Characterizing Global Ozonesonde Profile Variability from Surface to the UT/LS with a Clustering Technique and MERRA-2 Reanalysis

Ryan M. Stauffer[1], Anne M. Thompson[2], Jacquelyn C. Witte[3]

[1]Universities Space Research Association c/o Atmospheric Chemistry and Dynamics Lab, NASA/Goddard Space Flight Center, Greenbelt, Maryland, USA

[2]NASA/Goddard Space Flight Center, Greenbelt, Maryland, USA

[3]Science Systems and Applications, Inc., Lanham, Maryland, USA

Corresponding author: Ryan M. Stauffer (ryan.m.stauffer@nasa.gov)

Key Points:

- Analysis of self-organizing map cluster-based O_3 allows straightforward geophysical characterization of O_3 profiles based on profile shape

- UT/LS O_3 amount and geopotential height dominate variability among mid-latitude O_3 clusters. This relationship ceases within $\pm 30°$ latitude

- Tropical O_3 profile cluster shapes correspond to carbon monoxide amount, and convection indicated by velocity potential and the MJO

Abstract

Our previous studies employing the self-organizing map (SOM) clustering technique to ozonesonde data have found significant links among meteorological and chemical regimes, and the shape of the ozone (O_3) profile from the troposphere to the lower stratosphere. These studies, which focused on specific northern hemisphere mid-latitude geographical regions, demonstrated the advantages of SOM clustering by quantifying O_3 profile variability and the O_3/meteorological correspondence. We expand SOM to a global set of ozonesonde profiles spanning 1980-present from 30 sites to summarize the connections among O_3 profiles, meteorology, and chemistry, using the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) reanalysis and other ancillary data. Four clusters of O_3 mixing ratio profiles from the surface to the upper troposphere/lower stratosphere (UT/LS) are generated for each site, which show dominant profile shapes and typical seasonality (or lack thereof) that generally correspond to latitude (i.e. Tropical, Subtropical, Mid-Latitude, Polar). Examination of MERRA-2 output reveals a clear relationship among SOM clusters and covarying meteorological fields (geopotential height, potential vorticity, and tropopause height) for Polar and Mid-latitude sites. However, these relationships break down within ±30° latitude. Carbon monoxide satellite data, along with velocity potential, a proxy for convection, calculated from MERRA-2 wind fields assist characterization of the Tropical and Subtropical sites, where biomass burning and convective transport linked to the Madden-Julian Oscillation (MJO) dominate O_3 variability. In addition to geophysical characterization of O_3 profile variability, these results can be used to evaluate chemical transport model output and satellite measurements of O_3.

1 Introduction

1.1 Motivation: Geophysical Characterization of Ozonesonde Profiles

Ozonesondes provide high vertical resolution (~150 m) O_3 measurements from the surface to the mid-stratosphere with excellent precision and accuracy (Witte et al., 2017 and refs therein); several records from around the globe span multiple decades. They provide an independent source of validation for O_3 measurements from satellites (Bian et al., 2007; Boynard et al., 2009; Thompson et al., 2003a; 2012; 2017; Ziemke et al., 2011), output from chemical transport/climate models (CTM, CCM; Considine et al., 2008; Stone et al., 2016; Tilmes et al., 2016), and assimilated O_3 (Dethof and Holm, 2004; Wargan et al., 2016). Because of their high vertical resolution, ozonesondes are able to capture fine details in the geophysical features that control the variability in O_3. The characterization of meteorological and chemical regimes and their effect on O_3 aids the interpretation of various satellite and model products. On short timescales, frequent campaign-based sonde profiling has validated high-resolution model output of transport processes from Asia to the US (Lin et al., 2012a), stratosphere-to-troposphere exchange (STE; Lin et al., 2012b; Ott et al., 2016), and convective activity (Thompson et al., 2008). On greater timescales, long-term ozonesonde networks characterize seasonal and inter-annual impacts of STE (Lin et al., 2015), biomass burning, the El Niño Southern Oscillation (ENSO; Chandra et al., 2002), and monsoon convection (Yonemura et al., 2002; Thompson et al., 2011) on the tropospheric O_3 budget.

The thousands of sonde profiles from long-term stations help form O_3 climatologies from which seasonal variability in the O_3 profile can be quantified. Climatologies are typically derived from averaged O_3 profiles on a latitudinal and monthly or seasonal basis (Logan, 1999; McPeters et al., 1997; McPeters and Labow, 2012; Sofieva et al., 2014; Labow

et al., 2015). However, Stauffer et al. (2016; 2017) showed that analysis of averaged O_3 profile data conceals information on geophysical controls of short-term O_3 profile variability in the troposphere to lower stratosphere. Averaged O_3 climatologies by convention cannot fully capture variability on timescales shorter than months or seasons, or from climate oscillations like ENSO, the Madden-Julian Oscillation (MJO), or the Quasi-Biennial Oscillation (QBO). Ozone averages may also inhibit potential diagnoses for disagreement between sonde measurements, satellite products (see Stauffer et al., 2017 Figure A1), and model output.

1.2 Self-Organizing Map (SOM) Clustering of O_3 Profiles

One way to characterize short-term variability from large O_3 profile data sets in a geophysically meaningful way is to cluster the O_3 data into groups with similar profile shapes, disregarding information on profile month or season. The benefits of clustering large sets of ozonesonde profiles with SOM are made clear by several recent studies examining varying regions and O_3 regimes. Jensen et al. (2012) applied SOM to O_3 profiles at two tropical Atlantic sites, Ascension Island and Natal, Brazil. Clusters of O_3 profiles at the two sites represented a set of "typical" O_3 profile shapes, a set of convectively-influenced low O_3 profiles, and two clusters with O_3 enhancements from African biomass burning. Prototypical meteorological conditions and air parcel trajectories were linked to the dominant O_3 profile shapes, which were occasionally, but not always, seasonally-dependent. Stauffer et al. (2016) continued in a similar manner by analyzing sonde profiles from four US sites from the surface to 12 km. They found that the O_3 profile clusters depended more on synoptic-scale meteorology than seasonality. The O_3 profile clusters diverged from monthly O_3 climatologies by over ± 100 ppbv in the UT/LS at the US sites for several O_3 profile clusters.

In a companion paper, Stauffer et al. (2017) characterized SOM O_3 profile clusters at Trinidad Head, CA, and their links to meteorology, chemistry (namely carbon monoxide; CO), and surface O_3 pollution at inland CA sites. As in Jensen et al. (2012), clusters of O_3 profiles over the US were much more closely linked to transport and large-scale meteorological and chemical conditions than to season. Gallardo et al. (2016) used SOM to generate O_3 profile clusters at Rapa Nui (Easter Island) in the subtropical South Pacific. Rapa Nui's latitude near the meandering subtropical jet leads to highly variable O_3 profiles that show characteristics of both the mid-latitudes (tropopause height variations) and tropics (convectively-influenced). As a result, seasonal climatologies are poor representations of week-to-week changes in O_3 at Rapa Nui.

These previous applications of SOM clustering of O_3 profiles over limited geographical regions motivate us to expand our analyses to an assortment of prototypical global ozonesonde sites. We address two important science questions in this study: 1) How do O_3 profiles cluster for various regions, each of which are known to exhibit differing O_3 distributions? 2) What are the links among the O_3 profile clusters, meteorology, and chemistry, and how do they depend on latitude or region (e.g. mid-latitudes versus tropics)? Given that past SOM analyses of sonde O_3 profiles have yielded significant links to meteorological and chemical signatures, we expect lessons from this study to apply to future chemical model validation and diagnostics.

2 Data and Clustering Methods

2.1 Global Ozonesonde Data Set

The 30 ozonesonde stations chosen for this analysis are shown in Figure 1 and are detailed in Tables 1 and 2. In general, sites were selected based on their use in previous sonde-based climatologies (e.g. Logan, 1999; McPeters and Labow, 2012; Tilmes et al., 2012). Subtropical and Tropical stations were chosen for their sufficient number of records, requiring that the selected stations capture the longitudinal distribution of the tropospheric O_3 zonal wave-one feature over all seasons (Hudson and Thompson, 1998; Thompson et al., 2003b; Sauvage et al., 2006). For compatibility with the MERRA-2 output, we analyze only ozonesonde data from 1980 onward. Every site has more than 200 profiles, with the majority of stations contributing >1000 profiles (>46,000 samples in total). The locations are categorized by latitude: Polar ($\phi > \pm60°$), Mid-Latitude ($\phi = \pm30\text{-}60°$), Subtropical ($\phi = \pm15\text{-}30°$), and Tropical ($\phi = 0\pm15°$).

Ozonesonde data are obtained from the World Ozone and Ultraviolet Radiation Data Centre (WOUDC, WMO/GAW UV Radiation Monitoring Community, http://www.woudc.org) and the Southern Hemisphere Additional Ozonesondes (SHADOZ; Thompson et al., 2003a) archive (http://tropo.gsfc.nasa.gov/shadoz). Data from Boulder and Hilo are obtained from the NOAA ESRL GMD archive. Refer to Table 2 for a list of supporting institutions. Ozone profile data from Uccle and De Bilt, and the SHADOZ stations (except Samoa) have been reprocessed to reflect the latest guidelines for homogenization of historical ozonesonde data (Smit et al., 2012; Smit et al., 2014; Van Malderen et al., 2016; Deshler et al., 2017; Thompson et al., 2017; Witte et al., 2017). Reprocessed data from Uccle and De Bilt can be obtained directly from the station data managers (see Acknowledgments). Table 1 lists the various types of ozonesondes used in the historical data sets. The differing instrument types are not expected to influence our results. Sonde chamber tests (Smit et al., 2012) and evaluation of reprocessed data (Witte et al., 2017) show that instrument artifacts are greatest above the UT/LS.

2.2 MERRA-2 Reanalyses

A suite of variables (e.g. geopotential height, temperature, potential vorticity, relative humidity, and U and V wind) on pressure levels (925, 850, 700, 500, 250 hPa; 200 hPa for U and V winds) from the MERRA-2 reanalysis (GMAO, 2015; Bosilovich et al., 2015) provide meteorological context for the O_3 profile data. The instantaneous, assimilated meteorological data are available every three hours on a 0.5° x 0.625° horizontal grid from 1980-present. Statistics will be presented for MERRA-2 grid points closest in space (time) to each ozonesonde site (profile).

2.3 Satellite CO Data

Stauffer et al. (2017) showed correspondence between Atmospheric Infrared Sounder (AIRS; Aumann et al., 2003) CO data and O_3 SOM clusters from Trinidad Head, CA, distinguishing pollution and STE effects on the O_3 profiles. Therefore, we employ the daily L3 V6 (gridded to 1° x 1° horizontal resolution) CO total column data to help differentiate chemically- versus meteorologically-driven O_3, especially in the tropics where biomass burning pollution has the greatest impact. AIRS CO data is available from September 2002 – present; we restrict sonde and AIRS CO comparisons to these years.

2.4 SOM Specifics

The SOM (Kohonen, 1995) method carried out here is similar to that described in Stauffer et al. (2016; 2017). Briefly, the user determines the N number of clusters desired

and the dimensions of the SOM, the number of iterations of the SOM algorithm, and the

altitude range of data to cluster, among other inputs (see Appendix A of Stauffer et al., 2016

for examples). For each site, the entire 100 m averaged O_3 mixing ratio profile data set is

input into the SOM algorithm to obtain the initial nodes (analogous to cluster centroids) via a

linear interpolation between the two largest principal components of the O_3 mixing ratio set.

This interpolation ensures that the initial nodes, which are shaped like O_3 profiles, cover the

large variability of the whole O_3 profile set. With initial nodes generated, the entire O_3

profile set is continuously input into the SOM equation for the number of user-determined

iterations. Individual O_3 profiles are assigned to their best-matching node for each iteration

based on similarity in shape (minimum Euclidean distance), and the nodes are recalculated as

the mean of their member profiles. As the iterations approach the final user-chosen amount,

fewer and fewer profiles are reassigned nodes, until there are no cluster membership changes.

The result is N number of O_3 profile clusters, where the SOM cluster nodes are equivalent to

the mean of each O_3 profile cluster. For further details on SOM, as well as quantitative

comparisons to the similar k-means algorithm that summarize our preference for SOM, refer

to Stauffer et al. (2017) and Appendix A of Stauffer et al. (2016).

