Fire Safety Journal 95 (2018) 122–134

Contents lists available at ScienceDirect

Fire Safety Journal

journal homepage: www.elsevier.com/locate/firesaf

Exploration of NFIRS protected populations using geocoded fire incidents

Austin Anderson, Ofodike A. Ezekoye *

Fire Research Group, Department of Mechanical Engineering, The University of Texas at Austin, Austin, TX 78712, USA

ARTICLE INFO

Keywords:
NFIRS
NFPA
National estimates
Fire reporting
Fire
Geocoding
GIS processing
Fire department
Protected population
Fire problem
NFIRS non-reporting
NFIRS underreporting

ABSTRACT

The NFIRS database is one of the primary resources for data relevant to fire statistics in the United States. This database is populated by self reported fire incidents by U.S. fire departments. A major criticism of this database for statistical analysis is the problem of reporting bias. One cannot use basic summary statistics and other useful, simple metrics to characterize the U.S. fire problem because the NFIRS database does not mandate reporting from the entire nation. Thus, it only contains data from a subset of fire departments across the nation. Additionally, no major attempts have been made to remedy this problem because U.S. fire data consumers rely on the only other data source (NFPA estimates) at the national level. This paper explores approaches to better characterize the NFIRS reporting population. Public data release records from the NFIRS database were geocoded, and then spatial data processes were applied to obtain a reasonable estimate of the protected populations of NFIRS reporting fire departments. The analysis indicates that between 71 and 83% of the American population is covered by NFIRS reporting fire departments. Additionally, estimates of fire incident, fatality, and injury rates from NFIRS reporting populations appear to be roughly 20%–50% lower than the national rates as reported by the NFPA. These rates are confirmed to be feasible by additional analyses that also supply estimates of the non-NFIRS-reporting population's relevant incident and fatality rates. Regional estimates of the NFIRS-reporting population's fire problem are also generated. Until now, there had been no means of generating an independent assessment of the US fire problem. This new modeling approach provides researchers with a publicly available toolset to conduct rigorous sensitivity analysis on the impact of fire prevention and mitigation approaches at the national level.

1. Introduction

In 1973, the National Commission on Fire Prevention and Control published its seminal report on fire in the United States, *America Burning*. This report represented the first attempt by the U.S. government to quantify the fire problem in America and its results led to the formation of the United States Fire Administration (USFA). One of the primary charges of the USFA, as recommended in *America Burning*, was the development of a “fire data base for the Nation's fire services and the Federal and State governments … [which] will provide for a nationwide exchange of information pertaining to fire and life safety and have data collection, storage, retrieval, and dissemination capability.” [1]. The National Fire Incident Reporting System (NFIRS) is the USFA's implementation of this charge, and since 1978 it has served as what could be considered one of the two pillars of fire data in the United States. The other pillar is the NFPA's National Fire Department Experience Survey. While the NFIRS survey data are publicly available, the NFPA data are not.

There are valid criticisms of the NFIRS data. It has historically been plagued with dubious data quality, reporting inconsistencies across fire departments, and perhaps largest of all, not all fire departments in the nation are required to report, and thus the *information in NFIRS, while potentially possessing good depth, is ultimately not representative of the nation's fire problem*. Approaches that try to compensate for this weakness have been implemented. While the technique had been available for some time, in 1989, Hall and Harwood published a paper describing best practices for applying what was coined the “national estimates” approach for characterizing the U.S. fire problem. This approach used the NFPA's National Fire Department Experience Survey in conjunction with a scaling and one-step raking procedure to derive national estimates for NFIRS database fields [2]. The raking procedure additionally relies on an underlying assumption that missing or underreported incidents from fire departments who report to NFIRS are represented by the reported incidents.

The major caveat with using this particular approach is that direct scaling of NFIRS data to national estimates implies that the specific kinds of fires recorded by NFIRS reporting fire departments are the same as the specific kinds of fires experienced by the nation as a whole. Some exploration of this issue by Butry and Thomas [3] has indicated that there are significant differences in fire risk indicators between NFIRS reporting

* Corresponding author.
E-mail address: dezekoye@mail.utexas.edu (O.A. Ezekoye).

https://doi.org/10.1016/j.firesaf.2017.11.013
Received 20 November 2015; Received in revised form 14 November 2017; Accepted 17 November 2017

and non-reporting cities. These results shed doubt on the viability of the NFIRS "representativeness" assumption implied by the scaling procedure. A preferable alternative is to identify the actual NFIRS reporting population and to then use that population when discussing or analyzing specific kinds of fires present in NFIRS without adjustment.

Indeed, it is notable that in 1989 John Hall wrote the following when discussing the NFIRS database:

> There is, however, more to national estimates than finding one number. Just as it is not known how large a share of the total fire problem is contained in the sample that NFIRS represents, *so it also is not known how representative that sample is*, and this has implications. … Participation rates, therefore, are not necessarily uniform across regions and sizes of community, both of which are factors correlated with frequency and severity of fires. This means NFIRS may be susceptible to systematic biases. No one at present can quantify the size of these deviations from the ideal, representative sample, so no one can say with confidence that they are or are not serious problems [2]. [emphasis added]

There have been some attempts conducted by the USFA aimed at examining the representativeness of the NFIRS database, but these attempts largely focus on the subset of fire departments reporting to the NFIRS database, and do not consider demographic characteristics of fire departments reporting to the NFIRS database [4].

Despite some of these shortcomings, the NFPA estimate approach has been embraced and is still utilized in various forms to this day by the USFA, the Consumer Products Safety Commission (CPSC) and many researchers using fire statistics in the U.S. [5,6]. In fact, surveying manuscripts on U.S. fire statistics at the national level and written within the past 30 years show that these studies only use the NFPA national estimate to model U.S. fire incidents. Interestingly, no rigorous error analysis has been published on the veracity of the NFPA estimate. One of the reviewers of this paper expressly stated that there is no guarantee that the NFPA estimate is accurate. The reviewer is correct. The scope of this work, however, is not to conduct a new survey using the same approach as NFPA in order to assess the statistical reliability of the NFPA estimates, but instead to create as substantially different an approach as can be developed using the only publicly available data set (i.e., NFIRS). The goal, then, is to provide this methodology to other researchers who can conduct their own assessments of NFPA estimates for use in better defining the bounds of the US fire problem.

This paper seeks to define the characteristics and size of the populations protected by NFIRS reporting fire departments using modern tools unavailable to Hall at the time he discussed these issues. Namely, batch geocoding of NFIRS incidents using the address data in NFIRS, coupled with generation of concave hulls, is used to estimate the protected areas of every NFIRS reporting fire department, with the exception of forest service and wildland fire responses. These protected area estimates are then applied with NFIRS fire statistics to calculate incident, fatality, and injury rates for the NFIRS reporting U.S. fire departments. Additionally, potential explanations for the discrepancies between NFIRS reporting fire department rates and national rates as estimated by the NFPA are explored. Finally, rate calculations are presented for regions within the nation and compared against NFPA regional estimates.

2. Methodology

The method posed by this paper for estimating NFIRS reporting fire department coverage is summarized in Fig. 1, and outlined as follows:

1. Obtain fire department response data from the NFIRS database.
2. Geocode fire department responses using address data supplied in the NFIRS database for certain types of responses.
3. Generate a polygon representing each fire department's jurisdiction by drawing a concave hull around its set of geocoded fire responses.
4. Spatially intersect Census tract or block data with the fire department polygons.

The above operations are all readily performed by most GIS software today. The authors used PostGIS 2.1 on a PostgreSQL 9.3 server. PostGIS is an extension that spatially enables PostgreSQL servers, allowing spatial data to be stored and spatial operations to be performed. Additionally, PostGIS is packaged with a free TIGER/Line data geocoder, which enables the geocoding of the approximately 1.6 million incidents with some address data match in a reasonable amount of time.

Note that for geographic census tract operations and analysis, the 2012 TIGER/Line shape files available from the U.S. Census Bureau were used. Additionally, for geographic census block operations and analysis, the 2010 TIGER/Line shape files were used. Full geographic census block data are only available from decennial census years.

2.1. A note about NFIRS data

The database used in this paper was supplied by the USFA. It consists of the content and incidents contained in the 2002–2012 public data release files from the NFIRS database as supplied by the USFA. Included in these data are approximately 36.5 million recorded calls of which 6.6 million are fire incidents, of which roughly 3.4 million are structure fire incidents.

