Constraining emissions of volatile organic compounds from western US wildfires with WE-CAN and FIREX-AQ airborne observations

Lixu Jin¹, Wade Permar¹, Vanessa Selimovic¹,a, Damien Ketcherside¹, Robert J. Yokelson¹, Rebecca S. Hornbrook², Eric C. Apel², I-Ting Ku³, Jeffrey L. Collett Jr.³, Amy P. Sullivan³, Daniel A. Jaffe⁴,⁵, Jeffrey R. Pierce³, Alan Fried⁶, Matthew M. Coggon⁷, Georgios I. Gkatzelis⁷,⁸,b, Carsten Warneke⁷, Emily V. Fischer³, and Lu Hu¹

¹Department of Chemistry and Biochemistry, University of Montana, Missoula, MT, USA
²Atmospheric Chemistry Observations & Modeling Laboratory, National Center for Atmospheric Research, Boulder, CO, USA
³Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA
⁴School of Science, Technology, Engineering and Mathematics, University of Washington, Bothell, WA, USA
⁵Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA
⁶Institute of Arctic and Alpine Research, University of Colorado Boulder, Boulder, CO, USA
⁷Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA
⁸Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA

a now at: Department of Chemistry, University of Michigan, Ann Arbor, MI, USA
b now at: Institute of Energy and Climate Research, IEK-8: Troposphere, Forschungszentrum Jülich GmbH, Jülich, Germany

Correspondence: Lixu Jin (lixu.jin@umontana.edu)

Received: 15 October 2022 – Discussion started: 10 November 2022
Revised: 24 March 2023 – Accepted: 16 April 2023 – Published: 31 May 2023

Abstract. The impact of biomass burning (BB) on the atmospheric burden of volatile organic compounds (VOCs) is highly uncertain. Here we apply the GEOS-Chem chemical transport model (CTM) to constrain BB emissions in the western USA at ∼25 km resolution. Across three BB emission inventories widely used in CTMs, the inventory–inventory comparison suggests that the totals of 14 modeled BB VOC emissions in the western USA agree with each other within 30%–40%. However, emissions for individual VOCs can differ by a factor of 1–5, driven by the regionally averaged emission ratios (ERs, reflecting both assigned ERs for specific biome and vegetation classifications) across the three inventories. We further evaluate GEOS-Chem simulations with aircraft observations made during WE-CAN (Western Wildfire Experiment for Cloud Chemistry, Aerosol Absorption and Nitrogen) and FIREX-AQ (Fire Influence on Regional to Global Environments and Air Quality) field campaigns. Despite being driven by different global BB inventories or applying various injection height assumptions, the model–observation comparison suggests that GEOS-Chem simulations underpredict observed vertical profiles by a factor of 3–7. The model shows small to no bias for most species in low-/no-smoke conditions. We thus attribute the negative model biases mostly to underestimated BB emissions in these inventories. Tripling BB emissions in the model reproduces observed vertical profiles for primary compounds, i.e., CO, propane, benzene, and toluene. However, it shows no to less significant improvements for oxygenated VOCs, particularly for formaldehyde, formic acid, acetic acid, and lumped ≥C₃ aldehydes, suggesting the model is missing secondary sources of these compounds in BB-impacted environments. The underestimation of primary BB emissions in inventories is likely attributable to underpredicted amounts of effective dry matter burned, rather than errors in fire detection, injection height, or ERs, as constrained by aircraft and ground measurements. We
cannot rule out potential sub-grid uncertainties (i.e., not being able to fully resolve fire plumes) in the nested GEOS-Chem which could explain the negative model bias partially, though back-of-the-envelope calculation and evaluation using longer-term ground measurements help support the argument of the dry matter burned underestimation. The total ERs of the 14 BB VOCs implemented in GEOS-Chem only account for half of the total 161 measured VOCs (~75 versus 150 ppb ppm⁻¹). This reveals a significant amount of missing reactive organic carbon in widely used BB emission inventories. Considering both uncertainties in effective dry matter burned (× 3) and unmodeled VOCs (× 2), we infer that BB contributed to 10% in 2019 and 45% in 2018 (240 and 2040 Gg C) of the total VOC primary emission flux in the western USA during these two fire seasons, compared to only 1%–10% in the standard GEOS-Chem.

1 Introduction

Biomass burning (BB), including wild and prescribed fires, is estimated to be the largest primary source of fine particulate matter (PM) and the second-largest source of volatile organic compounds (VOCs) globally (Yokelson et al., 2008), impacting air quality, public health, and climate. In fire-prone areas such as the western United States of America (USA), the relative importance of BB emissions as a source of air pollution has been growing due to increased wildfire activity (Westerling, 2016; Higuera et al., 2021) and decreased anthropogenic emissions (Warneke et al., 2012; Simon et al., 2015). Wildfires have been suggested to account for up to half of the overall PM₂.₅ burden since 2012 and contribute to its increasing trend in the last 3 decades in the western USA (McClure and Jaffe, 2018; O’Dell et al., 2019; Burke et al., 2021). Wildfire impacts on VOC burdens are highly uncertain, in part due to the limited observational constraints on BB VOC emissions. Here we apply comprehensive VOC observations from two recent aircraft campaigns targeting fires, along with the GEOS-Chem chemical transport model (CTM), to examine our understanding of BB emissions in the western USA.

Current CTMs often poorly simulate the impact of wildfire smoke partly because of an incomplete description of the quantity and speciation of VOC emissions, along with poor representation of their spatial, temporal, and vertical distributions (Alvarado and Prinn, 2009; Jaffe and Wigder, 2012; Jaffe et al., 2018; Baker et al., 2016, 2018; Wolfe et al., 2022). BB emission estimates are typically derived from the product of a compound-and-biome-specific emission factor (EF, expressed as mass of species in grams per dry biomass burned in kilograms) and an effective amount of dry matter burned (effective DM burned, kg). Both EF and DM burned are subject to large uncertainties (Akagi et al., 2011; Andereae, 2019; Carter et al., 2020). EFs are either measured in laboratory burning experiments that attempt to simulate real-world fires or quantified from near-field measurements on the ground or air that may be influenced by atmospheric aging processes before sampling (e.g., Burling et al., 2010; Warneke et al., 2010; Wooster et al., 2011; Permar et al., 2021; Majluf et al., 2022). Recent efforts to reconcile the difference between laboratory and field measurements support the need to adjust lab EFs to the typical field combustion efficiency (Permar et al., 2021; Selimovic et al., 2018). However, the burn conditions throughout the course of a fire are currently not considered in inventories. In addition, commonly used global BB emission inventories often consider only three to six biome groups (Andreae and Merlet, 2001; Wiedinmyer et al., 2011; Akagi et al., 2011; Randerson et al., 2012; Kaiser et al., 2012; Koster et al., 2015; Andreae, 2019). For example, the Quick Fire Emissions Dataset version 2.4 (QFED2.4) inventory has three biome groups to represent all global biomass: tropical forest, extratropical forest, and savanna/grass (Koster et al., 2015). In the Global Fire Emissions Database version 4 with small fires (GFED4s), the extratropical forest biome is subdivided into the boreal forest and temperate forest, and additionally two biomes for peatland and agriculture/waste burning are considered, thus giving a total of six (van der Werf et al., 2017). Differences (and errors) in vegetation classifications among inventories can also lead to diverse assigned EFs, even though those EFs may come from the same experimental studies, thus resulting in different emission estimates.