Major differences between this and the previous Stauffer et al. studies analyzing SOM

for O_3 profiles are as follows: 1) We optimized our clustering by using a 2x2 SOM with 4

clusters to avoid clusters with too few or too many members, and to keep analysis of clusters

from 30 sites from becoming too cumbersome. When using SOM with 9 clusters as in

Stauffer et al. (2016; 2017), clusters at two of our Tropical sites contained a dozen or less

profiles, too small for meaningful analysis. 2) SOM clusters were calculated using O_3 mixing

ratio profiles from the surface to the average annual tropopause altitude (nearest km) at each

of the sites. This variable altitude was chosen to capture day-to-day and seasonal changes in

the tropopause height at higher latitude sites, as well as UT/LS O_3 dynamics at Tropical and

Subtropical sites. 3) Once the SOM was generated for each site, we numbered the clusters from 1 to 4 based on average O_3 amount in the cluster. This was done to simplify analysis so that cluster 1 is consistently defined as "low O_3" for each site, up to cluster 4, which contains the highest O_3 for each site.

3 Ozone Variability and Seasonality of Profile Clusters

An example of the four-cluster SOM output from representative sites in each of the latitude groups is shown in Figure 2. This demonstrates how O_3 profiles in the clusters compare against the overall median profile from the sites (cyan on every graph) and O_3 variability within the clusters. Membership statistics (profile numbers) are shown for each cluster. In general, profile shapes/cluster means from sites within each latitude region are similar, particularly outside of the tropics. Clusters from sites at high latitudes (Ny-Ålesund example in Figure 2a) are determined entirely by UT/LS O_3 variability/tropopause height, with a progressively lower tropopause altitude and dwindling membership as clusters progress from 1 to 4. Note that the altitude ranges for each latitude group vary but are low enough to avoid seasonal stratospheric O_3 depletion events near the poles. Profiles and cluster membership at Mid-Latitude sites (Wallops Island example in Figure 2b) resemble higher latitude sites (note different y-axis scales), with the exception of increased tropospheric O_3 variability and the emergence of pollution effects in cluster 2 at Wallops Island.

Clusters from Subtropical (Irene example in Figure 2c) and Tropical sites (Samoa example in Figure 2d) are much more sensitive to tropospheric O_3 amounts than are higher latitudes. At Irene, only one cluster exists with a notably lower tropopause (cluster 3), containing just 8% of all profiles. Profile clusters at lower latitudes can be defined largely by

tropospheric pollution enhancements; the difference between clusters 1 and 4 at Irene is a ~25 ppbv O_3 increase below 10 km. SOM O_3 clusters from Samoa in Figure 2d show how UT/LS O_3 is linked to lower tropospheric O_3 in the tropical Pacific. Profiles in cluster 1 contain low O_3 throughout the entire surface to 17 km altitude range, whereas clusters 3 and 4 display higher O_3 at all altitudes up to the UT/LS. Cluster 2 contains O_3 that represents the overall mean profile from Samoa. We find that clusters 2 or 3 typically contain O_3 that mirrors the overall average profile for most sites, though they rarely combine to comprise a majority of profiles for a given site (see Table 3).

3.1 Ozone Profile Cluster Types

The O_3 profile SOM cluster means (nodes) from all 30 sites are presented in Figure 3, and depict the transition of O_3 profiles from low to high latitudes (blue = tropics to red = high latitudes on Figure 3). Compare this figure to Table 3, which contains cluster statistics from all 30 sites. Figure 3 is a convenient way to visualize the dominant O_3 profile types and the variability in O_3 with latitude, and longitude in the case of the tropics. We observe the tropical wave-one pattern (Thompson et al., 2003b; 2017) in low- to mid-tropospheric O_3 from the tropical Pacific (low tropospheric O_3 amounts at Watukosek, Kuala Lumpur, and Samoa) to the tropical Atlantic (high tropospheric O_3 amounts at Natal and Ascension Island). The invariance of the tropopause altitude at low latitudes from clusters 1 to 4 is also evident. This contrasts with higher latitude sites, whose O_3 values extend beyond the x-axes in clusters 2, 3, and 4 in Figure 3. The "S-shape" in tropical tropospheric O_3 arising from convective transport is especially evident at sites with the lowest upper tropospheric O_3 like Kuala Lumpur and Watukosek. In terms of vertical structure in the O_3 profiles, SOM clusters show the covariance of near-surface and UT/LS O_3 in the tropics and subtropics. Low near-

surface O_3 tends to occur in conjunction with low UT/LS O_3, and vice versa. This feature and its links to geophysical processes are explored in Section 4.3. In this paper, we focus mostly on the vertical structure of O_3, but we note the role that horizontal transport may play in the O_3 profile clusters. Stauffer et al., (2017) demonstrated that SOM clusters can distinguish STE and horizontal pollution transport. That study focused only on surface to 6 km O_3, making it possible to identify O_3 laminae indicative of horizontal transport processes.

One consistency among sites from all latitudes is that the plurality of O_3 profiles is nearly always contained in the lowest O_3 cluster 1 (see Table 3). There are only 4 exceptions, all from sites in the Subtropics and Tropics (Naha, Watukosek, Ascension, and Irene). Additionally, the highest O_3 cluster 4 contains the lowest percentage of profiles except at Natal, Réunion, and Irene.

3.2 Seasonality of O_3 Profile Clusters

Our prior SOM cluster analyses (Jensen et al., 2012; Stauffer et al., 2016; 2017) indicate that clustering profiles is not identical to seasonal categorization, demonstrating the value in highlighting geophysical rather than seasonal variability. The histograms of profile months for each site are broken down by latitude group in Figure 4. Cluster 4 profiles display a clear seasonal cycle for all latitude groups, signaling that there is a preferred time of year for the highest O_3 amounts, whether the high O_3 is located in the UT/LS (higher latitudes) or in the troposphere (tropics). Sites from the subtropics to polar latitudes all observe the highest SOM cluster O_3 in their respective spring seasons, when tropopause altitudes are volatile. The Tropical sites (Figure 4a, cluster 4) also tend to display the highest O_3 in their respective spring seasons (cf Thompson et al., 2003a; 2012). Kuala Lumpur, just 2.7° north of the equator, shows a strong preference for higher O_3 in April through June as the ITCZ

moves northward, while the four Southern Hemisphere Tropical stations exhibit peaks in September to November (Austral spring). Histograms from the other clusters at most sites are flatter than cluster 4, indicating that those profiles occur scattered throughout the year. Exceptions are noted for certain Tropical sites in clusters 1 and 2, which show coherence likely related to convectively active seasons and low UT/LS O_3. The large percentage of low O_3 cluster 1 profiles that occur in January at Hanoi is also striking (Figure 4b). Coherent seasonality is also displayed by Subtropical and Mid-Latitude cluster 3 profiles, which mimic the cluster 4 seasonality, but with histogram peaks spread over a larger number of months. It is difficult to draw conclusions about seasonality for most non-tropical site profiles from clusters 1 and 2, as those clusters contain a large number of profiles that occur throughout the year.

4 Profile Links to Meteorological and Chemical Data

4.1 MERRA-2 Meteorological Reanalyses

We analyze meteorological fields from MERRA-2 at periods coincident with O_3 profiles in the SOM clusters to characterize the large-scale dynamic features associated with the profile shapes. Stauffer et al. (2016) showed strong links between 500 hPa geopotential height (Z) and O_3 profile clusters at four US sites. Figure 5 shows 500 hPa Z composite means and anomalies (from MERRA-2 1981-2010 daily climatology) corresponding to the O_3 SOM clusters at four example sites, one from each latitude group. At Polar and Mid-Latitude sites (Marambio and De Bilt examples on Figure 5a, b), a consistent pattern emerges relating synoptic-scale meteorology and the O_3 profile shape. Low O_3 in cluster 1 is associated with a mid-level ridge and above-average tropopause height (not shown). Cluster

2 (the "nominal" O_3 cluster) shows no significant anomalies, and clusters 3 and 4 each show increasingly significant trough features and a well-below-average tropopause height. This is true for all Polar and Mid-Latitude sites, and is best seen in Figure 6, which shows the distributions of 500 hPa Z anomalies for the MERRA-2 grid point closest to each site organized by latitude. The magnitude of the Z anomalies tends to increase with latitude up to about ±60°, with average Z anomalies of < -120 m in cluster 4 for many Mid-Latitude sites. Other meteorological variables that tend to co-vary with Z show similar distributions (i.e. temperature, relative humidity, and potential vorticity).

There are no distinct large-scale patterns in Z (or tropopause altitude variations) for locations within 30° of the equator (Réunion and Ascension Island in Figure 5). The distributions of 500 hPa Z anomalies in Figure 6 are centered about zero for all 10 sites in the Subtropical and Tropical regions. The meteorological dynamics that govern variability in the tropics and subtropics are entirely different from higher latitudes where temperature gradients are strong, so we look to other ancillary data to characterize the links between geophysical processes and O_3 profiles.

4.2 Convection and the MJO

MERRA-2 U and V were used to separate the wind fields into divergent (velocity potential; VP) and non-divergent (stream function) components. Stream function contours are nearly parallel to geopotential height contours in the mid-latitudes, but the VP contribution to the large-scale flow grows in the tropics, where the wind tends to deviate from geostrophic balance. As with the geopotential heights in Figure 6, stream function is a poor predictor of SOM O_3 cluster in the subtropics and tropics. An exception occurs at Irene, with

large stream function anomalies noted for cluster 3 profiles (Figure S1). This type of feature was not apparent in the geopotential height patterns (cf Figure 6).

Velocity potential, which describes tropical convective activity and is related to MJO evolution (Ventrice et al., 2013), is closely linked to SOM cluster number (Figure 7). Previous studies examining MJO phase and O_3 profiles/columns have noted the MJO association with large anomalies in subtropical lower stratospheric O_3 (Tian et al., 2007; Li et al., 2012) and in tropospheric O_3 in the tropical Pacific (Ziemke and Chandra, 2003). The MJO can modulate the total O_3 column (trough-to-peak changes) by about 5-10 DU. We show examples of MERRA-2 calculated 200 hPa VP (VP200) means and anomalies (from 1981-2010 daily climatology) from four sites (Hanoi, Kuala Lumpur, Natal, and Samoa; Figure 7). Negative VP anomalies (green colors) are associated with enhanced convection on regional scales, and are linked to cluster 1 low O_3 profiles. Convection lifts O_3-poor air from the tropical boundary layer to the UT/LS, leading to the classic tropical "S-shape" O_3 profiles in cluster 1. Positive VP anomalies indicate suppressed convection, and are linked to the higher O_3 profiles from clusters 3 (not shown on Figure 7) and 4.

Figure 8 shows the VP anomaly mean and 95% confidence intervals of the mean for the four O_3 profile clusters at all 30 sites. Confidence intervals were calculated using 10,000 bootstrap resamples of the VP anomaly distribution. The VP anomalies diverge farther from zero for sites near the equator, with low O_3 clusters generally associated with negative VP anomalies (enhanced convection), and higher O_3 associated with positive VP anomalies (suppressed convection). Note that the largest VP anomalies often do not occur directly over the site locations as in geopotential height for higher latitudes (cf Figures 5 and 7). Although correspondence between VP and SOM cluster exists for all Subtropical and Tropical sites to some degree (including some mid-latitude sites like Tateno and Kagoshima in Japan), there is a sharp difference in VP magnitude among West Pacific and other sites. Velocity potential

anomalies at sites in the West Pacific are much greater than at locations such as Hilo, Natal

(Figure 7c), Ascension Island, and Samoa, which are each closer to the equator than Hanoi

but in the opposite Western Hemisphere, where VP anomalies are typically smaller. We find

that the VP/SOM cluster linkages correspond to total O_3 column anomalies spanning ~15 DU

from profiles in lowest O_3 clusters to the highest, larger than the trough-to-peak changes that

previous studies have attributed to the MJO. Specifically, at the five Tropical sites, total

column O_3 in cluster 1 profiles (large negative VP anomalies) average -6.6 DU below

climatology, while cluster 4 profiles (large positive VP anomalies) average +8.5 DU above

climatology. This suggests that VP provides an excellent characterization of the dependence

of tropical O_3 on convection and the MJO.

