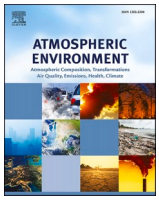
Atmospheric Environment 309 (2023) 119929

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv



Decoupling in the vertical shape of HCHO during a sea breeze event: The effect on trace gas satellite retrievals and column-to-surface translation

Amir H. Souri [a,b,*], Rajesh Kumar [c], Heesung Chong [d], Maryam Golbazi [c], K. Emma Knowland [b,e], Jeffrey Geddes [f], Matthew S. Johnson [g]

[a] Atmospheric Chemistry and Dynamics Laboratory, NASA Goddard Space Flight Center (GSFC), Greenbelt, MD, USA
[b] GESTAR II, Morgan State University, Baltimore, MD, USA
[c] Research Applications Laboratory, National Center for Atmospheric Research, Boulder, CO, 80305, USA
[d] Atomic and Molecular Physics (AMP) Division, Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA, USA
[e] Global Modeling and Assimilation Office, NASA GSFC, Greenbelt, MD, USA
[f] Earth and Environment Department, Boston University, Boston, MA, 02215, USA
[g] Earth Science Division, NASA Ames Research Center, Moffett Field, CA, 94035, USA

HIGHLIGHTS

- A sea breeze event is conducive to a decoupled relationship between HCHO vertical columns and the surface concentrations.
- A coarser scale atmospheric chemical transport model could not fully replicate the tendency.
- Large discrepancies in the vertical shapes of HCHO are problematic for applying trace gas remote-sensing data.

ARTICLE INFO

Keywords:
Sea breeze
Remote sensing
Formaldehyde
Geostationary
WRF-CMAQ

ABSTRACT

The effect of sea breeze circulation on stratification of the vertical formaldehyde (HCHO) concentration vertical profiles is explored using a regional atmospheric chemical transport model (CTM) for three synoptically stagnant days focused on the east coast of the U.S in June 2018. During this event, a significant thermal contrast between the Atlantic Ocean and the terrestrial regions (12–17 °C), observed by moderate resolution imaging spectroradiometer (MODIS) and well-captured by the WRF-CMAQ model (15–18 °C), is conducive to monsoon-like flow, perpendicular to the shorelines, carrying clean marine air masses over the land within a few hundreds of meters above the surface. In contrast, the westerly continental polluted air masses prevail in higher altitudes. These two conflicting flows result in atypical vertical shapes of HCHO concentrations increasing with altitude. This decoupling pattern is so pronounced that we observe total column HCHO negatively correlate with surface concentrations. Comparisons of an accredited global model, GEOS-CF, to surface wind measurements and MODIS skin temperature indicate its poor representation of the sea breeze timing and strength, resulting in GEOS-CF HCHO vertical shapes being drastically different from the WRF-CMAQ. Based on radiative transfer calculations, the differences in the vertical distribution of HCHO between the WRF-CMAQ and that of GEOS-CF in the first 3 km are sufficient to induce a 20–30% error in air mass factors (thus total vertical HCHO column abundances). Through an experiment involving converting HCHO total columns to surface mixing ratios, we demonstrate that GEOS-CF allocates noticeably more HCHO molecules (40–150%) to the surface layer due to the misrepresentation of the vertical shape of HCHO during the sea breeze event. It is known that a significant fraction of the human population lives in coastal areas prone to detrimental effects caused by air pollution, and elevated pollutant concentrations usually occur in synoptically stagnant atmospheric conditions when local circulation patterns come into play; accordingly, our experiments emphasize the importance of the effect a priori profiles can have on satellite-derived applications under such conditions. To ensure that the quantitative representation of satellite-based trace gas retrievals on a daily basis is trustworthy and useable for air quality applications, atmospheric models providing a priori profiles for satellite retrievals should be well-tuned to reproduce complex local circulation such as sea-land breezes.

* Corresponding author. Atmospheric Chemistry and Dynamics Laboratory, NASA Goddard Space Flight Center (GSFC), Greenbelt, MD, USA.
E-mail addresses: a.souri@nasa.gov, ahsouri@gmail.com (A.H. Souri).

https://doi.org/10.1016/j.atmosenv.2023.119929
Received 7 March 2023; Received in revised form 1 June 2023; Accepted 26 June 2023
Available online 28 June 2023
1352-2310/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Rapid developments in sensor design, calibration, and retrieval algorithms have enabled quantitatively measuring one of the most challenging trace gas absorbers in the UV range, formaldehyde (HCHO), from satellites (Chance et al., 1991; Chance et al., 2000; De Smedt et al., 2008, 2010, 2015, 2018, 2021; González Abad et al., 2015; Nowlan et al., 2023). This specific compound has important implications for understanding the spatiotemporal variability in volatile organic compound (VOC) emission (Palmer et al., 2006; Millet et al., 2006; Stavrakou et al., 2009; Marais et al., 2012; Bauwens et al., 2016; Souri et al., 2020a), hydroxyl radical (OH) budget (Wolfe et al., 2019; Anderson et al., 2022; Zhu et al., 2022), and tropospheric ozone non-linear chemistry (Duncan et al., 2010; Jin et al., 2017; Souri et al., 2020a, 2020b, 2023, Travis et al., 2022).

Satellite trace gas retrievals are a multifaceted problem involving various assumptions about surface reflectance, cloud and aerosol optical properties, and a priori profile shapes (Gonzalez Abad et al., 2019). Such a priori profile shapes are usually articulated by atmospheric chemical transport models (CTM) either in a free run or constrained by independent observations. Numerous studies have raised concerns over using simulated coarse a priori profiles causing large biases in vertical column densities (Russell et al., 2011; Souri et al., 2016; Goldberg et al., 2019). Not only do atmospheric models introduce biases due to unresolved scales (i.e., loss of spatial variance), uncertainty in emissions products, and atmospheric chemistry routines, but also unresolved physical processes such as convection and local circulations (Janjić et al., 2018). For example, Zhu et al. (2016) identified a tendency of a global weather model to underestimate the mixed-layer depth during the summertime over the southeast US region. This underestimation consequently led to a misrepresentation of HCHO concentrations within that specific part of the atmosphere. The associated inaccuracies in the vertical distribution of HCHO resulted in the introduction of biases of approximately 10–20% in the vertical column densities (Zhu et al., 2020). However, it is crucial to acknowledge that the studies by Zhu et al. (2016, 2020) primarily focused on averaged fields, undermining the influence of rapid atmospheric circulation variations such as those observed in sea-land breezes.

The sea-land wind circulation is a particular meteorological phenomenon that is challenging for coarse spatial-scale atmospheric models to replicate (Miller et al., 2003). This coastal weather phenomenon usually occurs during the day (night) when the land heats up (cools down) faster than the water resulting in a localized area of high (low) pressure over the water and low (high) pressure over the land. While this pattern can always be present in the atmosphere, it becomes more pronounced when synoptically stagnant conditions induced by high-pressure systems result in weak pressure gradients allowing localized gradients to dominate local wind patterns. Accurately representing this circulation has always been a challenge for atmospheric models due to spatial heterogeneity in surface properties, turbulence, synoptic forcing, and the topography of coastal regions. As a result, it is often advised to use ultra-high spatial resolution models (<100 m) to be able to explicitly resolve the momentum fluxes through large eddies (Antonelli and Rotunno, 2007). Nonetheless, there have been successful attempts using parameterized planetary boundary layer schemes to capture the general sea-land breeze flow at a much lower computational cost (Wang et al., 2017; Papanastasiou et al., 2010; Arrillaga et al., 2016; Caicedo et al., 2019).

A pivotal driver of air pollutant concentrations, and their spatiotemporal distribution, is atmospheric transport. Many studies have investigated the sea-land breeze condition responsible for abnormally high or low air pollution depending on the location of emissions, background conditions, and the specific pollutants (Clappier et al., 2000; Liu and Chan, 2002; Banta et al., 2005; Stauffer and Thompson, 2015; Caicedo et al., 2019; Dacic et al., 2020; Li et al., 2020; Geddes et al., 2021; Kotsakis et al., 2022). For instance, Banta et al. (2005) observed elevated surface ozone levels (approximately 200 ppbv) over Houston driven by the transported ozone from the land to the Gulf of Mexico during the night (land breeze) being recirculated back to the land on the next day (sea breeze). Another study by Li et al. (2020) showed high ozone levels between 0.5 and 1 km measured by aircraft over the same city on a sea-breeze day. They attributed this tendency to the non-local updraft flow caused by the sea breeze. Geddes et al. (2021) found low background levels of Ox (NO_2+O_3) during sea breeze days, occurring ~ 30 days in the summer season in the Boston area, accompanied by accumulated Ox due to local emissions existing far inland. These studies indicate complex gradients in air pollutants' vertical and horizontal distribution during sea-land breeze events, which might have significant implications for satellite retrievals (Geddes et al., 2021).

Utilizing a combination of in-situ observations conducted in the New York Bight region during the period of 2010–2020, McCabe and Freedman (2023) identified an annual occurrence of more than 32 sea-breeze days. These occurrences predominantly transpire during the summer season, coinciding with the presence of a surface high pressure system. In a separate study concentrating on the Delaware coastal areas, Hughes and Veron (2018) discovered that during the years 2005–2013, approximately 70–80% of summer days experienced a sea breeze at least once at a coastal meteorological station. However, this frequency dropped to 15% for stations located further inland, specifically exceeding 20 km from the shoreline. These investigations shed light on the considerable impact of sea breezes on local meteorology, given the substantial population residing in coastal regions, emphasizing their implications for both air quality and satellite trace gas retrievals.

Although the impact of the sea-breeze circulation on ozone has been well-studied, to the best of our knowledge, Dasgupta et al. (2005) is the only study tackling the vertical distribution of HCHO pertaining to this type of event. They observed unusual HCHO vertical concentration increasing with altitude based on aircraft measurements associated with land-sea breezes. However, more than sparse pointwise observations are indeed needed to understand the resulting impact on air mass factors used for the conversion of retrieved slant column densities to vertical column densities (Palmer et al., 2001). Therefore, this study will use a well-characterized CTM set up over the east coast of the U.S. in June 2018 to investigate how sea breeze circulation can alter the vertical distribution of HCHO. Because the accurate representation of the vertical shape of HCHO is key for satellite HCHO retrievals, we will study how the differences in the vertical shape of HCHO simulated by a well-characterized, high spatial resolution CTM against an operational product with coarser spatial scales can influence satellite retrievals of HCHO vertical column densities. Furthermore, we will quantify the discrepancies between estimated surface HCHO concentrations derived from total column vertical densities using two different model prior profiles.

2. Model, measurements, and method

2.1. WRF-CMAQ

Air quality simulations are performed using version 5.3 of the Community Multiscale Air Quality (CMAQ) model driven offline by the meteorological fields simulated using version 4.1 of the Weather Research and Forecasting (WRF) model. The WRF and CMAQ models are employed in a nested domain configuration with the outer and inner domains at a horizontal grid spacing of 12 km and 4 km, respectively. The CMAQ domains are a subset of the WRF domains. The outer domain of WRF covers the contiguous U.S., while the outer CMAQ domain only covers the northeastern U.S. The inner WRF and CMAQ domains zoom into the regions of interest for the NASA Ozone Water-Land Environment Transition Study-2 (OWLETS-2) and Long Island Sound Tropospheric Ozone Study (LISTOS) campaigns (Fig. 1). Physical parameterizations in WRF follow Appel et al. (2017), including analysis and soil moisture nudging. The Meteorology-Chemistry Interface Preprocessor (MCIP)

converts WRF-simulated meteorology to CMAQ-ready files. Meteorology-dependent anthropogenic emissions based on the National Emissions Inventory (NEI) representative of 2014 are created for both domains by feeding the MCIP output to the Sparse Matrix Operator Kernel Emissions (SMOKE). NEI 2014 was the latest emission inventory available during emission preparation for this study. The Biogenic Emissions Inventory System (BEIS) implemented for online calculation in CMAQ is used to estimate biogenic emissions of trace gasses and aerosols. The Carbon bond 6 version r3 chemical mechanism represents gas-phase chemistry, and AERO7 describes aerosol processes, including the lifecycle of secondary organic aerosols.

Meteorological boundary conditions for the outer WRF domain are based on the North American Mesoscale (NAM) 12 km Analysis (https://doi.org/10.5065/G4RC-1N91). In contrast, those for the inner domain are dynamically provided during the run-time from the outer domain. Hourly chemical lateral boundary conditions for the outer CMAQ domain are based on global GEOS-Chem simulations v12.0 with the full chemistry mechanism (NOx-Ox–HC–Aer-Br) in a 2° × 2.5° spatial resolution and 47 vertical profiles stretching from the surface to about 80 km with one year spin-up, and those for the inner CMAQ domain are based on the outer CMAQ domain simulation. The simulations were initiated on 15 May 2018 at 00 UTC for both domains, and the first 16 days of the model output have been discarded as the model spin-up. Hourly WRF-CMAQ model output is used for analysis in this study. We also output the WRF-CMAQ integrated process analysis representing the contribution of each physiochemical process to the amount of HCHO.

