SPARC
Stratosphere-troposphere Processes And their Role in Climate

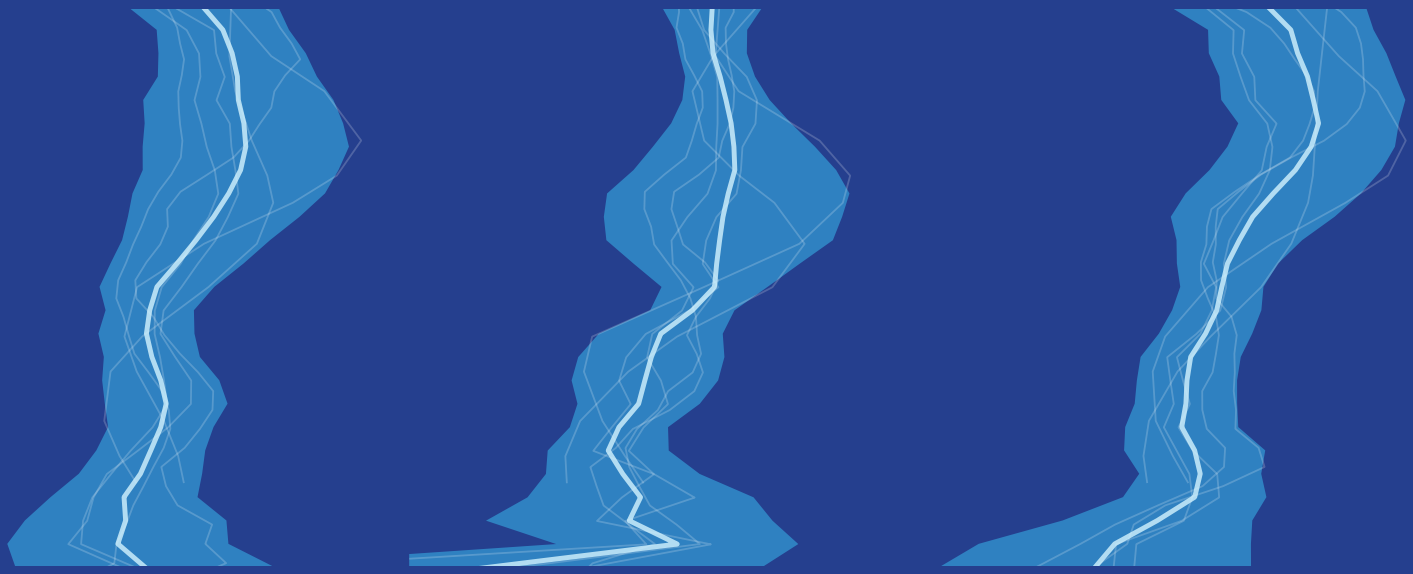
Core Project of the WMO/ISC/IOC
World Climate Research Programme

GLOBAL ATMOSPHERE WATCH

International Ozone Commission

SPARC/IO3C/GAW Report on Long-term Ozone Trends and Uncertainties in the Stratosphere

I. Petropavlovskikh, S. Godin-Beekmann, D. Hubert, R. Damadeo, B. Hassler, V. Sofieva



LONG-TERM OZONE TRENDS AND UNCERTAINTIES IN THE STRATOSPHERE

LOTUS

SPARC Report No. 9, GAW Report No. 241, WCRP Report 17/2018

WCRP World Climate Research Programme

WORLD METEOROLOGICAL ORGANIZATION

United Nations Educational, Scientific and Cultural Organization

Intergovernmental Oceanographic Commission

International Science Council

SPARC/IO3C/GAW Report on Long-term Ozone Trends and Uncertainties in the Stratosphere

February 2019

Prepared by the SPARC LOTUS Activity,
edited by I. Petropavlovskikh, S. Godin-Beekmann, D. Hubert, R. Damadeo, B. Hassler, and V. Sofieva

SPARC Report No. 9, GAW Report No. 241, WCRP Report 17/2018

SPARC
Stratosphere-troposphere Processes And their Role in Climate

GLOBAL ATMOSPHERE WATCH

IO3C International Ozone Commission

WCRP World Climate Research Programme

WORLD METEOROLOGICAL ORGANIZATION

United Nations Educational, Scientific and Cultural Organization

Intergovernmental Oceanographic Commission

International Science Council

Citation
SPARC/IO3C/GAW, 2019: SPARC/IO3C/GAW Report on Long-term Ozone Trends and Uncertainties in the Stratosphere. I. Petropavlovskikh, S. Godin-Beekmann, D. Hubert, R. Damadeo, B. Hassler, V. Sofieva (Eds.), SPARC Report No. 9, GAW Report No. 241, WCRP-17/2018, doi: 10.17874/f899e57a20b, available at www.sparc-climate.org/publications/sparc-reports.

Electronic version available at **http://www.sparc-climate.org/publications/sparc-reports/sparc-report-no9**

ISSN 2296-5785 (Print)

Editing, Design, and Layout: Mareike Kenntner and Brigitte Ziegele

Authors & Reviewers

SPARC LOTUS Activity leaders

Irina Petropavlovskikh	(1) NOAA Earth System Research Laboratory (2) CIRES, University of Colorado	USA
Sophie Godin-Beekmann	LATMOS-IPSL, CNRS/UVSQ/Sorbonne University	France
Daan Hubert	Royal Belgian Institute for Space Aeronomy (BIRA-IASB)	Belgium

Working group leaders of the SPARC LOTUS Activity

Robert Damadeo	NASA Langley Research Center	USA
Birgit Hassler	Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre	Germany
Viktoria F. Sofieva	Finnish Meteorological Institute	Finland

Review editors

Neil R.P. Harris	Cranfield University	United Kingdom
Fiona Tummon	Federal Office of Meteorology and Climatology, MeteoSwiss	Switzerland
Donald J. Wuebbles	University of Illinois	USA

Authors and Contributors

Executive Summary

Chapter lead authors

Irina Petropavlovskikh	(1) NOAA Earth System Research Laboratory (2) CIRES, University of Colorado	USA
Sophie Godin-Beekmann	LATMOS-IPSL, CNRS/UVSQ/Sorbonne University	France
Daan Hubert	Royal Belgian Institute for Space Aeronomy (BIRA-IASB)	Belgium
Robert Damadeo	NASA Langley Research Center	USA
Birgit Hassler	Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre	Germany
Viktoria F. Sofieva	Finnish Meteorological Institute	Finland
Stacey M. Frith	(1) Science Systems and Applications Inc. (2) NASA Goddard Space Flight Center	USA
Kleareti Tourpali	Aristotle University of Thessaloniki	Greece

Chapter 1: Introduction

Chapter lead authors

Sophie Godin-Beekmann	LATMOS-IPSL, CNRS/UVSQ/Sorbonne University	France
Irina Petropavlovskikh	(1) NOAA Earth System Research Laboratory (2) CIRES, University of Colorado	USA
Daan Hubert	Royal Belgian Institute for Space Aeronomy (BIRA-IASB)	Belgium

Chapter 2: Observations and model data

Chapter lead authors

Irina Petropavlovskikh	(1) NOAA Earth System Research Laboratory (2) CIRES, University of Colorado	USA
Viktoria F. Sofieva	Finnish Meteorological Institute	Finland

Co-authors

Stacey M. Frith	(1) Science Systems and Applications Inc. (2) NASA Goddard Space Flight Center	USA
Kleareti Tourpali	Aristotle University of Thessaloniki	Greece
Corinne Vigouroux	Royal Belgian Institute for Space Aeronomy (BIRA-IASB)	Belgium
Jeannette D. Wild	(1) Innovim, LLC, Greenbelt, MD (2) NOAA/NCEP/Climate Prediction Center	USA
William T. Ball	(1) Institute for Atmospheric and Climate Science, ETH Zurich (2) Physikalisch-Meteorologisches Observatorium Davos - World Radiation Center	Switzerland
Kai-Lan Chang	National Research Council Fellow at NOAA Earth System Research Laboratory	USA
Sean M. Davis	(1) NOAA Earth System Research Laboratory (2) CIRES, University of Colorado	USA
Doug A. Degenstein	University of Saskatchewan	Canada
Lucien Froidevaux	Jet Propulsion Laboratory, California Institute of Technology	USA
Sophie Godin-Beekmann	LATMOS-IPSL, CNRS/UVSQ/Sorbonne University	France
Daan Hubert	Royal Belgian Institute for Space Aeronomy (BIRA-IASB)	Belgium
Alexandra Laeng	KIT, Institute of Meteorology and Climate Research (IMK-ASF)	Germany
Thierry Leblanc	Jet Propulsion Laboratory, California Institute of Technology	USA
Eliane Maillard Barras	Federal Office of Meteorology and Climatology, MeteoSwiss	Switzerland
Herman G.J. Smit	Institute of Energy and Climate Research at Forschungszentrum Jülich	Germany
Wolfgang Steinbrecht	Deutscher Wetterdienst, Meteorological Observatory Hohenpeissenberg	Germany

Data providers

John Anderson	School of Science, Hampton University	USA
Omaira E. García	IARC, Agencia Estatal de Meteorología (AEMET)	Spain
James W. Hannigan	National Center for Atmospheric Research (NCAR)	USA
Bryan J. Johnson	NOAA Earth System Research Laboratory, Global Monitoring Division	USA
Nicholas B. Jones	University of Wollongong	Australia
Emmanuel Mahieu	University of Liège	Belgium
Ankie J. M. Piters	Royal Netherlands Meteorological Institute (KNMI)	The Netherlands
Françoise Posny	Laboratoire de l'Atmosphère et des Cyclones (LACy), UMR8105, Saint-Denis, Réunion	France
Richard Querel	National Institute of Water and Atmospheric Research (NIWA)	New Zealand
Dan Smale	National Institute of Water and Atmospheric Research (NIWA)	New Zealand
René Stübi	Federal Office of Meteorology and Climatology, MeteoSwiss	Switzerland
Daan P.J. Swart	National Institute for Public Health and the Environment (RIVM)	The Netherlands
David W. Tarasick	Environment and Climate Change Canada, Toronto	Canada
Anne M. Thompson	NASA Goddard Space Flight Center	USA
Matthew B. Tully	Bureau of Meteorology	Australia
Joanna A.E. van Gijsel	Faculty of Geo-Information Science and Earth Observation, University of Twente	The Netherlands
Roeland Van Malderen	Royal Meteorological Institute of Belgium (RMIB)	Belgium
Ray Wang	School of Earth and Atmospheric Sciences, Georgia Institute of Technology	USA
Jacquelyn C. Witte	(1) Science Systems and Applications Inc. (2) NASA Goddard Space Flight Center	USA

CCMI modelers

Slimane Bekki	LATMOS-IPSL, CNRS, Paris	France
Neal Butchart	Met Office Hadley Centre	United Kingdom
Martyn P. Chipperfield	University of Leeds	United Kingdom
Makoto Deushi	Meteorological Research Institute (MRI)	Japan
Sandip S. Dhomse	University of Leeds	United Kingdom
Steven C. Hardiman	Met Office Hadley Centre	United Kingdom

Michaela I. Hegglin	University of Reading	United Kingdom
Patrick Jöckel	Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre	Germany
Douglas Kinnison	National Center for Atmospheric Research (NCAR)	USA
Jean-François Lamarque	National Center for Atmospheric Research (NCAR)	USA
Martine Michou	Météo-France CNRM	France
Olaf Morgenstern	National Institute of Water and Atmospheric Research (NIWA)	New Zealand
Fiona M. O'Connor	Met Office Hadley Centre	United Kingdom
Luke D. Oman	NASA Goddard Space Flight Center	USA
Giovanni Pitari	Department of Physical and Chemical Sciences, Università dell'Aquila	Italy
David A. Plummer	Environment and Climate Change Canada, Montreal	Canada
Laura E. Revell	Bodeker Scientific	New Zealand
Eugene Rozanov	(1) Physikalisch-Meteorologisches Observatorium Davos - World Radiation Center (2) Institute for Atmospheric and Climate Science, ETH Zurich	Switzerland
David Saint-Martin	CNRM/GMGEC/AMACS	France
Robyn Schofield	(1) School of Earth Sciences, University of Melbourne (2) ARC Centre of Excellence for Climate System Science	Australia
Andrea Stenke	Institute for Atmospheric and Climate Science, ETH Zurich	Switzerland
Kane A. Stone	(1) School of Earth Sciences, University of Melbourne (2) ARC Centre of Excellence for Climate System Science *Now at: Department of Earth, Atmospheric, and Planetary Science, MIT	Australia
Simone Tilmes	National Center for Atmospheric Research (NCAR)	USA
Daniele Visioni	Department of Physical and Chemical Sciences, Università dell'Aquila	Italy
Guang Zeng	National Institute of Water and Atmospheric Research (NIWA)	New Zealand

Chapter 3: Challenges for trend studies

Chapter lead authors

Daan Hubert	Royal Belgian Institute for Space Aeronomy (BIRA-IASB)	Belgium
Viktoria F. Sofieva	Finnish Meteorological Institute	Finland

Co-authors

William T. Ball	(1) Institute for Atmospheric and Climate Science, ETH Zurich (2) Physikalisch-Meteorologisches Observatorium Davos - World Radiation Center	Switzerland
Robert Damadeo	NASA Langley Research Center	USA
Stacey M. Frith	(1) Science Systems and Applications Inc. (2) NASA Goddard Space Flight Center	USA
Kleareti Tourpali	Aristotle University of Thessaloniki	Greece
Jeannette D. Wild	(1) Innovim, LLC, Greenbelt, MD (2) NOAA/NCEP/Climate Prediction Center	USA
Christos S. Zerefos	Research Centre for Atmospheric Physics and Climatology, Academy of Athens	Greece

Contributing authors

Justin Alsing	Center for Computational Astrophysics, Flatiron Institute, NY *Now at : Oskar Klein Centre for Cosmoparticle Physics, Stockholm University	USA
Kostas Eleftheratos	Dep. of Geology and Geoenvironment, National and Kapodistrian University of Athens	Greece
Sophie Godin-Beekmann	LATMOS-IPSL, CNRS/UVSQ/Sorbonne University	France
John Kapsomenakis	Research Centre for Atmospheric Physics and Climatology, Academy of Athens	Greece
Jean-Christopher Lambert	Royal Belgian Institute for Space Aeronomy (BIRA-IASB)	Belgium
Thierry Leblanc	Jet Propulsion Laboratory, California Institute of Technology	USA
Gerald E. Nedoluha	Remote Sensing Division, Naval Research Laboratory	USA
Wolfgang Steinbrecht	Deutscher Wetterdienst, Meteorological Observatory Hohenpeissenberg	Germany
René Stübi	Federal Office of Meteorology and Climatology, MeteoSwiss	Switzerland

Chapter 4: The LOTUS regression model

Chapter lead authors

Robert Damadeo	NASA Langley Research Center	USA
Birgit Hassler	Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre	Germany

Co-authors

Daniel J. Zawada	University of Saskatchewan	Canada
Stacey M. Frith	(1) Science Systems and Applications Inc. (2) NASA Goddard Space Flight Center	USA
William T. Ball	(1) Institute for Atmospheric and Climate Science, ETH Zurich (2) Physikalisch-Meteorologisches Observatorium Davos - World Radiation Center	Switzerland
Kai-Lan Chang	National Research Council Fellow at NOAA Earth System Research Laboratory	USA
Doug A. Degenstein	University of Saskatchewan	Canada
Daan Hubert	Royal Belgian Institute for Space Aeronomy (BIRA-IASB)	Belgium
Stergios Misios	University of Oxford	United Kingdom
Irina Petropavlovskikh	(1) NOAA Earth System Research Laboratory (2) CIRES, University of Colorado	USA
Chris Z. Roth	University of Saskatchewan	Canada
Viktoria F. Sofieva	Finnish Meteorological Institute	Finland
Wolfgang Steinbrecht	Deutscher Wetterdienst, Meteorological Observatory Hohenpeissenberg	Germany
Kleareti Tourpali	Aristotle University of Thessaloniki	Greece
Christos S. Zerefos	Research Centre for Atmospheric Physics and Climatology, Academy of Athens	Greece

Contributing authors

Justin Alsing	Center for Computational Astrophysics, Flatiron Institute, NY *Now at : Oskar Klein Centre for Cosmoparticle Physics, Stockholm University	USA
Dimitris Balis	Aristotle University of Thessaloniki	Greece
Melanie Coldewey-Egbers	Deutsches Zentrum für Luft- und Raumfahrt (DLR), Inst. für Methodik der Fernerkundung	Germany
Kostas Eleftheratos	Dep. of Geology and Geoenvironment, National and Kapodistrian University of Athens	Greece
Sophie Godin-Beekmann	LATMOS-IPSL, CNRS/UVSQ/Sorbonne University	France
Aleksandr Gruzdev	A.M. Obukhov Institute of Atmospheric Physics, Russian Academy of Science	Russia
John Kapsomenakis	Research Centre for Atmospheric Physics and Climatology, Academy of Athens	Greece
Alexandra Laeng	KIT, Institute of Meteorology and Climate Research (IMK-ASF)	Germany
Marko Laine	Finnish Meteorological Institute	Finland
Eliane Maillard Barras	Federal Office of Meteorology and Climatology, MeteoSwiss	Switzerland
Michael Taylor	Department of Meteorology, University of Reading	United Kingdom
Thomas von Clarmann	KIT, Institute of Meteorology and Climate Research (IMK-ASF)	Germany
Mark Weber	Institute of Environmental Physics, University of Bremen	Germany
Jeannette D. Wild	(1) Innovim, LLC, Greenbelt, MD (2) NOAA/NCEP/Climate Prediction Center	USA

Chapter 5: Time series and trend results

Chapter lead authors

Birgit Hassler	Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre	Germany
Robert Damadeo	NASA Langley Research Center	USA

Co-authors

Kai-Lan Chang	National Research Council Fellow at NOAA Earth System Research Laboratory	USA
Viktoria F. Sofieva	Finnish Meteorological Institute	Finland
Kleareti Tourpali	Aristotle University of Thessaloniki	Greece

Stacey M. Frith	(1) Science Systems and Applications Inc. (2) NASA Goddard Space Flight Center	USA
William T. Ball	(1) Institute for Atmospheric and Climate Science, ETH Zurich (2) Physikalisch-Meteorologisches Observatorium Davos - World Radiation Center	Switzerland
Doug A. Degenstein	University of Saskatchewan	Canada
Sophie Godin-Beekmann	LATMOS-IPSL, CNRS/UVSQ/Sorbonne University	France
Daan Hubert	Royal Belgian Institute for Space Aeronomy (BIRA-IASB)	Belgium
Eliane Maillard Barras	Federal Office of Meteorology and Climatology, MeteoSwiss	Switzerland
Stergios Misios	University of Oxford	United Kingdom
Irina Petropavlovskikh	(1) NOAA Earth System Research Laboratory (2) CIRES, University of Colorado	USA
Chris Z. Roth	University of Saskatchewan	Canada
Wolfgang Steinbrecht	Deutscher Wetterdienst, Meteorological Observatory Hohenpeissenberg	Germany
Corinne Vigouroux	Royal Belgian Institute for Space Aeronomy (BIRA-IASB)	Belgium
Thomas von Clarmann	KIT, Institute of Meteorology and Climate Research (IMK-ASF)	Germany
Daniel J. Zawada	University of Saskatchewan	Canada
Christos S. Zerefos	Research Centre for Atmospheric Physics and Climatology, Academy of Athens	Greece

Contributing authors

Justin Alsing	Center for Computational Astrophysics, Flatiron Institute, NY *Now at : Oskar Klein Centre for Cosmoparticle Physics, Stockholm University	USA
Dimitris Balis	Aristotle University of Thessaloniki	Greece
Melanie Coldewey-Egbers	Deutsches Zentrum für Luft- und Raumfahrt (DLR), Inst. für Methodik der Fernerkundung	Germany
Kostas Eleftheratos	Dep. of Geology and Geoenvironment, National and Kapodistrian University of Athens	Greece
Aleksandr Gruzdev	A.M. Obukhov Institute of Atmospheric Physics, Russian Academy of Science	Russia
John Kapsomenakis	Research Centre for Atmospheric Physics and Climatology, Academy of Athens	Greece
Alexandra Laeng	KIT, Institute of Meteorology and Climate Research (IMK-ASF)	Germany
Marko Laine	Finnish Meteorological Institute	Finland
Michael Taylor	Department of Meteorology, University of Reading	United Kingdom
Mark Weber	Institute of Environmental Physics, University of Bremen	Germany
Jeannette D. Wild	(1) Innovim, LLC, Greenbelt, MD (2) NOAA/NCEP/Climate Prediction Center	USA

Reviewers

Gufran Beig	Indian Institute of Tropical Meteorology (Ministry of Earth Sciences, Govt. of India)	India
Greg E. Bodeker	Bodeker Scientific	New Zealand
Pablo O. Canziani	University of Buenos Aires	Argentina
Martin Dameris	Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre	Germany
Richard D. Eckman	NASA	USA
James Keeble	Centre for Atmospheric Science, University of Cambridge	United Kingdom
Paul A. Newman	NASA Goddard Space Flight Center	USA
Scott Osprey	NCAS, University of Oxford	United Kingdom
Richard S. Stolarski	Department of Earth and Planetary Sciences, Johns Hopkins University	USA
Susann Tegtmeier	GEOMAR Helmholtz Centre for Ocean Research Kiel	Germany
Matthew B. Tully	Bureau of Meteorology	Australia
Claus Zehner	European Space Agency (ESA/ESRIN)	Italy

Acknowledgements

We acknowledge SPARC, IO3C and WMO GAW for supporting and initiating the LOTUS activity (**http://lotus.aeronomie.be**), through which a diverse group of experts in satellite and ground-based ozone observation systems, ozone modelling and statistical analysis techniques were able to closely collaborate to produce the work described in this Report. These activities have been undertaken under the guidance and sponsorship of the World Climate Research Programme. The ground-based data used in this publication were obtained as part of WMO's Global Atmosphere Watch (GAW) and two of its main contributors, namely, the Network for the Detection of Atmospheric Composition Change (NDACC) and the Southern Hemisphere ADditional OZonesondes programme (SHADOZ). We acknowledge the meticulous and sustained work of the PIs and staff at all stations to acquire and maintain long-term ozone profile data records of high quality. The data records are publicly available via the NDACC Data Host Facility (**http://www.ndacc.org**), the SHADOZ archive (**https://tropo.gsfc.nasa.gov/shadoz**)and the World Ozone and Ultraviolet Data Centre (**http://www.woudc.org**). NDACC and SHADOZ are supported by meteorological services, space agencies and research organisms from many countries, with archives funded by NASA and NOAA. The Merged Ozone Data Set was constructed under the NASA MeaSUREs Project and is maintained under NASA WBS 479717 (Long Term Measurement of Ozone). We thank the SBUV instrument team for providing the SBUV Version 8.6 individual data records. The OSIRIS team would like to thank the Canadian Space Agency for over twenty years of support. The merged SAGE-CCI-OMPS data record was developed in the framework of ESA's Climate Change Initiative project on Ozone (Ozone_cci). Authors of the Report are indebted to the IGAC/SPARC Chemistry Climate Model Initiative (CCMI), the Working Group leaders and model/modelling groups PIs for the use of zonal mean ozone profiles in this Report. We also thank the British Atmospheric Data Center for assistance with the CCMI data Archive. W. T. Ball acknowledges financial support by the Swiss National Science Foundation (SNSF) grants 200020_163206 (SIMA). D. Hubert acknowledges financial support by the Belgian Science Policy Office through the ProDEx TROVA project and by ESA through the Climate Change Initiative project on Ozone (Phase 1 and 2) and the Multi-TASTE project (Phase F). C. Vigouroux acknowledges financial support by the EU H2020 project GAIA-Clim (No 640276). V.F. Sofieva acknowledges support by the Academy of Finland, Centre of Excellence of Inverse Modelling and Imaging. Sophie Godin-Beekmann acknowledges financial support by CNRS and CNES through the NDACC-France project. I. Petropavlovskikh would like to acknowledge support from the Cooperative Institute for Environmental Sciences (CIRES) at the University of Colorado Boulder, she was funded through the grants #NA17OAR4320101, NA15OAR4320137 and NA12OAR4320137 under the National Oceanic and Atmospheric Administration Cooperative agreement. K.-L. Chang would like to acknowledge financial support from the National Research Council Fellowship and the National Oceanic & Atmospheric Administration (NOAA), Global Monitoring Division, Boulder, Colorado, USA. The creation of merged data records for GOZCARDS was initially supported by NASA MeaSUREs, and this effort has continued thanks to the Aura MLS mission; we acknowledge the assistance of Ryan Fuller (at the Jet Propulsion Laboratory) in this regard. Work at the Jet Propulsion Laboratory, California Institute of Technology, was done under contract with NASA. SWOOSH data were provided by the NOAA Earth System Research Laboratory's Chemical Sciences Division.

