Estimating North American background ozone in U.S. surface air with two independent global models: Variability, uncertainties, and recommendations

A.M. Fiore a,*, J.T. Oberman c, M.Y. Lin d, L. Zhang e, f, O.E. Clifton a, D.J. Jacob e, V. Naik g, L.W. Horowitz d, J.P. Pinto h, G.P. Milly b

a Department of Earth and Environmental Sciences, Columbia University, 61 Route 9W, Palisades, NY 10964, USA
b Lamont-Doherty Earth Observatory of Columbia University, 61 Route 9W, Palisades, NY, USA
c Nelson Institute Center for Sustainability and the Global Environment (SAGE), University of Wisconsin-Madison, Madison, WI, USA
d NOAA Geophysical Fluid Dynamics Laboratory and Atmospheric and Oceanic Sciences, Princeton University, 201 Forrestal Road, Princeton, NJ, USA
e School of Engineering and Applied Sciences, Harvard University, 29 Oxford Street, Cambridge, MA, USA
f Department of Atmospheric and Oceanic Sciences & Laboratory for Climate and Ocean-Atmosphere Studies, School of Physics, Peking University, China
g UCAR/NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA
h U.S. EPA, National Center for Environmental Assessment, Research Triangle Park, NC, USA

HIGHLIGHTS

- North American Background (NAB) surface O3 provided by two independent global models.
- Models often bracket observations implying value in multi-model approach.
- NAB is highest at high-altitude sites in spring where it correlates with total O3.
- NAB is lowest over the Eastern U.S. in summer where it generally does not correlate with total O3.
- Largest model differences in NAB due to stratosphere, wildfires, lightning, isoprene.

ARTICLE INFO

Article history:
Received 26 December 2013
Received in revised form 24 July 2014
Accepted 26 July 2014
Available online 30 July 2014

Keywords:
Surface ozone
Background ozone
Air pollution
Air quality
Exceptional events

ABSTRACT

Accurate estimates for North American background (NAB) ozone (O3) in surface air over the United States are needed for setting and implementing an attainable national O3 standard. These estimates rely on simulations with atmospheric chemistry-transport models that set North American anthropogenic emissions to zero, and to date have relied heavily on one global model. We examine NAB estimates for spring and summer 2006 with two independent global models (GEOS-Chem and GFDL AM3). We evaluate the base simulations, which include North American anthropogenic emissions, with mid-tropospheric O3 retrieved from space and ground-level O3 measurements. The models often bracket the observed values, implying value in developing a multi-model approach to estimate NAB O3. Consistent with earlier studies, the models robustly simulate the largest nation-wide NAB levels at high-altitude western U.S. sites (seasonal average maximum daily 8-h values of ~40–50 ppb in spring and ~25–40 ppb in summer) where it correlates with observed O2. At these sites, a 27-year GFDL AM3 simulation simulates observed O3 events above 60 ppb and indicates that year-to-year variations in NAB O3 influence their annual frequency (with NAB contributing 50–60 ppb or more during individual events). During summer over the eastern United States (EUS), when photochemical production from regional anthropogenic emissions peaks, NAB is largely uncorrelated with observed values and it is lower than at high-altitude sites (average values of ~20–30 ppb). Four processes contribute substantially to model differences in specific regions and seasons: lightning NOx, biogenic isoprene emissions and chemistry, wildfires, and stratosphere-to-troposphere transport. Differences in the representations of these processes within the GFDL AM3 and GEOS-Chem models contribute more to uncertainty in NAB estimates, particularly in spring when NAB is highest, than the choice of horizontal resolution within a single model.
1. Introduction

The United States Environmental Protection Agency (U.S. EPA) sets National Ambient Air Quality Standards (NAAQS) to protect human health and welfare. Under the Clean Air Act, ground-level ozone (O$_3$) is regulated as a criteria air pollutant, reviewed every five years to assess and incorporate the best available scientific evidence. Following these reviews, the level for the O$_3$ NAAQS has been lowered over the past decade, from 0.08 ppm in 1997 to the current level of 0.075 ppm (75 ppb) in 2008, with proposals calling for even lower levels, within a range of 60–70 ppb on the basis of the latest health evidence (Federal Register, 2010). A location is considered to be in violation of the O$_3$ NAAQS when the three-year-average of the fourth highest daily maximum 8-h average O$_3$ (MDA8) exceeds the current 75 ppb level. In order to better understand how the O$_3$ NAAQS can be attained most effectively, a fundamental, quantitative understanding of the background O$_3$ – both its magnitude and variability – over the United States is needed.

The first draft of the current U.S. EPA Policy Assessment (EPA, 2013) and McDonald-Buller et al. (2011) describe the relevance of background O$_3$ in the U.S. national O$_3$ standard-setting process. Here we review recent model estimates for background O$_3$ (Table 1) and compare simulations from two independent models (GEOS-Chem and GFDL AM3) in the context of observational constraints with a focus on spatial, seasonal, and daily variability. Differences between the models provide a first estimate of the error in our quantitative understanding. A process-oriented multi-model approach, tied closely to in situ and space-based observations, can harness the strengths of individual models to provide information requested by air quality managers during both the standard-setting and implementation processes.

The term “background” is ambiguous, with several definitions used in practice to estimate it from observations and models (e.g., see discussion in Fiore et al., 2003). In the context of a review of the NAAQS, it is useful to define background O$_3$ concentrations in a way that distinguishes the O$_3$ produced from precursor emissions that are relatively less controllable versus from precursor emissions that are relatively more controllable through U.S. policies. The U.S. EPA thus defines a North American Background (NAB) as the O$_3$ levels that would exist in the absence of continental North American (i.e., Canadian, U.S., and Mexican) anthropogenic emissions (EPA, 2006). NAB includes contributions from natural sources (stratospheric intrusions, emissions of precursors from natural sources (e.g., wildfires, lightning, biogenic) throughout the globe, anthropogenic methane (CH$_4$), and emissions of anthropogenic pollutants from countries outside North America that contribute to global O$_3$ abundances. This definition restricts NAB to a model construct, estimated in simulations in which North American anthropogenic emissions are set to zero. The desire to quantify the impact of Canadian and Mexican emissions on NAB O$_3$ has led to the term “U.S. background”, a parallel model construct but estimated by setting only U.S. anthropogenic emissions to zero.

The development of effective State Implementation Plans (SIPs), by which states demonstrate how non-attainment regions will reach compliance with the NAAQS, requires an accurate assessment of the role of local, regional, and background sources in contributing to individual high-O$_3$ events. The Clean Air Act includes a provision for ‘exceptional events’, whereby high-O$_3$ events due to natural causes (such as wildfires or stratospheric intrusions) or foreign influence (e.g., Asian pollution) can be exempted from counting towards non-attainment status (Federal Register, 2007). Modeling the individual components of NAB can provide information to aid in attributing such events to specific sources.

In the previous review cycle (EPA, 2006) of the O$_3$ NAAQS, the Air Quality Criteria Document considered NAB estimates from the GEOS-Chem model for a single year (Fiore et al., 2003), the only estimates documented in the published literature at that time. Recent work has updated those estimates (Wang et al., 2009; Zhang et al., 2011), compared them with NAB in regional models using GEOS-Chem boundary conditions (Emery et al., 2012; Mueller and Mallard, 2011) and considered additional years. The first NAB estimates with an independent global model, (GFDL AM3; hereafter AM3; Table 2) were found to episodically reach 60–75 ppb over the Western United States in spring (Lin et al., 2012a). By contrast, GEOS-Chem estimated a maximum NAB of 65 ppb (Zhang et al., 2011) and the AM3 NAB was typically ~10 ppb higher than GEOS-Chem NAB on days when observations exceeded 70 ppb (Lin et al., 2012a, 2012b). Lin et al. (2012a, 2012b) are discussed in the Integrated Science Assessment (ISA) for the current O$_3$ NAAQS review cycle (EPA, 2013) but they focused on a different simulation year (2010) from the other studies. The ISA supplemental material does include a comparison of GEOS-Chem and GFDL AM3 for the same year (2006) at a measurement site in Gothic, CO, U.S.A. Here we extend that initial analysis by examining, in a consistent and process-oriented manner, the AM3 and GEOS-Chem NAB estimates during March through August of 2006. We additionally draw on a multi-decadal AM3 simulation to provide context for the single year inter-comparison. We include an evaluation of total surface O$_3$ in the base simulations with ground-based and space-based observations during 2006 to identify conclusions that are robust to the specific modeling system, as well as situations where observation-based constraints can be most effective in reducing uncertainty.

