# Revising the Solar Cycle Response in Stratospheric Ozone Using an Ensemble of Regression Models

Sandip Dhomse, Martyn Chipperfield, Wuhu Feng (Uni. Leeds)
Ryan Hossaini (Uni. Lancaster)

Introduction

I. Large uncertainty in our understanding of Solar Cycle Signal (SCS) in stratospheric ozone.

II. Our understanding has been based on SAGE II satellite instrument data (low temporal sampling).

III. MLS satellite instrument now provides about 16 yrs of continuous ozone profile observations.

IV. Historically, only Ordinary Least Square (OLS) regression models are used to estimate SCS.

V. Explanatory variables are generally correlated and finding optimum/suitable variables is tricky.

VI. Here we use ensemble of four regression models (OLS, Lasso, Ridge, ElasticNet) to estimate SCS.

VII. Both MLS and TOMCAT 3-D model SCS are consistent throughout the stratosphere.

VIII. Model simulations with SORCE solar irradiances shows better agreement with MLS-derived SCS.

11-year Solar Cycle in Sunspot Numbers
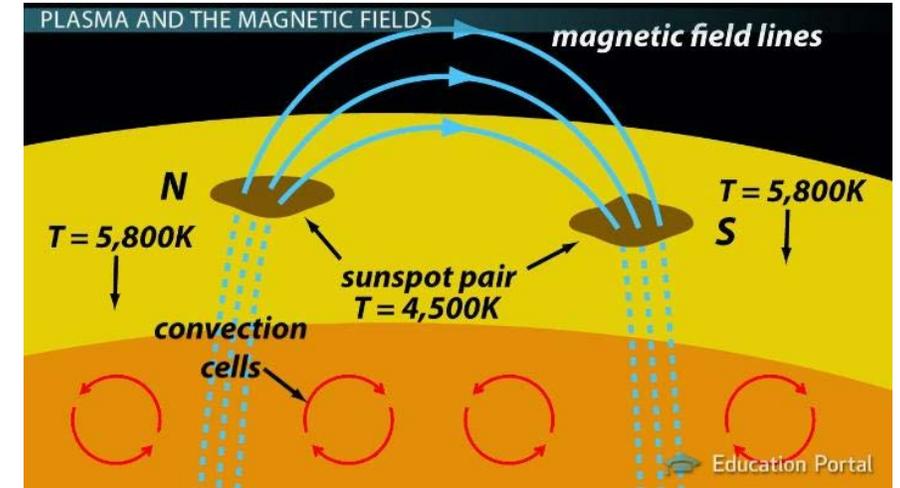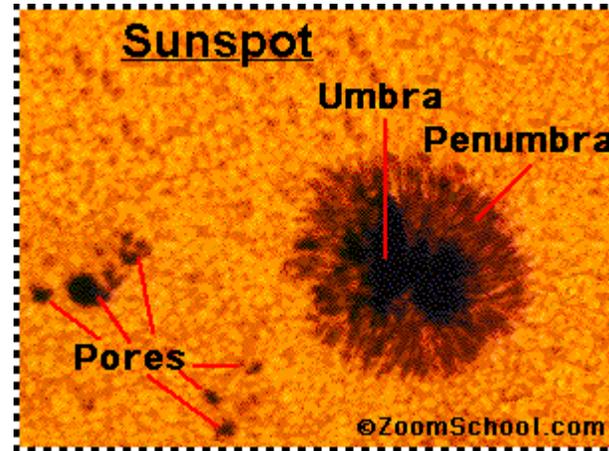
- Caused by Sun's magnetic field's strong torque due to rotating plasma.

- Lasts for few days to months.

- 11-year cycle causes very little variation in total solar irradiance.