Johnson et al. (2023) rigorously validated this model setup against aircraft HCHO observations in July and August 2018, suggesting a high degree of correlation in simulated vertical shapes within the first 2 km altitude near the surface. Unfortunately, aircraft data were unavailable for June 2018 when the sea breeze events of interest took place.

2.2. GEOS-CF

We use the output from the NASA GEOS Composition Forecast system (GEOS-CF; Keller et al., 2021) to compare our WRF-CMAQ results with a different model realization. GEOS-CF has been used in several studies to investigate local to regional-scale composition changes (e.g., Dacic et al., 2020; Gronoff et al., 2021; Johnson et al., 2021; Bernier et al., 2022; Sullivan et al., 2022). The GEOS-CF model system couples the GEOS-Chem chemistry transport model (Bey et al., 2001; version 12.0.1 in GEOS-CF version 1) to the GMAO GEOS atmospheric general circulation model (AGCM, GEOS-5, Molod et al., 2015). In addition to a daily 5-day forecast, the GEOS-CF first simulates retrospective estimates of the atmospheric composition for the previous 24 h, where the meteorological analysis increments are calculated using the "replay" technique (Orbe et al., 2017) to the previously-run GEOS Forward Processing for Instrument Teams (GEOS FP-IT) assimilation system (see Fig. S1 in Knowland et al., 2022). The outputs are hourly trace gasses concentrations and meteorological fields globally at 0.25° latitude by 0.25° longitude and 72 hybrid-eta model layers stretching from the surface to 0.01 hPa. Further details on the GEOS-CF model, its input emissions data, chemical processes, and output files can be found in Keller et al. (2021) and Knowland et al. (2022).

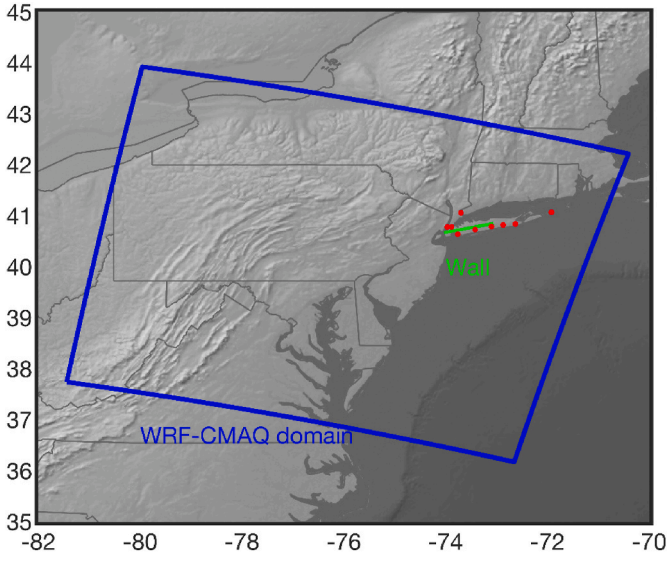


Fig. 1. The location of the case study, along with the 4-km WRF-CMAQ domain, ISD wind measurements (red dots), and a narrow band chosen over Long Island for cross-section analysis (the wall).

2.3. In-situ observations

To evaluate the performance of an atmospheric model's ability to capture near-surface wind conditions, we use integrated surface database (ISD) observations (Smith et al., 2011). We converted the wind speed and direction to the U and V components to be able to perform an ordinary arithmetic mean. The locations of these stations are shown in Fig. 1.

2.4. MODIS SST and LST

To validate the skin temperature simulated by WRF and GEOS-CF, we combined the sea surface temperature (SST) and land surface temperature (LST) retrievals from Moderate Resolution Imaging Spectroradiometer (MODIS) AQUA (~13:30 LT) sensor. As for SST, we used the collection 5 product by NASA Goddard Ocean Biology Processing Group (OBPG) (www.oceancolor.gsfc.gov). The SST retrieval leverages a split-window algorithm applied to two IR channels (11 and 12 μm) (Walton et al., 1998), minimizing the water vapor interference. The algorithm uses in situ buoy SSTs for further adjustment. The retrieval factors in the difference in the sub-surface sea temperature observed by MODIS IR (~1 mm) and the actual skin temperature, which for this study allowed for a direct comparison to skin temperatures simulated by weather forecast models (Kilpatrick et al., 2015). We removed bad pixels using the quality flags provided with the data. The MODIS LST is based on NASA's collection 6 MOD11 product (Wan, 2014). The LST algorithm also uses a split-window algorithm, but its calculations depend on the uncertain surface emissivity knowledge. Thus, the LST accuracy is generally inferior to that of SST. Therefore, we used the main quality flag provided with the data only to consider pixels with accuracy better than 3 °C.

2.5. RTM calculations

Scattering weights (SWs) in the UV spectral region depend on geophysical factors such as satellite observation geometries, O_3 profiles, surface reflectance, and cloud and aerosol properties. Deriving SWs involves radiative transfer calculations, which require those factors to be predetermined as input.

We calculate SWs for each grid box defined in our WRF-CMAQ 4-km domain, applying viewing geometries of Tropospheric Emissions: Monitoring Pollution (TEMPO) and hourly-varying solar geometries. Our method follows the framework developed by the Smithsonian Astrophysical Observatory (SAO), described in detail by Nowlan et al. (2023). In this study, we conduct radiative transfer calculations at 340 nm under aerosol- and cloud-free conditions using the Vector Linearized Discrete Ordinate Radiative Transfer (VLIDORT) model Version 2.8 (Spurr, 2006, 2008; Spurr and Christi, 2019). We use meteorological atmospheric profiles in addition to ozone from GEOS-CF as input to VLIDORT. The ozone profiles from GEOS-CF are preferred over WRF-CMAQ due to a better representation of stratospheric ozone and its contribution to the upper tropospheric region (Knowland et al., 2022).

To co-locate the GEOS-CF profiles with the CMAQ grid boxes, we perform the nearest-neighbor interpolation for GEOS-CF surface pressure and construct 72 layers at each grid box using hybrid sigma-pressure coordinates. Temperatures and ozone profiles are then interpolated horizontally. We determine surface reflectance using different approaches for land and water surfaces. For land, we use the bidirectional reflectance distribution function (BRDF). To estimate BRDF kernels at 340 nm, we extend visible-band retrievals from the MODIS product (MCD43C1 Version 6.1) (Schaaf and Wang, 2015) to the UV wavelengths. The extension employs spectral empirical orthogonal functions (EOFs) derived by combining climatological visible surface reflectance spectra (Zoogman et al., 2016) with the UV databases from the SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY (SCIAMACHY) (Tilstra et al., 2017). We fit the EOFs to BRDF kernels from the first four bands of MODIS (469–859 nm) to account for the surface reflectance at 340 nm (Nowlan et al., 2023). As the MODIS BRDF data are unavailable over the open ocean, we calculate the surface reflectance of water bodies using the Cox-Munk slope distribution (Cox and Munk, 1954). This approach requires wind speed/direction and salinity information. We obtain surface wind data from GEOS-CF (v1) and ocean salinity data from a monthly climatology from the World Ocean Atlas 2009 (Antonov et al., 2010). Additionally, we consider surface-leaving radiance for SW calculations over water surfaces. As input data, we acquire chlorophyll concentrations from the MODIS Terra monthly climatology (Hu et al., 2012).

3. Results and discussion

3.1. Narrative of the events from a model realization

Sea breeze patterns analyzed during this study are described based on a mixture of the information from WRF simulation and surface wind measurements. We refer to two types of sea breeze classified in this study as “pure” (the largest gradient wind perpendicular to the coastline) and “non-pure” (the largest gradient wind parallel to the coastline) (Miller et al., 2003). Both conditions transpire during atmospherically stagnant conditions under synoptic conditions dominated by high pressure, which result in weak isobaric gradients in the lower atmosphere (500 hPa). The absence of strong isobaric gradients at the synoptic scale permits the local pressure gradients induced by the difference in sea-land temperature to create mesoscale circulations. The magnitude of the sea breeze is believed to be proportional to the square root of the land-sea temperature difference (Mathews, 1982). As a rule of thumb, a minimum of 5 °C difference, and a maximum seaward wind speed of 7 m/s, are required to expect “deep” penetration onto land (Miller et al., 2003). This study focuses on a period of three consecutive days, from June 16, 2018, to June 18, 2018. At a synoptic scale, the Bermuda high periphery extended into the eastern U.S. setting up stagnant winds, clear sky, and excess heat (Fig. S1). During this time, the Icelandic low pressure stayed further east compared to its 30-year climatology based on NOAA’s NCEP reanalysis data (Fig. S2). The prevailing winds at 500 hPa were westerly/northwesterly, following the upper-level jet stream over New York City (NYC) (Fig. S3).

Fig. 2 depicts the WRF-simulated air temperature maps at 2 m, surface wind vectors at 20 m overlaid with 500 hPa geopotential height isobars for the three aforementioned days at two timeframes: 0500 to 1000 LST and 1000 to 1800 LST. Immediately evident is a persistent land-sea air temperature contrast between the NYC region (yellow box B) and the ocean (red box A). The contrast is weaker by half during the morning timeframe (5–9 °C) than the afternoon period (13–17 °C) due to reduced incoming solar radiation. The absence of a warmer ocean than the land removes the likelihood of experiencing a land breeze event. Because of the development of the high-pressure system, evident in the expansion of the geopotential isobars, the air temperature progressively increases over land from the beginning of the time period to the end. In particular, the air temperature at 2 m over NYC (box B) increases in the 1000 to 1800 LST timeframe by nearly 2.6 °C per day. On the other hand, the air temperature at 2 m over the ocean (box A) increases at the rate of 0.3 °C per day. As a result, the difference in the land-sea air temperature grows over time.

Concerning the low-level wind patterns, during the morning hours of 06/16 and 06/17, we observe winds parallel to the Long Island shoreline, suggesting that the temperature difference between the land and the ocean was insufficient to override the synoptic condition (i.e., a non-pure sea breeze). Conversely, significant differences in the air temperature between the land and ocean in the afternoon hours result in breezes perpendicular to the shoreline (i.e., a pure sea breeze). As the difference in the air temperature between box A and box B grows over

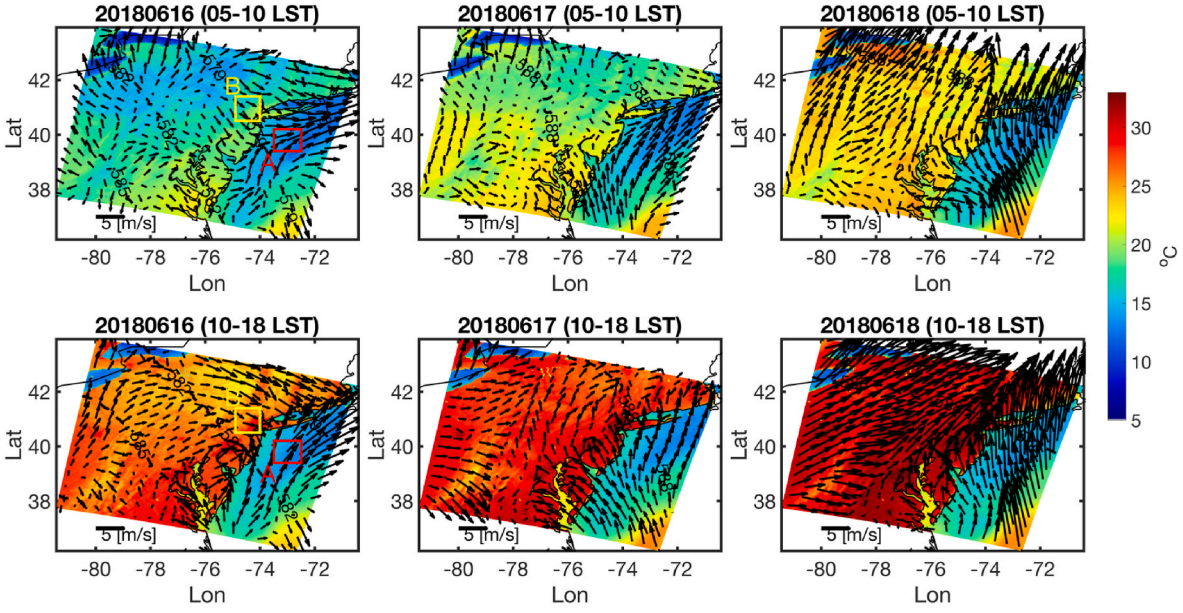


Fig. 2. The general circulation patterns during two timeframes: 0500 to 1000 local standard time (LST) and 1000 to 1800 LST from 06/16 to 06/18. The map shows the simulated air temperature at 2 m overlaid by wind vectors at 20 m and geopotential height isobars at 500 hPa (dashed lines). The unit for the geopotential height is decameter.

time, the sea breeze becomes more prominent. In particular, on 06/18, the sea breeze prevails throughout the day, potentially carrying clean air masses from the ocean over NYC and Long Island.

