Simplified rules for SAGE II ozone data usage

Stefanie Kremser[1], Larry Thomason[2] and Leroy Bird[1]

[1] Bodeker Scientific, Alexandra, New Zealand; [2] NASA, Langley Research Center, Hampton, USA

Motivation

While space-based measurements provide consistent, global, long-term measurements of radiatively active trace gases and aerosols, they are indirect (based on optical properties of their targets) and therefore confounded by non-target gases and particulate matter. As a result, our ability to accurately detect small, but important, changes in concentrations of constituents, such as ozone, is challenged by measurement uncertainties. SAGE II observations of ozone are a key part of understanding long-term trends. Since its mission began in 1984 a plethora of 'generally accepted' screening methods have been developed, tailored to each measurement system and to each anticipated use of the data. These methods are often inconsistent, ad-hoc, and untraceable and seldom revised even after significant revisions to the data themselves. Here we will present the newly developed and simplified SAGE II ozone data usage rules which are based on how the measurements were made.

Figure 1: A schematic diagram showing the solar occulation viewing geometry which is exploited by SAGE II. Z_t is the tangent altitude.

SAGE II ozone measurements

SAGE II measurements are the longest continued time series of ozone - 20 years of measurements performed by one instrument and are widely used in trend analyses. Measurements are derived from line-of-sight (LOS) transmission profile measurements from 0.5 to 100 km, which are based on solar occultation observing technique (Fig. 1). Ozone data quality is affected by (i) the noise in the LOS measurement, (ii) the ability to separate the ozone contribution from the total contribution, and (iii) the amount of ozone and aerosols above the tangent altitude along the path.

To avoid artefacts, outliers and inconsistencies in the SAGE II data, which can lead to uncertainties in trend estimates, data usage rules are necessary before these data are used in any application.

Generally accepted data usage rules

1. Exclude all values between 23.06.1993 and 11.04.1994 between 15km and 50km if the error is bigger than 10%[1].
2. Wang et al.4 suggested to remove all data if error is greater than 300%. This rule was later adapted to only exclude ozone values above 35km if error is bigger than 300%[2,3,5].
3. Exclude all ozone values below or at 35km if error is bigger than 200%[2,3,5].
4. Exclude ozone values below or at the level where 525 nm extinction > 1×10^{-3} km^{-1} and the ratio 525nm extinction/1020nm < 1.4.[2,3,5].
5. Exclude all values between 30 and 50km where the uncertainty is > 10%[2,3,5].
6. Exclusion of all data points at altitude and below the occurrence of an aerosol extinction (525 nm) value of greater than 6×10^{-3} km^{-1} [2,3,5].
7. Eliminate all data below 23 km between July 1991 and December 1993 for excessive aerosol[1,2].
8. Exclude all values between 10.5 and 24.5 km if ozone > 10 ppm. This rule removing large values which was based purely on visual inspection[1,6].
9. Exclude all values above 25 km if ozone > 100 ppm. This rule is removing large values which was based purely on visual inspection[1,6].
10. Exclude all value at pressure < 3 hPa if ozone > 50 ppm. This rule is removing large values which was based purely on visual inspection[1,6].
11. Outlier screening by removing all values that are more than 10σ away from the monthly mean value for a given latitude band (15° zones), longitude (90° quadrants) and altitude (0.5km grid)[1,6]. This rule was later adapted to remove values that are farther than 3σ away from the mean in 10° latitude bins[2,3].

New data usage rules

1. Remove all data where the relative uncertainty is **exactly** 200% which is based on an effort to capture bad data due to enhanced aerosol.
2. Remove all data where the LOS optical depth due to aerosol is greater than 3.
3. Last rule is based on a basic outlier test for skewed data[7].

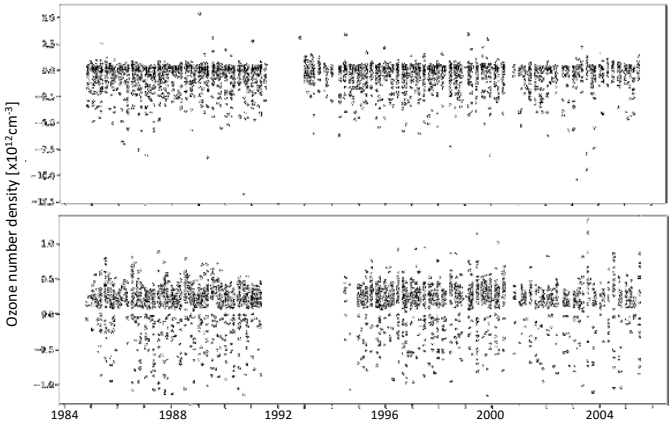

Figure 2: Ozone number density at 16 km altitude between 5° and 15°N. Top: SAGE II ozone; Bottom: Ozone after old rules were applied.