　　　We explore evidence for SOM O_3 profile cluster and MJO linkages further with an

explicit MJO index, the Real-time Multivariate MJO Index (RMM; Wheeler and Hendon,

2004). The RMM Index is formed via a principal component analysis of near-equatorially

averaged 850 hPa U wind, 200 hPa U wind, and interpolated satellite outgoing longwave

radiation data. RMM1 and RMM2, based on the two largest principal component time series

calculated from those meteorological variables, form the RMM Index that can be separated

into eight MJO phases, which indicates regions of the tropics where convection is most

active. Daily RMM values are obtained from the Australian Bureau of Meteorology and are

available from June 1974 onward

(http://www.bom.gov.au/climate/mjo/graphics/rmm.74toRealtime.txt). We plot the average

RMM1 and RMM2 values corresponding to each O_3 profile cluster for Hanoi (9a), Kuala

Lumpur (9b), Natal (9c), and Watukosek (9d), the same sites presented in Figure 7.

　　　Figure 9 shows similar patterns in RMM Index for the three Pacific sites (all but Natal

in Figure 9). At the Pacific sites, cluster number 1 (lowest O_3 amounts) tends to coincide

with MJO phases 4 and 5, when enhanced convection is expected in the Tropical West

Pacific/Indonesia. Cluster number 4 (highest O_3 amounts) tends to be associated with MJO

phases 1 and 8, when convection is most active in the Western Hemisphere, and is suppressed

in the Pacific. The different MJO phase space plot at Natal (Figure 9c) reflects its tropical

Atlantic location, with low O_3 (cluster 1) associated with MJO phases 1 and 8, and high O_3

(cluster 4) linked to enhanced convection in the opposite hemisphere, but suppressed

convection locally. We note that the RMM Index values shown in Figure 9 do not amount to

"strong" MJO events ($RMM1^2 + RMM2^2 \geq 1$; Wheeler and Hendon, 2004) on average.

However, we have computed the 95% confidence intervals on the average values shown

(error bars in Figure 9), again using 10,000 bootstrap resamples. At all four sites in Figure 9,

the average RMM Index for clusters 1 and 4 are significantly different from each other,

meaning that our linking of the MJO to SOM O_3 profile clusters is statistically rigorous. The

significant differences in average RMM between clusters 1 and 4 also exist at Ascension and

Samoa (6 of 7 sites between 21°N and 14°S except Hilo).

4.3 Atmospheric Infrared Sounder (AIRS) CO

To add to the meteorological context provided above, we look for chemical clues to

help describe the varying tropospheric O_3 amounts at the Subtropical and Tropical locations.

Carbon monoxide has a lifetime of several months and can be used as a tracer for biomass

burning, which has a strong influence on many of the Tropical ozonesonde sites used in this

study (e.g. Jensen et al., 2012; Thompson et al., 2012). AIRS CO data are incorporated and

analyzed in a way similar to the MERRA-2 data. Note that AIRS data only cover the period

back to 2002. Figures 10-12 and Figure S2 show total CO column composite means and

anomalies (from 2002-2016 monthly climatology) corresponding to O_3 SOM cluster at four

example sites: Kuala Lumpur (Figure 10), Watukosek (Figure 11), Irene (Figure 12), and

Sapporo (Figure S2). At Subtropical and Tropical locations, there is congruence between CO and SOM clusters with CO column anomalies averaging as high as $+4.5 \times 10^{17}$ molec./cm^2 (~25% above the mean) upwind of Irene in cluster 3 (red contours on Figure 12). In the case of Irene cluster 3, it appears a sharp gradient in CO is observed near the location of the subtropical jet/trough feature (cf Figure S1), with Irene affected by a lower tropopause altitude (see Figure 2c) and lower CO (blue contours) than what is observed south of the site. Similar large positive CO anomalies are observed in conjunction with higher tropospheric O_3 at Kuala Lumpur in clusters 3 and 4, and Watukosek cluster 4. Negative CO anomalies, though generally smaller in magnitude than positive anomalies, accompany the low O_3 clusters 1 and 2 at the low-latitude sites. The example from Sapporo (Figure S2) shows an example of how the correspondence between CO and O_3 cluster commonly breaks down at higher latitudes. This is due primarily to the dominance of the Z and tropopause height variation over other signals in the O_3 clusters.

The high regional CO amounts and seasonality of cluster 3 at Irene show evidence of typical biomass burning conditions throughout Central Africa (i.e. July-October; Figure 4b) but the seasonality information for Kuala Lumpur and Watukosek is less clear for the high CO clusters 3 (Kuala Lumpur: All months but June-September; Figure 4a) and 4 (Watukosek: May-July, October-November). This is because Watukosek and Kuala Lumpur O_3 responds to more sporadic biomass burning fires (CO anomalies; Thompson et al., 2001; Yonemura et al., 2002) and convective transport (VP anomalies) or the MJO cycle as we demonstrated here. ENSO, fire activity, and tropospheric O_3 amounts also tend to covary at some of the stations examined here (Chandra et al., 2002), though a full treatment of ENSO and related analyses are beyond the current study scope. At Irene, Reunion, Ascension, Natal, and Samoa the biomass burning enhancement of O_3 follows a more regular seasonal cycle with

most significant effects in Austral winter and spring (Oltmans et al., 2001; Jensen et al., 2012; Thompson et al., 2012).

The tendency for the MJO and biomass burning to covary in the tropical Pacific was described in Reid et al., (2012), who found that a minimum in fire counts and a maximum in precipitation on the Maritime Continent (near Watukosek and Kuala Lumpur) occurred during MJO Phase 4. This explains the cluster 1 (low O_3) association with MJO Phases 4 and 5, negative VP200 anomalies, and low CO at the Tropical Pacific sites. Over the course of an entire burning season, ENSO is the greatest indicator for total biomass burning activity. However, sub-seasonally, the MJO is the greatest predictor of the timing of biomass burning (Reid et al., 2012). We also find that precipitation (indicated by the MJO) and biomass burning (indicated by CO) follow a similar pattern, and show its link to O_3 profile shape and O_3 amounts in the tropical Pacific. This effect also leads to the covariance in near-surface and UT/LS O_3 amounts noted in Figure 3, as low O_3 air is convectively lofted during periods of low burning activity, and high O_3 accumulates during periods of greater burning activity and suppressed convection.

5 Conclusions

We used a combination of SOM statistical clustering, MERRA-2 meteorological reanalyses, a daily MJO index, and AIRS CO satellite data to characterize global ozonesonde records for 30 sites from 1980-present. Our results reveal the major geophysical contributors to O_3 profile variability at sites ranging from 69°S to 82°N, including ten sites within subtropical and tropical latitudes. Four clusters of O_3 profiles were generated independently for each site with SOM and arranged according to O_3 amount.

Descriptions of SOM cluster O_3 based on latitude (Polar, Mid-Latitude, Subtropical, and Tropical) can be summarized as follows: 1) Polar and Mid-Latitude O_3 profile clusters depend mostly on large-scale atmospheric dynamics diagnosed by geopotential (and tropopause) height. The SOM algorithm was run on O_3 data from the surface to the average tropopause height for each site. Thus, UT/LS O_3 variability drives the clustering at Polar and Mid-Latitude sites. 2) SOM clustering at Subtropical and Tropical locations within 30° of the equator showed strong dependence on mid-tropospheric O_3 variation, with less volatility in UT/LS O_3. Horizontal temperature and tropopause altitude gradients are small at these latitudes, so geopotential height was a poor indicator of SOM cluster. Velocity potential calculated from MERRA-2 wind fields at the low-latitude stations and an explicit MJO index (RMM) showed strong links between VP/convection and SOM clusters, especially at West Pacific sites (Hanoi, Kuala Lumpur, Watukosek). Satellite-based carbon monoxide data from AIRS also showed correspondence among SOM O_3 clusters and column CO at the Subtropical and Tropical sites, where biomass burning has significant effects on O_3 profiles. The AIRS CO and VP/MJO analyses indicate how convection and biomass burning vary together and drive tropospheric O_3 variability in the tropics, with suppressed convection and high CO leading to high amounts of tropospheric O_3, and vice versa.

Acknowledgments

References

Aumann, H. H., M. T. Chahine, C. Gautier, M. Goldberg, E. Kalnay, L. McMillin, H. Revercomb, P. W. Rosenkranz, W. L. Smith, D. Staelin, L. Strow, and J. Susskind (2003), AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products and processing systems, *IEEE Trans. Geosci. Remote*, 41, 253–264, doi:10.1109/TGRS.2002.808356.

Bian, J., A. Gettelman, H. Chen, and L. L. Pan (2007), Validation of satellite ozone profile retrievals using Beijing ozonesonde data, *J. Geophys. Res.*, 112, D06305, doi:10.1029/2006JD007502.

Bosilovich, M., et al. (2015), MERRA-2: Initial evaluation of the climate, NASA Tech. Rep. Series on Global Modelling and Data Assimilation NASA/TM–2015-104606, Vol. 43, 145 pp. [Available online at https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich803.pdf.].

Boynard, A., C. Clerbaux, P.-F. Coheur, D. Hurtmans, S. Turquety, M. George, J. Hadji-Lazaro, C. Keim, and J. Meyer-Arnek (2009), Measurements of total and tropospheric ozone from IASI: comparison with correlative satellite, ground-based and ozonesonde observations, *Atmos. Chem. Phys.*, 9, 6255-6271, https://doi.org/10.5194/acp-9-6255-2009.

Chandra, S., J. R. Ziemke, P. K. Bhartia, and R. V. Martin (2002), Tropical tropospheric ozone: Implications for dynamics and biomass burning, *J. Geophys. Res.*, 107(D14), doi:10.1029/2001JD000447.

Considine, D. B., J. A. Logan, and M. A. Olsen (2008), Evaluation of near-tropopause ozone distributions in the Global Modeling Initiative combined stratosphere/troposphere model with ozonesonde data, *Atmos. Chem. Phys.*, 8, 2365-2385, doi:10.5194/acp-8-2365-2008.

Deshler, T., R. Stübi, F. J. Schmidlin, J. L. Mercer, H. G. J. Smit, B. J. Johnson, R. Kivi, and B. Nardi (2017), Methods to homogenize electrochemical concentration cell (ECC)

ozonesonde measurements across changes in sensing solution concentration or

ozonesonde manufacturer, *Atmos. Meas. Tech.*, 10, 2021-2043,

https://doi.org/10.5194/amt-10-2021-2017.

Dethof, A., and E. V. Holm (2004), Ozone assimilation in the ERA-40 reanalysis project,

Q.J.R. Meteorol. Soc., 130, 2851-2872, doi:10.1256/qj.03.196.

Gallardo, L., A. Henríquez, A. M. Thompson, R. Rondanelli, J. Carrasco, A. Orfanoz-

Cheuquelaf, and P. Velásquez (2016), The first twenty years (1994–2014) of ozone

soundings from Rapa Nui (27°S, 109°W, 51 m a.s.l.), *Tellus B Chem. Phys. Meteorol.*,

68:1, 29484, doi:10.3402/tellusb.v68.29484.

GMAO (2015), MERRA-2 inst3_3d_asm_Nv: 3d, 3-Hourly, Instantaneous, Model-Level,

Assimilation, Assimilated Meteorological Fields V5.12.4. Goddard Space Flight Center

Distributed Active Archive Center, accessed November 2016,

doi:10.5067/WWQSXQ8IVFW8.

Hudson, R. D., and A. M. Thompson (1998), Tropical tropospheric ozone from total ozone

mapping spectrometer by a modified residual method, *J. Geophys. Res.*, 103(D17),

22129–22145, doi:10.1029/98JD00729.

Jensen, A. A., A. M. Thompson, and F. J. Schmidlin (2012), Classification of Ascension

Island and Natal ozonesondes using self-organizing maps, *J. Geophys. Res.*, 117, D04302,

doi:10.1029/2011JD016573.

Kohonen, T. (1995), The Basic SOM, in Self-Organizing Maps, pp. 77–130, Springer, New

York.

Labow, G. J., J. R. Ziemke, R. D. McPeters, D. P. Haffner, and P. K. Bhartia (2015), A total

ozone-dependent ozone profile climatology based on ozonesondes and Aura MLS data, *J.

Geophys. Res. Atmos.*, 120, 2537–2545. doi: 10.1002/2014JD022634.

Li, K.-F., B. Tian, D. E. Waliser, M. J. Schwartz, J. L. Neu, J. R. Worden, and Y. L. Yung (2012), Vertical structure of MJO-related subtropical ozone variations from MLS, TES, and SHADOZ data, *Atmos. Chem. Phys.*, 12, 425-436, doi:10.5194/acp-12-425-2012.