2. Review of prior model estimates for NAB and its components

We focus here on model estimates for NAB using the U.S. EPA definition, which relies on simulations with North American anthropogenic emissions set to zero. Earlier reviews synthesize observations relevant for evaluating base model simulations at remote sites (McDonald-Buller et al., 2011; Reid et al., 2008; Vingarzan, 2004). Even with the same approach, model estimates will differ due to different representations of natural emissions and the choice of different years since meteorological variability alters the balance between transported versus regionally produced O$_3$. In Table 1, we summarize published model estimates of various statistics for NAB, along with estimates from individual NAB sources (wildfires, lightning, the stratosphere, global anthropogenic CH$_4$ plus international anthropogenic emissions, and the sum of all natural sources).

Despite quantitative differences, a basic consensus emerges that the highest NAB levels generally occur during springtime and over western U.S. (WUS) high-altitude regions, with lowest NAB levels...
over EUS low-altitude regions in summer. The summertime minimum reflects the peak in regional photochemistry, which leads to accumulation of O₃ generated from regional precursors at the same time as it shortens the lifetime of O₃ mixing downward into the photochemically active boundary layer (see e.g., Fiore et al., 2002). At high-altitude WUS sites, models consistently indicate a day-to-day correlation between NAB levels and total O₃ during spring (Emery et al., 2012; Fiore et al., 2003; Lin et al., 2012a, 2012b; Zhang et al., 2011), implying that enhanced NAB levels play a role in raising total O₃, including above the level of the NAAQS. While these results are qualitatively consistent across several modeling platforms, the models vary in their quantitative attributions for NAB and its specific sources.

A few studies report the annual fourth highest MDA8 NAB value, the metric used to assess compliance with the O₃ NAAQS. Consideration of different metrics and different years complicates using the ranges across different modeling systems in Table 1 as error estimates. For example, mean values of NAB are unlikely to be static from year to year due to trends and variability in global anthropogenic emissions of O₃ precursors, in natural sources of NAB, and in the dominant regional transport patterns (Lin et al., 2014). A multi-model parameterization indicates an increase of ~4 ppb between 1960 and 2000 due to rising global CH₄ plus international anthropogenic emissions of non-CH₄ O₃ precursors (Wild et al., 2012). More recent increases in Asian emissions may have additionally raised WUS NAB by up to 3 ppb in spring between 2001 and 2006 (Zhang et al., 2008). The Asian, European, and global anthropogenic CH₄ components of NAB have received particular attention under the UNECE Task Force on Hemispheric Transport of Air Pollution (Fiore et al., 2009; Reidmiller et al., 2009; TFHTAP, 2010; Wild et al., 2012). Recent studies have documented the mechanisms by which Asian pollution can reach surface air over the WUS (e.g., Brown-Steiner and Hess, 2011; Lin et al., 2012b).

Wang et al. (2009) additionally estimated summertime U.S. Background (USB), which includes the in

<table>
<thead>
<tr>
<th>Study model</th>
<th>Study period; metric</th>
<th>NAB</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiore et al. (2003); GC (2° × 2.5°)</td>
<td>Mar−Oct 2001; 1−5 pm mean</td>
<td>Typically 15−35; up to 40−50 (highest in spring and WUS)</td>
<td>Natural: 18−23 (NW), 18−27 (SW), 13−20 (NE), 15−21 (SE)</td>
</tr>
<tr>
<td>Wu et al. (2008); GC with winds from GISS GCM (4° × 5°)</td>
<td>2000 Climatology; 1−5 pm mean</td>
<td>12−30 (summer); 22−40 (April); highest in WUS</td>
<td>Natural: 10−15 (EUS, summer); 15−25 (WUS, summer)</td>
</tr>
<tr>
<td>Wang et al. (2009); GC (1° × 1°)</td>
<td>Summer 2001; MDA8</td>
<td>26 ± 8</td>
<td>Natural: 18 ± 6 (low-alt); 27 ± 6 (high-alt); 34 −45 (4th highest).</td>
</tr>
<tr>
<td>Zhang et al. (2011); GC (5° × 2°)</td>
<td>Mar−Aug 2006−2008; MDA8</td>
<td>39−44 (spring); 35−45 (summer); low-alt 27 ± 8; high-alt 40 ± 7; 51−59 (4th highest)</td>
<td>Natural: 18 ± 6 (low-alt); 27 ± 6 (high-alt); 34−45 (4th highest).</td>
</tr>
<tr>
<td>Emery et al. (2012); CAMx (12 km²), GC boundary conditions</td>
<td>Mar−Aug 2006; MDA8</td>
<td>25−50 ppb (20−45 in GC); 35−100 (4th highest; 65 max without fires; 55 max in GC)</td>
<td>Fires: 10−50 ppb (events)</td>
</tr>
<tr>
<td>Lin et al. (2012a); GFDL AM3 (-50 km²)</td>
<td>Apr−Jun 2010; MDA8</td>
<td>15 WUS high-alt sites: 50 ± 11 (mean); 55 ± 11 (days when obs exceed 60 ppb)</td>
<td>Strat: 15 WUS high-alt sites: 22 ± 12 (mean); 15 −25 for obs O₃ @ 60−70; 17−40 for obs O₃ @ 70−85 Median, bias-corrected: 10−22 (W); 8−13 (NE); 3−8 (SE Max, bias-corrected: 35−55 (W); 30−45 (EUS)</td>
</tr>
<tr>
<td>McKeen et al. (2002); 3D regional model (60 km²)</td>
<td>Jun−Jul 1995; 1−4 pm mean</td>
<td>Fires: 10−30 ppb (event, Central and EUS)</td>
<td></td>
</tr>
<tr>
<td>Collins et al. (2003); STOCHEM driven by UM HadAM4 GCM</td>
<td>March 1991−1994 monthly mean</td>
<td>5−15 ppb (highest in WUS)</td>
<td></td>
</tr>
<tr>
<td>Kaynak et al. (2008); CMAQ (36 km²)</td>
<td>Jul−Aug 2004; MDA8</td>
<td>Lightning: up to 10 ppb; 14 ppb 4th highest; &lt;2 ppb 71% of the time</td>
<td></td>
</tr>
<tr>
<td>Mueller and Mallard (2011); CMAQ, GC boundary conditions (36 km²)</td>
<td>2002; MDA8</td>
<td>Fires: 30−50 (WUS, events)</td>
<td></td>
</tr>
<tr>
<td>Zhang et al. (2014); GC (5° × 2°)</td>
<td>Mar−Aug 2006; MDA8</td>
<td>Lightning: 6−10 ppbv (summer); Fires: &lt;20 (local events); 1−3 (WUS summer mean); Strat: 8−10 (WUS spring mean), up to 15 (events)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Model estimates for North American Background (NAB) ozone using the current U.S. EPA definition (North American anthropogenic emissions set to zero) and for specific components of NAB (ppb).
Table 2
GFDD AM3 and GEOS-Chem model configurations.