It is essential to assess if these simulated circulation patterns agree with observations. To evaluate WRF-CMAQ, we use the averaged wind information from 9 ISD measurements around Long Island (Fig. 1). Fig. 3 shows the diurnal evolution of the observed wind pattern and that from the models. On 06/16 and 06/17, both model and observations show weak winds in the early morning, transitioning northwesterly mid-day and forming a southerly sea breeze in the late afternoon (>4 p.m.). The strong sea breeze starting in the afternoon of 06/17, lasted until the end of 06/18. The root mean squared error (RMSE) of the modeled wind speed compared to observations is constantly below 0.7 m/s^{-1} for each day. The model can reproduce 85% of the variance in the observed wind speed. With all hourly model-observation data points combined, the model demonstrates a strong degree of correspondence in meridional winds ($r = 0.95$) and a moderate degree of correlation in zonal winds ($r = 0.77$). Overall, WRF-CMAQ has considerable skill at capturing both wind speed and direction observed from the in-situ measurements suggesting that it is a reliable realization.

3.2. Vertical distribution of HCHO

In this section, we examine the vertical distribution of HCHO over a narrow band over Long Island (the wall in Fig. 1) during the land-sea breeze event. Fig. 4a shows the curtain plot of the simulated HCHO mixing ratios over Long Island from midnight on 06/16 till late afternoon on 06/18. This figure also shows the model simulated horizontal wind vectors and the PBL height (PBLH). We rotate the wind vectors from U and V components in two new coordinates: in parallel (x-direction) and perpendicular (y-direction) to Long Island. For instance, in this new coordination, southerly winds from the Atlantic Ocean are displayed as an upward arrow parallel to the y-axis. The peaks in PBLH are consistently simulated around noontime when the surface heat fluxes (i.e., thermal eddies) are the greatest. A more evolved PBLH around midday manifesting in effective mixing induced by both large (non-local) and small eddies (local diffusivity) result in a relatively homogenous HCHO vertical distribution. During the evening (>1600 LST), the southerly flow from the Atlantic Ocean begins to clean marine air masses over land close to the surface (<1 km above ground level (agl)), followed by the PBLH reduction. In higher altitudes (>1 km agl), westerly winds dominated by the synoptic condition disperse HCHO along Long Island, while a shallower PBLH creates an inversion cap, effectively diminishing the mixing. As a result, elevated HCHO concentrations occur in the residual layer fully decoupled from clean air masses lingering in the shallow PBL. To our best knowledge, the only observational evidence of vertical distribution of HCHO during sea-land breeze was reported by Dasgupta et al. (2005). They observed HCHO mixing ratios increasing with altitudes over two coastal locations near Cedar Key, Florida. While they speculated that tendency resulted from a rising plume, the prevailing sea breeze likely contributed to the observed pattern. Unlike HCHO, more observational evidence of the decoupling pattern has been reported in the literature for ozone. For instance, during the DISCOVER-AQ Texas campaign, Li et al. (2020) observed an elevated ozone plume between 500 m and 1 km during a sea-breeze event in Houston, Texas. They attributed it to the convergence of two air masses, relatively clean air from the Gulf of Mexico and the northerly polluted continental air mass.

Fig. 4b shows the time series of simulated concentrations of surface HCHO (red) and columnar HCHO (blue) along the same cross-section in Fig. 4a. In response to the more rapid oxidation of VOCs by OH and O_3 and large biogenic isoprene emissions, daytime surface HCHO mixing ratios are higher than the nighttime (Dasgupta et al., 2005). During the intense sea breeze event on 06/17 between 1200 and 1800 LST, surface concentration rapidly declined (~0.4 ppbv per hr). Both inhibited chemical production of HCHO due to reduced photochemistry, and effective ventilation simultaneously contribute to this tendency. However, a process analysis will be discussed later to determine the exact processes driving this surface-level HCHO decrease. We see the columnar HCHO only moderately decreasing during the same time period. This is primarily due to elevated HCHO levels in the residual layer compensating for the decline in HCHO close to the surface. The

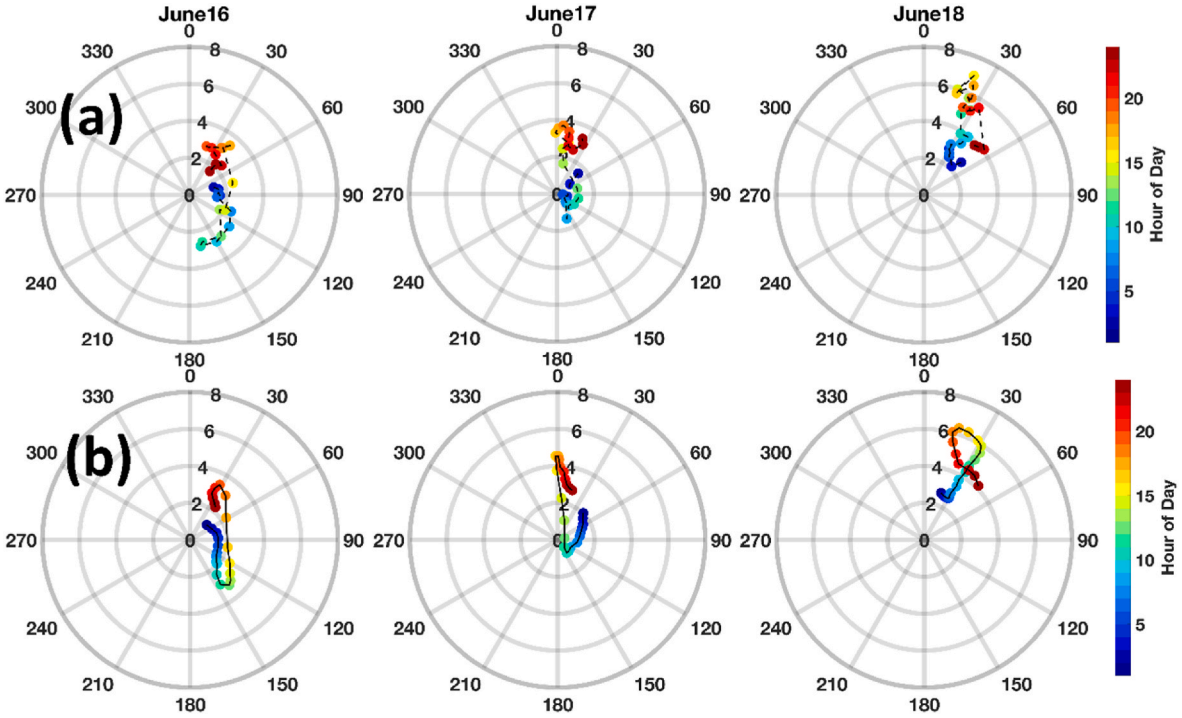


Fig. 3. (a) The averaged wind conditions measured by 9 ISD observations (Fig. 1) throughout the day in the proximity of Long Island, (b) the simulated values synched at the same time and location. Hours in LST. Here the wind condition with zero degrees denotes southerly wind.

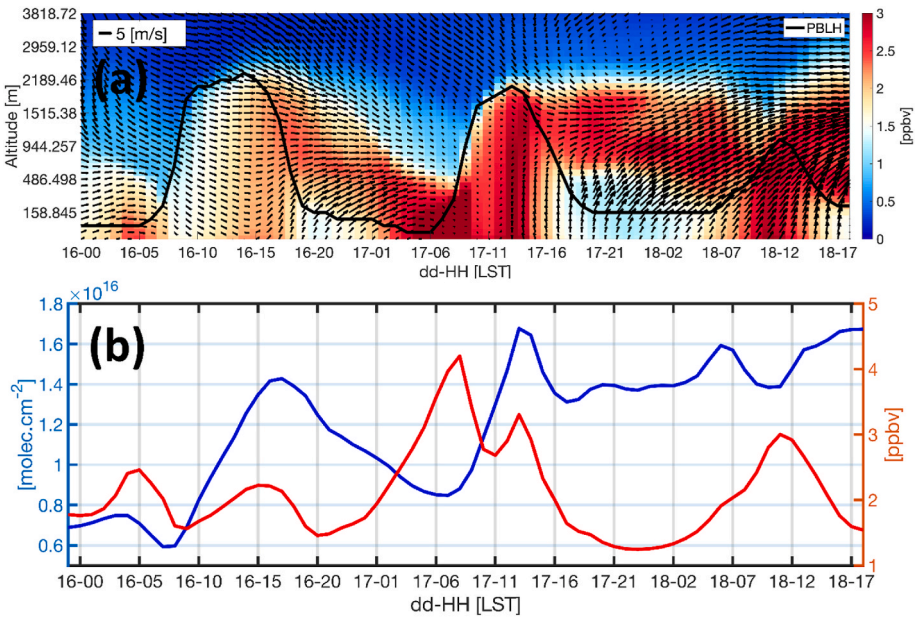


Fig. 4. (a) The curtain plot of WRF-CMAQ simulated HCHO mixing ratios along Long Island from 06/16 0000 LST till 06/18 1700 LST. The solid line is the PBLH. The vertical (horizontal) wind is perpendicular (parallel) to the coastline. (b) The time-series of the total HCHO columns (blue) and the surface HCHO concentrations along the same region in (a).

Pearson correlation coefficient between these two time-series is close to zero, confirming the complexity of the coastal meteorological pattern due to land sea breezes.

The 2-D correlation coefficients between the aforementioned quantities for the entire WRF-CMAQ domain during the same period (06/16 till 06/18) are shown in Fig. 5 which shows the metric for varying heights overlaid by statistically significant values ($p < 0.05$) displayed by dots. The figure is evidence of three distinctive patterns: i) statistically significant positive correlation prevailing over land far from the regions impacted by sea breezes (P1), ii) poor correlation over lakes and regions where the land-sea breezes and synoptic weather patterns interact (P2), and iii) statistically significant negative correlation next to Atlantic coastlines along the east coast, and regions around the Chesapeake Bay (P3). Fig. 5 shows that the higher the altitude, the more prevailing the P1 pattern (positive correlation between surface-level HCHO concentrations and total column abundances) becomes. This tendency indicates that decoupling primarily materializes in areas close to the surface where atmospheric chemistry and dynamics are the most convoluted. To gather more information on why these patterns exist, we calculated the fraction of HCHO columns integrated between the surface and 200 m (F1) and between 200 m and 3000 m (F2) to the total columns. These maps, along with their ratio, are shown in Fig. 6. In an absolute sense, there are always more molecules of HCHO in F2 than in F1, but in a relative sense, their ratio (F2/F1) varies from 4 to 10. Low ratios (<6) are seen primarily in areas where the total HCHO column and the surface concentrations are highly correlated (Fig. 5). This

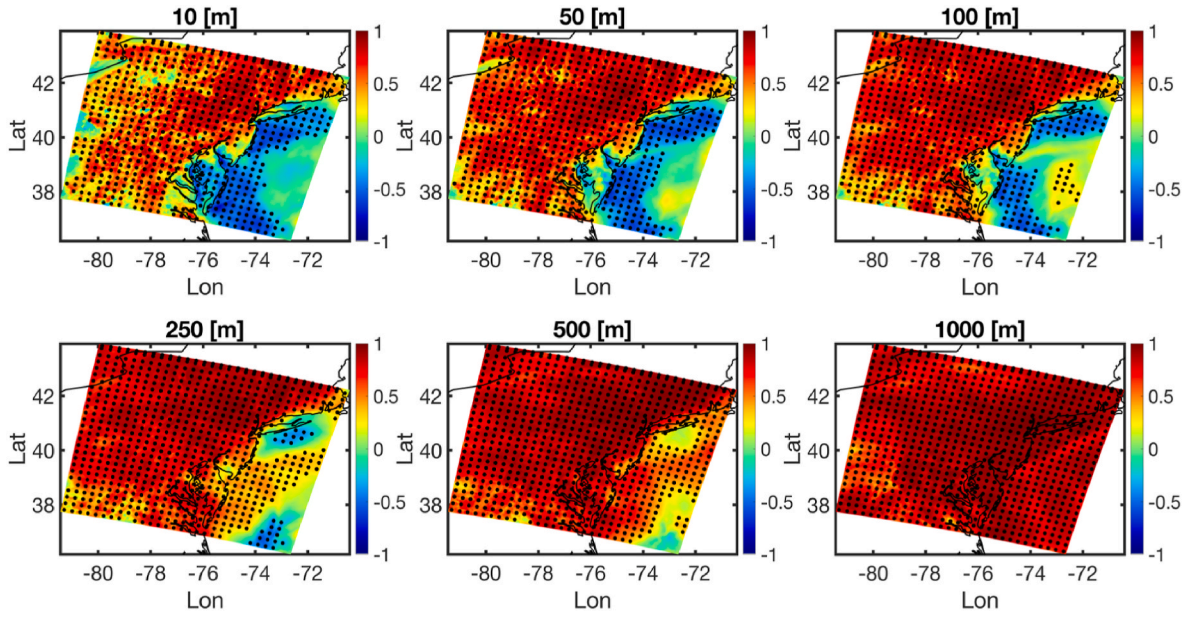


Fig. 5. The Pearson correlation coefficients were between simulated HCHO mixing ratios at different heights (10–1000 m) and simulated HCHO total columns between 06/16 to 06/18. Dots are statistically significant values ($p < 0.05$).

implies that the calm winds, homogeneously diffused HCHO, and effective vertical mixing cause the column to be more representative of the surface concentration. In contrast, high ratios (>6) dominantly occur in coastal and offshore areas where the northward advection of clean marine air conflicts with westerly polluted air masses aloft, creating a decoupled relationship between the surface and columnar concentrations manifesting in negative correlated values in Fig. 5.

