Connecting smoke plumes to sources using Hazard Mapping System (HMS) smoke and fire location data over North America

Steven J. Brey[1], Mark Ruminski[2], Samuel A. Atwood[1], Emily V. Fischer[1]

[1]Atmospheric Science, Colorado State University, Fort Collins, 80523, U.S.
[2]NOAA/NESDIS Satellite Analysis Branch, College Park, 20740, U.S.

Correspondence to: Steven J. Brey (sjbrey@rams.colostate.edu)

Abstract. Fires represent an air quality challenge because they are large, dynamic and transient sources of particulate matter and ozone precursors. Transported smoke can deteriorate air quality over large regions. Fire severity and frequency are likely to increase in the future, exacerbating an existing problem. Using the National Environmental Satellite, Data and Information Service (NESDIS) Hazard Mapping System (HMS) smoke data for North America for the period 2007 to 2014, we examine a subset of fires that are confirmed to have produced sufficient smoke to warrant the initiation of a U.S. National Weather Service smoke forecast. We find that gridded HMS analyzed fires are well correlated (r = 0.84) with emissions from the Global Fire Emissions Inventory Database 4s (GFED4s). We define a new metric, smoke hours, by linking observed smoke plumes to active fires using ensembles of forward trajectories. This work shows that the Southwest, Northwest, and Northwest Territories initiate the most air quality forecasts, and produce more smoke than any other North American region by measure of the number of HYSPLIT points analyzed, the duration of those HYSPLIT points, and the total number of smoke hours produced. The average number of days with smoke plumes overhead is largest over the north-central U.S. Only Alaska, the Northwest, the Southwest, and Southeast U.S. regions produce the majority of smoke plumes observed over their own borders. This work moves a new dataset from a daily operational setting to a research context, and it demonstrates how changes to the frequency or intensity of fires in the western U.S. could impact other regions.

1 Introduction

North American fires represent a major source of atmospheric pollutants (Wiedinmyer et al., 2006), and they contribute to elevated ground level ozone (O_3) and fine particulate matter ($PM_{2.5}$) in the U.S. (Baker et al., 2016; Brey and Fischer, 2016; Jaffe et al., 2008; Park et al., 2007; Saide et al., 2015). Exposure to wildfire smoke has been shown to have negative impacts on respiratory and cardiovascular health (Dennekamp and Carey, 2010; Haikerwal et al., 2015; Rappold et al., 2011). The relative importance of the contribution of smoke, particularly wildfire smoke, as both an O_3 precursor and as a direct source of $PM_{2.5}$ will grow as anthropogenic emissions decline (Russell et al., 2012; Val Martin et al., 2015). In addition to air quality considerations, smoke is also thought to make a major contribution to absorbing aerosols observed throughout the

troposphere over North America (Forrister et al., 2015; Liu et al., 2015; Liu, 2014; Washenfelder et al., 2015). Thus understanding the sources and prevalence of smoke is relevant for air quality, visibility, and radiative forcing calculations.

Millions of hectares of forest burn in North America each year (Randerson et al., 2012), the area burned by wildfires in the western U.S. has increased (Westerling, 2016; Westerling et al., 2006), and models predict this trend will continue in a warmer climate (Hurteau et al., 2014; Keywood et al., 2013; Randerson et al., 2012; Scholze et al., 2006; Yue et al., 2013). Agricultural fires are also a significant source of smoke in North America. Over 1.2 million hectares of cropland are burned in the contiguous U.S. each year (McCarty et al., 2009). Though often smaller in size and shorter in duration, these fires have been shown to increase the abundance of gas and aerosol species that can deteriorate air quality (Akagi et al., 2011; Dennis et al., 2002; Jimenez et al., 2006). There are a growing number of case studies documenting instances of smoke traveling thousands of kilometers to affect atmospheric composition far downwind from the fire locations (Baker et al., 2016; Creamean et al., 2016; Miller et al., 2011; Pfister et al., 2008; Stein et al., 2009; Val Martin et al., 2013a). Case studies demonstrate that regional and long-range transport of smoke causes elevated column and surface concentrations of aerosols and trace gases over extensive (continental) regions of the U.S., and the smoke can have implications for atmospheric composition over relatively long temporal scales (weeks) (Creamean et al., 2016; Park et al., 2003, 2007; Val Martin et al., 2013a; Vedal and Dutton, 2006).

In this paper, we show how smoke data from the National Environmental Satellite, Data and Information Service (NESDIS) Hazard Mapping System (HMS) can be used to provide context for the growing number of case studies linking compromised air quality or visibility to smoke. As an introductory example, Fig. 1 shows that during summer (June – September) over the Contiguous U.S. North Dakota, South Dakota, and Minnesota have more smoke plumes overhead than any other states, averaging between 8 and 12 days per month between June and September. The location with the most smoke overhead is approximately co-located with Fargo North Dakota. Figure 1 also incorporates data from $PM_{2.5}$ monitors. Fargo North Dakota lacks continuous ground level $PM_{2.5}$ observations. Nearby monitors in the north east, south east and west of Minnesota show that approximately 10% of these plume days are days where ground level $PM_{2.5}$ is one standard deviation above average summertime concentrations, approximately one day per summer month in these areas. In contrast, Washington and Oregon, both states with major summertime wildfire activity, average six plume days per month. On ~30% of days with smoke plumes in the atmospheric column (~two days per month), ground level $PM_{2.5}$ concentrations are one standard deviation above the summertime average. Figure 1 is not intended to validate the HMS smoke product. This figure is meant to set the stage for interpreting the analysis presented in later sections, *i.e.* a smoke plume overhead does not necessarily ensure elevated ground-level $PM_{2.5}$ concentrations.

The aggregate view provided by Fig. 1 is unique because it is observationally grounded. There are a number of challenges associated with using a chemical transport model (CTM) to produce an analogous representation of Fig. 1. Rastigejev et al. (2010) show that CTMs have difficulty representing plume dispersion, but the plan view result of this

process is captured in the HMS smoke plume analysis. There are other factors contributing to the modelling challenge (*e.g.* Lassman et al., 2017 and references within). For example, there have been a number of advances made in estimating the emissions inputs by improving burned area products (*e.g.* Randerson et al., 2012) and combining these with emission factors for a wider range of trace species (*e.g.* Akagi et al., 2011; Wiedinmyer et al., 2011). However, incorporating the full suite of emitted species into models and simulating the rapid chemical evolution of smoke remains a challenge (Alvarado et al., 2015), as is proper treatment of injection height (Paugam et al., 2016) and the timing of emissions (Saide et al., 2015). Models are also subject to uncertainty associated with meteorological inputs (Garcia-Menendez et al., 2013). Finally, running and analyzing a chemical transport model at the fine grid resolution appropriate to simulate all the individual smoke plumes of interest from North American fires over the scale of a decade is currently too computationally expensive to be practical. Thus other lenses are needed to examine how the smoke from North American fires is transported and dispersed in the atmosphere over seasonal and interannual timescales.

The two primary goals of this study are to 1) present the distribution and seasonality of fires that initiate National Weather Service (NWS) smoke forecasts, and 2) develop a regional climatology of smoke transport in the U.S. using operational data from the NESDIS HMS combined with forward trajectory calculations. Based on the subset of fires initiating smoke forecasts, HMS observations of smoke in the atmospheric column and trajectory calculations, we present an estimate of the relative frequency that smoke observed over U.S. regions is associated with fires throughout North America.

2 Data

2.1 Description of operational fire and smoke products

The HMS is an interactive environmental satellite image display and graphical interface system that was developed by NESDIS. Trained satellite analysts use the HMS to generate a daily operational list of fire locations and outline areas of smoke. As a part of this process, analysts manually compare automated fire detections to the mid-wave infrared (MWIR) images used to produce them to ensure each fire exists (Ruminski et al., 2006). Schroeder et al. (2008) assessed the automatic detections and estimated they have a false alarm rate of about (~2%) in the conterminous U.S. As in all satellite-based fire detection systems, small fires are more difficult to detect and are under-reported (*e.g.* Hu et al., 2016). Detections deemed to be false are removed, and fires that are not automatically detected are added manually. Visible satellite imagery is also used by analysts to identify fires that may be too small to be automatically detected, either because they do not produce sufficient heat or because they are obscured by a tree canopy (Rolph et al., 2009). In these cases, a smoke plume may be the only indication of a fire. The number of fire detections added manually can be significant. For example, over 50% of the total fire detections were added manually during a 12-month period in 2002-2003 examined by (Ruminski et al., 2006). Land-cover data and power-plant locations compliment satellite imagery to help HMS analysts confirm whether automatic

detections are fires (Ruminski et al., 2006). The HMS office makes a distinction between all detected fires (hereafter HMS hotspots) and fires an HMS analyst has confirmed to produce a substantial amount of smoke (hereafter HYSPLIT points).

HYSPLIT points are a subset of the HMS hotspots; they are fire detections where an analyst also visually confirms the presence of smoke using visible satellite imagery. HYSPLIT points are human-vetted because they are used to initialize
5   the NWS smoke forecasts (Rolph et al., 2009; Ruminski et al., 2006). Each HYSPLIT point is assigned a latitude, longitude, date, time, and duration. The locations of HYSPLIT points are estimated to be accurate to within 2-3 km. GOES visible band and polar orbiting satellites used by the HMS have 1 km resolution at nadir. The 2-3 km accuracy of the location of HYSPLIT points is an estimate based on the additional uncertainties introduced by navigation errors and loss of resolution for observations made away from nadir. Visible satellite imagery was available every 30 minutes during the study period,
10   HMS analysts estimate that this makes the start time of smoke emissions accurate to within 1-2 hours. However, fires can continue to generate smoke after sunset when visible imagery is no longer available. Single fires that produce notable amounts of smoke are associated with a cluster of co-located, or nearly co-located, HYSPLIT points assigned roughly in proportion to the amount of smoke observed. The intended operational consequence of designating HYSPLIT points this way is to allow the NWS smoke forecast model to generate more smoke for large fires and less smoke for smaller fires.

15   The HMS office does not make any distinctions between sources of smoke. HYSPLIT points can be associated with agricultural burning, prescribed burning, or wildfires (Ruminski et al., 2006). The number of HYSPLIT points analyzed, multiplied by their respective duration (Smoke Production Duration Hours or SPDH) is proportional to visual smoke produced for the land cover in which it is burning, and the analyst determines this relationship. The relationship between smoke quantity (as defined by particulate matter concentration [$\mu g\ m^{-3}$] or kg of emissions for example) and HYSPLIT
20   points is variable and does not account for land cover. Large wildfires are represented by dozens of HYSPLIT points spread over many square kilometers. These are typically in the western U.S., Canada, and Alaska. The start times of HYSPLIT points can vary within the cluster of points. Operationally, this serves to represent the variability in the amount of smoke observed at different times of the day. For example, during a large wildfire event, analysts may create several HYSPLIT points with a 24-hr duration starting at 08 UTC (middle of the night local time). They also create another set of HYSPLIT
25   points in the vicinity that are assigned a shorter duration and a start time of 20 UTC (early afternoon local time). The operational intention of this strategy is to force the NWS smoke forecast model to produce more smoke in the afternoon and evening and less overnight to replicate observed diurnal trends. HYSPLIT points can also be proxies for unobservable smoke producing fires. When analysts see a large number of HMS hotspots, but due to cloud cover do not directly observe smoke, they create HYSPLIT points in order to initiate the NWS smoke forecast model. This occurs most frequently in Kansas,
30   Oklahoma, the Northern Plains (Dakotas), and the lower Mississippi Valley (eastern Arkansas, eastern Louisiana, western Mississippi). The Servicio Meteorológico Nacional (SMN, the Mexican National Weather Service) provides most HYSPLIT points over Mexico. These locations are merged with the HMS product. Occasionally the HMS office will perform fire-

detection analysis in parts of Northern Mexico. The archives of smoke and fire locations spans from August 2005 to present day; but the analysis presented here spans 2007–2014. These archives include text files for HYSPLIT points and HMS hotspots and GIS shapefiles for smoke plumes.

This analysis uses data from 2007-2014 for two reasons. First, prior to April 2006 HYSPLIT points did not have duration or start-time estimates. Second, the HMS implemented a system that automatically generates HYSPLIT points in Northern Canada, Mexico, and Central America in the fall of 2014. This change in procedure resulted in more HYSPLIT points than each of the prior years, and there were many more instances of HYSPLIT points assigned durations of 24 hours. The majority of fires in Northern Canada are wildfires in boreal forests, so each of these automated HYSPLIT points is assigned a duration of 24 hours whereas prior to this it would have received a mix of 24 hour and lesser durations. The intention of the automated system was to reduce the workload of analysts. A similar automated system was implemented for Alaska in 2009. However, this implementation did not lead to a significant change in the proportion of HYSPLIT points with durations of 24 hours analyzed in North America. Prior to Fall 2014 the HYSPLIT points in Mexico and Central America were intended to be generated by SMN, but at some point in the years prior to 2014 SMN began performing HYSPLIT point analysis inconsistently. As a result, NESDIS developed an automated system based on HMS hotspots. The durations for these HYSPLIT points are estimated using the difference between the latest and earliest time for hotspots aggregated into a 20 km grid.

