Sources and implications of bias and uncertainty in a century of US wildfire activity data

Karen C. Short

USDA Forest Service, Missoula Fire Sciences Laboratory, 5775 US Hwy 10 W, Missoula, Montana, 59808, USA, kshort@fs.fed.us

Abstract
Analyses to identify and relate trends in wildfire activity to factors such as climate, population, land use/land cover, and wildland fire policy are increasingly popular in the United States (US). There is a wealth of US wildfire activity data available for such analyses, but users must be aware of inherent reporting biases, inconsistencies, and uncertainty in the data in order to maximize the integrity and utility of their work. Data for analysis are generally acquired from archival summary reports of the federal or interagency fire organizations; incident-level wildfire reporting systems of the federal, state, and local fire services; and, increasingly, remote-sensing programs. This paper provides an overview of each of these sources and the major reporting biases, inconsistencies, and uncertainty within them. Use of the national fire reporting systems of state and local fire organizations has been rising in recent decades, providing an improved set of incident-level data for all-lands analyses of wildfire activity. A recent effort to acquire, standardize, error-check, compile, scrub, and evaluate the completeness of US federal, state, and local wildfire records from 1992-2012 has been completed for the national Fire Program Analysis (FPA) application. The resulting FPA Fire-Occurrence Database (FPA FOD) currently includes nearly 1.7 million records from the 21-year period, 1992-2012, with values for at least the following core data elements: location (fire origin) at least as precise as a Public Land Survey System section (2.6 km² grid), discovery date, and final fire size. The FPA FOD is publicly available from the US Forest Service Research Data Archive (http://dx.doi.org/10.2737/RDS-2013-0009.2). While necessarily incomplete in some aspects, the database is intended to facilitate fairly high-resolution geospatial analysis of US wildfire activity over recent decades.

Keywords: wildfire occurrence, reporting, data, statistics, bias, uncertainty, USA

1. Introduction

The statistical analysis of wildfire activity is a critical component of national wildfire planning, operations, and research in the United States (US). Wildfire activity data have been collected in the US for over a century. Yet, to this day, no single, unified system of wildfire record-keeping exists. Data for analysis are generally harvested from archival summary reports of the federal or interagency fire organizations; incident-level wildfire reporting systems of the federal, state, and local fire services; and, increasingly, remote-sensing programs. It is typical for research into wildfire activity patterns for all or part of the last century to require data from several of these sources and perhaps others. That work is complicated by the disunity of the various datasets and potentially compromised by inherent reporting biases, inconsistencies, and errors or uncertainty in the data, as described here.

2. Data sources

Data for analyses of variables like wildfire numbers and area burned in the US are available in various forms from the early 20th century to the present. Figure 1 shows the general temporal coverage of key data sources, grouped into three categories: (1) archival summary reports, (2) documentary fire records, and (3) remotely sensed data.