSE for gradient boosting and random forest at each monitor location between May 6, 2008 and September 26, 2008 smoothed throughout the study region using a two-dimensional spline-on-sphere smoother.
PM$_{2.5}$ and ozone exposure estimates by ZIP code by day for the 2008 northern California wildfires

Fig. 3. Ozone levels by ZIP-code day during the study period with averages for some air basins.

Reid et al. 2019 *Env Int*
<table>
<thead>
<tr>
<th>Variables</th>
<th>Data Source</th>
<th>Temporal Resolution</th>
<th>Spatial Resolution</th>
<th>Buffer Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable</td>
<td>US EPA, states, Federal Land Manager Environmental Database, Fire Cache Smoke Monitor Archive, IMPROVE Network, academic research groups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM2.5 from monitoring stations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatiotemporal Variables</td>
<td>GOES Aerosol and Smoke Product (GASP) AOD NASA</td>
<td>Hourly 4 km</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD NASA</td>
<td>Daily 1 km</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MODIS Active Fire Detection NASA</td>
<td>Daily 1 km</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>VIIRS Fire Occurrence NASA</td>
<td>Daily 1 km</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hazard Mapping System (HMS) Smoke and Fire Product NOAA</td>
<td>Daily 4 km</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MODIS Normalized Difference Vegetation Index (NDVI) NASA</td>
<td>Daily 375 m</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14 Meteorological Variables NOAA: North American Mesoscale (NAM)</td>
<td>6-Hourly 12 km</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial Variables</td>
<td>Elevation (m) USGS</td>
<td></td>
<td>1 arc-second</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Percentages of land cover types National Land Cover Database 2011</td>
<td>Every 5 years 30 m</td>
<td>1 km</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kilometers of highway within buffer zones National Highways Planning Network</td>
<td>100, 250, 500, 1000 m, 2000 m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal Variables</td>
<td>Julian Date</td>
<td></td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weekend</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Maestas, Reid, Considine, Li et al. *Work In Progress*
Blended Models...

• Strengths
  • May be able to combine best of satellite and CTM
  • Full spatial coverage

• Weaknesses
  • There are assumptions of each model that should be understood – performance may vary spatially and temporally
  • There are errors in these models that should be considered
  • Results of epidemiological investigations may differ because of the exposure assessment method....
Different findings from different exposure models....

Henderson et al. (2011) *EHP*  
Gan et al. (2017) *GeoHealth*
Moving Forward....

• More work to improve and compare exposure models
• More work on integrating exposure measurement error into epidemiological models
• Possibly increased use of social media on its own or in blended models to estimate exposure (i.e., Ford et al. 2017 Atmos. Chem. Phys.)
• Cheaper sensors or personal monitoring
References

Thank You!!
Questions?
Colleen.Reid@Colorado.edu