Reply to Referee 2:

The manuscript compares a relatively short (2015 – 2019) observational record of June-July-August average ozone over eastern China to output from the GEOS-Chem model. With the ability of GEOS-Chem to adequately simulate ozone demonstrated, the authors analyze an extended GEOS-Chem simulation over 1980 – 2019 driven by MERRA-2 but using constant anthropogenic emissions. Through EOF analysis, they find a dipole in ozone variability (DP-O_3), with centers of opposite sign over North China (NC) and the Pearl River Delta (PRD). The interannual variability in DP-O_3 is then linked to variations in May sea-ice extent in the Arctic ocean to the north of Svalbard and Franz Josef Land (SI_{FJL}). Variability in sea-ice gives rise to anomalous Rossby wave forcing that propagates downstream and produces meteorological conditions over China that modify the photochemical production and accumulation of ozone. The authors similarly make a link between DP-O_3 and variability in sea surface temperatures over the Indian Ocean, the Subtropical Indian Ocean Dipole (SIOD). These links between SI_{FJL} and SIOD and meteorological conditions over China are then reinforced through analysis of the variability in a large ensemble (40 members) of the CESM model. To close, the SI_{FJL} and SIOD indices are combined to create a new index (SEI) that explains a larger percentage (~40%, r=0.62) of the variance in DP-O_3.

The paper is generally well organized and the development of the statistical relationships are easy to follow. I do not have a strong background in statistical analysis but the use of correlation and composites seems solid, particularly for derived impacts of sea-ice variability which has parallels with similar hypothesized links between sea-ice and interannual variability of the east Asian monsoon. The effects of the Subtropical Indian Ocean Dipole SST anomalies in January-February-March are statistically weaker and the authors suggest a mechanism involving subsurface heat content anomalies that migrate across the Indian ocean towards Sumatra by June-July-August that seems speculative. Overall, though, **I only have minor comments that are itemized here.**

1. Lines 89 – 90: It might be helpful for the reader if the mention of the fifth generation ECMWF reanalysis also included the phrase ‘ERA5’, as it is frequently referred to.

Reply:

According to the advice of the other reviewer, the data has been replaced with **MERRA-2 data** to enhance the consistency of data.

This phrase, i.e., MERRA-2, has been used in the manuscript.

Revision:

p. 3, line 89: The meteorological fields data with a horizontal resolution of

0.5° latitude by 0.625° longitude for the period 1980–2019 were taken from the **MERRA-2 dataset** (Gelaro et al., 2017), including geopotential height at 500 hPa (Z500) ......

2. Lines 123 – 126: The spatial correlation coefficient is definitely a significant aspect of assessing model performance, but since the paper is about the dipole pattern the temporal behavior will be important. Can the authors present one or two widely used metrics, such as root mean square error of MDA8 ozone for the 2015 – 2019 period? Preferably, this would be presented separately for both the NC and PRD regions.

Reply:

In the revised version, two widely used indicators, **root mean square error / mean and mean absolute error (MAE)**, were used to evaluate the simulation performance of GEOS-Chem model.

Compared the simulated and observed summer mean MDA8 O_3 concentrations in NC and the PRD, which had a low bias with a **MAE** of **5.7 μg m^{-3} and 12.1 μg m^{-3}** in the PRD and NC, respectively. The values of **root mean square error / mean** were **15.8 % and 8.1 %** in NC and the PRD, respectively, indicating the good performance of reproducing the O_3 concentration.

Revision:

p. 5, line 132: …… Compared the simulated and observed summer mean MDA8 O_3 concentrations in NC and the PRD, which had a low bias with a mean absolute error of **5.7 μg m^{-3} and 12.1 μg m^{-3}** in the PRD and NC, respectively. The values of root mean square error / mean were **15.8 % and 8.1 %** in NC and the PRD, respectively ……

3. Line 143: '…three years with the lowest and highest simulated SI in 143 each member.' Was sea-ice for a particular region, such as the Barents sea mentioned earlier, used or was it overall Arctic sea-ice extent? Was a particular month used?

Reply:

Based on the conclusions in the manuscript, the three years with the lowest and highest **SI anomalies near the Franz Josef Land in May** and **SIOD anomalies in January–February–March** were selected. We modified the manuscript to avoid ambiguity.

Revision:

p. 6, line 153: …… composite analyses were conducted based on the three years with the lowest and highest simulated **preceding climatic variability for a particular month** in each member.

4. Line 168: Just to clarify, the timeseries of the DP-O_3 timeseries shown in panel (a) of Figure 2 is the JJA average of the EOF each year? But the EOF is calculated daily?

Reply:

EOF is calculated using an **annual summer average**. The EOF approach was applied to summer-mean O_3 for the period 1980–2019.

