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

A B S T R A C T

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

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
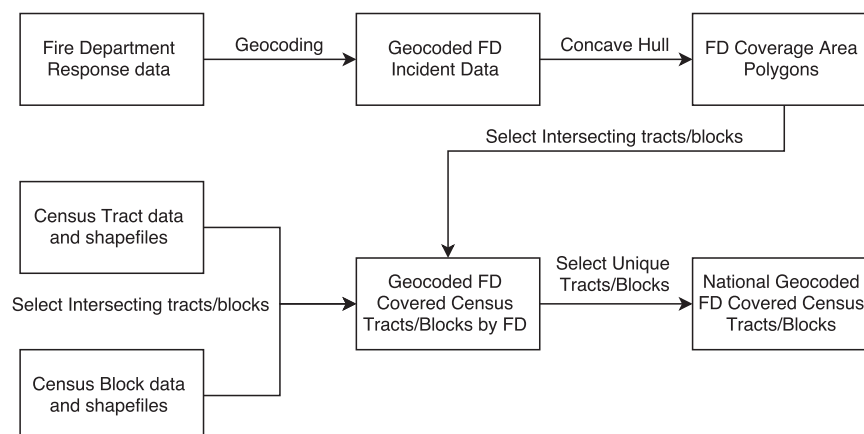

Fig. 1. Summary of NFIRS geocoding methodology.