

Supplementary Material

As shown in the Chapter 2.1, PCA was applied to the standardized ozone anomaly fields to reduce the dimensionality of the dataset while retaining the dominant modes of spatial variability. The first 40 principal components (PCs) explain 72.2% of the total variance, with the explained variance decreasing progressively from 8.1% for PC1 to less than 1% for each PC from PC23. The first 40 PCs were therefore retained as a compromise between preserving the dominant variability of the ozone fields and reducing the dimensionality of the dataset for the subsequent clustering analysis. The resulting PC matrix consists of 1281 daily observations and 40 principal components, which were used as input to the k-means clustering.

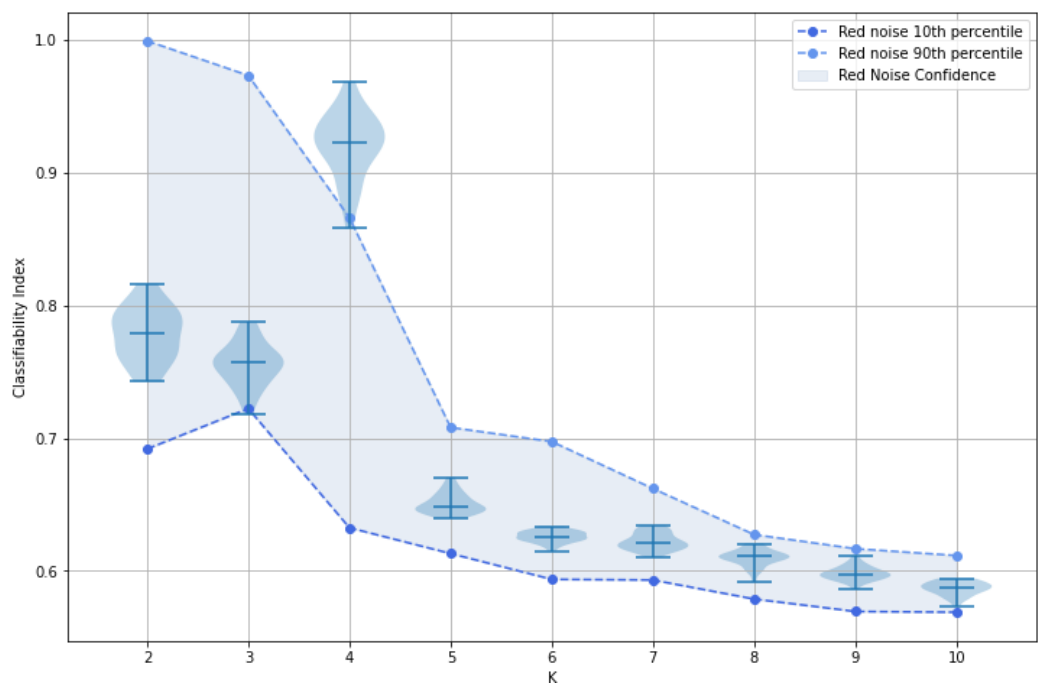


Figure S1: Classifiability indices resulting from the k-means clustering and the red noise test for k = [2,10].

10 Figure S1 shows the k-means clustering algorithm and red-noise test output as a distribution of classifiably numbers for each number of classes [2;10]. For CAMS ozone at 850 hPa in the selected period and domain, only the four-classes classification is statistically significant, as indicated by the classifiability index exceeding the Red Noise confidence level.