Revision:

p. 7, line 179: The time series of DP-O_3 showed a strong interannual variation ……

5. Lines 184 – 185: The caption on Figure 2 should have some brief description of what SI_{FJL}, SIOD and SEI are. I know these are all described in detail a bit further down in the text, but a brief description to give the reader some idea of what is being presented is helpful.

Reply:

In the revised version, we have added some brief description of SI_{FJL}, SIOD and SEI in the caption on Figure 2.

Revision:

p. 8, line 201: Figure 2. Variations in standardized DP-O_3 time series (black), **May SI near the Franz Josef Land** (SI_{FJL}, red), **January–February–March mean Subtropical Indian Ocean Dipole** (SIOD, blue), and synergistic effects index (SEI, green) from 1980 to 2019. **SEI defined as the weighted average of SI_{FJL} and SIOD**. The correlation coefficients of the DP-O_3 with SI_{FJL} (red), SIOD (blue), and SEI (green) were shown in the figure.

6. Lines 187 – 191: Some of the acronyms used in the caption for Figure 3 are easily

understood, but others like Ssr and MIcc are not.

Reply:

Some explanations have been **added to the acronyms** used in the caption for Figure 3 to make it easier to understand.

Revision:

p. 9, line 206: Figure 3. Composite summer atmospheric circulations associated with the DP-O_3 (DP-O_3P minus DP-O_3N) for the period 1980 to 2019, including (a) **surface air temperature** (**SAT**, unit: K, shadings) and geopotential height at 500 hPa (unit: gpm, contours), (b) **surface incoming shortwave flux** (**Ssr**, unit: W m^{-2}, shadings) and **low and medium cloud cover** (**Mlcc**, unit: 1, contours), and (c) **precipitation** (**Prec**, unit: mm, shadings) and surface wind (unit: m s^{-1}, arrows) ……

7. Lines 195 – 196: Was there any detrending of the sea-ice anomalies performed before the calculation of SI_{FJL}?

Reply:

We removed the trend before calculating all indices and, to be sure, the trend of sea ice anomalies is also **removed** before the calculation of SI_{FJL}.

8. Line 249: The presentation of May SST anomalies in panel (a) of Figure 5 show positive anomalies extending well north of 80N. Isn't there almost always sea-ice present in this area?

Reply:

Thank you for pointing out our mistake that the missing value was not properly handled when using SST data. The improper control of data quality was revised throughout the manuscript. SST in the Barents and Kara Sea (black box in Figure 5) is used to calculate sea-ice related SST indices. Therefore, the **negligence does not affect our conclusions** because we did **not use SST data north of 80°N**.

Revision:

p. 12, line 279:

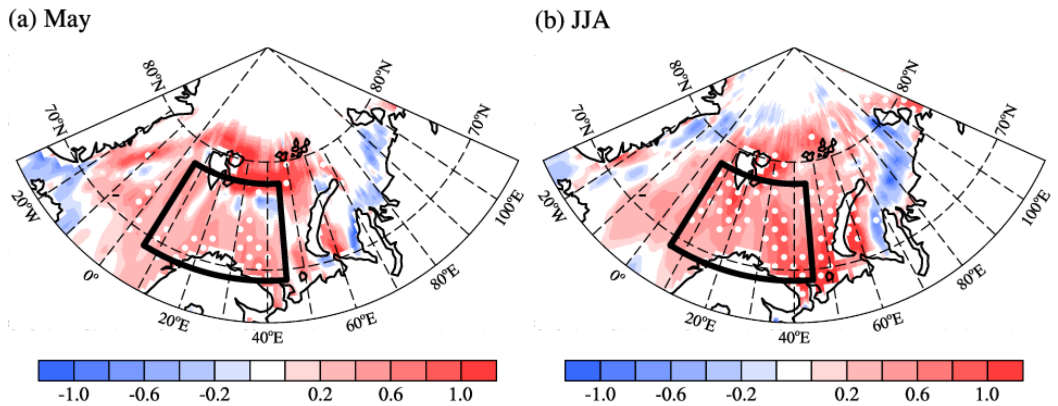

Figure 5. Composites of (a) May Arctic SST (unit: K) associated with SI_{FJL} index (negative SI_{FJL} years minus positive SI_{FJL} years) from 1980 to 2019. The back box in (a) and (b) represents the center of the SST associated with SI_{FJL}. The white dots indicate that the composites with shading were above the 90% confidence level.

9. Lines 272 – 273: Was there a range of years from which the three highest and lowest SI_{FJL} anomalies were taken from the CESM large ensemble?

Reply:

The period 1980–2019 was used to obtain three years with the highest and lowest SI_{FJL} anomalies. To make this clear, we added a range of years.

Revision:

p. 13, line 302: The relationship between the preceding May SI anomalies and the JJA EU-like pattern was also confirmed by large ensemble simulations of CESM **during 1980–2019**.