Executive summary

- The assessment of long-term observations by LOTUS confirms the significant decline of ozone concentrations in the upper stratosphere (at altitudes above the 10–5 hPa level) between January 1985 and December 1996. The strongest trends are observed near 2 hPa (~42 km) with values of 5.9–6.2 % per decade at mid-latitudes and 4.8 % per decade in the tropics. Trends are significant at more than 5 standard deviations in this altitude range.

- Trends derived from satellite and ground-based records in the pre-1997 time period agree with climate model simulations within respective uncertainties thus confirming our understanding of ozone loss processes in the upper stratosphere during that period.

- Between January 2000 and December 2016, positive trends are obtained throughout the upper stratosphere for satellite and ground-based records. The combined trends from six merged satellite records are larger in the Northern Hemisphere mid-latitudes (2–3 % per decade between ~5–1 hPa) than in the tropics (1–1.5 % per decade between ~3–1 hPa) and Southern Hemisphere mid-latitudes (~2 % per decade near 2 hPa). Statistical confidence is largest for trends in the Northern Hemisphere mid-latitudes.

- For altitudes below the 4 hPa level, ozone trends in the post-2000 time period are not significant. Though not significant, negative ozone trends of 0.5–1.5 % per decade are consistently detected by multiple satellite combined records in the 50–15 hPa altitude range over the tropics. Trends derived from ground-based data and Chemistry-Climate Model Initiative (CCMI) model simulations are generally consistent but more variable in this region. The mean CCMI model trend is negative at altitudes below 30 hPa, but the range of individual model trends is large; trends in ground-based records tend to be negative at 20 hPa but increase at lower altitudes (except trends from the microwave records). At mid-latitudes, the trends are close to zero down to 50 hPa.

- Larger differences in post-2000 trends from the various records are observed in the lowermost stratosphere (100–50 hPa) in all latitude bands. Non-significant negative trends are derived from merged satellite records over the tropics and the Northern Hemisphere mid-latitudes. Model simulations show positive trends in the mid-latitudes in both hemispheres in this altitude range, although the trends are not statistically significant.

- LOTUS estimates of past and recent ozone trends are in fairly good agreement with results from previous studies. For the post-2000 period, the largest differences are found throughout the middle stratosphere. These differences stem primarily from extensions of and revisions to existing data records, the addition of new data records, and in some cases the use of a different trend model.

- While trend values in recent studies are fairly similar, the uncertainties and hence significances of the combined trends in broad latitude bands differ substantially. The LOTUS approach, based on both error propagation and standard error of the mean, also explicitly accounts for correlation between the data sets, which results in more conservative uncertainties and thus lower, but more realistic, confidence in positive upper stratospheric trend values compared to the most recently published assessment of merged satellite data set trends.

Have ozone concentrations in the stratosphere significantly increased since the end of the 1990s when levels of ozone depleting substances (ODSs) started to decline? Finding an answer to this question is of great societal importance to ensure that the measures taken by the Montreal Protocol and subsequent amendments to reduce ODSs continue to adequately protect the ozone layer. However, the confidence with which we can assess changes in stratospheric ozone since the mid-1990s has been the subject of considerable scientific debate in recent years, as it depends on the data sets and the analysis methods used. Settling this scientific debate is one of the main objectives of the LOTUS activity, short for Long-term Ozone Trends and Uncertainties in the Stratosphere.

Below, we summarise the main results obtained during the first phase of LOTUS, which was primarily targeted at providing timely input to the 2018 World Meteorological Organization (WMO) Ozone Assessment (*WMO*, 2018). During this phase we reevaluated the satellite and ground-based data records as well as the time series analysis methods commonly used to derive long-term trends. Using a single "LOTUS regression" model, we reassessed past and recent trends in the vertical distribution of stratospheric ozone from the updated individual data records. We then developed a new approach for combining the individual trend estimates from satellite-based records into a single best estimate of ozone profile trends with associated uncertainty estimates. Finally, we compared the satellite-based profile trends in broad latitude bands to trends from ground-based data, from the collection of CCMI-1 model simulations, and from past evaluations of satellite-based trends in peer-reviewed literature.

Some regions in the stratosphere have not been considered (*e.g.*, polar) or have not been analysed in full detail (*e.g.*, lower stratosphere) because of the timeline for the 2018 WMO Ozone Assessment (*WMO*, 2018).

ES.1 New and improved data sets for trend analyses

The work performed in LOTUS and the resulting trends are based on the latest observations from single and merged satellite records as well as from ground-based instruments. In addition to including four additional years of data compared to the results published in the 2014 WMO Ozone Assessment and in the framework of the SI2N activity (*i.e.*, *Harris et al.*, 2015; and references therein), many of the records utilised in LOTUS have been improved for trend analyses (*i.e.*, new methods to combine/homogenise data sets, sampling corrections, and revised calibration and pointing stability). With nearly global coverage and durations spanning at least 30 years (1985–2016), the combined satellite records constitute the backbone of LOTUS trend analyses. With respect to ground-based measurements, LOTUS used individual records from passive and active remote sensing techniques as well as from ozonesondes, including the few homogenised sonde records that exist. In addition to observations, CCMI model simulations were used to test our understanding of ozone profile trends and results shown here represent the first analysis of trends in the vertical distribution of ozone from the CCMI-1 REF-C2 simulations. The details of all observational and model data used in LOTUS, as well as the methods for averaging and merging data records, are discussed in *Chapter 2* of this Report.

ES.2 Addressing the challenges with data for trend estimation

Any assessment of trends and uncertainties necessitates the investigation of the data sets themselves as well as the nuances of their creation, in the case of merged data, and utilisation. The intercomparisons of the satellite and ground-based ozone time series reported in *Chapter 3* reveal a number of measurement artifacts (*e.g.*, drifts, discontinuities, and spikes) but generally show good agreement. In fact, the agreement between observational records is better than for earlier versions of merged ozone records used for previous assessments (*i.e.*, *WMO*, 2014; *Harris et al.*, 2015; and references therein), which lends increased confidence in derived trends. Acknowledging and understanding potential anomalies is important for explaining differences in the trends and trend uncertainties and provides guidance on how to improve the data sets. Availability of data records by a number of complementary instruments is key in singling out these issues and attributing them to one of the data records. Recently proposed Bayesian analyses (*Section 3.1.5*) may further help to systematically identify artifacts in particular data sets, while Monte Carlo (MC) methods (*Section 3.1.4*) can help assess the impact of remaining uncertainties in the records on the final trend estimates. Sampling biases also come into play, since the ozone time series are regressed at an aggregate level and lead to systematic changes in derived trends by up to 1–2 % per decade in parts of the stratosphere, which constitute a considerable fraction of the estimates of post-2000 trends.

ES.3 Sensitivity testing for a consensus regression model

One of the primary goals of LOTUS is to assess the impact of analysis methods on derived ozone trends and their uncertainties. In that regard, a test of 15 previously published multiple linear regression (MLR) models applied to a common data set was performed to evaluate the sensitivity of derived trends to different methodologies. The results showed good agreement in the shape of retrieved trends but a general spread in derived trend values after 2000 of 1–2 % per decade, with overall differences as high as 3 % per decade, which revealed the need for additional sensitivity tests in order to create a consensus analysis method.

Chapter 4 of this Report details a series of sensitivity tests pertaining to the impact of different geophysical and empirical proxies used in MLR analyses on the derived trends and their uncertainties. Since work during the first phase of LOTUS was optimised towards estimating middle and upper stratospheric trends in satellite data sets, we focused on the sensitivity of proxies to ozone variability at these altitudes. Accordingly, results showed that short period proxies (*e.g.*, AO[1] , AAO[1], NAO[1], and EHF[1]) had negligible effects on trends and variably small impacts on uncertainties while excluding the solar cycle, QBO[1], or ENSO[1] proxies from the regression model had significant effects on the trend (1–2 % per decade difference) and uncertainty (around 1 % per decade) estimates. Different long-term trend proxies were also investigated, revealing the complexity of attempting to capture changes in ozone stemming from the influences of both ODSs and greenhouse gases. Ultimately a single consensus "LOTUS regression" model, based on a simple yet appropriate set of geophysical proxies and a trend proxy designed to capture mean trends in satellite data sets, was chosen for our analyses and was also packaged for public use (**https://arg.usask.ca/docs/LOTUS_regression**).

Lastly, while a more traditional MLR-based approach was chosen for the majority of work in LOTUS, a newer statistical approach, namely a dynamic linear model (DLM),

[1] Abbreviations for proxies: AO = Arctic Oscillation; AAO = Antarctic Oscillation; NAO = North Atlantic Oscillation; EHF = Eddy Heat Flux; QBO = Quasi-Biennial Oscillation; ENSO = El Niño Southern Oscillation.

was also tested. The DLM technique, as its name implies, does not constrain the shape of the likely nonlinear long-term trend and thus may ultimately be better suited for this type of work. However, this method was not fully evaluated during the first phase of LOTUS and a more comprehensive comparison of MLR and DLM for ozone trend estimates is still needed.

ES.4 Examination of trend results from individual data sets

The "LOTUS regression" model was first applied to the individual data sets at their native resolution. Results display expected patterns of ozone decline in the upper stratosphere prior to the late 1990s and a subsequent smaller increase since 2000 in generally good agreement with models, though the magnitude and statistical significance of these results vary between the different data sets. Most data records and model simulations point to continuously declining ozone levels in the tropical lower stratosphere in a fairly coherent yet generally not statistically significant way, but results at mid-latitudes in the lower stratosphere are variable and inconclusive. Additionally, the differences between trend results are discussed as they pertain to the differences in the various data sets and merging techniques (see *Chapter 3*). In order to place these results in context with previous comprehensive studies, the regressions are repeated with the data sets averaged over broad latitude bands (*i.e.*, 60°S–35°S and 35°N–60°N representing southern and northern mid-latitudes respectively and 20°S–20°N representing the tropics) prior to continuing with the overall analysis.

ES.5 Revised approach to combine trends

The typical desire for a single set of spatially resolved trend results (*e.g.*, as in previous Ozone Assessments) when faced with an ensemble derived from the analyses of multiple data sets creates the unique challenge of merging not only the ensemble of trend results but also their uncertainties. With an aim towards determining if these results are statistically significant, it is the latter component that is more important and often more complicated. Historically, a variety of techniques have been used to merge the uncertainties and the LOTUS Report introduces a new, statistically more robust method. This includes not only components involving simple error propagation, which captures uncertainties introduced from the data and the analyses, and the standard error of the mean, which captures systematic uncertainties such as those introduced by drifts between data sets, but also the correlations between the data sets themselves, estimated from the correlation of the fit residuals. Consequently, it is the nature of the independence of the data sets and their resulting trends that is the most important aspect of merging the trend uncertainties. *Section 5.3* in this Report details the complexities of this merging and the necessary assumptions chosen for the LOTUS work, which are balanced between not wanting to overestimate or underestimate the combined uncertainties. The results presented here have uncertainties that fall between previous comprehensive works but err towards a more conservative estimate. Ultimately, this work concludes that the most meaningful way to improve the uncertainties in future analyses would be to reconcile the discrepancies between the data sets themselves prior to the merging process.

ES.6 Assessment of combined ozone profile trends

Estimates of combined satellite trends are summarised in **Figure ES.1** and in **Table ES.1** (see *Section 5.6*). For this work, results are separated into two distinct time periods with "pre-1997" being defined as the period from January 1985 to December 1996, while "post-2000" refers to the period from January 2000 to December 2016. Comparisons of LOTUS trends (hereafter L19) with previously published trends (WMO, 2014, hereafter W14; Harris et al., 2015, hereafter H15; Steinbrecht et al., 2017, hereafter S17) are shown in **Figure ES.1** as well.

ES.6.1 1985–1996 trends

Negative trends are found across nearly the entire stratosphere in the pre-1997 period for almost all satellite and ground-based data records. Individual and combined satellite data show highly statistically significant evidence of declining ozone concentrations in the upper stratosphere (at altitudes above the 10–5 hPa level) since the mid-1980s and well into the 1990s. The depletion reaches a maximum rate near 2 hPa (~42 km) of 5.9–6.2 % per decade at mid-latitudes and 4.8 % per decade in the tropics (see **Table S5.1** in the *Supplement*). Ozone decline rates in the middle stratosphere (30–15 hPa) are considerably smaller, with statistically insignificant values of at most 1–2 % per decade. Negative trends are found across the lower stratosphere (down to 50 hPa), while in the lowermost stratosphere (down to 100 hPa) trends differ according to latitude, with large significant negative trends of about 5 % per decade in the Northern Hemisphere. However, confidence in trend results is reduced in the lower stratosphere due to large natural variability, low ozone values, and decreased sensitivity of satellite observations. Trends derived from ground-based measurements generally corroborate satellite trend results. However, due to their larger sparseness in space and time, especially during this early period, the significance of the trends is not as high and trend values differ. Results agree well with those of model simulations (within 1 % per decade) throughout the middle and upper stratosphere at all latitudes, lending confidence that these losses in ozone were the result of chemical forcing from ODSs according to model predictions. However, larger differences exist between satellite and model results in the lowermost stratosphere, with disagreements outside the large uncertainties only in the Southern Hemisphere.

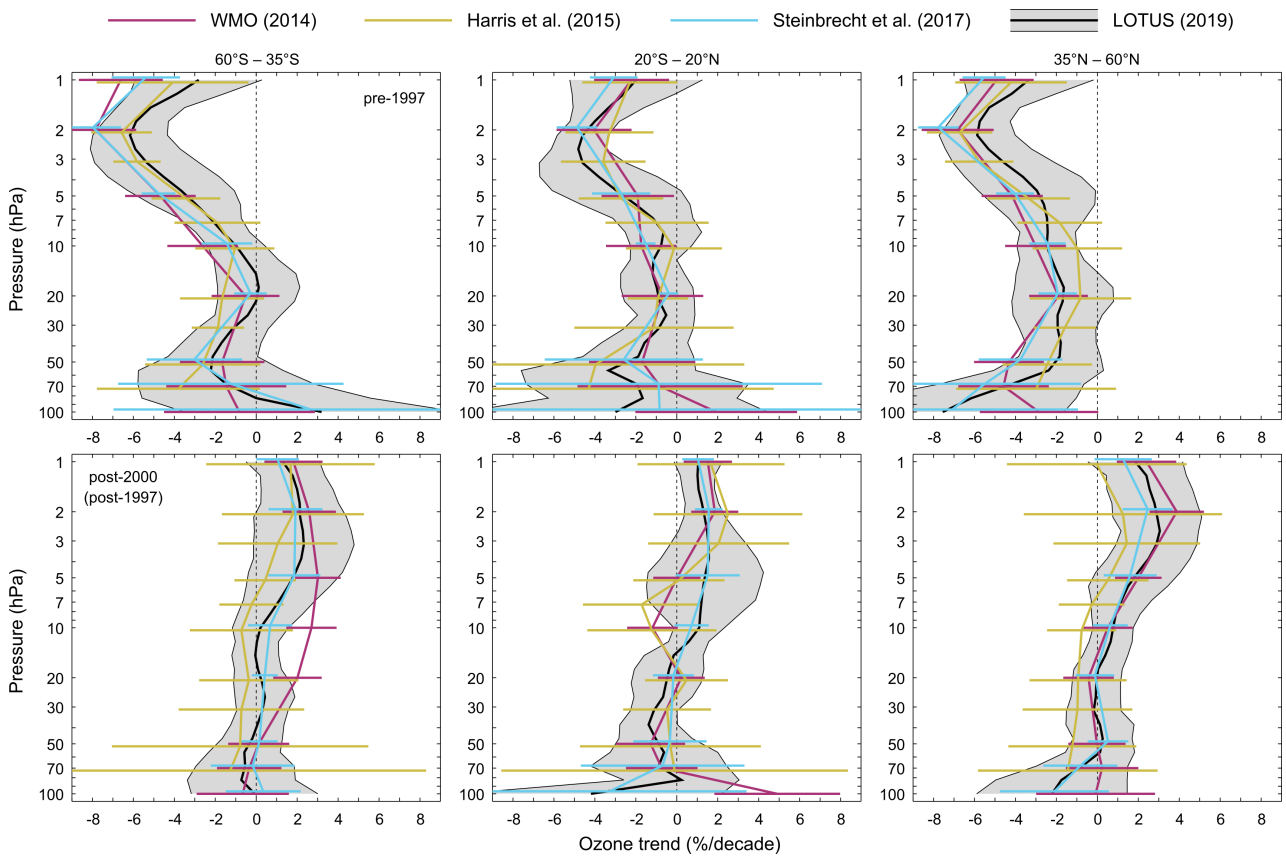


Figure ES.1: *Overview of ozone profile trends from past and recent assessments: WMO (2014), Harris et al. (2015), Steinbrecht et al. (2017), and LOTUS (this work) are shown in red, orange, blue, and black respectively. Top row shows trends before the turnaround of ODSs and bottom row since the turnaround (analysis time period differs by assessment). Shaded area and error bars represent the 95 % confidence interval for the combined trend. Coloured profiles are slightly offset on the vertical axis for display purposes. This figure is also shown in Chapter 5 as* ***Figure 5.12****. LOTUS results are tabulated for each pressure level in* ***Table S5.1*** *in Supplement. Steinbrecht et al. (2017) did not report or discuss pre-1997 trends, but results shown here were obtained from that work (private communication).*

	Ozone trend Jan 1985 - Dec 1996 (% per decade, ± 2 σ)			**Ozone trend Jan 2000 - Dec 2016** (% per decade, ± 2 σ)		
Pressure (hPa)	**60-35°S**	**20°S-35°N**	**35-60°N**	**60-35°S**	**20°S-35°N**	35-60°N
1	-2.8 ± 3.1	-2.0 ± 3.2	-3.3 ± 3.2	1.3 ± 1.8	1.0 ± 1.2	1.9 ± 2.3
2	-6.0 ± 1.7	-4.4 ± 1.2	-5.8 ± 1.7	2.1 ± 2.1	1.3 ± 0.9	2.9 ± 2.1
5	-3.4 ± 2.4	-2.6 ± 3.0	-2.8 ± 2.7	1.8 ± 2.2	1.4 ± 2.8	1.8 ± 2.0
7	-2.2 ± 1.5	-1.1 ± 1.9	-2.5 ± 1.8	1.1 ± 1.7	1.2 ± 2.5	1.1 ± 1.4
10	1.0 ± 1.4	-0.8 ± 1.5	-2.5 ± 1.5	0.2 ± 1.4	1.1 ± 1.2	0.8 ± 1.0
20	0.0 ± 1.9	-0.9 ± 1.7	-1.7 ± 2.5	0.2 ± 1.3	-0.5 ± 1.7	0.0 ± 1.2
50	-2.2 ± 2.7	-2.4 ±3.3	-2.1 ± 2.3	-0.3 ± 1.6	-0.9 ± 2.1	0.2 ± 1.5
70	-1.2 ± 4.5	-1.9 ± 5.3	-4.5 ± 3.8	-0.6 ± 2.4	-0.7 ± 3.3	-0.9 ± 2.4

Table ES.1: *Overview of LOTUS combined satellite trends in three latitude bands and two time periods. Central values and uncertainties representing the 95 % confidence interval are listed in the table. Trend results that are statistically significant at the 2-sigma level are highlighted in grey cells. See also* ***Figure ES.1****. Please note that trends and uncertainties are interpolated onto pressure levels that are common to other studies (e.g., WMO, 2014; Steinbrecht et al., 2017) to facilitate comparisons between these studies and LOTUS. Trends discussed in Chapter 5 are presented on the LOTUS pressure levels, which have a higher vertical resolution, and these are tabulated in* ***Table S5.1*** *in Supplement.*

ES.6.2 2000–2016 trends

Positive trends are found throughout the upper stratosphere and part of the middle stratosphere in the post-2000 period for both satellite and ground-based trends, though results vary for ground-based data depending upon the observation technique. Results from satellites show statistically significant positive trends in the Northern Hemisphere at mid-latitudes of 2–3 % per decade in the upper stratosphere (between ~5–1 hPa) and 1–1.5 % per decade in the tropics (between ~3–1 hPa).

Positive trends of ~2 % per decade are also found in the Southern Hemisphere near 2 hPa at mid-latitudes though the statistical confidence is smaller. In the Northern Hemisphere mid-latitudes, trends in the upper stratosphere are significant down to 4 hPa. At altitudes below 4 hPa, mid-latitude trends are no longer statistically significant, dropping from positive 1.8 % per decade at 5 hPa to near zero between 50–20 hPa. In the tropics, trends become negative below 15 hPa though the estimates of 0.5–1.5 % per decade are statistically insignificant as well. Generally, these satellite-based results are in agreement with ground-based observations and model simulations. The persistent negative trends in the middle and lower stratosphere over the tropics are likely the consequence of radiative and dynamical forcing from greenhouse gases according to model predictions (*WMO*, 2014; and references therein). Derived trends differ considerably in the lowermost stratosphere, below 50 hPa, depending on the data set and latitude. For example, satellite-based results show statistically insignificant negative trends (or near zero in the Southern Hemisphere) and ground-based trends agree in sign except in the tropics where there are significant positive trends. Model simulations, however, predict positive trends in mid-latitudes in both hemispheres in this altitude range.

ES.6.3 Comparison with previous assessments

LOTUS estimates of past and recent ozone trends are in fairly good agreement with results from previous assessments (*e.g.*, W14, H15, S17, and references therein). L19 and S17 trends differ by less than 0.5 % per decade in the post-2000 period, which is expected since similar data sets and a similar regression model were used for both studies. Trends by W14 and by H15 are in reasonable agreement with L19 as well, though larger differences are noted for the post-2000 period at Southern mid-latitudes, in the tropical middle stratosphere, and, for H15, in the Northern mid-latitudes as well. These differences stem primarily from extensions of and revisions to existing data records, the addition of new data records, and the use of a different trend proxy (*e.g.*, H15 assumed an inflection point at 1997).

While trend values in W14, H15, S17, and L19 are fairly similar, the uncertainties and hence significances of the combined trends differ substantially. This is the most critical component for the detection of the relatively small post-2000 trends. Even though both S17 and L19 use similar data records, the L19 approach, based on both error propagation and standard error of the mean, yields different uncertainties as compared to S17. In the upper and middle stratosphere, uncertainties from the standard error dominate and thus the estimate of the independence of the data records is critical in testing the trend null hypothesis. S17 trends are statistically significant across the entire upper stratosphere while in the L19 trend analysis, which derives a smaller degree of data independence from the correlation of fit residuals, high significance is found only at Northern mid-latitudes, and less significant trends are found in the tropics and at Southern mid-latitudes. There is not sufficient information in the trend analyses that can help determine exactly how independent the different data sets are. However, it is concluded that the real trend uncertainty lies in between S17 and L19 uncertainty estimates in the upper stratosphere. On the other hand, analysis of results of the two other assessments considered in the study suggested that they used either a too conservative (H15) or too optimistic (W14) approach to estimate combined uncertainties.

In the lower stratosphere, ozone trends are affected by large atmospheric variability and decreased sensitivity of satellite measurements. The L19 approach, which includes a term for error propagation from the regression coefficients, is not capable of capturing all sources of uncertainty and most importantly measurement drift, which leads to the conclusion that uncertainties derived from the analysis may be underestimated.

ES.6.4 Open issues and future work

The LOTUS Report assessment of satellite and ground-based ozone data sets (*Chapter 2*) builds the foundation for reconciling the discrepancies in ozone trends estimated from the individual climate data records. Understanding the causes of these differences would create improvements not only in the internal consistency of data sets, but also in the uncertainties of overall ozone trends. Further, development of techniques to directly assess uncertainties in the merged records resulting from discrepancies that cannot be completely reconciled, such as small relative drifts and differences resulting from coordinate transformations and sampling differences, would allow for a more precise estimate of significance of the mean trend.

For the satellite and ground-based data used in the LOTUS Report, information on stability and drifts of the measurement is still incomplete (*Chapter 3*).

The homogenisation of ozonesonde records was not finished prior to their use in the LOTUS assessment, and thus the ozonesonde trends and their uncertainties (especially in the lower stratosphere) may change in the future. In addition, in order to properly combine instrument-specific trends, a common matrix for providing error budget information for each ozone record is needed. Work developing a common approach to assessing errors in Level 2 satellite data is ongoing under the SPARC "Towards Unified Error Reporting (TUNER)" activity and ozone record uncertainties are addressed in other SPARC (Stratosphere-troposphere Processes And their Role in Climate) activities. Standardised error budgets have also been defined within the Network for the Detection of Atmospheric Composition Change (NDACC) and are in the process of being included in the data records.

The common statistical linear regression trend model (*Chapter 4*) used in the LOTUS Report was optimised for analyses of the zonally averaged satellite data sets. However, analyses of the ground-based data require reconsideration of additional proxies (*i.e.*, lag for ENSO, AO, AAO, NAO, EHF, *etc.*) and optimisation methods that can improve interpretation of the processes that impact ozone changes over the limited geophysical region and reduce trend uncertainties (*Chapter 5*).

The first attempt to evaluate representativeness of the ground-based station records for the middle and upper stratosphere using Solar Backscatter Ultraviolet Radiometer (SBUV) data was done under the LOTUS Report activity and discussed in *Chapter 4*. Comparisons of trends derived from satellite data selected under overpass criteria against zonally averaged trends will help with interpretation of stability in all observing systems and determine ozone recovery with high confidence.