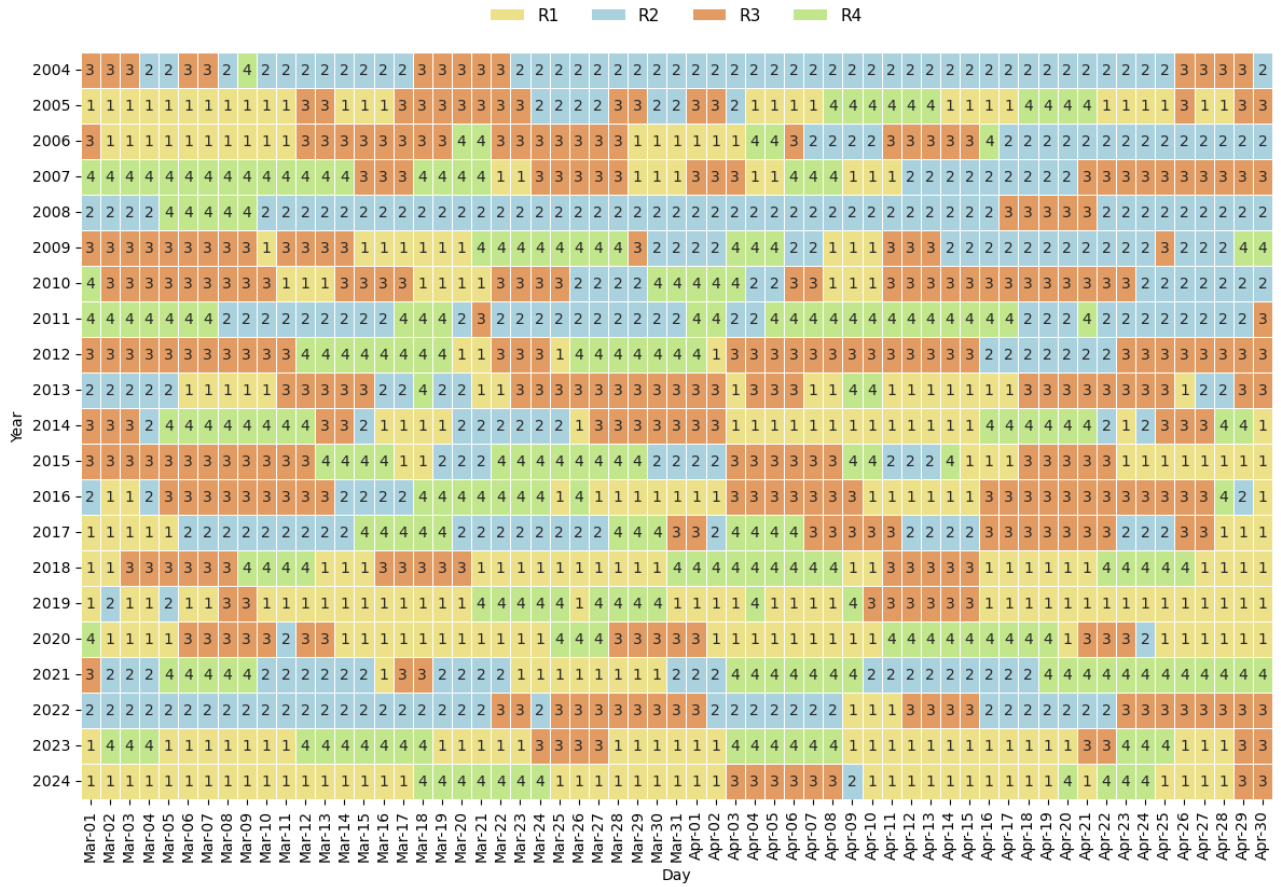


Figure S2: Daily classification (2004-2024, March-April) in ozone regimes according to the k-mean clustering.

15 Figure S2 represents the calendar according to the selected classification of each daily ozone field in four classes.

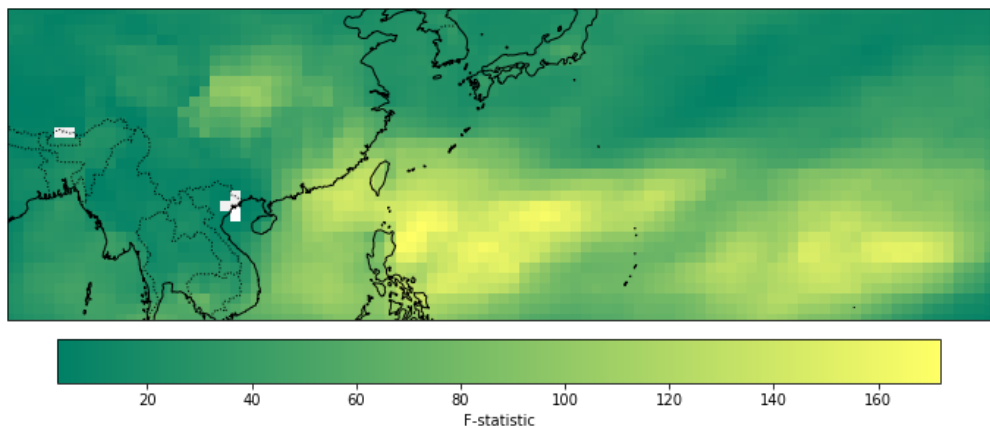
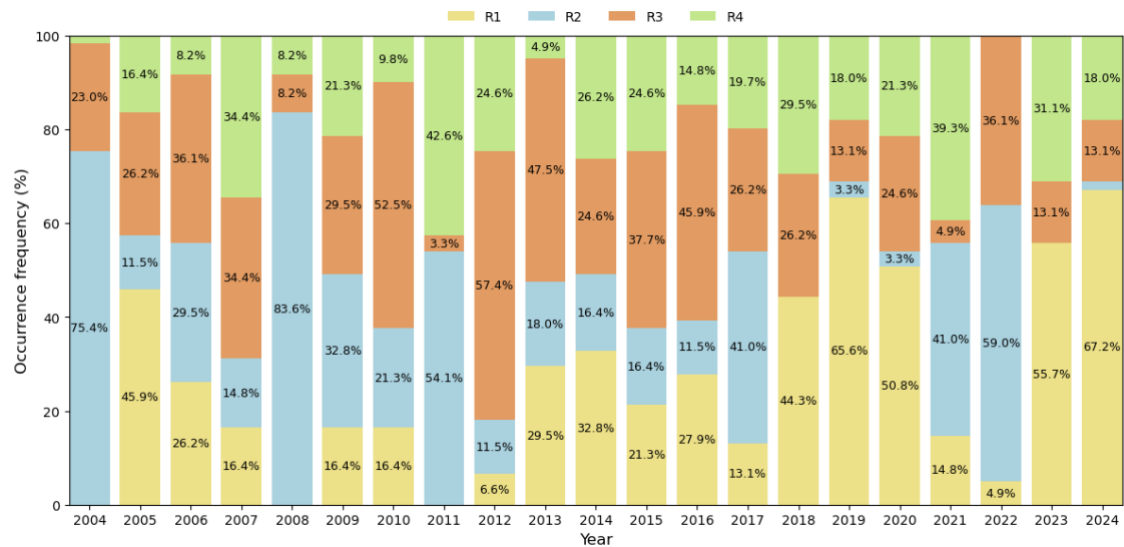


Figure S3: F-values resulting from the ANOVA test. White pixels represent non-significant areas with $p > 0.005$.

Figure S3 show the ANOVA test for the four selected clusters. According to the high F values, most of the spatial variability is explained by differences between clusters ($p<0.05$).



20

Figure S4: Interannual variability of ozone regimes during March–April over the 2004–2024 period. Stacked bars show the relative occurrence frequency (%) of each regime for each year.

Figure S4 shows the occurrence of each ozone regime by year. R1, R2 and R3 exhibit a strong interannual variability, whereas R4 is more stable.