<table>
<thead>
<tr>
<th>Model</th>
<th>GFDD AM3</th>
<th>GEOS-Chem <a href="http://acmg.seas.harvard.edu/geos/">http://acmg.seas.harvard.edu/geos/</a></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Donner et al., 2011)</td>
<td>(Zhang et al., 2011)</td>
</tr>
<tr>
<td></td>
<td>(Rasmussen et al., 2012)</td>
<td>(Bey et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>(Naik et al., 2013)</td>
<td>(Park et al., 2004)</td>
</tr>
<tr>
<td></td>
<td>(Lin et al., 2012b)</td>
<td></td>
</tr>
<tr>
<td>Grid</td>
<td>Cubed sphere with 48 x 48 cell faces, approximately 2 x 2 horizontal resolution. Vertical coordinate is a 48-level hybrid sigma grid, with the top level at 0.01 hPa; lowest 5 layers extend to 60, 130, 220, 330, and 470 m for surface pressure of 1013.25 hPa and scale height of 7.5 km.</td>
<td>Continental North American nested (Wang et al., 2004) simulation at 0° latitude by 2° longitude using boundary conditions from boundary conditions from a 2 x 2.5° global simulation. Vertical grid has 47 levels to 0.01 hPa with lowest 5 layers centered at 70, 200, 370, 470, 600 m for a column at sea level.</td>
</tr>
<tr>
<td>Meteorology</td>
<td>Online, nudged to NCEP u and v (Kalny et al., 1996). The nudging timescale is inversely proportional to pressure (Lin et al., 2012b)</td>
<td>Assimilated from NASA GEOS-5</td>
</tr>
<tr>
<td>Isoprene nitrate yield and fate</td>
<td>Observationally-constrained 8% yield with 40% NOx recycling (Horowitz et al., 2007 and references therein)</td>
<td>18% yield with no NOx recycling (permanent sink for NOx)</td>
</tr>
<tr>
<td>Lightning NOx distribution</td>
<td>Parameterized based on convective cloud top height (Price and Rind, 1992), and described in Horowitz et al. (2003); global source in 2006 is 4.9 Tg N a⁻¹, range over 1981–2007 is 4.4–4.9 Tg N a⁻¹</td>
<td>Scaled to match a top-down global constraint of 6.5 Tg N a⁻¹ (Martin et al., 2007) and spatially redistributed based on the LEI/OTD flash climatology (Murray et al., 2012) and includes a yield of 500 mol N flash⁻¹ at northern mid-latitudes and 125 mol N flash⁻¹ elsewhere (Hudman et al., 2007)</td>
</tr>
<tr>
<td>Anthropogenic emissions</td>
<td>ACCMIP (Lamarque et al., 2010) with annual interpolation after 2000 to RCP4.5 2010 value (Lamarque et al., 2011). NOx emissions are 27 (global), 5.6 (North America), 5.9 (East Asia) Tg N a⁻¹. CO are 640 (global), 100 (North America), 160 (East Asia) Tg CO a⁻¹. CH₄ are 2.9 (global), 0.3 (North America), 0.5 (East Asia) Tg C a⁻¹.</td>
<td>EDGAR (Olier et al., and Berdowski, 2001) with U.S. emissions from 2005 National Emissions Inventory (NEI-05). NOx emissions are 27.9 (global), 6.4 (North America), 7.9 (East Asia) Tg N a⁻¹. CO are 582 (global), 91 (North America), 192 (East Asia) Tg CO a⁻¹, CH₄ are 27 (global), 4.55 (North America), 4.1 (East Asia) Tg C a⁻¹.</td>
</tr>
<tr>
<td>Biogenic emissions</td>
<td>Model of Emissions of Aerosols and Gases from Nature (MEGAN) 2.1 (Guenther et al., 2006), implemented as described by Emmons et al. (2010) and Rasmussen et al. (2012). Emissions are 385 (global) and 32 (North America) Tg C a⁻¹.</td>
<td>MEGAN 2.0 (Guenther et al., 2006)</td>
</tr>
<tr>
<td>Biomass burning emissions</td>
<td>As for anthropogenic emissions but distributed vertically as recommended by Dentener et al. (2006). NOx emissions are 5.3 (global), 0.18 (North America), 0.14 (East Asia) Tg N a⁻¹. CO are 440 (global), 18 (North America), 14 (East Asia) Tg CO a⁻¹. CH₄ are 1.5 (global), 0.09 (North America), 0.09 (East Asia) Tg C a⁻¹.</td>
<td>Emissions are 404 (global) and 30 (North America) Tg C a⁻¹.</td>
</tr>
</tbody>
</table>

| a | North American domain is 15–55° N and 60–125° W. |
| b | East Asian domain is 15–50° N, 95–160° E. |
| c | CH₄ is provided as an example non-CH₄ volatile organic compound (NMVOC) |

sparse observational network. Models are useful for quantifying the frequency of these events and for determining the contribution of these events to seasonal mean O3 levels. For decades, quantifying the stratospheric contribution to the troposphere, and particularly to surface air, has been contentious, with controversy rooted in the imprecise methods for quantifying accurately this component, as summarized in Lin et al. (2012a) (see their Section 2.3). Lin et al. (2012a) demonstrate that stratospheric intrusions play an important role in driving variability, including high-O3 events, at high-altitude WUS sites during spring. High-altitude greatly increases susceptibility to stratospheric influence; for days when observed O3 exceeds 70 ppb at monitoring sites in the Intermountain West during April–June of 2010, Lin et al. (2012a) find that median values of stratospheric O3 in the AM3 model are 10 ppb lower at the lower elevation sites than at high-elevation sites. Episodic wildfires have also been shown to contribute to high-O3 events (e.g., Jaffe and Wigder, 2012; McKeen et al., 2002; Mueller and Mallard, 2011), though Singh et al. (2010) found little O3 production in wildfire plumes over California unless mixing with an urban plume occurred.

3. North American background estimates from two independent global models

We compare background estimates for March through August of 2006 from two independent global models: the GEOS-Chem global chemistry-transport model (CTM) and the GFDD AM3 chemistry-climate model nudged to re-analysis winds. The models include different representations of the processes contributing to the abundance and distributions of tropospheric O3 (Table 2). We evaluate the base O3 simulations with hourly measurements from a ground-based network of monitoring sites and with monthly averaged retrievals from satellite instruments that are sensitive to O3 in the middle troposphere. We compare the models for March through August of 2006, the period analyzed previously by Zhang et al. (2011), drawing on a 27-year AM3 simulation to place the 2006 NAB estimates in the context of inter-annual variability. We note that the inter-annual variability may be underestimated in AM3 over some regions due its use of climatological inventories for soil NOx and wildfire emissions.

3.1. Model simulations, observations and analysis methods

Table 2 describes the GEOS-Chem and AM3 model configurations for the base simulations (meteorological year 2006). The GEOS-Chem CTM has been applied in various configurations over the past decade to estimate NAB and its various components for the summer of 1995 (Fiore et al., 2002), the 2001 O3 season (Fiore et al., 2003; Wang et al., 2009), and the 2006–2008 O3 seasons (Zhang et al., 2011, 2014) including evaluation with in situ and satellite observations. The AM3 model has previously been applied at ~50 km horizontal resolution globally to estimate the impacts of Asian pollution and stratospheric intrusions on surface O3 over the WUS from March through June of 2010; evaluation with in situ and space-based observations for that period shows it represents the subsidence of Asian and stratospheric O3 plumes over the WUS (Lin et al., 2012a, 2012b). The AM3 simulation used here is ~200 km horizontal resolution and is multi-decadal (1980–2007; first year is discarded as initialization), enabling us to place the year 2006 in the context of inter-annual variability (Table 2 and Section 4). Both
models estimate NAB in U.S. surface air by setting North American anthropogenic emissions of aerosol and O₃ precursors to zero. Anthropogenic sources include fossil and biofuel combustion (including aircraft and ship emissions within the domain), agricultural waste burning, and fertilizer application.

For anthropogenic emissions inventories, GEOS-Chem uses the 2005 National Emissions Inventory for the U.S., while AM3 uses the historical Atmospheric Chemistry-Climate Model Intercomparison Project (ACCMIP) emissions developed in support of the IPCC AR5 (Lamarque et al., 2011, 2010). Global, North American, and East Asian annual emissions for 2006 are provided in Table 2. Differences in the North American anthropogenic emissions inventories (5.58 and 6.67 Tg N a⁻¹ in AM3 and GEOS-Chem, respectively; 4.85 and 5.32 Tg N a⁻¹ for the United States), while crucial to the standard simulation for comparison with observations, should be irrelevant for the NAB simulations. Shortcomings in model representation of anthropogenic emissions or isoprene chemistry do not necessarily imply shortcomings in the NAB simulations, which may still be useful for quantifying daily to inter-annual variability driven by the transported components of NAB, such as O₃ associated with stratospheric intrusions, production from lightning NOₓ and CH₄.

The ground-based U.S. EPA Clean Air Status and Trends Network (CASTNet) sites were situated in order to provide regionally representative measurements and to minimize the influence of polluted urban air (Baumgardner et al., 2002) and thus are useful for evaluating O₃ simulated by coarse grid models. Our evaluation focuses on MDA8 O₃ concentrations, the statistic currently used by the U.S. EPA to assess compliance with the O₃ NAAQS, at 77 CASTNet sites. Simulated MDA8 O₃ is calculated from archived hourly average O₃ concentrations in the model surface layer. The models use a terrain-following sigma-coordinate for near-surface layers, with the lowest layer centered at approximately 60 m and 70 m for a column where the lowest layer is at sea level in GFDL AM3 and GEOS-Chem, respectively. All statistics are calculated by sampling the models at the locations of CASTNet sites with bilinear interpolation from the four nearest model grid cells to the latitude and longitude at each station.