DM burned in commonly used emission inventories is estimated by two different satellite remote sensing approaches. The “fire-detection-based and/or burned-area-based (FD-BA)” method estimates DM burned from the product of fire burn areas (BAs) and fuel consumption (i.e., loading, type, timing, and rate). Global BB emission inventories using this method include GFED4s (van der Werf et al., 2017) and the Fire INventory from NCAR version 1.5 (FINv1.5; Wiedinmyer et al., 2011). Another approach uses fire radiative power (FRP, radiant energy released per time by burning fuel) and its empirical relationship with biomass burned. Some widely used BB emission inventories using this approach include QFED2.4 (Koster et al., 2015) and the Global Fire Assimilation System version 1.2 (GFASv1.2; Kaiser et al., 2012). Both FD-BA-based and FRP-based inventories share common sources of uncertainties, such as missing fire detections and/or FRP observations used to initialize DM burned estimates. Additionally, those fire products are mostly from polar-orbiting satellites with a low temporal coverage (i.e., once or twice daily at a fixed local time) and can be ob-
scurred by clouds and smoke, resulting in assumptions often having to be made to fill both temporal and spatial gaps in the observations (Wang et al., 2018; Wiggins et al., 2020; Stockwell et al., 2022). Current operational BB emission inventories can produce monthly CO and aerosol fluxes that vary by a factor of 5 or even 20 for a specific region (Al-Saadi et al., 2008; Zhang et al., 2014; Koster et al., 2015). These differences in global total emissions averaged over longer periods are smaller but still on the order of a factor of 2–4 (Stroppiana et al., 2010; Granier et al., 2011; Carter et al., 2020; Liu et al., 2020; Pan et al., 2020). The discrepancy could be even larger in VOC emission estimates due to different speciation among inventories (i.e., GFED4s has 21 VOCs, while QFED2.4 has 9 VOCs). Different input data used to drive BB emissions, such as EFs, fire detections, fire burned area, and the amount of biomass burned, are all thought to contribute to the divergent estimates among emission inventories. Recently, Carter et al. (2020) suggested that, at least for aerosol, the BB emission uncertainties are mostly from DM burned at both regional and global scales and that differences in EFs across inventories are smaller than differences in DM burned. These errors in estimating DM burned will also affect VOC emission estimates; thus their uncertainty is thought to be at least on a similar order to that of aerosol and CO.

When compared to observations, previous model evaluation studies (again mostly focusing on CO and aerosol) often point to a general underestimation of BB emissions in the commonly used inventories and a factor of 2 as the global BB emission uncertainty (Kopacz et al., 2010; Wang et al., 2018; Carter et al., 2020; Pan et al., 2020; Bela et al., 2022). For example, various degrees of negative model bias are found in aerosol optical depth and CO near BB source regions when compared to corresponding satellite and ground observations, though the FRP-based BB inventories often provide higher emissions than the FD-BA-based estimates (Yurganov et al., 2011; Petrenko et al., 2012, 2017; Zhang et al., 2014; Reddington et al., 2016; Pan et al., 2020; Liu et al., 2020; Bela et al., 2022). For the western USA, Pfister et al. (2011) suggested that an early version of FINN (version 1) underestimated BB emissions by a factor of 4 over California, as revealed by constraints from aircraft and satellite measurements. More recently, Carter et al. (2021) found that the GEOS-Chem model driven by GFED4s is biased low for CO but captures the carbonaceous BB aerosol when compared to the 2018 WE-CAN (Western Wildfire Experiment for Cloud Chemistry, Aerosol Absorption and Nitrogen) airborne observations. Another recent study by Bela et al. (2022) found that the daily mean emission estimates from seven existing inventories for a case study of a western US wildfire varied by a factor of 83, despite bracketing the observed BB CO fluxes. Even with observational constraints on certain input parameters (e.g., for relating FRP to the quantity of biomass burned or emissions released), their uncertainty range is still a factor of ~2 compared to the direct CO flux measurements in fire plumes (Bela et al., 2022). A similar case study also suggested a wide spread of the hourly emission estimates (spanning a factor of >33) from nine satellite-based inventories in the FIREX-AQ (Fire Influence on Regional to Global Environments and Air Quality) airborne observations (Stockwell et al., 2022).

Here we aim to improve current understanding of VOC emissions from wildfires in the western USA. Leveraging the comprehensive VOC observations from the WE-CAN airborne campaign, we evaluate a 0.25° × 0.3125° nested version of the GEOS-Chem CTM driven by three commonly used global BB emissions inventories (Sect. 4). We assess the potential reasons for model and observation discrepancies including the fire detections, emission ratios, and plume injection heights used in the emission inventory/CTM (Sects. 5 and 6). We further apply independent measurements from ground sites and the FIREX-AQ airborne campaign to test the regional representativeness and interannual variability in our findings (Sects. 7 and 8). Finally, we discuss the potential implications of our findings, taking into account the uncertainties associated with the model results (Sect. 9).

2 Methods

2.1 WE-CAN aircraft campaign

The WE-CAN airborne campaign systematically characterized emissions and the chemical evolution of western US wildfire smoke with the NSF/NCAR C-130 research aircraft. The campaign was mainly based in Boise, ID, in July–September 2018 and sampled 27 fire plumes in the near field (some fires measured multiple times on different days) and various cases of regional and aged smoke (Lindaas et al., 2021; Permar et al., 2021). Table S1 in the Supplement summarizes the sampling time, fire location, and acres burned for specific fires sampled during WE-CAN.

Four sets of complementary VOC measurements were utilized to constrain BB emissions, including a proton-transfer-reaction time-of-flight mass spectrometer (PTR-ToF-MS, or PTR), trace organic gas analyzer (TOGA), advanced whole-air sampler (AWAS), and iodide (I⁻) adduct high-resolution time-of-flight chemical-ionization mass spectrometer (I-CIMS). The four instruments have different strengths and weaknesses in terms of analytical and separation powers, uncertainty, and measurement frequencies (Apel et al., 2010; Andrews et al., 2016; Palm et al., 2019; Permar et al., 2021).

We primarily focus on 14 VOCs or lumped VOC groups that are represented in the standard GEOS-Chem version 12.5.0 with observations assigned to the model speciation (Tables 1 and S2). Among them, 3 VOCs were mostly measured by discrete sampling with AWAS and in emission transects (nearest downwind with <2 h aging). Thus, we limit their model evaluation to emission ratios. These include ethane, lumped alkanes with four or more carbon atoms (or lumped ≥C₄ alkanes), and lumped ≥C₃ alkenes. The
other 11 VOC measurements used higher-frequency instruments, allowing for more comprehensive model evaluations along the C-130 (and DC-8) flight tracks. For these, we follow the data reduction described in Permar et al. (2021), mainly using PTR data with interferences corrected using co-deployed TOGA measurements and laboratory observations (Koss et al., 2018).

Figure S1 in the Supplement compares key VOCs measured by higher-frequency instruments in the entire WE-CAN (and FIREX-AQ) datasets. We find that PTR agrees with I-CIMS within ±20%–40% for formalic acid. PTR agrees with TOGA measurements within ~20% for formaldehyde, acetaldehyde, acetone, MEK, benzene, and toluene, with high correlation between each instrument (r = 0.93 to 0.99; and similar agreements are found in the FIREX-AQ dataset). PTR-measured xylenes are ~60% higher than in TOGA during WE-CAN (and lower by 20% in FIREX-AQ), but again they are highly correlated to each other. The difference in xylene measurements is possibly due to unknown fragmentation and/or under-characterized instrument sensitivity from likely varying isomer fractions in smoke plumes in PTR. For those reasons, we used TOGA xylene measurements whenever data are available (Sect. 4).

The emission ratios relative to CO in WE-CAN emission transects identified in Permar et al. (2021) are used to evaluate this key input in emission inventories. CO was measured at 1 Hz with 1 ppb accuracy with a Picarro G2401-m WS-CRDS analyzer during WE-CAN. All observations were taken from the WE-CAN 1 min merge data unless otherwise noted (version 4; https://www-air.larc.nasa.gov/cgi-bin/ArcView/firexaq?MERGE=1TS8, last access: 21 May 2023).