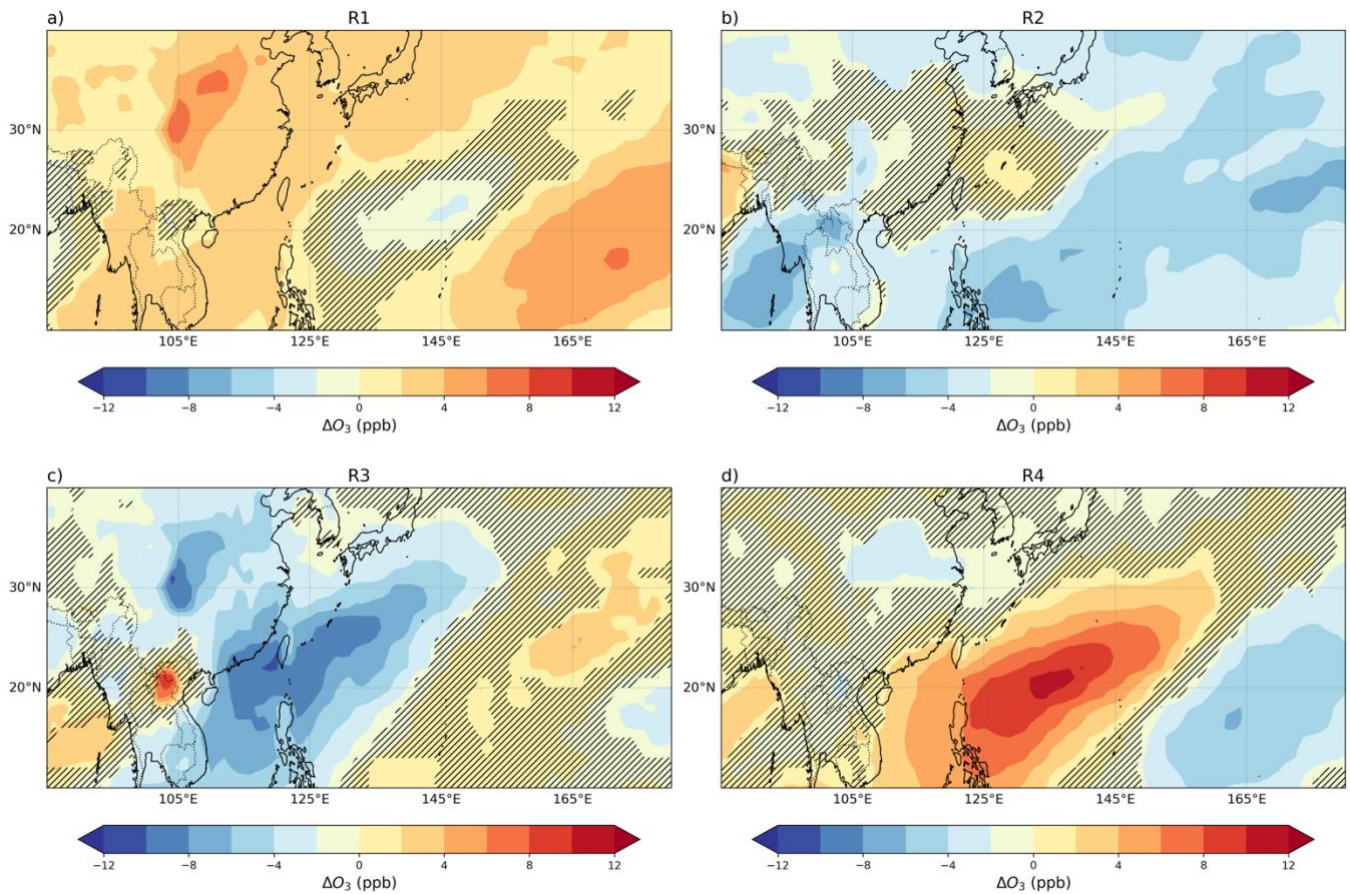
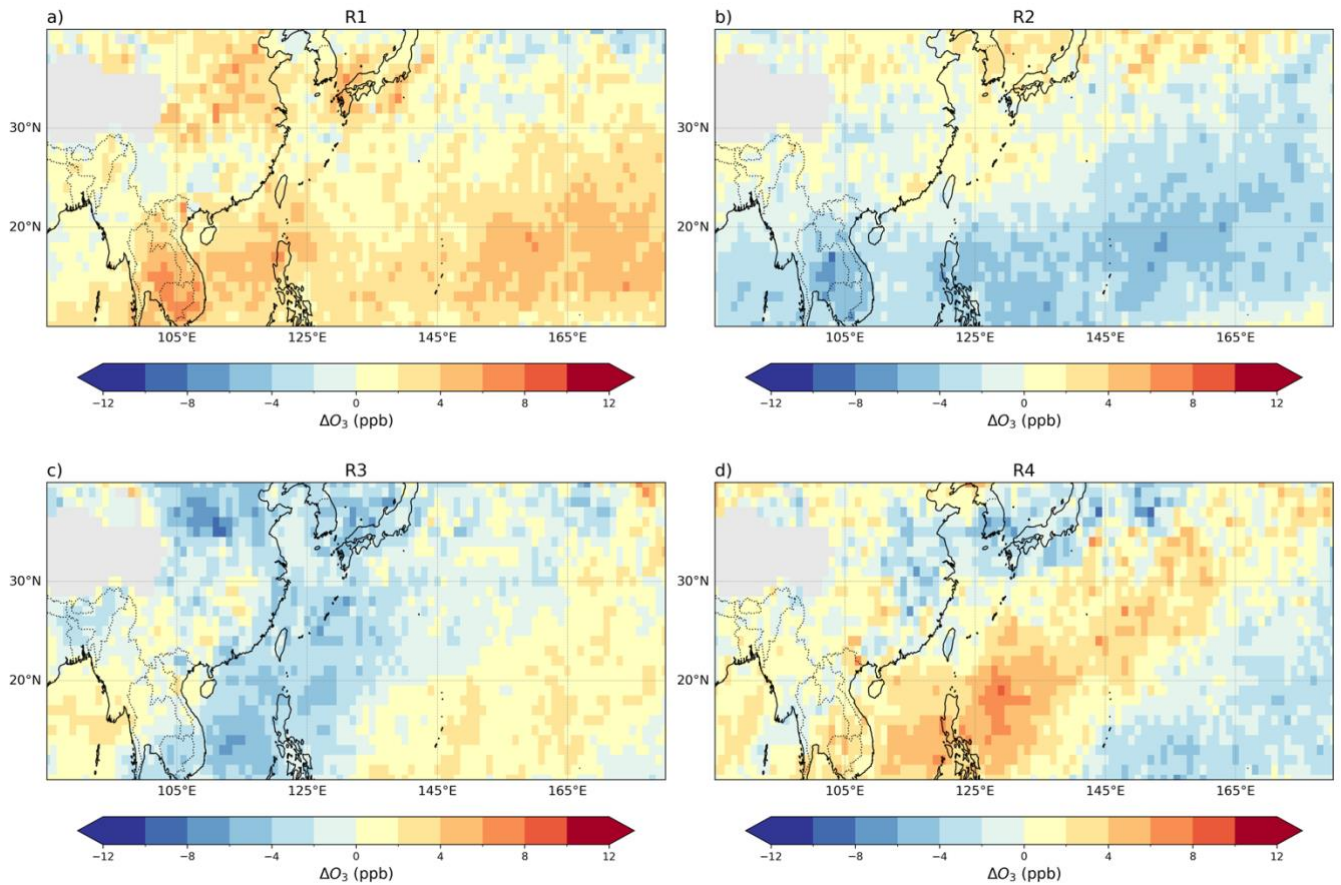


Figure S5: The four ozone regimes obtained by averaging daily ozone anomalies (ppb) from CAMS at 850 hPa during March–April 2017–2024. The non-shaded areas represent values significant at the 95% level of confidence after the T-test. Red colours represent positive anomalies in ppb and blue colours negative anomalies.