- However, large changes in UV part of the spectrum, hence stratospheric ozone changes provide helpful guide to find solar variability – climate interaction.
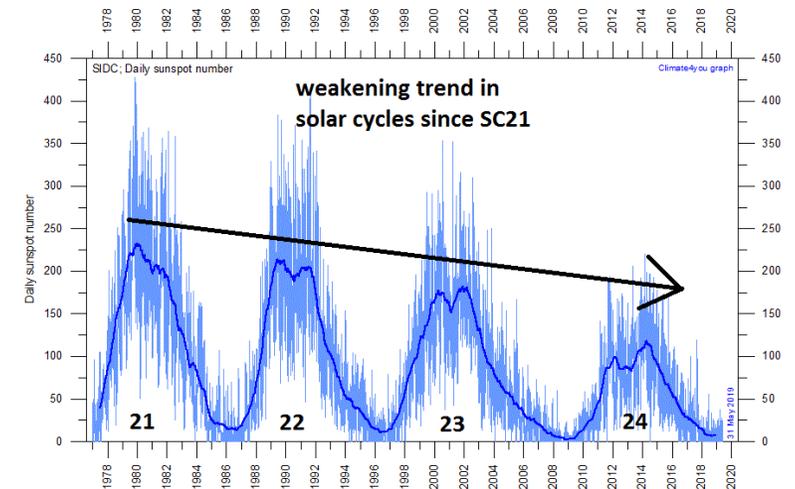
# Ambiguous nature of solar cycle Signal (Dhomse et al., GRL, 2016)
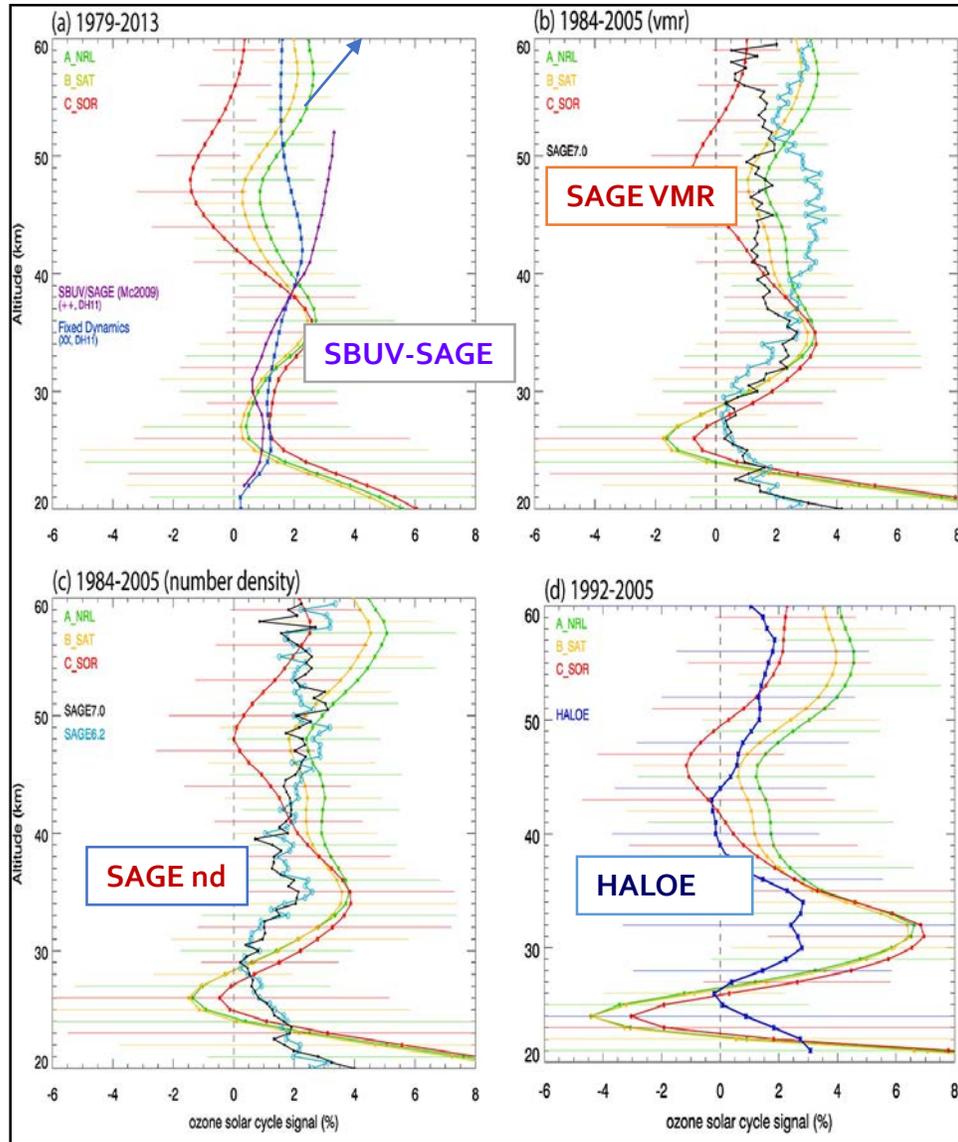
- SBUV data suggested up to 3% SCS near stratopause (Hood & Soukharev, 2006)

- Model and HALOE shows much weaker SCS in the upper stratosphere (Remsberg et al., 2009; Dhomse et al., 2011)

- SAGE retrieves O_3 number density (nd) and shows larger SCS (Dhomse et al., 2016)

- If nd is converted to mixing ratio (vmr) then SCS is a bit weaker

- Different time period → different signal

- SORCE satellite solar fluxes suggested negative SCS in the upper stratosphere (Haigh et al., 2010), but we now know that initial versions of both SIM and SOLSTICE had issues due to instrument degradation

Historically Ordinary Least Square (OLS) Regression models are used to estimate signal but 'one size' does not fit all


Simple Linear Regression Model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Dependent Variable — Y_i
Population Y intercept — β_0
Population Slope Coefficient — β_1
Independent Variable — X_i
Random Error term — ε_i

$\underbrace{\beta_0 + \beta_1 X_i}_{\text{Linear component}}$ $\underbrace{\varepsilon_i}_{\text{Random Error}}$