HYSPLIT points are sometimes analyzed at hours when no visible satellite imagery is available to confirm smoke production. This only occurs when the thermal signal in the MWIR imagery is significant in terms of intensity and duration (*e.g.* strong downslope winds at night can cause significant nighttime fire activity). Wildfires can burn for weeks or even months. For the long-lived wildfire scenario, HMS analysts will get two looks daily at a fire from a given polar satellite. One look overnight (in the Western U.S. between 08 – 10 UTC) and another in the afternoon (20 – 22 UTC). Analysts will add HYSPLIT points at both of those times. For the nighttime pass, fires are often not burning as actively (fewer hotspots). Analysts will add HYSPLIT points based on the fewer number of hotspots and typically assign them a 24-hr duration (*i.e.* total hours smoke production observed), creating a baseline for emissions. In the afternoon, fires are more active and generate more smoke so analysts will add additional HYSPLIT points based on the afternoon satellite data. In this case, they only assign durations of 10 or 12 hours. The operational significance of these procedures is to attempt to account for the diurnal variations in smoke production in initializing smoke forecasts.

HYSPLIT Points are analyzed in all 50 U.S. States as well as all Canadian Provinces and Territories. Figure 2 shows the location of all HYSPLIT points analyzed between 2007-2014 shaded by the analyzed duration. Dense clusters of 24-hour duration HYSPLIT points are analyzed in regions with active wildfire seasons (*e.g.* Rocky Mountains, Cascade Mountains, Northern California, Northwest Territories). The Midwest, Great Plains, and Southeast U.S. have many smaller short-lived fires where most HYSPLIT point durations are 12 hours or less. The smoke observed from small short-lived fires

in the Mississippi River valley is consistent with other studies that consider burn area from small fires *(Randerson et al., 2012)*.

2.2 Description of HMS smoke analysis

After identifying HYSPLIT points, HMS analysts use imagery from multiple NOAA and NASA satellites to identify the geographic extent of smoke-plumes (Rolph et al., 2009; Ruminski et al., 2006). Smoke detection is done with visible-band imagery occasionally assisted by infrared to distinguish between clouds and smoke when possible (Ruminski et al., 2006). Geostationary GOES imagery, with its frequent refresh rate (typically every 15 minutes for each spacecraft), is used almost exclusively for smoke detection, although on rare occasions polar orbiting satellite imagery is used. Given the limitations of the satellite data (mostly obscuration of smoke due to cloud cover), the number and extent of smoke plumes within this dataset represents a conservative estimate. Due to the limitations of visible satellite imagery and time constraints, no information about the vertical location or extent of smoke plumes is provided. In 2006 HMS analysts began providing estimates of smoke-plume concentrations (*e.g.* 5 μg/m^3, 16 μg/m^3, and 27 μg/m^3). These plumes of varying concentrations are often nested (*e.g.* 27 μg/m^3 within 16 μg/m^3 and 16 μg/m^3 within 5 μg/m^3). Smoke plumes vary in size considerably; small plumes cover areas < 100 km^2 and others cover several western states. Between 5 August 2005 and 21 December 2015 there are only 80 days (~2%) where there are either no smoke plume GIS files available or no smoke-plumes analyzed. Most of these days occur during winter months. Rolph et al. (2009) used the HMS smoke plumes to validate NOAA's operational smoke forecasting system. There is no published validation of the HMS smoke plume extent data. It could have systematic biases because discrimination of smoke is easier over some areas, seasons, and under certain synoptic weather conditions. For our aggregate smoke plume analysis, the largest uncertainty is at the edges of the plumes. The temporal information is also a source of uncertainty. During the day, the extent can be assessed hourly. However, this is not possible overnight. Smoke is sometimes transported to areas with anthropogenic haze pollution. In some cases the smoke will mix with and become indistinguishable from the anthropogenic haze pollution. The greater the distance travelled by a smoke plume, the more challenging it is to distinguish between smoke and anthropogenic haze. This challenge is particularly pronounced for aged smoke impacting the Southeastern U.S. For this work, we use the archived GIS smoke-plume files available at the following URL ftp://satepsanone.nesdis.noaa.gov/FIRE/HMS/GIS/.

2.3 Description of land cover characteristics database

We assign each HYSPLIT point a land cover type using land-cover classifications from the U.S. Geological Survey (USGS) 2002 North American Land-Cover Characteristics 1 km grid-spacing dataset, created by the National Center for Earth Resources Observation and Science (EROS) as part of the Global Land Cover Characterization Project (Brown, 2016). Land-cover characteristics are assessed using 1 km Advanced Very High Resolution Radiometer (AVHRR) data between 1992 and

1993 using the methods described in (Anderson et al., 1976). For this work, we use the GeoTIFF file projected to latitude-longitude grid as a geospatial raster (Brown, 2016). The latitudinal extent is 18°N - 72°N and 66°W - 172°W. We assign land-cover classifications to HYSPLIT points based on the nearest grid-point center that is not classified as urban or water. When urban or water is the closet grid-point, we substituted the assigned land-cover type with the most common other land-cover classification in the surrounding 0.06 degrees (~5 km) land-cover grid-cells. If all surrounding grid cell land-cover classifications are urban or water, no land-cover assessment is made. Data are available at https://nationalmap.gov/small_scale/mld/landcvi.html. We point readers to Hansen et al., 2000 for a discussion of the weaknesses of the 1992 land-cover dataset. We have also completed a comparison to a 2010 land-cover dataset, and this comparison is available in the Supplemental Information Sect 9.

3 Comparison to Global Fire Emissions Database Version 4s (GFED4s)

The number of HYSPLIT points assigned to a fire is intended to be proportional to the amount of smoke observed. We gridded HYSPLIT points and compared them to the Global Fire Emissions Inventory Database (GFED; http://www.globalfiredata.org/data.html) version 4s (Giglio et al., 2013; Randerson et al., 2012; van der Werf et al., 2010). To compare HMS emissions to GFED4s we count the total duration associated with all HYSPLIT points that fall within each GFED4s grid cell for each month between 2007 and 2014. The units of our new gridded emission inventory are thus duration hours per 0.25° x 0.25° grid box, hereafter referred to as HMS smoke production duration hours (SPDH). GFED4s produces emissions estimates in kg of carbon. Figure 3 shows a comparison of the cumulative gridded SPDH and monthly GFED4s kg carbon emissions over North America for June to September 2007 to 2014 (32 months). The dark red areas in Fig. 3 show the locations where SPDH and GFED4s kg carbon emissions have maximum cumulative values for this time period. Figure 3 shows agreement between the two datasets for the locations of their respective maxima in summer, with some notable differences. For example, GFED4s shows the lower Mississippi river valley as an area with significant fire emissions, but this area is not as prominent in the HMS derived dataset. The HMS dataset also show larger extents of local maximums of fire activity compared to GFED4s (*e.g.* North Central Idaho, North Cascades, Gila National Forest). The spatial agreement for locations with maximum emissions holds when we compare individual summers between 2007 and 2014 (not shown). For example, both emission inventories agree on the geographical location and extent of the 2013 Quebec fires, an anomalously high fire year for Quebec.

Figure 3c presents the cumulative June though September 2007 to 2014 SPDH (Fig. 3a) divided by GFED4s kg of carbon (Fig. 3b) for each grid cell. Areas represented by the same color can be interpreted as locations where the total HMS smoke production duration hours assigned to observed fires is consistent with a given quantity of carbon emitted according to GFED4s. Fig. 3c shows that the relationship between SPDH and GFED4s kg of carbon varies by several orders of

magnitude. The largest ratios are observed in North Dakota and southern Manitoba, the southern Mississippi River Valley, Sinaloa Mexico, and over interior Alaska. The GFED4s estimate of grams of carbon emitted is a function of the estimated burn area, combustion efficiency, soil moisture, land cover type, fuel loading, and emission factors (Giglio et al., 2013; Randerson et al., 2012; van der Werf et al., 2017). The observed difference between SPDH and GFED4s total grams of carbon emitted could be due to any of these factors.

When considering the entire North American domain (15 - 80°N, 50 - 170°W), summed monthly emissions of both SPDH and GFED4s kg of carbon time series, we observe that the magnitudes of the two inventories are also correlated temporally. Figure 4 shows the normalized domain summed monthly time series for all months between 2007 and 2014. The Pearson correlation coefficient for the two time series is 0.842, confirming that at large scales the two emission inventories are well-correlated. In 2007 and 2008 the normalized HMS duration hours greatly exceed normalized GFED4s emissions. The difference in 2007 is due to large smoke production in the Northwest and Rocky Mountain regions present in the HMS data. In 2008 extreme smoke production associated with fires in the Southwest are responsible for the discrepancy. Figure 5 shows the Pearson correlation coefficient for the monthly 8-year time series for each grid cell. Locations with large wildfires (Fig. 2) are well correlated between the two datasets (r > 0.90). Locations with small fires like the Southeast, Southern Plains, and Northern Great Plains have lower correlation values (r < 0.20). Small fires are difficult to detect from satellite. The high correlation between SPDH and GFED4s kg carbon demonstrate that spatial and temporal aggregates of the duration field of HYSPLIT points is roughly proportional to the amount of smoke produced by analyzed fires, though that proportionality varies.

4 HYSPLIT point regional distributions

Figure 6 designates 10 U.S. regions that closely align, but do not exactly overlap, with EPA regions. Figure 7a presents the total number of HYSPLIT points aggregated for each region. The largest numbers of HYSPLIT points are identified in the Southwest, Northwest, Northwest Territories, and Southeast regions. Figure 7b shows the total SPDH produced by each region and the percentage of SPDH that are produced by HYSPLIT points analyzed on cropland. For most regions over the course of the entire year, the percent of SPDH owed to cropland is less than 5%. However, the relative importance of cropland fires has a strong seasonality. Notable exceptions are the U.S. Southeast and the Southern Plains, where the SPDH from cropland > 25%. Figure 8 presents the seasonality of North American HYSPLIT points between 2007-2014, and it shows that the majority of HYSPLIT points are identified between June and August. During these months, the dominant land cover classification for HYSPLIT points is evergreen needleleaf forest, followed by a nearly equal share of scrubland and mixed forest. When viewed in aggregate, the contribution to HYSPLIT points from cropland is largest in the months of April, September, and October. When we assess the land cover classification of HYSPLIT points using a 2010 land cover

dataset, the most common alternative assignment for the most abundant cropland assignment type ("cropland/grassland mosaic") is "temperate subpolar grassland". Regional differences for alternative assignments exist. For example, in the U.S. Midwest and Great Plains, the most common alternative land cover assignment for cropland is "temperate subpolar grassland" while in the Southeast U.S. the most common alternatives are "temperate subpolar needleleaf forests" and "wetland". These 2010 data show that the most common alternative land cover assignment for forested lands in North America is "subpolar shrubland". Maps and bar graphs showing the differences in land cover assignments for HYSPLIT points between the 1992 and 2010 data are available in the Supplemental Information Sect 9.

Figure 9 shows the locations, seasonality, and duration of HYSPLIT points in the 9 contiguous U.S. regions shown in Fig. 8 between 2007 and 2014. The land cover classification for the seasonality and duration histograms is indicated by color using the same color scheme as Fig. 8. HYSPLIT points with durations of 24 hours are usually associated with wildfires. HYSPLIT points with shorter durations represent smaller fires, and often occur on cropland. At the national scale shown in Fig. 8, grassland and cropland land cover classifications make up a small proportion of the total HYSPLIT points throughout the year. However as discussed above at the regional scale, grassland and cropland HYSPLIT points can represent a significant fraction or even dominate the total number of HYSPLIT points. Figure 9 shows this is the case for the Great Plains, Midwest, and Southern Plains. These regions also have the fewest HYSPLIT points analyzed in the summer months, whereas regions dominated by evergreen forest have a minimum in the winter. The Southwest has more HYSPLIT points than any other U.S. region, followed by the Northwest, Southeast, and Southern Plains. The Northeast has the fewest HYSPLIT points, which occur mostly on cropland in southern New Jersey. The Mississippi river valley has some of the most densely analyzed HYSPLIT points in the U.S, and these points are located most commonly on forest and cropland. These points are split between the Southeast and Southern Plains regions. The same HYSPLIT point duration and monthly histograms presented in Fig. 9 are available for all regions in the supplemental information.

There are more HYSPLIT points analyzed in the Southern Plains (n=41,846) than there are in the Rocky Mountains region (n=35,371). However, this does not indicate that the Southern Plains generate more smoke, because the number of points does not include information on fire duration or land cover (discussed in Sect 2.1 and 3.0). The total smoke produced in a region should be roughly proportional to the number of HYSPLIT points multiplied by their respective durations (Fig. 7b). Most of the fires in the Southern Plains have durations less then nine hours while the most-common duration of HYSPLIT points in the Rocky Mountains region is 24 hours. Alaska (not shown) is an extreme example. Almost all of the HYSPLIT points in Alaska have 24-hour durations. This is by design, since 2009 all HYSPLIT points in Alaska have been assigned 24-hour durations, which is consistent with the types of fires that occur within the state.

5 Smoke-transport analysis

5.1 Forward trajectories

Forward HYSPLIT trajectories are initialized from each of the HYSPLIT points over the duration of the fires and they represent likely smoke transport pathways. We initialize HYSPLIT trajectories at three different altitudes that span the range of injection heights expected for North American fires based on a prior analysis. Val Martin et al. (2010) present a climatology of smoke-plume heights by land biome for North America. The smoke-plume heights were estimated with the Multi-angle Imaging SpectroRadiameter (MISR) data, and thus could represent an underestimate of the actual injection height due to the timing of the MISR overpass. A given fire can inject smoke into the atmosphere at many different altitudes. We initialize each HYSPLIT trajectory start hour at three different altitudes spanning the range of injection heights in Val Martin et al. (2010): 500, 1500, and 2500 m amgl. In total, 3,925,932 trajectories are associated with the 517,214 HYSPLIT points analyzed between 2007 and 2014. For this analysis, trajectories are only considered for the hours that the calculated height above ground level is > 0. Each trajectory was run using the Global Data Assimilation System (GDAS) 1-degree meteorology data and the Eta Data Assimilation System (EDAS) 40km meteorology data. Further specifics on the meteorology data, the HYSPLIT model, and how we ran trajectories for this analysis can be found in Sect S1 of the Supplemental Information.