Figure S5 shows the same maps as Figure 2, but the averaging period is restricted to March–April 2017–2024 to match the period covered by the IASI+GOME2 observations. Therefore, Figure S5 is used as comparison with the IASI+GOME2 ozone regimes shown in Figure 4. Using the same averaging period allows a consistent comparison between the CAMS and satellite-derived ozone regimes.



35 **Figure S6: Composite mean O₃ anomalies (ppb) from IASI+GOME2 satellite data, obtained using the spatial projection method for the four CAMS-derived regimes: (a) R1, (b) R2 (c) R3 and (d) R4. Each satellite day was assigned to the regime with the smallest Euclidean distance to the corresponding CAMS composite mean map.**

To further evaluate the temporal consistency between CAMS and IASI+GOME2, we applied a spatial projection method. This approach consists in projecting the daily IASI+GOME2 O₃ observations onto the four regimes previously identified from CAMS. For each daily satellite O₃ dataset, we computed the Euclidean distance to the four CAMS composite means
 40 and assigned the day to the regime with the smallest distance. New composite means were then built using only the satellite data (Figure S6). These maps show a strong spatial similarity to the original CAMS composites (Figure S5), confirming that the four regimes are also well captured by the satellite observations.

We also checked the typical time lag between CAMS and IASI+GOME2 classifications for each regime, computing how many days apart the two datasets agree on the same regime (Table S1). All four regimes show a low no-match rates ($\leq 1.7\%$)
 45 and good temporal agreement, with a median lag of about 0 days for R1 and about 1 day for the other regimes. The exact same-day agreement is stricter and lower overall (36–52%), showing that CAMS and satellite regimes do not always match on the exact same day.

Overall, these results confirm that the four regimes identified from CAMS are robust and consistently reproduced in the independent IASI+GOME2 satellite observations, both in space and in time.

Regime	Number Days	No Match Lag (%)	Median Lag (%)	Exact Match (%)
R1	126	0.0	0.3	39.6
R2	152	1.7	1.5	51.8
R3	110	0.7	1.2	52.2
R4	85	0.0	1.1	35.8

50 **Table S1: Temporal co-variability between CAMS and IASI+GOME2 classifications for each regime. Number Days indicates the number of IASI+GOME2 observations assigned to each regime. No Match Lag is the percentage of IASI+GOME2 days for which no CAMS day belonging to the same regime within ± 30 . Median lag expresses the median temporal lag (in days) between IASIGOME2 and the closest match with CAMS for the same regime. Exact Match indicates the percentage of IASI+GOME2 days with the exact same-day agreement with CAMS.**