	Underfitting	Just right	Overfitting
Symptoms	• High training error • Training error close to test error • High bias	• Training error slightly lower than test error	• Very low training error • Training error much lower than test error • High variance
Regression illustration			
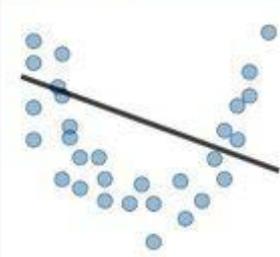
- Multivariate regression models are widely used to quantify the influence of a particular process

- In the stratosphere (~10-50 km) ozone concentrations are controlled by different processes at different altitudes/latitudes, but almost all previous studies used same regression model for all locations

- Complicated to find suitable proxies that control ozone variations at all the levels

Why Lasso (L1) and Ridge (L2) Regression Models?
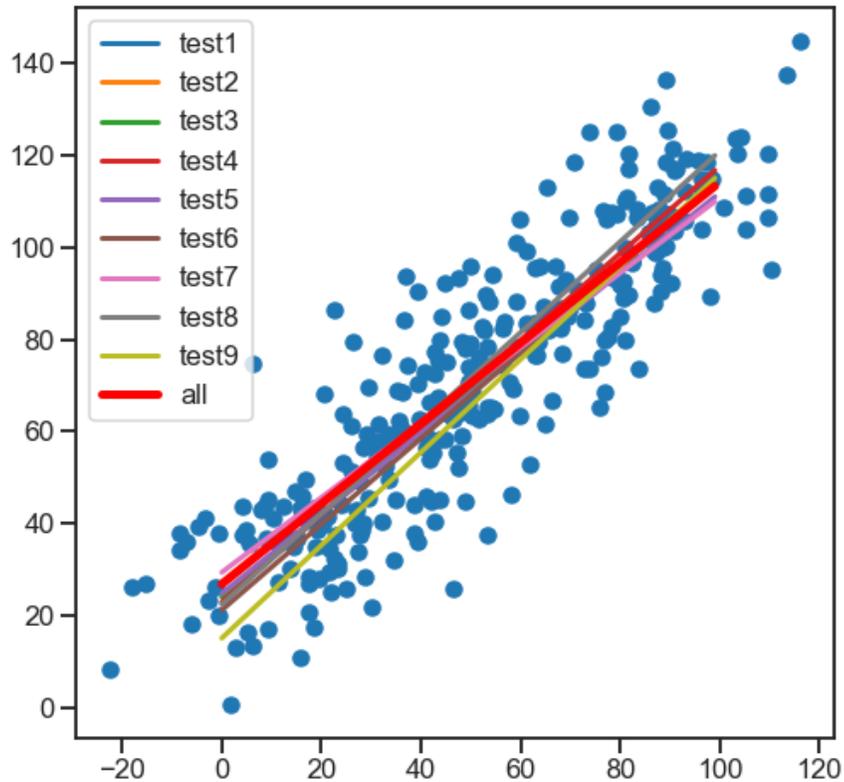
For trend analysis, generally we use OLS linear regression or BLUE (Best Linear Unbiased Estimates)

$Y = constant + m1x1 + m2x2 + m3x3 + errors$, $m(i)$ are regression coefficients and $x(i)$ are explanatory variables

Most atmospheric processes are correlated. Hence if we use different combinations of data points, regression coefficients will not be unbiased (or Gaussian)

e.g. 90% training data and 10% testing data

→ Linear regression model: minimum residuals but high variance (coefficients vary if data points change)
→ Lasso regression: tries to reduce (Residual square) + (alpha) x slope
→ Ridge regression: tries to reduce (Residual square) + (alpha) x slope^2
→ ElasticNET: is combination of both

Data and methodology

- Data : MLS level 2 (daily profile data) v5 data from

    https://search.earthdata.nasa.gov/search?q=ML2O3_005

- Model setup is similar to *Dhomse et al.,* (Nature Comm. 2019) with ERA5 dynamical forcing

- Three simulations are performed with NRL v2, SATIRE and SORCE solar irradiances (2005-2020)

- MLS L2 data is binned onto TOMCAT latitude bins (2.8^o)

- Calculate zonal mean monthly mean percentage anomalies 2005-2020 (model and MLS)

- Multivariate regression model is modified version of that used in Dhomse et al., (2016)

    dOzone = linear trends (12) + QBO terms (Q30 and Q50, 24) + Age of Air (12) +
    + Solar + SOI + AO + AAO   (total 52 terms)

- QBO, Southern Oscillation, AO and AAO indices from Climate Prediction Center, Solar (Mg ii index) from IUP Bremen, Age-of-Air is from TOMCAT simulation