5.2 Combined analysis of HYSPLIT points and forward trajectories

To build smoke source-receptor relationships for 10 U.S. regions, we define a "smoke hour" as an hourly latitude-longitude HYSPLIT trajectory location (hereafter trajectory point) that overlaps a HMS smoke plume. We use these smoke hours to represent the relative abundance of probable smoke transport pathways. Smoke plume overlap assessments are made using a two-step process which reduces the number of trajectories included in the analysis initialized at heights that do not represent the smoke injection height of a given fire. Spatial overlap analysis is performed using the R Sp package *over* function (Pebesma et al., 2016). This code and further information on the procedure used can be found in Supplemental Information Sect S3 and S11.

       1) We confirm that at least one of the first 49 trajectory points (2 days + start hour) overlap a smoke plume analyzed for the same two dates. If none of the first 49 trajectory points overlap a smoke plume, the trajectory is immediately discarded. If any of the first 49 trajectory points overlap a smoke plume, plume-overlap analysis is performed for the entire trajectory (145 trajectory points). We determine the relevance of a trajectory using two days of smoke plumes because smoke plumes only represent smoke perimeters during daylight hours. By including the next-dates smoke plumes in the test, trajectories associated with fires that start in the afternoon are not evaluated more stringently than trajectories associated with

HYSPLIT points for fires in the morning. The results do not change significantly when validating trajectories with only the first-days smoke plumes.

2) When trajectories meet our first criteria, a point over polygon calculation is done for each matching date trajectory point and smoke plume. There are two weaknesses to this approach: 1) Smoke-plume boundaries are only representative of smoke-plume perimeters during daylight hours, while trajectories exist at both day and night. 2) Some HMS smoke plumes on individual days are very large; during extreme-smoke events plumes can cover most of the continental U.S., and thus this criteria is not always meaningful. Smoke plumes with different concentration estimates are often nested. For this analysis, these smoke plumes are merged, so a trajectory point can only overlap a single plume and contribute one smoke hour. Each trajectory smoke hour is associated with the source region and land cover classification from its initialization point, allowing a source region and land classification analysis of the total number of smoke hours impacting or emanating from a region. This methodology does not provide information about smoke concentration ($\mu g/m^3$), and we have not placed additional altitude constraints on the trajectories. A further weaknesses of smoke hours is that they have no way of accounting for secondary organic aerosol particle formation, which previous studies have shown can be significant (*e.g.* Janhäll et al., 2010; Sakamoto et al., 2015, 2016). However, these processes remain poorly understood and represented in chemical transport models, so an appreciation for the fact that these processes exist, but not attempting to account for them, is sufficient for understanding the results presented in this work.

5.3 Smoke production and frequency by region

Figure 10 shows the total number of smoke hours produced by and present over each region. Years with elevated fire activity in each region can be easily identified using Fig. 10. This figure provides context for isolated case studies of smoke transport associated with extreme periods including the summer 2013 Quebec wildfires (Laffineur et al., 2014), the summer 2012 wildfires in the Rocky Mountain region (Val Martin et al., 2013b), and the 2008 California wildfires (SW Region) (Gyawali et al., 2009). Supplemental Table S3 shows the total smoke hours produced by and over each U.S. region and the differences between using the GDAS and EDAS meteorology datasets for the trajectory calculations.

5.4 Smoke transport climatology: Attribution, age and altitude

Figure 11 shows a summary of smoke transport to contiguous U.S. regions between 2007 and 2014 for the months of June, July, August, and September. The first histogram associated with each region shows the distribution of the age (hour in HYSPLIT trajectory) of smoke hours in the atmospheric column above each region separated by the region of origin. The second histogram associated with each region shows the distribution of the height of the smoke in the column above each region (height of the HYSPLIT trajectories). For a given fire, older smoke is more likely to deliver lower $PM_{2.5}$ concentrations than smoke that is only a few hours old. For example, the Great Plains region has more smoke hours (12.3

million hours) than the Northwest (12.1 million hours) and more than the Southwest (10.1 million); however, the average age of smoke hours over the Great Plains is ~one day older than in the Northwest and Southwest.

On average regions that produce the most smoke hours during the summer have more smoke at lower altitudes (*e.g.* Northwest and Rocky Mountains), and the smoke is located at higher altitudes over regions far downwind of summertime

5   fire activity (*e.g.* Northeast). The impact of our choices of 500, 1500, and 2500 m trajectory starting height is apparent in the panels showing trajectory altitude; the majority of the smoke-hours defined over every U.S. region are between these altitudes. There are clear maxima in the number of smoke hours defined at 500, 1500, and 2500 m over source regions that are not receptors of aged smoke hours.

Regions with the largest number of HYSPLIT points tend to contribute the largest proportion of the total smoke

10   hours within their own borders. For example, the smoke impacting the Southwest originates almost exclusively within the Southwest. Smoke present over the Northwest is nearly equally likely to be associated with fires from the Northwest or Southwest. Regions with comparatively little local fire activity have a diverse set of source regions contributing to their total column smoke hour budgets; examples include the Southern Plains, the Northeast, and the Midwest. Due to the difficulty in detecting small fires, it is possible that the smoke hour budget in these regions underestimates local contributions of smoke

15   in these regions. The Northwest, Southwest, and Rocky Mountains dominate most other regions total smoke hours; fires in Canada make a major contribution to smoke hours, particularly over the Midwest and the Northeast. The only regions where the three largest contributors of smoke hours are not the Southwest, Northwest, and Rocky Mountains, are the Southeast, Northeast, the Southern Plains and Alaska (not shown). The only regions that contribute more smoke hours to their own total budget then any outside region are the Southwest, the Northwest, and the Southeast. When the analysis presented in Fig. 11

20   is generated with trajectories run using EDAS meteorology data, the interpretation of the results is nearly identical and they are available in Sect S7 of the Supplemental Information.

5.5 Common regional smoke transport pathways

The Northwest, Southwest, and Rocky mountains produce more smoke hours than any other U.S. regions (Fig. 10). Figure 12 shows the average smoke-transport pathways for fires in these regions for June to September between 2007 and 2014, and

25   it was generated by summing smoke hours. This figure is not a simple trajectory climatology, rather it is a trajectory climatology when smoke-producing fires occur. Smoke produced by fires in California is transported most frequently over Northern California and Eastern Oregon (Fig. 12b). Based on our analysis of the HYSPLIT points these fires occur primarily in evergreen needleleaf forests. Smoke originating from fires in the Northwest is transported most frequently over Eastern Washington, Eastern Oregon, Northern Idaho, and Montana (Fig. 12a). Smoke from fires in Alaska impact every U.S. and

30   Canadian region (not shown). The dominant transport pathway for the smoke crosses Alaska, the Northwest Territories, and

Yukon Territory. Smoke traveling from Alaska to Texas has been observed previously (Morris et al., 2006), but this situation is relatively rare when viewed in an aggregate context (Supplemental Information Fig. S3a).

Figure 13 summarizes the climatology of the abundance of smoke hour transport for U.S. regions between June and September 2007 to 2014. The Northwest, Southwest, and Rocky Mountains are responsible for the majority of U.S. summertime smoke hour production. They are also among the smokiest regions by measure of smoke hours overhead and smoke hour age. These three regions also dominate the total smoke hours overhead for all U.S. regions excluding the Southeast and Alaska.

Figure 1 shows that the U.S. Midwest has more smoke plumes overhead than the Northwest and Southwest while Fig. 13 shows that the Northwest and Southwest are the largest U.S. producers and receptors of smoke hours. This occurs because smoke hours are proportional to total smoke while the plumes in Fig. 1 represent the total count of overhead plumes and contains no information about smoke concentration. This contrast shows that there are many dilute smoke plumes over the Midwest since it is downwind of so many smoke producing regions. The Northwest and Southwest by comparison have fewer plumes overhead but are located much closer to fire activity. Plumes over the Northwest and Southwest represent higher smoke concentrations than the plumes observed over Midwest.

6 Conclusions

This work moves a new dataset from a daily operational setting to a research context. We define smoke hours, a quantity designed to be proportional to total column smoke, by linking observed smoke plumes to observed fires using HYSPLIT trajectories. The location of smoke hours is restricted to locations where HMS analysts have drawn smoke plumes while the height is estimated from HYSPLIT forward trajectories emanating from HYSPLIT points. This work shows that the Southwest, Northwest, and Northwest Territories initiate the most air quality forecasts, and produce more smoke than any other North American region by measure of the number of HYSPLIT points analyzed, the duration of those HYSPLIT points, and the total number of smoke hours produced (Fig. 3, 10-13). This dataset confirms that there is a substantial amount of smoke-producing fire activity in the Southeast, particularly along the lower Mississippi River Valley. This is likely an underestimate of the contribution of small fires to smoke abundance in this region given that small fires within this region can be substantially under-detected (Hu et al., 2016). In addition, it can be challenging to differentiate anthropogenic haze from smoke in this region.

During the summer wildfire season, the largest smoke plume source regions are located in the west, while receptor regions for smoke plumes are primarily located in the east. The majority of smoke plumes located over western source regions are less than 24 hours old. During summer, our analysis implies that the receptor regions have very little smoke less than 24 hours old with most smoke in the column older than 48 hours. The Southeast is a unique exception. There is an

abundance of fresh smoke hours (peak near zero hours in Fig. 11) due to the numerous small fires within the region, and this region receives aged smoke hours (<48 hours of atmospheric aging) from several upwind regions. Due to the challenge of distinguishing aged smoke plumes from anthropogenic haze in the Southeast, the abundance of aged smoke hours is likely underestimated in this region. Though not the focus of the paper, this dataset shows that the Northeast and Mid Atlantic receive more smoke from fires in Canada than regions in the U.S. Midwest.

We present a smoke transport climatology for the summer wildfire season. Based on our metric of smoke hours, the U.S. regions that produce the most smoke are the Northwest, Southwest, and Rocky Mountains (Fig. 10 and 13). Heavily populated locations in the eastern U.S. (Northeast, Mid Atlantic, Southeast) are routinely impacted by wildland fire smoke from western regions. Thus, changes to the frequency or intensity of fires in the U.S. west or Canada is likely to impact the entire U.S. airshed.

7 Code availability

All analysis code associated with this project is available online at the following subversion repository: http://salix.atmos.colostate.edu/svn/smokeSource/. The code for the Python module used to execute HYSPLIT Trajectories is available at the following GitHub repository: https://github.com/samatwood/HYSPLITm/. All figures were creating using R and Python. Scripts used to make figures can be identified by a leading "plot" in script names (*e.g.* R/plotRegionSmokeTotals.R). These scripts explicitly reference the data files used to create the plots.

8 Data availability

All data for this project are publicly available at the URLs specified in Sect 2. Formatted analysis level data and the code used to format the raw data described in Sect 2, are available in the subversion code repository.

9 Author contributions

S. Brey and E. Fischer prepared the manuscript with contributions from all co-authors. E. Fischer was the leader of this research project. S. Brey performed the majority of analysis, including writing analysis code, creating figures, and executing HYSPLIT trajectories. M. Ruminski is responsible for creating, describing, and properly interpreting Hazard Mapping System HYSPLIT points and smoke plume data. S. Atwood wrote the Python module used to run HYSPLIT Trajectories and served as our expert for best practices in the use of HYSPLIT source-receptor analysis.

10 Competing interests

The authors declare that they have no conflict of interest.

11 Acknowledgements

This publication was developed under Assistance Agreement No. 83588401 awarded by the U.S. Environmental Protection Agency to Emily V. Fischer. It has not been formally reviewed by EPA. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the Agency. EPA does not endorse any products or commercial services mentioned in this publication. Bonne Hotmann made significant science and style contributions to Fig. 1. We gratefully acknowledge the skill and dedication of NESDIS Hazard Mapping System analysts; their careful operations are responsible for the dataset presented in this paper. The authors would like to thank Ana Rappold for her work using HMS, which strengthened and guided the early stages of this analysis, and for a brainstorming phone call in January 2016 which helped spark the authors interest in performing this work. Parts of this work were included in Steven Brey's masters thesis. His graduate committee, Elizabeth Barnes, Jeffery Pierce, and Monique Rocca made this work stronger. The authors gratefully acknowledge Matt Bishop for his help in enabling us to execute millions of HYSPLIT trajectories on our local computer cluster. The authors gratefully acknowledge the NOAA Air Resources Laboratory (ARL) for the provision of the HYSPLIT transport and dispersion model and READY website used in this paper. The authors would also like to thank the core development teams of R and Python for making open source languages that are free and help advance scientific discovery.

References

Akagi, S. K., Yokelson, R. J., Wiedinmyer, C., Alvarado, M. J., Reid, J. S., Karl, T., Crounse, J. D. and Wennberg, P. O.: Emission factors for open and domestic biomass burning for use in atmospheric models, Atmos Chem Phys, 11(9), 4039–4072, doi:10.5194/acp-11-4039-2011, 2011.

Alvarado, M. J., Lonsdale, C. R., Yokelson, R. J., Akagi, S. K., Coe, H., Craven, J. S., Fischer, E. V., McMeeking, G. R., Seinfeld, J. H., Soni, T., Taylor, J. W., Weise, D. R. and Wold, C. E.: Investigating the links between ozone and organic aerosol chemistry in a biomass burning plume from a prescribed fire in California chaparral, Atmos Chem Phys, 15(12), 6667–6688, doi:10.5194/acp-15-6667-2015, 2015.

Anderson, J. R., Hardy, E. E., Roach, J. T. and Witmer, R. E.: A land use and land cover classification system for use with remote sensor data, USGS Numbered Series. [online] Available from: http://pubs.er.usgs.gov/publication/pp964 (Accessed 23 August 2016), 1976.