### 2.2 FIREX-AQ aircraft campaign and ground sites data

Two additional datasets are used to examine the broader representativeness and the year-to-year variability in our findings from WE-CAN. We use ground-level CO mixing ratios from nine western US sites measured during WE-CAN 2018 to assess the model prediction of regional BB emissions at the surface over the fire season. These include a mountaintop site at Mt. Bachelor Observatory, OR; a long-term ground station in Missoula, MT; and seven available Environmental Protection Agency (EPA) monitoring stations across the western states (Table S3 and Fig. 1; Laing et al., 2017; Selimovic et al., 2020; https://www.epa.gov/aqs, last access: 21 May 2023).

We also repeat the WE-CAN analyses using FIREX-AQ DC-8 aircraft observations that took place in July–September 2019. FIREX-AQ was a joint field campaign led by NOAA and NASA that investigated the chemistry and transport of smoke from both wildland and agricultural fires in 2019. Here we focus on the western US portion of FIREX-AQ, which represents 64% of the entire campaign data (Fig. 1 and Table S4). The DC-8 in FIREX-AQ systemically sampled 18 wildfires in the western USA, and here we use the 1 min merge data unless otherwise noted (version R1; https://www-air.larc.nasa.gov/cgi-bin/ArcView/firexaq, last access: 21 May 2023). The wildfire emission sizes during FIREX-AQ were less than during WE-CAN as reflected in the ZIP code.
by the GFAS total VOC emissions (20 versus 190 Gg C in the western USA on campaign-specific days) and the distribution of measured acetonitrile abundance in both campaigns (Fig. S2). Together with the surface CO measurements, they provide independent evidence to test if the model emission biases found from WE-CAN in 2018 are representative across the western USA and in different years.

2.3 GEOS-Chem chemical transport model

We employ GEOS-Chem nested grid simulations (version 12.5.0; Bey et al., 2001; http://www.geos-chem.org, last access: 21 May 2023, https://doi.org/10.5281/zenodo.3403111) to interpret the recent airborne observations and ground measurements in terms of new constraints on western US VOC emissions from wildfires. GEOS-Chem is driven by assimilated meteorology from the NASA Goddard Earth Observing System (GEOS). Here we use GEOS Forward Processing (GEOS-FP) meteorological inputs to drive GEOS-Chem nested grid simulations over North America for the WE-CAN (24 July–14 September 2018) and FIREX-AQ (22 July–5 September 2019) periods. The nested domain covers 10°–70° N and 140°–60° W, with 0.25° × 0.3125° (~25 km × 30 km, latitude × longitude) horizontal resolution and 47 vertical layers extending up to 0.01 hPa (Wang et al., 2004; Kim et al., 2015). The model boundary conditions are obtained from the global simulation at 4° × 5° resolution every 3 h. The transport/convection and emission/chemistry time steps of the nested simulation are 5 and 10 min, respectively. We carry out a full-year spin-up simulation at 4° × 5° resolution followed up by another 1-week spin-up at the nested resolution prior to the time periods of interest, to minimize effects from initial conditions.

The GEOS-Chem chemical mechanism includes detailed \( \text{HO}_x–\text{NO}_x–\text{VOC}–\text{ozone–halogen–aerosol} \) chemistry with a fully coupled troposphere and stratosphere (Park, 2004; Mao et al., 2010; Eastham et al., 2014; Schmidt et al., 2016). Dry deposition uses a standard resistance-in-series model (Wesley, 1989). Wet deposition includes scavenging of soluble tracers in convective updrafts, as well as rainout and washout of soluble tracers (Liu et al., 2001). Emissions are computed using the HEMCO module described by Keller et al. (2014). These include biogenic VOC emissions from MEGAn2.1 (Guenther et al., 2012) as implemented by Hu et al. (2015a) and anthropogenic emissions from the CEDS global emission inventory overwritten with the EPA’s national emission inventory 2011 (NEI 2011) for the USA (Hoesly et al., 2018). Below we describe aspects of the model configurations that are most relevant to this work.

We carried out several simulations driven with four different global BB emission inventories. An initial result sug-
suggests that the FINnv1.5 emission inventory predicted only 4%–8% of western US BB VOCs or CO emissions compared to those from the other three inventories, even though their total global emission estimates agree within 40%. This is likely due to fuel characterization errors for this region in FINnv1.5; thus we focus on simulations with GFED4s, GFASv1.2, and QFED2.4 for the analysis in this work. A recent study found that these three inventories strongly correlate with aircraft-derived hourly total carbon emissions during FIREX-AQ but generally underpredict BB and cannot capture the observed fire-to-fire variability (Stockwell et al., 2022). For simplification, we denote these three BB inventories as GFED4, GFAS, and QFED in the following discussion. We also note that the BB emission inventories, in the standard GEOS-Chem, may not contain a complete list of VOCs in the model. We thus implement their BB emissions in the base simulation (GEOS-Chem + GFAS; Table 2) by scaling the CO BB flux with WE-CAN field-measured emission ratios (ERs) from Permar et al. (2021). These species include MEK, formic acid, acetic acid, and lumped ≥ C_3 aldehydes in GEOS-Chem + GFAS (Table 1).

The standard GEOS-Chem version also implements different emission injection height schemes for each BB inventory, providing an opportunity to examine the impact of various plume height assumptions on the vertical distribution of trace gases. Specifically, GFED4 (and FINnv1.5) emissions, as incorporated in GEOS-Chem, are prescribed in the model surface layer and mainly rely on diffusion and convection (which depends on atmospheric turbulence and stability) for mixing before the chemistry operator. QFED prescribes 65% of BB emissions by mass evenly from the surface to the top of the planetary boundary layer (PBL), and the remaining 35% are evenly distributed between the PBL face to the top of the planetary boundary layer (PBL), and 35% are evenly distributed from the sur-

Table 2. Description of the model experiments.

<table>
<thead>
<tr>
<th>Simulation name</th>
<th>Biomass burning inventory</th>
<th>Injection height scheme</th>
<th>Diurnal representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inventory experiments</td>
<td>GFAS (base)</td>
<td>sf2mami^a</td>
<td>WRAP (2005)</td>
</tr>
<tr>
<td></td>
<td>GFED4</td>
<td>surface^b</td>
<td>Mu et al. (2011)</td>
</tr>
<tr>
<td></td>
<td>QFED</td>
<td>35% FT, 65% PBL^c</td>
<td>WRAP (2005)</td>
</tr>
<tr>
<td>Injection experiments</td>
<td>GFAS</td>
<td>35% FT, 65% PBL^c</td>
<td>WRAP (2005)</td>
</tr>
<tr>
<td></td>
<td>GFAS</td>
<td>surface^b</td>
<td>WRAP (2005)</td>
</tr>
<tr>
<td></td>
<td>GFAS</td>
<td>sf2mami^a</td>
<td>WRAP (2005)</td>
</tr>
<tr>
<td></td>
<td>GFAS</td>
<td>mami^d</td>
<td>WRAP (2005)</td>
</tr>
<tr>
<td></td>
<td>GFAS</td>
<td>apb_apt^e</td>
<td>WRAP (2005)</td>
</tr>
<tr>
<td>noBB</td>
<td>BB emissions turned off</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>3 × GFAS</td>
<td>Tripled BB emissions</td>
<td>sf2mami^a</td>
<td>WRAP (2005)</td>
</tr>
</tbody>
</table>

^a BB emissions are evenly distributed from the surface to the mean altitude of maximum injection (“mami”) from GFASv1.2.