Columns retrieved from satellite instruments are sensitive to free tropospheric O₃ and enable an evaluation on a continuous spatial scale of the simulated background available to subside into surface air. We use here direct tropospheric O₃ retrievals from both the Ozone Monitoring Instrument (OMI) (Liu et al., 2010) and the Tropospheric Emission Spectrometer (TES) (Beer, 2006). All data are processed using a single fixed a priori profile as described in Zhang et al. (2010). Previous validation of these retrievals against in situ and aircraft measurements indicate an accuracy to within 5 ppb at 500 hPa (Zhang et al., 2010 and references therein). Annual biases averaged over 2005–2007 relative to northern mid-latitude O₃ sondes (Zhang et al., 2010) have been uniformly subtracted from the retrieved products prior to evaluating the AM3 and GEOS-Chem base simulations. As evident from Zhang et al. (2010) (see their Fig. 5), the bias is not uniform and thus the real model error may deviate at any particular location from the true O₃ abundance differently than implied by the comparison with the satellite products reported here. We apply the appropriate satellite averaging kernels to the model daily O₃ fields for direct comparison with the retrieved satellite O₃ columns (Zhang et al., 2010). While the averaging kernels for the 500 hPa retrieved product for both the TES and OMI instruments are most sensitive to the mid-troposphere, there is a broad vertical sensitivity throughout the troposphere, but generally very little information is retrieved from the boundary layer (see example averaging kernels in Fig. 1 of Zhang et al. (2010)).

3.2. Regional and seasonal NAB estimates

Seasonal mean MDA8 NAB O₃ is consistently higher over the WUS than the EUS in both models (Fig. 1). During spring, AM3 simulates higher NAB over the high-altitude WUS than GEOS-Chem, which we attribute at least partially to a larger stratospheric influence in AM3 (Lin et al., 2012a) than in GEOS-Chem (Zhang et al., 2011). The diagnostics necessary to determine whether AM3 actually simulates more stratosphere-to-troposphere exchange of O₃, or whether it mixes free tropospheric air (including the stratospheric component) into the planetary boundary layer more efficiently, were not archived from these simulations. However, an evaluation of AM3 O₃ profiles at sonde launch locations during the 2010 CalNex field campaign indicates that the model captures much of the observed vertical structure of O₃ throughout
the troposphere and lower stratosphere, including its day-to-day variability (Lin et al., 2012a, 2012b). During summer, the different simulated spatial patterns for NAB over the WUS are influenced by differences in the lightning NO\textsubscript{x} sources (see Section 3.5).

Fig. 2 shows the spatial patterns of the fourth highest NAB value between March 1 and August 31. As the O\textsubscript{3} seasonal cycle is typically highest during the summer in polluted regions, we expect the fourth highest during this six-month period to represent reasonably this statistic over a full year. AM3 simulates the highest values over Colorado whereas GEOS-Chem indicates that the highest values occur over New Mexico (Fig. 2), reflecting excessive NAB associated with lightning NO\textsubscript{x} and subsequent O\textsubscript{3} production and transport (Zhang et al., 2014). Due to different timing of these processes, AM3 simulates the fourth highest NAB values during spring over much of Colorado but GEOS-Chem simulates peak values over much of New Mexico during August (Fig. 2). Over Minnesota and Wisconsin, GEOS-Chem generally produces the fourth highest values in spring, but AM3 indicates that they occur in summer. Over the northeastern states and west coast, the fourth highest MDA8 concentrations generally occur during spring, though they occur later in the year over southeastern states, with occurrences generally later in GEOS-Chem than AM3.

The fourth highest values often occur during months when model biases are largest (Section 3.4), indicating that bias-correction techniques may be necessary for quantitatively accurate NAB estimates at specific locations and times. In the following sections, we analyze the model NAB estimates in the context of evaluating the total surface O\textsubscript{3} simulations with both space- and ground-based observations, a first step towards developing the process-level knowledge needed for accurate bias correction.

### 3.3. Constraints from space-based observations

With the exception of O\textsubscript{3} produced within the U.S. boundary layer from CH\textsubscript{4} or natural NMVOC and natural NO\textsubscript{x} NAB in surface air mixes downward from the free troposphere. We use 500 hPa products retrieved from both the OMI and TES instruments aboard the NASA Aura satellite to evaluate the potential for space-based constraints on simulated mid-tropospheric total O\textsubscript{3} distributions. Our comparison thus evaluates the reservoir of mid-tropospheric O\textsubscript{3} of any origin, that can mix into the planetary boundary layer. For context, the ratio of NAB to total O\textsubscript{3} over North America at 500 hPa in AM3 varies spatially and seasonally, with highest values generally in spring and a minimum contribution within any model grid cell of 70% (not shown).

During spring, AM3 estimates a stronger north-to-south O\textsubscript{3} gradient in the mid-troposphere than GEOS-Chem (compare first and third rows of Fig. 3). The satellite retrievals from both instruments suggest a stronger gradient than simulated with GEOS-Chem, which generally underestimates O\textsubscript{3} in the northern half of the United States compared both to TES (5–15 ppb) and OMI (up to 10 ppb). In contrast, AM3 mid-tropospheric O\textsubscript{3} is higher than the satellite products in the northern half of the domain, with a closer match to the OMI retrievals (generally within 5 ppb over the United States) than TES (positive biases up to 10–20 ppb). Prior direct evaluation of AM3 with O\textsubscript{3} sondes indicates biases of up to 10 ppb in AM3 at the high northern latitude measurement sites of Alert and Resolute at 500 and 800 hPa with little bias in spring at the mid-latitude North American sites of Edmonton, Trinidad Head, Boulder and Wallops Island (Naik et al., 2013), roughly consistent with the biases relative to OMI.

Both satellite instruments indicate a general decrease from spring into summer over the western and northern United States, but an increase over several southeastern states, northern Mexico, and the Gulf of Mexico (compare Figs. 3 and 4). The summertime spatial pattern of U.S. O\textsubscript{3} observed from space is broadly consistent with that estimated by interpolating upper tropospheric ozone-sonde measurements during August of 2006 (Cooper et al., 2007). While the increases from spring to summer in the mid-troposphere over the EUS may include a contribution from lofting of regional anthropogenic O\textsubscript{3} production, there is likely also a contribution from the larger lightning NO\textsubscript{x} source in the free troposphere during summer. GEOS-Chem estimates a summertime mid-tropospheric O\textsubscript{3} enhancement at mid-latitudes, centered over the United States whereas AM3 simulates a gradient with O\textsubscript{3} generally increasing along the southwest-to-northeast direction (Fig. 4). Over Canada,
the AM3 model tends to be higher than both retrievals by up to 15–20 ppb during summer, as occurs during spring versus TES (but springtime biases are smaller compared to OMI).

We expect discrepancies between AM3 and observations during summer over forested boreal regions due to the use of a climatological wildfire inventory and the vertical distribution used to prescribe those emissions (Dentener et al., 2006), which lofts fire effluents into the mid-troposphere where they can efficiently produce O₃ and PAN (see also Section 3.5). GEOS-Chem includes fire emissions representative of the year 2006 and emits them only the planetary boundary layer. These model differences in wildfire treatment are likely reflected in the mid-tropospheric O₃ distributions shown in Figs. 3 and 4. They may also contribute to the different spatial distributions of simulated NAB at the surface, specifically the higher NAB estimated with AM3 over the northern United States and Canada relative to the NAB estimated with GEOS-Chem (Figs. 1 and 2).

In both Figs. 3 and 4, the models are generally more consistent with the OMI retrievals, which may reflect differences in the vertical sensitivity of the TES and OMI instruments. While the satellite retrievals provide useful qualitative constraints on the simulated mid-tropospheric distributions, the disagreement between OMI and TES over many locations (gray boxes in Figs. 3 and 4) hinders their quantitative utility. The higher sampling frequency possible from instruments on geostationary satellites such as TEMPO (Hilsenrath and Chance, 2013) should improve the potential for space-based constraints on free-tropospheric and near-surface distributions.