Simplified rules for SAGE II ozone

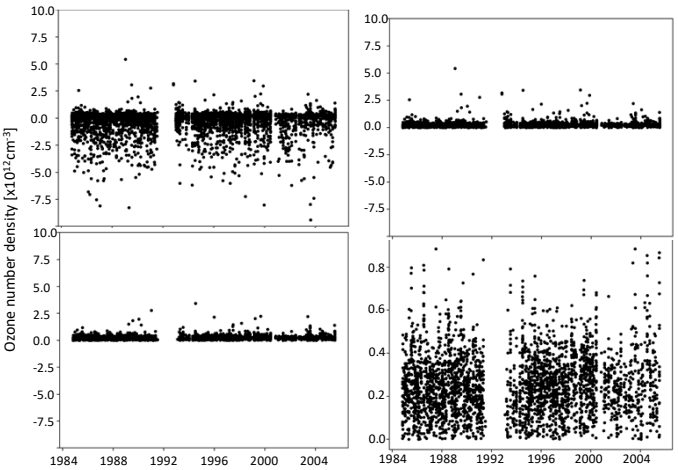

Figure 3: Ozone number density as measured by SAGE II at 16 km between 5°N to 15°N. Top left: SAGE II ozone data; Top right, SAGE II ozone after applying new rule #1; Bottom left: SAGE II ozone after applying rule #1 and #2; Bottom right: SAGE II ozone data that remain after applying all 3 new rules.

Comparison/Conclusion

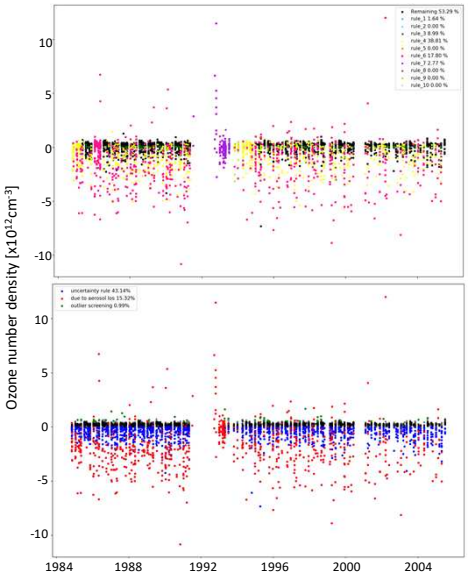

Figure 4: Ozone number density at 16 km altitude between 5°S and 5°N. Top: Old rules applied, Bottom: New rules applied to SAGE II ozone data.

- Fig. 4: new and old rules remove about the same amount of data, but differences are apparent:

	Old Incl.	Old Rem.
New Incl.	38.4%	15.7%
New Rem.	13.8%	32.0%

- In other cases, the new rules lead to more data points available for further study; especially during the Pinatubo period.
- Relative error rules can lead to biases in the data (Fig. 2).
- New rules focus on how measurements were made, i.e. they reflect that ozone measured that one altitude depends on everything above it.

References:

[1] Hassler , B. et al., Technical Note: A new global database of trace gases and aerosols from multiple sources of high vertical resolution measurements, *ACP*, 2008; [2] Hassler, B. et al., An updated version of a gap-free monthly mean zonal mean ozone database, *ESSD*, 2018; [3] Davis, S.M. et al., The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database: a long-term database for climate studies, *ESSD*, 2016; [4] Wang, H.J., Assessment of SAGE version 6.1 ozone data quality, *JGR*, 2002; [5] SAGE II release notes (https://eosweb.larc.nasa.gov/project/sage2/sage2_release_v7_notes); [6] Rind, D. et al., A complementary analysis for SAGE II data profiles, *GRL*, 2005; [7] Hubert, M. and Van der Veeken, S., Outlier detection for skewed data, *J. Chemometrics*, 2008.