^b BB emissions are released into the model surface layer and mixed into the atmospheric boundary layer via diffusion before advection and chemistry operators. ^c A total of 65% of BB emissions by mass are released within the planetary boundary layer (PBL), and 35% are released between the top of PBL and 5500 m in the free troposphere (FT). ^d BB emissions are released to the mean altitude of maximum injection from GFASv1.2. ^e BB emissions are evenly distributed from the bottom to the top of the plume from GFASv1.2. n/a: not applicable.

height and 5500 m. This approach was based on the distribution pattern of aerosol smoke plume heights from 5-year Multi-angle Imaging SpectroRadiometer (MISR) observations and was suggested to improve PAN simulations at high latitudes (Val Martin et al., 2010; Fischer et al., 2014).

GFAS, as implemented in the standard version of GEOS-Chem, releases emissions evenly from the model surface to the mean altitude of maximum injection (“mami”). GFAS also provides estimates of the top and the bottom of the plume at its native resolution (0.1° × 0.1° and daily). All three products are derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) FRP product and a plume rise model (PRM) at GFAS native pixels (Latham, 1994; Freitas et al., 2007). The PRM uses atmospheric profiles of meteorological parameters and fire information from European Centre for Medium-Range Forecasts (ECMWF) and MODIS observations to derive a full smoke detrainment profile and further to be translated into injection height information (Rémy et al., 2017). The BB-free region is regarded as plume heights of zero in the model. Thus, the plume heights would be artificially reduced when averaging to the coarser-than-native resolution (i.e., 0.25° × 0.3125° here). To account for this grid-dependent issue, we calculate emission-flux-weighted averages for those GFAS plume height products at the corresponding GEOS-Chem resolution.

For GFAS and QFED with temporal resolutions that vary daily, the standard GEOS-Chem prescribes a climatologically diurnal distribution profile that emits the majority (∼ 85%) of the daily BB emissions in the afternoon (local time) (WRAP, 2005). For GFED4 with monthly temporal resolution, the model distributes the daily fraction using MODIS active fire products and climatological mean diur-
nal cycles (Mu et al., 2011). These temporal patterns are in general consistent with observations in the western USA as wildfires tend to be most active in the afternoon. A recent study found that varying diurnal distribution using FRP observed from a geostationary satellite (so that diurnal cycles of BB emissions vary from grid to grid and from day to day) shows little improvement compared to the climatological approach representation at a campaign average for both WE-CAN and FIREX-AQ, at least for the western USA (Tang et al., 2022). Thus, in this work, we do not attempt to constrain the diurnal distribution of BB emissions as that would require continuous observations in the near field or from space.

Table 2 summarizes all the simulation experiments used in this study. These include three default simulations driven by the different emission inventories which all have different plume height schemes in the standard GEOS-Chem (“Inventory experiments”). In addition, we employ five different plume injection schemes in combination with the GFAS to test assumptions regarding BB emission vertical distribution (“Injection experiments”). Further, we carry out one simulation with BB emissions turned off (“noBB”) and another simulation with 3 times the default GFAS BB emissions (“3 × GFAS”) as additional sensitivity tests to examine the BB impact in the western USA. All the simulations were performed for the summer of 2018 and 2019, covering both the WE-CAN and the FIREX-AQ campaign periods.

To directly compare the model to the aircraft observations, the model outputs are sampled along the C-130 and DC-8 flight tracks (same location and altitude) and at the time of the flights for every minute. Then both observations and model results are averaged to the center of the model horizontal grid boxes and to transport resolution (0.25° × 0.3125°; 5 min). It is a general concern that Eulerian model results are not properly accounting for sub-grid features, partly resulting from point source emissions being needed to be diluted instantly to relatively coarse model grid sizes, particularly when compared to observations from aircraft campaigns targeting fresh fire smoke plumes. Thus, our model evaluation does not focus on individual fire cases but rather on the campaign average conditions, no-flow smoke environments, and trace : trace ratios. In addition, we apply the ground-based measurements over a longer term as an additional test as they should be less sensitive to any potential model biases due to the various assumptions or adjustments made for fire persistence, small fires, or fires obscured by clouds and haze in the inventories (Liu et al., 2020).

<table>
<thead>
<tr>
<th>Simulation</th>
<th>BB emissions turned off (“noBB”)</th>
<th>BB emissions 3 times default GFAS (“3 × GFAS”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFAS inventories</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BB inventory</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>GFED4 inventories</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>GFED5 inventories</td>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>

Figure 1 shows the VOC primary emissions over the western USA in the base simulation during the 2018 fire season (June–September, or JJAS). This 4-month period typically accounts for 70–90% of the annual acreage burned in this region (Jaffe et al., 2008). In GFAS, the JJAS contributes 85% of the 2018 annual BB emissions in the western USA. According to emission inventories chosen in GEOS-Chem, biogenic emissions are thought to be the dominant VOC source in summer in this region (2200 Gg C or ~75% of the total VOC emissions), followed by anthropogenic emissions (405 Gg C or ~15%) and BB emissions. In JJAS 2018, the total BB emissions from 14 VOCs in the model range from 220 to 340 Gg C (or 10% of the total). As we will show later, our model–observation comparisons suggest that the role of BB is significantly underestimated in the current CTMs.

In the 2019 fire season, the BB VOC emissions (40–75 Gg C or 1–3% of total primary VOC emissions) are 10%–30% of those in 2018 for this region and about 40%–50% of the 2019 annual BB emissions depending on which BB inventory is used. This shows the WE-CAN and FIREX-AQ aircraft campaigns sampled two distinct fire seasons which may reflect upper and lower bounds of wildfire activity in this region (https://www.nifc.gov/fire-information/statistics/wildfires, last access: 21 May 2023). Despite the large interannual variability in wildfire emissions, the western USA accounted for ~90% of BB VOC emissions in the contiguous United States (CONUS) in 2018 and ~60% in 2019 according to GFAS, confirming a significant fire influence exists in the western USA, which could also affect the rest of CONUS downwind (O’Dell et al., 2021).

The total BB VOC emission estimates in the western USA differ by 20%–40% across the three global inventories examined for the 2018 fire season. All emission inventories show similar spatial distributions as they all use MODIS satellite products such as active fire, burned areas, or FRP as inputs. However, larger differences between inventories occur for emission estimates for individual fires on specific days (more than a factor of 20), also shown in Bela et al. (2022) and Stockwell et al. (2022). These differences likely reflect the various assumptions or adjustments made for fire persistence, small fires, or fires obscured by clouds and haze in the inventories (Liu et al., 2020).

All three global BB inventories suggest aldehydes, alkanes, and alkenes are the most abundantly emitted VOCs from western US wildfires, largely consistent with recent field measurements (Perram et al., 2021). However, emission estimates for individual VOCs disagree by a factor of 1–5 in the western US fire season (Figs. 2 and S3). Emission estimates for xylenes show the largest difference (5 Gg C in GFED4 versus 1 Gg C in GFAS), while propane emissions agree within ±20% across the three inventories (8–10 Gg C). De-
Figure 2. Biomass burning VOC emission estimates for the 2018 fire season (JJAS) (black) and emission ratios (red) over the western USA in three global emission inventories. The emission ratios are regionally averaged from each inventory and are calculated from the regression of daily mean VOC and CO BB emission fluxes at each grid cell for the region. Error bars represent 95% confidence intervals from the bootstrapping resampling of the regression. We note that regionally averaged emission ratios derived from inventories might differ from those for individual fires derived from the full chemistry simulations used in Sect. 6. Values of zero indicate the species were not included in the BB emission inventory.

Despite the same FRP products being used, QFED is lower than GFAS by a factor of 2–3 for emission estimates of acetaldehyde, lumped $\geq C_4$ alkanes, and lumped $\geq C_3$ alkenes. This discrepancy can likely be explained by different emission ratios and speciation used for lumped compounds in these two inventories, in which QFED tends to have simpler speciation (Kaiser et al., 2012; Koster et al., 2015). For instance, $C_3$–$C_6$ and $C_8$ alkenes are incorporated in the GFAS, while only $C_3$ alkene is considered in the QFED. Besides, the emission ratios in QFED are primarily sourced from Andreae and Merlet (2001), whereas GFAS has incorporated updates from literature through 2009.