Baker, K. R., Woody, M. C., Tonnesen, G. S., Hutzell, W., Pye, H. O. T., Beaver, M. R., Pouliot, G. and Pierce, T.: Contribution of regional-scale fire events to ozone and PM2.5 air quality estimated by photochemical modeling approaches, Atmos. Environ., 140, 539–554, doi:10.1016/j.atmosenv.2016.06.032, 2016.

Brey, S. J. and Fischer, E. V.: Smoke in the City: How Often and Where Does Smoke Impact Summertime Ozone in the United States?, Environ. Sci. Technol., 50(3), 1288–1294, doi:10.1021/acs.est.5b05218, 2016.

Brown, J.: USGS Small-scale Dataset - North American Land Cover Characteristics - 1-Kilometer Resolution 200212 GeoTIFF - ScienceBase-Catalog, [online] Available from: https://www.sciencebase.gov/catalog/item/535fe572e4b078dca33ae61f (Accessed 12 July 2016), 2016.

Creamean, J. M., Neiman, P. J., Coleman, T., Senff, C. J., Kirgis, G., Alvarez, R. J. and Yamamoto, A.: Colorado air quality impacted by long range transport: A set of case studies during the 2015 Pacific Northwest fires, Atmospheric Chem. Phys. Discuss., 1–35, doi:10.5194/acp-2016-280, 2016.

Dennekamp, M. and Carey, M.: Air quality and chronic disease: why action on climate change is also good for health, New South Wales Public Health Bull., 21(5–6), 115–121, doi:10.1071/NB10026, 2010.

Dennis, A., Fraser, M., Anderson, S. and Allen, D.: Air pollutant emissions associated with forest, grassland, and agricultural burning in Texas, Atmos. Environ., 36(23), 3779–3792, doi:10.1016/S1352-2310(02)00219-4, 2002.

Forrister, H., Liu, J., Scheuer, E., Dibb, J., Ziemba, L., Thornhill, K. L., Anderson, B., Diskin, G., Perring, A. E., Schwarz, J. P., Campuzano-Jost, P., Day, D. A., Palm, B. B., Jimenez, J. L., Nenes, A. and Weber, R. J.: Evolution of brown carbon in wildfire plumes, Geophys. Res. Lett., 42(11), 2015GL063897, doi:10.1002/2015GL063897, 2015.

Garcia-Menendez, F., Hu, Y. and Odman, M. T.: Simulating smoke transport from wildland fires with a regional-scale air quality model: Sensitivity to uncertain wind fields, J. Geophys. Res. Atmospheres, 118(12), 6493–6504, doi:10.1002/jgrd.50524, 2013.

Giglio, L., Randerson, J. T. and van der Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), J. Geophys. Res. Biogeosciences, 118(1), 317–328, doi:10.1002/jgrg.20042, 2013.

Gyawali, M., Arnott, W. P., Lewis, K. and Moosmüller, H.: In situ aerosol optics in Reno, NV, USA during and after the summer 2008 California wildfires and the influence of absorbing and non-absorbing organic coatings on spectral light absorption, Atmos Chem Phys, 9(20), 8007–8015, doi:10.5194/acp-9-8007-2009, 2009.

Haikerwal, A., Akram, M., Monaco, A. D., Smith, K., Sim, M. R., Meyer, M., Tonkin, A. M., Abramson, M. J. and Dennekamp, M.: Impact of Fine Particulate Matter (PM2.5) Exposure During Wildfires on Cardiovascular Health Outcomes, J. Am. Heart Assoc., 4(7), e001653, doi:10.1161/JAHA.114.001653, 2015.

Hansen, M. C., Defries, R. S., Townshend, J. R. G. and Sohlberg, R.: Global land cover classification at 1 km spatial resolution using a classification tree approach, Int. J. Remote Sens., 21(6–7), 1331–1364, doi:10.1080/014311600210209, 2000.

Hu, X., Yu, C., Tian, D., Ruminski, M., Robertson, K., Waller, L. A. and Liu, Y.: Comparison of the Hazard Mapping System (HMS) fire product to ground-based fire records in Georgia, USA, J. Geophys. Res. Atmospheres, 121(6), 2015JD024448, doi:10.1002/2015JD024448, 2016.

5 Hurteau, M. D., Westerling, A. L., Wiedinmyer, C. and Bryant, B. P.: Projected Effects of Climate and Development on California Wildfire Emissions through 2100, Environ. Sci. Technol., 48(4), 2298–2304, doi:10.1021/es4050133, 2014.

Jaffe, D., Chand, D., Hafner, W., Westerling, A. and Spracklen, D.: Influence of fires on O-3 concentrations in the western US, Environ. Sci. Technol., 42(16), 5885–5891, doi:10.1021/es800084k, 2008.

Janhäll, S., Andreae, M. O. and Pöschl, U.: Biomass burning aerosol emissions from vegetation fires: particle number and mass emission factors and size distributions, Atmos Chem Phys, 10(3), 1427–1439, doi:10.5194/acp-10-1427-2010, 2010.

10 Jimenez, J., Wu, C.-F., Claiborn, C., Gould, T., Simpson, C. D., Larson, T. and Sally Liu, L.-J.: Agricultural burning smoke in eastern Washington—part I: Atmospheric characterization, Atmos. Environ., 40(4), 639–650, doi:10.1016/j.atmosenv.2005.09.071, 2006.

Keywood, M., Kanakidou, M., Stohl, A., Dentener, F., Grassi, G., Meyer, C. P., Torseth, K., Edwards, D., Thompson, A. M., Lohmann, U. and Burrows, J.: Fire in the Air: Biomass Burning Impacts in a Changing Climate, Crit. Rev. Environ. Sci.
15 Technol., 43(1), 40–83, doi:10.1080/10643389.2011.604248, 2013.

Laffineur, Q., Delcloo, A., De Backer, H., Adam, M. and Klugmann, D.: Observation of an intercontinental smoke plume over Europe on June 2013: some ambiguity in the determination of the source, vol. 16, p. 2173. [online] Available from: http://adsabs.harvard.edu/abs/2014EGUGA..16.2173L (Accessed 22 August 2016), 2014.

Lassman, W., Ford, B., Gan, R. W., Pfister, G., Magzamen, S., Fischer, E. V. and Pierce, J. R.: Spatial and temporal
20 estimates of population exposure to wildfire smoke during the Washington state 2012 wildfire season using blended model, satellite, and in situ data, GeoHealth, 1(3), 2017GH000049, doi:10.1002/2017GH000049, 2017.

Liu, J., Scheuer, E., Dibb, J., Diskin, G. S., Ziemba, L. D., Thornhill, K. L., Anderson, B. E., Wisthaler, A., Mikoviny, T., Devi, J. J., Bergin, M., Perring, A. E., Markovic, M. Z., Schwarz, J. P., Campuzano-Jost, P., Day, D. A., Jimenez, J. L. and Weber, R. J.: Brown carbon aerosol in the North American continental troposphere: sources, abundance, and radiative
25 forcing, Atmos Chem Phys, 15(14), 7841–7858, doi:10.5194/acp-15-7841-2015, 2015.

Liu, Y.: A Regression Model for Smoke Plume Rise of Prescribed Fires Using Meteorological Conditions, J. Appl. Meteorol. Climatol., 53(8), 1961–1975, doi:10.1175/JAMC-D-13-0114.1, 2014.

McCarty, J. L., Korontzi, S., Justice, C. O. and Loboda, T.: The spatial and temporal distribution of crop residue burning in the contiguous United States, Sci. Total Environ., 407(21), 5701–5712, doi:10.1016/j.scitotenv.2009.07.009, 2009.

30 Miller, D. J., Sun, K., Zondlo, M. A., Kanter, D., Dubovik, O., Welton, E. J., Winker, D. M. and Ginoux, P.: Assessing boreal forest fire smoke aerosol impacts on U.S. air quality: A case study using multiple data sets, J. Geophys. Res. Atmospheres, 116(D22), D22209, doi:10.1029/2011JD016170, 2011.

Morris, G. A., Hersey, S., Thompson, A. M., Pawson, S., Nielsen, J. E., Colarco, P. R., McMillan, W. W., Stohl, A., Turquety, S., Warner, J., Johnson, B. J., Kucsera, T. L., Larko, D. E., Oltmans, S. J. and Witte, J. C.: Alaskan and Canadian

forest fires exacerbate ozone pollution over Houston, Texas, on 19 and 20 July 2004, J. Geophys. Res. Atmospheres, 111(D24), D24S03, doi:10.1029/2006JD007090, 2006.

Park, R. J., Jacob, D. J., Chin, M. and Martin, R. V.: Sources of carbonaceous aerosols over the United States and implications for natural visibility, J. Geophys. Res. Atmospheres, 108(D12), 4355, doi:10.1029/2002JD003190, 2003.

5   Park, R. J., Jacob, D. J. and Logan, J. A.: Fire and biofuel contributions to annual mean aerosol mass concentrations in the United States, Atmos. Environ., 41(35), 7389–7400, doi:10.1016/j.atmosenv.2007.05.061, 2007.

Paugam, R., Wooster, M., Freitas, S. and Val Martin, M.: A review of approaches to estimate wildfire plume injection height within large-scale atmospheric chemical transport models, Atmos Chem Phys, 16(2), 907–925, doi:10.5194/acp-16-907-2016, 2016.

10   Pebesma, E., Bivand, R., Rowlingson, B., Gomez-Rubio, V., Hijmans, R., Sumner, M., MacQueen, D., Lemon, J. and O'Brien, J.: sp: Classes and Methods for Spatial Data. [online] Available from: https://cran.r-project.org/web/packages/sp/index.html (Accessed 8 July 2016), 2016.

Pfister, G. G., Wiedinmyer, C. and Emmons, L. K.: Impacts of the fall 2007 California wildfires on surface ozone: Integrating local observations with global model simulations, Geophys. Res. Lett., 35(19), L19814, 15   doi:10.1029/2008GL034747, 2008.

Randerson, J. T., Chen, Y., van der Werf, G. R., Rogers, B. M. and Morton, D. C.: Global burned area and biomass burning emissions from small fires, J. Geophys. Res. Biogeosciences, 117(G4), G04012, doi:10.1029/2012JG002128, 2012.

Rappold, A. G., Stone, S. L., Cascio, W. E., Neas, L. M., Kilaru, V. J., Carraway, M. S., Szykman, J. J., Ising, A., Cleve, W. E., Meredith, J. T., Vaughan-Batten, H., Deyneka, L. and Devlin, R. B.: Peat Bog Wildfire Smoke Exposure in Rural North 20   Carolina Is Associated with Cardiopulmonary Emergency Department Visits Assessed through Syndromic Surveillance, Environ. Health Perspect., 119(10), 1415–1420, doi:10.1289/ehp.1003206, 2011.

Rastigejev, Y., Park, R., Brenner, M. P. and Jacob, D. J.: Resolving intercontinental pollution plumes in global models of atmospheric transport, J. Geophys. Res. Atmospheres, 115(D2), D02302, doi:10.1029/2009JD012568, 2010.

Rolph, G. D., Draxler, R. R., Stein, A. F., Taylor, A., Ruminski, M. G., Kondragunta, S., Zeng, J., Huang, H.-C., Manikin, 25   G., McQueen, J. T. and Davidson, P. M.: Description and Verification of the NOAA Smoke Forecasting System: The 2007 Fire Season, Weather Forecast., 24(2), 361–378, doi:10.1175/2008WAF2222165.1, 2009.

Ruminski, M., Kondragunta, S., Draxler, R. and Zeng, J.: Recent changes to the Hazard Mapping System, 15th Int. Emiss. Inventory Conf, (Reinventing Inventories) [online] Available from: https://www.researchgate.net/publication/228625934_Recent_changes_to_the_Hazard_Mapping_System (Accessed 8 July 30   2016), 2006.

Russell, A. R., Valin, L. C. and Cohen, R. C.: Trends in OMI NO2 observations over the US: Effects of emission control technology and the economic recession, Atmospheric Chem. Phys., 12(6), 15419–15452, doi:10.5194/acpd-12-15419-2012, 2012.

Saide, P. E., Peterson, D. A., da Silva, A., Anderson, B., Ziemba, L. D., Diskin, G., Sachse, G., Hair, J., Butler, C., Fenn, M., 35   Jimenez, J. L., Campuzano-Jost, P., Perring, A. E., Schwarz, J. P., Markovic, M. Z., Russell, P., Redemann, J., Shinozuka,

Y., Streets, D. G., Yan, F., Dibb, J., Yokelson, R., Toon, O. B., Hyer, E. and Carmichael, G. R.: Revealing important nocturnal and day-to-day variations in fire smoke emissions through a multiplatform inversion, Geophys. Res. Lett., 42(9), 2015GL063737, doi:10.1002/2015GL063737, 2015.

Sakamoto, K. M., Allan, J. D., Coe, H., Taylor, J. W., Duck, T. J. and Pierce, J. R.: Aged boreal biomass-burning aerosol size distributions from BORTAS 2011, Atmos Chem Phys, 15(4), 1633–1646, doi:10.5194/acp-15-1633-2015, 2015.

Sakamoto, K. M., Laing, J. R., Stevens, R. G., Jaffe, D. A. and Pierce, J. R.: The evolution of biomass-burning aerosol size distributions due to coagulation: dependence on fire and meteorological details and parameterization, Atmos Chem Phys, 16(12), 7709–7724, doi:10.5194/acp-16-7709-2016, 2016.