55

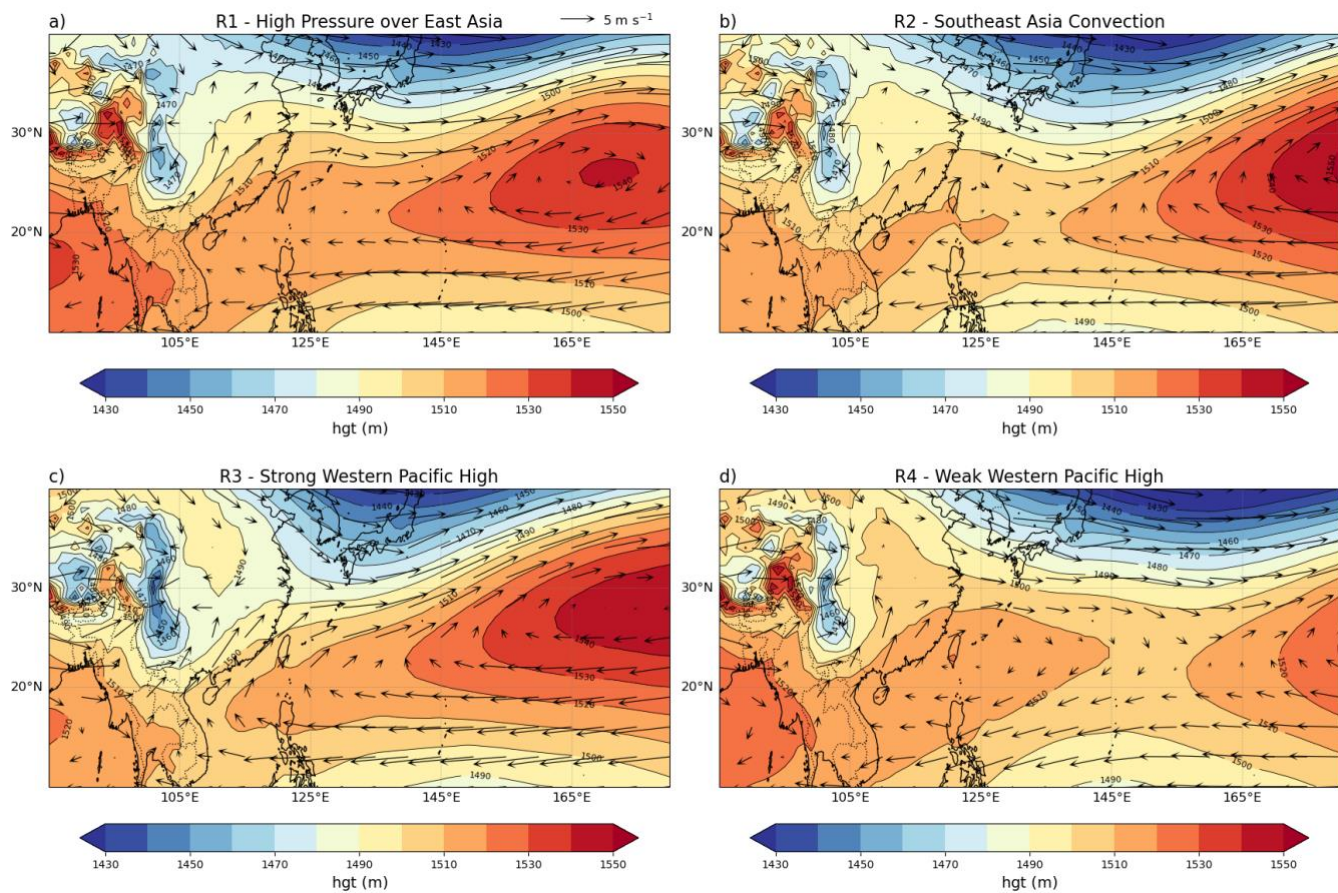


Figure S7: Geopotential height and wind fields at 850 hPa from ERA5 for the four ozone regimes: High Pressure (R1, panel a), Convection (R2, panel b), Strong WPSH (R3, panel c), and Weak WPSH (R4, panel d). Wind vectors represent both zonal and meridional wind anomaly components. The vector length indicates the magnitude of the wind anomaly, with the reference vector corresponding to 5 m s⁻¹.

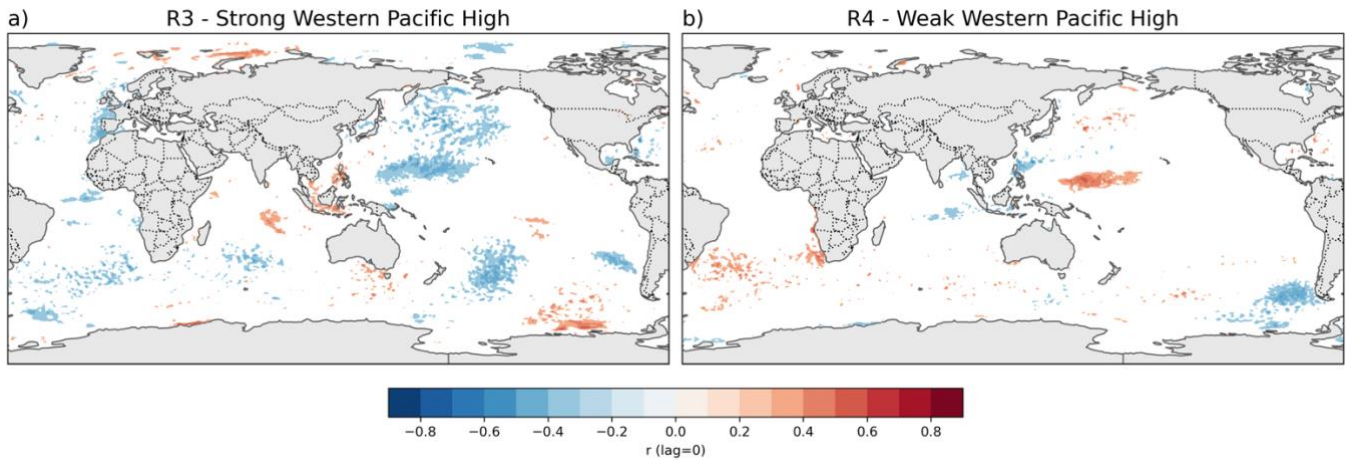


Figure S8: Global maps of the linear correlation between monthly SST anomalies and the monthly averaged occurrence of the ozone regimes Strong WPSH regime (R3, panel a) and Weak WPSH regime (R4, panel b). White areas indicate regions where the correlation is not statistically significant according to $p < 0.05$.

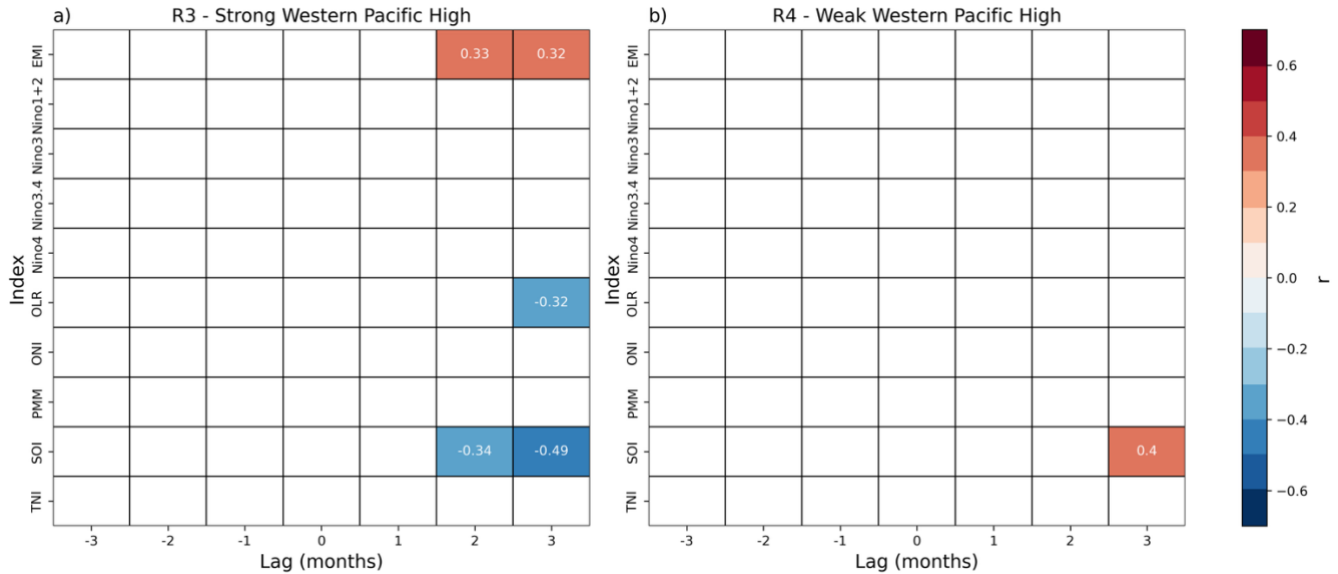


Figure S9: Lagged Pearson correlations between monthly frequencies of ozone regimes over East Asia and selected ENSO indices for (a) Strong WPSH regime (R3) and (b) weak WPSH regime (R1). ENSO indices include SST-based indicators (Niño1+2, Niño3, Niño3.4, Niño4, EMI, ONI) and atmospheric/coupled indicators (OLR, SOI, PMM, TNI). Only statistically significant correlations ($p < 0.05$) are shown; non-significant values are masked with white.