There is a clear need for future activities of the CCMI modeling community, with experiments designed with the view on the verification of simulated trends. A large number of models is absolutely necessary in order to be able to assess the ozone variability associated with chemistry and dynamical transport mechanisms. Moreover, an assessment of model sensitivity to uncertainties in the volcanic aerosols, solar cycle, QBO, ENSO and other mechanisms is considered of great importance in order to advance our understanding of the ozone layer variability and associated response to natural variability.

In this Report, the ozone trends are analysed at low and middle latitudes, with a focus on the upper and middle stratosphere. Future works would explore trends in polar regions and in the lower stratosphere, which can be done in conjunction with the SPARC activity Observed Composition Trends And Variability in the Upper Troposphere and Lower Stratosphere (OCTAV-UTLS), dedicated to the assessment of the composition of the upper troposphere and lower stratosphere (UTLS) and identification of atmospheric processes that impact UTLS changes on the decadal scales. Similarly, the trends derived from total column data are also left for future work.

Assessments similar to the LOTUS activity need to be regularly repeated, preferably in collaboration with other SPARC and WMO/GAW (Global Atmospheric Watch) activities.

Abbreviations

AAO	Antarctic Oscillation
ACE-FTS	Atmospheric Chemistry Experiment - Fourier Transform Spectrometer (instrument)
AO	Arctic Oscillation
AOD	Aerosol Optical Depth
BASIC	BAyeSian Integrated and Consolidated (data set)
BESOS	Balloon Experiment on Standards for Ozone Sondes (campaign)
BM	Brewer-Mast ozonesonde (instrument)
BUV	Backscatter Ultraviolet Radiometer (instrument)
CCI	ESA's Climate Change Initiative
CCM	Chemistry-Climate Model
CCMI	Chemistry-Climate Model Initiative
CCMVal-2	Chemistry-Climate Model Validation Activity
CF	Correction Factor
CTM	Chemistry-Transport Model
DIAL	Differential Absorption Lidar (instrument)
DLM	Dynamic Linear Model (method)
DOFS	Degrees Of Freedom for Signal
ECC	Electrochemical Concentration Cell (instrument)
EESC	Equivalent Effective Stratospheric Chlorine
EHF	Eddy Heat Flux
EMD	Empirical Mode Decomposition (method)
ENSO	El Niño–Southern Oscillation
EnviSat	ENVIronmental SATellite
EOF	Empirical Orthogonal Function
EOS	Earth Observing System
ERA- Interim	European Center for Medium-Range Weather Forecast Re-Analysis (data set)
ERBS	Earth Radiation Budget Satellite
ESA	European Space Agency
FTIR	Fourier-Transform InfraRed spectrometer (instrument)
GAW	Global Atmosphere Watch (programme)
GCOS	Global Climate Observing System
GDR	German Democratic Republic
GHG	GreenHouse Gas
GISS	NASA Goddard Institute for Space Studies
GOMOS	Global Ozone Monitoring by Occultation of Stars (instrument)
GOZCARDS	Global OZone Chemistry And Related trace gas Data records for the Stratosphere (data set)
HALOE	HALogen Occultation Experiment (instrument)

HITRAN	HIgh Resolution TRANsmission (data set)
HMC	Hamiltonian Monte Carlo (method)
IGAC	International Global Atmospheric Chemistry
IGACO	Integrated Global Atmospheric Chemistry Observations
ILT	Independent Linear Trend
IMF	Intrinsic Mode Function (method)
IO3C	International Ozone Commission
JOSIE	Jülich Ozone Sonde Intercomparison Experiment (campaigns)
KC	Carbon iodine cell ozonesonde (instrument)
Lidar	LIght Detection And Ranging (instrument)
LOTUS	Long-term Ozone Trends and Uncertainties in the Stratosphere (research activity)
MC	Monte Carlo (method)
MCMC	Markov Chain Monte Carlo (method)
MERRA	Modern-Era Retrospective analysis for Research and Applications
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding (instrument)
MLO	Mauna Loa Observatory
MLR	Multiple Linear Regression (method)
MLS	Microwave Limb Sounder (instrument)
MWR	MicroWave Radiometer (instrument)
MZM	Monthly Zonal Mean
NAM	Northern Annular Mode
NAO	North Atlantic Oscillation
NASA	National Aeronautics and Space Administration
NCEP	National Centers for Environmental Prediction
NDACC	Network for the Detection of Atmospheric Composition Change
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration
NOAA ESRL	NOAA Earth System Research Laboratory
O3S-DQA	OzoneSonde Data Quality Assessment (research activity)
OCTAV-UTLS	Observed Composition Trends And Variability in the Upper Troposphere and Lower Stratosphere (research activity)
ODS	Ozone Depleting Substance
OHP	Observatoire de Haute-Provence
OMPS-LP	Ozone Mapping and Profiler Suite (instrument) - Limb Profiler (technique)
OSIRIS	Optical Spectrograph and InfraRed Imaging System (instrument)
Ozone_cci	Ozone Climate Change Initiative (research activity)
PC	Principal Component
Post-2000	January 2000 – December 2016 (or end of data record)
Pre-1997	January 1985 – December 1996
PWLT	Piecewise Linear Trend
QBO	Quasi-Biennial Oscillation
SABER	Sounding of the Atmosphere using Broadband Emission Radiometry (instrument)

SAGE II, III	Stratospheric Aerosol and Gas Experiment II and III (instruments)
SAM	Southern Annular Mode
SBUV, SBUV 2	Solar Backscatter Ultraviolet Radiometer (instrument)
SBUV MOD	SBUV Merged Ozone Data Set (data set)
SBUV COH	SBUV Cohesive (data set)
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (instrument)
SCISAT	SCIence SATellite (instrument)
SFIT2, SFIT 4	Spectral Fitting Algoritm version 2, version 4
SH	Southern Hemisphere
SHADOZ	Southern Hemisphere ADditional OZonesondes (network)
SI2N	SPARC/IO3C/IGACO/NDACC initiative
SMM	Station Monthly Mean
SMMa	Station Monthly Mean anomaly
SMO	SAGE II - MIPAS - OMPS-LP (data set)
SOO	SAGE II - OSIRIS - OMPS-LP (data set)
SPARC	Stratosphere-troposphere Processes And their Role in Climate
SSC	Station Seasonal Cycle
SST	Sea Surface Temperature
STS	Simultaneous Temporal and Spatial (regression method)
Suomi-NPP	Suomi National Polar-orbiting Partnership
SVD	Singular Value Decomposition (method)
SWOOSH	Stratospheric Water and OzOne Satellite Homogenized (data set)
SZA	Solar Zenith Angle
TUNER	Towards Unified Error Reporting (research activity)
UARS	Upper Air Research Satellite
UMK04	Umkehr ozone retrieval version of 2004 (algorithm)
UNEP	United Nations Environment Programme
USask 2D	University of Saskatchewan OMPS-LP 2D data set (algorithm)
UTLS	Upper Troposphere Lower Stratosphere
UV	UltraViolet
VMR	Volume Mixing Ratio
WMO	World Meteorological Organization
WOUDC	World Ozone and Ultraviolet Radiation Data Centre
WT	Wavelet Transform (method)
ZMMa	Zonal Monthly Mean anomaly

Table of contents

Chapter 1: Introduction

Ozone is a key species in the atmosphere, as it protects life on Earth by filtering out damaging ultraviolet (UV) radiation from the sun. The evaluation of the effect of ODSs on the long-term evolution of stratospheric ozone levels dates back to the late 1970s when the threat linked to increased levels of chlorine in the stratosphere due to industrial products started to emerge. Since then, ozone research has played a pioneering role in alerting on the impact of human activities on the global environment, promoting a direct link between science and policy action. In the early 1980s, atmospheric models predicted the maximum impact of ODSs on ozone in the upper stratosphere, around 40 km (*WMO,* 1985). The discovery of the Antarctic ozone hole changed this understanding of the ozone equilibrium in the stratosphere (*Farman et al.,* 1985; *Solomon et al.,* 1986). It became clear that chemical processes involved in polar ozone depletion could also affect ozone in the lower stratosphere at the global scale. Following the signing of the Montreal Protocol in 1987, a strategy was devised to accurately monitor the evolution of stratospheric ozone. Since 1989, the WMO/United Nations Environment Programme (UNEP) Ozone Assessments report on the state of the ozone layer and on the attribution of long-term changes in both the total column and the vertical distribution of ozone to ODSs, as well as on other direct and indirect processes that affect ozone levels, for example changing atmospheric temperature due to the accumulation of greenhouse gases (GHGs), or volcanic eruptions that inject aerosols into the stratosphere. In order to support these assessments, several international evaluation exercises have been organised. After the International Ozone Trend Panel (1988) that addressed the evaluation of trends from total column ozone measurements, the first SPARC/WMO report (*SPARC,* 1998) tackled the issue of trends in the vertical distribution of ozone. It provided a detailed description of both satellite and ground-based ozone profile measurement techniques, assessed their quality, and provided the first estimates of decreasing trends in stratospheric ozone since 1980 based on a combination of various data sources. This study showed statistically significant negative trends in ozone levels in the whole stratosphere, with two clear maxima in the upper stratosphere at around 40 km altitude and in the lowermost stratosphere at about 15 km (*SPARC,* 1998).

The accumulation of halogen compounds peaked in the stratosphere between the mid-1990s and the beginning of the 21st century, with a turnaround time that depends on altitude and latitude (*WMO,* 2014). As the decrease of ODS levels is becoming more apparent, it is now crucial to evaluate the success of the Montreal Protocol with regards to the recovery of the ozone layer. In addition, after 2000, the launch of several new satellite platforms dedicated to atmospheric composition measurements (*e.g.*, Earth observing system (EOS) Aura and EnviSat (Environmental Satellite)) provided new global ozone records. This resulted in the necessity of merging the various data records to determine the long-term evolution of ozone vertical distribution. The SPARC[1], IO3C[2], IGACO-O3[3], and NDACC[4] (SI2N) activity that started in 2011, aimed to evaluate the ozone profile trends from ground-based and satellite observations, including the new merged satellite data records. The activity was intended to contribute to the WMO/UNEP 2014 Assessment on the state of the ozone layer, but concluded only after the release of the report. From the combination of available ozone trend profiles, the Assessment reported a significant increase of ozone values by 2.5–5 % per decade around 35–45 km altitude at mid-latitudes and in the tropics during the period 2000–2013. These findings were in agreement with global climate model simulations that attributed the upper stratospheric ozone increase to both declining ODS levels and stratospheric cooling by increasing GHGs (*WMO,* 2014).

The SI2N team used another approach to determine the significance of ozone profile trends compared to that used for the WMO/UNEP 2014 Assessment. SI2N analyses re-evaluated long-term ozone profile trends from all available ozone records over the period 1979–2012 and results were published in *Harris et al.* (2015). For the period before the ODS peak, the reported trends were in good agreement with those reported in the 2014 Assessment. For the potential "recovery" period (*e.g.*, 1998–2012), positive trends of ~2 % per decade in mid-latitudes and ~3 % per decade in the tropics were found in the upper stratosphere from the combination of ozone profile trends, in somewhat broader agreement with *WMO* (2014). However, the significance of these combined increasing trends was investigated using several methods. The first one, similar to that used in *WMO* (2014), estimated the uncertainty of average trends from the weighted mean of the individual trends' standard deviations. The second used the joint distribution of the individual variances around the arithmetic mean of the estimators (*e.g.*, *SPARC,* 2013). The addition of uncertainty related to satellite drift, estimated from ground-based ozone profile measurement records (*Hubert et al.,* 2016), increased the total uncertainty of the trend and resulted in insignificant positive trends throughout the stratosphere, in contrast to results of the *WMO* (2014). The study concluded that it was too early to confirm the significant increase of ozone in the upper stratosphere, considering the length of the "recovery" period since 1998 and uncertainties in combined ozone trends.

[1] SPARC: Stratosphere-troposphere Processes And their Role in Climate; [2] IO3C: International Ozone Commission; [3] IGACO-O3: Integrated Global Atmospheric Chemistry Observations (Ozone); [4] NDACC: Network for the Detection of Atmospheric Composition Change.

The new SPARC, WMO and IO3C activity entitled "Long-term Ozone Trends and their Uncertainties in the Stratosphere (LOTUS)" was initiated in 2016 to revisit the causes of differences in the conclusions regarding the significance of post-turnaround ozone trends between the SI2N study and the 2014 WMO/UNEP Assessment. Since the end of the SI2N activity, historic ground-based and satellite time series have been revised and new merged satellite data sets have been produced. The main objectives of LOTUS are thus to assess the new long-term ozone profile records, provide a better understanding of all relevant uncertainties in the records, and revise the methods used to derive trends and their associated uncertainties.

This Report summarises the main results obtained during the first year of the LOTUS activity, which was targeted at providing timely inputs to the 2018 WMO/UNEP Ozone Assessment. This deadline defined the scope of this Report and focused the LOTUS activities primarily on changes in ozone levels in the middle and the upper stratosphere outside the polar regions as observed by merged satellite records. These studies were complemented by an analysis of ground-based data and recent model data provided by the CCMI (*Morgenstern et al.,* 2017). The short timeline of this work necessitated leaving several pressing topics for the second phase of LOTUS, most notably: the significance of recent trends in the lower stratosphere, the attribution of trends to ODSs and GHGs using model data, and estimates of trends in the polar regions.

The structure of this Report is as follows: *Chapter 2* describes improvements in the historic data records and the newly released data sets for both ground-based and satellite measurements. Challenges for trend studies are addressed in *Chapter 3* and include the evaluation of sampling biases and drifts between the various time series. Methodologies designed to assess the stability of the various satellite records are also described. *Chapter 4* describes the regression methodology used in LOTUS and related sensitivity tests, while *Chapter 5* provides updated ozone profile trends from the various records. A substantial part of this chapter is devoted to an assessment of the methods used to combine trends from the individual records and estimate the significance of the combined trends. The main conclusions of the LOTUS study are drawn in the *Executive Summary* and remaining open issues are listed.

Chapter 2: Observations and model data

This chapter describes the data sets used to create the ozone profile climate data records that were intercompared and analysed for trends in later chapters in this Report. *Section 2.1* describes the ozone profile data sets from ground-based and in situ instruments, *Section 2.2* is dedicated to ozone data records from satellite instruments. Finally, *Section 2.3* describes the ozone profiles produced by the chemistry-climate and chemistry-transport models of the CCMI.

2.1 Ground-based observations

We start with a brief review of the measurement techniques, data characteristics, and recent changes in the ozone profile data records collected by ground-based instrumentation. More detailed information can be found in *Hassler et al.* (2014) and references therein. The second part of this section presents the methods used to create monthly zonal mean data from these ground-based records, highlighting data set-specific limitations in spatial and temporal sampling. These broad-band, zonally averaged anomaly time series are the input to the trend analyses in *Chapter 5*.

2.1.1 Measurement techniques

2.1.1.1 Ozonesonde

Ozonesondes are a widely used method for measuring in situ ozone vertical distributions up to altitudes of 30–35 km. The balloon-borne electrochemical ozonesondes are small, lightweight, and compact instruments and ozonesonde records at several measurement stations provide the longest ozone profile time series available, with some starting in the 1960s. Ozone profiles are obtained with a height resolution of about 100–150 m. The sensing device is interfaced to a standard meteorological radiosonde for data transmission to the ground station and additional measurements of pressure, temperature, and wind speed. Three major types of ozonesondes have been in use since the 1970s (*e.g.*, *Smit,* 2012a): Electrochemical concentration cell (ECC), Brewer-Mast (BM), and carbon iodine cell (KC). Nowadays most stations have adopted the ECC ozonesonde type developed by *Komhyr* (1969).

A comprehensive review of the performance of the different ozonesondes in terms of precision and accuracy is given in *SPARC-IOC-GAW Assessment of Trends in the Vertical Distribution of Ozone* (*SPARC*, 1998). The assessment also showed inconsistencies in trends derived from data gathered from different sounding stations. A summary and update of the review have been given by *Hassler et al.* (2014) as part of the SI2N assessment. Overall, in recent decades, the random error component of sonde measurements is generally within ±5–10 % between the tropopause and altitudes less than 26 km for all types of sondes. Systematic biases between all types of ozonesondes or compared to other ozone sensing techniques are smaller than ±5–10 %. Above about 26 km altitude the results are not conclusive and the measurement behavior of the sonde types differs. The uncertainty at the top of the measured profile depends on the type of ozonesonde and sensor solution. For example, BM sondes systematically underestimate ozone with increasing altitude (*i.e.*, -15 % at 30 km altitude) (*De Backer et al.,* 1998; *Stuebi et al.,* 2008), while KC sondes tend to overestimate ozone by 10–20 % at altitudes above 30 km (*SPARC,* 1998; *Deshler et al.,* 2008). Intercomparison studies (*e.g.*, *Smit et al.,* 2007; *Smit and ASOPOS panel,* 2012b) indicate that the response of ECC sondes between 28 km and 35 km depends on the type of ECC sonde and sensing solution applied (*i.e.*, 10–20 % differences at altitudes near 35 km).

However, laboratory studies (*Johnson et al.,* 2002) and international intercomparisons like the Jülich Ozone Sonde Intercomparison Experiment (JOSIE; *Smit et al.,* 2007) and the Balloon Experiment on Standards for Ozone Sondes (BESOS; *Deshler et al.,* 2008) have also clearly demonstrated that even small differences in sensing techniques, sensor types, or sensing solutions can introduce significant inhomogeneities in the long-term sounding records between different sounding stations or within each station individually. Therefore, existing artifacts in long-term sounding records have to be resolved by homogenisation either in space (between different stations) or in time (long-term changes) through the use of generic transfer functions which have been derived from intercomparison experiments (*e.g.*, JOSIE or BESOS) and dual balloon soundings (*Deshler et al.,* 2017). A major goal of the Ozone Sonde Data Quality Assessment (O3S-DQA), which is part of this LOTUS assessment, is to reduce the uncertainties between long-term sounding records from 10–20 % down to 5–10 % through the use of generic transfer functions (*Smit and O3S-DQA panel,* 2012). Currently, a total of about 30 long-term station records have been reevaluated and homogenised through resolving known instrumental bias effects, thereby reducing the uncertainties down to 5–10 % (*Tarasick et al.,* 2016; *Van Malderen et al.,* 2016; *Deshler et al.,* 2017; *Sterling et al.,* 2018; *Witte et al.,* 2017, 2018). Some of these recently homogenised ozonesonde data sets are part of this LOTUS assessment.

There is still a potential for sudden changes in future records that can be created by abrupt radiosonde changes (often due to financial burden) or manufacturing changes, which has caused problems in the past. The ozonesonde community, including sonde manufacturers and station operators, recently performed a new JOSIE campaign where they assessed the methods and techniques used by stations in the Southern Hemisphere ADditional OZonesondes (SHADOZ) network. These exercises help to identify inconsistencies in operations and resolve changes to the stability of the record.

2.1.1.2 Lidar

Ozone lidar (Light Detection and Ranging) vertical distribution measurements are based on the Differential Absorption Lidar (DIAL) method that uses the emission of two laser wavelengths (so-called "on" and "off" wavelengths) characterised by a different ozone absorption cross-section. Range resolved measurements are provided by the use of pulsed lasers. The ozone number density is retrieved from the slope of the lidar signals originating from the atmospheric scattering of both laser wavelengths towards the optical receiving system. These signals have to be corrected for differential Rayleigh and Mie scattering as well as for differential absorption by other constituents. The laser wavelengths are chosen so that these corrections represent less than 10 % of the main term linked to ozone absorption. For stratospheric ozone measurements, the on-wavelength is usually generated by an XeCl excimer laser at 308 nm. For the non-absorbed wavelength, different techniques are used, among which the most common are the generation of a wavelength at 353 nm by stimulated Raman scattering in a cell filled with hydrogen or the use of the third harmonic of a Nd:Yag laser emission (355 nm). A more detailed description of the ozone lidar measurement technique can be found, for example, in *Mégie and Menzies* (1980), *Pelon et al.* (1986), and *Godin-Beekmann et al.* (2003). Long-term ozone lidar measurements are currently performed at several stations of the NDACC. Data records of more than 20 years are available at Haute-Provence Observatory (France), Hohenpeissenberg (Germany), Table Mountain (California, USA), Mauna Loa Observatory (MLO; Hawaii, USA) and Lauder (New Zealand). In addition, standardised definitions for the vertical resolution and uncertainty budget of the NDACC lidar ozone measurements were recently published (*Leblanc et al.,* 2016a, 2016b). The uncertainty in lidar ozone profiles ranges from a few percent below 20 km to more than 10–15 % above 45 km with vertical resolution decreasing as a function of altitude, ranging from ~0.5 km below 20 km to several kilometers above 40 km (*Godin et al.,* 1999; *Leblanc and McDermid,* 2000; *Leblanc et al.,* 2016b). Instrumental artifacts that can affect the stability of a long-term lidar record include changes in optical receiver configuration and alignment of the laser beams within the field of view of the telescope (impacting the slope of the lidar signals), changes in laser power, and changes in telescope area (impacting mainly the top of the profiles). Concerning the ozone number density retrieval, undocumented changes of ozone absorption cross-section values between data processing versions can introduce systematic biases throughout the profile.

2.1.1.3 Microwave radiometer

Microwave ozone radiometers (MWR) measure the spectra of emission lines produced by thermally excited, purely rotational ozone transitions at millimeter wavelengths. The pressure broadening effect of the line allows the retrieval of a vertical ozone profile from the measured spectrum by the use of an a priori profile, a radiative transfer simulation and the optimal estimation method based on *Rodgers* (2000). The rotational ozone transitions are measured at either 142.175 GHz or 110.836 GHz depending on the instrument. The instrument principally consists of a millimeter wave receiver and multichannel spectrometer. The measured signal is amplified and down-converted to a lower intermediate frequency which can be processed by a spectrometer. The instruments are calibrated by substituting the radiation from the sky with the thermal radiation from two black body sources at the receiver input. One source is at ambient temperature or heated and stabilised (~300 K) and the second source is cooled with liquid nitrogen at 77 K. The attenuation of the ozone signal in the troposphere is determined by measuring the tropospheric thermal emission and relating the tropospheric opacity to its emission using a radiative transfer model (*Hocke et al,* 2007; *Hassler et al,* 2014).

Ozone profiles between 20 km and 70 km altitude are given in volume mixing ratio (VMR; given in ppmv) and the pressure grid on which data are provided varies by instrument. The vertical resolution is typically 8–10 km between 20 km and 40 km, increasing to 15–20 km at 60 km (*Studer et al.,* 2013; *Nedoluha et al,* 2015, *Maillard-Barras et al.*, 2009).

The total error includes systematic error, random error, and the smoothing error term, which can be determined for each ozone profile. Based on a standard integration time of 1 h, the random and systematic errors are on the order of 3–5 % while the total error is on the order of 7–10 % in the stratosphere. The total error increases up to 20 % at 20 km and to 30–35 % at 70 km (*Studer et al.,* 2014). Lauder MWR agrees within 5–10 % with lidars and the Stratospheric Aerosol and Gas Experiment (SAGE) II between 22 km and 43 km (*McDermid et al.,* 1998). MLO MWR agrees within 10 % with lidars, Dobson Umkehr, and the Upper Air Research Satellite (UARS) Microwave Limb Sounder (MLS) at almost all altitudes, and it agrees within 5 % in the 20–45 km region (*McPeters et al.,* 1999). The 15-year climatological mean difference between the Bern (*Studer et al.,* 2013) and Payerne MWRs is within 7 % from 25 km to 65 km (*Eliane Maillard-Barras, private comm.*). Additional information on microwave radiometers can be found in *Hassler et al.* (2014).

2.1.1.4 FTIR

The ground-based FTIR (Fourier-Transform InfraRed) ozone observations are coordinated by the Infrared Working Group of NDACC. Within this network, the measurements are performed over the 600–4500 cm^{-1} spectral range, using primarily high-resolution spectrometers such as the Bruker 120M (or 125M) or Bruker 120HR (or 125HR), which can achieve a spectral resolution of 0.0026 cm^{-1}. The source of light being the sun, the spectra are recorded only during day-time and under clear sky conditions. The average number of measurements per day among all the stations is about two, with a mean of eight days of measurements per month. Despite the dependence on weather conditions, the average number of measurements per month remains very stable over the full FTIR time series.

In addition to total columns retrieved from the absorption line areas, low vertical resolution profiles can be obtained from the temperature and pressure dependence of the line shapes. The absorption line shapes also depend on the instrumental line shape, with the latter needing to be monitored regularly using gas cell measurements, and are retrieved in a harmonised way within the network (*Hase et al.,* 1999).