2.2. Geocoding

Geocoding is the process of converting address data to some spatial coordinate system. The particular software and configuration used to

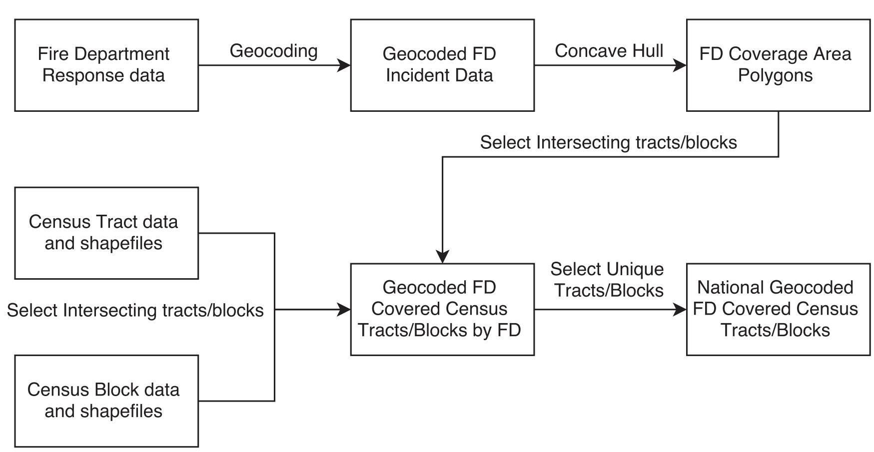


Fig. 1. Summary of NFIRS geocoding methodology.

geocode the data were discussed earlier. The authors attempted to geocode the 3.4 million structure fire incidents with specified addresses using the PostGIS geocoder. Of these 3.4 million incidents, 1.6 million had some form of address match located, and of these 1.6 million incidents, 1 million were matched to an address exactly. These address matches are generally more likely to match structure addresses in urban areas than in rural areas. Thus, it is expected that some rural FDs will be excluded not necessarily just on the basis of too few incidents reported in general, but also because too few of their available incidents could be geocoded.

Of these 1 million geocoded addresses that were matched exactly to the TIGER/Line data, 1000 were randomly sampled and then geocoded using Google maps to confirm acceptable levels of agreement. The average distance between the TIGER/Line geocoded addresses with “exact” matches and the google results was 0.07 miles, with a standard deviation of 0.24 miles. Note that of the 1000 points sampled, the results of two points were excluded on account of Google's string matching algorithm specifying an address within the wrong city or state, resulting in gross disagreement. Normally, that level of error might give pause if one's task were to, as an example, estimate fire incident density within communities to generate insight regarding where to position resources. However, this study used the aggregate fire incident responses within communities to estimate a given fire department's population coverage. Thus, the spatial error incurred by using TIGER/LINE data versus more precise, and expensive, NavTEQ data was deemed acceptable.

Note that raw geocoding of fire incidents is not the advancement in this study. Many metropolitan fire departments geocode their call responses. However, the next step of generating fire department jurisdiction polygons on a national basis to examine the demographics of NFIRS reporting fire department populations is an important and useful extension of this process for analysis.

2.3. Polygon and tract coverage generation

Mathematically, a concave hull is more appropriately referred to as an α-shape. α-shapes were discussed in detail by Edelsbrunner in Ref. [7], which supplies a solid mathematical treatment. For the purposes of this paper, the colloquial term “concave hull,” is used, as it is typically preferred by the GIS community for these purposes. Practically, the operation of constructing a concave hull around a set of points is akin to shrink-wrapping the border of the points. Fig. 2 displays this operation for the fire incidents representing the city of Austin, Texas's fire department (AFD). One can see that the region simply connects the exterior points of the set with straight lines to form an estimated jurisdiction. Note that as a result of the mathematical constraints of the process, only fire departments with 3 or more geocoded fire incidents could have their coverage estimated. The results of this process for the nation are depicted in Fig. 3. These polygons represent estimations of these fire department's jurisdictions based on actual fires they reported to have responsed to.

Not surprisingly, note the vast area for which no fire reporting occurs. As expected, high population density areas are disproportionately represented. Interesting exceptions are large coverage areas in the Western U.S. and in Alaska. Many of the large coverage areas in the Western U.S. are associated with either the larger geographic county sizes related to greater coverage from County fire departments, or fire and rescue districts, which are fire districts that cover multiple incorporated areas. For example, the large fire department covering most of Northern Alaska, labeled (a) in Fig. 3 is the North Slope Borough FD, whose jurisdictional area is the North Slope Borough, a large county-like administrative division in Alaska that covers 229,720 sq. kms. Another reason for large coverage areas relates to mutual aid incidents in which one fire department responds to an incident within another fire department's jurisdiction. The large fire department in Wyoming, labeled (b) in Fig. 3, is actually headquartered out of the town of Rawlins, Wyoming, located at the southernmost tip of the triangle. It responded to three incidents well outside its “designated” jurisdiction in two other cities in Wyoming, corresponding to the other two points of the triangle.

The next step of the process is to document the census tracts intersecting the given fire department's polygon, assuming that those tracts effectively represent the population protected by the fire department. In the U.S. a census tract is a geographic region containing a population of approximately 4000 people. For qualitative comparison, Fig. 4 displays the results of this operation for the city of Austin, compared against the city's actual 2008 public safety response jurisdictions. Austin Texas is a city of approximately 1 million people that covers 711 sq. kms and intersects roughly 219 census tracts. Note that the holes (white regions) in Fig. 4a represent two separate governmental units whose boundaries are entirely encapsulated by the city of Austin: the town of West Lake Hills and the town of Sunset Valley. The populations of these two towns are technically not covered by AFD, an issue that will be discussed in the following section. The results of this process on a national scale are depicted in Fig. 5. This gives the reader a sense, from a population perspective, of how much of the nation is covered. Note however, that the land area covered by fire departments as depicted in Fig. 5, is likely tremendously overestimated for Western states, where census tracts are large enough that very small volunteer fire departments can appear to cover large land areas based on the operations performed. This issue is also discussed in the following section.

2.4. Assumptions and caveats in the current approach

There are a few assumptions implicit within the present approach that limit its application at all scales of analysis, and this section will discuss them.

As mentioned earlier, the assumption underpinning the operation of intersecting census tracts with a given fire department jurisdiction is that any census tract touching or containing a fire department polygon is assumed to be “covered” for the purposes of estimating its protected population. This assumption presents issues to consider. Notably, as demonstrated in Fig. 4, a large metropolitan city may contain “holes” in its jurisdiction indicating independent populations not covered by its fire department. Likewise, and perhaps more pronounced, a county fire department does not primarily serve the populations of a large metropolitan city within its borders, but instead serves the unincorporated

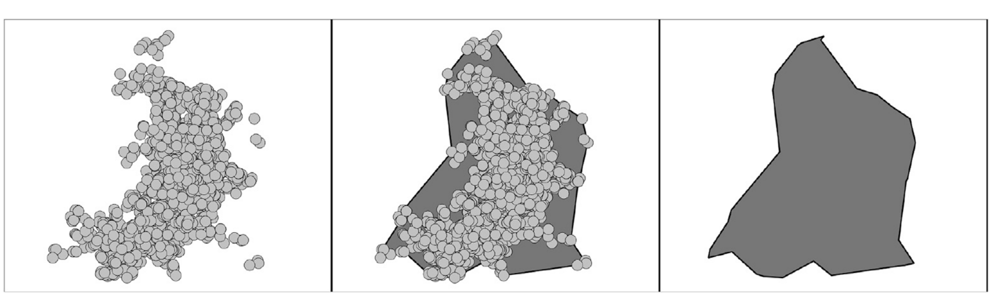


Fig. 2. Depiction of concave hull creation process. Respectively from the left: Points representing fire incidents, points representing fire incidents overlapped with the generated concave hull, and the concave hull itself. Notice that the border is “shrink-wrapped” around the exterior points of the point cloud.

Fig. 3. Estimate of the portion of America covered by NFIRS reporting fire departments. Gray regions are covered, white regions are uncovered. Points of interest include (a) The North Slope Borough FD in Alaska, and (b) City of Rawlins FD in Wyoming.

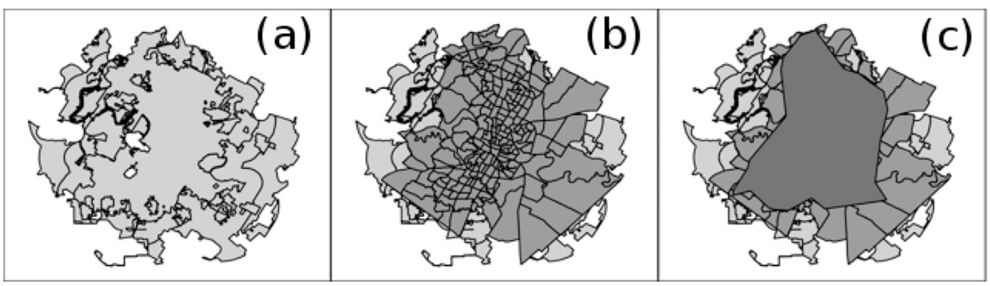


Fig. 4. Depiction of the census tracts intersection process for the City of Austin, TX. From the left: (a) Actual 2008 public safety response jurisdiction from the City of Austin, (b) Actual Jurisdiction overlapped with census tracts that intersected the CoA polygon, (c) Polygon that generated the census tract overlaps on top.

areas of the county.