Scholze, M., Knorr, W., Arnell, N. W. and Prentice, I. C.: A climate-change risk analysis for world ecosystems, Proc. Natl. Acad. Sci., 103(35), 13116–13120, doi:10.1073/pnas.0601816103, 2006.

Schroeder, W., RUMINSKI, M., CSISZAR, I., GIGLIO, L., GIGLIO, E., SCHMIDT, C. and MORISETTE, J.: Validation analyses of an operational fire monitoring product: The Hazard Mapping System, Int. J. Remote Sens., 29(20), 6059–6066, 2008.

Stein, A. F., Rolph, G. D., Draxler, R. R., Stunder, B. and Ruminski, M.: Verification of the NOAA Smoke Forecasting System: Model Sensitivity to the Injection Height, Weather Forecast., 24(2), 379–394, doi:10.1175/2008WAF2222166.1, 2009.

Val Martin, M., Logan, J. A., Kahn, R. A., Leung, F.-Y., Nelson, D. L. and Diner, D. J.: Smoke injection heights from fires in North America: analysis of 5 years of satellite observations, Atmos Chem Phys, 10(4), 1491–1510, doi:10.5194/acp-10-1491-2010, 2010.

Val Martin, M., Heald, C. L., Ford, B., Prenni, A. J. and Wiedinmyer, C.: A decadal satellite analysis of the origins and impacts of smoke in Colorado, Atmos Chem Phys, 13(15), 7429–7439, doi:10.5194/acp-13-7429-2013, 2013a.

Val Martin, M., Heald, C. L., Ford, B., Prenni, A. J. and Wiedinmyer, C.: A decadal satellite analysis of the origins and impacts of smoke in Colorado, Atmos Chem Phys, 13(15), 7429–7439, doi:10.5194/acp-13-7429-2013, 2013b.

Val Martin, M., Heald, C. L., Lamarque, J.-F., Tilmes, S., Emmons, L. K. and Schichtel, B. A.: How emissions, climate, and land use change will impact mid-century air quality over the United States: a focus on effects at national parks, Atmos Chem Phys, 15(5), 2805–2823, doi:10.5194/acp-15-2805-2015, 2015.

Vedal, S. and Dutton, S. J.: Wildfire air pollution and daily mortality in a large urban area, Environ. Res., 102(1), 29–35, doi:10.1016/j.envres.2006.03.008, 2006.

Washenfelder, R. A., Attwood, A. R., Brock, C. A., Guo, H., Xu, L., Weber, R. J., Ng, N. L., Allen, H. M., Ayres, B. R., Baumann, K., Cohen, R. C., Draper, D. C., Duffey, K. C., Edgerton, E., Fry, J. L., Hu, W. W., Jimenez, J. L., Palm, B. B., Romer, P., Stone, E. A., Wooldridge, P. J. and Brown, S. S.: Biomass burning dominates brown carbon absorption in the rural southeastern United States, Geophys. Res. Lett., 42(2), 2014GL062444, doi:10.1002/2014GL062444, 2015.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y. and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos Chem Phys, 10(23), 11707–11735, doi:10.5194/acp-10-11707-2010, 2010.

van der Werf, G. R. van der, Randerson, J. T., Giglio, L., Leeuwen, T. T. van, Chen, Y., Rogers, B. M., Mu, M., Marle, M. J. E. van, Morton, D. C., Collatz, G. J., Yokelson, R. J. and Kasibhatla, P. S.: Global fire emissions estimates during 1997–2016, Earth Syst. Sci. Data, 9(2), 697–720, doi:https://doi.org/10.5194/essd-9-697-2017, 2017.

Westerling, A. L.: Increasing western US forest wildfire activity: sensitivity to changes in the timing of spring, Phil Trans R Soc B, 371(1696), 20150178, doi:10.1098/rstb.2015.0178, 2016.

Westerling, A. L., Hidalgo, H. G., Cayan, D. R. and Swetnam, T. W.: Warming and Earlier Spring Increase Western U.S. Forest Wildfire Activity, Science, 313(5789), 940–943, doi:10.1126/science.1128834, 2006.

Wiedinmyer, C., Quayle, B., Geron, C., Belote, A., McKenzie, D., Zhang, X., O'Neill, S. and Wynne, K. K.: Estimating emissions from fires in North America for air quality modeling, Atmos. Environ., 40(19), 3419–3432, doi:10.1016/j.atmosenv.2006.02.010, 2006.

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J. and Soja, A. J.: The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning, Geosci. Model Dev., 4(3), 625–641, doi:10.5194/gmd-4-625-2011, 2011.

Yue, X., Mickley, L. J., Logan, J. A. and Kaplan, J. O.: Ensemble projections of wildfire activity and carbonaceous aerosol concentrations over the western United States in the mid-21st century, Atmospheric Environ. Oxf. Engl. 1994, 77, 767–780, doi:10.1016/j.atmosenv.2013.06.003, 2013.
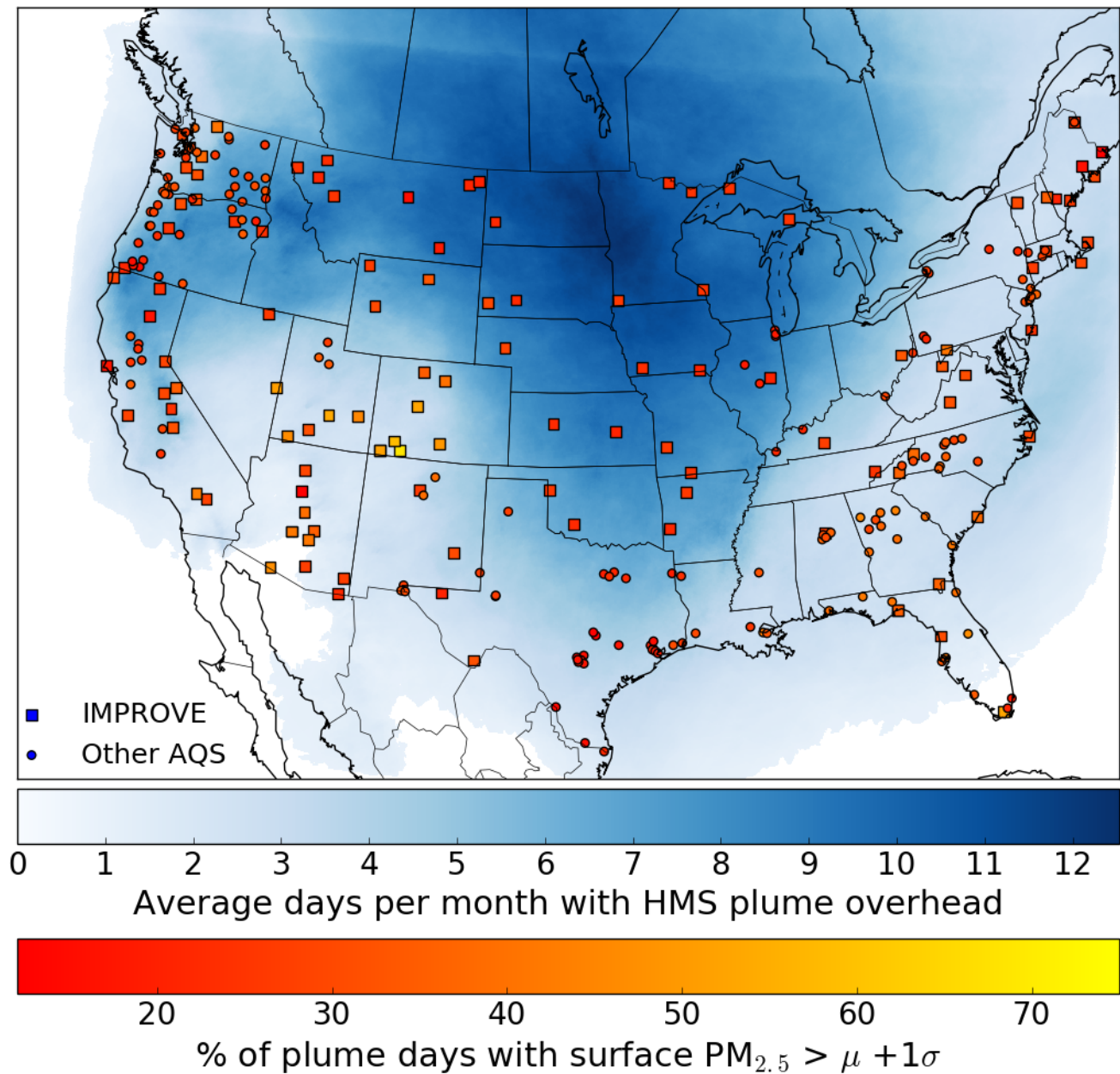
Figure 1: North American smoke climatology for June-September 2007-2014. Blue shading indicates the average days per month HMS smoke plumes are analyzed overhead. Markers show non FRM (Federal Reference Method) PM$_{2.5}$ monitors shaded by the percent of plume days where surface 24-hr average PM$_{2.5}$ is greater than the mean (μ) plus one standard deviation (σ) value for the monitors mean value in June-September. Only monitors with data on 90% of days between June-September 2007-2014 are presented. IMPROVE monitors are marked by squares; all other monitors from the EPA Air Quality System (AQS) are marked by circles. IMPROVE monitors record data every third day. We make the assumption that smoke plumes are equally likely to occur on measurement days as non-measurement days, and thus multiply the number of ground level smoke-impacted plume days

at IMPROVE sites by three. Monitors that fall within white areas are influenced by smoke < 1 day per month. The climatology presented here does not relate to smoke plume or ground level $PM_{2.5}$ concentrations, only the presence of smoke in the column and proportion of those days when surface $PM_{2.5}$ is elevated.
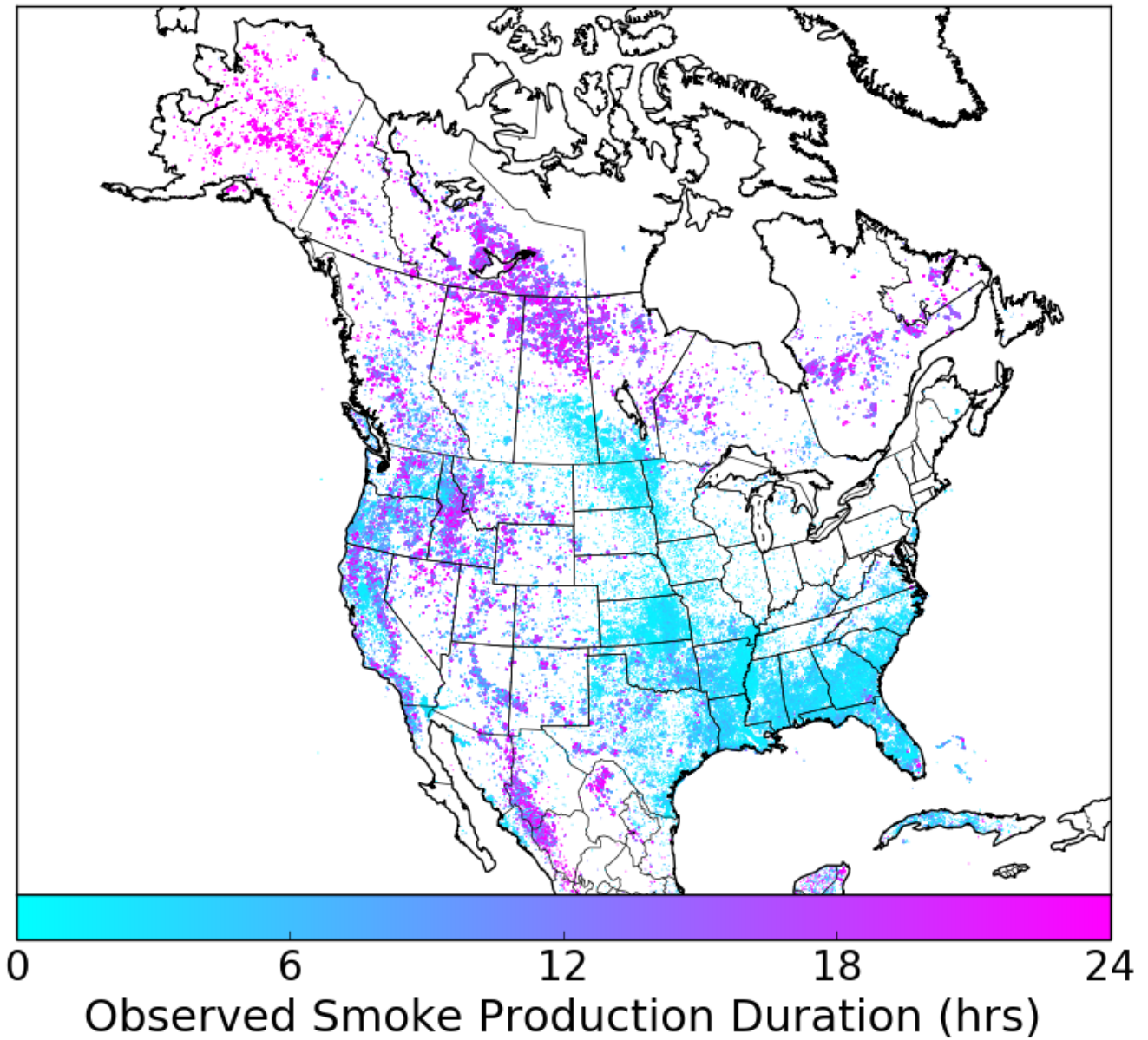
Figure 2: The locations of all HYSPLIT points analyzed between 2007-2014 shaded by the duration assigned by the analyst. HYSPLIT Points are assigned a duration between 1 and 15, or 24 hours.