We can nevertheless glean additional insights into the model vertical distributions of O₃ by examining differences in the models sampled with the two different averaging kernels (Figs. 3 and 4). For example, over Canada, GEOS-Chem indicates that OMI would measure higher O₃ than TES whereas AM3 indicates that TES should retrieve higher O₃ than OMI during both seasons. In the spring, the retrieved OMI product is generally higher than TES over this region. GEOS-Chem is generally within 10 ppb of the OMI product with a tendency to underestimate springtime mid-tropospheric O₃ over Canada, whereas AM3 is generally within 5 ppb of OMI over much of the United States and Canada, with a tendency towards a positive bias. During summer, TES is higher than OMI over Canada. The high O₃ bias over the EUS in AM3 is confined close to the surface (Fig. 5) since AM3 tends to underestimate free tropospheric O₃, particularly over the convectively active Gulf of Mexico region where lightning NOₓ is expected to be an important source of NAB O₃. We conclude
that the estimates from the models could bracket the true NAB in many cases, but caution that the ability of the models to bracket the satellite measurements does not necessarily imply accuracy in their NAB estimates.

3.4. Constraints from ground-based measurements

We use the CASTNet MDA8 O3 observations to further constrain the model NAB estimates through an evaluation of the base simulations, which include all anthropogenic emissions, to simulate total surface O3. Since NAB depends strongly on altitude (Fig. 1; references in Table 1), the remainder of our analysis separates the data by altitude to gain insight into the different processes shaping NAB distributions. Specifically, we divide the CASTNet sites into two groups: (1) below 1.5 km in elevation (low-altitude sites), primarily sites in the EUS, and (2) Intermountain West CASTNet sites with elevation greater than 1.5 km (high-altitude sites). This second category includes all high-altitude CASTNet sites except for those in California.

3.4.1. Seasonal variability

Fig. 5 shows the observed and simulated seasonal cycles at the CASTNet sites. At the high-altitude sites, both models are generally within 5 ppb of the regional mean observed values and usually fall within one standard deviation of the observed monthly mean values at the sites within the region. Consistent with the evaluation in Section 3.3, the models tend to bracket the observations, but with notably different seasonal cycles. AM3 peaks in early spring, overestimating observed values in March but then declines to slightly underestimate observed values in June and July. In contrast, GEOS-Chem underestimated observed values from March through July but increases to overestimate observed values in August. The model differences are amplified in the NAB estimates: AM3 simulates a large seasonal decline in NAB from springtime (near 50 ppb) into summer (below 35 ppb) while GEOS-Chem estimates little seasonality in NAB (monthly mean values around 40 ppb).

At the low-altitude sites, AM3 exhibits a large positive bias in total surface O3 in all months, most evident during summer. The exacerbation of the bias in summer at low-altitude sites implies a problem with O3 produced from regional emissions, with isoprene–NOx–O3 chemistry a likely culprit given its different treatment in the models (Table 2; see Section 3.5). Both models show declining NAB levels from spring into summer, though the GEOS-Chem amplitude of the NAB seasonal cycle is smaller than that of AM3. The AM3 discrepancy with observations is much larger than the difference between the GEOS-Chem and AM3 NAB estimates except for March and April. If we assume that the model biases during March and April at both the high and low altitude sites are entirely due to NAB and correct the NAB estimates accordingly, the model NAB estimates would become more similar. While we conclude that the AM3 NAB at low-altitude sites is too high in March since we expect NAB to be lower than the observed value, it...
is possible that NAB could actually be higher in an atmosphere with lower NOx than under current conditions, due to more efficient O3 production and slower chemical loss.

At the high-altitude sites in summer, the GEOS-Chem overestimate of observed O3 has been attributed previously to an overestimate of O3 associated with production from lightning NOx plus subsequent transport when prescribing a higher production of NOx from flashes at mid-latitudes and spatially scaling the source to match LIS-OTD climatological flash counts (Murray et al., 2012), which may lead to regional errors for a specific year (Zhang et al., 2014). The larger difference between the NAB estimates from the two models in August than between the simulated and observed total O3 implies that the agreement with observations does not sufficiently constrain the NAB estimates.

### 3.4.2. Daily variability

Fig. 6 shows probability density distributions constructed from observed and simulated MDA8 O3 in spring (top) and summer (bottom) sampled at the high-altitude (left) versus low-altitude (right) CASTNet sites, and statistics are summarized for the AM3 and high-resolution GEOS-Chem simulations in Table 3. We additionally include in Fig. 6 estimates from a coarse resolution version of the GEOS-Chem model (green) in order to examine the extent to which differences in horizontal resolution contribute to the different NAB and total O3 estimates in AM3 versus GEOS-Chem. In all cases, the NAB (dotted lines) differ more between the GEOS-Chem and AM3 models than between the high- versus low-resolution versions of GEOS-Chem. This conclusion also holds for the total O3 distributions in spring. In summer, however, the total O3 distributions in GEOS-Chem are more sensitive to the choice of horizontal resolution, presumably reflecting the larger contributions from local-to-regional photochemical production during this season and the importance of spatially resolving domestic anthropogenic and natural emissions distributions. Emery et al. (2012) found that the higher resolution CAMx model generally simulated higher WUS NAB than a coarse resolution version of GEOS-Chem, and better agreement has been noted between CAMx and the higher resolution version of GEOS-Chem (EPA, 2013). Simulation of higher WUS NAB by higher resolution models (Emery et al., 2012; Lin et al., 2012a) likely reflects improved resolution of topography and mesoscale meteorology at higher resolution and the damping of vertical eddy transport at coarser resolution (Wang et al., 2004; Zhang et al., 2011).

AM3 simulates a wider NAB range than GEOS-Chem (Fig. 6 and Table 3). This wider range of NAB may contribute to the wider total surface O3 distribution in the AM3 versus GEOS-Chem standard simulations, which aligns more closely with the observed variability, except for O3 simulated with the high-resolution GEOS-Chem model in summer at high-altitude sites. The relative skill of AM3 in capturing the variability of NAB despite its generally high bias implies that AM3 is useful for process-level analysis and for quantifying day-to-day variability.

In Table 3, we partition statistics for total and NAB O3 in surface air into average versus high-O3 days. We use observed values, rather than simulated values used in Zhang et al. (2011), to select for high-O3 days in order to sample the same temporal subset from both models. Using the simulated total O3 values would lead to subsets of different sizes given the individual model biases. During spring, the models robustly estimate NAB to be ~10 ppb higher on average at high-altitude than at low-altitude CASTNet sites, but AM3 estimates higher NAB levels than GEOS-Chem. During summer, the models also estimate higher NAB at high-altitude than at low-altitude sites, and average NAB levels decrease from spring to summer at low-elevation sites. GEOS-Chem suggests little change from spring to summer in average high-altitude NAB whereas AM3 simulates a decrease of over 10 ppb. At the high-altitude sites, both models suggest that NAB increases as total O3 increases, although the sample size is small for events above 75 ppb and the average values for the different data subsets all fall within one standard deviation each other. At the low altitude sites, there is little change in the average NAB when selecting for observed values exceeding 60, 70, or 75 ppb. The variability in NAB, as measured by the standard deviation in Table 3, is similar in the two models at the low-elevation sites, but AM3 simulates more variability in NAB at the high-altitude sites than GEOS-Chem, particularly on high-O3 days.