3.3. Driving factors in the vertical distribution of HCHO

In the previous section, we realized that two contrasting flows were possibly conducive to a decoupled correlation in HCHO concentrations between the surface and the rest of the atmospheric column. However, other processes, such as the chemical production of HCHO, can contribute to shaping the vertical distribution of HCHO. To quantitatively determine the primary physiochemical drivers of the vertical profile of HCHO over Long Island, we leverage our numerical representation of the atmospheric state by the WRF-CMAQ model to gauge the importance of each process (i.e., chemistry, transport, deposition, etc.). Fig. 7a shows the dominant process (in terms of the absolute magnitude) among six model processes, namely direct HCHO emission, horizontal transport (horizontal advection plus diffusion), vertical transport (vertical advection plus diffusion), cloud chemistry, dry deposition, and chemistry (i.e., secondarily formed HCHO) over the wall. For the most part, the three primary processes driving the vertical distribution of HCHO are vertical and horizontal transport and chemistry.

As for the chemistry, WRF-CMAQ simulations predict a plume-like shape of HCHO maxing out around midday due to active photochemistry (see Fig. 7b). The shape of the plume follows the PBLH where precursor species to HCHO formation reside. On all days, photochemical production of HCHO occurs from early morning till 1600 LST. After that, the photochemical pathways of HCHO formation drop rapidly, contributing more to the decoupling pattern seen in Fig. 4.

Vertical transport dominates the perturbation of HCHO vertical profiles during the development of PBLH in the early mornings. Due to the model grid size being coarse, it does not explicitly resolve vertical momentum fluxes taking place at fine scales (50–300 m); accordingly, the vertical advection is nearly zero in the model (i.e., hydrostatic state), giving rise to the parameterized vertical diffusion dominance. Two processes are involved in the PBL scheme (i.e., ACM2): i) the nonlocal convective mixing induced by large thermal eddies resulting in strong upward diffusion from the surface to non-surface layers and downward mixing from each layer to the one below (i.e., compensation), and ii) the local eddy diffusivity between adjacent vertical layers. The first process vastly surpasses the second one during statically unstable environments when turbulence is large. The sign of the vertical diffusion tendency is interconnected with the compound's vertical gradients. If the concentration decreases by altitude, the vertical diffusion results in the dilution of near-surface concentration and an increase in areas close to PBLH. This is why we see positive (negative) tendencies in altitude above (below) 1 km during times when the HCHO vertical gradient is negative. However, as the gradient turns around in the late afternoon, we see an opposite pattern. An example of this was seen in Souri et al. (2021), which studied the impact of vertical diffusion on surface ozone in Europe. Because ozone increases with altitude, causing vertical diffusion to enhance surface concentrations, Souri et al. (2021) simulated positive tendencies in vertical diffusion. We observe neutral tendencies in mid-PBLH altitudes from Fig. 4, where the two conflicting tendencies cancel out.

The impact of horizontal transport on HCHO distribution is complex because winds and boundary conditions simultaneously contribute to these vertical profiles. We overlay three boxes in Fig. 7d when the sea breeze from the Atlantic Ocean advects clean air masses. During those times, the dominant process controlling HCHO vertical distributions is horizontal advection. In particular, on 06/17, our quantitative analysis suggests that the sea breeze is the main factor in reducing afternoon HCHO below 1 km. On 06/17 and 06/18 days, we see polluted air being advected in higher altitudes enhancing the decoupling pattern seen in Fig. 4. In summary, we find that times that undergo a sea breeze have low HCHO concentration near the surface (<1 km) but experience elevated HCHO aloft due to polluted continental air masses.

3.4. Contrasting GEOS-CF and WRF-CMAQ

This study has demonstrated the physiochemical processes driving a decoupling relationship between HCHO near the surface and the total column during sea breeze events. The land seas breeze circulation is primarily caused by low-level circulations induced by thermal gradients between the ocean and terrestrial regions, which can be challenging for weather forecast models to reproduce, especially models with relatively coarse spatial grid boxes/inputs and less sophisticated physics. The impact the resolution of a model can have on resolving mesoscale processes is considerable (e.g., Wang et al., 2017; Souri et al., 2020c). In this section, we compare a global model with our WRF-CMAQ results. We choose GEOS-CF because it has been officially selected for vertical HCHO shapes used for the TEMPO HCHO retrieval with hourly temporal resolution throughout the daytime.

Fig. 8 shows the same curtain plot used to show the vertical distribution of HCHO over Long Island during the case study (Fig. 4) but with GEOS-CF results. Above 1 km agl height, both models strongly agree with respect to wind patterns. Likewise, both models can capture the midday peaks in HCHO due to photochemistry. These two datasets cross-sections are mildly correlated ($r = 0.77$). However, there are substantial differences in simulated HCHO mixing ratios in the late afternoon to early morning between the models. In particular, GEOS-CF shows largely polluted air masses (>6 ppbv) close to the surface at night. It is believed that HCHO tends to reside longer at night resulting from the lack of photochemistry, reduced OH, and vegetation stomatal closure (i.e., minimal dry deposition rates). While the exact reason for the model tendency in the late afternoon to early morning is unknown due to the lack of process analysis results from GEOS-CF, westerly continental polluted air is very likely to be the cause. In fact, during sea breezes, GEOS-CF shows more frequent westerly winds (non-pure breezes) than

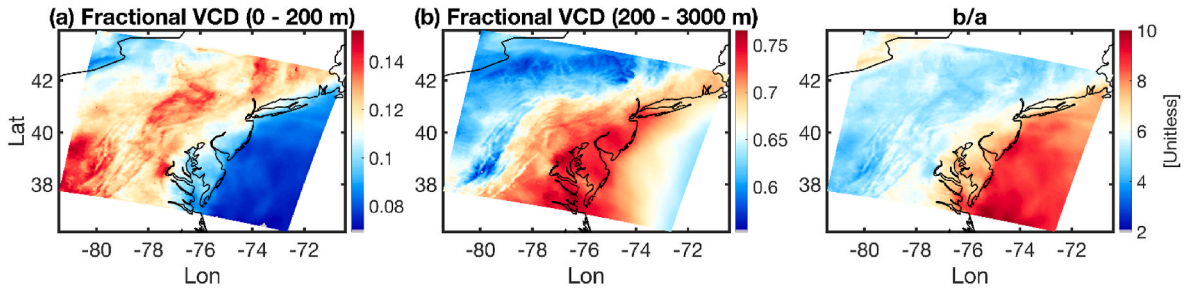


Fig. 6. The fraction of HCHO molecules integrated into (a) 0–200 m and (b) 200–3000 m partial columns compared to the total atmospheric column. The last panel shows the ratio of (b) to (a).

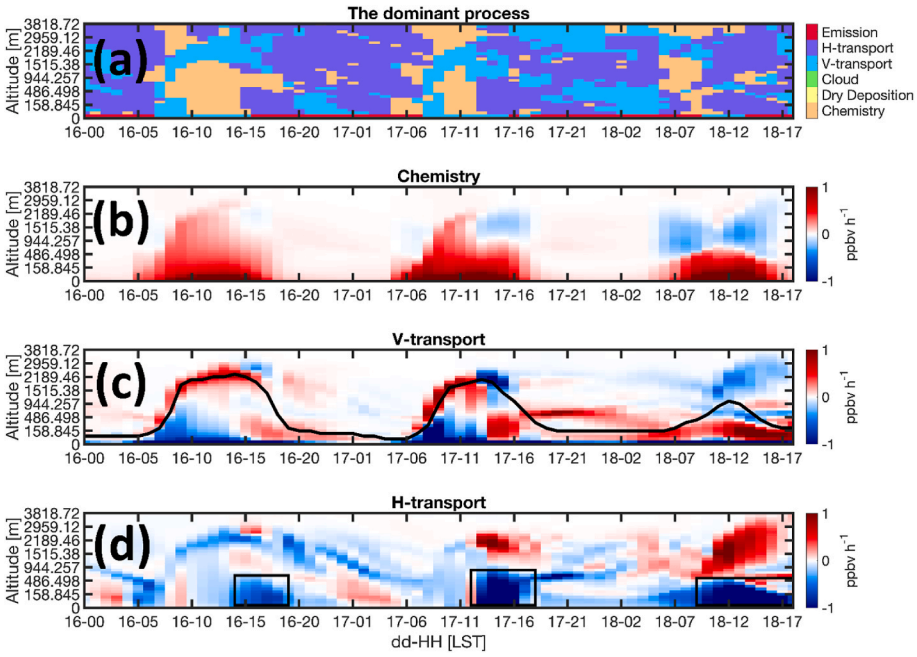


Fig. 7. (a) The dominant physiochemical process in shaping HCHO concentrations over Long Island from 06/16 0000 LST to 06/18 1700 LST. Process-specific HCHO tendencies (ppb/hr) for (b) chemistry (i.e., HCHO chemical production minus loss), (c) the vertical transport (vertical advection plus vertical diffusion), and (d) the horizontal transport (horizontal advection plus horizontal diffusion) are displayed. The black boxes in (d) represent the times of clear advection of clean marine air masses from the Atlantic Ocean. The black line in (c) is the PBLH.

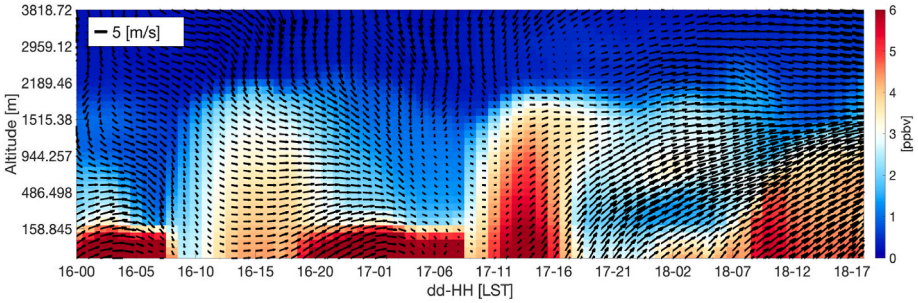


Fig. 8. Same as Fig. 4 but with GEOS-CF results.

WRF (pure breeze) (Fig. S4), carrying a mixture of polluted and clean air masses over Long Island, resulting in more polluted HCHO near the surface.

To gain insight into why GEOS-CF simulates less pure sea breezes, we compare MODIS LST + SST retrievals with the models' skin temperature (Fig. 9) at 1:30 LT. We did not include the results on 06/18 because of widespread cloudiness. Immediately evident from MODIS is the very large surface temperature (>40 °C) over urban settings, including NYC, due to urban heating and abundant sunshine. Encouragingly, WRF can reproduce some of these surface temperature enhancements, but it is biased low. Notably, GEOS-CF surface temperatures over land are too low and fail to recreate these hotspots. The coarser grid spatial resolution of GEOS-CF compared to WRF-CMAQ is unlikely to explain this disagreement since several cities, including NYC, are spatially large enough to cover a GEOS-CF grid box (roughly 25 × 25 km^2) fully. GEOS-CF shows an excellent performance at reproducing the SST over the Atlantic Ocean, possibly due to satellite-based SSTs retrievals being used as boundary conditions in the model. Conversely, WRF SSTs are too cold compared to SST observations. As noted, the thermal gradient between land and sea drives sea breeze circulations. The MODIS retrievals suggest that the difference in the skin temperature between the ocean and NYC is ~12–17 °C around midday. This difference is largely underrepresented by GEOS-CF (8–9 °C) due to the underestimation of land surface temperature. While the WRF SSTs are too low, the thermal contrast between the land/ocean is more reasonably simulated (15–18 °C), thus allowing for the accurate simulation of the sea breeze circulation.

In addition to validating the skin temperature, we directly compare GEOS-CF simulated surface wind vectors at 10 m to the observations used in Fig. 3. GEOS-CF underperforms the simulation of wind speed with RMSEs ranging from 1.3 to 2.6 m/s compared to the WRF-CMAQ (<1 m/s). The afternoon wind directions in GEOS-CF are inclined towards southwesterly rather than southerly winds (Fig. S5). This, in turn, contributes to less effective ventilation. Given that WRF-CMAQ better-simulated surface wind speeds during the land-sea breeze analyzed in this study, the vertical shape of WRF-simulated HCHO will be the benchmark against GEOS-CF for the subsequent analysis.