Total HMS SPDH

GFED4s kg Carbon Emissions (10^8)

(GFED4s kg Carbon) / (HMS SPDH)
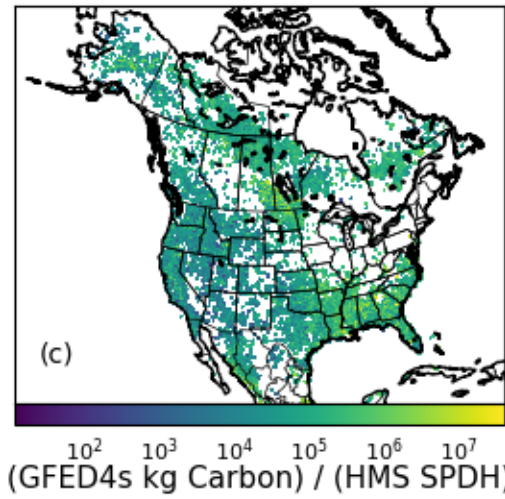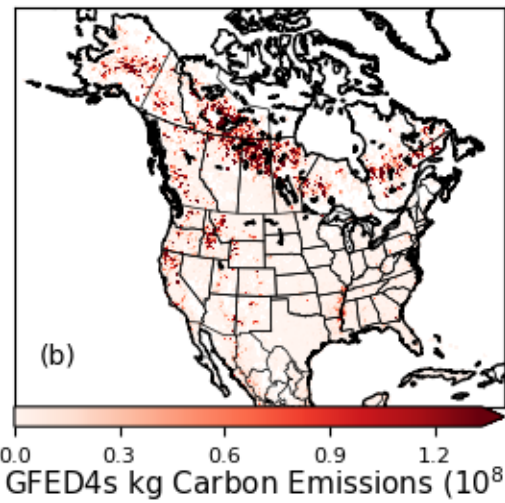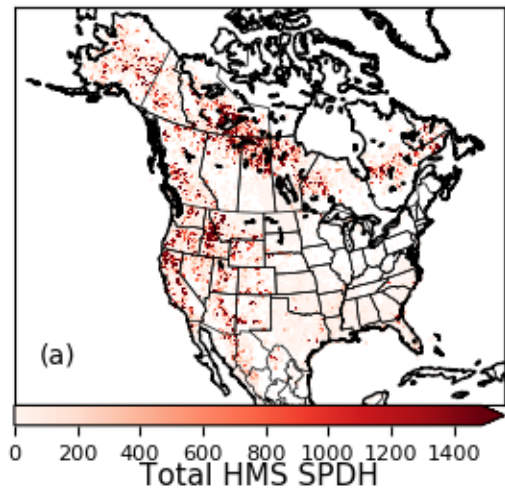
Figure 3: (a) Cumulative June-September 2007-2014 HMS smoke production duration hours (SPDH) and (b) GFED4s kg carbon per 0.25° x 0.25° grid cell. Both emission inventories color bars saturate (darkest red) at the 99th percentile of non-zero values, meaning all values above the 99th percentile value for both datasets are represented by the darkest color on the color bar. (c) Total GFED4s kg of carbon emitted divided by the total HMS smoke production duration hours per grid cell for months June-September 2007-2014. The color bar shows the values on a $\log_{10}$ scale where yellow values represent grid cells with more GFED4s kg of carbon per HYSPLIT Point smoke production duration hour and purple indicate fewer. Locations where either GFED4s or HMS emissions are equal to zero are not plotted. The mean kg carbon per HYSPLIT Point smoke production duration hour is 300,123.
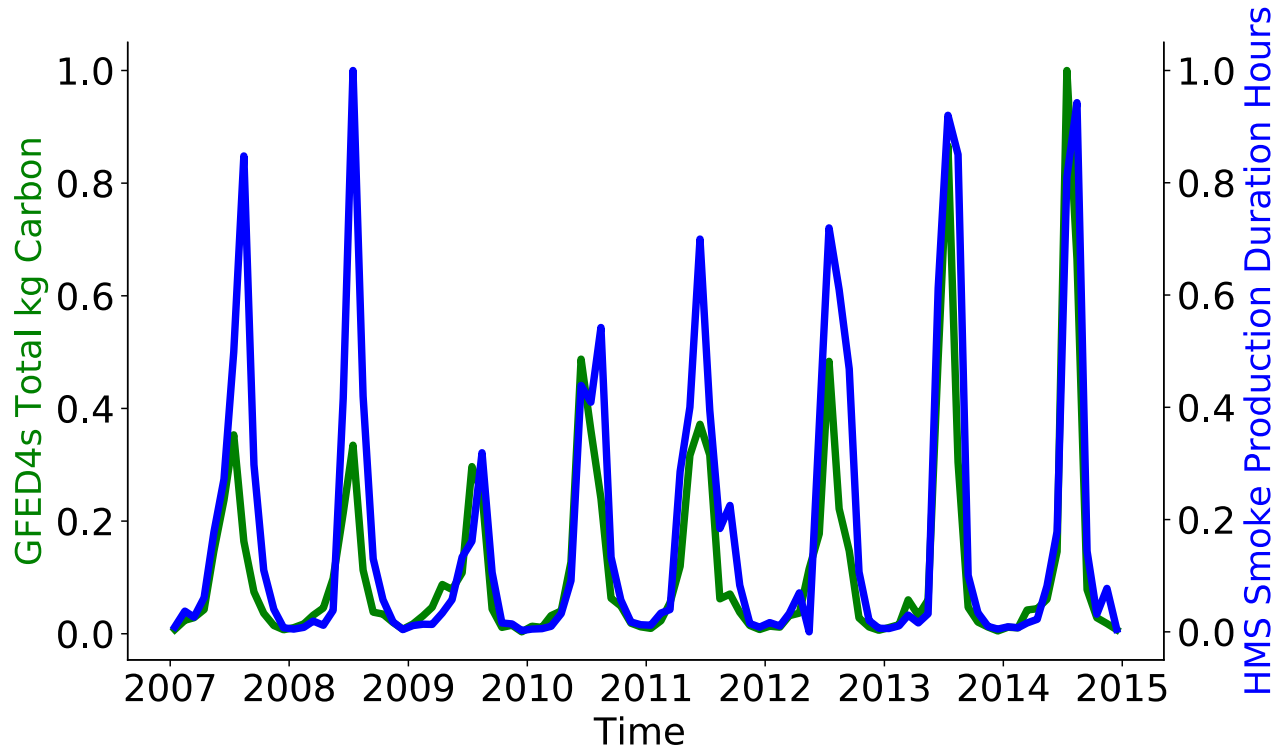
Figure 4: Time series of monthly domain (20-80° N, 50-170° W) summed kg carbon (left axis) and HMS-hours (right axis). Monthly totals have been divided by the maximum monthly value observed between 2007-2014. The Pearson correlation coefficient for the presented monthly time series is 0.842.
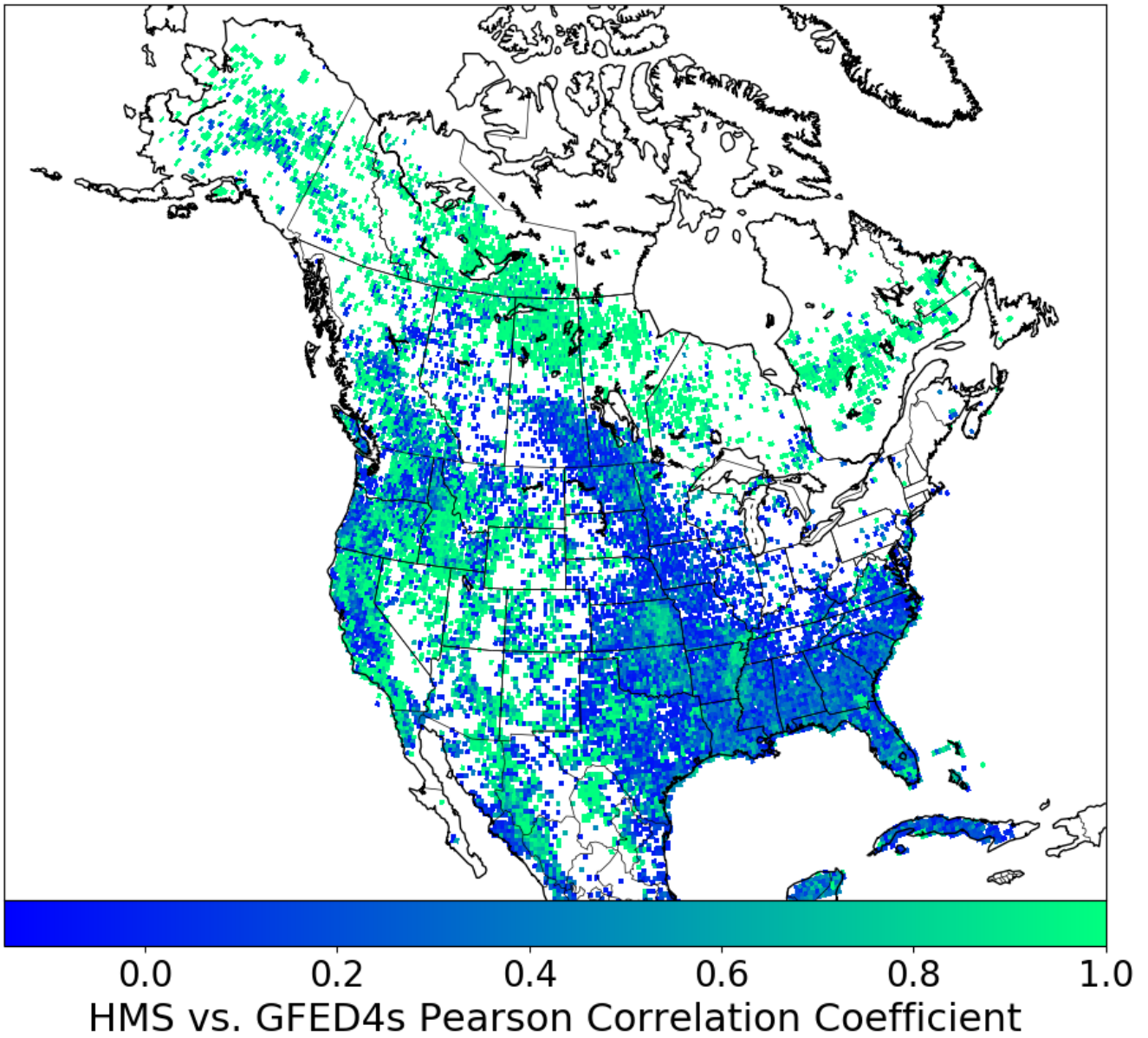
Figure 5: Plan view monthly time series Pearson correlation coefficient for monthly HMS-hours and GFED4s kg carbon for each 0.25 x 0.25 degree grid cell. White grid cells are locations where one or both of the inventories have zero emissions for all 96 months.
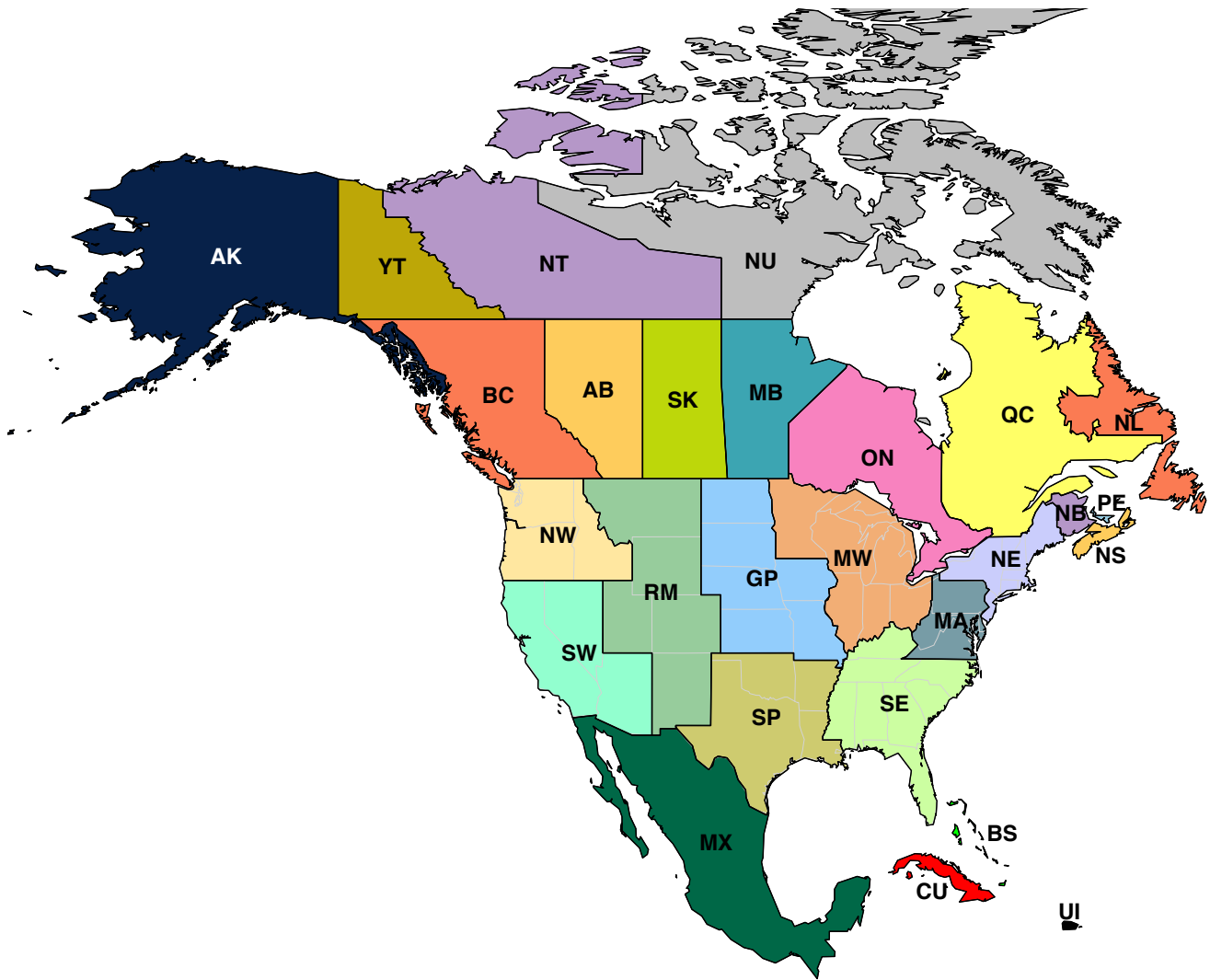
Figure 6: Smoke source and receptor regions used in this analysis. Northeast (NE), Mid Atlantic (MA), Southeast (SE), Midwest (MW), Southern Plains (SP), Great Plains (GP), Rocky Mountains (RM), Southwest (SW), Northwest (NW), Alaska (AK), U.S. Islands (UI), Mexico (MX), Quebec (QC), Nova Scotia (NS), Saskatchewan (SK), Alberta (AB), Newfoundland and Labrador (NL), British Columbia (BC), New Brunswick (NB), Prince Edward Island (PE), Yukon Territory (YT), Manitoba (MB), Ontario (ON), Nunavut (NU), Northwest Territories (NT), Cuba (CU), and Bahamas (BS).