70 Figures S8 and S9 show that the correlation between ENSO and ozone regime strong WPSH regime and weak WPSH regime is not significant. This supplementary analysis is useful also for validating the methodology used. Although the Strong WPSH and Convection regimes show opposite spatial ozone patterns and may appear to be associated with opposite signs of

the same PCA spatial pattern, the absence of a robust ENSO influence in strong WPSH regime (contrary to what is observed for convection regime) support their interpretation as two different physical regimes.

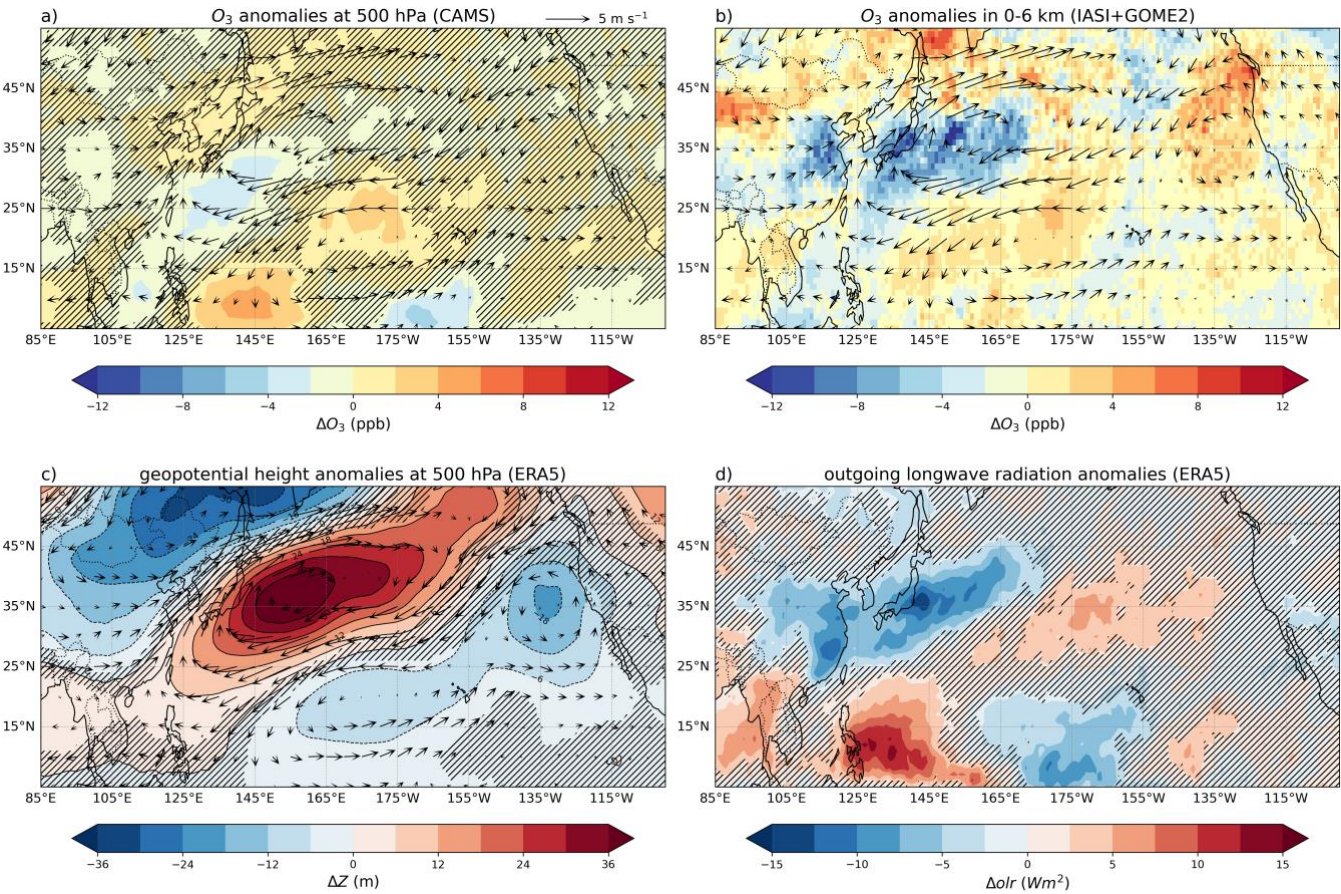


Figure S10: a) Ozone anomalies at 500 hPa (ppb) from CAMS for ozone regime Strong WPSH regime (R3). b) Columnar ozone anomalies in the 0-6 km layer (ppb) from IASI+GOME2. c) Geopotential height anomalies (m). d) Outgoing longwave radiation anomalies (W/m2) at the top of the atmosphere from ERA5. Arrows indicate zonal wind anomalies at 500 hPa from ERA5. Shaded areas indicating values significant at the 95% confidence level (T-test).