We next evaluate the ability of the models to capture observed day-to-day variability by correlating observed and simulated MDA8 O3 in surface air at each CASTNet site, separately for the spring and summer seasons of 2006 (Fig. 7). During spring, the correlation coefficients between observations and total simulated surface O3 over WUS sites are generally higher in GEOS-Chem, but the level of correlation in summer is maintained or improved in AM3 whereas it decreases in GEOS-Chem (Fig. 7). Over the EUS, the models show similar spatial patterns in their ability to reproduce the observed day-to-day variability. GEOS-Chem is generally better than AM3 over northern sites, whereas AM3 captures more of the variability over the southeastern sites in summer (Fig. 7).
Fig. 6. Probability density curves calculated via kernel (Gaussian) density estimation with a bandwidth of 2 ppbv from surface MDA8 O₃ data over the CONtiguous United States (CONUS) during spring (top) and summer (bottom) and at high (left, excluding California sites) and low (right) elevation CASTNet sites: observed (black) and GFDL AM3 (red) and GEOS-Chem at low (green; 2° × 2.5') and high (blue; ½° × ½°) horizontal resolution (LR vs. HR) models sampled at the CASTNet sites for total (solid lines) and NAB (dashed lines) O₃. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 8 shows the correlation coefficients of the NAB estimates versus the simulated total O₃ separately for spring and summer, for each model sampled at the CASTNet sites that were used in Fig. 7 to evaluate the simulated daily variability. Over the Intermountain WUS sites, the models robustly indicate that variability in NAB drives a substantial portion of the total surface O₃ variability in both seasons, but with a stronger influence (higher correlations) during spring. AM3 generally indicates a stronger role for NAB in contributing to variability in total surface O₃ at the sites in Florida and Texas. Over the central states and over the inland mid-Atlantic region, the NAB is more strongly correlated with total surface O₃ in GEOS-Chem than in AM3; the stronger correlation in GEOS-Chem in these regions may arise from soil NOₓ emissions, which respond to meteorological variability in GEOS-Chem whereas AM3 uses a monthly climatology. At the EUS sites in Fig. 8, the NAB in both models is poorly correlated, and in some cases, anti-correlated with the total simulated surface O₃. The highest total surface O₃ events over the EUS are thus generally decoupled from the highest NAB events, consistent with the current understanding that regional pollution is the dominant influence on total O₃ distributions in this region.

We select four sites, encircled in black in Figs. 7 and 8, to probe more deeply the day-to-day variability in NAB and total surface O₃. The time series at these sites (Fig. 9a and b) provide evidence at the local scale for our assessment of regional and seasonal biases. At the two western U.S. sites (Gothic, CO and Grand Canyon NP, AZ) in Fig. 9a, the 6-month average NAB is nearly the same in both models, but this reflects little seasonal variation in the GC NAB (thin blue line) versus a sharp seasonal decline from spring into summer in AM3 (thin red line). The standard deviation is twice as large in AM3 as in GEOS-Chem, consistent with the frequency distributions of NAB in Fig. 6 (left side); AM3 also captures the observed variability. Despite the summertime high bias in AM3 at the two EUS sites (M.K. Goddard, PA and Georgia Station, GA), AM3 correlates at least as well with the observations as GEOS-Chem (Figs. 7 and 9).

### Table 3

Summary statistics of seasonal mean MDA8 total and NAB O₃ in surface air (ppb) as observed and estimated with the GFDL AM3 and GEOS-Chem (GC) models, segregated by altitude, season, and observed values. California is excluded to focus on the Intermountain West.

<table>
<thead>
<tr>
<th>Season Filter</th>
<th>N OBS</th>
<th>AM3 base</th>
<th>GC base</th>
<th>AM3 NAB</th>
<th>GC NAB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above 1.5 km (excluding CA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAM None</td>
<td>993</td>
<td>57 ± 7</td>
<td>60 ± 7</td>
<td>54 ± 6</td>
<td>48 ± 8</td>
</tr>
<tr>
<td>MAM Obs ≥ 60</td>
<td>300</td>
<td>64 ± 4</td>
<td>63 ± 7</td>
<td>58 ± 6</td>
<td>52 ± 8</td>
</tr>
<tr>
<td>MAM Obs ≥ 70</td>
<td>33</td>
<td>73 ± 4</td>
<td>66 ± 6</td>
<td>62 ± 4</td>
<td>55 ± 7</td>
</tr>
<tr>
<td>MAM Obs ≥ 75</td>
<td>7</td>
<td>80 ± 4</td>
<td>65 ± 7</td>
<td>61 ± 2</td>
<td>56 ± 8</td>
</tr>
<tr>
<td>JJA None</td>
<td>899</td>
<td>58 ± 7</td>
<td>55 ± 6</td>
<td>57 ± 8</td>
<td>35 ± 8</td>
</tr>
<tr>
<td>JJA Obs ≥ 60</td>
<td>344</td>
<td>65 ± 4</td>
<td>58 ± 5</td>
<td>59 ± 7</td>
<td>38 ± 8</td>
</tr>
<tr>
<td>JJA Obs ≥ 70</td>
<td>38</td>
<td>73 ± 5</td>
<td>61 ± 4</td>
<td>62 ± 7</td>
<td>43 ± 8</td>
</tr>
<tr>
<td>JJA Obs ≥ 75</td>
<td>9</td>
<td>80 ± 6</td>
<td>64 ± 4</td>
<td>64 ± 6</td>
<td>47 ± 6</td>
</tr>
<tr>
<td>Below 1.5 km</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAM None</td>
<td>5769</td>
<td>49 ± 11</td>
<td>57 ± 8</td>
<td>48 ± 8</td>
<td>39 ± 8</td>
</tr>
<tr>
<td>MAM Obs ≥ 60</td>
<td>969</td>
<td>65 ± 6</td>
<td>64 ± 8</td>
<td>57 ± 8</td>
<td>37 ± 8</td>
</tr>
<tr>
<td>MAM Obs ≥ 70</td>
<td>175</td>
<td>75 ± 6</td>
<td>69 ± 8</td>
<td>63 ± 10</td>
<td>36 ± 10</td>
</tr>
<tr>
<td>MAM Obs ≥ 75</td>
<td>58</td>
<td>82 ± 6</td>
<td>71 ± 10</td>
<td>68 ± 12</td>
<td>36 ± 11</td>
</tr>
<tr>
<td>JJA None</td>
<td>5583</td>
<td>51 ± 15</td>
<td>69 ± 15</td>
<td>54 ± 14</td>
<td>29 ± 9</td>
</tr>
<tr>
<td>JJA Obs ≥ 60</td>
<td>1509</td>
<td>69 ± 9</td>
<td>76 ± 13</td>
<td>63 ± 11</td>
<td>30 ± 9</td>
</tr>
<tr>
<td>JJA Obs ≥ 70</td>
<td>537</td>
<td>78 ± 8</td>
<td>77 ± 13</td>
<td>67 ± 12</td>
<td>30 ± 9</td>
</tr>
<tr>
<td>JJA Obs ≥ 75</td>
<td>294</td>
<td>83 ± 8</td>
<td>76 ± 14</td>
<td>69 ± 14</td>
<td>31 ± 9</td>
</tr>
</tbody>
</table>

3.5. Processes contributing to inter-model differences in total and NAB surface O₃

We use the sites in Fig. 9a to examine the role of specific processes in contributing to differences in the GEOS-Chem and AM3...
Base and NAB simulations. Superimposed in Fig. 9a and b are results from a separate simulation (Lin et al., 2014) in which a stratospheric O3 tracer (O3Se90) was available (orange line), tagged relative to the e90 tropopause (Prather et al., 2011) as described in Lin et al. (2012a). The high summertime correlation of O3Se90 and NAB at the WUS sites (Fig. 9a) does not imply that stratospheric O3 intrusions are the dominant factor. The magnitude of the O3Se90 enhancements in summer do not account for all of the episodic NAB enhancements. Rather, this result implies that other sources enhance NAB free tropospheric O3, which then mixes into the surface air alongside the O3Se90 tracer in the model. We interpret the high correlations in both seasons at the EUS sites ($r^2$ ranges from 0.72 to 0.86) in a similar manner: the O3Se90 indicates a larger influence of O3 mixing down from the free troposphere.

### 3.5.1 Deep stratospheric intrusions over the WUS in spring

As described by Lin et al. (2012a), stratospheric O3 drives a substantial portion of the daily variability in observed springtime...
O₃ over the WUS. Inspection of Fig. 9a (top two panels) shows that the episodic enhancements in the O₃Se₉₀ tracer can explain much of the episodic enhancements in NAB. A caveat is that the magnitude of the stratospheric contribution is an upper limit due to the definition of the O₃Se₉₀ tracer, which could be tagging O₃ in the lower stratosphere that originated in the troposphere (see discussions in Lin et al. (2012b) and Zhang et al. (2014)). Nevertheless, the strong correlation of the orange and red lines in Fig. 9a implies a key role for mixing of free tropospheric air into the planetary boundary layer in driving day-to-day variability in NAB O₃ levels.