Estimating jurisdictions and population protected correctly can be quite complex within metropolitan areas, where there can be sizeable numbers of jurisdictions overlapping with rather complex interactions. Fig. 6 displays this issue for jurisdictions in and around the city of Houston, Texas. Houston Texas is a city of approximately 2 million people that covers an area of 1551 sq. kms. Houston is encompassed by Harris county. Every gray polygon represents some form of separate fire department responding to incidents within the areas of their respective polygons. Considerable overplotting hides some underlying jurisdictions as well. Notable interactions include the overlap between the Harris County and Houston fire departments, as well as a very small polygon indicating the very small town of West University Place's fire department, which, like West Lake Hills in the city of Austin, is its own municipality. Additional processing would be required to correctly attribute protected populations to the right jurisdictions in these incidents, but is rather

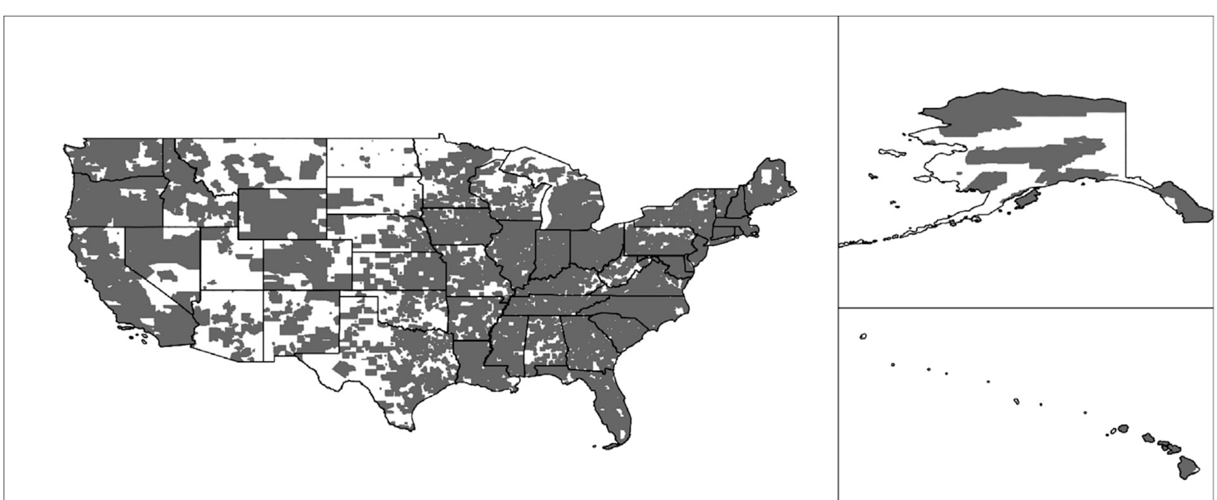


Fig. 5. Estimate of the portion of census tracts in America that are "covered" by NFIRS reporting fire departments, subject to the assumption that any contact with a tract constitutes coverage. This figure gives a sense of *population protected* by NFIRS reporting fire departments. Refer to Fig. 3 to obtain a qualitative sense of land area protected.

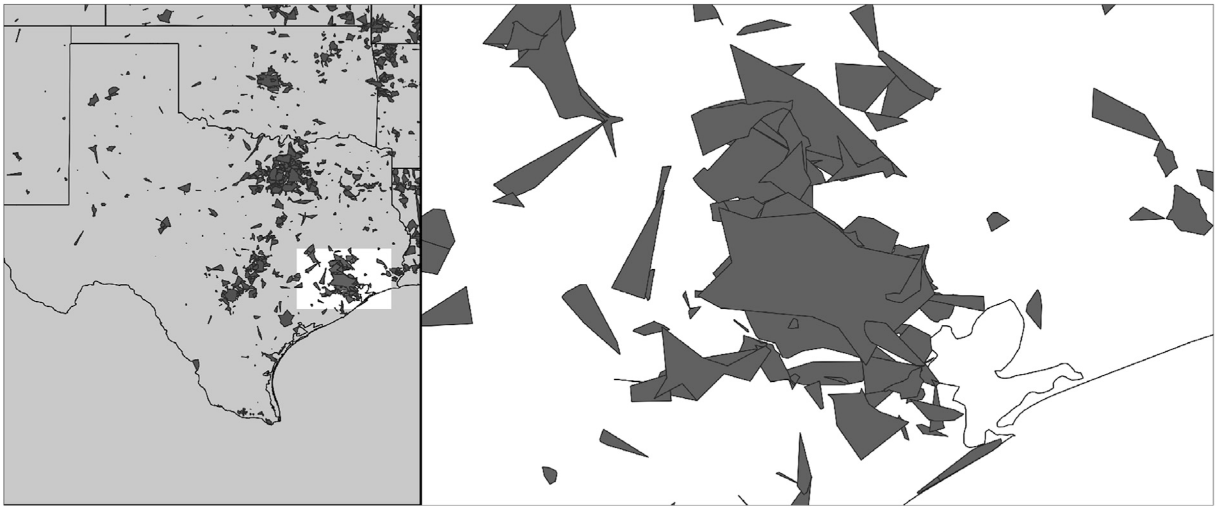


Fig. 6. Overlapping jurisdictions in and around the City of Houston, Texas Note the number of small jurisdictions (small polygons) that appear to operate exclusively within a bigger jurisdiction's "realm of influence." Usually these are actually independent jurisdictions responsible for the population inside the "hole" of the larger jurisdiction.

computationally and time intensive in most cases. Because this paper examines NFIRS coverage and incidents at the Regional and National levels, this additional level of characterization was not made, as the approach in this paper requires only data on whether the population of a given census tract is covered by *any* fire department.

Another issue with using the intersection of census tracts with fire department response polygons is presented by Fig. 7. This figure displays a very small fire department operating in Northern Nevada, the Winnemucca volunteer fire department. Notice from the figure that the actual fire department polygon is very small, but just barely intersects three additional census tracts, which are then included in the department's protected population. Recall that census tracts contain populations between roughly three to five thousand individuals. Thus, errors of this sort will have a relatively small impact on protected population estimates at a state to national level. Once again, if one desires more precise estimates at the jurisdiction level, and is mindful of the aforementioned overlapping issues discussed above, a viable, though more computationally intensive, approach would be to use census blocks rather than census tracts to estimate population protected. Census blocks tend to be smaller in area than census tracts as they are the smallest census geographic entity in the U.S. and are defined not by population coverage but by terrain features and property demarcations. In fact many census blocks contain zero population. Fig. 8 displays the improvement in resolution afforded by census blocks for the Winnemucca example. Note that the census blocks are fine enough to outline the street network of the city. This level of resolution results in very close geographic fits of the fire department polygons to census areas for population accounting. To supply a rough scale of the additional resolution provided by census blocks, note that the decennial census subdivided the United States into 74,002 census tracts which consisted of aggregations of 11,078,297 census blocks.

Finally, examining Fig. 3, one may express concern regarding the polygon process for errors in fire department reporting data, or errors in the geocoding process, when examining, for example, some of the "spikier" departments in the Western United States. In particular, there are two fire departments whose polygons are depicted in more detail in Fig. 9. These polygons represent two fire departments in Oregon that possessed sizeable spatial outlier fires, on the order of crossing the entire state. Examination of the address data for these fires, however, showed that they lined up flawlessly with TIGER/LINE address data, in other words there was zero correction necessary to the address, and it could be positioned entirely accurately. In lieu of that, there was no truly compelling reason to assert that these fires were not actually responded to by their respective fire departments, and thus to assert that the

Fig. 7. Winnemucca volunteer fire department's estimated influence. Note the manner in which the polygon slightly touches two census tracks that are then considered "covered" by the small fire department.

Fig. 8. Winnemucca volunteer fire department's estimated influence using census blocks. Note the polygon can be very finely matched by the blocks.

population in these areas is not covered in some sense by the fire departments that responded to those fires.

3. Simple analyses using this methodology

The geocoding process captured coverage estimates for 15,575 career and volunteer fire departments in the NFIRS database responding to structure fire incidents from 2002 to 2012. These 15,575 departments responded to 5.6 million out of 6.6 million reported *fire incidents*, or roughly 86% of all fire incidents reported to the NFIRS database between 2002 and 2012. The 1 million orphan incidents belong to fire departments for which no polygon could be constructed due to a lack of structure fire incidents with address data that could be exactly matched. The summary statistics discussed in this section that are derived from NFIRS thus represent the universe of all populations whose census tracts intersect a concave hull estimate of NFIRS reporting fire departments' area covered based on actual fire response, for those fire departments for which 3 or more structure fire incidents could be accurately geocoded using TIGER/LINE census data. Note also that the summary statistics, such as incident, fatality, and injury rates, derived in this section utilize *all* fire incidents, fatalities, and injuries reported by these 15,575 fire departments, and not just the geocoded structure fire incidents used to estimate their population protected according to the methodology outlined earlier.