- Except trend terms, all the explanatory variables are detrended and normalised

Model vs MLS ozone comparison

polar O3 (63N-90N) 480 K
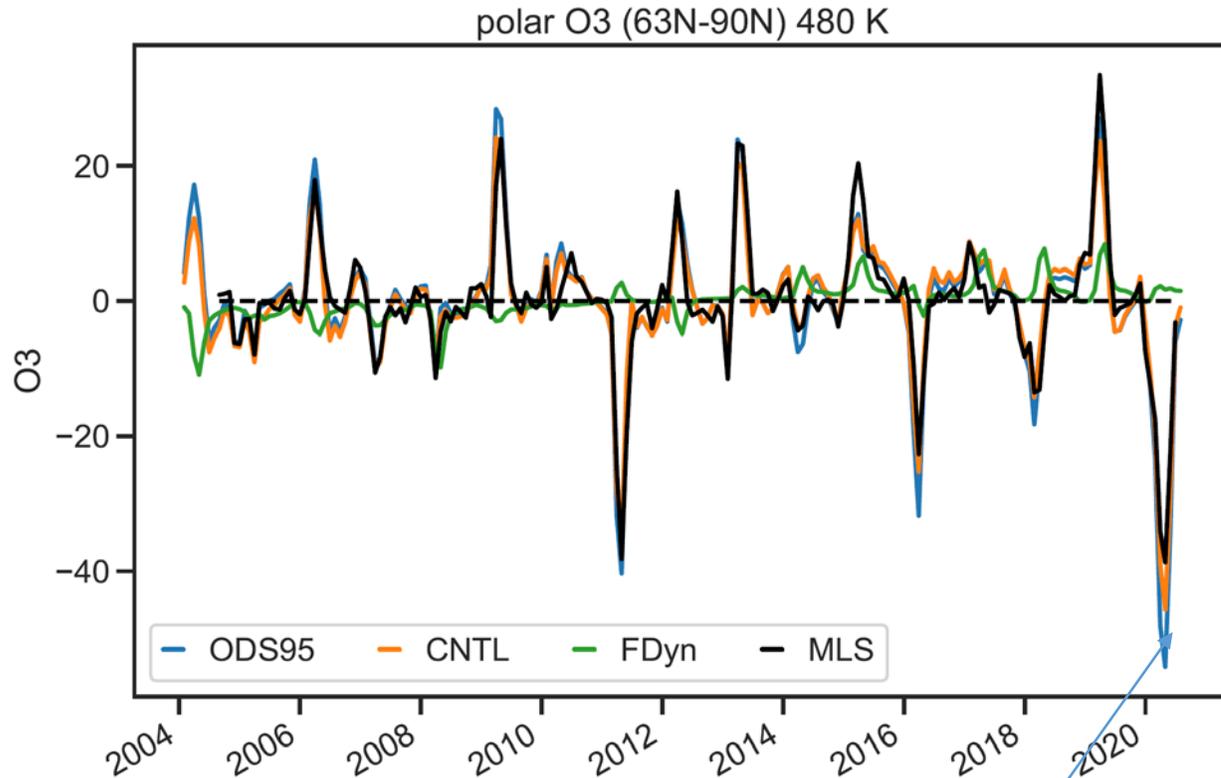
Ozone hole in the Arctic 2020
Feng et al., (GRL, 2021)
Weber et al., (JGR, 2021)
Bognar et al., (JGR, 2021)

- Model ozone variability agrees better with satellite and ground-based ozone measurements
- Model simulations are performed using NRL v2 solar flux data
- Model is forced with ERA5 reanalysis meteorology
- Fixed dynamics (FDyn) means repeating annual dynamical fields are used to force the model
- Modelled ozone anomalies show excellent agreement with MLS data
- Observed ozone depletion in 2011 and 2020 is well captured by the model