For illustrative purposes, we focus on an event during late May of 2006 at the Gothic and Grand Canyon sites, during which the AM3 model NAB (thin red line in Fig. 9a and b) spikes, with an associated increase in the simulated total O₃ (thick red line). In contrast, GEOS-Chem NAB (thin blue line) decreases, as does total O₃ (thick blue line) during this event. The opposing trends in the models during this event raise questions as to whether both models simulate a mixing event but import different NAB levels, or whether the boundary layer in AM3 mixes more vigorously with the free troposphere than in GEOS-Chem. The observations (black) increase during this period, as in AM3, but the model overestimates the observed values on May 28 and 29. The O₃Se₉₀ tracer (orange line in Figs. 9a and b) suggests that the AM3 model is simulating surface O₃ enhancements associated with a stratospheric intrusion; the spatial pattern of these enhancements aligns with the observed spatial pattern of enhanced ground-level O₃ at the CASTNet sites (not shown). Fig. 9a further suggests that these events drive much of the variability in NAB at high-altitude western sites in spring, consistent with earlier findings for April through June of 2010 (Lin et al., 2012a).

3.5.2. Wildfires over the EUS in spring and summer

There are several EUS events during spring and summer where AM3 simulates a localized spike in NAB that is not simulated by GEOS-Chem, which we attribute at least partially to the different treatment of wildfire emissions in the models (Table 2). For example, there is an extreme “NAB event” in the AM3 model on June 28, 2006 at the Pennsylvania CASTNet site in Fig. 9a. AM3 estimates NAB values above 60 ppb, exceeding the total observed value of about 60 ppb, while GEOS-Chem simulates NAB below 20 ppb (Fig. 9a, yellow highlight). We find that the use of a year-specific fire inventory (as is done in the GEOS-Chem simulations) versus a climatology leads to differences of 10 ppb for the June 28, 2006 event in AM3 (not shown). In AM3, the recommendations from Dentener et al. (2006) are applied to vertically distribute...
biomass burning emissions north of 25°N, placing 40% of the total emissions between 3 and 6 km (see their Table 4) over boreal North America. This lofting of fire effluents likely contributes to the summertime O$_3$ overestimates at 500 hPa over Canada (Fig. 4). Vertical mixing of NAB O$_3$ from the free troposphere into surface air in the AM3 model is indicated by associated enhancements of the O3se90 tracer on days with high NAB. In contrast, biomass burning is emitted only in the boundary layer in GEOS-Chem, likely resulting in less efficient O$_3$ production and subsequent long-range transport. The GEOS-Chem approach appears more consistent with the observations.

3.5.3. Lightning NO$_x$ over the southwestern United States in summer

During August at the two WUS sites in Fig. 9a, the models reverse their relative rankings of simulated NAB relative to springtime, with the GEOS-Chem NAB much as 10–20 ppb higher than AM3 NAB. In notable contrast to the spring, GEOS-Chem overestimates the observed O$_3$ values. We attribute the summertime overestimate and poor correlations of GEOS-Chem with the observed values over WUS sites in Figs. 7 and 9a to the lightning NO$_x$ source and subsequent transport. GEOS-Chem produces approximately 10 times more lightning NO$_x$ than AM3 over the southwestern states during summer (0.018 Tg N in AM3 versus 0.159 Tg N in GC within the region 26°N–42°N, 124°W–97°W) and the models further differ in their spatial distributions of the lightning NO$_x$ source (Table 2). This source has been reduced in a newer version of GEOs-Chem, decreasing simulated NAB O$_3$ over these regions (Zhang et al., 2014).

3.5.4. Isoprene oxidation chemistry over the EUS in summer

Earlier work (e.g. Fiore et al., 2002, 2003) demonstrated that NAB is fundamentally different between the EUS and the WUS, with the EUS more strongly controlled by regional photochemistry, where the O$_3$ lifetime in the planetary boundary layer is as short as 1–2 days and isoprene–NO$_x$–O$_3$ chemistry dominates much of the region from May through September (Jacob et al., 1995). At the two EUS sites in Fig. 9a (M.K. Goddard, PA and Georgia Station, GA), we attribute some of the differences in the summertime simulations to the isoprene oxidation mechanism (Table 2) that would tend to reduce O$_3$ production in GEOS-Chem relative to AM3 due to isoprene ozonolysis serving as a more important loss pathway for NAB in GEOS-Chem (Fiore et al., 2002; Mickley et al., 2001). These differences in isoprene oxidation chemistry could at least partially explain the higher NAB in AM3 during the isoprene emission season (i.e., a longer O$_3$ lifetime in the AM3 boundary layer). The largest inter-model differences in NAB, however, occur in spring when transported sources are more important than regional production involving natural sources.

The isoprene oxidation chemistry likely also contributes to the large bias in AM3 total surface O$_3$. GEOS-Chem assumes a much higher yield of isoprene nitrates from the reaction of isoprene hydroxyperoxides radicals with NO and assumes that they are a permanent sink for NO$_x$ (Table 2). In contrast, AM3 assumes an 8% isoprene nitrate yield and allows 40% of the products to recycle back to NO$_x$ on the basis of observational constraints from field campaigns (Horowitz et al., 2007; Perring et al., 2009). Earlier work with predecessors of the models used here suggests that these differences may explain over 10 ppb of the high bias in AM3 relative to GEOS-Chem in May over the EUS in summer (Fiore et al., 2005). The fact that GEOS-Chem best captures the observations implies that the additional O$_3$ production from isoprene oxidation using the field-based constraints on isoprene nitrates must be offset by larger O$_3$ losses, such as may occur through additional HO$_x$ uptake by aerosol (Mao et al., 2013) and halogen-induced O$_3$ destruction (Parrella et al., 2012).

4. Inter-annual variability in NAB MDA8 O$_3$ estimates in surface air

The 27-year AM3 NAB simulation (1981–2007) enables us to define spring and summer climatologies of seasonal mean NAB O$_3$ in surface air, and to quantify the year-to-year variability as the standard deviation of the annual seasonal mean values (Fig. 10). The seasonal mean spatial patterns are similar to those in 2006 (Fig. 1), with little year-to-year variation over much of the country. Fig. 10 also includes the climatological fourth highest MDA8 value between March 1 and August 31 over the multi-decadal simulation. We emphasize that these estimates are subject to the biases diagnosed above in comparison to observations. In particular, NAB estimates over the EUS are probably too high in AM3. The variability over central Texas and central Mexico in the fourth highest values may indicate year-to-year variations in events involving NAB production from lightning NO$_x$ and convective mixing. Large variability in both mean NAB levels and the highest events is simulated over Western Colorado in spring, with standard deviations of 2–3 ppb, likely reflecting variability in year-to-year influence from stratospheric O$_3$ intrusions.

Jaffe (2011) noted regionally coherent year-to-year variability in the number of high-O$_3$ events at high-altitude western U.S. measurement sites in both spring and summer and we examine here the potential contribution of NAB to this observed variability. Specifically, Jaffe (2011; see their Fig. 6) found that the number of O$_3$ events above levels of 65, 70, and 75 ppb varied together, with the lowest and highest number of springtime events occurring in 1997 and 1999, respectively; for summer, the lowest and highest years were 1997 and 2002. We follow the approach of Wang et al. (2009; see their Fig. 5) to illustrate simultaneously the model skill at capturing the observed values, and the simulated NAB contribution to observed levels within specific ranges for total surface O$_3$. Fig. 11 shows the AM3 NAB contributions throughout the overall observed distributions for 2006 in comparison to a low versus high year for observed high-O$_3$ events at the 11 high-altitude WUS sites in Figs. 5 and 6. Note that the highest years in Fig. 11 differ for spring and summer, but the lowest year is 1997 in both seasons.

For observed O$_3$ events above 60 ppb, AM3 tends to overestimate observations during spring but does not exhibit any systematic bias during summer. Furthermore, the model captures events up to 80 ppb during spring of 1999, though in other years there is a general tendency to underestimate events above 75 ppb. This finding is in contrast to those from higher-resolution models including the GEOS-Chem version used here, which underestimates events above 60 ppb (Zhang et al., 2011). During all years and both seasons shown in Fig. 11, there is a trend for the median simulated NAB contribution to increase from observed values of 40 ppb to those in the 70 ppb range, with 75th percentile values reaching 50–60 ppb for observed values above 60 ppb during 2006 and 1999, implying that enhanced NAB levels contribute to the higher observed values. This interpretation is consistent with the findings of Lin et al. (2012a) that stratospheric O$_3$ intrusions over the high-altitude WUS drive much of the observed day-to-day variability in spring, as well as with Jaffe (2011) who suggests that a large-scale process drives coherent variability at the monitoring sites in this region.