This section will explore two simple analyses using the population estimates derived from this methodology. First, national statistics regarding fire incident, injury, and death *rates* were examined and compared against the national estimates derived by the NFPA, to obtain a sense of how “representative” NFIRS reported estimates are of the national fire problem. Additionally, potential explanations for departures between NFIRS and NFPA national rates are explored using supplemental data sources from the NFPA and U.S. census. Finally, regional estimates are computed for the northeast, midwest, south, and west regions of the United States and compared against NFPA estimates for these regions.

Note that for the analyses in this section, the comparisons to NFPA data cannot be evaluated statistically as error bounds are not available on the NFPA data. Further, given that these are statistical models, as compared to deterministic/physical models, there is no way to develop first principles models for these results. The numbers presented represent point estimates. A detailed evaluation of the effects the deviations between the NFIRS derived data and the NFPA data is provided in the following sections. This evaluation process will be shown to bound the results.

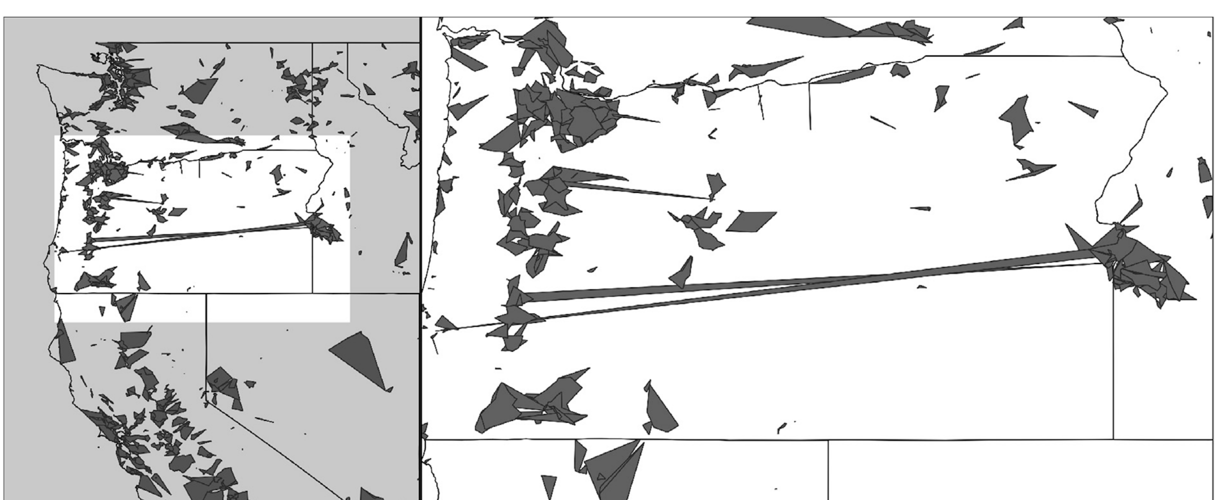


Fig. 9. Map of estimated fire department jurisdictions in Oregon, displaying two fire department's whose estimates feature large spatial outliers. Examination of the outlier fires indicates no obvious reason to disregard their validity aside from the apparent absurdity of the outcome.

3.1. National NFIRS reporting fire departments compared against national estimates of fire

To obtain comparable estimates to NFPA's national loss reports, the following approximations were assumed. First, an individual fire department's census tract coverage, as estimated by the methodology, was assumed constant over the 2002–2012 period. However, there is some variation in which fire departments reported to NFIRS in a given year, and thus the total population "covered" by NFIRS reporting fire departments does fluctuate year-by-year. These fluctuations were accounted for by identifying, on a yearly basis, fire departments that submitted at least one fire incident to the NFIRS database. Failure to submit an incident meant a given department's estimated coverage was not included in a given year's estimates. Second, at the census tract level that this methodology operates, only 5-year American Community Survey estimates are available and only comprehensively for 4 years (2005–2009 estimates are missing some census tracts). The American Community survey is an ongoing survey that progressively samples a selection of the American population yearly with more detailed questions than the decennial census includes. As a result, it also provides updates to the decennial census's demographic and population information during interim years between censuses. Since a 5-year ACS estimate represents the population in an area averaged over 5 years, the estimates are assumed to be a reasonable representation of the middle year, owing to the fact that, in aggregate, the population of the U.S. tends to grow over time, and thus the 5-year average is likely close to the median year populations for most communities not undergoing tremendously swift populations swings. Using these approximations, estimates of the percentage of population covered by NFIRS reporting fire departments were computed on the basis of census tracts for the years 2005–2009.

In order to bound how far estimates using census tracts can deviate from estimates using the far more resolved, but computationally intensive census block calculation, population percentage coverage estimates were also calculated for NFIRS reporting fire departments on the basis of census blocks using 2010 census block data. Because census block data are only available during decennial census years, the percentage of population covered by geocoded NFIRS reporting fire departments was computed on the basis of census block data across the years 2002–2012.

Fig. 10 shows that a finely resolved (i.e. census block) estimate of the U.S. NFIRS geocoded protected population is approximately 70%. Fig. 10 also compares the census block and census tract estimates. As one might expect given the illustration in Fig. 7, the census tracts approach estimates a population coverage that is consistently about 10% higher than the estimates derived using census blocks. It may seem surprising that the discrepancy is only roughly 10%. The discrepancy is low likely because sizeable departures in estimation can generally only occur in areas of the nation where a small, rural department is the sole fire department within a very large census tract, and even then the error is capped at the general aggregation limit of census tracts, about 3000 to 5000 individuals.

In an attempt to supply the most finely resolved estimates possible, subsequent comparisons between this methodology and NFPA national estimates are performed using the census block estimates.

Using the population coverage estimates derived from the methodology on the basis of census blocks, it was possible to calculate incident, civilian fatality, and civilian injury rates for the population of the U.S. protected by geocoded NFIRS reporting fire departments. Fig. 11 displays the comparison of these rates to the NFPA national rates as reported by the NFPA for 2002–2012.

Notable from Fig. 11 is that the estimates of incident, fatality, and injury rates in the NFIRS population are uniformly lower than the NFPA national estimates by a fair margin. There are two meaningful explanations for this observation. First, the populations of those communities protected by non-geocoded NFIRS reporting fire departments experience substantially higher incident, fatality, and injury rates than those that report to NFIRS. This notion has some truth to it, as Hall even noted in Ref. [2] that "Participation rates, therefore, are not necessarily uniform across regions and sizes of community." This assertion has been explored by various authors, and Butry and Thomas [3] identified differences in fire risk factors between NFIRS reporting and non-reporting cities, noting that such differences could easily lead to differences in fire problems.

A second explanation might be that this methodology, while capturing *population* coverage well for NFIRS reporting fire departments, does not account for the possibility that not all *fires*, *fatalities*, and *injuries* reported in these "covered" regions are accounted for. That is, there may either be non-reporting fire departments whose areas of influence lie inside or overlap jurisdictional boundaries of other fire departments using this paper's methodology, or reporting fire departments may fail to exhaustively report the full extent of their incidents, fatalities, and injuries. Imagine for example that none of the smaller fire departments in Fig. 6 reported their fires, fatalities, or injuries to NFIRS. Such a situation would indicate that the *population* of the City of Houston, while "covered" by an NFIRS reporting jurisdiction, is not fully *reporting* all of its fire incidents, fatalities, and injuries due to the NFIRS non-reporting fire departments that also cover portions of said population. Likewise, Houston Fire Department may simply fail to report all of its data to the NFIRS database, but instead only submit a portion of it.

The following two subsections explore these assertions in more detail, utilizing additional NFPA and census resources to determine how much validity each viewpoint may have.

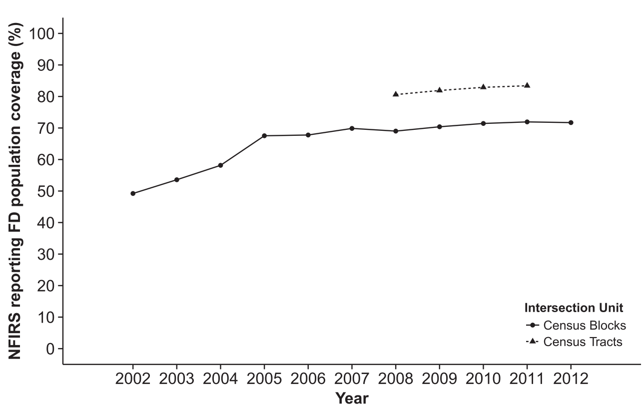


Fig. 10. Comparison of percentage of U.S. population covered by Geocoded NFIRS reporting fire departments estimated using census tracts or census blocks as the basis for intersection with FD influence polygons, for 2002–2012. Note that the 2010 U.S. block populations were used to compute census block estimates for the 2002–2012 period, while appropriate 5-year ACS estimates (available for 4 years) were used for the census tract estimates.