The profile retrievals are derived using one of the two different algorithms: PROFITT9 (*Hase,* 2000) and SFIT2 or its recent update SFIT4 (*Pougatchev et al.,*1995), both based on the optimal estimation method (*Rodgers,* 2000). The retrieval settings (*i.e.*, spectral window optimised for ozone, a priori information, *etc.*) have been harmonised within the network (see *Vigouroux et al.,* 2015 for details). There are four or five degrees of freedom for signal (DOFS) from the ground up to about 45 km. Four layers with about one DOFS can therefore be defined to provide partial columns with almost independent information (see **Figure 1** of *Vigouroux et al.,* 2015): roughly one in the troposphere and three in the stratosphere. The three partial columns used in LOTUS are located in the following altitude ranges: 12–21 km; 21–29 km; and 29–48 km, according to the FTIR vertical resolution which is between 7 km and 15–20 km depending on altitude.

The random uncertainties on these three partial columns is about 5 % (*Vigouroux et al.,* 2015; details on error budget can also be found in *García et al.,* 2012). The systematic uncertainty is dominated by the uncertainties on the spectroscopic parameters (HIgh Resolution TRANsmission (HITRAN) 2008 for the current analysis) and is about 3 % for each partial column.

Instruments and data archives	Stations (start of data record)		
	60°S –35°S	**20°S – 20°N**	**35°N – 60°N**
Ozonesonde (0–30km) http://www.ndacc.org, http://www.woudc.org/data, https://tropo.gsfc.nasa.gov/shadoz/Archive.html	Lauder (1986), Macquarie Island (1994), Broadmeadows (1999)	Hilo (1982), Ascension Island (1998), Nairobi (1998), Natal (1998), Pago Pago (1998), Kuala Lumpur (1998), Suva (1998), Hong Kong Observatory* (2000, 22.3°N)	Goose Bay (1963), Uccle (1965), Hohenpeißenberg (1966), Payerne (1968), Edmonton (1970), Wallops Island (1970), Lindenberg (1975), Legionowo (1979), Praha (1979), Boulder (1991), De Bilt (1992), Valentia (1994), Huntsville* (1999, 34.7°N)
Lidar (10–50km) http://www.ndacc.org	Lauder (1994)	Mauna Loa (1993)	OHP (1985), Hohenpeißenberg (1987), Table Mountain (1988)
Microwave (20–70km) http://www.ndacc.org	Lauder (1992)	Mauna Loa (1995)	Bern (1994), Payerne (2000)
FTIR (0–50km) http://www.ndacc.org	Lauder (2001), Wollongong (1996)	Izana* (1999, 28.3°N)	Jungfraujoch (1995)
Dobson/Brewer Umkehr (0–50km) ftp://aftp.cmdl.noaa.gov/data/ozwv/DobsonUmkehr/Stray%20light%20corrected/monthlymean	Perth (1984), Lauder (1987)	Mauna Loa (1984)	Arosa (1956), Boulder (1984), OHP (1984)

Table 2.1: *Overview of the sources of ozone profile observations by ground-based techniques used for the monthly zonal mean data considered in this Report. Stations are sorted chronologically by start year of the record; those with an asterisk are located slightly outside the attributed latitude zones.*

2.1.1.5 Umkehr

The Umkehr measurement makes use of the zenith sky UV radiation changes during the sunset or sunrise hours of the day. The earliest measurements were recorded by *Götz* (1931) at Arosa, Switzerland. Two spectral UV wavelength regions (311.4 nm and 332.5 nm) selected for the Umkehr method are subject to different levels of ozone absorption. The zenith sky ratio between two spectral channels changes with the elevation of the sun. Measurements begin at 60° solar zenith angle (SZA); the ratio gradually increases up to 85° SZA and then decreases between 85° and 90° SZA. Umkehr measurements from Dobson instruments have been collected operationally since the 1957 International Geophysical Year at a few select stations, but additional Dobson observing stations became available in the 1980s. The trend-optimised algorithm was developed by *Petropavlovskikh et al.* (2005) to derive morning and afternoon daily ozone profiles in 10 Umkehr layers based on a pressure layer system. A priori information is used to solve the optimum statistical inverse problem (*Rodgers,* 2000). The method is designed to derive ozone profiles with a vertical smoothing technique (defined by averaging kernels). This approach affects the accuracy of the retrieved ozone in a particular layer by weighting ozone variability from adjacent layers. Therefore, the method adds error in the layer-retrieved ozone amount, which is estimated to be about 5 % in the stratosphere. However, this error does not impact trend analyses as it is constant in time. Time series of Umkehr ozone profiles are retrieved with the UMK04 algorithm (*Petropavlovskikh et al.,* 2005). A generic stray light correction is applied to reduce systematic biases in Umkehr retrieved profiles (*Petropavlovskikh et al.,* 2009, 2011).

2.1.2 Deseasonalised monthly mean time series

2.1.2.1 Procedure

The trend analyses in *Chapter 5* are performed on monthly averaged deseasonalised data collected at a number of stations and by five types of instruments (ozonesonde, lidar, FTIR, MWR, and Umkehr). Profiles from each instrument record are first averaged by month, separately for each station. Months with an insufficient number of profiles are discarded from further analysis. The selection criteria depend on the instrument technique (*Sections 2.1.2.2–5*). These time series created for each instrument are referred to as the Station Monthly Mean (SMM) dataset. Subsequently, the deseasonalisation process is performed in two steps.

In step 1, a site-specific seasonal cycle is computed as the average, for each calendar month (Jan, Feb, …, Dec), of all SMM data in the reference period (Jan 1998 – Dec 2008). Months with an insufficient number of years over the reference period (typically <7 but also measurement technique dependent, see below) are flagged and excluded from further analysis. This requirement ensures a more accurate determination of the observed seasonal cycle but is only satisfied for a select number of sites. This data set is referred to as the Station Seasonal Cycle (SSC) data set, one per site and per instrument.

In step 2, we compute the relative difference of each monthly mean value (SMM) to the observed climatological mean value (SSC) for that month. These deseasonalised relative anomaly time series are referred to as the Station Monthly Mean Anomaly data set (SMMa) and are defined as

$$SMMa(site,p,t) = 100 \times \frac{SMM(site,p,t) - SSC\big(site,p,m(t)\big)}{SSC\big(site,p,m(t)\big)} \quad (2.1),$$

where p stands for vertical grid level (pressure or altitude), t represents time (*i.e.*, month) and *m(t)* the corresponding calendar month (*i.e.*, Jan, Feb, …). Hence, by construction, the (dominant part of the) seasonal cycle is removed from SMMa data and the absolute level averages to zero over the reference period. In addition, any instrument-related constant multiplicative offsets (*i.e.*, bias) are thereby removed as well.

The deseasonalisation step is motivated by the need to combine, for a wide latitude belt, the data from multiple sites, each potentially exhibiting a different bias. A Zonal Monthly Mean Anomaly data set (ZMMa) is obtained by averaging the SMMa data from each station located within the broad zonal bands. We create ZMMa's for three broad latitude bands: 60°S–35°S, 20°S–20°N, and 35°N–60°N. Only the stations listed in **Table 2.1** are used for the broad zonal bands. **Figure 2.1** demonstrates ozone anomaly time series of ZMMa ozonesonde records in the 35°N–60°N (top), 20°S–20°N (centre), and 60°S–35°S (bottom) latitude bands as a function of altitude (ground to ~30 km).

Site-dependent instrument biases can generate, in a multi-station average of SMM data sets, not only random uncertainty but also discontinuities (due to differences in time coverage). However, such sources of error are suppressed in a multi-station average of SMMa data sets (the ZMMa data sets). The intercomparisons described in *Chapter 3* (*Section 3.1.1*) identify a number of stations with clear inhomogeneities in the time series. The availability of multiple sites is expected to reduce the impact of spatial and temporal inhomogeneities in the combined ground-based data records. However, it is important to realise that it is not uncommon that a latitude belt contains just one site for the considered measurement technique.

ZMMa are created for each instrument technique separately and with equal weight given to all sites within the band (**Figures 2.2–2.6**). This effectively gives more weight to regions with more stations (*e.g.*, Europe and North America). The data from different instrument techniques are not combined in this study due to complications associated with differences in sampling frequencies, vertical smoothing and the use of different measurement units. The time series of ground-based station and zonally averaged ozone anomalies are available from the LOTUS Report data depository.

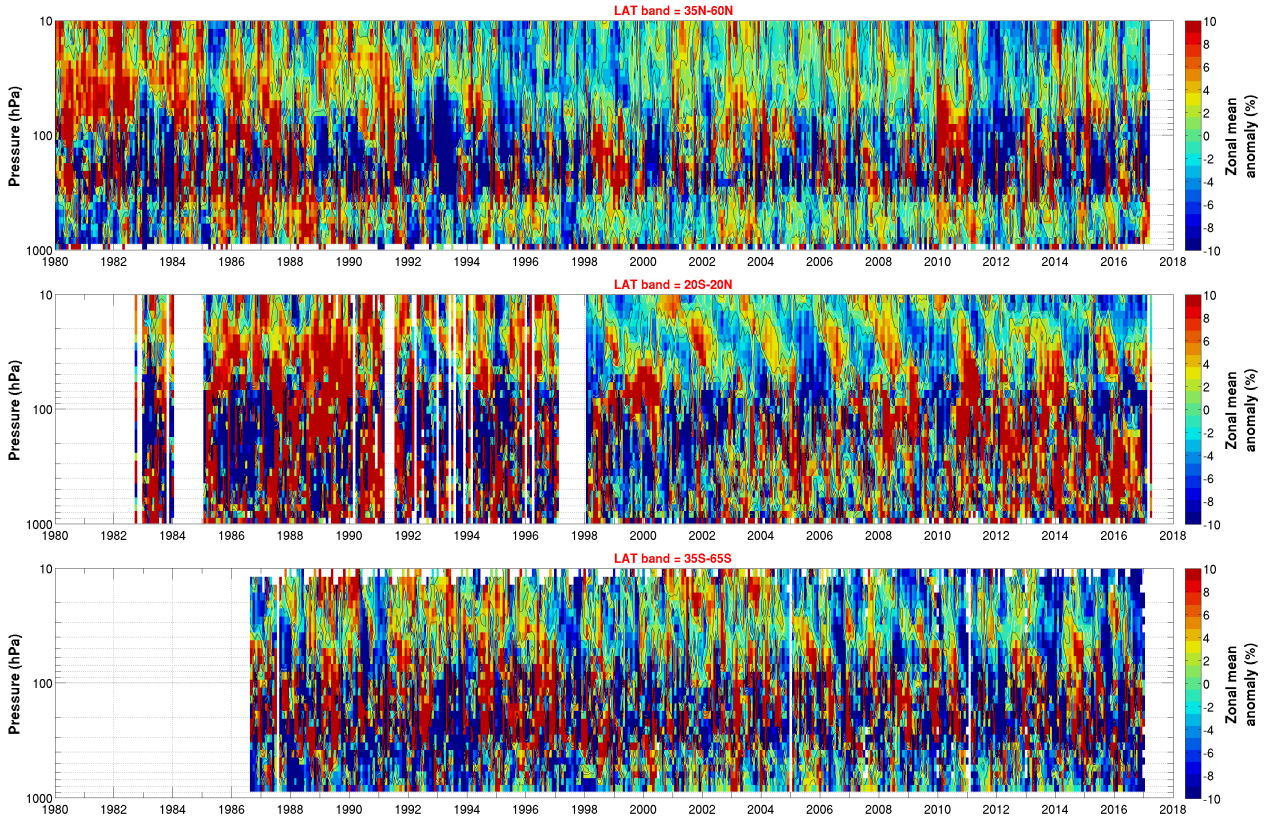


Figure 2.1: *Example time series of monthly zonal mean relative deseasonalised anomalies computed from ozonesonde data in the 35°N–60°N (top), 20°S–20°N (centre), and 60°S–35°S (bottom) latitude bands as a function of altitude (ground to ~30 km). The sonde stations used for each band are listed in* ***Table 2.1****.*

Trends derived from each ZMMa time series are reported in *Chapter 5*, *Section 5.4*. The only ground-based records considered individually in this Report are the ozonesonde records from Hilo, Hawaii and Lauder, New Zealand. The trends derived from these two ozonesonde records are used for discussion of consistency in trends obtained from multiple instruments co-located at these locations (see *Chapter 5*, **Figure 5.10**).

2.1.2.2 Ozonesonde

Ozonesonde observations were retrieved from the public NDACC, World Ozone and Ultraviolet Radiation Data Centre (WOUDC), and SHADOZ data archives. The station data record differs sometimes between archives, due to different processing settings, different time periods covered, *etc*. Therefore, for a given site, the data from different archives was not mixed in order to avoid introducing inhomogeneities. Only half of the sites report total ozone normalisation correction factor (CF) values, some of these have applied the CF to the original profiles while others have not. To avoid losing a large number of sites where the CF data is missing, this information is not used to correct the reported data nor to screen the observations. Instead, the data are screened according to the criteria outlined in *Hubert et al.* (2016). German Democratic Republic sondes (GDR), mainly flown prior to the 1990s in Eastern Europe, have larger uncertainties and these profiles are hence not used (*Liu et al.*, 2009). Flights that do not reach 20 hPa are rejected as well, to avoid additional uncertainty in case the profile was normalised to a total ozone column. The VMR profile is then integrated in the pressure domain to obtain ozone partial columns of ~1 km thickness from the surface to 30 km altitude. The entire profile is discarded if at least 10 out of 30 layers are missing (quality-screened) input data.

2.1.2.3 Lidar and microwave radiometer

The monthly mean ozone profiles for lidar and microwave observations are obtained by averaging the ozone profiles available in the NDACC database (**www.ndacc.org**). For most stations, we used the profiles from the (monthly) National Aeronautics and Space Administration (NASA) Ames data files, while the recent profiles by the Bern MWR were taken from the hierarchical data format data files. Monthly mean ozone profiles for the Hohenpeißenberg lidar were obtained in a slightly different way, by retrieving the monthly mean lidar return signal (which results in some improvement above 40 km). Individual profiles are weighted by measurement length. Most stations report profiles as number density (10^{16} m^{-3}) versus altitude. For the historic microwave ozone data from Bern and Payerne stations, however, only VMR versus altitude are available.

The altitude resolution of individual microwave profiles is on the order of 10 km to 20 km. For the lidars it varies between ~0.5 km (at 15 km) to more than 5 km (above 40 km).

For monthly means, the altitude resolution is less relevant, because atmospheric changes tend to average out and tend to be coherent over many kilometers.

Three lidar and two microwave stations are available for the 35°N–60°N broad-band averages, whereas for 20°S–20°N and 60°S–35°S broad bands only single station records are available for comparisons with satellite records (see *Section 5.4*).

2.1.2.4 FTIR

As mentioned previously, FTIR solar absorption measurements are taken during the day only and only during clear-sky conditions. There are on average about three measurements per day and eight days of measurements per month. The random errors are determined by the smoothing error, which is one of the dominating error sources in FTIR profile retrievals (*Vigouroux et al.,* 2015) and is about 5 % for the three layers provided for LOTUS analyses. The systematic errors are about 3 % for the three layers. The standard deviation of the monthly means and the number of measurements used in the monthly means is also provided in the FTIR datafiles.

Two FTIR records are averaged to represent 60°S–35°S broad-band ozone variability and trends. Single station records are available for comparisons in the other two broad latitude bands (see case study in *Section 5.4*).

2.1.2.5 Umkehr

Monthly averages for Umkehr time series are calculated from all data that have passed the quality assurance (*i.e.*, iterations less than three, standard deviation of the difference between Umkehr simulated and observed values in the final retrieval less than observation uncertainty). Umkehr measurements in the years following the eruptions of El Chichón (1982–1984) and Pinatubo (1991–1993) were affected by scattering from aerosols injected into the stratosphere. These effects are not taken into account by the forward model, thus creating erroneous ozone profile retrievals. The post-processing corrections do not remove errors completely. Therefore, for trend analyses the monthly averaged Umkehr data during volcanic periods are marked as missing.

Three Umkehr records are available for the 35°N–60°N broad-band averages and two stations are used to represent the 60°S–35°S belt, whereas for 20°S–20°N only a single station record is available for comparisons with satellite records (see case study in *Section 5.4*).

2.1.2.6 Instrument and station measurement frequency

Figures 2.2 to **2.6** show the number of measurements per month for the ozonesonde, lidar, MWR, FTIR, and Umkehr techniques at all stations that are used for trend analyses in this Report. Frequency of observations varies from station to station over the records, which likely depends on the fluctuation in funding available from the supporting national programs. The minimum number of observations (two or more) required to accept a monthly mean value in the SMM dataset (see *Section 2.1.2.1*) depends on the instrument technique. In part, these rather low numbers (when compared to what is used by the satellite community) reflect limitations due to observational conditions and the sonde launch schedule.

Ozonesondes (**Figure 2.2**) are launched in all weather conditions, typically following a fixed schedule on the same day(s) of the week or month. Three European stations (Payerne, Uccle, and Hohenpeissenberg) launch sondes three times a week, while most stations do so once a week. The SHADOZ sites, located in the tropics, launch twice a month. Uncertainties in the derived monthly mean values are reduced by rejecting months and grid levels with <2 (tropics) or <3 (elsewhere) observations. Seasonal cycle entries for ozonesonde records are discarded for months and grid levels that contain <6 years (tropics) or <7 years (elsewhere) of SMM data over the reference period.

For the ground-based observations, at least two measurements are required for lidar (**Figure 2.3**), microwave (**Figure 2.4**), and Umkehr (**Figure 2.5**), while at least three measurements are required for FTIR (**Figure 2.6**). Lidars measure during clear-sky nights only and report just one profile per night. Microwave radiometers, on the other hand, measure continuously under most weather conditions and report half hourly, hourly, or six hourly profiles depending on the site. Umkehr profiles are retrieved on days of (mostly) clear sky conditions and can have two measurements per day. However, each station will have a different maximum number of days per month depending on local weather conditions (*e.g.*, overcast). The FTIR measurements also require fair weather conditions and therefore have a similar limitation on the number of profiles per month, which vary for latitude and season. Seasonal cycle entries for ground-based records are discarded for months and grid levels that contain <6 years (tropics) or <7 years (elsewhere) of SMM data over the reference period.

The non-uniform temporal sampling can have an impact on the seasonal cycle derived from each instrument record and its ability to capture the true atmospheric variability. Since composition in the lower stratosphere is strongly affected by meteorological scale variability (*Lin et al.,* 2015), the impact of the sampling frequencies on the station record seasonal cycle should be assessed for each ground-based instrument in this part of the atmosphere. Prior to trend analyses, each ground-based and ozonesonde record is deseasonalised separately prior to combining anomalies; thus, a sampling bias is expected to have small impact on the combined records and derived trends. At the time of this writing, no detailed studies were available on the impact of sampling on differences in ground-based trends. These are recommended for future analyses.

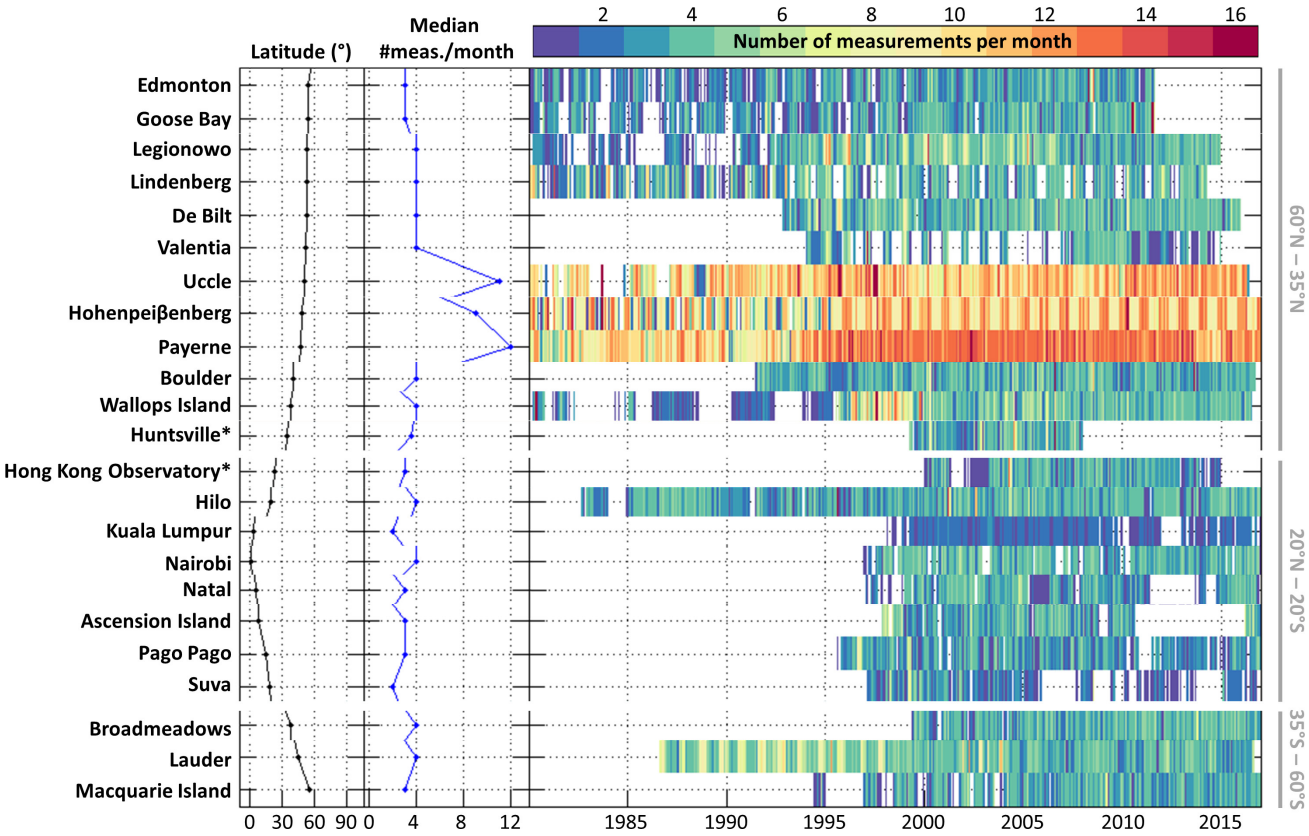


Figure 2.2: *Sampling statistics for ozonesonde station records retrieved from the NDACC, WOUDC, and SHADOZ data archives, sorted North to South. The figure shows the median number of measurements per month over the entire data record (centre) and the number of measurements for each month since 1980 (right, colour scale). Stations with an asterisk are located slightly outside the attributed latitude zones.*

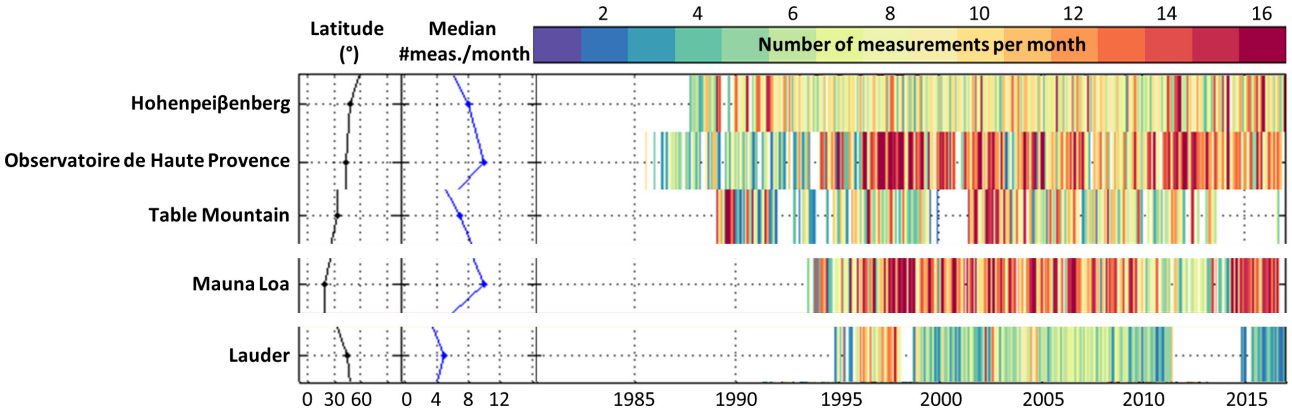


Figure 2.3: *As **Figure 2.2** but for the stratospheric ozone lidar station records retrieved from the NDACC data archive.*

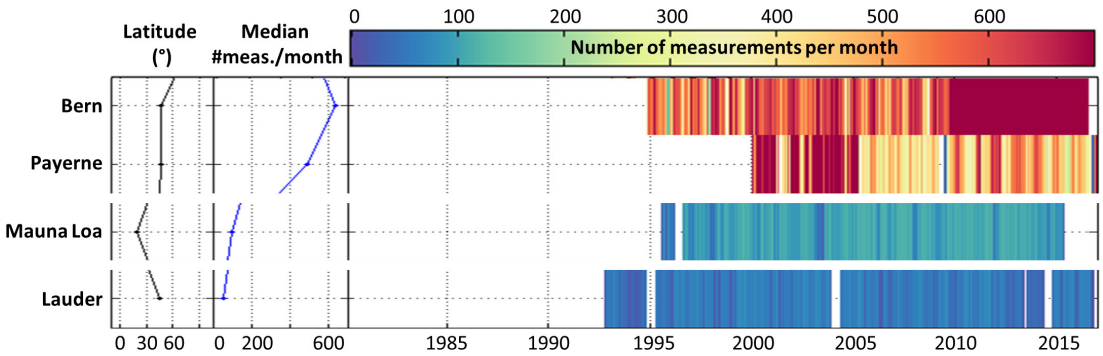


Figure 2.4: *As **Figure 2.2** but for ozone microwave radiometer station records retrieved from the NDACC data archive. Stations report half hourly, hourly, or six-hourly profiles.*