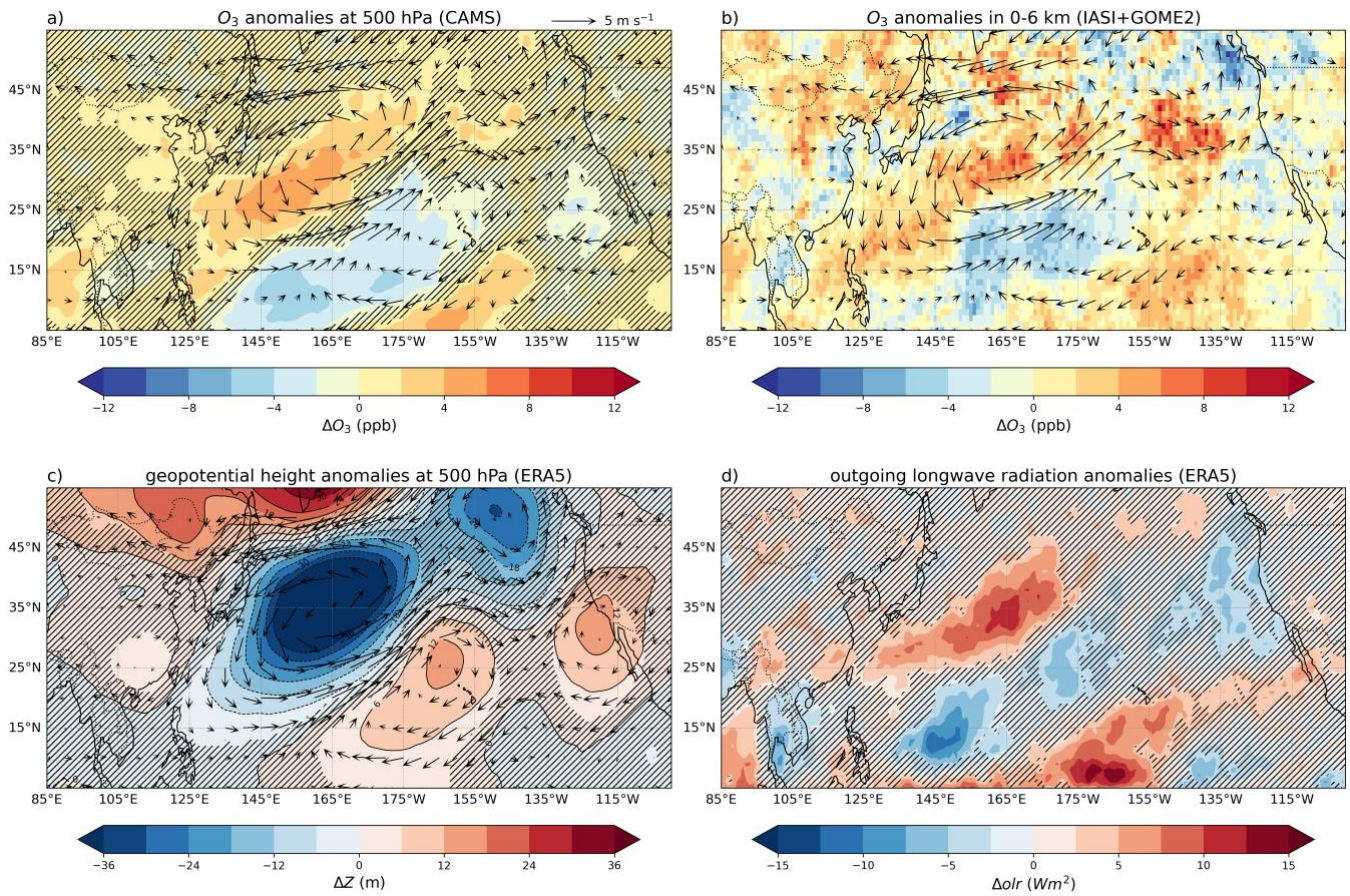


Figure S11: a) Ozone anomalies at 500 hPa (ppb) from CAMS for Weak WPSH regime (R4). b) Columnar ozone anomalies in the 0-6 km layer (ppb) from IASI+GOME2. c) Geopotential height anomalies (m). d) Outgoing longwave radiation anomalies (W/m²) at the top of the atmosphere from ERA5. Arrows indicate zonal wind anomalies at 500 hPa from ERA5. Shaded areas indicating values significant at the 95% confidence level (T-test).

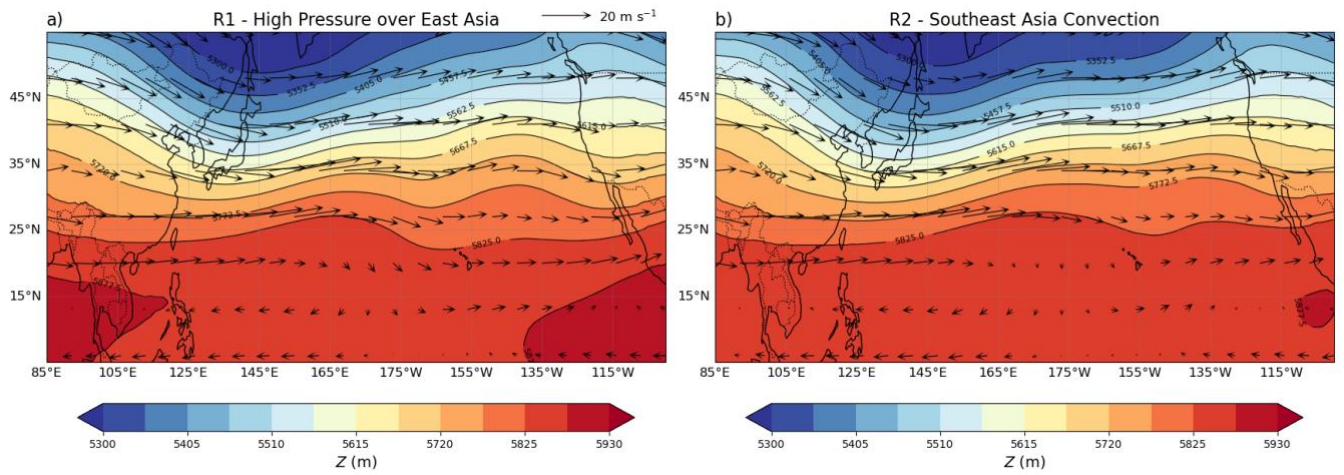


Figure S11: Geopotential height and wind fields at 500 hPa from ERA5 for the regimes: (a) High Pressure (R1), and (b) Convection (R2). Wind vectors represent both zonal and meridional wind components. The vector length indicates the magnitude of the wind anomaly, with the reference vector corresponding to 20 m s⁻¹.