3.1.1. Non-reporting NFIRS population estimates

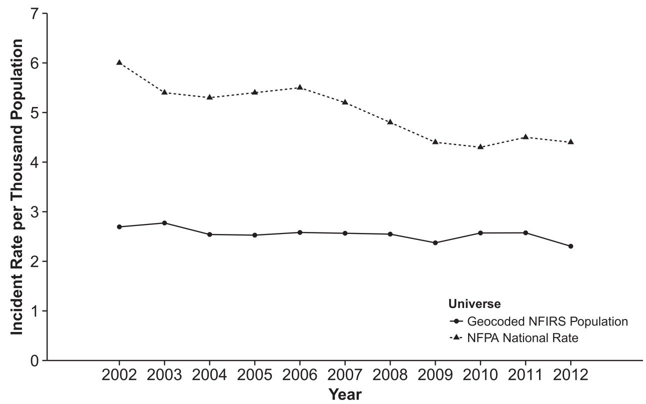
One may note, from Fig. 11 that these estimates, if valid, accurately describe the fire problem experienced by the geocoded NFIRS reporting and NFPA populations. If NFPA estimates are the complete picture of the U.S. fire problem, it would be possible to derive the incident, fatality, and injury rates of the non-reporting NFIRS population using knowledge of the NFIRS reporting rates and their population coverage. This derivation is outlined below:

$$R_{National} = p_{NFIRS} \times R_{NFIRS} + (1 - p_{NFIRS}) \times R_{non-reporting}$$

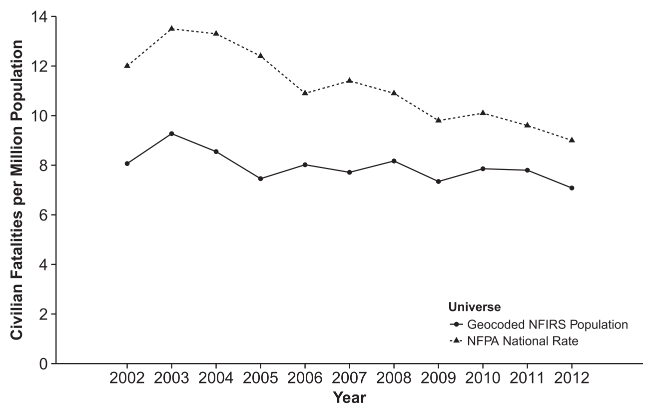
$$R_{non-reporting} = \frac{R_{National} - p_{NFIRS} \times R_{NFIRS}}{1 - p_{NFIRS}}$$

where R represents a rate of interest for the subtexted population, and p_{NFIRS} represents the proportion of the population belonging to geocoded NFIRS reporting fire departments, from Fig. 10.

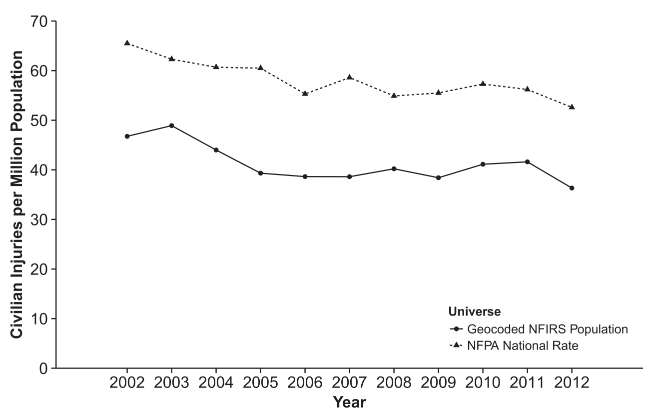
Using this formula, estimates of incident, fatality, and injury rates for the non-reporting NFIRS populations were constructed. If one believes the assertions made by Hall that smaller communities are less likely to



(a) Incident Rates per thousand population



(b) Civilian Fatality Rates per million population



(c) Civilian Injury Rates per million population

Fig. 11. Incident, Civilian Fatality, and Civilian Injury rates for Geocoded NFIRS reporting and national populations for 2002–2012.

report, then it could be expected that these "non-reporting" population estimates should be, in a sense, some weighted combination of the fire problems of smaller communities. In other words,

$$R_{non-reporting} = \sum_{i=1}^{n} (p_i R_i) \quad (1)$$

where p_i is the proportion and R_i is the rate of non-reporting population covered by communities within size range i. The NFPA outlines nine such size ranges for community population protected by fire departments. the smallest four ranges pertain to communities of 25,000 persons or less. For purposes of comparison, the minimum and maximum estimates of incident, fatality, and injury rates among these bottom four community sizes were obtained for the years 2003–2012 from Ref. [8] and its 9 preceding annual reports. Mathematically, these estimates are calculated as follows:

$$\text{NFPA High} = \max\left(R_{pop<2500}, R_{2500\leq pop<4999}, R_{5000\leq pop<9999}, R_{10000\leq pop<25000}\right)$$

$$\text{NFPA Low} = \min\left(R_{pop<2500}, R_{2500\leq pop<4999}, R_{5000\leq pop<9999}, R_{10000\leq pop<25000}\right)$$

where $R_{pop<2500}$ would for example refer to the relevant NFPA estimate for a community with a population of less than 2500 persons. Thus these "NFPA High" and "NFPA Low" estimates represent bounds on the incident/fatality/injury rates for "smaller" communities of less than 25,000 persons for a given year.

Fig. 12 displays the results of this comparison. It is notable that the non-reporting population estimate for incident rate falls nicely between the high and low NFPA estimates, and in fact seems to follow the general trend of the high estimate, which for this subfigure represents communities under 2500 persons. In other words, for this subfigure, the "NFPA High" calculation chose $R_{pop<2500}$ every year. Since this trend logically obeys the constraints of equation (1), one can assert that incident rate can be explained by incident non-reporting. The non-reporting population rate calculation depends only on knowledge of NFPA's national rates, and thus this sort of agreement is promising. This agreement begins to break down for fatalities, with three years of points falling outside of the NFPA's estimates for given community sizes, and breaks down completely for injuries, where the non-reporting rate estimate completely violates the NFPA's upper bound for these communities. It should be noted that discrepancies in fatality rates between NFIRS and NFPA may be attributed to differences in reporting practices. For example, the NFPA follows up on vehicle fire deaths to ensure that the deaths were the result of a fire and not physical trauma. NFIRS does not always likewise follow up, thus generating a discrepancy between the two data sources.

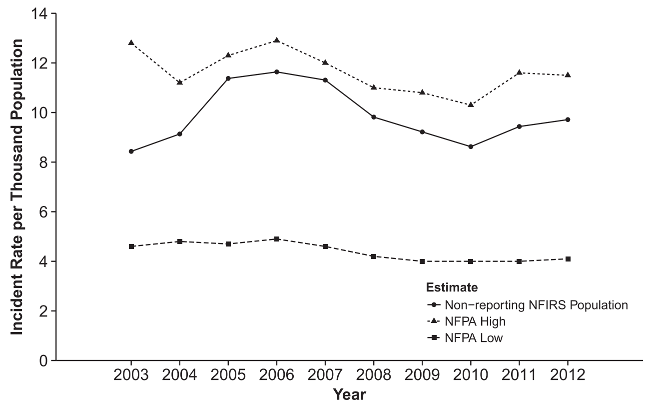
Regardless of underlying cause, the fatality and injury subfigures in Fig. 12 present encouraging information. They tend to respect the general trends of the upper and lower bounds. Thus, they appear to indicate that, at least for fatalities and injuries, there is likely general underreporting from these "covered" populations, which, if corrected for, could bring the NFIRS non-reporting population estimates in line with NFPA estimates. Indeed, this underreporting concept is further explored in the following section.