Consistent with earlier work (Fiore et al., 2003), Fig. 11 shows that summertime NAB levels are typically much lower than in spring, with maximum values nearly always below 60 ppb and 75th percentile values generally below 50 ppb. Jaffe (2011) suggested...
Fig. 10. Climatological (1981–2007) average (left) and standard deviation (right) of spring (top) and summer (middle) seasonal mean MDA8 NAB O₃, and of the fourth highest value between March 1 and August 31 (bottom) as estimated with the GFDL AM3 model simulation in which North American anthropogenic emissions are set to zero.

Fig. 11. GFDL AM3 simulated daily maximum 8-h (MDA8) surface O₃ versus observed values (black) and AM3 NAB statistics (green) at 11 Intermountain Western U.S. CASTNet sites above 1.5 km altitude for a "low-O₃" year (left column) and "high-O₃" year (right column) to provide context for the year 2006 (middle column) during spring (top panel) and summer (bottom panel), following the approach of Wang et al. (2009; see their Fig. 5). The 1:1 line (solid black) and a 60 ppb threshold (dashed line) are shown. Box and whisker plots show the median (triangle), 25th–75th range (box) and minimum and maximum NAB values (vertical lines) for 10 ppb bins of observed O₃ values. The "low" and "high" years are selected from Fig. 6 of Jaffe (2011). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
that summertime inter-annual variability is strongly influenced by wildfire activity. The lack of year-to-year variations wildfires in this version of the AM3 model may contribute to its underestimate of the highest events in 2002 and 2006, which were the first and second highest fire activity years for the 1997–2006 period analyzed by Jaffe (2011).

5. Conclusions and recommendations

On the basis of health evidence, the level of the National Ambient Air Quality Standard (NAAQS) for ground-level O3 has been lowered in recent years, pushing closer to “background” levels. In the past, the U.S. Environmental Protection Agency considered model-based estimates of background O3 as part of the process for setting the NAAQS. These model-based estimates, previously called “Policy-Relevant Background”, are now termed “North American Background” (NAB), which is defined to be background levels that would exist in the absence of North American anthropogenic emissions. Identifying high-background events is crucial for determining whether an observation merits consideration for “exceptional event” status, which exempts a particular observation from counting towards non-attainment if it can be shown that the event occurred due to processes beyond the control of U.S. air quality management options. The model simulations presented here can provide information on the frequency of such events and the individual components contributing to NAB, including O3 originating from international pollution, wildfires, or the stratosphere.

As a first step towards assessing our understanding of NAB and its components, we briefly reviewed recent model estimates (Table 1). We then evaluated total surface O3 and NAB estimates from two independent models (GEOS-Chem and AM3) for March through August of 2006, using comparisons between the base simulations and space-based and ground-based measurements to place constraints on the model estimates. A 27-year NAB simulation in the AM3 model provides context for our two-model analysis and indicates that 2006 is a typical year in terms of its spatial and seasonal patterns in NAB, though 2006 NAB levels are generally higher than the climatological averages (compare Fig. 10 with 1 and 2). The largest variability in mean NAB MDA8 estimated with AM3 occurs over Idaho, western Colorado and Wyoming, and New Mexico, with standard deviations of over 2 ppb; the largest variability in the fourth highest MDA8 NAB occurs over Colorado and Texas (Fig. 10). A comparison of low- versus high-O3 years at high-altitude western U.S. (WUS) sites indicates a role for NAB in driving year-to-year differences in the frequency of springtime high-O3 events (Fig. 11).

At high-altitude WUS sites, the GEOS-Chem and AM3 models consistently show higher NAB than at low-altitude sites, but the magnitude and day-to-day variability often differs (Figs. 1, 5, 6 and 9a, Table 3). In some months (e.g., August), the larger differences between the NAB estimates from the two models than between the simulated and observed total O3 imply that agreement with observations does not sufficiently constrain the NAB estimates. While AM3 indicates a seasonal decline of NAB into summer over this region, GEOS-Chem suggests a relatively weak seasonal cycle associated with an increase of influence from lightning NOx in that model during the late summer (Figs. 5 and 9). Higher stratosphere–troposphere exchange in AM3 may explain the springtime NAB enhancement in the free troposphere relative to GEOS-Chem (Fig. 5), which, followed by more vigorous mixing between the free troposphere and boundary layer, may explain the higher NAB in surface air during this season in AM3 (Fig. 1).

At low-altitude sites, such as over the EUS, the models consistently show lower NAB levels than at high-altitude sites, as in earlier work (Table 1). We find that the highest total surface O3 events over the EUS are often decoupled from the highest NAB events (Fig. 8), consistent with the understanding that regional pollution is the dominant influence on total O3 distributions there. Over the EUS, uncertainties in isoprene–NOx–O3 chemistry (Table 2) likely contribute to differences in simulated total O3, and to a lesser extent, NAB estimates.

We find little evidence that horizontal resolution is a major contributor to differences in mean NAB estimates in the models (Fig. 6), consistent with EPA (2013). Higher resolution refines spatially local NAB estimates, including at the tails of the distribution. It also resolves better the impact from local and regional emissions, as evidenced by the larger differences associated with resolution in summertime distributions when photochemical production peaks over many U.S. regions (Fig. 6). We conclude that simulated NAB distributions reflect large-scale synoptic transport that is resolved sufficiently at the relatively coarse scale of global models, with the NAB differences mainly stemming from different treatments of NAB sources such as stratospheric O3, boreal fires, and lightning NOx. The regional and seasonal variability in these driving processes further manifests as differences in the model timings of the fourth highest NAB over many regions (Fig. 2).

Future efforts to determine the processes contributing to model differences, and to the biases in individual models versus observations, would benefit from evaluation with daily O3 vertical profiles as measured by sondes, consistently defined tracers of stratospheric influence (e.g., the O3Se90 tracer in AM3), as well as daily three-dimensional archival of other chemical species (e.g., CO, PAN, H2O) that can aid in disentangling tropospheric versus stratospheric origins and from meteorological variables (e.g., mixing depth, mass fluxes) to diagnose the role of mixing processes. The routine use of synthetic tracers could further aid in distinguishing between model differences in transport, dilution, and mixing versus chemical evolution during transport. Improved estimates of NAB in a given region and season will require better constraints on, for example: lightning NOx for central and southwestern states in summer; transported stratospheric O3 over the high-altitude WUS in spring; isoprene chemistry and its impact on chemical processing and NAB lifetime over the EUS in summer; and wildfires, which may influence NAB throughout the nation from late spring into summer.

We propose that future multi-model studies target limited time periods to enable process-oriented analysis during field campaigns when ground-based and satellite observations are supplemented with a broader suite of observations from intensive aircraft flights and balloon launches. If combined with a thorough evaluation of O3 precursors, such analysis should hasten progress towards understanding the impact of specific sources on NAB O3. We further recommend developing bias-correction techniques, such as those routinely applied in numerical weather prediction, to improve the accuracy of local NAB estimates. As a first step, simple assumptions assuming the bias is entirely driven by one process (e.g., as applied to the stratospheric O3 estimates from the AM3 model by Lin et al. (2012a)) can be applied to individual models and then used to generate a multi-model estimate with uncertainties. The two models analyzed here often bracket the observations (Figs. 3–6 and 9), thereby indicating different sources of error, which leads us to conclude that a multi-model approach can harness unique capabilities of different modeling systems and thus provide more accurate NAB estimates than a single model.

Acknowledgments

We acknowledge support from the NASA Air Quality Applied Sciences Team (NNX12AF15G to AMF; NNX11AH93G to DJJ), and
the NOAA Ernest F. Hollings Scholarship Program (JTO). We are grateful to P. Dolovich (U.S. EPA) for useful discussions and comments, and suggestions from anonymous review. The information in this article has been subjected to review by the National Center for Environmental Assessment, U.S. Environmental Protection Agency, and approved for publication. The views expressed in this manuscript are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

References