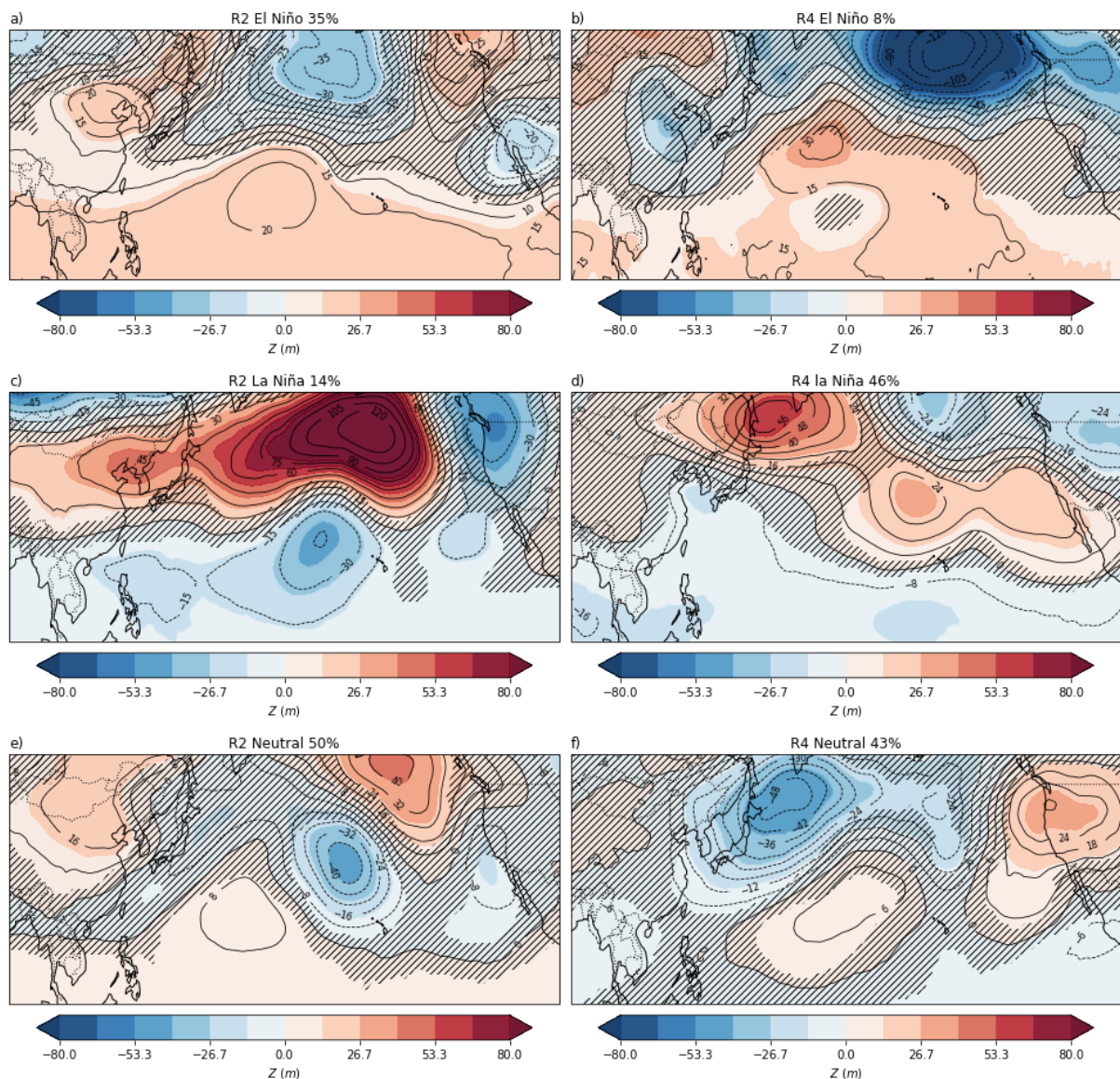


Figure S12: Geopotential anomalies at 500 hPa (m) from ERA5 for ozone regimes high pressure and convection during a) and d) El Niño, b) and e) La Niña and c) and f) Neutral conditions according to Niño 3.4 index. Shaded areas indicate non-significant values at the 95% confidence level (T-test). The percentages of occurrences of each regime under the different ENSO phases are indicated for each figure.

95 The circulation patterns associated with both high-pressure and convection regimes (Figures 14c and 15c) show wave-like structures, characterized by alternating positive and negative anomalies of geopotential height at 500 hPa across the North Pacific Ocean. Due to the influence of ENSO on the occurrences of these ozone regimes we analyse the role of ENSO in modulating these circulation patterns. Figure S13 shows anomaly composites of geopotential height at 500 hPa divided by ENSO phase (El Niño, La Niña and neutral conditions), based on the Niño 3.4 index. For both ozone regimes, the circulation

100 patterns show differences depending on the seasonal climate background. El Niño composites (Fig. S13a-b) exhibit some large-scale anomalies typical of the ENSO circulation, including positive geopotential height anomalies in the tropics and negative anomalies in high latitudes. However, the typical ENSO teleconnection over the North Pacific Ocean associated with a Rossby wave pattern extending from the tropical Pacific to North America is not completely represented. Instead, the El-Niño circulation observed during the ozone regimes is a combination of ENSO background anomalies and other high-
105 frequency anomalies. La Niña composites (Fig. S13c-d) are characterized by negative anomalies in the tropics and positive anomalies over the North Pacific. As for El Niño, the observed La Niña circulation differs from the typical ENSO response.

The reasons why these patterns do not fully represent the canonical ENSO teleconnection structures may derive from different factors: physically, spring is a period of decreasing ENSO magnitude, but also methodologically there is superposition of regime-related and ENSO-related circulation features. Anomalies are stronger for convection regime during
110 El Niño and high-pressure regime during La Niña because the low occurrences, 14% and 8% respectively.

Neutral conditions composites are very similar to those obtained using all occurrences combined, indicating that the structure of both regimes is largely influenced by neutral ENSO years, which account for about 50% of the high-pressure regime occurrences and 43% of the convection regime. However, high-pressure regime pattern looks to be explained by neutral circulation anomalies, with El Niño (35%) strengthening the positive anomalies in the subtropics. For convection regime, the
115 neutral-condition circulation anomalies are influenced by La Niña (46%) in the tropical and mid- latitudes.

To conclude, these results suggest that ENSO plays a role in modulating the low-frequency circulation associated with ozone regimes, affecting their amplitude and spatial distribution. The ozone regimes are identified independently of ENSO phases, their occurrences are statistically linked to ENSO variability, then the observed circulation anomalies do not coincide with the North Pacific ENSO teleconnection pattern.