A recent global study comparing BB aerosol emissions from inventories suggests that the effective DM burned is the biggest contribution to divergent emission estimates across inventories (Carter et al., 2020). In contrast, we find that the regionally averaged ERs dominate disagreement in emission estimates for most VOCs across the three inventories (Figs. 2 and S4). These ERs are regionally averaged from each inventory and thus are functions of both assigned ERs for specific biome and vegetation classifications and are calculated from the regression of daily mean VOC and CO BB emission fluxes at each grid cell for the region from inventories.

We also find that, at least for the western USA, these three inventories agree on the amount of effective DM burned to within 40% (47–67 Tg in 2018, as calculated by dividing VOC and CO emission estimates by the corresponding regionally averaged EF). Even though the amount of effective DM does not drive the inventory–inventory discrepancy of emission estimates in the region, our model–observation evaluations in the following sections infer that it is likely too low. Indeed, using the National Interagency Fire Center burned area report ($\sim 2.42 \times 10^6$ ha for 2018 in the west), back-of-the-envelope calculation suggests that these global inventories’ effective DM burned per area is 19–28 Mg ha$^{-1}$. These values of biomass burned per area are at the low end of the range of estimates from the United States Forest Service (USFS) fuel consumption models, such as FOFEM and Consume (25–193 Mg ha$^{-1}$ for western US wildfires; Reinhardt et al., 1997; Drury et al., 2014). Besides, limited field fuel consumption measurements of western mixed conifer forest wildfires, ranging from 32–44 Mg ha$^{-1}$ (Campbell et al., 2007; Hyde et al., 2015), are also higher than values from the global BB emission inventories. Taken together, these calculations suggest that the three global BB inventories underestimate the amount of DM burned in the western USA for
the inventories capture the enhancement patterns observed except for surface toluene and CO within ±30 %–40 %. All the inventories capture the enhancement patterns observed by the C-130, both elevated altitudes and timing with high correlations with observations (r = 0.7 to 1.0 in 5 min averaged data). The sensitivity run with no BB emissions (noBB) indicates that wildfire is a significant source for CO and primary VOCs including propane, benzene, and toluene during WE-CAN (enhanced by 2–3 times compared to noBB) but a lesser source for oxygenated VOCs (OVOCs), especially for formaldehyde (Figs. 4 and S5). The model driven by GFAS (GEOS-Chem + GFAS) tends to simulate slightly higher and better VOCs than GFED4 and QFED, possibly reflecting that GFAS has more accurate ERs as discussed later in Sect. 6.

All three inventory experiments significantly underestimate observed CO and VOCs, except for MEK. In the middle to lower troposphere (>500 hPa), simulations reproduce 40 %–70 % of the observed abundance of CO, benzene, toluene, and acetone and 30 %–40 % of the observed propane, formaldehyde, acetaldehyde, and lumped ≥C₃ aldehydes but only 0 %–10 % of the observed organic acids. The model suggests mixed performance for xylenes, i.e., a high bias of 0 %–100 % in the lower troposphere and a low bias of 50 %–100 % in middle troposphere. In a relatively clean environment (<500 hPa), the simulations show relatively small negative biases for all compounds and tend to match observations in generally clean or well-mixed environments during WE-CAN.

Unlike other VOCs and CO, MEK is systematically overestimated by 50 %–300 % throughout the middle to lower troposphere in all simulations including noBB but is reproduced in a relatively clean environment (<500 hPa). Similar positive model bias has been reported in a recent study comparing GEOS-Chem to a comprehensive suite of airborne datasets over North America (Chen et al., 2019). This is likely due to the overestimation of MEK or its precursors in the EPA NEI and/or MEGAN inventories (Yáñez-Serrano et al., 2016), as such large high model bias exists even when the BB influence is removed (Fig. 5). Thus, further evaluation is needed for the sources of MEK and its precursors in anthropogenic and biogenic emission inventories.

We further refine the analysis in low-/no-smoke conditions by filtering out data when either the observed acetonitrile (CH₃CN) mixing ratio, a known BB tracer, is >159 ppt (25th quantile of CH₃CN) or the enhancement ratio of CH₃CN relative to CO is >2.01 ppb/ppm⁻¹ (Huangfu et al., 2021). The vertical profiles after applying this filter are shown in Figs. 5 and S6 and represent about one-third of the sampling time during WE-CAN, allowing us to examine the non-BB-related processes/emissions. Compared to the full campaign data, the observations of CO and all VOCs in low-/no-smoke conditions are lower by a factor of 2 or more, confirming the important influence from BB in the western USA during WE-CAN. The simulations capture the observed CO, benzene, and toluene in this clean environment but still underestimate the rest of the VOCs (especially OVOCs) by 10 %–90 %, except for MEK. The low model bias for formaldehyde in the free troposphere can
Figure 4. Median vertical profiles of observed VOC mixing ratios in the western USA during WE-CAN. GEOS-Chem simulations driven by three different biomass burning emission inventories (GFED4s, GFASv1.2, and QFED2.4) are compared to observations. Also shown are two model sensitivity tests with biomass burning emission turned off (noBB) and with tripling GFASv1.2 biomass burning emission (3×GFAS). Model results are sampled along the flight tracks at the time of the flights, and observations are regridded to model resolution. Profiles are binned to the nearest 30 hPa. Horizontal bars show the 25th–75th percentile range of measurements in each vertical bin. The number of observations in each bin is given on the right side of each panel. Results are filtered to include only data where the number of data points for the pressure bin is larger than 10.

be partly due to underestimated oxidation of CH$_4$ or other precursors (Zhao et al., 2022). The negative model bias for acetaldehyde, formic acid, and acetic acid in the PBL may be related to reasons including missing or underestimated precursors from biogenic emissions (Millet et al., 2010, 2015; Paulot et al., 2011). The negative bias for acetone in the middle–upper troposphere may reflect a poorly constrained global background from ocean sources in GEOS-Chem (Wang et al., 2020). Nevertheless, the negative model bias in the low-/no-smoke conditions sampled during WE-CAN (Fig. 5) is much smaller than the BB-influenced environment. Thus, underestimation in the low-/no-smoke conditions does not explain model underestimation across compounds in the full campaign dataset (Fig. 4).

We calculate the average model biases that are due to BB processes for each species using the enhancements between the full campaign dataset and the low-/no-smoke conditions. Given the calculation of primary trace gases (CO, propane, benzene, and toluene), we conclude that the model potentially underestimates BB emissions or related processes by a factor of 3–7 in the GFAS while the bias can slightly vary in the GFED or QFED. Thus, we further carry out a sensitivity run by tripling the GFAS emissions in the model (GEOS-Chem + 3×GFAS) as a test of the BB impact in the western USA. Figures 3 and S5 show that tripling BB primary emissions results in evident improvements and reproduces the observed levels for CO and most primary VOCs (propane, benzene, and toluene). The improvement for xylenes is moderate due to other model errors in the averaged OH reaction rate constant and ER (Sect. 6).