Lin, M., et al. (2012a), Transport of Asian ozone pollution into surface air over the western United States in spring, *J. Geophys. Res.*, 117, D00V07, doi:10.1029/2011JD016961.

Lin, M., A. M. Fiore, O. R. Cooper, L. W. Horowitz, A. O. Langford, H. Levy II, B. J. Johnson, V. Naik, S. J. Oltmans, and C. J. Senff (2012b), Springtime high surface ozone events over the western United States: Quantifying the role of stratospheric intrusions, *J. Geophys. Res.*, 117, D00V22, doi:10.1029/2012JD018151.

Lin, M., A. M. Fiore, L. W. Horowitz, A. O. Langford, S. J. Oltmans, D. Tarasick, and H. E. Rieder (2015), Climate variability modulates western US ozone air quality in spring via deep stratospheric intrusions, *Nature Comm.*, 6, 7105, doi:10.1038/ncomms8105.

Logan, J. A. (1999), An analysis of ozonesonde data for the troposphere: Recommendations for testing 3-D models and development of a gridded climatology for tropospheric ozone, *J. Geophys. Res.*, 104(D13), 16115–16149, doi:10.1029/1998JD100096.

McPeters, R. D., G. J. Labow, and B. J. Johnson (1997), A satellite-derived ozone climatology for balloonsonde estimation of total column ozone, *J. Geophys. Res.*, 102, D7, 8875–8885, doi:10.1029/96JD02977.

McPeters, R. D., and G. J. Labow (2012), Climatology 2011: An MLS and sonde derived ozone climatology for satellite retrieval algorithms, *J. Geophys. Res.,* 117, D10303, doi:10.1029/2011JD017006.

Oltmans, S. J., et al. (2001), Ozone in the Pacific tropical troposphere from ozonesonde observations, *J. Geophys. Res.*, 106(D23), 32503–32525, doi:10.1029/2000JD900834.

Ott, L. E., B. N. Duncan, A. M. Thompson, G. Diskin, Z. Fasnacht, A. O. Langford, M. Lin, A. M. Molod, J. E. Nielsen, S. E. Pusede, et al. (2016), Frequency and impact of

summertime stratospheric intrusions over Maryland during DISCOVER-AQ (2011): New
evidence from NASA's GEOS-5 simulations, *J. Geophys. Res. Atmos.*, 121, 3687–3706,
doi:10.1002/2015JD024052.

Reid, J. S., et al. (2012), Multi-scale meteorological conceptual analysis of observed active
fire hotspot activity and smoke optical depth in the Maritime Continent, *Atmos. Chem.
Phys.*, 12, 2117-2147, doi:10.5194/acp-12-2117-2012.

Sauvage, B., V. Thouret, A. M. Thompson, J. C. Witte, J.-P. Cammas, P. Nédélec, and G.
Athier (2006), Enhanced view of the "tropical Atlantic ozone paradox" and "zonal wave
one" from the in situ MOZAIC and SHADOZ data, *J. Geophys. Res.*, 111, D01301,
doi:10.1029/2005JD006241.

Smit, H. G. J., and the Panel for the Assessment of Standard Operating Procedures for
Ozonesondes (ASOPOS) (2012), Guidelines for homogenization of ozonesonde data,
SI2N/O3S-DQA activity as part of "Past changes in the vertical distribution of ozone
assessment". [Available at http://www-
das.uwyo.edu/%7Edeshler/NDACC_O3Sondes/O3s_DQA/O3S-DQA-
Guidelines%20Homogenization-V2-19November2012.pdf.]

Smit, H. G. J., and the Panel for the Assessment of Standard Operating Procedures for
Ozonesondes (ASOPOS) (2014), Quality assurance and quality control for ozonesonde
measurements in GAW, World Meteorological Organization, GAW Report 201.
[Available at
http://www.wmo.int/pages/prog/arep/gaw/documents/FINAL_GAW_201_Oct_2014.pdf.]
.

Sofieva, V. F., J. Tamminen, E. Kyrölä, T. Mielonen, P. Veefkind, B. Hassler, and G. E.
Bodeker (2014), A novel tropopause-related climatology of ozone profiles, *Atmos. Chem.
Phys.*, 14, 283-299, doi:10.5194/acp-14-283-2014.

Stauffer, R. M., A. M. Thompson, and G. S. Young (2016), Tropospheric ozonesonde profiles at long-term U.S. monitoring sites: 1. A climatology based on self-organizing maps, *J. Geophys. Res. Atmos.*, 121, 1320-1339, doi:10.1002/2015JD023641.

Stauffer, R. M., A. M. Thompson, S. J. Oltmans, and B. J. Johnson (2017), Tropospheric ozonesonde profiles at long-term U.S. monitoring sites: 2. Links between Trinidad Head, CA, profile clusters and inland surface ozone measurements, *J. Geophys. Res. Atmos.*, 122, 1261–1280, doi:10.1002/2016JD025254.

Stone, K. A., O. Morgenstern, D. J. Karoly, A. R. Klekociuk, W. J. French, N. L. Abraham, and R. Schofield (2016), Evaluation of the ACCESS – chemistry–climate model for the Southern Hemisphere, *Atmos. Chem. Phys.*, 16, 2401-2415, doi:10.5194/acp-16-2401-2016.

Thompson, A. M., J. C. Witte, R. D. Hudson, H. Guo, J. R. Herman, M. Fujiwara (2001), Tropical tropospheric ozone and biomass burning, *Science*, 291, 5511, 2128-2132, doi: 10.1126/science.291.5511.2128.

Thompson, A. M., et al. (2003a), Southern Hemisphere Additional Ozonesondes (SHADOZ) 1998–2000 tropical ozone climatology 1. Comparison with Total Ozone Mapping Spectrometer (TOMS) and ground-based measurements, *J. Geophys. Res.*, 108, 8238, doi:10.1029/2001JD000967, D2.

Thompson, A. M., et al. (2003b), Southern Hemisphere Additional Ozonesondes (SHADOZ) 1998–2000 tropical ozone climatology 2. Tropospheric variability and the zonal wave-one, *J. Geophys. Res.*, 108, 8241, doi:10.1029/2002JD002241, D2.

Thompson, A. M., J. E. Yorks, S. K. Miller, J. C. Witte, K. M. Dougherty, G. A. Morris, D. Baumgardner, L. Ladino, and B. Rappenglück (2008), Tropospheric ozone sources and wave activity over Mexico City and Houston during MILAGRO/Intercontinental

Transport Experiment (INTEX-B) Ozonesonde Network Study, 2006 (IONS-06), *Atmos. Chem. Phys.*, 8, 5113-5125, https://doi.org/10.5194/acp-8-5113-2008.

Thompson, A. M., A. L. Allen, S. Lee, S. K. Miller, and J. C. Witte (2011), Gravity and Rossby wave signatures in thetropical troposphere and lower stratosphere based on Southern Hemisphere Additional Ozonesondes (SHADOZ), 1998–2007, *J. Geophys. Res.*, 116, D05302, doi:10.1029/2009JD013429.

Thompson, A. M., et al. (2012), Southern Hemisphere Additional Ozonesondes (SHADOZ) ozone climatology (2005–2009): Tropospheric and tropical tropopause layer (TTL) profiles with comparisons to OMI-based ozone products, *J. Geophys. Res.*, 117, D23301, doi:10.1029/2011JD016911.

Thompson, A. M., et al. (2017), First reprocessing of Southern Hemisphere Additional Ozonesondes (SHADOZ) profiles (1998-2016). 2. Comparisons with satellites and ground-based instruments, *J. Geophys. Res. Atmos.*, doi:10.1002/2017JD27406.

Tian, B., Y. L. Yung, D. E. Waliser, T. Tyranowski, L. Kuai, E. J. Fetzer, and F. W. Irion (2007), Intraseasonal variations of the tropical total ozone and their connection to the Madden-Julian Oscillation, *Geophys. Res. Lett.*, 34, L08704, doi:10.1029/2007GL029451.

Tilmes, S., et al. (2012), Technical Note: Ozonesonde climatology between 1995 and 2011: Description, evaluation and applications, *Atmos. Chem. Phys.*, 12, 7475-7497, doi:10.5194/acp-12-7475-2012.

Tilmes, S., et al. (2016), Representation of the Community Earth System Model (CESM1) CAM4-chem within the Chemistry-Climate Model Initiative (CCMI), *Geosci. Model Dev.*, 9, 1853-1890, doi:10.5194/gmd-9-1853-2016.

Van Malderen, R., M. A. F. Allaart, H. De Backer, H. G. J. Smit, and D. De Muer (2016), On instrumental errors and related correction strategies of ozonesondes: possible effect on

calculated ozone trends for the nearby sites Uccle and De Bilt, *Atmos. Meas. Tech.*, 9, 3793-3816, doi:10.5194/amt-9-3793-2016.

Ventrice, M. J., M. C. Wheeler, H. H. Hendon, C. J. Schreck III, C. D. Thorncroft, G. N. Kiladis (2013), A modified multivariate Madden-Julian Oscillation index using velocity potential, *Mon. Wea. Rev.*, 141, 4197-4210, doi:10.1175/MWR-D-12-00327.1.

Wargan, K., G. Labow, S. Firth, S. Pawson, N. Livesey, G. Partyka (2017), Evaluation of the ozone fields in NASA's MERRA-2 reanalysis, *J. Clim.*, 30, 2961-2988, doi:10.1175/JCLI-D-16-0699.1.

Wheeler, M. C., H. H. Hendon (2004), An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction, *Mon. Wea. Rev.*, 132, 1917-1932, doi:10.1175/MWR-D-12-00327.1.

Witte, J. C., et al. (2017), First reprocessing of Southern Hemisphere ADditional OZonesondes (SHADOZ) profile records (1998–2015): 1. Methodology and evaluation, *J. Geophys. Res. Atmos.*, 122, doi:10.1002/2016JD026403.

WMO/GAW UV Radiation Monitoring Community, World Meteorological Organization-Global Atmosphere Watch Program (WMO-GAW)/World Ozone and Ultraviolet Radiation Data Centre (WOUDC) [Data]. Retrieved October, 2016, from http://woudc.org. A list of all contributors is available on the website. doi:10.14287/10000008.

Yonemura, S., H. Tsuruta, S. Kawashima, S. Sudo, L. C. Peng, L. S. Fook, Z. Johar, and M. Hayashi (2002), Tropospheric ozone climatology over Peninsular Malaysia from 1992 to 1999, *J. Geophys. Res.*, 107(D15), doi:10.1029/2001JD000993.

Ziemke, J. R., and S. Chandra (2003), A Madden-Julian Oscillation in tropospheric ozone, *Geophys. Res. Lett.*, 30, 2182, doi:10.1029/2003GL018523, 23.

Ziemke, J. R., S. Chandra, G. J. Labow, P. K. Bhartia, L. Froidevaux, and J. C. Witte (2011),

A global climatology of tropospheric and stratospheric ozone derived from Aura OMI and

MLS measurements, Atmos. Chem. Phys., 11, 9237-9251, doi:10.5194/acp-11-9237-

2011.

Table 1. Ozonesonde site locations and associated metadata. Sondes listed are Electrochemical Concentration Cell (ECC), Brewer-Mast (BM), Carbon-Iodide (CI), and a Brewer-Mast type sonde made in the German Democratic Republic (GDR). Note that only data after 1980 are considered in this paper and in the data samples column (N).