Multivariate linear regression model (e.g. OLS Regression- MLS)
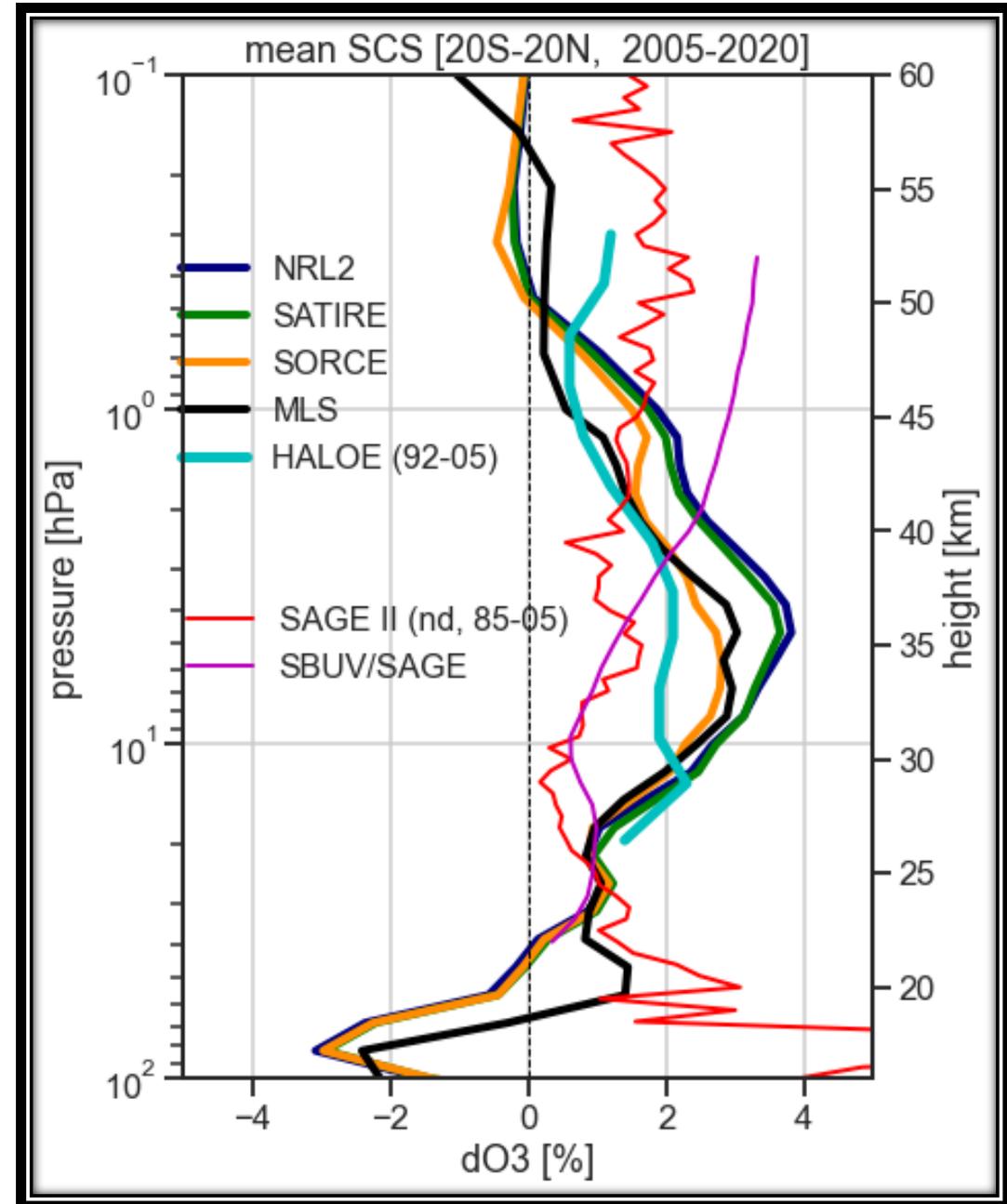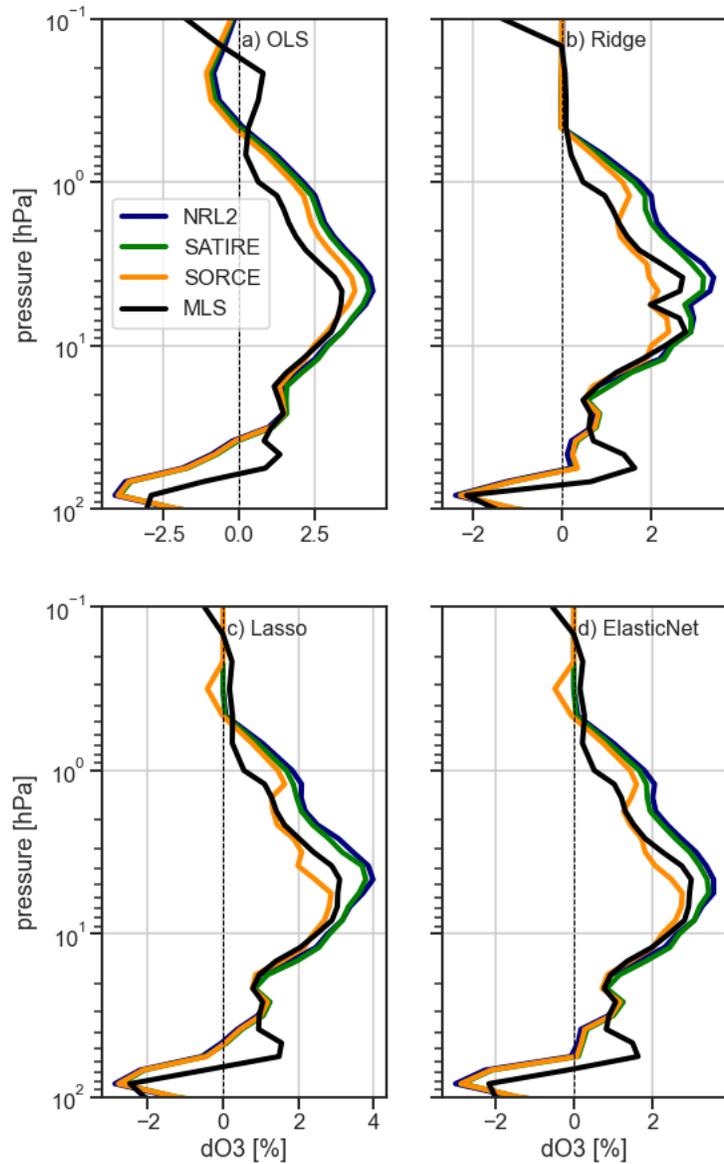