GEOS-Chem + 3×GFAS has elevated simulated abundance for OVOCs to various degrees compared to the base run. For acetaldehyde and acetone, we find that 3×GFAS brings the model close to the measurement uncertainty. For formaldehyde, formic acid, acetic acid, and lumped ≥C$_3$ aldehydes, tripling the primary BB emissions of these species (and their precursors that are included in the model) does not significantly improve the model–observation discrepancy (the improvement is within 5%). Since 3×GFAS mostly corrects the model error in primary BB emissions, this underestimation suggests the current model likely misses large secondary sources of these compounds in BB plumes due
Figure 5. Median vertical profiles of observed VOC mixing ratios in the western USA for low-/no-smoke conditions sampled in WE-CAN. GEOS-Chem simulations driven by three different biomass burning emission inventories (GFED4s, GFASv1.2, and QFED2.4) are compared to observations. Results are filtered to include only data coincident with the bottom 25th percentile of observed acetonitrile, where $\Delta$CH$_3$CN/$\Delta$CO is less than 2.01 ppb ppm$^{-1}$ and where the number of data points for the pressure bin is larger than 10. Model results are sampled along the flight tracks at the time of flights, and observations are regridded to model resolution. Profiles are binned to the nearest 30 hPa. Horizontal bars show the 25th–75th percentile range of measurements in each vertical bin. The number of observations in each bin is given on the right side of each panel.

Eulerian models are known to have trouble preserving sub-grid features such as concentrated fire plumes over time due to rapid dissipation by numerical diffusion (Eastham and Jacob, 2017; Rastigejev et al., 2010). Campaigns targeting plumes like WE-CAN can get particularly intense and thus deviate from the climatologically diurnal distribution of BB emissions used in the model, resulting in low model bias when compared to aircraft measurements. In addition, any wind direction or plume height errors in the model would result in the model’s aircraft diagnostics missing the fire plume when the real aircraft sampled it, contributing to some amount of low bias. Finally, if the plumes are narrower than $\sim$25 km (and the aircraft transect lengths are also narrower than $\sim$25 km), then the plume will be diluted in the model grid box more than the plane observed (even when including the transect portions outside of the plumes), also contributing to a low model bias. In addition to the BB emissions, those above factors due to fire sub-grid features may all have contributed to the low model bias in the aircraft analysis, but it is difficult to fully tease them out if at all possible. We thus consider the model bias revealed here as the upper limit of BB emission negative bias in the global inventories (Sects. 3 and 8).

5 Model uncertainties in fire detection and emission injection heights

To explore causes for the underestimation of BB emissions for these three emission inventories, we first determine if the inventories have detected the 27 individual fire plumes sampled in WE-CAN. A fire is considered to be detected if the inventory has registered any CO emissions in the model surface grid box at its location when the C-130 arrives. Table S1 shows that all the BB inventories (including FINNv1.5) capture all the sampled fires. BB emission inventories typically rely on space-based observations of burned area or FRP (i.e., MODIS Terra and Aqua fire products) for fire detections. For example, MCD64A1 burned area products are applied in GFED4 and MOD14–MYD14 FRP products are used in
GFAS and QFED. During WE-CAN, wildfires were mostly sampled in the late afternoon when fires were the most active. The fires sampled by the C-130 tended to have developed well-defined plumes that were visible from geostationary GOES-16 or GOES-17 GeoColor images in the morning of the same day when flight planning was finalized. Our finding suggests that the fire detection products from low-orbit satellites commonly used in global BB emission inventories are efficient at detecting large fires in the western USA that tend to burn for several days if not weeks or months.

We further examine the impact of the assumed injection altitude of BB emissions by conducting sensitivity tests using five different BB injection height schemes (Table 2; Sect. 2.3). Figure S7 shows almost identical model vertical profiles in the five plume injection experiments, particularly in the free troposphere. In the middle troposphere, the simulations with higher plume injection heights tend to show larger enhancements; in the PBL, releasing BB emissions at the surface tends to result in the highest surface mixing ratios among the experiments. But the differences across simulations are within ±10 % except for benzene and toluene (about ±40 % near the surface). The model does not appear to be highly sensitive to assumptions regarding BB injection heights in the western USA at ~25 km resolution. This insensitivity is likely because the trace gas emissions from large wildfires are efficiently lifted into the free troposphere by strong vertical mixing in the summer (Chen et al., 2009; Jian and Fu, 2014). However, the choice of plume injection heights can still be important for secondary production and downwind areas (Tang et al., 2022). For example, daily mean ozone concentrations vary by up to 14 % or 4 ppb at the surface in our injection experiments. Thus, the impact of various BB emission injection schemes on surface air quality needs further investigation, especially for populated downwind regions.

6 Model uncertainties in emissions ratios

Emission ratios (ERs; often interchangeable with emission factors or EFs) can be a source of uncertainty in BB emissions estimates if they are poorly characterized or unmeasured (e.g., Akagi et al., 2011; Urbanski et al., 2011). We calculate ERs from the slope of the reduced major axis regression of VOCs and CO measured (and simulated) in emission samples. In order to calculate ERs, plume samples with physical ages less than 2 h in the WE-CAN campaign and less than 1 h in the FIREX-AQ campaign are used; these are deemed to be relatively fresh, with minimal or no secondary production. We note the observed ERs derived here using the 5 min averaged data tend to agree with what Permar et al. (2021) reported within 20 %, despite Permar et al. (2021) calculating ERs from 1 s observations and using the integration approach. Also, calculating observed and simulated ERs in a consistent way and according to the temporal and spatial resolution of the model can provide a valuable constraint on the overall model processes in terms of BB emission locations, timing, transport, and chemistry in fire-influenced environments.

Figure 6 illustrates this approach with scatterplots of a subset of observed VOCs and CO in emission transects and their comparison to the simulated relationship in GEOS-Chem + GFAS. The model shows the strong correlations between VOCs and CO (r = 0.7 to 1.0), suggesting GFAS captures the regional BB locations and timing sampled by the C-130 (Sect. 5). We find GFAS ERs agree with observed ERs within 30 % or better for formaldehyde, acetaldehyde, benzene, toluene, and lumped ≥C4 alkanes. GFAS is either too high or too low by 50 %–70 % for ethane, propane, and acetone. Overall, GEOS-Chem + GFAS tends to produce higher and more accurate ERs than the other two inventories (Figs. 7 and S8). Some notably large errors in simulated ERs (≥ a factor of 2) include acetaldehyde in QFED and acetone, MEK, benzene, and toluene in GFED4.

The modeled abundance and ERs of xylenes and lumped ≥C3 alkenes are significantly underestimated across all inventory experiments. These two lumped VOC groups are highly reactive, with lifetimes of ~1 h (assuming an average in-plume OH concentration of 1 × 107 molec. cm−3 and an OH reaction rate constant kOH of 23.1 × 10−12–25.0 × 10−12 cm3 molec.−1 s−1). Errors in their loss via OH reactions due to incorrect OH concentration or kOH could distort their simulated abundance and ERs. Model bias in OH concentration would affect all primary species in the same direction, and reactive VOCs would be particularly sensitive to such error. Thus, we use aromatic hydrocarbon–hydrocarbon relationships to diagnose if there are any major model OH biases in the current version. Figure S9 shows that the base model can capture the observed toluene–benzene relationship, in terms of both emission ratios and their relative decay rates. This agreement indicates the good reproduction of the OH level in the model, and future analysis is needed for evaluating current kOH in the model.

Further, we find that kOH for xylenes in recent GEOS-Chem versions has been updated based on new assumptions. The GEOS-Chem version 12.5.0 used in this analysis assigns 23.1 × 10−12 cm3 molec.−1 s−1 as kOH for xylenes, based on the assumption that m-xylene is the dominant isomer (Fischer et al., 2014). Other studies using the fractions of xylene isomers observed in urban atmospheres for a weighted kOH have suggested values of 13.2 × 10−12–17.0 × 10−12 cm3 molec.−1 s−1, about 25 %–40 % lower than used here, which, if updated, would result in higher simulated xylenes (Atkinson and Arey, 2003; Hu et al., 2015b; Bates et al., 2021). Therefore, correcting kOH could partly reconcile the negative model bias for xylene ERs (i.e., 0.15 in corrected simulations vs. 0.32 ppb ppm−1 in observations). The isomer fractional information for other lumped species and their chemistry in various environments
Figure 6. Biomass burning VOC emission ratios from wildfire emission transects sampled on the C-130 during WE-CAN (black). Also shown are the corresponding GEOS-Chem + GFAS simulations (blue). Model results are sampled along the flight tracks at the time of flights every 1 min, and observations (and model outputs) are regridded to model resolution (5 min and $0.25^\circ \times 0.3125^\circ$). Lines represent the best fit of the data using the reduced major axis regression, with the regression parameters given in the equations.