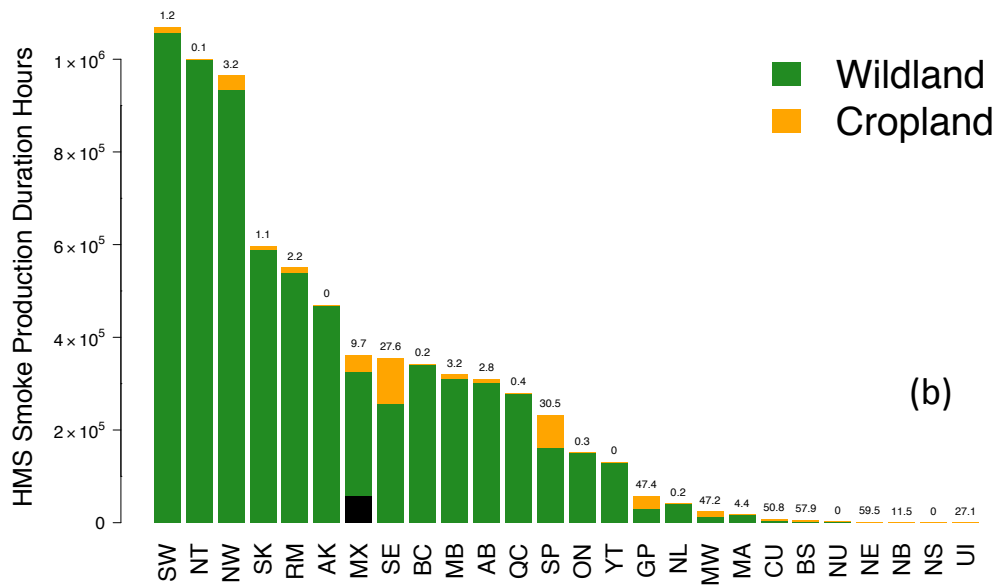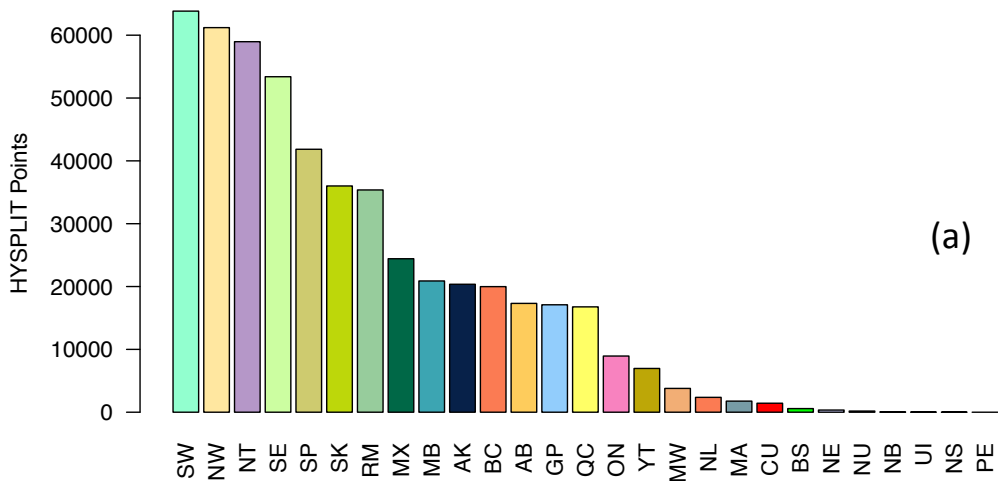
Figure 7: (a) Total HMS HYSPLIT point detections for all regions and months between 2007-2014. Region colors and abbreviations are defined in Fig. 6. (b) Total HMS smoke production duration hours (SPDH) for the same time period. The number on top of each bar indicates the percentage of SPDH that are from croplands. The black portion of Mexico (MX) represents fires that occurred on land outside of the land cover data extent used in this analysis.
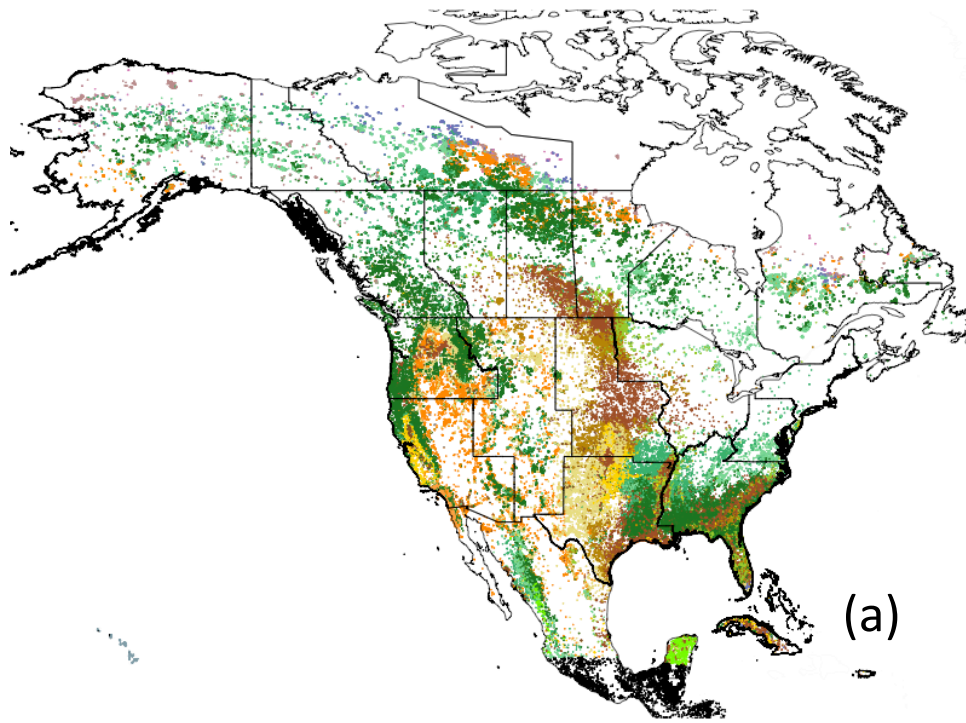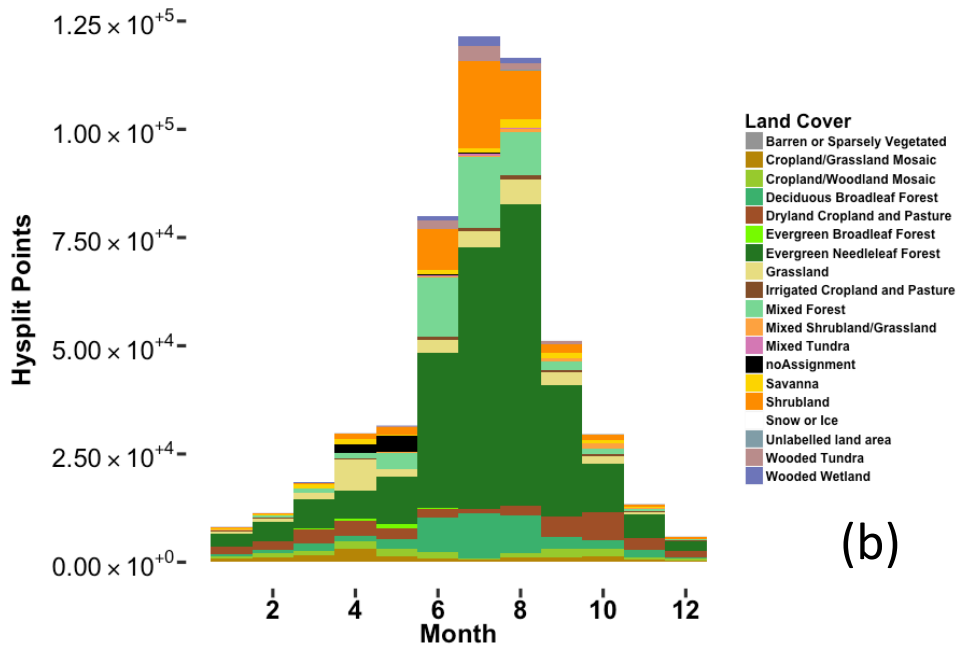
Figure 8: Location, (a) seasonality, (b) and land cover classification assignment (color) for all HYSPLIT points analyzed over North America between 2007-2014 (n=517,214). No land cover assignment is made for Southern Mexico (MX) and U.S. Islands (UI) due to the latitudinal range of land cover data.
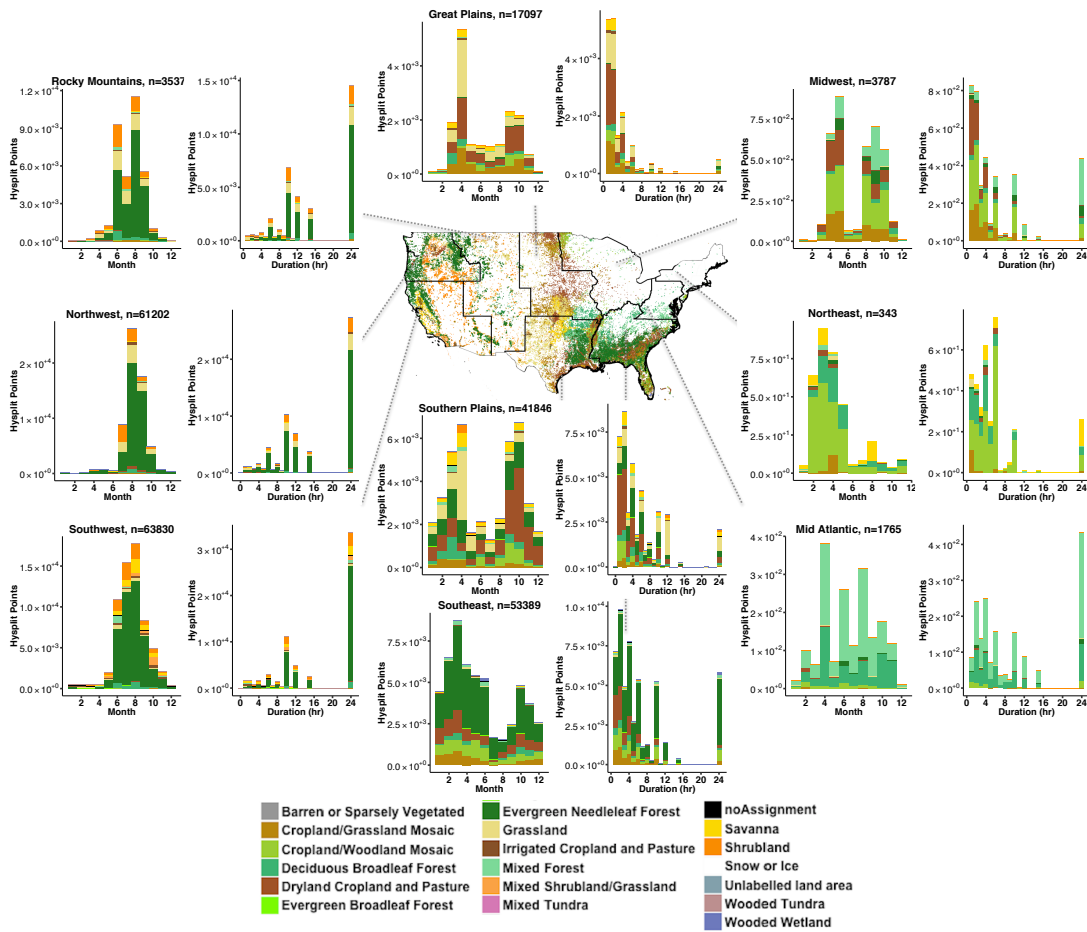
Figure 9: HYSPLIT point locations and land cover classification for the nine continental U.S. regions. The map in the middle shows the regions and the locations of all HYSPLIT points analyzed between 2007 and 2014. The HYSPLIT points are shaded by land cover classification. Two histograms are shown for each region. The total number of HYSPLIT points analyzed each month is shown on the left histograms. The HYSPLIT point duration and land cover classification is shown on the right. The bars for both histograms are shaded by the land cover classification assigned to the HYSPLIT points. Non U.S. regions and Alaska are available in the supplemental information.
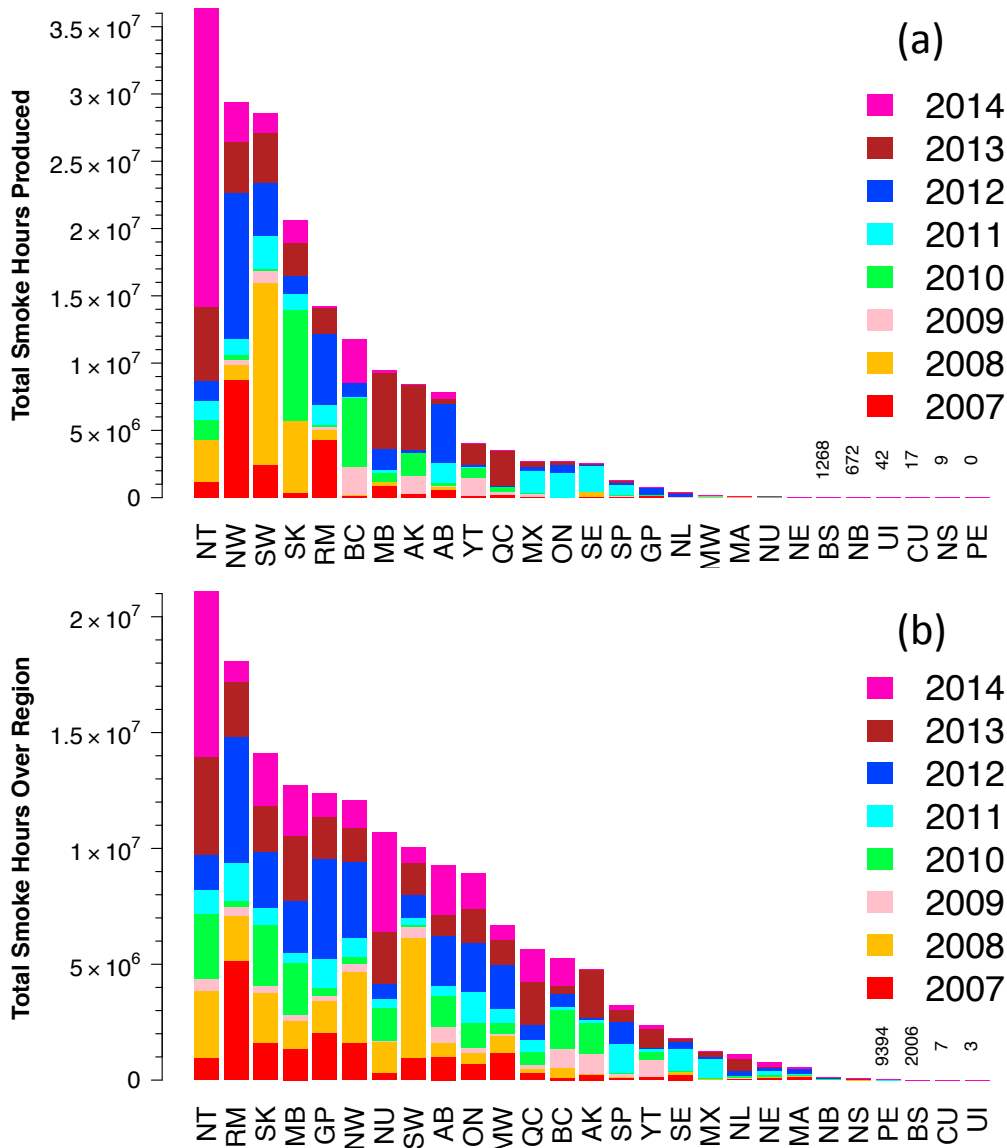
Figure 10: (a) Total number of smoke hours produced (anywhere) by each region for months June-September between 2007 and 2014 using GDAS1 meteorology. **(b)** Total number of smoke hours over each region for months June-September between 2007 and 2014 using GDAS1 meteorology.