Station Name	WOUDC #	Lat (°)	Lon (°)	Elevation (m)	Record (Month/Year)	N	Sonde Type
Alert	18	82.49	-62.34	62	12/87 - 12/14	1388	ECC
Ny-Ålesund	89	78.93	11.95	10	11/90 - 7/13	2235	ECC
Resolute	24	74.70	-94.96	64	1/66 - 12/14	1324	ECC
Churchill	77	58.74	-94.07	35	10/73 - 12/13	1358	ECC
Edmonton	21	53.54	-114.10	668	10/70 - 12/14	1546	ECC
Goose Bay	76	53.31	-60.36	44	9/80 - 12/14	1573	ECC
Legionowo	221	52.41	20.96	96	2/79 - 5/93; 6/93 - 5/16	1725	GDR; ECC
Lindenberg	174	52.21	14.12	98	1/75 - 6/92; 7/92 - 8/14	2184	GDR; ECC
De Bilt	316	52.10	5.18	4	11/92 - 10/16	1260	ECC
Uccle	53	50.80	4.35	100	11/66 - 3/97; 3/97 - 11/16	4710	BM; ECC
Hohenpeissenberg	99	47.80	11.00	975	11/66 - 7/16	4573	BM
Payerne	156	46.49	6.57	491	9/68 - 8/02; 9/02 - 3/16	5061	BM; ECC
Sapporo	12	43.06	141.33	19	1/69 - 11/09; 12/09 - 7/16	1164	CI; ECC
Boulder	67	40.00	-105.25	1742	3/79 - 11/16	1572	ECC
Wallops Island	107	37.93	-75.48	4	5/70 - 7/16	1420	ECC
Tateno	14	36.06	140.13	31	12/68 - 11/09; 12/09 - 7/16	1536	CI; ECC
Kagoshima	7	31.60	130.60	283	2/69 - 3/05	665	CI; ECC
Naha	190	26.21	127.69	27	10/89 - 10/08; 11/08 - 7/16	1052	CI; ECC
Hanoi	330	21.01	105.80	7	9/04 - 9/15	221	ECC
Hilo	109	19.43	-155.04	11	9/82 - 11/16	1498	ECC
Kuala Lumpur	443	2.73	101.27	60	1/98 - 6/16	396	ECC
Natal	219	-5.42	-35.38	30	1/98 - 7/16	579	ECC
Watukosek	437	-7.50	112.60	50	6/01 - 10/13	243	ECC
Ascension	328	-7.58	-14.24	85	7/90 - 10/92; 7/97 - 8/10; 3/16 - 11/16	718	ECC
Samoa	191	-14.23	-170.56	77	1/98 - 8/16	626	ECC
Reunion	436	-21.06	55.48	20	1/ 98 - 4/16	582	ECC
Irene	265	-25.90	28.22	1520	7/90 - 10/93; 11/98 - 4/16	468	ECC
Lauder	256	-45.00	169.68	370	8/86 - 4/15	1680	ECC
Marambio	233	-64.23	-56.62	198	11/88 - 9/16	1120	ECC
Syowa	101	-69.01	39.58	21	4/66 - 3/10; 4/10 - 7/16	1547	CI; ECC

Table 2. Supporting institutions for the ozonesonde sites used in this study.

Station Name	Local Supporting Institution
Alert	Environment and Climate Change Canada
Ny-Ålesund	Alfred-Wegener Institute
Resolute	Environment and Climate Change Canada
Churchill	Environment and Climate Change Canada
Edmonton	Environment and Climate Change Canada
Goose Bay	Environment and Climate Change Canada
Legionowo	Polish Institute of Meteorology and Water Management
Lindenberg	Deutscher Wetterdienst (DWD)
De Bilt	Royal Netherlands Meteorological Institute (KNMI)
Uccle	Royal Meteorological Institute of Belgium (KMI/IRM)
Hohenpeissenberg	Deutscher Wetterdienst (DWD)
Payerne	MeteoSwiss
Sapporo	Japan Meteorological Agency
Boulder	NOAA/ESRL Global Monitoring Division
Wallops Island	NASA/GSFC WFF
Tateno	Japan Meteorological Agency
Kagoshima	Japan Meteorological Agency
Naha	Japan Meteorological Agency
Hanoi	Japan Agency for Marine-Earth Science and Technology
Hilo	NOAA/ESRL Global Monitoring Division
Kuala Lumpur	Malaysian Meteorological Department
Natal	Instituto Nacional de Pesquisas Espaciais (INPE)
Watukosek	Lembaga Penerbangan dan Antariksa Nasional (LAPAN)
Ascension	NASA/GSFC and USAF
Samoa	NOAA/ESRL Global Monitoring Division
Reunion	Université de La Réunion
Irene	South African Weather Service
Lauder	National Institute of Water and Atmospheric Research
Marambio	National Meteorological Service of Argentina/Finnish Meteorological Institute
Syowa	Japan Meteorological Agency

Table 3. Ozonesonde profile cluster membership by site. The number and percentage of each site's total profiles are provided, with the average tropospheric O_3 mixing ratio for each SOM cluster.