3.1.2. Underreporting of NFIRS covered population estimates

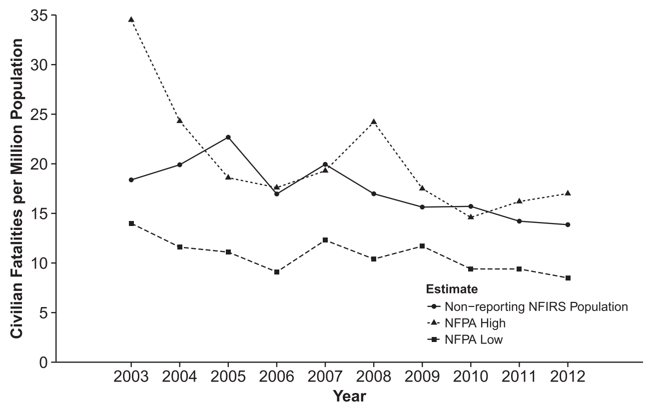
A potential issue outlined above in the methodology is that there may be underreporting from fire departments that lie within these "geocoded NFIRS covered" populations. Such an issue can be difficult to test, since NFIRS naturally does not contain any information from fire departments that do not report to it. However, with the help of NFPA estimates and census population estimates of communities, it may be possible to obtain at least a sense of the underreporting within this methodology's NFIRS protected populations.

The approach adopted to test for underreporting is to map the NFPA's national community size rate estimates to just the NFIRS reporting population, by attempting to quantify the proportion of non-reporting NFIRS population in the U.S. at various community size categories.

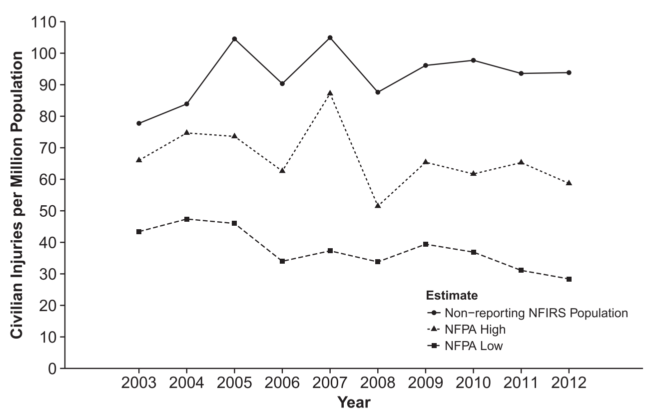
There are three major pieces of information necessary to perform this calculation, outlined in Table 1. The first is an estimate of the percentage of U.S. population that lies within each given community population range estimated by the NFPA in Ref. [8]. It should be noted that the population reported in Table 1 is calculated as the union of all incorporated places, and minor civil divisions (MCDs) in the United States, using [9]. In the U.S., incorporated places are cities, towns, etc. that are self-governed. MCDs are subdivisions within a State's counties, and may include smaller towns and townships that are not incorporated (i.e. not self-governed). Second, the NFPA has published estimates of the number of fire departments protecting various community sizes in the U.S. [10]. Finally, from the NFIRS geocoding methodology, one can obtain



(a) Incident Rates per thousand population



(b) Civilian Fatality Rates per million population



(c) Civilian Injury Rates per million population

Fig. 12. Incident, civilian fatality, and civilian injury rates for NFIRS non-reporting populations for 2003–2012. NFPA High and NFPA low refer to the highest and lowest rates observed in the bottom 4 NFPA size categories, representing populations below 25,000 persons.

estimates of fire departments protecting various community sizes. While, as mentioned earlier, these estimates are expected to be inflated upward for large area coverage fire departments like county fire departments, the main interest lies in the number of smaller fire departments that protect communities of 24,999 or fewer persons. While some fire departments that would otherwise belong in these categories are likely pushed upward into higher categories by the limitations of the geocoding and intersection methodology, it is unlikely to tremendously affect the results of this analysis.

Note in Table 1 that the number of actual U.S. communities differs from the estimates of FDs protecting various sized communities. This departure occurs because fire departments are not necessarily linked one-to-one with U.S. communities. Some counties and geographic areas utilize fire and rescue districts to cover multiple towns and portions of a county with a single fire department. Hence, it is possible to have 17 fire departments that protect populations over 1,000,000 persons, but only 10 incorporated places in the United States that technically have over 1,000,000 persons. Another note for Table 1 is that not all of the population in the United States belongs to an incorporated place or MCD. Indeed, roughly 10% of the U.S. population is not accounted for by incorporated places or MCDs. For the purposes of this analysis, this “remainder” population is considered so rural as to not belong to a community, and thus is not considered as protected by fire departments.

The “Estimated reporting rate,” (r_i), and “Estimated reporting population” (p_i), columns of Table 1 were constructed according to the following calculations:

$$r_i = \min\left(\frac{FD_{NFIRS}}{FD_{NFPA}}, 1\right)$$

$$p_{i,adj} = r_i \times p_i$$

where r_i represents the estimated percentage of NFIRS reporting fire departments from communities with population range i, p_i represents the percentage of U.S. population belonging to communities with population range i, and $p_{i,adj}$ represents the percentage of U.S. population belonging to communities with population range i that are covered by NFIRS reporting fire departments. The calculation of r_i involves some assumptions. First, since the NFIRS FD field uniformly overestimates the number of FDs in larger population categories compared to the NFPA, it is assumed that the actual NFPA surveyed fire departments are a subset of this overestimate. This assumption applies to the population ranges of 25,000 persons and above. For populations below 25,000, the NFPA estimate is assumed to be the total universe of FDs in the nation, of which the NFIRS FDs are a subset. Finally, the NFPA identified FDs are assumed to cover the entirety of the population within a given population range. It is encouraging to note that despite all these assumptions, the percentage of U.S. population covered by “geocoded NFIRS-reporting” FDs is estimated at roughly 76%, which falls within the region bracketed by the census tract and census block approaches as depicted in Fig. 10.

With the estimate of the percentage of the U.S. population that is covered by NFIRS reporting fire departments at various community sizes from Table 1, it is possible to use the NFPA's estimates of incident, fatality, and injury rates for communities of various population ranges in conjunction with the NFIRS reporting population covered by these community ranges to derive an “NFPA estimate” of incident, fatality, and injury rates for NFIRS reporting fire departments. Table 2 supplies these estimates and the outcomes of the following calculation:

$$R_{NFPA,adj} = \sum_{i=1}^{n} (p_{i,adj} R_i)$$

where $R_{NFPA,adj}$ is the rate of interest estimated by the NFPA for the “geocoded-NFIRS-adjusted” population. R_i is the rate of interest estimated by the NFPA for communities within population range i, as presented in the rate columns in Table 2, and the sum is over the i population ranges presented in both Tables 1 and 2.

Examining Table 2, one can observe the approximate underreporting in the NFIRS population from the methodology compared against NFPA estimates on the same population. For incident rates and injury rates, the differences of 0.45 incidents per thousand population and 8.1 injuries per million population would correspond to an underreporting of about 138,000 fire incidents and 2413 injuries respectively. Interestingly, the calculations show that fatality rate is “over-reported” for the population covered compared to NFPA estimates, corresponding to approximately 106 more fatalities reported by the NFIRS-geocoded population than

Table 1
Relevant metrics for estimating NFIRS reporting population coverage for the year 2012. NFPA estimates from Ref. [10], population and community information calculated from Ref. [9].

Population of Community	U.S. Population (%) [9] (p_i)	Community (#) [9]	NFPA FDs (#) [10] (FD_{NFPA})	NFIRS FDs (#) (FD_{NFIRS})	Estimated NFIRS reporting rate (r_i)	Estimated Reporting Population (%) ($p_{i,adj}$)
1,000,000 or more	8.5	10	17	19	1	8.5
500,000 to 999,999	6.1	28	40	51	1	6.1
250,000 to 499,999	5.85	52	62	160	1	5.85
100,000 to 249,999	12.06	250	268	620	1	12.06
50,000 to 99,999	13.49	603	530	1049	1	13.49
25,000 to 49,999	12.66	1136	1318	1715	1	12.66
10,000 to 24,999	13.32	2633	3567	3004	0.84	11.22
5000 to 9999	6.85	2991	4384	2207	0.5	3.45
2500 to 4999	4.92	4286	5807	1664	0.29	1.41
Under 2500	6.58	28684	14059	2525	0.18	1.18
Total	90.33	40673	30052	13014	–	75.92

expected by NFPA estimates. These raw estimates were derived in the following manner:

$$\left|\text{Raw Difference}\right| = \frac{N_{U.S.}}{\text{offset}}\left(p_{NFPA,adj}R_{NFPA,adj} - p_{NFIRS}R_{NFIRS}\right)$$

where $N_{U.S.}$ is the 2012 U.S. population and offset is the adjustment for the "per thousand" and "per million" population bases of the three rates of interest.

The results from Fig. 12 and Table 2 examine the potential issues with the methodology as though it had to be either one explanation or the other. Realistically, however, the answer likely lies somewhere in the middle. One could potentially, for example, define a uniform under-reporting rate as the number of incidents, fatalities, and casualties that must be added to the methodology's estimates in order to constrain the estimates of Fig. 12 to fall between the appropriate community brackets. However, it is unclear how one might validate the results of such an operation without additional information since it would then be using the NFPA's community estimates as constraints.