Lat = 1.5 N

Residuals -20
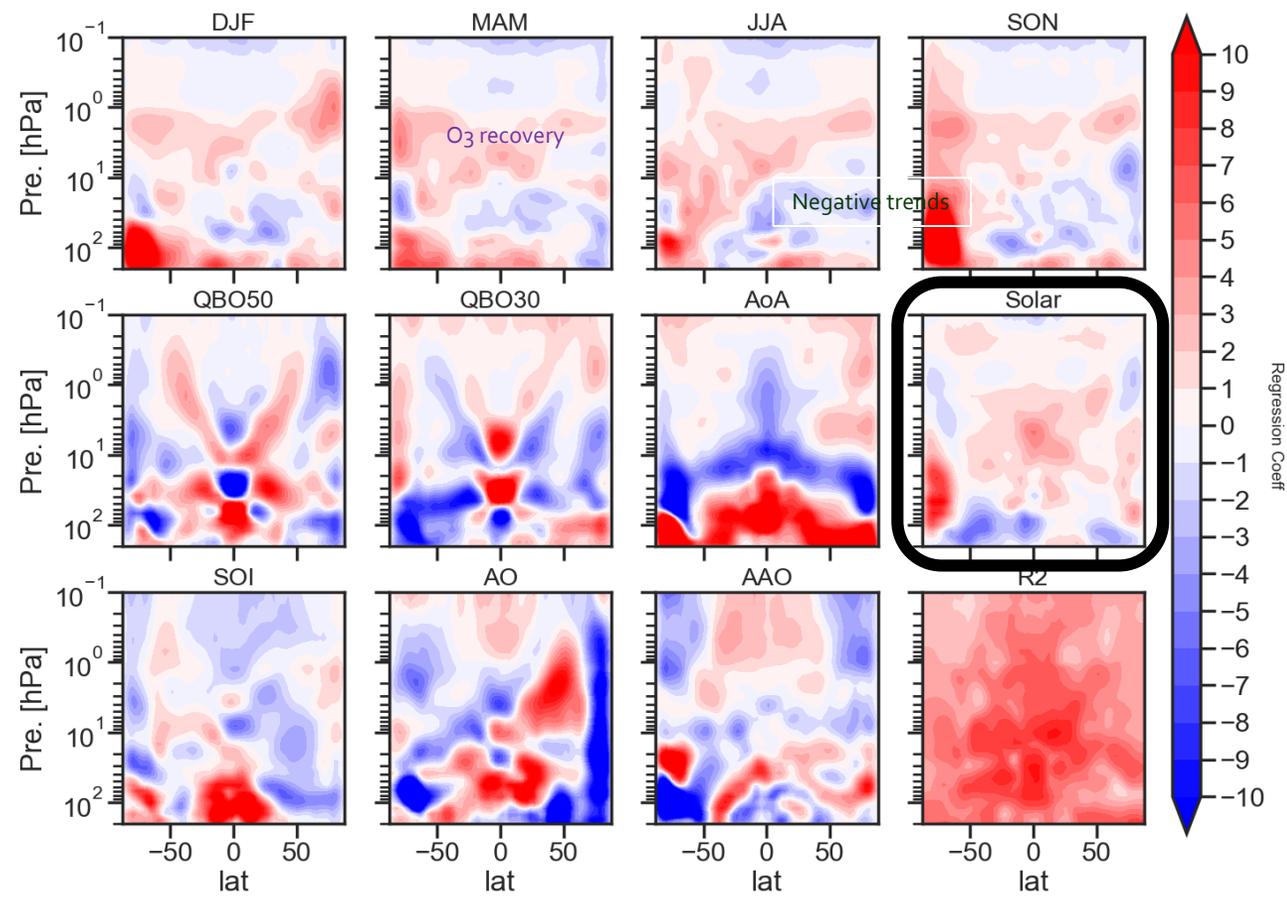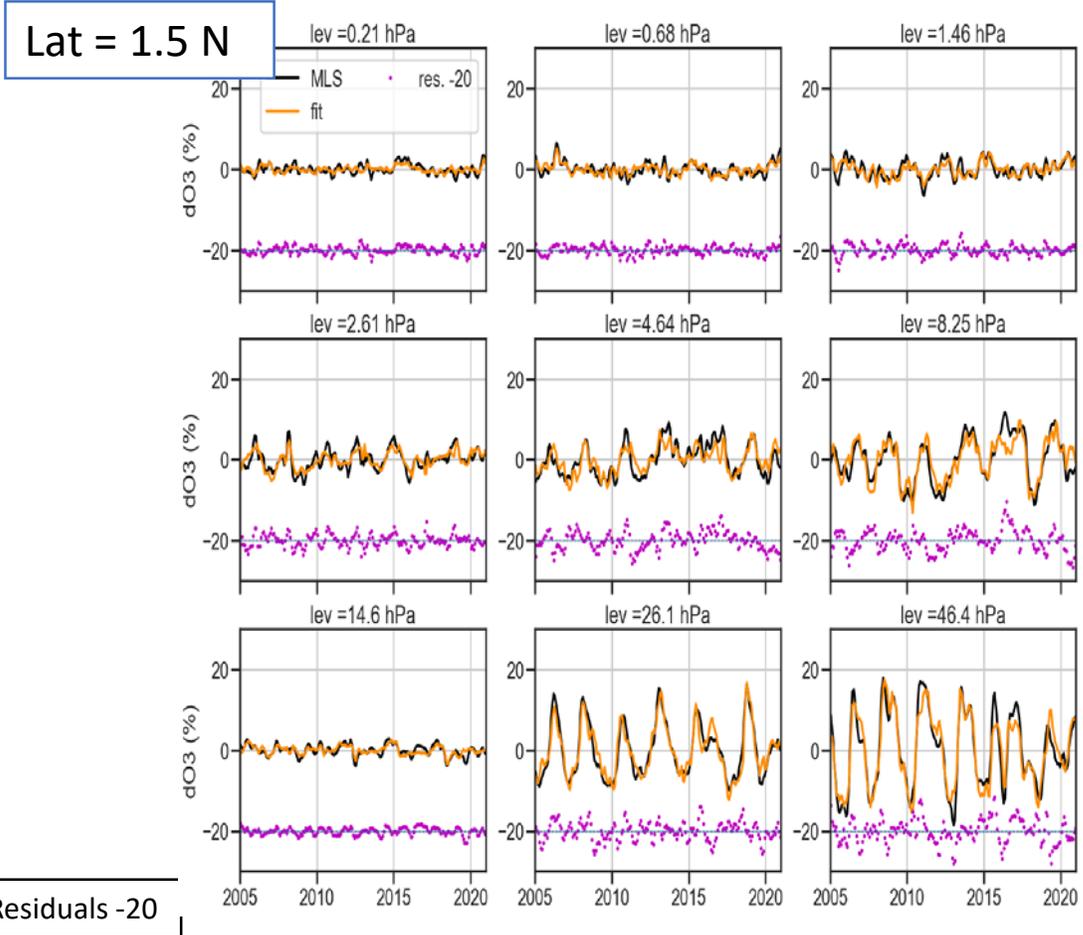
Regression model performs better (R2 values) than one used in Dhomse et al., (2011, 2013, 2016)