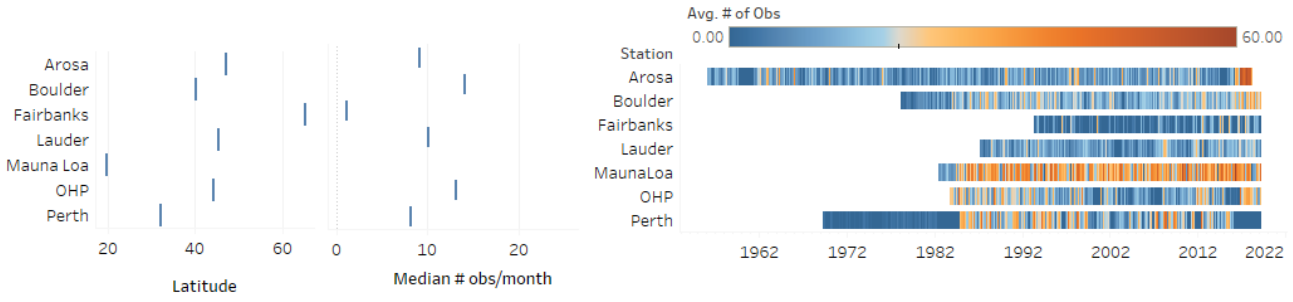


Figure 2.5: *As* ***Figure 2.2*** *but for Dobson Umkehr station records submitted by the record PIs to the LOTUS data archive. Stations report profiles once or twice a day. Note that the time axis differs from that of previous figures.*

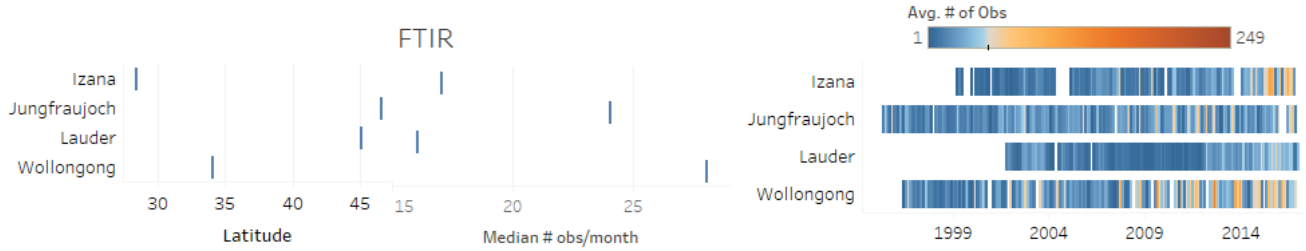


Figure 2.6: *As* ***Figure 2.2*** *but for FTIR station records submitted by the record PIs to the LOTUS data archive. Stations report profiles several times per day. Note that the time axis differs from that of previous figures.*

2.2 Satellite observations

2.2.1 General remarks

The main advantage of satellite instruments is their global coverage. For ozone trend analyses, long-term ozone data sets are needed in order to separate long-term trends from other sources of ozone variability such as solar activity. For the 2014 ozone assessment (*WMO,* 2014), several merged satellite data sets were created: SBUV Merged Ozone Data Set (SBUV MOD) and the SBUV Cohesive data set (SBUV COH), Global OZone Chemistry And Related trace gas Data records for the Stratosphere (GOZCARDS) and Stratospheric Water and OzOne Satellite Homogenized (SWOOSH), as well as SAGE-GOMOS (Global Ozone Monitoring by Occultation of Stars), and SAGE-OSIRIS (Optical Spectrograph and InfraRed Imaging System). Detailed information about these data sets and their intercomparison can be found in *Tummon et al.* (2015). An overview of satellite instruments can be found in for example *Hassler et al.* (2014).

Since the 2014 WMO Ozone Assessment, some of these merged data sets have been extended to 2016 and updated with the most recently processed versions of the ozone profile data sets from the individual satellite instruments. In addition, new merged data sets have been generated. These new merged data sets use revised data records from the individual instruments and rely on improved merging methods. The LOTUS Report evaluates these new data sets for improved accuracy and stability.

This section briefly describes the long-term merged ozone profiles data sets used in the LOTUS study. General information about the merged data sets and their main parameters is summarised in **Table 2.2**. According to measurement technique and ozone representation, the merged satellite data sets are grouped as (1) ozone profiles from nadir sensors, (2) ozone profiles from limb instruments in mixing ratio on a pressure grid, and (3) ozone profiles from limb instruments in number density on an altitude grid. In addition to measurement principles and specific features of retrieval algorithms, such a grouping is also made because ozone trends can be different in different representations due to the influence of stratospheric cooling (*McLinden and Fioletov,* 2011). The influence of the ozone representation on evaluated trends is discussed in *Chapter 5, Section 5.1.2* of the Report. For all satellite data sets, monthly zonal mean ozone profiles are used.

2.2.2 Nadir profile data records

The two nadir-based merged profile data sets in this Report are both based on the series of nine solar backscatter UV (Backscatter Ultraviolet Radiometer (BUV), SBUV and SBUV/2) nadir instruments flown over the period from 1970 to the present on NASA (*i.e.*, Nimbus 4 and Nimbus 7) and National Oceanic and Atmospheric Administration (NOAA; *i.e.*, NOAAs 9, 11, 14, 16, 17, 18, and 19) satellite platforms. The instruments are of

similar design and measurements are processed using the same retrieval algorithm (Version 8.6; *McPeters et al.,* 2013; *Bhartia et al.,* 2013). Radiance measurements are calibrated using a variety of hard and soft calibration techniques, including cross-instrument calibration during periods of measurement overlap to further ensure consistency over the record (*DeLand et al.,* 2012). However, despite the instrument similarity and common retrieval algorithm, each instrument experienced unique operational conditions (*e.g.*, instrument degradation, specific on-orbit problems) and orbital characteristics, including measurement time of day, which contribute to differences among the individual records.

SBUV instruments ideally operate in late morning-early afternoon sun synchronous orbits such that measurements are made at small solar zenith angles and at the same local time each orbit. While most instruments were launched into ~2 pm local time orbits, Nimbus 4 and Nimbus 7 measured near noon local time, and NOAA 17 was launched into a ~10 am orbit. Furthermore, NOAA satellite orbits slowly drift towards the terminator, and in some cases drift through the terminator, such that the instrument evolves from making late afternoon measurements to making early morning measurements. Thus the various SBUV instruments are measuring at different local times. This can introduce differences between overlapping measurements due to both real geophysical noise (*e.g.*, diurnal variation) and instrument noise, as the data uncertainty increases when the orbit approaches the terminator (*DeLand et al.,* 2012; *Kramarova et al.,* 2013a; *McPeters et al.,* 2013). The latter is particularly true of the NOAA-9, -11, and -14 instruments, whose orbits drifted faster than other instruments in the series (*DeLand et al.,* 2012; *Kramarova et al.,* 2013a).

The primary source of error in the SBUV retrieval is the smoothing error due to the instrument's limited vertical resolution, particularly in the troposphere and lower stratosphere (*Kramarova et al.,* 2013b; *Bhartia et al.,* 2013). The SBUV instrument has a resolution of 6–7 km near 3 hPa, degrading to 15 km in the troposphere and ~10 km above 1 hPa (*Bhartia et al.,* 2013). *Kramarova et al.* (2013a) showed that SBUV ozone profiles are generally consistent to within 5 % with data from UARS and Aura MLS, SAGE II, ozonesondes, microwave spectrometers, and lidar in the region between 25 hPa and 1 hPa (also see *Frith et al.,* 2017 for updated comparisons with AURA MLS).

Data set	Satellite instruments	Ozone representation	Latitude coverage and resolution	Altitude coverage and vertical sampling	Temporal coverage
SBUV MOD v8.6 (NASA) https://acd-ext.gsfc.nasa.gov/Data_services/merged/index.html	BUV, SBUV and SBUV-2 on Nimbus 4, 7 and NOAAs 11, 14, 16, 17, 18 ,19	Mixing ratio on a pressure grid	80S–80N, 5 deg	50–0.5 hPa, 15 layers (from ~6 to ~15 km)	01/1970 – 12/2016
SBUV COH v8.6 (NOAA) ftp://ftp.cpc.ncep.noaa.gov/SBUV_CDR	SBUV and SBUV-2 on Nimbus- 7 and NOAAs 9, 11, 16, 17, 18, 19		80S–80N, 5 deg	50–0.5 hPa, 15 layers (from ~6 to ~15 km)	01/1978 – 12/2016
GOZCARDS v2.20 https://gozcards.jpl.nasa.gov	SAGE I v5.9_rev, SAGE II v7, HALOE v19, Aura MLS v4.2		90S–90N, 10 deg	215–0.2 hPa, 6 or 12 levels per pressure decade (~3 km)	01/1979 – 12/2016
SWOOSH v2.6 https://data.noaa.gov/dataset/dataset/stratospheric-water-and-ozone-satellite-homogenized-swoosh-data-set	SAGE II v7, HALOE v19, UARS MLS v5, SAGE III v4, Aura MLS v4.2		90S–90N, 10 deg (also 5 and 2.5 deg)	316–1 hPa, 6 or 12 levels per pressure decade (~3 km)	01/1984 – 12/2016
SAGE-OSIRIS-OMPS LOTUS ftp	SAGE II v7, OSIRIS v5.10, OMPS-LP USask 2D v1.0.2	Number density (anomaly) on an altitude grid	60S–60N, 10 deg	10–50 km, 1 level per km	10/1984 – 12/2016
SAGE-CCI-OMPS http://www.esa-ozone-cci.org/?q=node/167	SAGE II v7 , OSIRIS v5.10, GOMOS ALGOM2s v1, MIPAS IMK/IAAv7, SCIAMACHY UB v3.5, ACE-FTS v3.5/3.6, OMPS-LP USask2D v1.0.2		90S–90N, 10 deg	10–50 km, 1 level per km	10/1984 – 07/2016
SAGE-MIPAS-OMPS v2 https://www.imk-asf.kit.edu/english/304_2857.php	SAGE II v7, MIPAS IMK/IAA v7, OMPS-LP NASA v2.5, ACE-FTS v3.5/3.6		60S–60N, 10 deg	6–60 km, 1 level per km	10/1984 – 12/2016

Table 2.2: *General information about merged satellite data sets.*

Inter-instrument biases among the later instruments NOAAs 16–19 (since late 2000) are mostly within 3 %, while biases involving NOAA-9, NOAA-11 descending and NOAA-14 are mostly within 5 % (*Frith et al.,* 2017, see **Figure 5**; *Wild et al.,* 2019).

2.2.2.1 SBUV MOD v8.6

The SBUV MOD time series includes data from all SBUV instruments except NOAA-9, which are excluded due to increased differences with other SBUV and external data sources (*Frith et al.,* 2014; *Frith et al.,* 2017; *DeLand et al.,* 2012; *Kramarova et al.,* 2013a). The combined record provides continuous coverage of ozone profile data since late 1978. As the data have already been inter-calibrated and all known instrument problems resolved, we have no physical rational to choose one data set over another. Therefore, when constructing the merged data set no external calibration adjustments are applied, but rather the data are simply averaged during periods when more than one instrument is operational. This approach relies on the average of multiple measurements to mitigate the effects of small offsets and drifts in individual data sets rather than attempting to choose a single record as a reference calibration. To account for higher uncertainty when orbits approach the terminator, only the subset of measurements with the equator crossing time between 8 am and 4 pm are accepted into the MOD combined time series. The exception to this selection criteria is the record from NOAA-11 ascending (1989–1995) that is entirely accepted to avoid a data gap. Small remaining biases and drifts in the merged record are accounted for in the MOD uncertainty estimates (*Frith et al.,* 2017; also see *Section 3.1.4*). *Tummon et al.* (2015) showed that the MOD record agrees with the mean of other merged ozone data sets within 5 %. The MOD monthly zonal mean data are available at: **https://acd-ext.gsfc.nasa.gov/Data_services/merged/index.html.**

Monthly means are computed for each SBUV instrument separately in 5-degree wide zonal bands. Only bin averages in which the average latitude of the profiles in the bin is within 1 degree from the bin centre and the average time of the profiles is within four days from the centre of the month are included in the MOD record. Measurements are removed for a year after the El Chichón volcanic eruption and for 18 months after the eruption of Mt. Pinatubo to avoid periods when volcanic aerosols likely interfered with the algorithm (*Bhartia et al.,* 2013).

2.2.2.2 SBUV COH v8.6

The SBUV MOD approach of averaging data from all available satellites during an overlap period results in the loss of characteristics of the measurement (*e.g.*, time of measurement). Alternatively, the SBUV COH merging approach is to identify a representative satellite for each time period, thus preserving knowledge of orbital characteristics for each measurement period. Additionally, data in the overlap periods are examined to determine a correction for some satellite records. In the later period of the combined record, the overlaps between NOAA-16 to -19 ozone records are long, and each satellite can be compared and adjusted directly to NOAA-18 (*Wild et al.,* 2019). For example, NOAA-16 at 4–2.5 hPa can differ from NOAA-17 and NOAA-18 by up to 3 % at all latitudes; while the NOAA-17 record differs from NOAA-18 in the mid-latitudes especially in the upper atmosphere at 4 hPa and above where diurnal issues become significant. Recent studies (*Wild et al.,* 2019) show that NOAA-19 also differs from NOAA-18 by approximately 1–2 %. The difference is mostly found in the equatorial regions and between 10 hPa and 6.4 hPa pressure levels. Strong drifts in the early satellites (NOAA-9, -11 and -14) and poor quality of NOAA-9 and NOAA-14 data can create unphysical trends when a successive head-to-tail adjustment scheme is used in the early period (*Tummon et al.,* 2015). The current SBUV-COH data set does not adjust the Nimbus-7 or NOAA-11 data, nor does it include the NOAA-9 ascending node. Only the NOAA-9 descending data is adjusted to fit between the ascending and descending nodes of the NOAA-11 record. NOAA-14 data do not appear in the final data set, but it is used to enable a fit of NOAA-9 descending to NOAA-11 descending where no overlap exists (*Wild et al.,* 2019). The COH data is available at **ftp://ftp.cpc.ncep.noaa.gov/SBUV_CDR** as monthly or daily zonal means both as mixing ratio on pressure level, or as layer data.

The lower quality data from NOAA-9, NOAA-11 descending, and NOAA-14 lead to larger uncertainties (10–15 %) in the mid-1990s (at the time of peak halogen loading and the expected "turn-around" in ozone trends) in both merged data sets and complicate efforts to establish a long-term calibration over the full record (from 1980s to 2000s). Error propagation and trend uncertainty estimates for the SBUV merged records are discussed in more detail in *Chapter 3*, *Section 3.1.4*.

2.2.3 Limb profile data records in mixing ratio on pressure grid

2.2.3.1 GOZCARDS v2.20

The Global OZone Chemistry And Related trace gas Data records for the Stratosphere (GOZCARDS) v1.01 data set, used in the previous ozone assessment (*WMO,* 2014; *Froidevaux et al.,* 2015), has been extended to the present. Recently, a GOZCARDS merged data set v.2.20 has been created. GOZCARDS provides VMRs on a pressure grid for 10-degree latitude bins (starting at 0–10 degrees) and is a combination of various high quality space-based monthly zonal mean ozone profile data. The GOZCARDS pressure levels are regularly spaced in log-space, with 12 (6) levels for each decade change in pressure for pressures larger (smaller) than 1 hPa. The recommended data range is 215 hPa to 0.2 hPa; at tropical latitudes the recommended range

is 100 hPa to 0.2 hPa to ensure only stratospheric data are considered. Caution is recommended for the upper stratospheric / lower mesospheric levels, given the existence of incompletely accounted for diurnal and seasonal effects (for both source and merged data, particularly when considering occultation data sets). The GOZCARDS monthly mean ozone record includes SAGE I (version 5.9), SAGE II (v7), the HALogen Occultation Experiment (HALOE; v19) and Aura MLS (v4.2), and covers the period from 1979-2016. SAGE II data are used as a reference for adjusting/debiasing the HALOE and Aura MLS measurements (using overlapping time periods of observation). Details of the screening criteria for each data set, the merging procedure, as well as estimated uncertainties (random and systematic) are provided by *Froidevaux et al.* (2015). This new GOZCARDS version utilises a reduced number of data sources and a finer stratospheric retrieval pressure grid, in comparison to v1.01 (*Froidevaux et al.*, 2015). UARS MLS data are not used, since they are not available on the finer vertical grid of GOZCARDS v2. While interpolation could have been used, an exact treatment of the retrieved uncertainties is not feasible. Data from the Atmospheric Chemistry Experiment - Fourier Transform Spectrometer (ACE-FTS) instrument are not used either, as the updated ACE-FTS v3.6 data version was not available in time for the data creation deadlines. The most significant change is the effect of using the SAGE II v7 data, which uses Modern-Era Retrospective analysis for Research and Applications (MERRA) temperature profiles in the retrievals, and the actual impact of the those temperatures (rather than National Centers for Environmental Prediction (NCEP) temperatures) on the conversion of SAGE II ozone (density on altitude grid) to the VMR on pressure grid used for GOZCARDS ozone. Additionally, Aura MLS v4.2 data are now included (instead of v2.2), along with HALOE v19 profiles which are interpolated to the finer pressure grid before merging. The SWOOSH record also uses SAGE II v7 ozone data and there is now closer agreement and better correlation between the SWOOSH and GOZCARDS v2.20 time series than between SWOOSH and GOZCARDS v1.01 data.

2.2.3.2 SWOOSH v2.6

The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database was created by Chemistry Sciences division of NOAA/ESRL (NOAA Earth System Research Laboratory) in Boulder, Colorado, USA. It includes vertically resolved ozone and water vapor data from a subset of the limb profiling satellite instruments operating since the 1980s. An overview of SWOOSH is provided by *Davis et al.* (2016). The primary SWOOSH products are monthly zonal mean time series of water vapor and ozone mixing ratio on 12 pressure levels per decade from 316 hPa to 1 hPa, the same levels as from the Aura MLS instrument. SWOOSH is provided on several zonal mean grids (2.5°, 5°, and 10°), and additional products include two coarse 3D griddings (30° lon x 10° lat, 20° x 5°) as well as a zonal mean isentropic product. Here, the 10° zonal mean product is used.

SWOOSH includes data from SAGE II v7, UARS HALOE v19, UARS MLS v5, SAGE III v4, and Aura MLS v4.2. Data are compiled from both individual satellite source data as well as a merged data product. For SWOOSH, all records provided in units of number density on altitude grid (*i.e.,* SAGE II and III) are converted to mixing ratio on pressure using MERRA reanalyses, similar to the process used in GOZCARDS v2.20. A key aspect of the merged product is that the source records are homogenised to account for inter-satellite biases and to minimise artificial discontinuities in the record. The SWOOSH homogenisation process involves adjusting the satellite data records to a "reference" satellite using coincident observations during time periods of instrument overlap. The reference satellite is chosen based on the best agreement with independent balloon-based sounding measurements, with the goal of producing a long-term data record that is both homogeneous (*i.e.,* with minimal artificial discontinuities in time) and accurate (*i.e.*, unbiased). For ozone the reference instrument is SAGE II. The SWOOSH v2.6 data are publicly available at **https://data.noaa.gov/dataset/dataset/stratospheric-water-and-ozone-satellite-homogenized-swoosh-data-set**.

2.2.4 Limb profile data records in number density on altitude grid

2.2.4.1 SAGE-OSIRIS-OMPS

The merged SAGE-OSIRIS-OMPS time series has been created at the University of Saskatchewan. The basic construction technique used for this merged time series of deseasonalised anomalies is described in *Bourassa et al.* (2014). For the merged time series the data for each of the three instruments are first treated separately. They are averaged within 1 km altitude and 10° latitude bins and then individually deseasonalised. The resulting zonal mean, deseasonalised anomalies are then merged after biases are removed. The time series spans the period from 1984, when the first SAGE II measurements were made, up to the present where both OSIRIS and the Ozone Mapping and Profiler Suite - Limb Profiler (OMPS-LP) continue to produce high quality data records.

The three observation data sets merged in this data record are: SAGE II v7.0, the recently released OSIRIS v5.10 with improved pointing stability (*Bourassa et al.,* 2018), and the University of Saskatchewan OMPS-LP 2D data set (USask 2D) v1.0.2 (*Zawada et al.,* 2018). This work also further describes the merging process used to create the data record and presents results from preliminary trend analyses based on these data. These preliminary analyses indicate that the addition of the OMPS-LP data to the original SAGE II-OSIRIS merged anomaly data record only slightly changes the magnitude of the derived trends, but the additional data enhances the significance of these results. Since the OSIRIS and OMPS-LP instruments are still operational, this merged data set will be updated regularly as new measurements become available (access from the LOTUS web page).

Two versions of the merged SAGE-OSIRIS-OMPS data set are produced: One uses SAGE II data with corrected sampling effects (*Damadeo et al.,* 2018) further called corr-SAGE and another relies on the standard SAGE II v7 data (*Damadeo et al.,* 2013).

2.2.4.2 SAGE-CCI-OMPS

The merged SAGE-CCI-OMPS data set has been developed in the framework of the European Space Agency (ESA) Climate Change Initiative on Ozone (Ozone_cci). It includes data from several satellite instruments: SAGE II on the Earth Radiation Budget Satellite (ERBS), GOMOS, the SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY) and the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) on EnviSat, OSIRIS on Odin, ACE-FTS on the SCIence SATellite (SCISAT), and OMPS-LP on the Suomi National Polar-orbiting Partnership (Suomi-NPP). The data set is created specifically with the aim of analysing stratospheric ozone trends. For the merged data set, the latest versions of the original ozone data sets are used. Detailed information about the individual data sets is presented in *Sofieva et al.* (2017). Data sets from the individual sensors have been extensively validated and inter-compared (*e.g.*, *Rahpoe et al.,* 2015*; Hubert et al.,* 2016); only those data sets that are in good agreement and that do not exhibit significant drifts, with respect to collocated ground-based observations and with respect to each other, are used for merging. The intercomparison of data records from individual instruments is presented in (*Sofieva et al.,* 2017) and can also be found in *Section 3.1.3* of the current Report. The long-term data set is created by computation and merging of deseasonalised anomalies, which are estimated using monthly zonal mean profiles from individual instruments (*Sofieva et al.,* 2017).

The merged SAGE II v7, Ozone_cci, and OMPS-LP (USask 2D v1.0.2) data set consists of merged monthly deseasonalised anomalies of ozone in 10° latitude zones from 90°S to 90°N. The data are provided on an altitude grid from 10 km to 50 km, during the period from October 1984 to July 2016. The best quality of the SAGE-CCI-OMPS data set is expected in the stratosphere at latitudes between 60°S and 60°N. Ozone trends in the stratosphere have been evaluated based on the created data sets (*e.g.*, *Sofieva et al.,* 2017; *Steinbrecht et al.,* 2017). The data set is available at the LOTUS website and at the Ozone_cci website (**http://www.esa-ozone-cci.org**).

2.2.4.3 SAGE-MIPAS-OMPS v2

The SAGE-MIPAS-OMPS data set consists of deseasonalised ozone anomalies from the SAGE II v7 (1984–2005), MIPAS IMK/IAA v7 (2002–2012) and OMPS-LP v2.5 (April 2012 – March 2017) data sets which are merged using the ACE-FTS v3.6 data record as a transfer standard. Namely, time series of parent instruments are debiased by minimising the root mean square of uncertainty-weighted differences with time series of ACE-FTS (where overlapping), taking the standard error of the mean as the uncertainty. This procedure removes biases between the different data sets, including those resulting from different altitude resolutions or different prior information, sampling issues, and limited or no overlap between different data sets. The merging in overlapping periods is performed via weighted means, with weights inversely proportional to standard errors of the means of corresponding monthly means from individual data sets. Two periods of MIPAS measurements, 2002–2004 and 2005–2012, are treated as two independent data sets.

The data set is provided along with uncertainty estimates. The data are provided in 10° latitude bins, from 60°S to 60°N for the period from October 1984 to March 2017. The main differences to the SAGE-CCI-OMPS data set are:

- the OMPS data are from the NASA processor, instead of the USask 2D processor
- the MIPAS data from 2002–2004 are included in the record
- the ACE-FTS data are used as the transfer standard.

The first release of this merged data record used version 2 of the NASA OMPS-LP profile retrievals and was used in the assessment by *Steinbrecht et al.* (2017). The SAGE-MIPAS-OMPS data record used for the LOTUS assessment incorporates the newer OMPS-LP NASA v2.5 data described by *Kramarova et al.* (2018).