Table 2
NFPA community size estimates from Ref. [8], and "geocoded-NFIRS-adjusted" NFPA estimates of incident, fatality, and injury rates, compared against actual NFIRS-geocoded estimates for the year 2012. The NFIRS geocoded estimate row is previously calculated and shown in Fig. 11.

Population of Community	Estimated U.S. Reporting Population (%)	Incidents per thousand population	Fatalities per million population	Injuries per million population
1,000,000 or more	8.5	3.0	6.6	54.4
500,000 to 999,999	6.1	3.0	6.6	54.4
250,000 to 499,999	5.85	3.3	7.4	62.9
100,000 to 249,999	12.06	3.3	8.5	61.0
50,000 to 99,999	13.49	3.3	6.6	64.7
25,000 to 49,999	12.66	3.3	8.7	64.0
10,000 to 24,999	11.22	4.1	8.5	58.7
5000 to 9999	3.45	4.8	13.2	36.5
2500 to 4999	1.41	7.0	17.0	28.4
Under 2500	1.18	11.5	15.6	30.4
NFPA adjusted estimate	75.92	2.75	6.24	44.4
NFIRS geocoded estimate	71.7	2.30	7.08	36.3

3.2. Regional fire problems compared against NFPA regional estimates

It was considered prudent to also compare the results of this methodology at the regional level with NFPA estimates. The relevance of such a comparison lay in examining whether NFIRS reporting fire departments exhibited the same regional trends as NFPA regional estimates reported for given regions.

Fig. 13 provides an examination of yearly population coverage from NFIRS reporting fire departments from 2002 to 2012, on the basis of census blocks. As indicated by the figure, reporting coverage has improved over time. It was suggested through discussions with personnel at the California All Incident Reporting System (CAIRS), which collects NFIRS format fire incident data for the California State Fire Marshall [11], that the State of California did not widely report to NFIRS the fire statistics it was collecting until 2007. It was surmised that data back-logged to 2005 was input into the national database. Such an update likely explains why there is a large jump in participation seen in the western region between 2004 and 2005. The majority of California NFIRS data starts in 2005.

Figs. 14–16 provide a yearly regional breakdown of the national overview provided earlier by 11. it is encouraging to note the relative stability of the estimates from year to year, despite changes in coverage of reporting populations over the course of those years.

Examining Fig. 14, general trend agreement is present for the northeast, south, and west. There appears to be a mismatch between the midwest incident rates, with the NFPA estimates indicating some upturns in incident rates during two periods in the timeline.

Examining Fig. 15, the variability of the NFPA estimates becomes much higher owing to the rarity of fire fatalities. In this sense the stability of the NFIRS-geocoded estimates is interesting, as it indicates that reporting of fatalities to the NFIRS database is relatively consistent from year to year

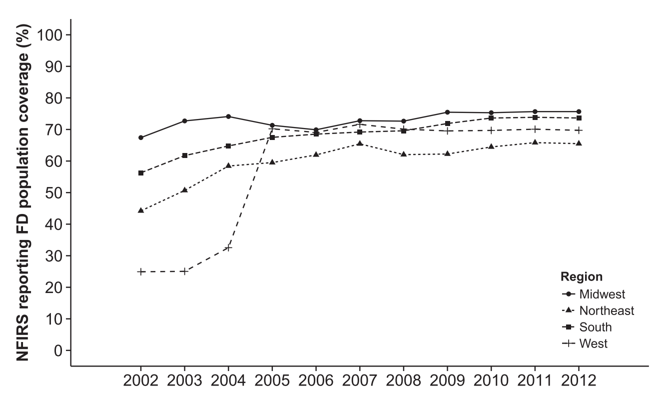


Fig. 13. Estimates of population covered for the northeast, midwest, south, and west regions of the U.S., 2002–2012.

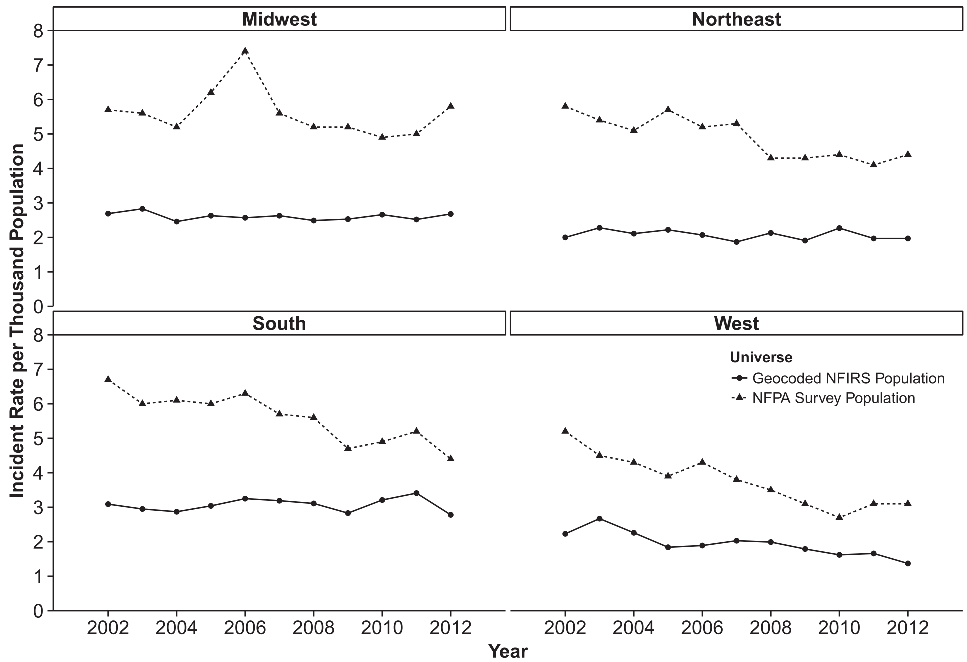


Fig. 14. Incidents per thousand population for Geocoded NFIRS reporting and NFPA survey populations by region for 2002–2012. National estimates obtained from Refs. [8] and [12].

across regions. Also notable in Fig. 15 is that there is one year in the northeast and two years in the west where the fatality rate estimates for the NFPA and the NFIRS-geocoded estimates coincide. For the Northeast this would appear to be something of an outlier but the estimates in the West appear to steadily converge as reporting from NFIRS increases in the region. This tendency is also observed for the midwest and the south.

Examining Fig. 16, agreement in trends between the midwest, northeast, and the south are all quite good. Notably, the injury rates in the west appear to be impacted by the sudden increase in reporting that occurred from 2003 to 2004, marked by a sudden drop in injury rate during that period. Afterward, the injury rate in the west is largely stagnant, a trend roughly reflected in the NFPA national estimates, though not as strongly.

Overall, the geocoded NFIRS estimates at the region resolution appear to be consistent, though lower compared against the NFPA estimates for the national rates. This is not entirely unexpected given the prior analyses of the national data, and in fact the increased volatility seen in the NFPA estimates could likely be explained by the fact that they include all of the

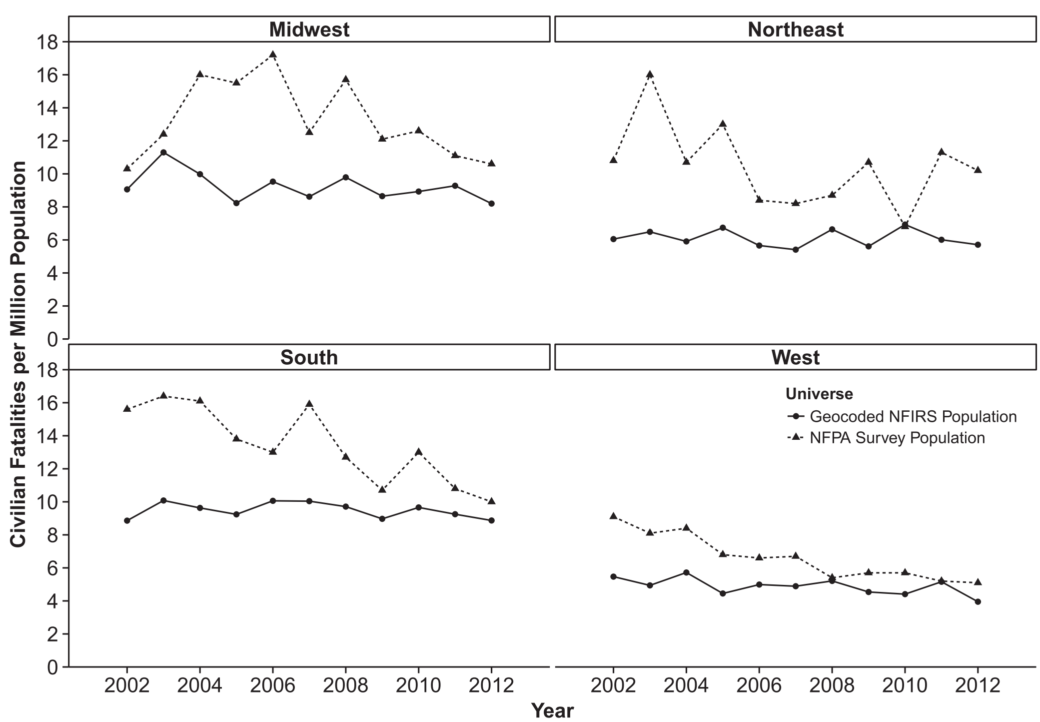


Fig. 15. Civilian fatalities per million population for geocoded NFIRS reporting and national populations by region for 2002–2012. National estimates obtained from Refs. [8] and [12].