|	**Cluster 1**			**Cluster 2**			**Cluster 3**			**Cluster 4**		
Station Name	N	%	O_3 (ppbv)	N	%	O_3 (ppbv)	N	%	O_3 (ppbv)	N	%	O_3 (ppbv)
Alert	619	45	49.1	369	27	59.1	262	19	72.1	137	10	95.1
Ny-Ålesund	1042	47	52.4	584	26	61.5	409	18	75.6	196	9	98.9
Resolute	624	47	49.5	340	26	60.7	240	18	76.9	120	9	104.5
Churchill	624	46	53.9	318	23	65.1	274	20	85.8	142	10	120.3
Edmonton	724	47	51.0	376	24	62.0	292	19	79.5	154	10	107.4
Goose Bay	854	54	53.7	337	21	67.0	253	16	85.6	129	8	122.3
Legionowo	739	43	55.8	485	28	76.4	342	20	92.4	159	9	137.4
Lindenberg	833	38	55.6	689	32	76.3	421	19	91.2	241	11	132.3
De Bilt	547	43	52.4	360	29	69.5	218	17	85.0	134	11	126.7
Uccle	2005	43	52.6	1472	31	70.0	791	17	85.0	414	9	122.7
Hohenpeissenberg	1934	42	50.6	809	18	69.0	1402	31	65.1	428	9	97.0
Payerne	1981	39	46.8	815	16	71.6	1789	35	67.4	469	9	103.0
Sapporo	601	52	61.4	254	22	74.0	196	17	95.0	113	10	131.8
Boulder	794	51	56.5	383	24	63.3	257	16	73.2	136	9	113.6
Wallops Island	656	46	58.5	407	29	80.8	223	16	101.0	134	9	158.2
Tateno	719	47	56.8	263	17	83.2	394	26	76.3	160	10	124.1
Kagoshima	318	48	47.9	197	30	66.8	113	17	78.9	37	6	105.2
Naha	224	21	52.4	363	35	44.6	290	28	62.5	175	17	80.7
Hanoi	80	36	45.2	56	25	59.6	48	22	64.2	37	17	76.7
Hilo	666	44	40.0	329	22	52.3	356	24	68.4	123	8	95.2
Kuala Lumpur	159	40	34.4	93	23	40.0	81	20	46.0	63	16	53.6
Natal	185	32	40.6	115	20	55.9	122	21	59.2	157	27	73.4
Watukosek	63	26	27.6	82	34	36.5	60	25	38.3	38	16	50.9
Ascension	187	26	46.0	239	33	59.4	146	20	69.2	146	20	80.1
Samoa	217	35	24.4	165	26	33.4	147	23	42.5	94	15	52.6
Reunion	180	31	43.0	166	29	57.4	109	19	64.8	127	22	80.4
Irene	97	21	64.0	205	44	60.6	39	8	78.7	127	27	82.2
Lauder	847	50	37.3	437	26	45.0	234	14	54.1	160	10	76.9
Marambio	470	42	37.3	263	23	42.8	236	21	51.9	150	13	67.9
Syowa	505	33	36.1	433	28	42.2	384	25	49.4	225	15	63.1
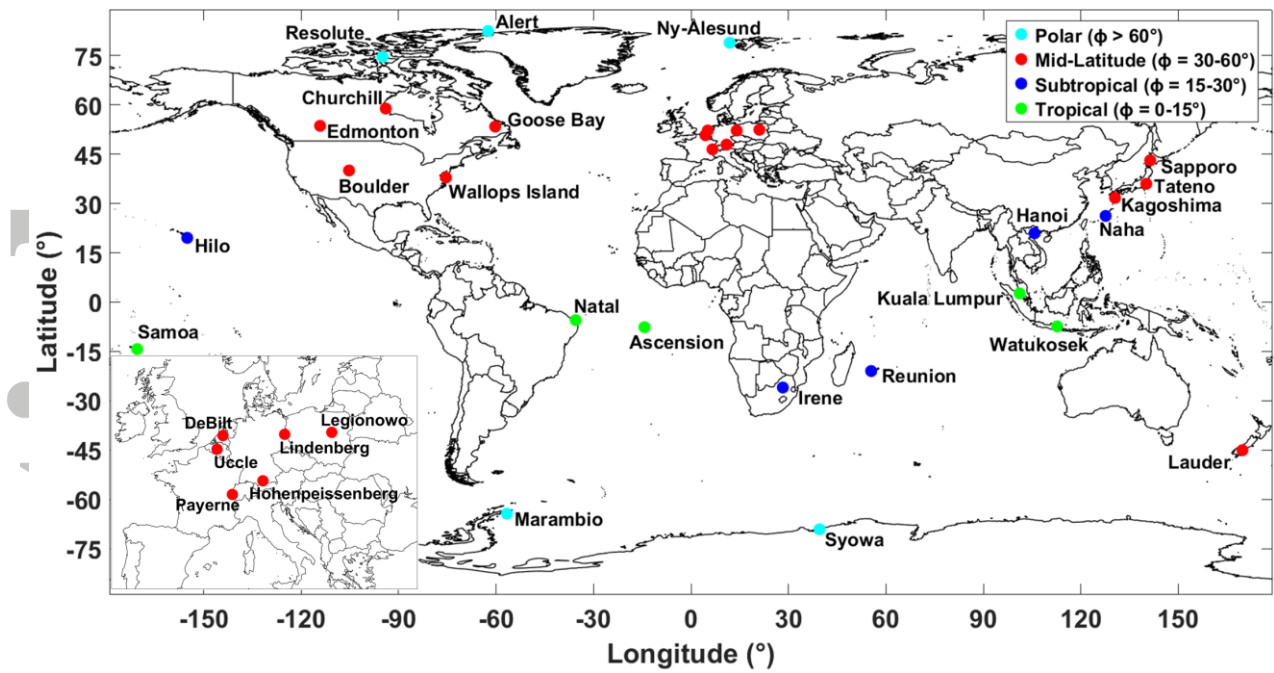
Figure 1. Locations of ozonesonde sites colored by latitude region. European site locations are provided in the figure inset.
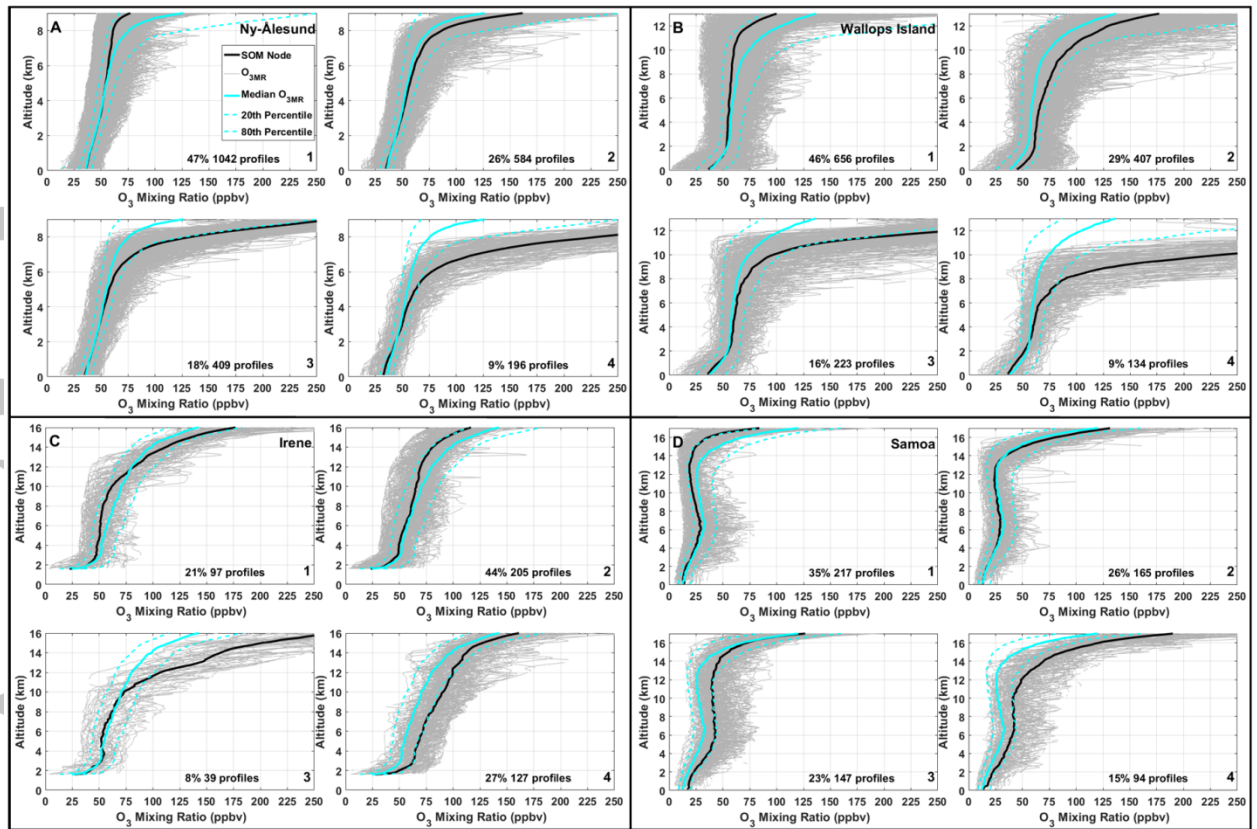
Figure 2. Examples of O$_3$ profile SOM clusters from each of the four latitude regions: Polar (A, Ny-Ålesund), Mid-Latitude (B, Wallops Island), Subtropical (C, Irene), and Tropical (D, Samoa). Individual profiles are shown in grey, with SOM cluster mean in black. For reference, the median and 20th and 80th percentile O$_3$ for the entire site data set is shown in cyan. Note the different vertical axes on each plot.
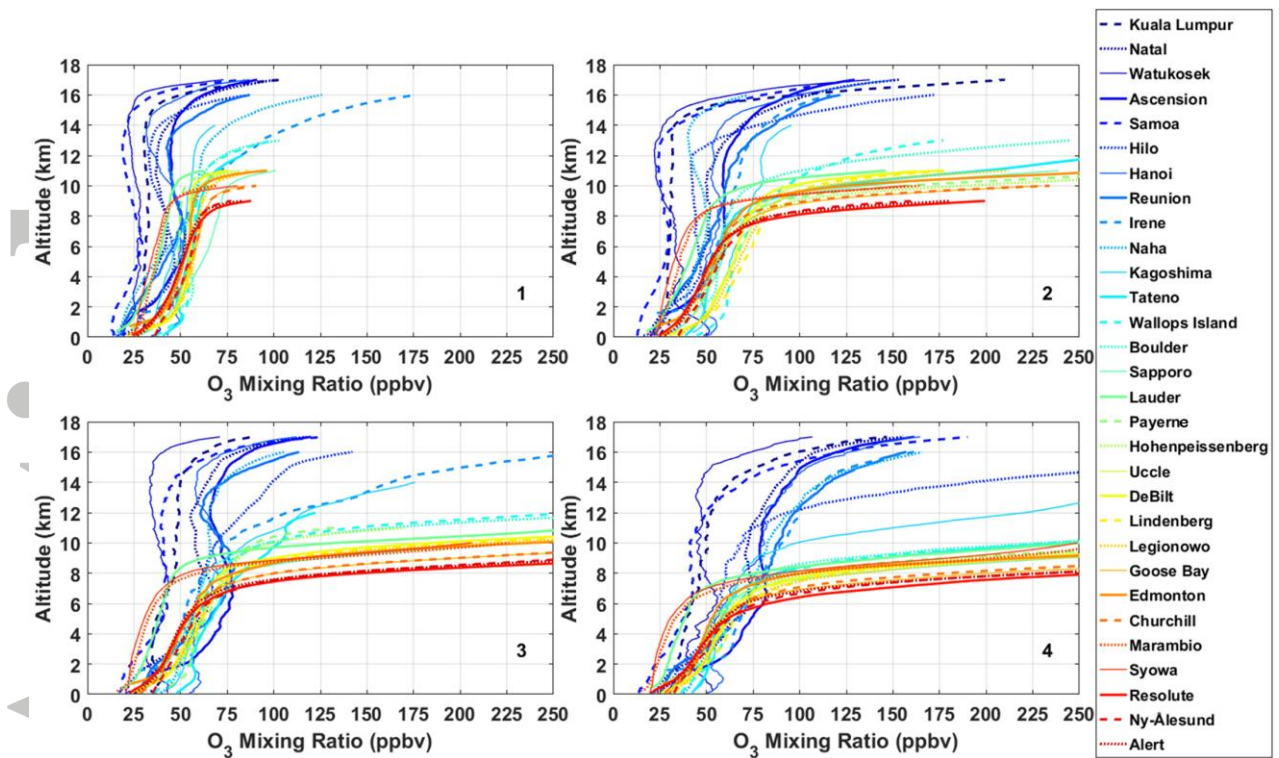
Figure 3. SOM cluster means (nodes) for all 30 ozonesonde sites arranged by cluster (numbers on plots). Sites are organized and colored by absolute value of latitude from Tropical (blues) to Polar (reds).
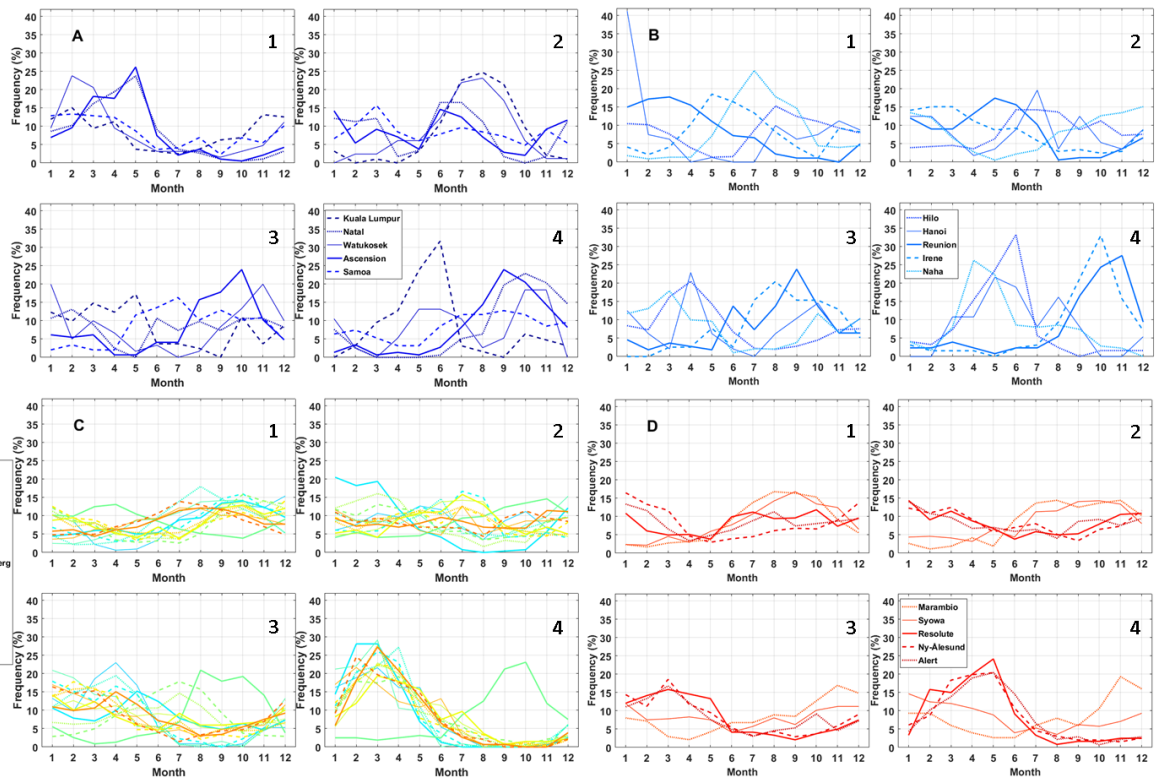