There exists a version of the SAGE-MIPAS-OMPS data set which uses MLS as a transfer standard, but it is not considered in this Report due to time limitations. The SAGE-MIPAS-OMPS is described in detail in *Laeng et al.* (2019) and the data set is available at **https://www.imk-asf.kit.edu/english/304_2857.php.**

2.2.5 Satellite data in broad latitude bands

In *Sections 5.2* and *5.3*, we discuss profile time series and trends in three broad latitude bands: 60°S–35°S, 20°S–20°N, and 35°N–60°N. For GOZCARDS, SWOOSH, SBUV MOD, and SBUV COH, we first computed the deseasonalised monthly anomalies (in percent) with respect to their own 1998–2008 climatology for each 5° or 10° latitude belt (**Table 2.2**), then averaged over the broader latitude zones with equal weights. The SAGE-CCI-OMPS and SAGE-OSIRIS-OMPS data records were provided as deseasonalised. However, instead of using 1998–2008 as the base period, the entire time period of the record was used for normalisation. In these two cases, we averaged the reported deseasonalised monthly anomalies (in percent) over the three belts, then offset the result to zero mean value in 1998–2008.

2.3 CCMI model data

2.3.1 Description of model data sets

We have used output from the chemistry–climate models (CCMs) and chemistry-transport models (CTMs) participating in phase 1 of the CCMI (*Eyring et al.,* 2013). CCMI is a joint activity of the International Global Atmospheric Chemistry (IGAC) and Stratosphere–troposphere Processes And their Role in Climate (SPARC) projects, with CCMI-1 being the first phase of this initiative and a continuation of previous CCM intercomparisons (CCM Validation Activity; CCMVal) such as CCMVal-1 and CCMVal-2. Model output from both CCMVal intercomparisons have been widely used in previous WMO Ozone Assessments (*WMO,* 2007, 2011, 2014).

Models participating in CCMI-1 are coupled chemistry–climate and chemistry-transport models, which are able to capture the coupling between the stratosphere and troposphere in terms of composition and physical climate processes more consistently than previous model generations. An overview of the models used in the first phase of CCMI-1, together with details particular to each model, and an overview of the available CCMI-1 simulations is given in *Morgenstern et al.* (2017).

For this Report we have used data from the REF-C2 simulation of CCMI-1. Although the most appropriate reference simulation set would have been the REF-C1, which reproduces the past, we opted to use the REF-C2 simulation, as the last year of the REF-C1 was as early as 2010 (or even earlier for some models) and would therefore not cover the entire period when observations are available. Our interest is to provide information about the long-term evolution of ozone changes until the present, and this seamless simulation from 1960–2100 was considered appropriate. REF-C2 is analogous to the REF-B2 experiment of CCMVal-2 but with a number of new and/or improved CCMs. The experiments follow the *WMO* (2011) A1 scenario for ozone depleting substances and the RCP 6.0 for other greenhouse gases, tropospheric ozone precursors, and aerosol and aerosol precursor emissions. Ocean conditions can be either modeled (from a separate climate model simulation, in 9 of the models used here), or internally generated, in the case of ocean-coupled models (7 of the models used). Details can be found in **Table S1** of the Supplement of *Morgenstern et al.* (2017). For the solar forcing, the recommendation was to use the forcing data from 1960–2011 (as in the hindcast REF-C1 simulations) and a sequence of the last four solar cycles (solar cycle numbers 20–23) until the end of the simulations. Finally, the QBO was either internally model-generated or nudged from the data set provided by Freie Universität Berlin. No volcanic forcings were used in this reference simulation. For a detailed description of the full forcings used in the reference simulations see *Eyring et al.* (2013), *Hegglin et al.* (2016), and *Morgenstern et al.* (2017).

In this work, we used a total number of 16 models submitted to the REF-C2 archive (see **Table S2.1** in the Supplemental Material that summarises the models and number of analysed runs). This final selection was based on availability of zonal averaged ozone profile data at the models' native latitude resolution between 60°S–60°N and full vertical coverage from troposphere to stratosphere. Our analysis required (a) zonal wind profiles (zonal means), (b) sea surface temperatures (SSTs, over the tropical Pacific) and (c) ozone as total column and profile (zonal means). We used all pressure levels provided by the models (at standard levels, a total of 31) and all model latitudes, using also the multiple simulations provided by many of the participating models.

2.3.2 Model data in broad latitude bands

As an initial step in this analysis, we have transferred the zonal mean ozone profile data for each model (and every ensemble member) to a common five degree latitude grid, keeping the 31 vertical levels as initially provided. All ozone profiles (at the corresponding pressure level/latitude bin) were then deseasonalised to their climatology, using 1998–2008 as the base period.

In order to create the time series analysed in *Chapter 5,* we first averaged all individual ensemble simulations for each model, so that only one time series for each model/modelling group is included in the average. This is done to avoid unequal weighting caused by the larger number of ensemble members provided by some models/groups. The deseasonalised model time series shown in *Section 5.2* and regressed in *Section 5.3* were computed as the equally weighted average over the appropriate latitude bands: 60°S–35°S, 20°S–20°N, 35°N–60°N, and 60°S–60°N. For each latitude band, the mean, standard deviation and median were calculated. The range of the model results is provided as the 10th (lower) and 90th (upper) percentiles. Moreover, the absolute minimum and maximum values at each time/level/latitude bin were calculated. The time series of model annual averages are presented in *Chapter 5* for comparisons with observations. The model data shown in **Figures 5.4** and **5.5** are smoothed with a 1-2-1-year filter to eliminate any possible shorter term natural variability, as it was included in a number of models but not in all.

2.4 Summary

Chapter 2 provides a description of long-term ozone profile data sets made available for the trend analyses discussed in *Chapter 5*. In order to be considered for trend analyses, the data have to be available from 1985 through 2016 and have no significant gaps (less than a year). In multiple regression analyses (see *Chapter 4* and *5*), longer data sets allow for a more robust fit, particularly for slowly varying proxies (*i.e.,* solar), which might otherwise alias into the trend.

Analyses published in this Report take advantage of four additional years in the long-term ozone records as compared to the results published in the 2014 WMO Ozone Assessment, *Tummon et al.* (2015), and *Harris et al.* (2015).

An exception for the length of the record was provided for several ground-based data sets in order to obtain adequate spatial distribution of trends across a wide range of latitudes. Discussion of the representativeness of individual ground-based records for broad-band trend assessment can be found in *Chapter 5*. Comparisons between satellite and ground-based trends averaged within a broad zonal band are discussed in *Chapter 5*.

The combined satellite records (*Section 2.2*) feature the addition of new satellite records (*i.e.,* two versions of the OMPS ozone data set) and recently reevaluated and stabilised historical data sets from well-established instruments (*i.e.,* removal of the drift in OSIRIS and MIPAS records). These new merged data sets are expected to be more accurate and stable due to the use of revised data records from the individual instruments. The combined data sets' stability also relies on improvements in the merging methods. Improved methods for combining satellite records became available in recent years (GOZCARDS v2.20, SAGE-CCI-OMPS, SWOOSH, and others), thus reducing unexplained features (*i.e.*, discontinuities) in the combined records and their impacts on the derived trends. Moreover, assessment of methods used to combine short satellite records led to improved understanding of the sources for propagation of errors in the combined trends and impact on the trend uncertainties (*i.e.*, see discussion about differences in the two SBUV combined records in *Section 3.1.4*). The LOTUS Report evaluates these new data sets for improved accuracy and stability.

Well-maintained long-term ozone records are also important for validation of the CCMI retrospective model runs (*Section 2.3*). Agreement between models and observational records (further discussed in *Chapter 5*) assure complete understanding of the processes that impact past ozone changes, such that we have trust in the scenarios for future ozone changes and attribution to GHG and ODS variability, though such a study of the future is not included in this Report.

Although the new and improved records are an integral part of understanding stratospheric ozone changes, there are some remaining issues that are not fully resolved in this Report. For example, comparisons of the coincident and collocated satellite and ground-based records suggest remaining intermittent drifts in the combined data sets (see *Chapter 3*). Drifts (or discontinuities) can also be found in ground-based records, which has inspired the homogenisation effort for the ozonesonde records (*Smit et al,* 2012a, 2012b). Unfortunately, only a handful of homogenised ozonesonde records were ready in time for the analyses done within the LOTUS activity. The trend analyses of the broad-band ozonesonde records will need to be repeated after all homogenised records are ready to update the broad-band averages. Other ground-based records, especially those available from the same location, need to be reevaluated to understand the causes for discrepancies and how the changes in the observational sampling or processing of the ozone measurements can potentially impact the derived trends.

Assessment of sampling biases for all satellite combined records used in this Report is not available (see discussion in *Chapter 3*). It is important to understand how sampling biases can affect the deseasonalised anomaly records. In addition, assessment of uncertainties in the combined ground-based records is needed for analyses of propagation of measurements errors in the trends analyses.

Additionally, errors in ozone satellite and ground-based combined records can be caused by their conversion to a different coordinate system and impact the resulting trends (*McLinden and Fioletov,* 2011). Error propagation is needed to evaluate the impact of the non-homogenised temperature time series (mostly prior to 2000; *Long et al.,* 2017) on the accuracy and spatial distribution of ozone records converted to new coordinates (*Douglass et. al,* 2017). Depending on the assimilation, this conversion can potentially introduce intermittent drifts and thus degrade the stability of the converted ozone record. Assessment of the impact of the conversion on trends should be addressed in the future.

Availability of satellite overpass data over all ground-based stations is needed to understand the spatial and temporal sampling limitations of the ground-based data sets. Comparisons between the overpass and broad-band derived trends is needed to understand representativeness of the ground-based and sonde ozone trends over the broad regions. Representativeness of all ground-based records for zonal averaged trends was not not fully assessed in this Report, although a limited case study of lidar records is discussed in *Chapter 3* (see *Section 3.2.2*).

Chapter 3: Challenges for trend studies

This chapter aims to provide information that, in *Chapter 5*, assists the interpretation of the long-term trends derived from the ozone profile data records described earlier. Profile trends can differ for a variety of reasons and this represents a real challenge to assess the long-term evolution of ozone (*e.g.*, *WMO,* 2014; *Harris et al.,* 2015; and references therein).

One of the primary drivers of differences in trends comes from the data sets themselves. First of all, single-instrument data records can differ in terms of their stability, their accuracy, and their sampling and smoothing properties in the spatial and temporal domain. The stability of a data record can be affected by aging of or changes in instrumentation, which stems from evolutions in the operation or calibration procedures, *etc*. Secondly, inter-instrument biases lead to discontinuities in merged data records for each transition from one set of instruments to another. The ability to adjust for offsets between instrument records depends on how they are merged and especially on the sampling properties and data quality of the records (*Tummon et al.,* 2015). This merging step is unavoidable for satellite data since no single instrument provides both the spatial and the temporal coverage needed to study multi-decadal trends at the near-global scale. In *Section 3.1*, we report the results of intercomparisons of different single-sensor or multi-sensor data records. These studies aim to identify potential artefacts in the time series, which can help us understand the cause of discrepancy between trends.

Another possible cause of differences in trends from intercomparisons lies in the reduction of the single profile data to monthly zonal mean data. Changes in the sampling pattern introduce a changing bias in the presence of spatio-temporal gradients in the ozone field (*e.g.*, diurnal and seasonal cycles or meridional structure). Differences in sampling properties also affect the comparison of ground-based and satellite trends. The satellite monthly zonal mean time series are not necessarily representative of the monthly mean values observed above the station. Both issues are studied in more detail in *Section 3.2*.

3.1 Consistency of ozone profile data records

3.1.1 Homogeneity of ground-based network data

In this section, we investigate inhomogeneities in the ground-based data records; these may occur in time (*e.g.*, due to changes in instrumentation, instrument performance, or calibration) and in space (*e.g.*, due to differences in instrumentation, instrument performance, or calibration between sites). We describe the key results of exploratory work by *Hubert et al.* (2019) on the homogeneity of ozone profile observations gathered by ground-based networks between 2002 and 2016. They used an ensemble of complementary high quality satellite data records as a transfer standard to investigate the data at 60 ozonesonde, 8 lidar, and 5 microwave radiometer stations operating within the NDACC, GAW, and SHADOZ networks. A description of these ground-based instruments can be found in *Section 2.1.1* of this Report. More detailed findings, discussion and conclusions of this exploration can be found in *Hubert et al.* (2019).

The analysis of *Hubert et al.* (2019) starts off with the following steps. Profile data from a ground-based record X_L (*e.g.*, ozonesonde data at a given location *L*) and a space-based record Y_M (*e.g.*, of satellite mission *M*) are first cleared of spurious measurements, then co-located in time and space (the window is detailed in the next paragraph), then converted to the same profile representation (ozone unit and vertical coordinate), and then finally interpolated to a common vertical grid. Ozonesonde and lidar data are smoothed to the vertical resolution of the satellite data, which differs for each instrument *M* (**Table 1b** in *Hassler et al.,* 2014). The second part of the analysis consists of computing the relative difference $\Delta_{LM}(z,t_i) = 100*(X_{L,i} - Y_{M,i})/Y_{M,i}$ for each profile pair *i*. For the sake of brevity, the indices *L* and *M* are excluded from the formulae that follow. These $\Delta(z,t_i)$ form a time series of relative differences (expressed as a percentage) which are smoothed using an N-month running median filter centered around the middle of each month *j*, which we denote $\Delta^*(z,t_j)$. What are used in the final analysis are time series of relative difference anomalies δ (expressed as a percentage),

$$\delta(z,t_j) = \Delta^*(z,t_j) - \underline{\Delta}(z) \qquad (3.1).$$

Here, $\underline{\Delta}(z)$ represents the median value of $\{\Delta(z,t_i)\}$ for all t_i in the reference period, which is 2005–2011 for all instruments except OMPS-LP where 2012–2016 was chosen (see grey horizontal line in **Figures 3.1 – 3.4**). By removing the satellite- and grid level-dependent median value, the values of δ at different grid levels (*z* and *z'*) and from different satellite records (*M* and *M'*) are on a comparable scale. The δ time series represents anomalies of the relative difference of ground-based minus satellite observations with respect to their median value over the reference period. Positive anomalies indicate that the ground-based bias relative to a satellite is more positive (or less negative) than usual during the reference period and vice versa for negative anomalies.

By construction, these positive and negative anomalies average to zero over the reference period, but this is not necessarily the case outside of this period. The median was preferred over the mean for $\underline{\Delta}(z)$ to avoid the impact of single outliers on the absolute scale of the anomaly time series. The uncertainty (2σ) of the anomalies (δ) is estimated as half the 95 % interpercentile of the Δ values in the running window divided by the square root of the number of pairs in that window. The absolute scale holds valuable information but is mostly disregarded here, as it can be offset for different M due to differences in the reference period or differences in the sampling of the co-located data set.

Different filter window lengths N were investigated. We only show results for a 12-month wide window, which balances the need to reduce the noise in the comparisons in order to identify small anomalies and the desire to preserve information to localise identified anomalies sufficiently well in time. The influence of natural variability is considered negligible because of the large smoothing window but especially because the pairs are well co-located in space (< 300km for comparison to MWR; all others < 500km) and time (< 1h for MWR stations with hourly data; < 6h for MIPAS and Aura MLS; all other comparisons < 12h).

Inhomogeneities in the ground-based data reveal themselves in the temporal and vertical structure of the relative difference anomalies. Increased confidence that anomalies are caused by (originate in) the ground-based record L is found when the δ_{LM} values are significant for multiple independent satellite references {M} over the same altitude region and during the same period in time. The six limb and occultation satellite records selected for this study represent complementary measurement techniques (different spectral ranges, viewing geometries, sampling properties, calibration and retrieval methods) and can therefore be considered independent. The consideration of six complementary and independent satellite records is a vital asset in attributing common features in δ_{LM} to the ground-based data. We considered OSIRIS, GOMOS, MIPAS, SCIAMACHY, Aura MLS, and OMPS-LP as references (**Table 3.1**), which are all fairly dense samplers that have been thoroughly validated and intercompared in recent years (*Tegtmeier et al.,* 2013; *Hassler et al.,* 2014; *Rahpoe et al.,* 2015; *Hubert et al.,* 2016).

Figure 3.1 shows anomaly time series δ for the MLO lidar station in Hawaii, USA and the six satellite records. Stippled cells are not statistically different from zero at the 2-sigma level. The temporal and vertical structure is fairly featureless with only a few areas showing significant anomalies, which shows the good and stable agreement between this lidar record and the different satellite references. While significant positive anomalies are apparent from 2002 to 2004 for OSIRIS, GOMOS, and SCIAMACHY, the statistically significant region is not fully coherent between the three references, which leaves ambiguity as to whether the differences are caused by the ground-based or the satellite record. Significant negative anomalies of ~4 % are apparent at 40–45 km starting in early 2013 for both OSIRIS and Aura MLS. Unfortunately, OMPS-LP cannot confirm this finding as it only started taking measurements in 2012. However, this feature resides at the upper end of the lidar profile where the uncertainty of the measurement becomes larger and hence larger anomalies are expected (*e.g.*, see **Figure 16** of *Leblanc et al.,* 2016b).

Figures 3.2 to **3.4** show selected results for ozonesonde, lidar, and microwave radiometer station records. It is not the purpose of this Report to discuss each station record individually as these can be found in *Hubert et al.* (2019). Instead, we focus on the general performance of the ground-based networks. The majority of the ground station data records exhibit one or more temporal features in their anomaly field with a magnitude of 5 % and often more. These features manifest themselves as (a series of) sudden discontinuities or as transient events over a variety of timescales, but they do not necessarily occur over the entire vertical range.

Some observed discontinuities coincide with, and are caused by, known changes in instrumentation. For instance, the switch from the KC-96 to the ECC ozonesonde at Naha station in November 2008 (*Morris et al.,* 2013) is clearly visible as a +10 % discontinuity in **Figure 3.2** (panel B). A correction scheme has been developed (*Section 2.1.1.1*) which should

Instrument	Platform	Analysis period	Level-2 data version	Viewing geometry	Spectral range	Analysis
SAGE II	ERBS	1984-2005	v7	solar occultation	UV-VIS-NIR	S
HALOE	UARS	1991-2005	v19	solar occultation	NIR-SWIR	S
OSIRIS	Odin	2001-2016	v5.10	limb scattered	UV-VIS-NIR	G, S
GOMOS	EnviSat	2002-2011	ALGOM2s v1	stellar occultation	UV-VIS-NIR	G, S
MIPAS		2005-2012	IMK/AA v7	limb emission	MIR-TIR	G, S
SCIAMACHY		2003-2012	IUP v3.5	limb scattered	UV-VIS-NIR	G, S
ACE-FTS	SciSat-1	2004-2016	v3.5/v3.6	solar occultation	MIR-TIR	S
MLS	EOS/Aura	2005-2016	v4.2	limb emission	MW	G, S
OMPS-LP	Suomi-NPP	2012-2016	USask-2D v1.0.2	limb scattered	UV-VIS	G, S

Table 3.1: *Characteristics of nine limb/occultation satellite data records. The last column indicates whether the data were used for the study of the homogeneity of ground-based data records (G; see Section 3.1.1) and/or for the estimation of satellite drift (S; see Section 3.1.2).*

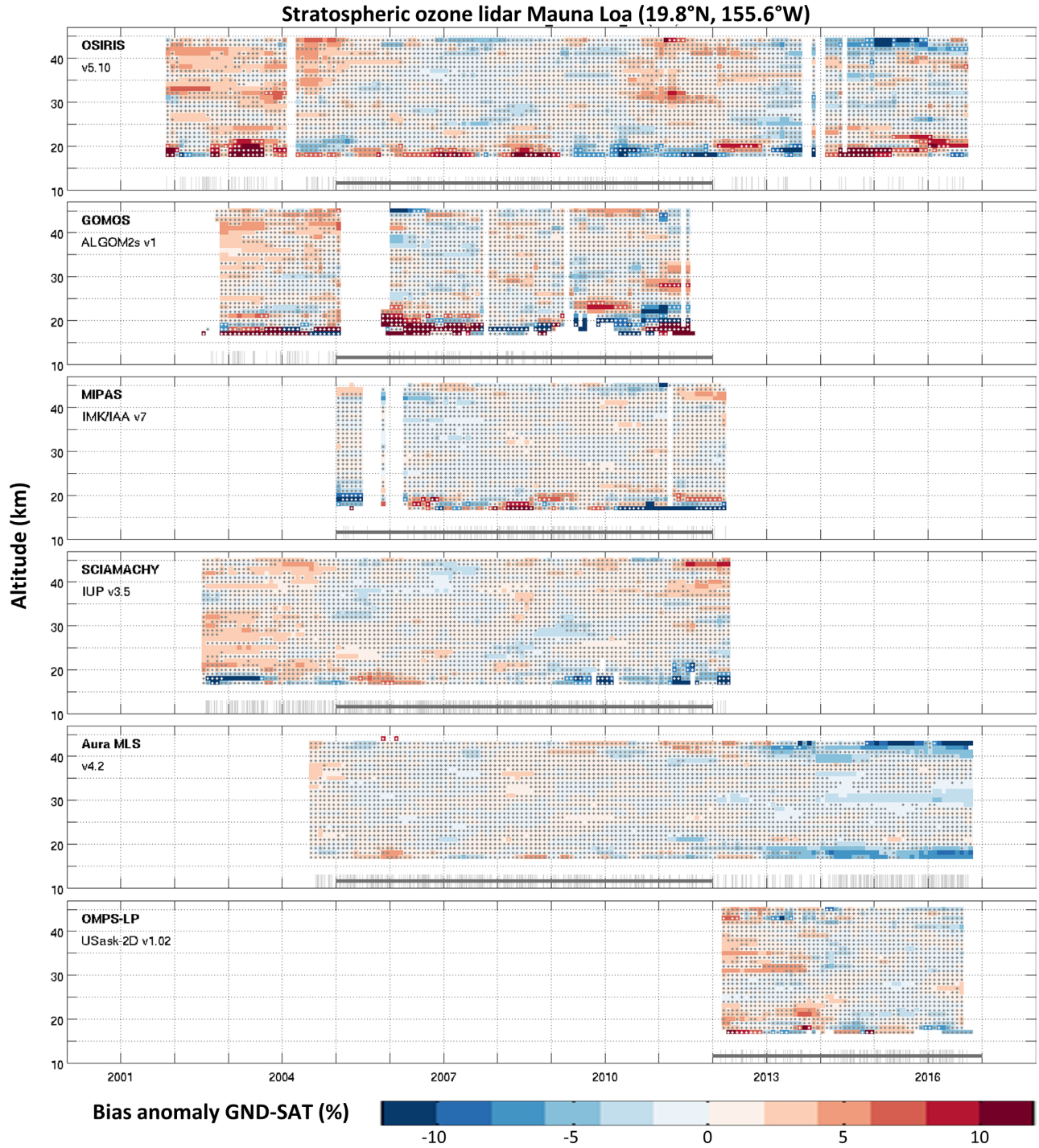


Figure 3.1: *Smoothed anomaly time series (δ, see Eq. 3.1) of the relative difference of MLO lidar and six satellite ozone profile data records (top to bottom). Red values indicate regions in which lidar measurements are biased more positive (or less negative) compared to satellite than their median value during the reference period. Stippled areas denote δ values that are not statistically different from zero at the 2-sigma level. A running average with a 12-month window was used to smooth the time series. Thin grey vertical lines show the sampling of the co-located profile data records; the grey horizontal lines indicate the reference period for each comparison. Adapted from Hubert et al. (2019).*

result in a more homogeneous time series. Another example is the positive (negative) bias excursion above (below) the ozone maximum starting in 2010 and ending in 2012 in the Hohenpeissenberg lidar record (**Figure 3.3**; panel A). These are likely related to an aging device that fired the laser until it was replaced in early 2013 (*W. Steinbrecht,* private comm.). However, in many cases further investigations are needed to understand these anomalies, to find the cause of the changes, and to ultimately develop a correction strategy.

In this respect, we comment that simply adjusting to another observational data record will lead to a loss of independence between records. Clearly, this should be avoided as much as possible.