3.5. HCHO air mass factors

An important implication of the different vertical distribution of HCHO between the two models is the effect it might have on satellite-based vertical column retrievals. This analysis is important for TEMPO as GEOS-CF will be used as the a priori profile of HCHO in AMF calculations (Personal communication, 2022). In principle, nadir satellite remote-sensing provides slant column densities (SCD) along the light path. For optically thin absorbers such as HCHO, this observed line-of-sight can be converted into vertical column densities (VCD) using air mass factor (AMF):

$$AMF = \frac{SCD}{VCD} = \frac{\tau_s}{\tau_v} \quad (1)$$

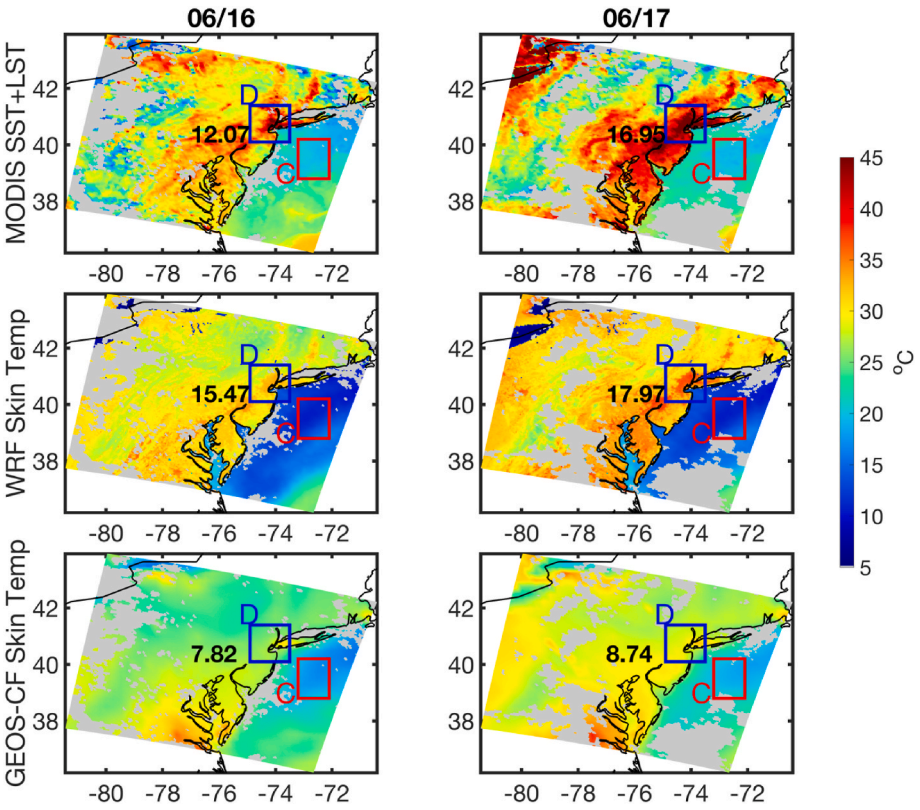


Fig. 9. The comparison of skin temperature observed by MODIS AQUA (~1:30 LT), WRF, and GEOS-CF on 06/16 and 06/17. The number embedded in the maps is the difference in skin temperature (°C) between the land (box D) and the ocean (box C). The gray areas are missing areas in MODIS due to large errors in the retrieval or cloudiness.

where τ_s and τ_v are the unitless optical thickness of the slant and vertical columns. We determine AMFs based on the Beer-Lambert law of extinction (Palmer et al., 2001):

$$AMF = AMF_G \int_0^\infty \omega(z) S(z) dz \tag{2}$$

where AMF_G is the geometric AMF defined in the absence of any extinction in the backscattered radiance along the light path, $\omega(z)$ are the SWs describing the sensitivity of the backscattered radiance to the optical thickness of the gas absorber, and $S(z)$ is the shape factor describing the normalized shape of HCHO partial columns to the total columns:

$$S = \frac{x(z)}{\int_0^\infty x(z) dz} \tag{3}$$

where $x(z)$ is the partial HCHO columns provided by a model. We calculated the hourly SWs using the viewing geometry for the upcoming TEMPO mission using the radiative transfer model described in Sect. 2.5. Regarding shape factors, we will use two prior vertical HCHO distributions simulated by the WRF-CMAQ and GEOS-CF models. Because the simulated vertical distribution of HCHO was found to be the most inhomogeneous on 06/17 due to sea/land breezes, our analysis will focus on this particular day. To isolate the effect of the lower tropospheric inhomogeneity on AMFs, we match HCHO mixing ratios between the two models above 3 km.

Fig. 10 shows four rows, including the simulated HCHO mixing ratios, the resultant shape factors, the SWs multiplied by the AMF_G, and box AMFs (i.e., vertically-resolved AMFs) calculated along Long Island (i.e., the wall). Both models show decreasing HCHO concentrations with altitude in the early hours. As the decoupled relationship between surface and non-surface regions evolves, the two models show conflicting trends, especially between 1400 and 1700 LT when GEOS-CF continues to decrease with height while WRF-CMAQ simulates the decoupled pattern due to the sea breeze. In the late evening (>6 p.m.), GEOS-CF simulates similar HCHO distributions compared to WRF-CMAQ. Due to the lower HCHO mixing ratios in the lower tropospheric region (<2 km) in WRF-CMAQ compared to GEOS-CF driven primarily by the capability to simulate the sea breeze, WRF-CMAQ shape factors are shifted towards greater values in the free tropospheric region. Between 1300 and 1500 LT, when the discrepancies between the two models are the highest, we observe an enhancement in the WRF-CMAQ shape factors around 2 km where the elevated HCHO mixing ratios reside. The SWs increase by altitudes resulting from less sensitivity of the backscattered radiation the further it travels into the atmosphere. Furthermore, the SWs increase rapidly due to longer light paths for higher solar zenith angles. Concerning the box AMFs, the product of SWs, and shape factors, it is worth noting that the AMF calculations are extended to the top of the atmosphere but are not shown in the figure. We observe negligible differences in box AMFs at 1000 LST and 1900 LT when the shape factors from both models agree well. However, at other times, particularly at 1500 LT, when the shape factors disagree the most, the resultant AMFs disagree up to nearly 20%. This error is uniquely caused by the differences in the vertical shape of HCHO below 3 km altitude between the two model's capability to simulate sea breeze circulation.

It is important to acknowledge that the calculations of SWs in our study did not factor in aerosol interferences. This omission was primarily due to the region exhibiting generally low AOD values, which typically do not result in noticeable interferences on the optical path, as noted by Jung et al. (2019). Nevertheless, in areas experiencing significant aerosol loading, the decoupling pattern can cause a vertically heterogeneous distribution of aerosol optical properties and concentration.

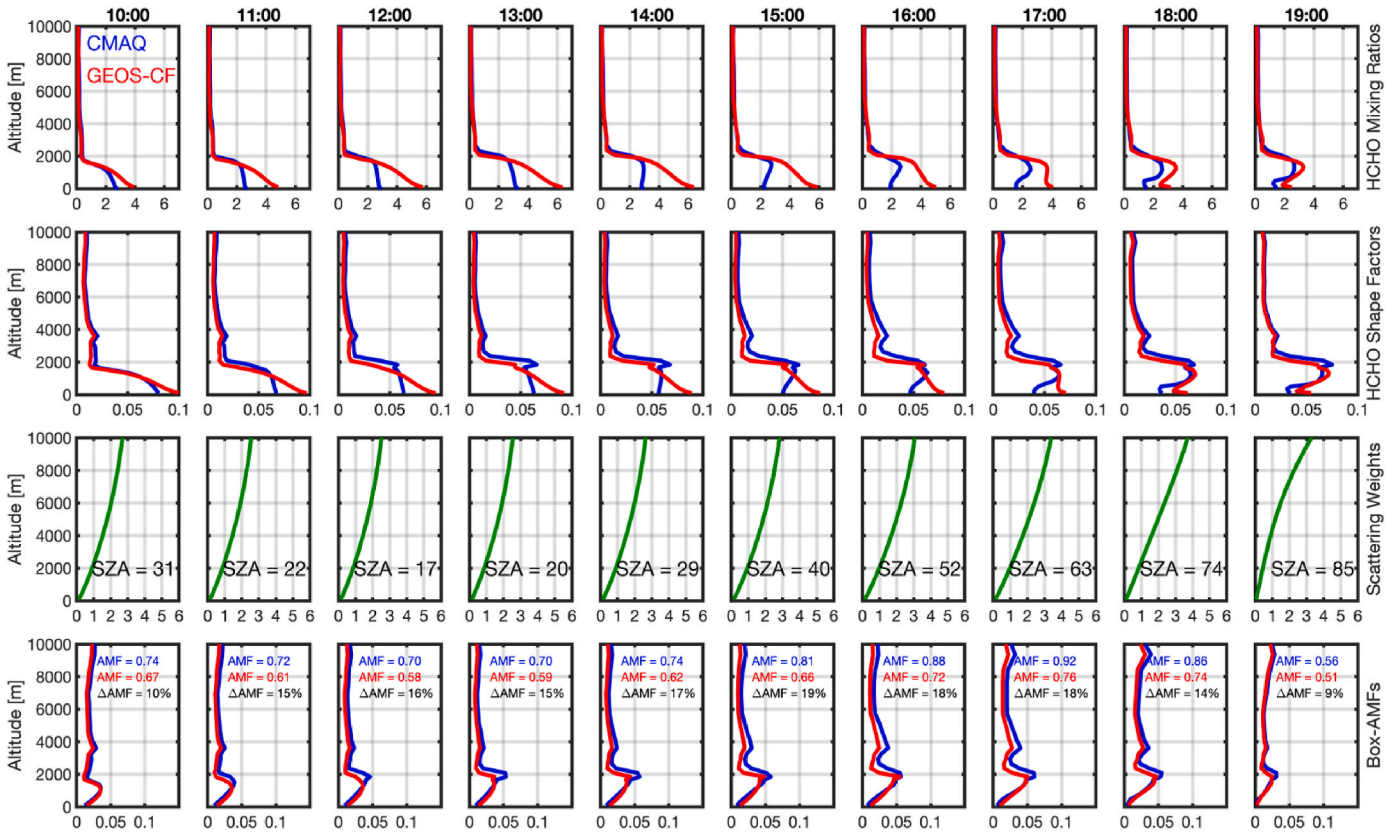


Fig. 10. (first row) The comparison of vertical mixing ratios of HCHO in WRF-CMAQ (blue) and GEOS-CF (red), (second row) HCHO shape factors from the two models, (third row) calculated SWs multiplied by geometric AMFs, and (fourth row) box AMFs from the two models. WRF-CMAQ HCHO mixing ratios are replaced by GEOS-CF above 3 km altitude to isolate the lower tropospheric effect. The location of these profiles is over Long Island, shown as a wall in Fig. 1.

This introduces additional layers of complexity in the calculations of AMFs. Consequently, it is imperative to conduct a separate follow-up study dedicated to investigating this particular aspect.

To gather a better picture of the extent to which the differences in AMF can affect VCDs, we compare WRF-CMAQ HCHO total columns to those recalculated by AMFs using the GEOS-CF shape factors using:

$$VCD_{CMAQ-recal} = \frac{VCD_{CMAQ} AMF_{CMAQ}}{AMF_{GEOS-CF}} \quad (4)$$

where AMF_{CMAQ} and $AMF_{GEOS\text{-}CF}$ are AMFs computed with WRF-CMAQ and GEOS-CF shape factors, respectively. We again matched their mixing ratios above 3 km altitude to isolate the lower tropospheric effect. Fig. 11 demonstrates the comparison of VCD_{CMAQ} and $VCD_{CMAQ\text{-}recal}$ at 1500 LT on 06/17 in the region where the decoupling of near-surface and total column HCHO issue caused by the sea breeze is present. We observe disagreements up to 30% due to the differences in the model shape factors. GEOS–CF–derived AMFs tend to overestimate VCDs because they are smaller than those of WRF-CMAQ. This result is evidence of the complexities associated with the vertical shape of HCHO simulated differently by models and the impact this has on satellite-retrievals of column HCHO.

3.6. Column-to-surface conversion

The two different models analyzed in this study provide two different realizations of the vertical shapes of HCHO over the region with a complex atmospheric circulation pattern. A major concern in translating the satellite-based columnar observations to the near-surface condition is the dependence upon these vertical shapes. In this section, we quantify

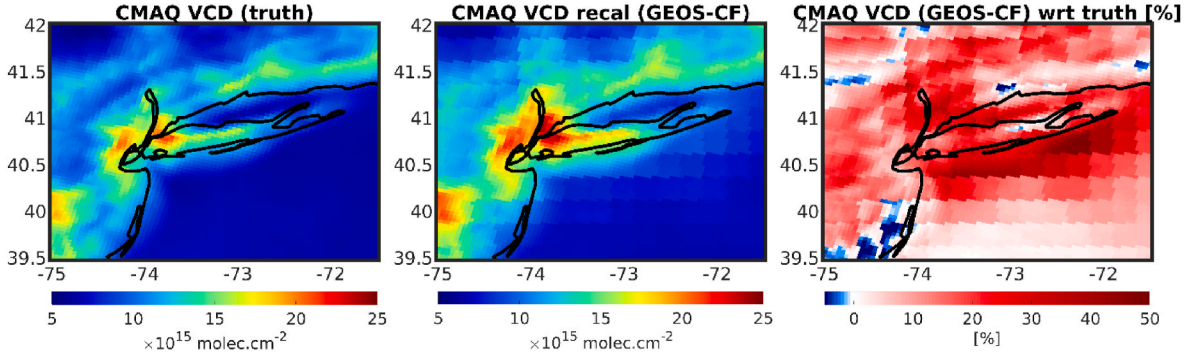


Fig. 11. Total column HCHO abundances (left) from WRF-CMAQ on 06/17 at 15:00 LST, (middle) WRF-CMAQ VCDs recalculated with AMFs using GEOS-CF shape factors, and (right) the normalized differences.

the impact of the differences in GEOS-CF and WRF-CMAQ vertical HCHO shapes on this translation. HCHO total columns can be converted to mixing ratios at the near surface using:

$$HCHO^{surface} = VCD_{sat}^{total}\frac{VCD_{model}^{surface}}{VCD_{model}^{total}} \times f(T,P,dz) \quad (5)$$

where VCD_{sat}^{total} and VCD_{model}^{total} are the VCD of a satellite and a model, respectively; $VCD_{model}^{surface}$ represents the partial VCDs of the model at the lowest layer close to the surface. f is a conversion factor used for converting surface partial columns to surface mixing ratios which is inversely proportional to the total number of air molecules residing in the closest layer to the surface. T, P, and dz are temperature, pressure, and layer thickness.