Figure 4. Histograms of launch months within each SOM cluster (numbers on plots) for profiles at all 30 ozonesonde sites. Plots are arranged in groups of Tropical (A), Subtropical (B), Mid-Latitude (C), and Polar (D) sites for clarity.
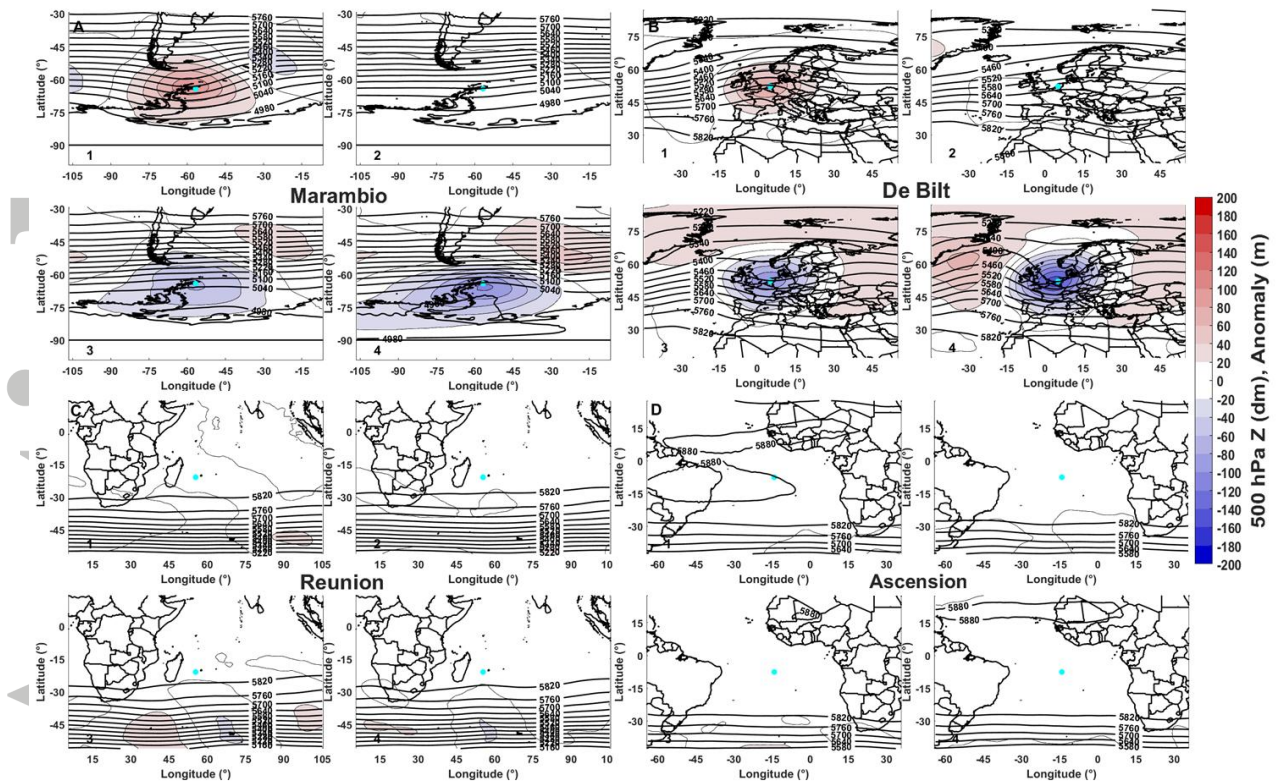
Figure 5. Contoured map of average MERRA-2 500 hPa geopotential heights (Z; black contours; dm) and anomalies (colors; m) from climatology (1981-2010) corresponding to each SOM cluster (numbers on lower left of plots). Examples from Polar (A, Marambio), Mid-Latitude (B, De Bilt), Subtropical (C, Reunion), and Tropical (D, Ascension) regions are shown. The site location is marked by the cyan dot in the center of the panels.
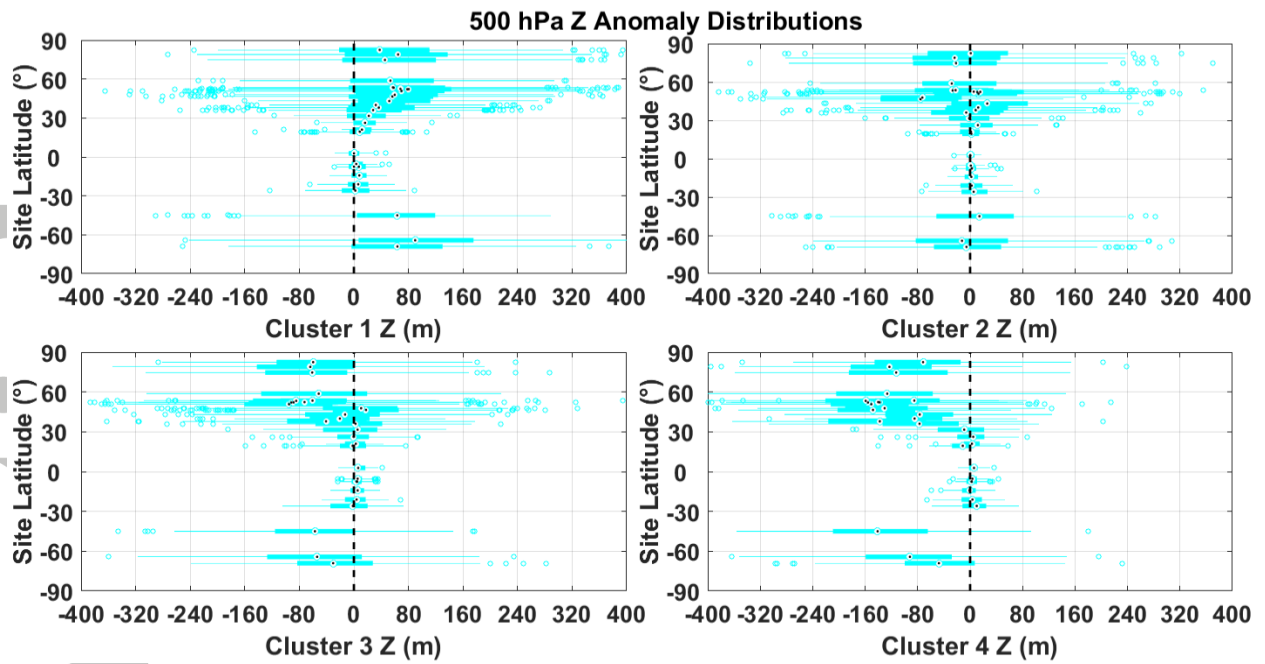
Figure 6. Box and whisker plots of MERRA-2 500 hPa geopotential height anomaly distributions for all 30 ozonesonde sites arranged by latitude and SOM cluster number. The black dot represents the distribution median, with box edges representing the 25th and 75th percentile interquartile range. Whiskers are drawn to 1.5 times the interquartile range from each box edge, with any points beyond that marked as outliers (cyan dots).
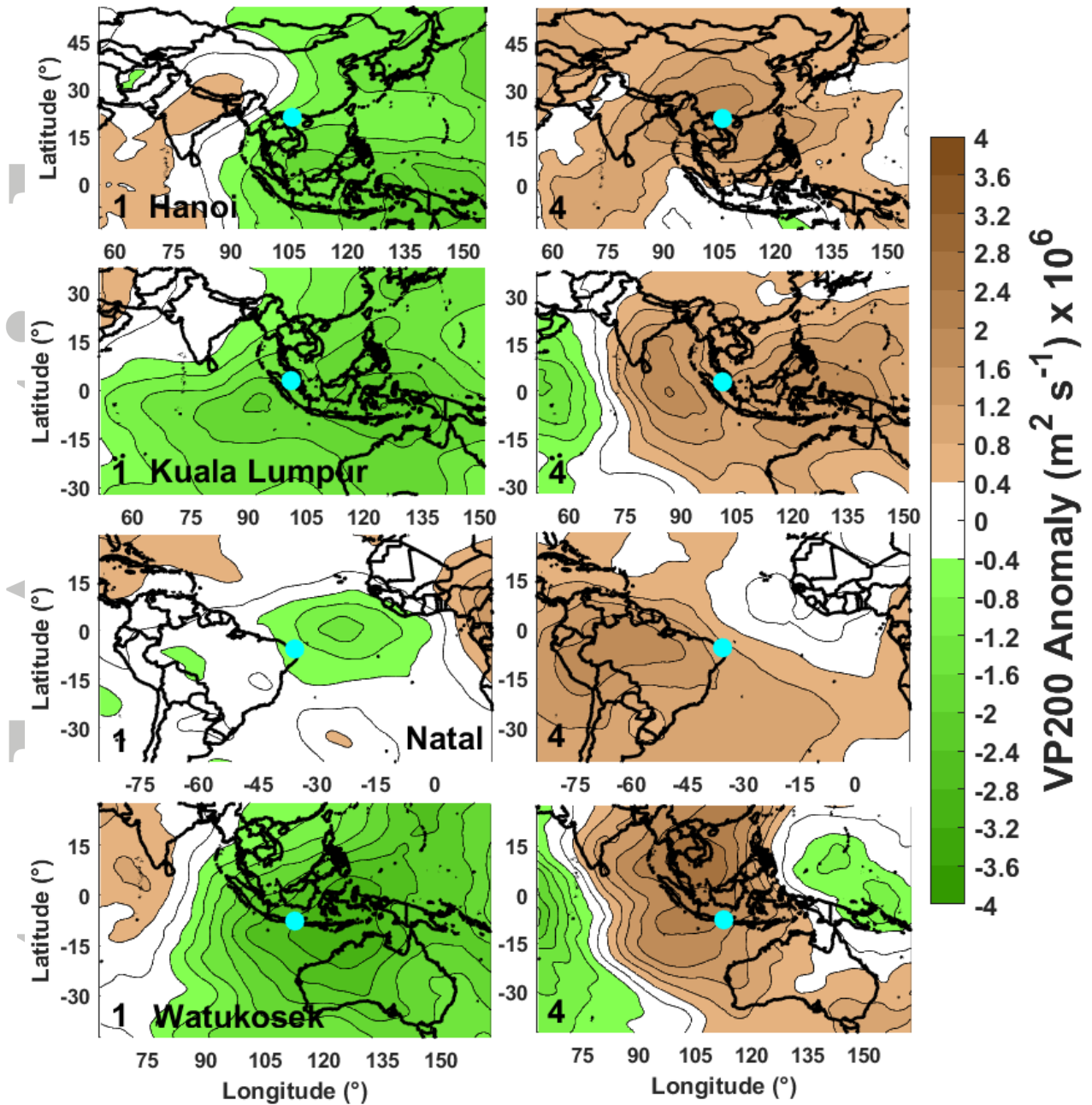
Figure 7. Contoured map of average MERRA-2 calculated 200 hPa velocity potential anomalies (colors) from climatology corresponding the 1st (lowest O$_3$) and 4th (highest O$_3$) SOM clusters (numbers on lower left of plots) at Hanoi, Kuala Lumpur, Natal, and Watukosek. The site location is marked by the cyan dot in the center of the panels.
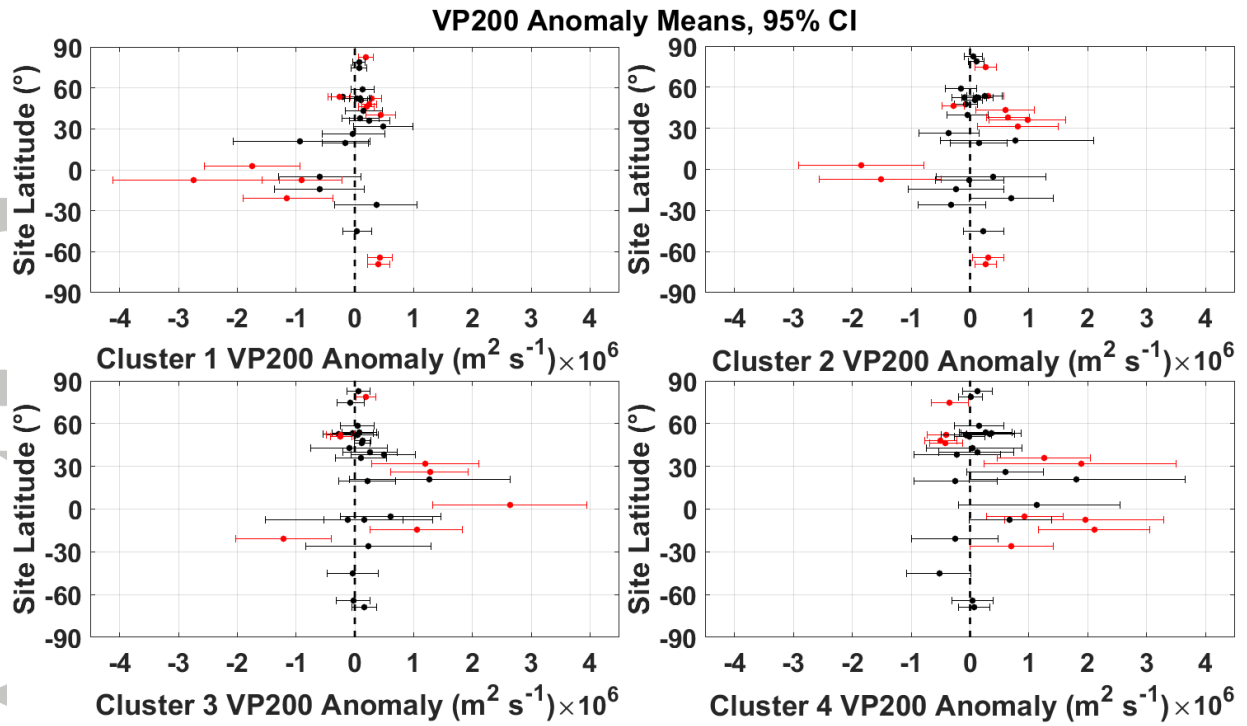
Figure 8. Mean and 95% confidence intervals of the mean MERRA-2 calculated 200 hPa velocity potential anomalies for all 30 ozonesonde sites arranged by latitude and SOM cluster number. The dot represents the distribution mean, with the 95% confidence interval of the mean shown by the error bars. Site means that are significantly different from a zero VP200 anomaly are colored red.
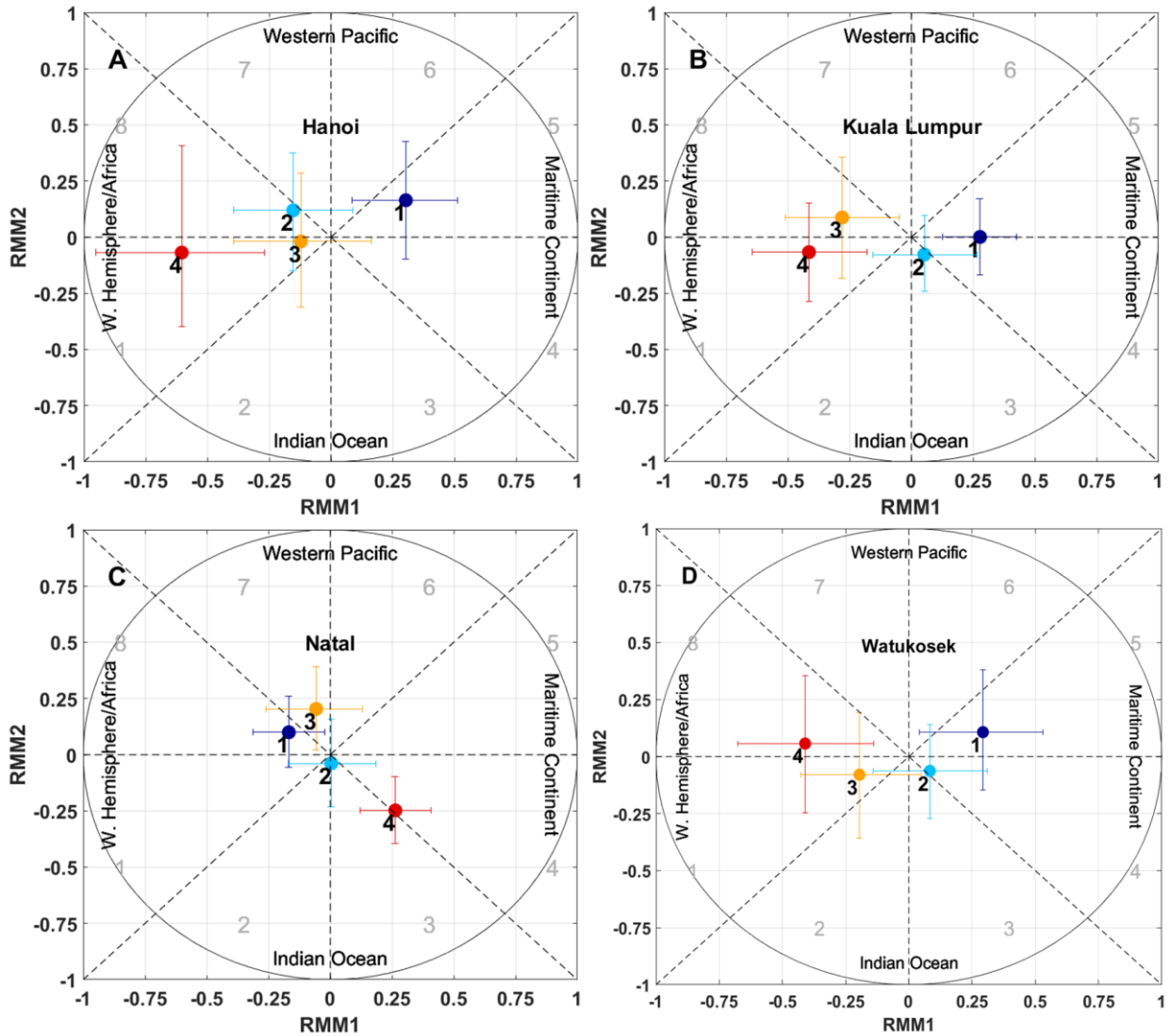
Figure 9. Real-time Multivariate MJO Index (RMM) phase space diagrams associated with each of the four O_3 profile clusters (numbers next to colored dots) for Hanoi (A), Kuala Lumpur (B), Natal (C), and Watukosek (D). The colored dots are the mean for each cluster, with the error bars representing the 95% confidence interval of the mean. The regions of enhanced convective activity (e.g. Indian Ocean) associated with each MJO phase 1-8 are labeled on the panels. The black circle represents "strong" MJO events ($RMM1^2 + RMM2^2 \geq 1$).
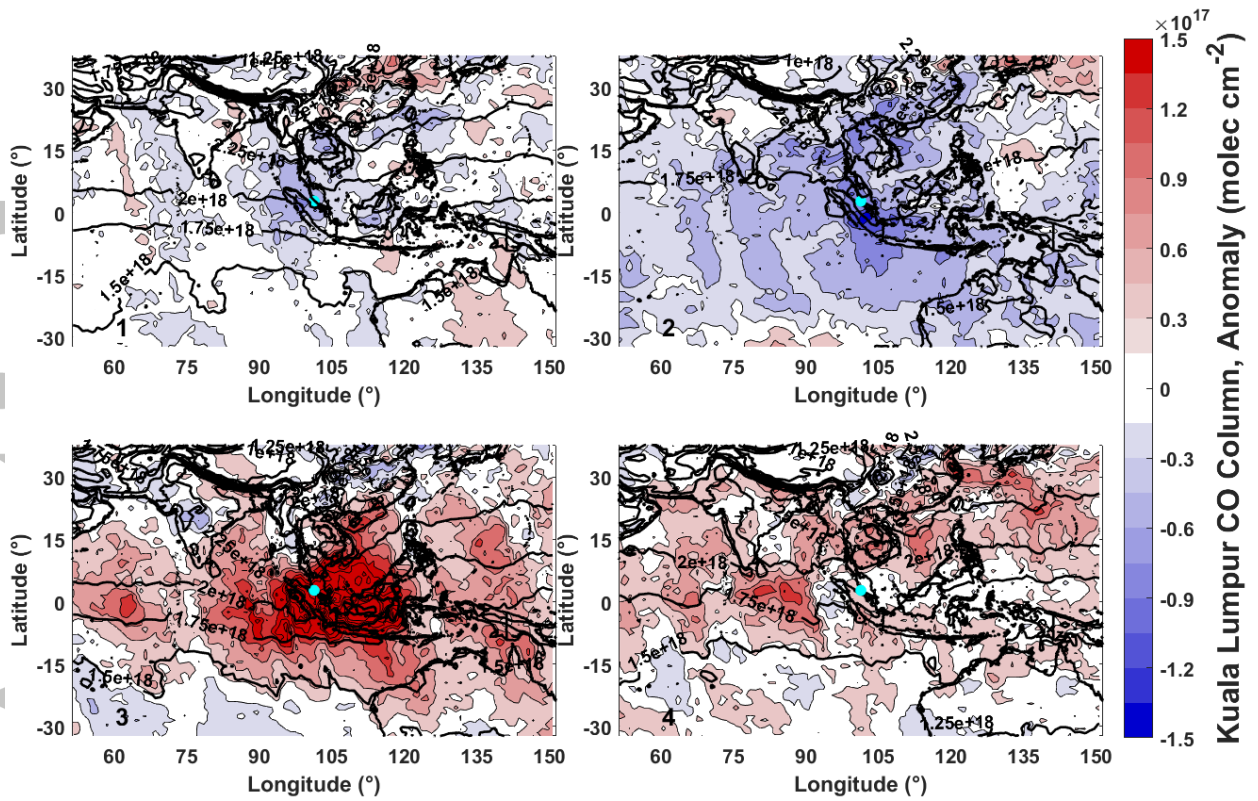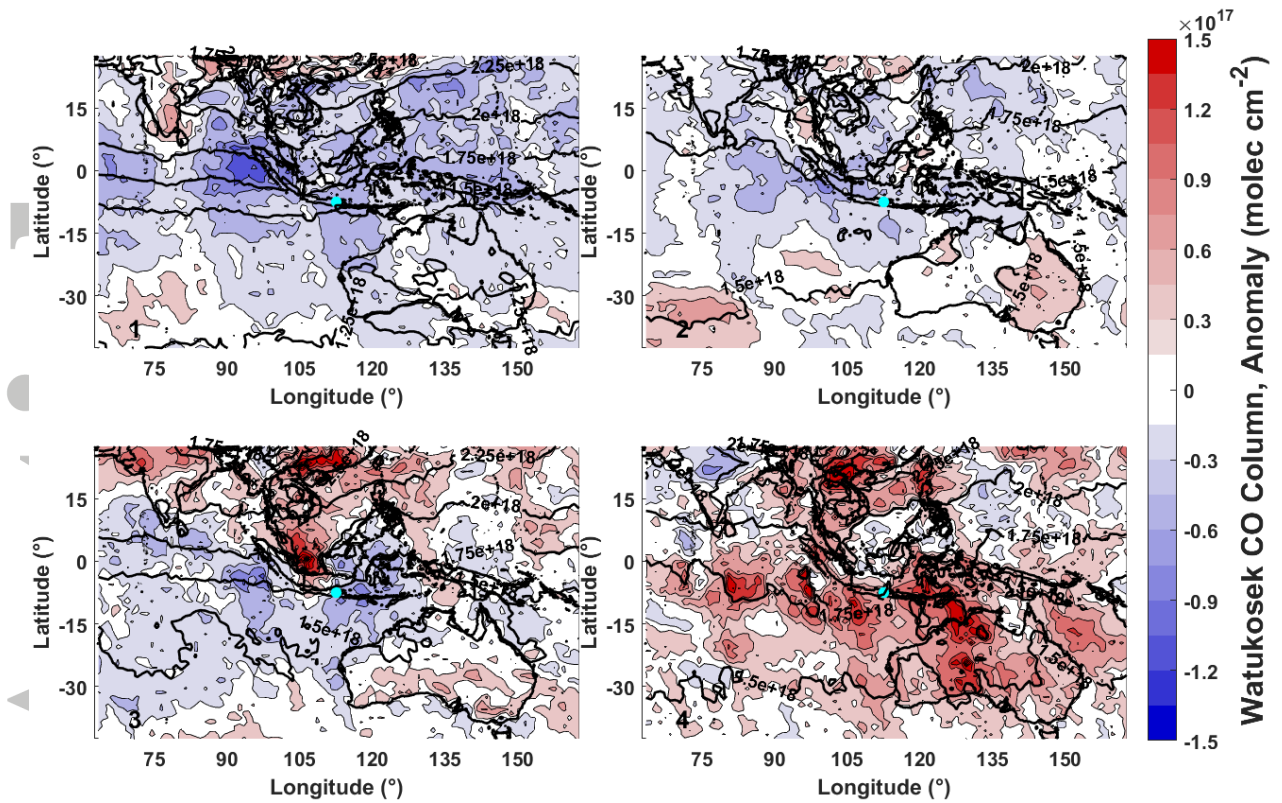
Figure 10. Contoured map of average AIRS total CO columns and anomalies from monthly climatology (2002-2016) corresponding to each SOM cluster (numbers on plots) at Kuala Lumpur. The site location is marked by the cyan dot in the center of the panels.