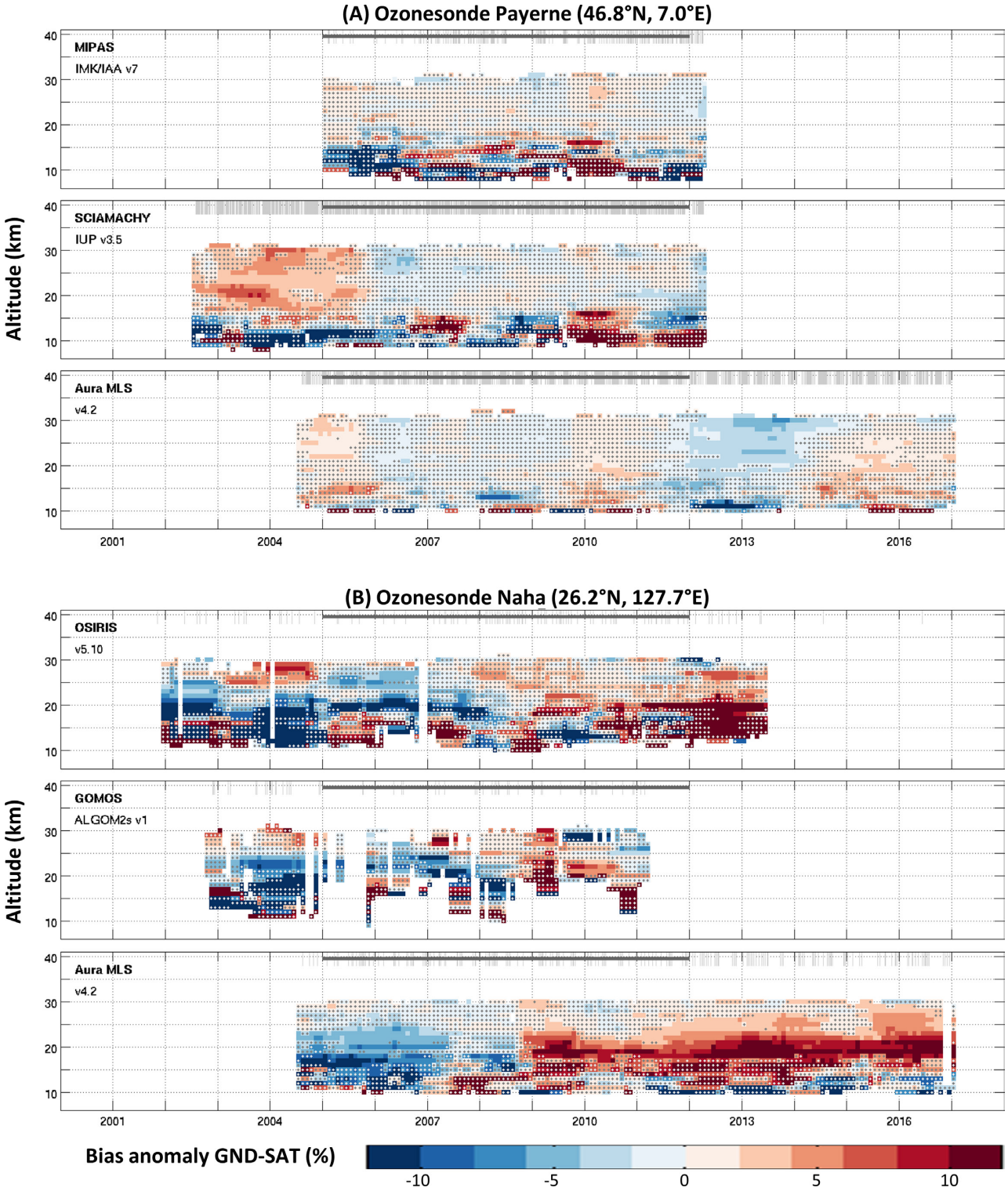


Figure 3.2: *As* ***Figure 3.1*** *but for two ozonesonde sites each with a different selection of three satellite references. Stippled areas denote non-significant δ values. Comparisons to all six satellite records for both stations are shown in* ***Figures S3.1*** *and* ***S3.2*** *in the Supplement. Adapted from Hubert et al. (2019).*

Our results indicate that most of the 73 considered station records (60 ozonesonde, 8 lidar, and 5 MWR) have one or more periods with inhomogeneities over part of the vertical range of the data record. Such artefacts are noted even in ground-based data records that are generally considered as "golden" time series for trend studies (because of their length and/or supposedly better stability). Examples are shown in **Figures 3.2** (A), **3.3** (A&B), and **3.4** (A). The magnitude of the anomalies is broadly consistent with the quoted 5–10 % systematic uncertainty of the ground-based measurement techniques (*Section 2.1*). Nonetheless, it illustrates the importance of clarifying to data users that the quoted systematic uncertainty is pertinent to every single ozone profile and that the sign and magnitude may change

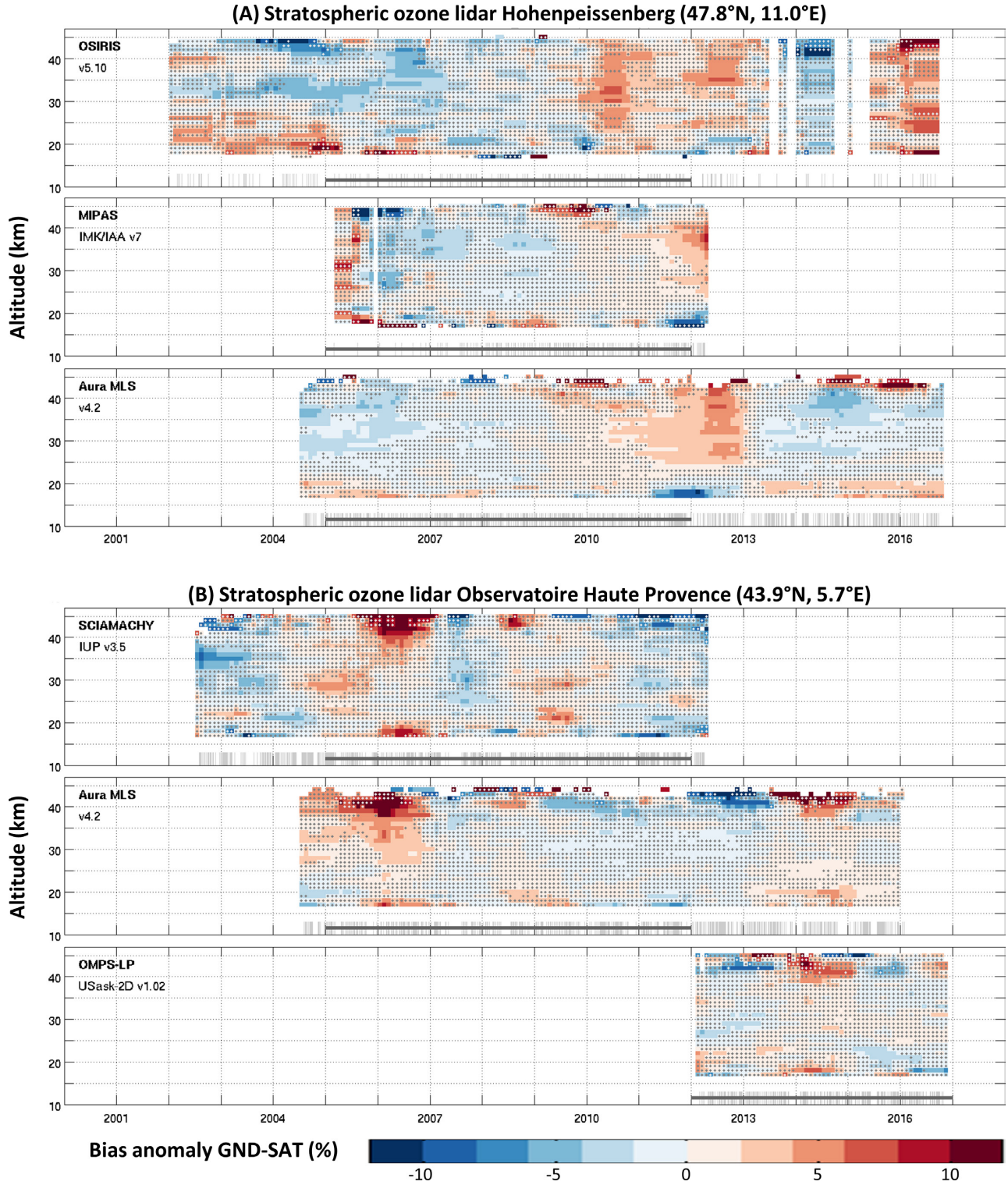


Figure 3.3: *As **Figure 3.2** but for two stratospheric ozone lidar sites. Stippled areas denote non-significant δ values. Comparisons to all six satellite records for both stations are shown in **Figures S3.3** and **S3.4** in the Supplement. Adapted from Hubert et al. (2019).*

over the data record, thereby effectively representing a non-systematic uncertainty component in the time domain.

Measurement artefacts are generally not modelled in regression analyses, thereby introducing random and systematic uncertainty in profile trends. These artefacts are furthermore dependent on the station and the vertical level. Single large discontinuities or multiple discontinuities with the same sign constitute a low frequency signal which will clearly bias the derived trend. The likelihood of such trend biases decreases with an increasing number of excursions as long as their sign and magnitude is sufficiently random in time. The random uncertainty of the trend estimate, on the other hand, will unavoidably accrue contributions from the unmodelled variance.

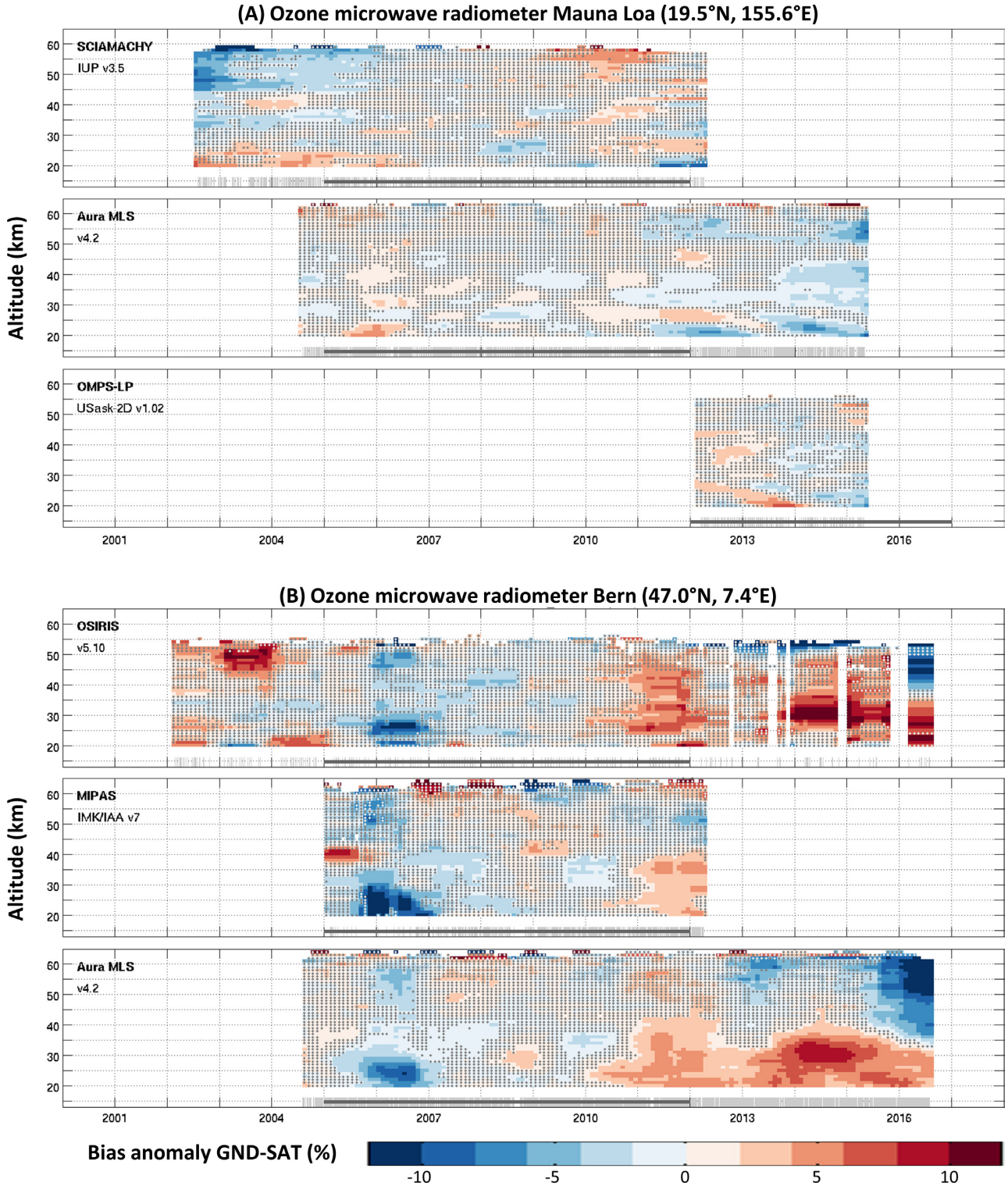


Figure 3.4: *As* ***Figure 3.2*** *but for two microwave radiometer sites. Stippled areas denote non-significant δ values. Comparisons to all six satellite records for both stations are shown in* ***Figures S3.5*** *and* ***S3.6*** *in the Supplement. Adapted from Hubert et al. (2019).*

A detailed time series analysis is needed to quantify the possible impact of measurement inhomogeneities on ozone profile trends. However, since both occurrence and magnitude of the artefacts depend on the site and the ground-based instrument, it is expected that the profile trends will (also) differ as a result of the additional bias and variance. Such differences between neighbouring sites and between ground-based instruments at one site have been reported repeatedly in past and recent analyses (*e.g.*, *Steinbrecht et al.*, 2006; *Logan et al.*, 2012; *Nair et al.*, 2013, 2015; *Tarasick et al.*, 2016; *Van Malderen et al.*, 2016).

A comparison of trends from different instruments at Lauder and Hawaii using the same regression model will be shown in *Section 5.4*.

A pragmatic approach to reduce the impact of these measurement inhomogeneities would be to average the station records over several sites, perhaps the entire ground-based network (*Logan et al.,* 1994, 1999a, 1999b; *Terao and Logan,* 2007). This will effectively decrease the relative importance of systematic effects as many measurement artefacts across the network are of varying magnitude and occur randomly in time and space. However, some artefacts can be attributed to changes that occurred fairly simultaneously across parts of the network, in particular for the ozonesonde data records. For instance, the ten sites in the Canadian subnetwork transitioned from BM to ECC ozonesonde models in the early 1980s and had a further series of simultaneous changes in the following decades (*Tarasick et al.,* 2016), the five Japanese sites switched from the KC to the ECC ozonesonde around 2009 (*Morris et al.,* 2013), and there are additional changes in the NOAA and SHADOZ subnetworks (*Witte et al.,* 2017; *Sterling et al.,* 2018). Such simultaneous changes in the sonde network will not be fully averaged out, so it is vital to have as many independent station records as possible. Measurement operations at lidar and MWR stations, on the other hand, are fairly independent, but the number of sites is much smaller than for the sonde network. This re-emphasises the need to sustain the current number of stations in the ground-based networks.

3.1.2 Stability of limb data records relative to ground-based networks

Bottom-up calculations of the stability of ozone profile data records, that is from first principles or from the propagation of low level monitoring data through the retrieval chain, are contentious as they rarely lead to a realistic perception of the long-term performance of the measurement systems. Top-down approaches compare profile measurements to a reference data record and ultimately derive estimates of the stability (also called "drift") relative to that reference (*e.g.*, *Nair et al.*, 2012; *Rahpoe et al.*, 2015; *Hubert et al.*, 2016). These estimates approximate absolute stability if the reference is sufficiently stable. However, past validation and intercomparison exercises have shown the challenges in establishing one (or more) ozone profile data records as a stable data record at the level required by profile trend assessments, which lies around 1 % per decade (*GCOS*, 2011).

Results from intercomparisons between different satellite records are described in *Sections 3.1.3–3.1.5*. In this section we use the ground-based networks of ozonesonde, lidar, and microwave radiometer measurements as a reference to assess the decadal stability of single-sensor single-profile data (Level-2) and of gridded monthly zonal mean data (Level-3) from single sensors and for one multi-sensor data record. Only limb/occultation sounders are considered here; ground-based comparisons for the SBUV nadir profilers have been reported by *Kramarova et al.* (2013a). The method follows that of *Hubert et al.* (2016) where regressions are made to the different time series of satellite and ground-based data, and the linear slope estimates, interpreted as satellite drift, are subsequently averaged over the network to obtain pseudo-global estimates. This reduces the impact of noise and inhomogeneities in the ground-based records on the satellite drift estimate (see *Section 3.1.1*). However, the uncertainties resulting from the linear fit do not fully take into account inhomogeneities across the network, so these are inflated using a χ^2-scheme (see *Section 4.1.2* in *Hubert et al.*, 2016). Finally, the uncertainty of the network-averaged satellite drift is obtained by propagating the χ^2-corrected uncertainties of the linear term through the weighted average.

The first analysis investigates single satellite ozone profiles (*i.e.*, Level-2) co-located in space (< 300 or 500 km) and time (< 1, 6, or 12 hours) to ground-based observations by ozonesonde, stratospheric lidar, and MWR networks. The co-location requirement reduces the number of compared measurements considerably in favour of a smaller mismatch uncertainty (*Verhoelst et al.*, 2015) in the comparison time series. **Figure 3.5** shows the vertical dependence of the drift relative to ozonesonde (bottom left), lidar (top left), and MWR (bottom right) for eleven limb/occultation sounder data records. Nine of these are part of a merged data record in this Report: SAGE II, HALOE, OSIRIS, GOMOS, MIPAS, SCIAMACHY, ACE-FTS, Aura MLS, and OMPS-LP (detailed version information can be found in **Table 3.1**). **Figure 3.6** shows the corresponding significance, with 2σ chosen as threshold for detection (*i.e.*, 95 % confidence level).

Longer time series are available here compared to *Hubert et al.* (2016), but revised or different satellite retrieval algorithms were also considered for most instruments (except for SAGE II and HALOE, for which no new data were available), and one entirely new instrument record was added to the analysis (OMPS-LP). The main conclusions of *Hubert et al.* (2016) still hold and a fairly consistent picture emerges from the ozonesonde, lidar, and microwave radiometer results. These show that, generally, the limb/occultation data records are stable within 5 % per decade in the middle and upper stratosphere. For some records, for example SAGE II and Aura MLS, the constraints on stability are even better, with an upper bound on drift of less than 2 % per decade. No significant drift was found for MIPAS and ACE-FTS. SCIAMACHY data prior to August 2003 were removed from the analysis (see *Section 3.1.3* and *Sofieva et al.*, 2017). The negative drift around 35 km is now no longer statistically significant and decreased from 5 % to 3 % per decade. However, statistically significant deviations from zero were found for a few instruments in different regions of the atmosphere. In chronological order, HALOE data around 25–30 km drift to lower ozone mixing ratios by 3–4 % per decade. Improvements in the pointing stability for OSIRIS have clearly reduced the positive drift from 8 % to 4 % per decade, but the latter result remains statistically significant. And GOMOS occultation data drift to lower ozone values by 5 % per decade and more below around 25 km.

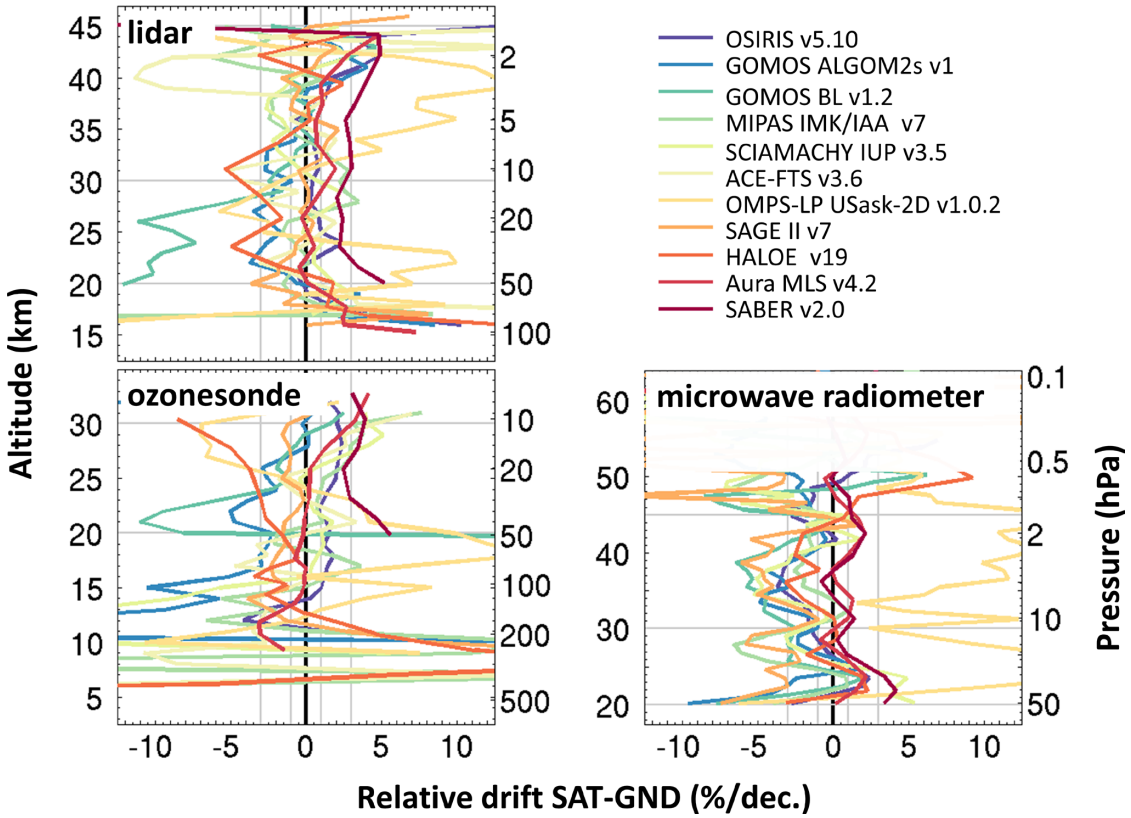


Figure 3.5: *Vertical profile of network-averaged satellite drift (Level-2) relative to co-located ground-based measurements by ozonesonde (bottom left), lidar (top left) and microwave radiometer (bottom right). Colours represent different limb/occultation data records (see legend).*

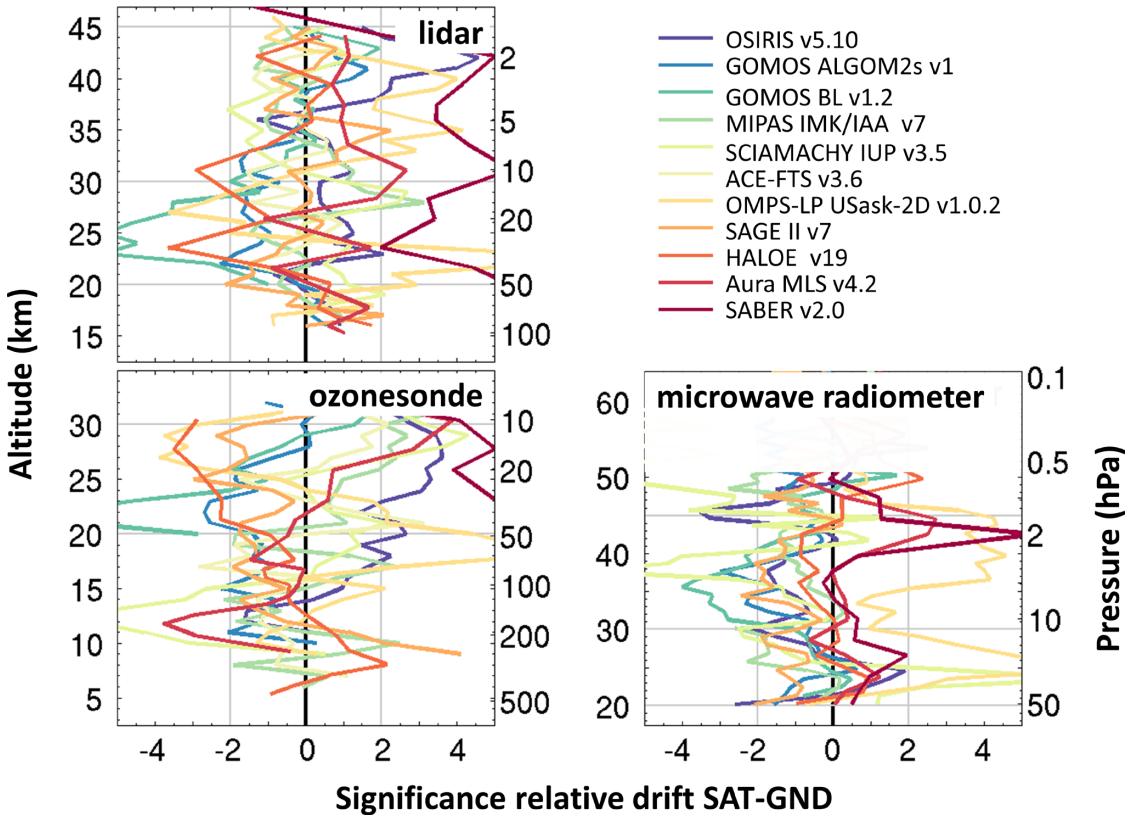


Figure 3.6: *As **Figure 3.5** but for the significance of the drift estimates. The 2σ detection threshold is indicated by grey vertical lines.*