For our experiment, we made a few assumptions: both models followed the exact mixing ratios above 3 km altitude, WRF-CMAQ HCHO mixing ratios are resampled at the GEOS-CF vertical grid using a bilinear interpolation, VCD_{CMAQ} is considered the true state of the atmosphere observed by a hypothetical satellite (VCD_{sat}^{total}), the f conversion factor is based on GEOS-CF values. In this experiment, we first calculate the surface HCHO mixing ratios using WRF-CMAQ outputs (Fig. 12, left panel). Because VCD_{model}^{total} and VCD_{sat}^{total} are identical based on our assumptions, the output is simply the partial surface WRF-CMAQ column converted to HCHO mixing ratios (truth). In a different scenario, we use GEOS-CF in VCD_{model}^{total} and convert VCD_{sat}^{total} (i.e., VCD_{CMAQ}) to the surface HCHO mixing ratios (Fig. 12, middle panel). The difference between these two outputs (Fig. 12, right panel) explains the effect of varying vertical shapes of HCHO on the column-to-surface translation. In response to large HCHO concentrations near the surface in GEOS-CF and its decreasing trend with altitude, we observe that WRF-CMAQ HCHO concentrations become overestimated when GEOS-CF information is used. In other words, GEOS-CF overly allocates the total amount of HCHO molecules from VCD_{sat}^{total} to the near-surface region because the clean air mass driven by the sea breeze is not fully developed in the model. The difference map highly correlates with the one shown in Fig. 11. However, the differences here are larger. Retrieval sensitivities (i.e., SWs) cause this difference; in the former experiment involving recalculating AMFs, smaller SWs close to the surface partly diminish the effect of varying vertical shapes of HCHO between two models. In contrast, this experiment does not depend on the sensitivity reinforcing the importance of vertical shapes in the application of surface estimation using satellite-based VCDs.

4. Conclusions

Satellite retrievals of trace gasses such as formaldehyde (HCHO), a critical compound widely used for monitoring volatile organic compounds (VOC) emissions, rely on the assumptions made for a priori profiles. These a priori profiles are commonly generated by chemical transport models (CTMs) subject to unresolved scales and processes, potentially introducing large biases in the retrieval. One of the unique challenges in coastal areas is the sea-land breeze circulation, which can create strong vertical and horizontal gradients in atmospheric composition. Accurately representing this phenomenon requires various aspects of a numerical model, including synoptic forcing, thermal differences, and surface properties, to be simulated reasonably. To understand how differently the vertical shape of HCHO behaves under such conditions, we set up a skillful CTM over the east coast U.S. in June 2018 when the synoptic conditions favored the development of sea breezes.

Between June 16 and 18, a dominant high-pressure system evident in calm night winds enabled the significant thermal difference between the land and the Atlantic Ocean observed by moderate resolution imaging spectroradiometer (MODIS) skin temperature and well captured by WRF-CMAQ to generate monsoon-like flows from the ocean. This model accurately produced these patterns compared to coastal surface wind measurements. The southerly breeze over Long Island and New York City (NYC) carried clean marine air masses, with low HCHO concentrations, over the land, cleaning out the lower few hundred meters above the surface in the afternoons. Conversely, in higher altitudes, the westerly winds dictated by synoptic conditions transported continental polluted air masses to the region, causing a strong decoupling relationship between near-surface HCHO mixing ratios and integrated vertical columns. To understand this pattern, we used model process analysis diagnostics and found that horizontal advection (i.e., sea breeze) was the most dominant process responsible for decreasing HCHO levels close to the surface. This atypical tendency was conducive to positive vertical gradients in HCHO, which differs drastically from typical decreasing vertical trends in this trace gas.

We selected a state-of-the-art global transport model planned to provide a priori profile information for the upcoming Tropospheric Emissions: Monitoring Pollution (TEMPO) mission retrievals to verify if a coarser resolution model can adequately capture the sea breeze circulation conducive to the unusual HCHO vertical shapes. While GEOS-CF could pick up the sea breeze in late afternoons, its timing and strength were insufficient, resulting in HCHO vertical shapes drastically different from WRF-CMAQ. Similar errors with GEOS-CF under-representing land-sea breeze events were reported in previous studies (Dacic et al., 2020; Bernier et al., 2022); the inability of GEOS-CF version 1 to capture such events was attributed in this study to the underrepresentation of land surface temperature over cities (e.g., Fig. 9), which is likely linked to a cold bias in the GEOS FP-IT meteorological fields in this region (not shown). A remedy to this cold bias is an ongoing development effort for GEOS-CF.

We leveraged the difference in the vertical shapes of HCHO simulated by WRF-CMAQ against GEOS-CF to investigate the effect they could have on the AMF calculations used for converting slant column density (SCD) to vertical column density (VCD) in satellite retrievals. Based on our radiative transfer model simulations pinned to the TEMPO

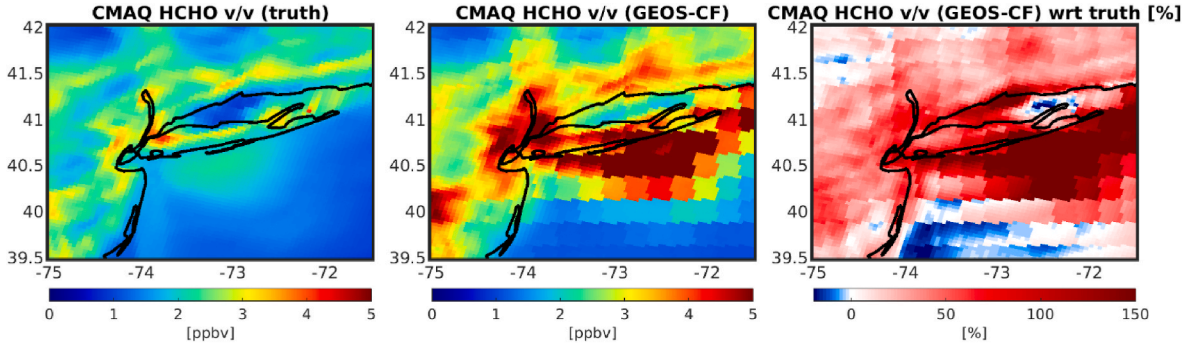


Fig. 12. (left) The surface HCHO mixing ratios from WRF-CMAQ, (middle) the surface HCHO mixing ratios estimated based on GEOS-CF profile shapes applied to WRF-CMAQ total HCHO columns, and (right) their normalized differences.

viewing condition, we noticed that their differences in the first 3 km altitude at the peak of sea breeze contribution were sufficient to introduce a 20–30% error in AMFs, making it difficult to represent HCHO retrievals accurately. Additionally, when comparing the translation of total columns to surface concentrations, we found that the differing vertical shapes resulted in significant biases. Specifically, the profile shapes in GEOS-CF incorrectly allocated 50–150% more HCHO molecules to the surface layer due to not fully capturing the sea breeze effects evaluated in this study.

It is intuitively clear that many people reside in coastal environments globally and are subject to degraded air quality and unique atmospheric phenomena such as the sea-land breeze. There is a recent shift in trace gas satellite observation techniques from low Earth orbit (LEO) systems to constellations of geostationary Earth orbit (GEO) sensors, enabling more efficient and prompt detection and response to air quality incidents. However, errors in a priori profiles can easily degrade the reliability of these measurements, often underrepresented by operational products. The GEO systems put more strain on the CTMs to provide accurate a priori profiles. These new retrievals require the models to accurately simulate trace gas profiles at higher spatial resolution and hourly timesteps through the diel cycle. We showed that the sea breeze circulation is one of the obstacles to achieving the accurate representation of the vertical shapes; however, other processes, such as convective transport, varying emissions, chemical mechanisms, and vertical diffusion, need to be factored in. Our work highlights the challenges the retrieval and atmospheric modeling communities should overcome to ensure the quantitative measurements of trace gasses are biased-free, allowing for accurate and effective decision-making for emission mitigation.

CRediT authorship contribution statement

Amir H. Souri: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft. **Rajesh Kumar:** Methodology, Validation, Writing – review & editing. **Heesung Chong:** Methodology. **Maryam Golbazi:** Validation. **K. Emma Knowland:** Data curation, Writing – review & editing. **Jeffrey Geddes:** Funding acquisition, Writing – review & editing. **Matthew S. Johnson:** Writing – review & editing, Project administration, Funding acquisition.

Declaration of competing interest

The contact author has declared that none of the authors has any competing interests.

Data availability

Data will be made available on request.

Acknowledgments

Amir Souri, Matthew Johnson, and Heesung Chong were funded for this work through NASA's Aura Science Team (grant no. 80 NSSC21K1333). Two authors (Kumar and Golbazi) work at the National Center for Atmospheric Research, which the National Science Foundation sponsors. K. Emma Knowland acknowledges support by the NASA Modeling, Analysis and Prediction (MAP) Program (Project manager, David Considine). Resources supporting the GEOS-CF were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2023.119929.

References

Anderson, D.C., Follette-Cook, M.B., Strode, S.A., Nicely, J.M., Liu, J., Ivatt, P.D., Duncan, B.N., 2022. A machine learning methodology for the generation of a parameterization of the hydroxyl radical. Geosci. Model Dev. (GMD) 15, 6341–6358. https://doi.org/10.5194/gmd-15-6341-2022.

Antonelli, M., Rotunno, R., 2007. Large-eddy simulation of the onset of the sea breeze. J. Atmos. Sci. 64, 4445–4457. https://doi.org/10.1175/2007JAS2261.1.

Antonov, J.I., Seidov, D., Boyer, T.P., Locarnini, R.A., Mishonov, A.V., Garcia, H.E., Baranova, O.K., Zweng, M.M., Johnson, D.R., 2010. World Ocean Atlas 2009. In: Levitus, S. (Ed.), NOAA Atlas NESDIS 69, 2. U.S. Gov. Printing Office, Washington, D.C., p. 184. Salinity.

Antonov, J.I., Seidov, D., Boyer, T.P., Locarnini, R.A., Mishonov, A.V., Garcia, H.E., et al., 2010. In: Levitus, S. (Ed.), World Ocean Atlas 2009, 2. U.S. Government Printing Office. Salinity (vol. NOAA Atlas NESDIS 69).

Appel, K.W., Napelenok, S.L., Foley, K.M., Pye, H.O.T., Hogrefe, C., Luecken, D.J., Bash, J.O., Roselle, S.J., Pleim, J.E., Foroutan, H., Hutzell, W.T., Pouliot, G.A., Sarwar, G., Fahey, K.M., Gantt, B., Gilliam, R.C., Heath, N.K., Kang, D., Mathur, R., Schwede, D.B., Spero, T.L., Wong, D.C., Young, J.O., 2017. Description and evaluation of the community Multiscale air quality (CMAQ) modeling system version 5.1. Geosci. Model Dev. (GMD) 10, 1703–1732. https://doi.org/10.5194/gmd-10-1703-2017.

Arrillaga, J.A., Yagüe, C., Sastre, M., Román-Cascón, C., 2016. A characterisation of sea-breeze events in the eastern Cantabrian coast (Spain) from observational data and WRF simulations. Atmos. Res. 181, 265–280. https://doi.org/10.1016/j.atmosres.2016.06.021.

Banta, R.M., Senff, C.J., Nielsen-Gammon, J., Darby, L.S., Ryerson, T.B., Alvarez, R.J., Sandberg, S.P., Williams, E.J., Trainer, M., 2005. A bad air day in Houston. Bull. Am. Meteorol. Soc. 86, 657–670. https://doi.org/10.1175/BAMS-86-5-657.

Bauwens, M., Stavrakou, T., Müller, J.-F., De Smedt, I., Van Roozendael, M., van der Werf, G.R., Wiedinmyer, C., Kaiser, J.W., Sindelarova, K., Guenther, A., 2016. Nine years of global hydrocarbon emissions based on source inversion of OMI formaldehyde observations. Atmos. Chem. Phys. 16, 10133–10158. https://doi.org/10.5194/acp-16-10133-2016.

Bernier, C., Wang, Y., Gronoff, G., Berkoff, T., Knowland, K.E., Sullivan, J.T., Delgado, R., Caicedo, V., Carroll, B., 2022. Cluster-based characterization of multi-dimensional tropospheric ozone variability in coastal regions: an analysis of lidar measurements and model results. Atmos. Chem. Phys. 22 (23), 15313–15331. https://doi.org/10.5194/acp-22-15313-2022.