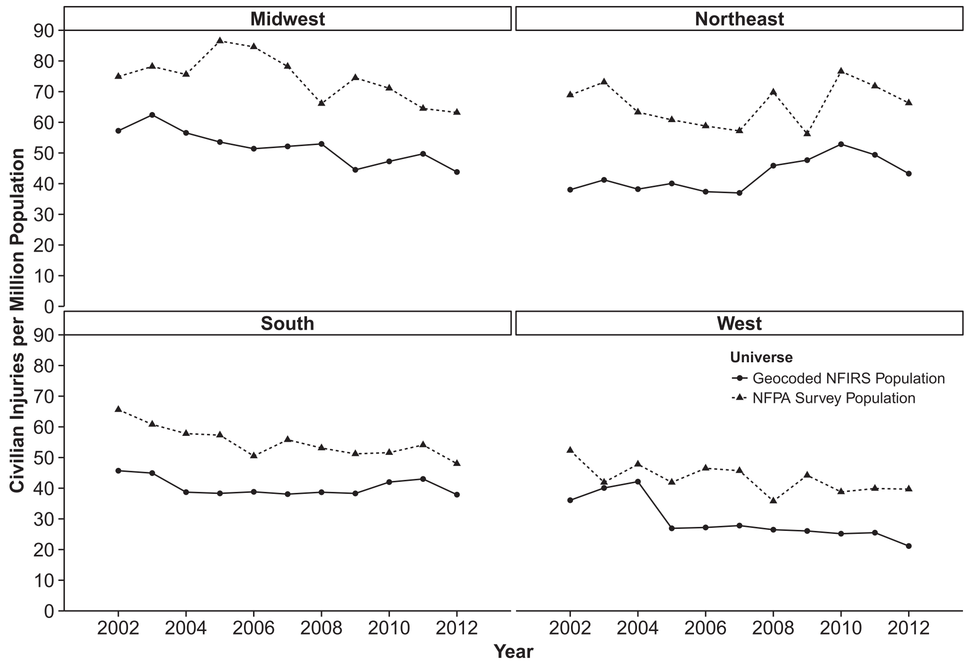


Fig. 16. Civilian injuries per million population for geocoded NFIRS reporting and national populations by region for 2002–2012. National estimates obtained from Refs. [8] and [12].

rural population of the U.S., which tends to have the most volatility from year to year. Thus, it seems that the apparent stability of the NFIRS-geocoded estimates is perhaps just an indication of the stability of the fire problems in their underlying larger community base. Such an assertion is certainly backed by examination of Table 2, where one can see that, especially for fatalities, there are huge differences in the NFPA's estimation of fatality rates between the 25,000 persons or above communities that appear to be well represented in NFIRS and those communities with less than 25,000 persons that have substantially less representation.

4. Conclusions

The methodology presented in this paper supplies estimates of the population protected by NFIRS reporting fire departments utilizing a process involving geocoding, generation of polygons using concave hulls, and spatial joining of said polygons to census data at the desired resolution. These estimates are critical to most analyses of NFIRS data that would rely on any form of rate calculations, such as incident, fatality, or injury rates, because it is necessary to know the universe from which NFIRS data is being drawn. This universe in the past has not been determined due to the incomplete reporting in NFIRS, and thus direct analysis of the data has been invalid as an approach. The supplied methodology addresses this major issue, and arms future analysts of fire statistics with information needed to avoid major statistical pitfalls in their analyses. Additionally, by directly exposing the population protected by NFIRS reporting fire departments, this methodology meets a secondary need for certain forms of analysis, that of outlining all geographic areas of the U.S. that are not covered by NFIRS reporting fire departments. Such information opens up opportunities for analyses related to examining the fire problems in non-NFIRS reporting populations, of which the analysis in Section 3.1.1 is an example.

However, this methodology is not without its weaknesses at present. As discussed earlier, it is highly dependent on the gathered data, and certain fire departments reported 3 or fewer structure fire incidents even in an 8 year period, leading these departments to be either excluded or very coarsely defined. Such issues could possibly be overcome by using the much richer EMS data available from fire departments. Unfortunately, as of 2004, the USFA no longer includes EMS call data in their PDRs, due in part to the sheer volume of calls, so the data would need to be explicitly requested from them.

Another problem discussed is that of truly representing population covered at a jurisdictional level. As noted in Fig. 6, many different fire department jurisdictions can overlap and interact at the individual jurisdiction level. Thus, to obtain clear population covered estimates for individual fire department jurisdictions, it is necessary to account for situations where a “larger” jurisdiction, e.g. a county level jurisdiction, contains a large metropolitan city, but does not cover that population due to the presence of that metropolitan city's jurisdiction handling said population.

In spite of these issues, the estimates produced by this methodology appear largely consistent with the NFPA's actual estimates. Using the NFPA's national estimates in conjunction with this methodology's estimates, it was possible to produce a secondary estimate of the fire problem of the non-reporting NFIRS population of the U.S., as displayed in Fig. 12. When compared against the NFPA's community size incident, fatality, and injury rates, these secondary estimates of the non-reporting population appear feasible, excluding non-reporting population injury rates, which are perhaps more heavily linked to underreporting of injuries among the NFIRS-geocoded population. In this sense, this analysis serves as a validated confirmation of the notions of Hall [2], and Butry and Thomas [3], that the fire problem of the non-reporting NFIRS population is different than that of the reporting NFIRS population, at least for incident and fatality rates.

Possible underreporting from the NFIRS-geocoded population estimated by this methodology was also examined by using this methodology's community population protected estimates to obtain an NFIRS reporting FD rate for various community sizes. This reporting rate was used in conjunction with an estimate of the U.S. population living in each of these community size categories to obtain an approximation of the fire problem of the NFIRS reporting fire department population using NFPA estimates. From this analysis, the potential under or over-reporting of the population evaluated from this methodology was quantified in Table 2.

Finally, this methodology appears consistent even when examining finer resolutions such as regions, and produces regional incident, fatality, and injury estimates that mostly agree in trend with the NFPA's regional estimates for these locations. The apparent lack of stability in the NFPA's estimates for these regions compared against this methodology's estimates could largely be attributed to their full inclusion of the smaller community sizes, which tend to exhibit higher volatility in rates from year to year.

Acknowledgements

This work was funded by the Federal Emergency Management Agency's Assistance to Fire-Fighters Grant Program under grant number AFG-EMW-2015-FP-00312.

References

[1] National Commission on Fire Prevention and Control, America Burning, 1973.
[2] J.R. Hall, B. Harwood, The National Estimates Approach to U.S. Fire Statistics, Fire Technology, May 1989.
[3] D.T. Butry, D.S. Thomas, Evaluating Potential Bias in Non-randomly Reported Fire Incident Data, September 2012. NIST TN 1770, NIST.
[4] U.S. Fire Administration, NFIRS Representativeness Study, U.S. FEMA, August 2008.
[5] D. Miller, 2009-2011 Residential Fire Loss Estimates, U.S. CPSC, July 2013.
[6] U.S. Fire Administration National Fire Data Center, National Estimates Methodology for Building Fires and Losses, U.S. FEMA, August 2012.
[7] H. Edelsbrunner, D.G. Kirkpatrick, S. Raimund, On the shape of a set of points in the plane, IEEE Trans. Inf. Theory 29 (4) (July 1983) 551–559.
[8] M.J. Karter, Fire Loss in the United States during 2012, NFPA, September 2013.
[9] Population Estimates Program, Incorporated Places and Minor Civil Divisions Datasets: Subcounty Resident Population Estimates: April 1, 2010 to July 1, 2013, U.S. Census Bureau, 2013. URL, https://www.census.gov/popest/data/cities/totals/2013/SUB-EST2013.html. (Accessed 29 September 2015).
[10] H.J.G. Haynes, G.P. Stein, US Fire Department Profile 2013, NFPA, November 2014.
[11] Office of the State Fire Marshall, California All Incident Reporting System (CAIRS), 2013. URL, http://osfm.fire.ca.gov/cairs/cairs. (Accessed 18 July 2016).
[12] M.J. Karter, U.S. Fire Experience by Region, NFPA, January 2013.