Figure 11. Contoured map of average AIRS total CO columns and anomalies from monthly climatology (2002-2016) corresponding to each SOM cluster (numbers on plots) at Watukosek. The site location is marked by the cyan dot in the center of the panels.
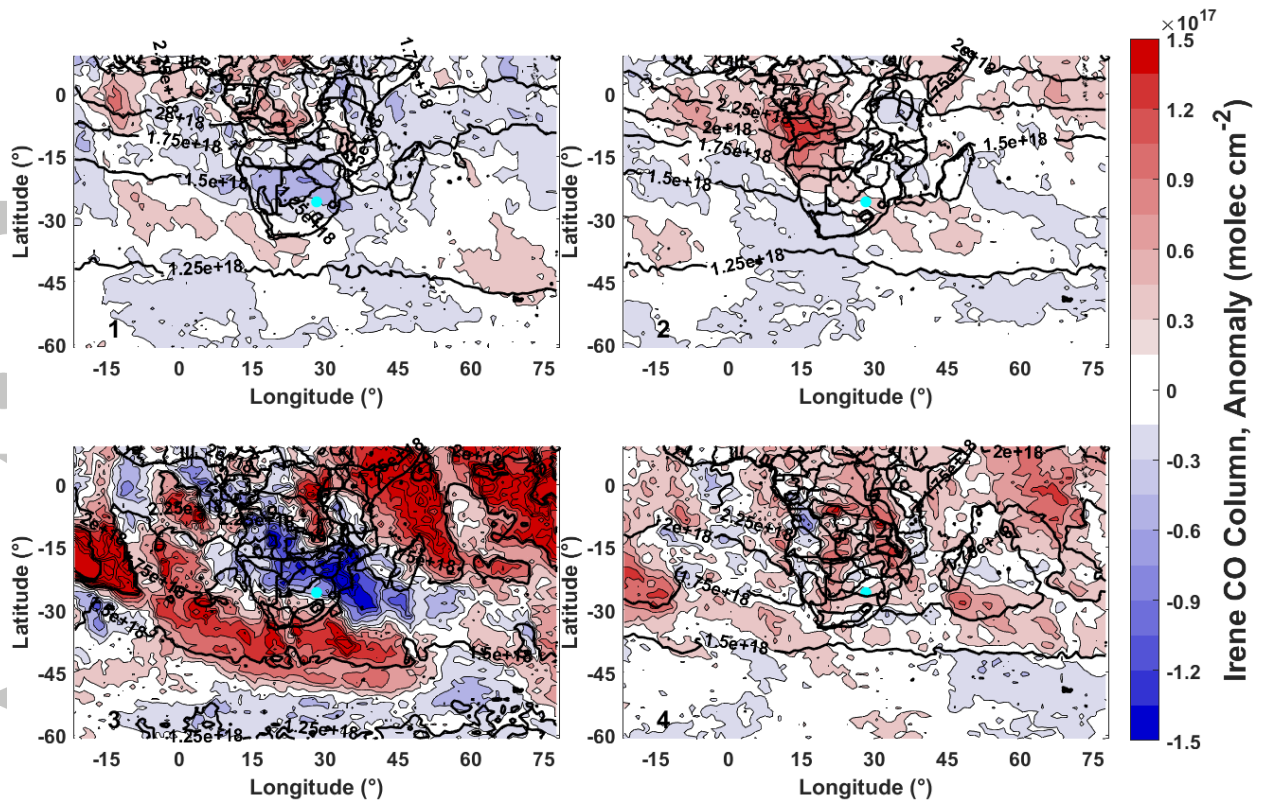
Figure 12. Contoured map of average AIRS total CO columns and anomalies from monthly climatology (2002-2016) corresponding to each SOM cluster (numbers on plots) at Irene. The site location is marked by the cyan dot in the center of the panels.