Bey, I., Jacob, D.J., Yantosca, R.M., Logan, J.A., Field, B.D., Fiore, A.M., Li, Q., Liu, H.Y., Mickley, L.J., Schultz, M.G., 2001. Global modeling of tropospheric chemistry with assimilated meteorology: model description and evaluation. J. Geophys. Res. Atmos. 106, 23073–23095. https://doi.org/10.1029/2001JD000807.

Caicedo, V., Rappenglueck, B., Cuchiara, G., Flynn, J., Ferrare, R., Scarino, A.J., Berkoff, T., Senff, C., Langford, A., Lefer, B., 2019. Bay breeze and sea breeze circulation impacts on the planetary boundary layer and air quality from an observed and modeled DISCOVER-AQ Texas case study. J. Geophys. Res. Atmos. 124, 7359–7378. https://doi.org/10.1029/2019JD030523.

Chance, K., Palmer, P.I., Spurr, R.J.D., Martin, R.V., Kurosu, T.P., Jacob, D.J., 2000. Satellite observations of formaldehyde over North America from GOME. Geophys. Res. Lett. 27, 3461–3464. https://doi.org/10.1029/2000GL011857.

Chance, K.V., Burrows, J.P., Schneider, W., 1991. Retrieval and molecule sensitivity studies for the global ozone monitoring experiment and the scanning imaging absorption spectrometer for atmospheric chartography. In: Remote Sensing of Atmospheric Chemistry. Presented at the Remote Sensing of Atmospheric Chemistry, SPIE, pp. 151–165. https://doi.org/10.1117/12.46657.

Clappier, A., Martilli, A., Grossi, P., Thunis, P., Pasi, F., Krueger, B.C., Calpini, B., Graziani, G., van den, Bergh H., 2000. Effect of sea breeze on air pollution in the greater athens area. Part I: numerical simulations and field observations. J. Appl. Meteorol. Climatol. 39, 546–562. https://doi.org/10.1175/1520-0450(2000)039<0546:EOSBOA>2.0.CO;2.

Cox, C., Munk, W., 1954. Measurement of the roughness of the sea surface from photographs of the sun's glitter. J. Opt. Soc. Am., JOSA 44, 838–850. https://doi.org/10.1364/JOSA.44.000838.

Dacic, N., Sullivan, J.T., Knowland, K.E., Wolfe, G.M., Oman, L.D., Berkoff, T.A., Gronoff, G.P., 2020. Evaluation of NASA's high-resolution global composition simulations: understanding a pollution event in the Chesapeake Bay during the summer 2017 OWLETS campaign. Atmos. Environ. 222, 117133 https://doi.org/10.1016/j.atmosenv.2019.117133.

Dasgupta, P.K., Li, J., Zhang, G., Luke, W.T., McClenny, W.A., Stutz, J., Fried, A., 2005. Summertime ambient formaldehyde in five U.S. Metropolitan areas: nashville, atlanta, Houston, philadelphia, and tampa. Environ. Sci. Technol. 39, 4767–4783. https://doi.org/10.1021/es048327d.

De Smedt, I., Müller, J.-F., Stavrakou, T., van der A, R., Eskes, H., Van Roozendael, M., 2008. Twelve years of global observations of formaldehyde in the troposphere using GOME and SCIAMACHY sensors. Atmos. Chem. Phys. 8, 4947–4963. https://doi.org/10.5194/acp-8-4947-2008.

De Smedt, I., Pinardi, G., Vigouroux, C., Compernolle, S., Bais, A., Benavent, N., Boersma, F., Chan, K.-L., Donner, S., Eichmann, K.-U., Hedelt, P., Hendrick, F., Irie, H., Kumar, V., Lambert, J.-C., Langerock, B., Lerot, C., Liu, C., Loyola, D., Piters, A., Richter, A., Rivera Cárdenas, C., Romahn, F., Ryan, R.G., Sinha, V., Theys, N., Vlietinck, J., Wagner, T., Wang, T., Yu, H., Van Roozendael, M., 2021. Comparative assessment of TROPOMI and OMI formaldehyde observations and validation against MAX-DOAS network column measurements. Atmos. Chem. Phys. 21, 12561–12593. https://doi.org/10.5194/acp-21-12561-2021.

De Smedt, I., Stavrakou, T., Hendrick, F., Danckaert, T., Vlemmix, T., Pinardi, G., Theys, N., Lerot, C., Gielen, C., Vigouroux, C., Hermans, C., Fayt, C., Veefkind, P., Müller, J.-F., Van Roozendael, M., 2015. Diurnal, seasonal and long-term variations of global formaldehyde columns inferred from combined OMI and GOME-2 observations. Atmos. Chem. Phys. 15, 12519–12545. https://doi.org/10.5194/acp-15-12519-2015.

De Smedt, I., Stavrakou, T., Müller, J.-F., van der A, R.J., Van Roozendael, M., 2010. Trend detection in satellite observations of formaldehyde tropospheric columns. Geophys. Res. Lett. 37 https://doi.org/10.1029/2010GL044245.

De Smedt, I., Theys, N., Yu, H., Danckaert, T., Lerot, C., Compernolle, S., Van Roozendael, M., Richter, A., Hilboll, A., Peters, E., Pedergnana, M., Loyola, D., Beirle, S., Wagner, T., Eskes, H., van Geffen, J., Boersma, K.F., Veefkind, P., 2018. Algorithm theoretical baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project. Atmos. Meas. Tech. 11, 2395–2426. https://doi.org/10.5194/amt-11-2395-2018.

Duncan, B.N., Yoshida, Y., Olson, J.R., Sillman, S., Martin, R.V., Lamsal, L., Hu, Y., Pickering, K.E., Retscher, C., Allen, D.J., Crawford, J.H., 2010. Application of OMI observations to a space-based indicator of NOx and VOC controls on surface ozone formation. Atmos. Environ. 44, 2213–2223. https://doi.org/10.1016/j.atmosenv.2010.03.010.

Geddes, J.A., Wang, B., Li, D., 2021. Ozone and nitrogen dioxide pollution in a coastal urban environment: the role of sea breezes, and implications of their representation for remote sensing of local air quality. J. Geophys. Res. Atmos. 126, e2021JD035314 https://doi.org/10.1029/2021JD035314.

Goldberg, D.L., Saide, P.E., Lamsal, L.N., de Foy, B., Lu, Z., Woo, J.-H., Kim, Y., Kim, J., Gao, M., Carmichael, G., Streets, D.G., 2019. A top-down assessment using OMI NO2 suggests an underestimate in the NOx emissions inventory in Seoul, South Korea, during KORUS-AQ. Atmos. Chem. Phys. 19, 1801–1818. https://doi.org/10.5194/acp-19-1801-2019.

González Abad, G., Liu, X., Chance, K., Wang, H., Kurosu, T.P., Suleiman, R., 2015. Updated smithsonian astrophysical observatory ozone monitoring instrument (SAO OMI) formaldehyde retrieval. Atmos. Meas. Tech. 8, 19–32. https://doi.org/10.5194/amt-8-19-2015.

Gonzalez Abad, G., Souri, A.H., Bak, J., Chance, K., Flynn, L.E., Krotkov, N.A., Lamsal, L., Li, C., Liu, X., Miller, C.C., Nowlan, C.R., Suleiman, R., Wang, H., 2019. Five decades observing Earth's atmospheric trace gases using ultraviolet and visible backscatter solar radiation from space. J. Quant. Spectrosc. Radiat. Transf. 238, 106478 https://doi.org/10.1016/j.jqsrt.2019.04.030.

Gronoff, G., Berkoff, T., Knowland, K., Lei, L., Shook, M., Fabbri, B., Carrion, W., Langford, A., 2021. Case study of stratospheric Intrusion above Hampton, Virginia: lidar-observation and modeling analysis. Atmos. Environ., 118498 https://doi.org/10.1016/j.atmosenv.2021.118498.

Hu, C., Lee, Z., Franz, B., 2012. Chlorophyll aalgorithms for oligotrophic oceans: a novel approach based on three-band reflectance difference. J. Geophys. Res.: Oceans 117. https://doi.org/10.1029/2011JC007395.

Hughes, C.P., Veron, D.E., 2018. A characterization of the Delaware sea breeze using observations and modeling. J. Appl. Meteorol. Climatol. 57, 1405–1421. https://doi.org/10.1175/JAMC-D-17-0186.1.

Janjić, T., Bormann, N., Bocquet, M., Carton, J.A., Cohn, S.E., Dance, S.L., Losa, S.N., Nichols, N.K., Potthast, R., Waller, J.A., Weston, P., 2018. On the representation error in data assimilation. Q. J. R. Meteorol. Soc. 144, 1257–1278. https://doi.org/10.1002/qj.3130.

Jin, X., Fiore, A.M., Murray, L.T., Valin, L.C., Lamsal, L.N., Duncan, B., Folkert Boersma, K., De Smedt, I., Abad, G.G., Chance, K., Tonnesen, G.S., 2017. Evaluating a space-based indicator of surface ozone-NOx-VOC sensitivity over midlatitude source regions and application to decadal trends. J. Geophys. Res. Atmos. 122 (10) https://doi.org/10.1002/2017JD026720, 439-10,461.

Johnson, M.S., Strawbridge, K., Knowland, K.E., Keller, C., Travis, M., 2021. Long-range transport of Siberian biomass burning emissions to North America during FIREX-AQ. Atmos. Environ. 252, 118241 https://doi.org/10.1016/j.atmosenv.2021.118241.

Johnson, M.S., Souri, A.H., Philip, S., Kumar, R., Naeger, A., Geddes, J., Judd, L., Janz, S., Chong, H., Sullivan, J., 2023. Satellite remote-sensing capability to assess tropospheric-column ratios of formaldehyde and nitrogen dioxide: case study during the Long Island Sound Tropospheric Ozone Study 2018 (LISTOS 2018) field campaign. Atmos. Meas. Tech. 16, 2431–2454. https://doi.org/10.5194/amt-16-2431-2023.

Jung, Y., González Abad, G., Nowlan, C.R., Chance, K., Liu, X., Torres, O., Ahn, C., 2019. Explicit aerosol correction of OMI formaldehyde retrievals. Earth Space Sci. 6, 2087–2105. https://doi.org/10.1029/2019EA000702.

Keller, C.A., Knowland, K.E., Duncan, B.N., Liu, J., Anderson, D.C., Das, S., Lucchesi, R. A., Lundgren, E.W., Nicely, J.M., Nielsen, E., Ott, L.E., Saunders, E., Strode, S.A., Wales, P.A., Jacob, D.J., Pawson, S., 2021. Description of the NASA GEOS composition forecast modeling system GEOS-CF v1.0. J. Adv. Model. Earth Syst. 13, e2020MS002413 https://doi.org/10.1029/2020MS002413.

Kilpatrick, K.A., Podestá, G., Walsh, S., Williams, E., Halliwell, V., Szczodrak, M., Brown, O.B., Minnett, P.J., Evans, R., 2015. A decade of sea surface temperature from MODIS. Remote Sens. Environ. 165, 27–41. https://doi.org/10.1016/j.rse.2015.04.023.

Kotsakis, A., Sullivan, J.T., Hanisco, T.F., Swap, R.J., Caicedo, V., Berkoff, T.A., Gronoff, G., Loughner, C.P., Ren, X., Luke, W.T., Kelley, P., Stratton, P.R., Delgado, R., Abuhassan, N., Shalaby, L., Santos, F.C., Dreessen, J., 2022. Sensitivity of total column NO2 at a marine site within the Chesapeake Bay during OWLETS-2. Atmos. Environ. 277, 119063 https://doi.org/10.1016/j.atmosenv.2022.119063.

Knowland, K.E., Keller, C.A., Wales, P.A., Wargan, K., Coy, L., Johnson, M.S., Liu, J., Lucchesi, R.A., Eastham, S.D., Fleming, E., Liang, Q., Leblanc, T., Livesey, N.J., Walker, K.A., Ott, L.E., Pawson, S., 2022. NASA GEOS composition forecast modeling system GEOS-CF v1.0: stratospheric composition. J. Adv. Model. Earth Syst. 14, e2021MS002852 https://doi.org/10.1029/2021MS002852.

Li, W., Wang, Y., Bernier, C., Estes, M., 2020. Identification of sea breeze recirculation and its effects on ozone in Houston, TX, during DISCOVER-AQ 2013. J. Geophys. Res. Atmos. 125, e2020JD033165 https://doi.org/10.1029/2020JD033165.

Liu, H., Chan, J.C.L., 2002. An investigation of air-pollutant patterns under sea–land breezes during a severe air-pollution episode in Hong Kong. Atmos. Environ. 36, 591–601. https://doi.org/10.1016/S1352-2310(01)00504-0.

Marais, E.A., Jacob, D.J., Kurosu, T.P., Chance, K., Murphy, J.G., Reeves, C., Mills, G., Casadio, S., Millet, D.B., Barkley, M.P., Paulot, F., Mao, J., 2012. Isoprene emissions in Africa inferred from OMI observations of formaldehyde columns. Atmos. Chem. Phys. 12, 6219–6235. https://doi.org/10.5194/acp-12-6219-2012.