Figure 7. Summary of biomass burning VOC emission ratios for western US wildfires observed on the C-130 during WE-CAN and the DC-8 during FIREX-AQ. Also shown are the emission ratios in simulations driven by three different BB emission inventories. Model results are sampled along the flight tracks at the time of flights every 1 min, and observations (and model outputs) are regridded to model resolution (5 min and $0.25^\circ \times 0.3125^\circ$). Emission ratios are calculated from the reduced major axis regression (RMA) of VOC and CO, with error bars representing the 95% confidence interval from the bootstrapping resampling of the regression. Values of zero indicate either the species were not included in the BB emission inventory in the standard GEOS-Chem or the ER calculation fails to reach the statistical threshold ($r^2 < 0.4$) in the RMA regression.
is less known; thus future investigation is needed to refine and assess the chemical impact of these lumped species.

7 Model evaluation with ground-based observations

The national wildland fire burned area in 2019 was only about half that in 2018 (https://www.nifc.gov/fire-information/statistics/wildfires, last access: 21 May 2023). This is also reflected in the different acetone distributions measured between the two aircraft campaigns (median 295 ppt during WE-CAN versus 205 ppt during FIREX-AQ; Fig. S2). To examine the year-to-year variability and regional representativeness of findings inferred from the WE-CAN C-130 measurements, we expand the analysis to observations from nine ground-based sites in 2018 and the FIREX-AQ DC-8 aircraft in 2019. The ground stations span several urban areas that are regularly affected by wildfire smoke. More importantly, the longer-term stationary measurements are further downwind in a better-mixed environment and physically unable to target plumes, and they can thus provide a counter-test to the contribution of the other factors from fire sub-grid features to model bias relative to aircraft observations that target the plumes (Sect. 4).

Figure 8 shows that most of the nine ground sites were heavily impacted by wildfire smoke in the 2018 summer, as indicated by elevated CO mixing ratios up to 250 ppb or higher lasting for a few days at times, while the general urban background CO is about 150–200 ppb (Ptister et al., 2011; Kim et al., 2013; Lopez-Coto et al., 2020; Gonzalez et al., 2021). Using the noBB and the base simulations, we define “BB-impacted days” as days when the modeled CO daily mean is increased by more than 20% relative to the noBB run, and the rest of the days are termed low-/no-smoke days. By this definition, Seattle and Denver were least affected by BB in 2018 among the nine sites but still experienced 7–8 BB-impacted days out of 55 d. The rest of the sites all experienced ≥ 25 BB-impacted days, according to GEOS-Chem + GFAS. In general, the base model captures the daily variation in the observed CO (r > 0.40 at all sites, with six sites having r > 0.65). In Seattle and Denver, anthropogenic emissions dominated local CO abundance and variability in 2018. The US EPA NEI appears to have spatial biases as the model background bias (loosely calculated by the 5th-percentile CO mixing ratio) tends to be less than 70 ppb. Thus, similar to the findings in Sect. 4, correcting the model background CO bias (due to anthropogenic emissions or global background) is not enough to reconcile the large model–observation discrepancy. We find that the 3 × GFAS simulation systematically improves the model mean bias to various degrees across the western USA for the seven fire-influenced sites without degraded correlation coefficients with observations.

8 Model evaluation with FIREX-AQ aircraft observations

Figures 9 and S10 show the model evaluation with FIREX-AQ DC-8 VOC observations for the western USA. Observed VOC mixing ratios during FIREX-AQ are lower than in WE-CAN for this region partly due to fewer BB emissions in 2019 (Sect. 3). Overall, our findings for 2019 FIREX-AQ are consistent with the 2018 WE-CAN evaluation: the base simulation tends to underestimate all observed VOCs but MEK by a factor of 2–12 in the middle to lower troposphere. When we restrict the analysis to the low-no-smoke environment, the base model also underestimates OVOCs and these negative model biases tend to be 40%–100% for the entire campaign average (Fig. S11). The model improvement for primary VOCs from tripling BB emissions is significant across the troposphere but not as obvious as during WE-CAN due to smaller BB emissions in 2019 (Sect. 3). Both WE-CAN and FIREX-AQ observations imply that the model misses substantial sources for OVOCs, particularly formaldehyde, formic acid, acetic acid, and lumped ≥ C₃ aldehydes.

We do not attempt to evaluate the modeled ERs for FIREX-AQ because the inventories do not update ERs for different years. Figure 7 shows the observed ERs in WE-CAN and FIREX-AQ are consistent within the combined instrument uncertainty (±40%) for a majority of VOCs in western fuel types. Given the observational constraints in ERs and primary BB emissions, we infer that the above missing OVOC sources in the model are most likely from photo-chemical reactions in smoke plumes.

9 Implications for total biomass burning VOC emissions in the western USA

We infer the systematic underestimation of simulated CO and individual VOCs in the western USA is mostly driven by the low bias of effective dry matter burned in fire-detected areas.
Figure 8. Time series of daily averaged CO mixing ratios from nine ground sites in the western USA during the 2018 WE-CAN campaign. Also shown are three GEOS-Chem simulations (the base simulation GFAS in blue, 3 × GFAS in gray, and noBB in pink). Biomass burning emissions are injected evenly from the surface to the mean altitude of maximum injection height in the model (Table 2). The shaded area represents BB-impacted days as defined in the text. The locations of the nine ground sites are provided in Fig. 1 and Table S3. The date format is month-day.

Sensitivity tests with tripled BB emissions result in better agreement between observations and model outputs, particularly for primary VOCs. Thus, our best estimate of the BB primary emissions of the 14 modeled VOCs for the western US 2018–2019 fire seasons is 120–1020 Gg C, which is 3 times the default emission estimates in three BB inventories. This is also ∼5%–30% of the total VOC emissions from primary sources for the western US fire seasons. However, the model still underpredicts OVOCs, even with tripled BB primary emissions; we are thus unable to constrain secondary production of BB VOCs in this work.

The above BB emission estimates are derived from 14 modeled VOCs with BB representation in three BB inventories (Table 1). However, the total ER of these 14 BB VOCs only accounts for half of the total measured VOC ERs from 161 species observed during WE-CAN (75 versus 150 ppb ppm⁻¹; Permar et al., 2021). The uncharacterized BB VOCs in the model mean that there is a significant quantity of missing reactive organic carbon fluxes in many major BB emission inventories and CTMs. Their chemical
and health impacts on the regional and global scale remain largely unexplored (Carter et al., 2022; Permar et al., 2023). Considering both underpredicted dry matter burned and uncharacterized VOCs, we infer that BB contributed ∼10%–45% (or 240–2040 Gg C) of the total VOC primary emissions in the western USA during the 2018–2019 fire seasons, which is far more significant than common model representations as in Fig. 1.

10 Conclusions

We performed nested GEOS-Chem simulations and compared them with observations from two recent airborne campaigns and nine surface sites to constrain the BB CO and VOC emissions in the western USA. We evaluated three widely used global BB emission inventories including potentially significant errors in their dry matter burned, fire detection efficiency, injection heights, and emission ratios. Based on the model–observation comparison, we provided updated emission estimates of BB VOCs for both modeled and uncharacterized VOCs during two different fire seasons in the western USA.