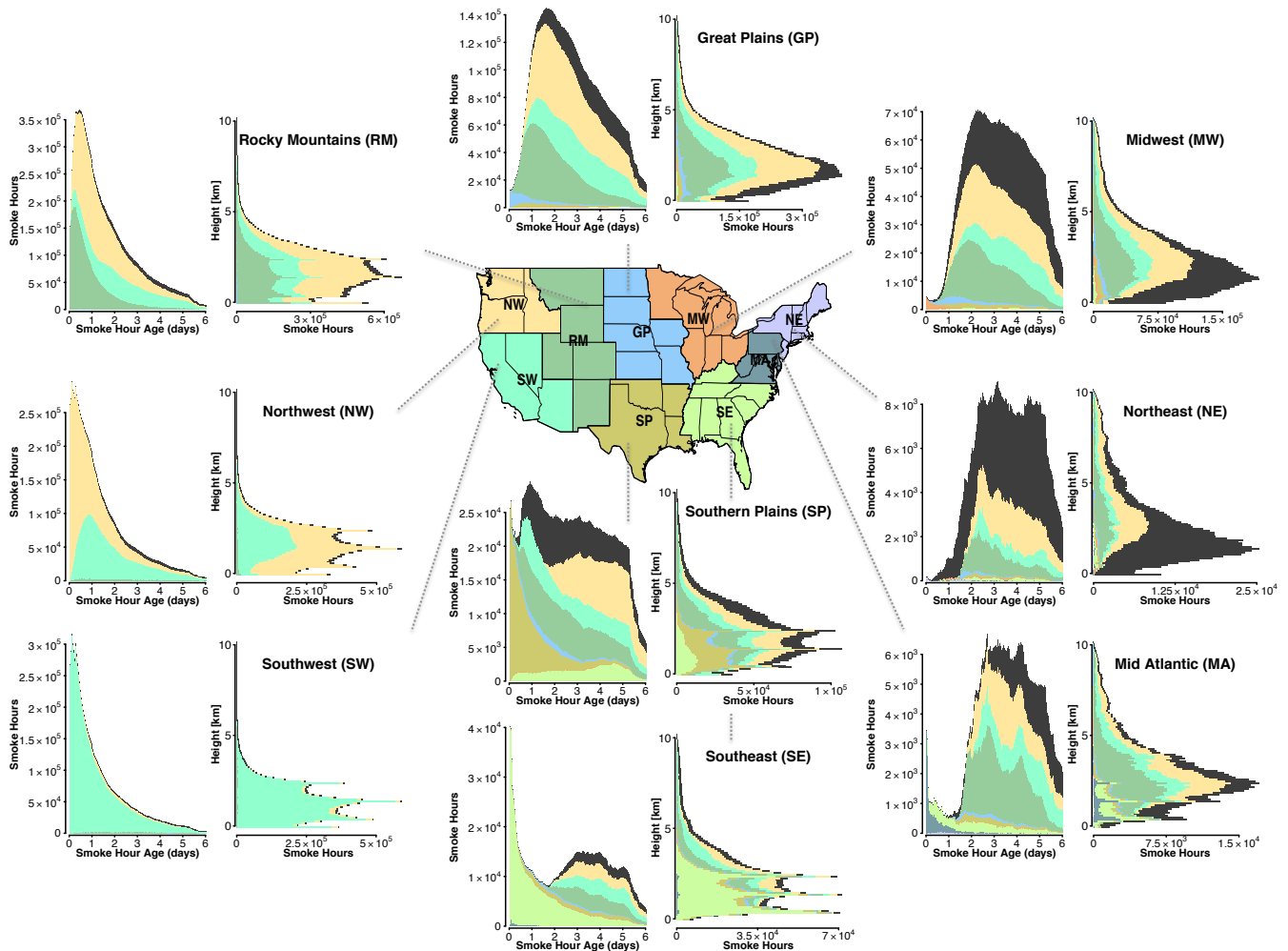
Figure 11: Summertime (June-September) regional smoke hour transport summaries for the nine continental U.S. regions between 2007 and 2014 using GDAS1 meteorology. Two histograms are shown for each region. The histogram on the left shows the distribution of smoke hour age. The color of the smoke hours matches the region of origin as indicated by the map in the center. Column 2 shows the distribution of smoke hour height segregated by source region color. Smoke hours contributed by regions outside of the continental U.S. are dark gray. Histograms of smoke hour age, height, and region of origin for all regions shown in Fig. 6 are available in the supplemental information.
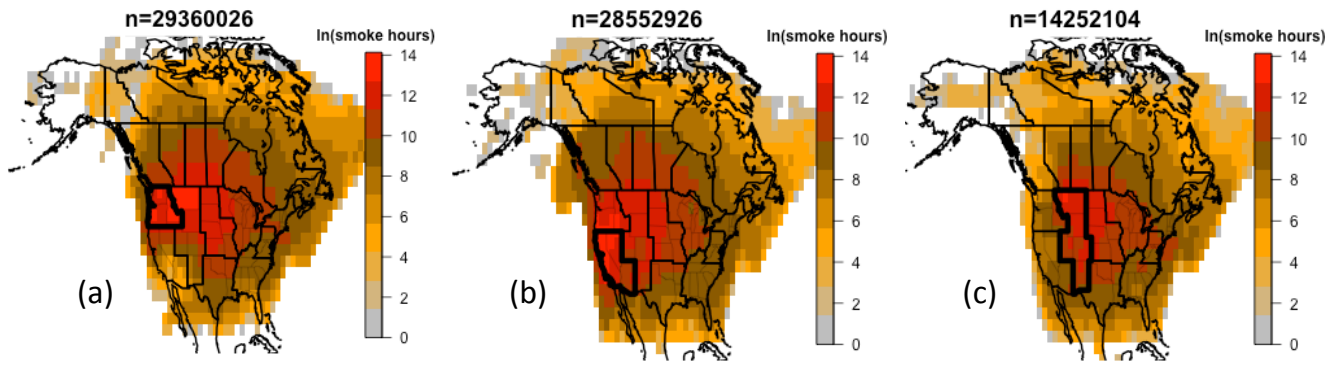
Figure 12: Summertime (June-September) regional smoke hour transport heat maps for the Northwest (a), Southwest (b), and Rocky Mountains (c). Each column shows the count of smoke hours produced by each region on a 2° by 2° degree grid with a consistent color bar for all three regions. The grid spans 18-180° W and 18-90° N, a domain covering all five sectors where HMS analyzes smoke plumes (only a subset plotted). Shaded values are the natural log of the number of smoke hours in each grid cell (min=1, max=1.2 million). All figures were generated using GDAS1 meteorology data.
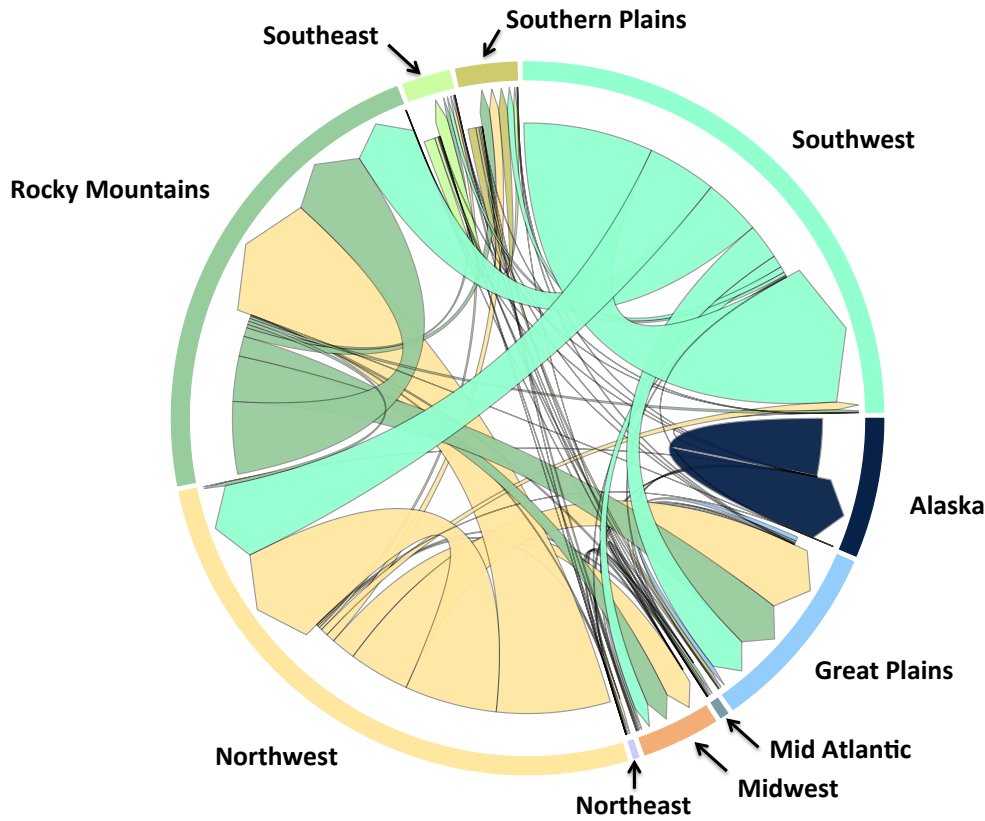
Figure 13: U.S. regions smoke hour climatology for June-September 2007-2014 using GDAS1 meteorology. Arrows indicate the smoke hours source region and destination region. The width of arrows are proportional to total smoke hours transported between regions. Each of the 10 U.S. regions makes up a fraction of the total circumference of the circle, which is proportional to the total

smoke hours over and produced by all U.S. regions. The widths of the arrows are proportional to the number of smoke hours in a given transport pathway. The colors of the arrows match the source region of the smoke hours (Fig. 6). This figure places the individual histogram of Fig. 11 into context.