McCabe, E.J., Freedman, J.M., 2023. Development of an objective methodology for identifying the seasea-breeze circulation and associated low-level jet in the New York Bight. Weather Forecast. 38, 571–589. https://doi.org/10.1175/WAF-D-22-0119.1.

Miller, S.T.K., Keim, B.D., Talbot, R.W., Mao, H., 2003. Sea breeze: structure, forecasting, and impacts. Rev. Geophys. 41 https://doi.org/10.1029/2003RG000124.

Millet, D.B., Jacob, D.J., Turquety, S., Hudman, R.C., Wu, S., Fried, A., Walega, J., Heikes, B.G., Blake, D.R., Singh, H.B., Anderson, B.E., Clarke, A.D., 2006. Formaldehyde distribution over North America: implications for satellite retrievals of formaldehyde columns and isoprene emission. J. Geophys. Res. Atmos. 111 https://doi.org/10.1029/2005JD006853.

Nowlan, C.R., González Abad, G., Kwon, H.-A., Ayazpour, Z., Chan Miller, C., Chance, K., Chong, H., Liu, X., O'Sullivan, E., Wang, H., Zhu, L., De Smedt, I., Jaross, G., Seftor, C., Sun, K., 2023. Global formaldehyde products from the ozone mapping and profiler suite (OMPS) nadir mappers on suomi NPP and NOAA-20. Earth Space Sci. 10, e2022EA002643 https://doi.org/10.1029/2022EA002643.

Orbe, C., Oman, L.D., Strahan, S.E., Waugh, D.W., Pawson, S., Takacs, L.L., Molod, A.M., 2017. Large-scale atmospheric transport in GEOS replay simulations. J. Adv. Model. Earth Syst. 9 (7), 2545–2560. https://doi.org/10.1002/2017MS001053.

Palmer, P.I., Abbot, D.S., Fu, T.-M., Jacob, D.J., Chance, K., Kurosu, T.P., Guenther, A., Wiedinmyer, C., Stanton, J.C., Pilling, M.J., Pressley, S.N., Lamb, B., Sumner, A.L., 2006. Quantifying the seasonal and interannual variability of North American isoprene emissions using satellite observations of the formaldehyde column. J. Geophys. Res. Atmos. 111 https://doi.org/10.1029/2005JD006689.

Palmer, P.I., Jacob, D.J., Chance, K., Martin, R.V., Spurr, R.J.D., Kurosu, T.P., Bey, I., Yantosca, R., Fiore, A., Li, Q., 2001. Air mass factor formulation for spectroscopic measurements from satellites: application to formaldehyde retrievals from the Global Ozone Monitoring Experiment. J. Geophys. Res. Atmos. 106, 14539–14550. https://doi.org/10.1029/2000JD900772.

Papanastasiou, D.K., Melas, D., Lissaridis, I., 2010. Study of wind field under sea breeze conditions; an application of WRF model. Atmospheric Res., Clouds, Aerosols and Radiation 98, 102–117. https://doi.org/10.1016/j.atmosres.2010.06.005.

Personal Communication with Xiong Liu at Smithsonian Astrophysical Observatory, 2022. Cambridge, MA.

Russell, A.R., Perring, A.E., Valin, L.C., Bucsela, E.J., Browne, E.C., Wooldridge, P.J., Cohen, R.C., 2011. A high spatial resolution retrieval of NO 2 column densities from OMI: method and evaluation. Atmos. Chem. Phys. 11, 8543–8554. https://doi.org/10.5194/acp-11-8543-2011.

Schaaf, C., Wang, Z., 2015. MCD43A1 MODIS/Terra+Aqua BRDF/albedo model parameters daily L3 global - 500m V006. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MCD43A1.006.

Smith, A., Lott, N., Vose, R., 2011. The integrated surface database: recent developments and partnerships. Bull. Am. Meteorol. Soc. 92, 704–708. https://doi.org/10.1175/2011BAMS3015.1.

Souri, A.H., Chance, K., Bak, J., Nowlan, C.R., González Abad, G., Jung, Y., Wong, D.C., Mao, J., Liu, X., 2021. Unraveling pathways of elevated ozone induced by the 2020 lockdown in Europe by an observationally constrained regional model using TROPOMI. Atmos. Chem. Phys. 21, 18227–18245. https://doi.org/10.5194/acp-21-18227-2021.

Souri, A.H., Choi, Y., Jeon, W., Li, X., Pan, S., Diao, L., Westenbarger, D.A., 2016. Constraining NOx emissions using satellite NO2 measurements during 2013 DISCOVER-AQ Texas campaign. Atmos. Environ. 131, 371–381. https://doi.org/10.1016/j.atmosenv.2016.02.020.

Souri, A.H., Choi, Y., Kodros, J.K., Jung, J., Shpund, J., Pierce, J.R., Lynn, B.H., Khain, A., Chance, K., 2020c. Response of Hurricane Harvey's rainfall to anthropogenic aerosols: A sensitivity study based on spectral bin microphysics with simulated aerosols. Atmos. Res. 242, 104965. https://doi.org/10.1016/j.atmosres.2020.104965.

Souri, A.H., Johnson, M.S., Wolfe, G.M., Crawford, J.H., Fried, A., Wisthaler, A., Brune, W.H., Blake, D.R., Weinheimer, A.J., Verhoelst, T., Compernolle, S., Pinardi, G., Vigouroux, C., Langerock, B., Choi, S., Lamsal, L., Zhu, L., Sun, S., Cohen, R.C., Min, K.-E., Cho, C., Philip, S., Liu, X., Chance, K., 2023. Characterization of errors in satellite-based HCHO/NO2 tropospheric column ratios with respect to chemistry, column-to-PBL translation, spatial representation, and retrieval uncertainties. Atmos. Chem. Phys. 23, 1963–1986. https://doi.org/10.5194/acp-23-1963-2023.

Souri, A.H., Nowlan, C.R., González Abad, G., Zhu, L., Blake, D.R., Fried, A., Weinheimer, A.J., Wisthaler, A., Woo, J.-H., Zhang, Q., Chan Miller, C.E., Liu, X., Chance, K., 2020a. An inversion of NOx and non-methane volatile organic compound (NMVOC) emissions using satellite observations during the KORUS-AQ campaign and implications for surface ozone over East Asia. Atmos. Chem. Phys. 20, 9837–9854. https://doi.org/10.5194/acp-20-9837-2020.

Souri, A.H., Nowlan, C.R., Wolfe, G.M., Lamsal, L.N., Chan Miller, C.E., Abad, G.G., Janz, S.J., Fried, A., Blake, D.R., Weinheimer, A.J., Diskin, G.S., Liu, X., Chance, K.,

2020b. Revisiting the effectiveness of HCHO/NO2 ratios for inferring ozone sensitivity to its precursors using high resolution airborne remote sensing observations in a high ozone episode during the KORUS-AQ campaign. Atmos. Environ. 224, 117341 https://doi.org/10.1016/j.atmosenv.2020.117341.

Spurr, R., 2008. LIDORT and VLIDORT: linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems. In: Kokhanovsky, A.A. (Ed.), Light Scattering Reviews 3: Light Scattering and Reflection. Springer Praxis Books. Springer, Berlin, Heidelberg, pp. 229–275. https://doi.org/10.1007/978-3-540-48546-9_7.

Spurr, R., Christi, M., 2019. The LIDORT and VLIDORT linearized scalar and vector discrete ordinate radiative transfer models: updates in the last 10 years. In: Kokhanovsky, A. (Ed.), Springer Series in Light Scattering: Volume 3: Radiative Transfer and Light Scattering, Springer Series in Light Scattering. Springer International Publishing, Cham, pp. 1–62. https://doi.org/10.1007/978-3-030-03445-0_1.

Spurr, R.J.D., 2006. VLIDORT: a linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media. J. Quant. Spectrosc. Radiat. Transf. 102, 316–342. https://doi.org/10.1016/j.jqsrt.2006.05.005.

Stauffer, R.M., Thompson, A.M., 2015. Bay breeze climatology at two sites along the Chesapeake bay from 1986–2010: implications for surface ozone. J. Atmos. Chem. 72, 355–372. https://doi.org/10.1007/s10874-013-9260-y.

Stavrakou, T., Müller, J.-F., De Smedt, I., Van Roozendael, M., van der Werf, G.R., Giglio, L., Guenther, A., 2009. Evaluating the performance of pyrogenic and biogenic emission inventories against one decade of space-based formaldehyde columns. Atmos. Chem. Phys. 9, 1037–1060. https://doi.org/10.5194/acp-9-1037-2009.

Sullivan, J.T., Apituley, A., Mettig, N., Kreher, K., Knowland, K.E., Allaart, M., Piters, A., Van Roozendael, M., Veefkind, P., Ziemke, J.R., Kramarova, N., Weber, M., Rozanov, A., Twigg, L., Sumnicht, G., McGee, T.J., 2022. Tropospheric and stratospheric ozone profiles during the 2019 TROpomi vaLIdation eXperiment (TROLIX-19). Atmos. Chem. Phys. 22 (17), 11137–11153. https://doi.org/10.5194/acp-22-11137-2022.

Tilstra, L.G., Tuinder, O.N.E., Wang, P., Stammes, P., 2017. Surface reflectivity climatologies from UV to NIR determined from Earth observations by GOME-2 and SCIAMACHY. J. Geophys. Res. Atmos. 122, 4084–4111. https://doi.org/10.1002/2016JD025940.

Travis, K.R., Judd, L.M., Crawford, J.H., Chen, G., Szykman, J., Whitehill, A., Valin, L.C., Spinei, E., Janz, S., Nowlan, C.R., Kwon, H.-A., Fried, A., Walega, J., 2022. Can column formaldehyde observations inform air quality monitoring strategies for ozone and related photochemical oxidants? J. Geophys. Res. Atmos. 127, e2022JD036638 https://doi.org/10.1029/2022JD036638.

Walton, C.C., Pichel, W.G., Sapper, J.F., May, D.A., 1998. The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites. J. Geophys. Res.: Oceans 103, 27999–28012. https://doi.org/10.1029/98JC02370.

Wan, Z., 2014. New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. Remote Sens. Environ. 140, 36–45. https://doi.org/10.1016/j.rse.2013.08.027.

Wang, Y., Di Sabatino, S., Martilli, A., Li, Y., Wong, M.S., Gutiérrez, E., Chan, P.W., 2017. Impact of land surface heterogeneity on urban heat island circulation and sea-land breeze circulation in Hong Kong. J. Geophys. Res. Atmos. 122, 4332–4352. https://doi.org/10.1002/2017JD026702.

Wolfe, G.M., Nicely, J.M., St Clair, J.M., Hanisco, T.F., Liao, J., Oman, L.D., Brune, W.B., Miller, D., Thames, A., González Abad, G., Ryerson, T.B., Thompson, C.R., Peischl, J., McKain, K., Sweeney, C., Wennberg, P.O., Kim, M., Crounse, J.D., Hall, S.R., Ullmann, K., Diskin, G., Bui, P., Chang, C., Dean-Day, J., 2019. Mapping hydroxyl variability throughout the global remote troposphere via synthesis of airborne and satellite formaldehyde observations. Proc. Natl. Acad. Sci. USA 116, 11171–11180. https://doi.org/10.1073/pnas.1821661116.

Zhu, L., González Abad, G., Nowlan, C.R., Chan Miller, C., Chance, K., Apel, E.C., DiGangi, J.P., Fried, A., Hanisco, T.F., Hornbrook, R.S., Hu, L., Kaiser, J., Keutsch, F. N., Permar, W., St Clair, J.M., Wolfe, G.M., 2020. Validation of satellite formaldehyde (HCHO) retrievals using observations from 12 aircraft campaigns. Atmos. Chem. Phys. 20, 12329–12345. https://doi.org/10.5194/acp-20-12329-2020.

Zhu, L., Jacob, D.J., Kim, P.S., Fisher, J.A., Yu, K., Travis, K.R., Mickley, L.J., Yantosca, R. M., Sulprizio, M.P., De Smedt, I., González Abad, G., Chance, K., Li, C., Ferrare, R., Fried, A., Hair, J.W., Hanisco, T.F., Richter, D., Jo Scarino, A., Walega, J., Weibring, P., Wolfe, G.M., 2016. Observing atmospheric formaldehyde (HCHO) from space: validation and intercomparison of six retrievals from four satellites (OMI, GOME2A, GOME2B, OMPS) with SEAC4RS aircraft observations over the southeast US. Atmos. Chem. Phys. 16, 13477–13490. https://doi.org/10.5194/acp-16-13477-2016.

Zhu, Q., Laughner, J.L., Cohen, R.C., 2022. Combining machine learning and satellite observations to predict spatial and temporal variation of near surface OH in North American cities. Environ. Sci. Technol. 56, 7362–7371. https://doi.org/10.1021/acs.est.1c05636.

Zoogman, P., Liu, X., Chance, K., Sun, Q., Schaaf, C., Mahr, T., Wagner, T., 2016. A climatology of visible surface reflectance spectra. J. Quant. Spectrosc. Radiat. Transf. 180, 39–46. https://doi.org/10.1016/j.jqsrt.2016.04.003.