In the standard GEOS-Chem, BB VOC emissions in the western USA rank as third in the rankings of total VOC primary sources (including biogenic and anthropogenic emissions). Despite large interannual variability, the western USA accounted for 60%–90% of BB VOC emissions over the CONUS in 2018 and 2019. Across three global BB inventories, total BB VOC emission estimates in the western USA agreed with each other within 30%–40%. However, estimates for individual VOCs can differ by a factor of 1–5, mostly driven by regionally averaged emission ratios (reflecting a combination of assigned ERs for specific biome and vegetation classifications) rather than effective biomass burned.

We found that simulations driven by three different BB inventories produce similar CO and VOC abundances. The model reproduced the plume enhancements in the locations observed in WE-CAN but showed negative biases for CO and VOCs (except MEK). Better model performance was found in relatively clean environments. By comparing BB-
impacted abundance enhancements between no-/low-smoke times and the entire campaign, we found that the model, regardless of which BB inventory was used, underestimated the BB emissions for primary compounds by a factor of 3–7; these include CO, benzene, toluene, and propane. For OVOCs that have both primary and secondary sources including formaldehyde, formic acid, acetic acid, and lumped ≥C_{3} aldehydes, tripling the BB emissions cannot fully explain the negative model bias; the model–observation comparison likely pointed to a large amount of missing secondary production in BB-impacted conditions in GEOS-Chem, which could account for the remaining bias. Unlike other VOCs, MEK was overestimated by a factor of 2–4 throughout the middle to lower troposphere, due to the overestimation of MEK itself or its precursors in the EPA NEI and MEGAN emission inventories. Tripling the BB emissions in GFAS reproduced observed mixing ratios for primary compounds but showed no or less significant improvement for OVOCs.

We found that the fire detection products in all the inventories detected the large fires sampled in the WE-CAN campaign. GEOS-Chem vertical profiles were not strongly sensitive to the various tested BB injection height schemes, as constrained by the observed VOC vertical profiles during WE-CAN. This is likely because strong and efficient vertical mixing during hot and dry summers in the western USA dominates the vertical transport processes. However, different injection height assumptions influenced the modeled downwind surface ozone mixing ratios (i.e., daily mean ozone differed by up to 14 % or 4 ppb); thus, the influence of injection heights on surface air quality requires further investigations.

We evaluated modeled ERs with WE-CAN and FIREX-AQ observations and found that GFAS performs slightly better than the QFED or GFED4 inventories for both VOC-CO correlations and ER values. GEOS-Chem + GFAS captured the observed ERs in aircraft emission transsects within 30 % for formaldehyde, acetaldehyde, benzene, toluene, and lumped ≥C_{4} alkanes and within 50 %–70 % for ethane, propane, and acetone. We also found the modeled abundance and ERs of xylens and lumped ≥C_{3} alkenes are significantly underestimated across all inventory experiments, likely reflecting the overestimation of the OH reaction rate constant k_{OH} used in the model.

Given that the errors in fire detection, plume injection, and ERs are relatively small, we infer that the underestimation of BB emissions in these inventories (a factor of 3–7) is likely due to underpredicted dry matter burned, which is also supported by our back-of-the-envelope calculation of effective DM burned. However, we cannot rule out the uncertainties in the nested GEOS-Chem (0.25° × 0.3125°) not being able to fully resolve the sub-grid features of BB emissions. Therefore, the above findings revealed by 2018 WE-CAN observational constraints are further tested for their regional representativeness and interannual variability with observations from nine western US ground sites and the 2019 FIREX-AQ airborne campaign. Compared to the ground-based “downwind” CO measurements, GEOS-Chem + GFAS captures the observed BB smoke events but underpredicts the mixing ratios in most cases. Tripling the BB emissions reduces the model’s negative bias across the western USA without degrading the correlation coefficients with observations. Repeating the analyses with FIREX-AQ observations also confirms the above conclusions.

Constrained by 2018 and 2019 airborne and ground measurements, the 14 BB VOCs included in the model contributed to 120–1020 Gg C of primary emissions in the western US 2018–2019 fire seasons. However, the total emission ratio relative to CO for these 14 VOCs in GEOS-Chem only accounted for half of that from the 161 measured VOCs in wildfire smoke, pointing to a significant quantity of uncharacterized reactive organic carbon fluxes that were missing in many current BB emission inventories and CTMs. Thus, accounting for both these missing species and underestimated DM burned, the total BB VOC emission estimates can reach 240–2040 Gg C or 10 %–45 % of the total primary VOC emissions in the western US fire seasons, highlighting a significant role of wildfires in US air quality.

**Code availability.** The GEOS-Chem model is publicly available at https://doi.org/10.5281/zenodo.3403111 (The International GEOSChem User Community, 2019). Corresponding analysis codes are available on request.

**Data availability.** The WE-CAN and the FIREX-AQ campaign data are available at: https://www-air.larc.nasa.gov/cgi-bin/ArcView/firexaq?MERGE=1 (last access: 21 May 2023).

The 2018 MBO data are available at: https://digital.lib.washington.edu/researchworks/handle/1773/46659 (last access: 21 May 2023).

**Supplement.** The supplement related to this article is available online at: https://doi.org/10.5194/acp-23-5969-2023-supplement.

**Author contributions.** WP, VS, DK, RJY, RSH, ECA, ITK, JLC Jr., APS, DAJ, AF, MMC, GIG, CW, and EVF measured CO and VOC data and provided input on the manuscript. LJ performed the modeling with plume injection inputs from JRP. LH and LJ formulated the research question and prepared the manuscript with contributions from all co-authors.

**Competing interests.** The contact author has declared that none of the authors has any competing interests.
Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Special issue statement. This article is part of the special issue “The role of fire in the Earth system: understanding interactions with the land, atmosphere, and society (ESD/ACP/BG/GMD/NHESS inter-journal SI)”. It is a result of the EGU General Assembly 2020, 4–8 May 2020.

Acknowledgements. The WE-CAN data were collected using NSF’s Lower Atmosphere Observing Facilities, which are managed and operated by NCAR’s Earth Observing Laboratory.

The authors acknowledge high-performance computing resources and support from Cheyenne (https://doi.org/10.5065/D6RX99HX) provided by the NCAR Computational and Information Systems Laboratory, sponsored by the NSF, and the University of Montana’s Griz Shared Computing Cluster (GSCC). We also thank Joel A. Thornton, Teresa L. Campos, Glenn S. Diskin, Dirk Richter, Patrick R. Veres, Joshua P. Schwarz, and Donald R. Blake for providing other WE-CAN and FIREX-AQ measurements used in this work.

Financial support. This study was supported by NASA (grant no. 80NSSC20M0166); NSF (EPSCoR Research Infrastructure grant no. 1929210 and AGS grant nos. 2144896 and 1950327); Montana NASA EPSCoR Research Initiation funding; and the NOAA Climate Program Office’s Atmospheric Chemistry, Carbon Cycle and Climate program (grant no. NA20OAR4310296). The 2018 WE-CAN field campaign was supported by the US NSF (AGS grant no. 1650275, University of Montana; grant no. 1650786, Colorado State University; grant no. 1650288, University of Colorado Boulder; grant no. 1650493, University of Wyoming; grant no. 1652688, University of Washington; and grant no. 1748266, University of Montana) and NOAA (grant no. NA17OAR4310010, Colorado State University, and grant no. NA16OAR4310100, University of Montana). The Mt. Bachelor Observatory was supported by the NSF (grant no. AGS-1447832) and NOAA (contract no. RA-133R-16-SE-0758). This material was also based upon work supported by the NCAR, which is a major facility supported by the NSF under cooperative agreement no. 1852977.

Review statement. This paper was edited by Holger Tost and reviewed by two anonymous referees.

References


Western Regional Air Partnership: 2002 fire emission inventory for the WRAP region – Phase II, Air Sciences Inc, Denver, CO, 2005.