Multivariate linear regression model (e.g. OLS Regression- TOMCAT NRL)
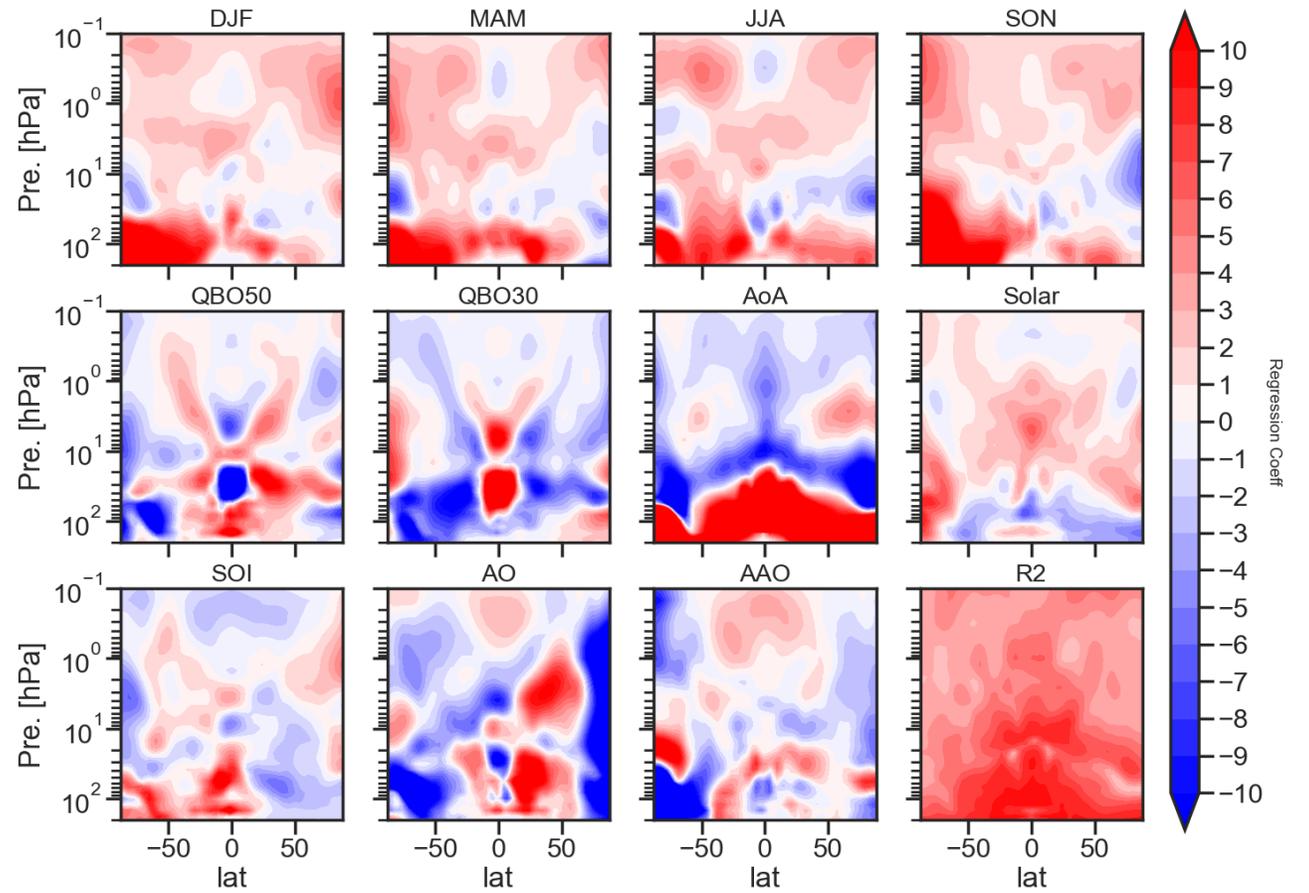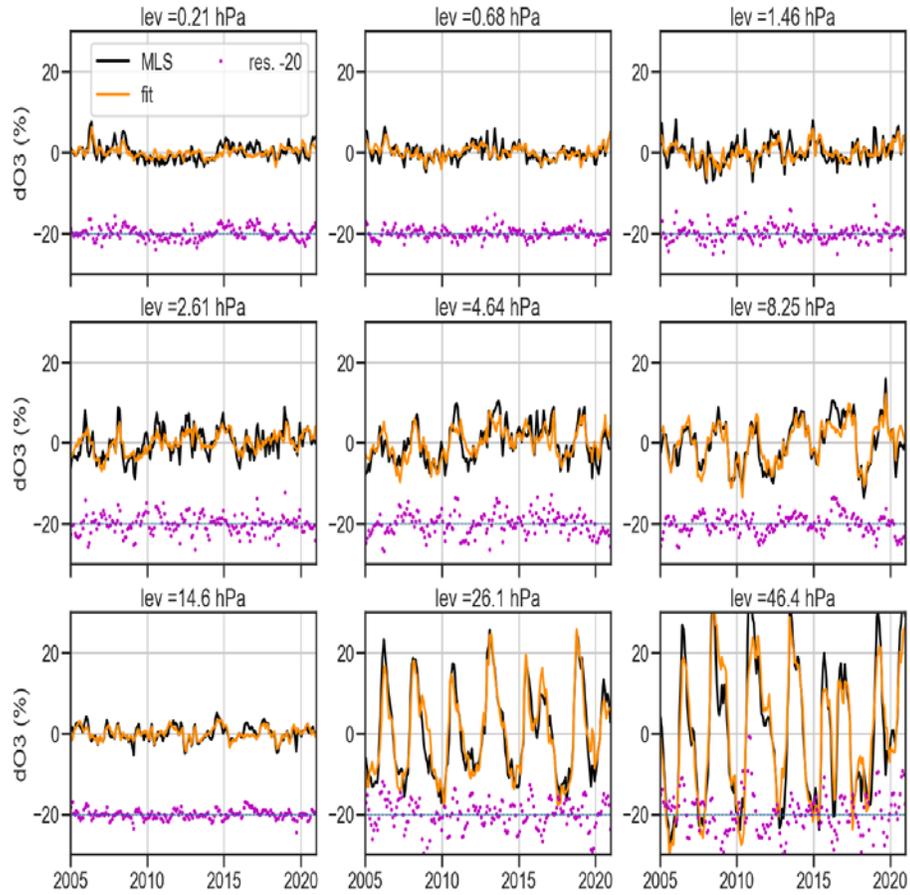
SCS – Tropical stratosphere

Summary and Outlook

- We have used various forms of linear regression models to (OLS, Ridge, Lasso) to estimate SCS from MLS satellite data and TOMCAT CTM simulations.

- For stratospheric transport, we use age-of-air as a new dynamical proxy.

- Updated regression model seems to perform better than previous studies.

- Nearly all the regression models show consistent SCS throughout the stratosphere.

- Updated analysis shows significantly different SCS compared to previous studies.

- MLS and TOMCAT-simulated ozone show almost 4% SCS in the middle stratosphere and negligible SCS in the upper stratosphere.

- Our analysis suggests significantly different SCS compared to SAGE/SBUV data.