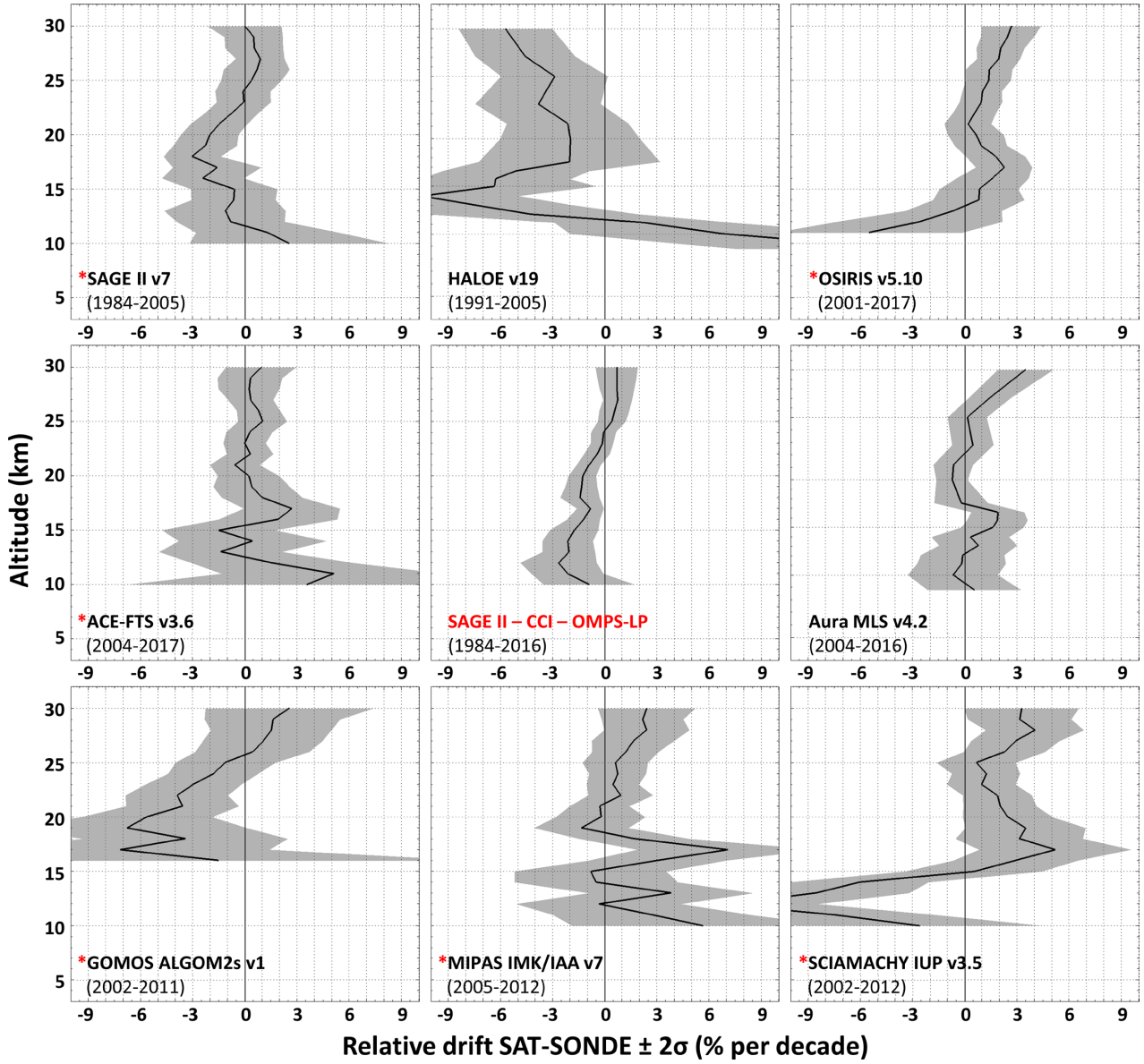


Figure 3.7: *Drift estimates and 95 % confidence interval of monthly zonal mean satellite data relative to the ground-based ozonesonde network. Eight limb/occultation records and the merged SAGE-CCI-OMPS (central panel) are shown. Satellite records contributing to the merged record are indicated with a red asterisk.*

The most striking result comes from OMPS-LP whose vertical drift profile oscillates between negative values (-6 % per decade) at 27 km and positive values around 18 km (+7 % per decade) and 40 km (+9 % per decade). These oscillations are clear from comparisons to each of the three ground-based data records. Even though the OMPS time series is only five years long, the ample co-location statistics allow for drift estimates with comparable precision to that of many other limb/occultation sounders. Instabilities in the altitude registration may be at the origin of the drift in the ozone record (*Moy et al.*, 2017; *Zawada et al.*, 2018; *Kramarova et al.*, 2018).

We stress that statistically insignificant results should not be blindly interpreted as instances where no drift in the data is guaranteed. This is because the adopted method, the available comparison statistics, and the quality of the ground-based reference data only allow us to probe satellite drift to levels that are comparable to (or larger than) the geophysical ozone profile trend expected since the mid-1990s. The lower bound to detect drift is at best 1 % per decade for just a few single-sensor records. Typically, detection thresholds are closer to 2–3 % per decade in the middle stratosphere and 3–4 % per decade elsewhere. These results do not necessarily apply to multi-sensor records since the very combination of different data sets will affect the resulting long-term stability.

The second analysis considers space- and time-gridded limb/occultation data (*i.e.*, Level-3) and gridded ozonesonde data. This approach takes advantage of the complete time series of satellite and ground-based records but at the cost of leaving mismatch or sampling uncertainty in the time series. However, the latter source of error becomes less important with increasing numbers of single profiles averaged by month in 5° latitude zones.

The time series analysis and derivation of drift follows the method described in *Hubert et al.* (2016), but the input data are different. Essentially, we consider the anomalies of monthly zonal mean data relative to its seasonal cycle in a reference period. Ozonesonde data are gridded following the procedure outlined in *Section 2.1.2.1* with slightly different selection criteria and reference period (2004–2011). Comparison time series are constructed as the absolute difference of the gridded satellite anomaly minus the gridded ozonesonde anomaly, with both terms expressed in percent. As before, drift is regressed for each latitude band that contains a ground-based record and then averaged over the entire latitudinal range of the sonde network. The central network-averaged drift estimates and 95 % confidence intervals are shown in **Figure 3.7**. These Level-3 results generally confirm what was observed in Level-2 results, but it appears that the precision of the Level-3 drift estimates is slightly better (*e.g.*, for ACE-FTS). The availability of the entire time series, instead of a subset of co-located measurements, contributes to this improvement, but the smaller impact of reducing station-to-station inhomogeneities helps as well. Indeed, the latter are mostly avoided by constructing the reference as an average of deseasonalised anomaly station data, where the seasonal cycle is derived individually from each sonde record. If ozone profiles at a certain site are, on average, 5 % higher, then this multiplicative bias will be present in both the monthly and seasonal cycle data, and, therefore not in the deseasonalised monthly relative anomaly data. This step brings the average level of the anomaly time series of all stations to zero over the reference period, which avoids artificial steps in the station-combined sonde time series where a measurement gap starts or ends for a particular station. A second benefit of this deseasonalisation procedure is that it reduces the variance in the comparison time series caused by differences in the seasonal cycle of satellite and sonde data.

The central panel of **Figure 3.7** also shows drift results for the merged SAGE-CCI-OMPS data record described in *Section 2.2.4.2* and by *Sofieva et al.* (2017). Non-significant, positive values of 0.8 % per decade are found between 25–30 km. The negative values of ~1.5 % per decade below 24 km are statistically significant, but we advise great care in interpreting significance in the lower part of the stratosphere. The variability of the ozone field, the lower ozone concentrations and the fading sensitivity of limb sounders make it very difficult to obtain precise uncertainty estimates in this part of the atmosphere. Comprehensive studies are therefore needed to further quantify these errors. For now, the result of lower stratospheric drift is inconclusive. Future work will consider other merged profile data records and extend the analysis to the lidar network data.

3.1.3 Intercomparisons of limb satellite measurements

In the context of data validation studies, intercomparisons of satellite measurements are typically performed using profile data that are co-located in space and in time. Many such analyses have been performed in recent years for limb sounders (*e.g.*, *Adams et al.*, 2013, 2014; *Laeng et al.*, 2014; *Kyrölä et al.*, 2013) including analyses of relative drifts (*Rahpoe et al.*, 2015). For the creation of merged data sets, on the other hand, the intercomparison of monthly zonal mean data is more relevant since the combination of different satellite records is usually done at the level of monthly zonal mean data or at the level of monthly deseasonalised anomalies. Such studies have been performed recently by the groups that created merged data sets from limb instruments (*e.g.*, *Bourassa et al.*, 2014; *Froidevaux et al.*, 2015; *Davis et al.*, 2016; *Sofieva et al.*, 2017). Results of intercomparison studies for the SBUV nadir profile sounders are reported in *Section 3.1.4*.

During the preparation of the merged SAGE-CCI-OMPS data set of ozone profiles, the deseasonalised anomalies of the individual instruments (SAGE II, GOMOS, MIPAS, SCIAMACHY, OSIRIS, ACE-FTS, and OMPS-LP USask 2D) have been extensively intercompared by computing and visualising the time series of the difference between the single-sensor anomalies and the median anomaly of the seven data records. This method is sensitive to detecting anomalous features (*i.e.*, large or increasing deviations from the median) in the time series of single sensors.

In particular, it was found that the deseasonalised anomalies for SCIAMACHY are larger at the beginning of the mission, for nearly all latitude bands and at many altitude levels (**Figure S3.7** in the supplement). Similarly, OMPS anomalies are lower in the first three months of the mission (**Figure S3.8** in the supplement). Note that the sampling of the OMPS-LP was significantly coarser in the first three months of the mission. The data from these early periods of SCIAMACHY and OMPS operation are therefore not included in the merged SAGE-CCI-OMPS data set. After the data selection, the anomalies from individual instruments are found to be in good agreement with each other. This is illustrated in **Figure 3.8,** which shows the deviations of deseasonalised anomalies of each instrument relative to the median anomaly of all limb records for 30°S–40°S. Deviations from the median anomalies are small, less than 5 % for the majority of data, and do not have statistically significant drifts with respect to the median anomaly (see also illustrations in the Supplement of *Sofieva et al.*, 2017).

3.1.4 Stability of limb data records relative to ground-based networks

The two nadir instrument-based merged ozone data sets used in this Report (SBUV MOD and SBUV COH; see *Chapter 2*, *Sections 2.2.1.1* and *2.2.1.2*) are constructed from the same initial set of SBUV satellite records. The SBUV series of instruments have similar design and data are retrieved using the same Version 8.6 algorithm (*McPeters et al.*, 2013). Furthermore, as part of the Version 8.6

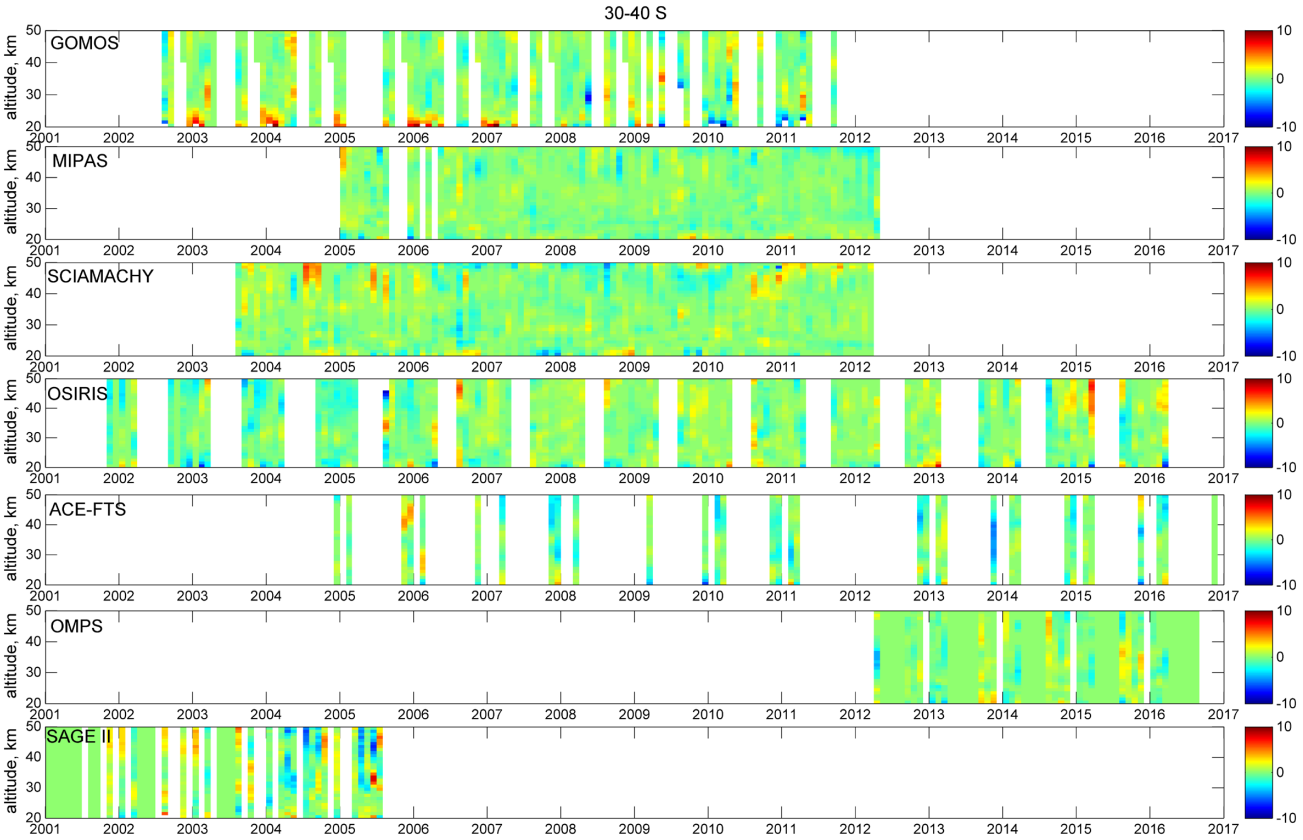


Figure 3.8: *Deviations (in %, colour) of deseasonalised anomalies for GOMOS, MIPAS, SCIAMACHY, OSIRIS, ACE-FTS, OMPS, and SAGE II (indicated in the panels) from the median deseasonalised anomalies computed using all data sets. Latitude band is 30°S–40°S. From Sofieva et al. (2017).*

processing, SBUV measurements from individual instruments were inter-calibrated at the radiance level based on comparisons during instrument overlap periods (*DeLand et al.*, 2012). However, despite the instrument similarity and common retrieval algorithm, each instrument experienced unique operational conditions (*e.g.*, instrument degradation or specific on-orbit problems) and orbital characteristics (including measurement time of day), which contribute to differences among the individual records. Therefore, differences in how the data are selected and merged in the combined records can lead to differences between the merged SBUV products.

In general, the SBUV ozone profile measurements agree to within ±5 % when compared to external satellite and ground-based instruments, with similar or better agreement among the SBUV instruments themselves (*Kramarova et al.*, 2013a; *Frith et al.*, 2017; *Wild et al.*, 2019). However, lower quality data from NOAA-9, NOAA-11 descending, and NOAA-14 lead to larger uncertainties (10–15 %) in the mid-1990s and complicate efforts to establish a long-term calibration over the full record (from 1980s to 2000s) (*DeLand et al.*, 2012; *Kramarova et al.*, 2013a; *Tummon et al.*, 2015; *Ball et al.*, 2017). The SBUV MOD merging approach is to average all available data after removing portions of individual records found to be inferior (*e.g.*, data from drifting orbits, NOAA-9 SBUV/2). This approach relies on the average of multiple measurements to mitigate the effects of small offsets and drifts in individual data sets rather than attempting to choose a single record as a reference calibration. The SBUV COH merging approach is to identify a representative satellite for each time period, thus preserving knowledge of orbital characteristics for each measurement period. Additionally, data after 2001 are adjusted directly to NOAA-18 in SBUV COH, removing small inter-satellite differences. Each approach has advantages and disadvantages. SBUV MOD is sensitive to successively increasing or decreasing biases in the instrument series that might alias into a trend. SBUV COH is sensitive to drifts in the reference instruments that might be propagated to other periods in the record. This was the case in the previous version of SBUV COH used in the SI2N report (*Tummon et al.*, 2015). The potential for unphysical drifts is greatly reduced in the current version of the SBUV COH data set, which limits inter-instrument corrections to periods where long overlaps of high quality data exist.

Frith et al. (2014; 2017) analysed the differences in monthly zonal mean time series from the individual SBUV data sets during periods of overlap in an effort to characterise the uncertainty associated with the merging process. Given the numerous instruments and overlaps, many reasonable approaches could be chosen based on different selections of data (*e.g.*, instrument and time period) and different means of determining inter-satellite adjustments (*e.g.*, mean offset, offset and drift, no adjustments).

The authors used the distribution of measured offsets and drifts between SBUV instruments during times of overlap to construct 10000 MC simulations of potential instrument error (see *Frith et al.*, 2017, **Figure 7**). In essence the collection of SBUV inter-instrument offsets and drifts were used to define an SBUV-system uncertainty, in an effort to account for both relative and absolute uncertainties. The MC simulations were structured to account for the larger observed uncertainty of instruments operating in the 1990s and time dependence of the absolute calibration procedures used within the SBUV retrieval algorithm (*DeLand et al.*, 2012). Previous studies using MC simulations suggest that the long-term drift uncertainty in a record constructed from multiple data sets is less than that for a single instrument because the introduction of new data "resets" the drift (*Stolarski and Frith*, 2006; *Weber et al.*, 2016), but adding too many data sets increases uncertainty as a result of multiple potential discontinuities in the record (*Weber et al.*, 2016). By applying the multiple regression model to the MC simulations we can test the degree to which potential time-dependent uncertainties alias into individual regression terms and assess the additional uncertainty due to the merging process itself. The 2-sigma variation of terms from a regression model fit to the MC simulations defines the "merging uncertainty" for each term.

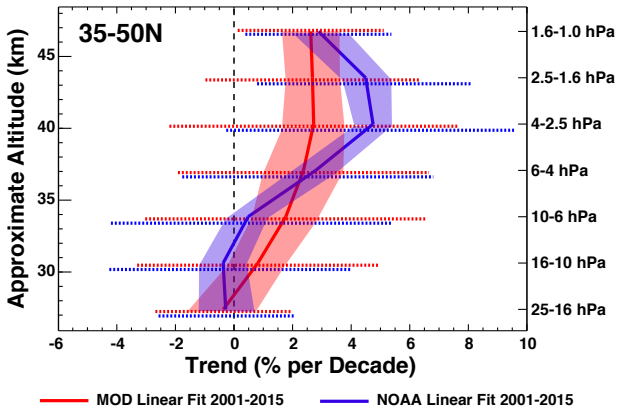


Figure 3.9: *ILT trend proxy fit to 35S–50S monthly zonal mean SBUV MOD (red) and SBUV COH (blue, referred to as "NOAA" in Figure) records over the 2001–2015 time period. The shaded regions indicate the 2-sigma statistical uncertainty estimated from the unexplained variability in the multiple regression analysis. The dotted error bars show the total trend uncertainty when the SBUV MOD 2-sigma merging uncertainty is included. The uncertainties are combined using the root sum of squares of each error term. For comparison, the estimated MOD uncertainty is also added to the SBUV COH error bars. From Frith et al. (2017).*

Frith et al. (2017) also compared regression analysis results between the SBUV MOD and SBUV COH data sets, treating each as equally valid approaches to merging the data record from the SBUV instrument suite. The authors report differences in the post-2000 trend with SBUV COH trends being generally more positive than SBUV MOD at altitudes above the 5 hPa level and less positive below 5 hPa, consistent with the results of this Report (*e.g.*, **Figures 5.1** and **5.2**). When only statistical error is included the results are statistically significantly different from each other, but when the merging uncertainty is taken into account the trend error bars overlap (see **Figure 3.9**). Direct comparisons between both data sets show the differences below 5 hPa are largely a result of a positive bias in NOAA-19 at the end of the record, which is adjusted in SBUV COH but not in SBUV MOD (**Figures 2** and **10** of *Frith et al.*, 2017). Above 5 hPa, a small positive drift in NOAA-18, used as the reference in SBUV COH, leads to a more positive trend relative to SBUV MOD (**Figures 3** and **10** of *Frith et al.*, 2017). **Figure 3.10** shows the annual drift (percent per year) relative to Aura MLS v4 for SBUV MOD and SBUV COH computed from October 2004 to December 2016. As described above, comparisons using Aura MLS as a transfer standard show SBUV COH with a more positive drift above 5 hPa and SBUV MOD with a more positive drift below 5 hPa.

3.1.5 The BASIC composite and its use for intercomparisons of merged data records

All merged (hereafter also called composite) data sets suffer from artefacts and/or drifts inherent to the instrument data used in their construction, or from absolute offsets and discontinuities when instrument data are combined. As has been extensively discussed in previous sections, the presence of these artefacts can lead to inaccurate and/or more uncertain trend estimates. The BAyeSian Integrated and Consolidated (BASIC) composite is a set of algorithms that, using only the data available, collectively merges multiple ozone composites into one. This Bayesian approach provides a principled way to incorporate prior information about data artefacts (see below), with a Gaussian mixture likelihood that together allows for a robust estimate of true ozone given both the information available and the design of the statistical model (see *Ball et al.* (2017) for details). The approach is designed to take advantage of the common variability present in all the ozone composite data sets to inform, within a probabilistic framework, the most likely ozone time series. Since each composite contains both the real ozone time series and additional composite- and instrument-specific artefacts such as drifts, spikes and discontinuities, the availability of multiple co-temporal and co-spatial time series allows BASIC to account for many of these issues that might remain in any individual ozone composite (*i.e.*, sampling differences, satellite drifts, biases between data sets merged in the composites, and resolution differences between instruments within composites).

The BASIC approach is thoroughly documented in *Ball et al.* (2017), but we briefly describe the steps here. First, errors provided with each composite are formed using different approaches and statistics and therefore cannot be directly compared. Thus for BASIC, the uncertainties for each composite time series are derived independently using singular value decomposition (SVD). The underlying assumption for using SVD is that each composite contains the true ozone time series and a set of

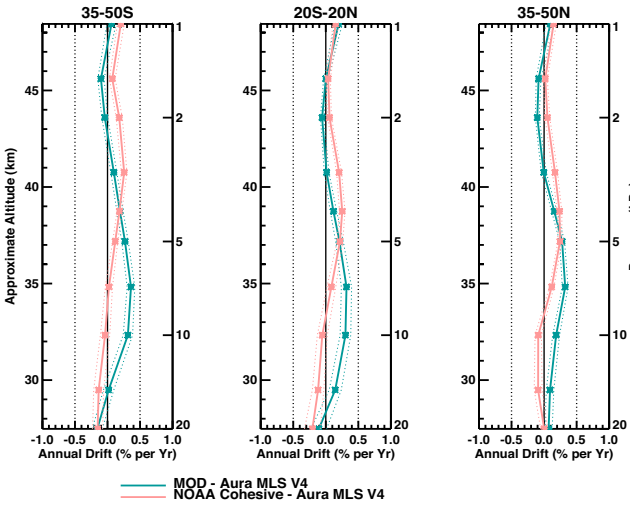


Figure 3.10: *Drift (in % per year) of zonally averaged profile data from SBUV MOD (turquoise) and SBUV COH (red) relative to Aura MLS v4 for 50°S–35°S (left), 20°S–20°N (middle) and 35°N–50°N (right).*

instrument and composite-construction artefacts that should be unique to each composite. Through SVD, the common modes of variability are identified, and if behaviour deviates from these then this is assigned as an uncertainty; ignoring the first SVD mode as the one common to all composites, the higher modes that describe deviations from the first mode are used to form an uncertainty: (see discussion in *Section 3.1.1* about the benefits of multiple datasets to identify and assign the source data causing a discrepancy). Use of SVD does not provide true uncertainties, but assigns an uncertainty based on common behaviour where a deviation of a composite, or group of composites, from the common behaviour leads to an estimated larger uncertainty. Indeed, enhanced uncertainties often correspond to known problems in specific composites or underlying instrument data, so *Ball et al.* (2017) consider this a reasonable assignment of belief, or uncertainty, in the accuracy of the composites. Often the spread in the ozone composites is well within the uncertainties provided with each composite. Second, any prior information about artefacts or drifts in the individual instrument or composite data are incorporated by inflating the SVD-estimated uncertainties; such information can be at times when instruments are known to change in each composite (leading to step-function changes in the time series), or when orbital drifts are known to induce an artificial trend in the time series (*e.g.*, some SBUV instruments in the later 1990s; see *Section 3.1.4*). We note that the choice of using an inflation factor of two is subjective, but we see little difference in the impact of choosing larger values (see *Ball et al.*, 2017). Third, we then form a Gaussian mixture likelihood for each month that allows for the probability distribution of ozone to form multiple peaks. As such, combining information about the most likely state f the ozone in the current month with information available in the preceding and following months leads to a posterior distribution that provides the most likely ozone time series given the information available. To form the posterior estimate of the monthly ozone, we must sample what is a high-dimensional problem using an efficient method such as Markov chain Monte Carlo (MCMC), which we do using Hamiltonian Monte Carlo (HMC; *Neal,* 1993; *Carpenter et al.*, 2016).

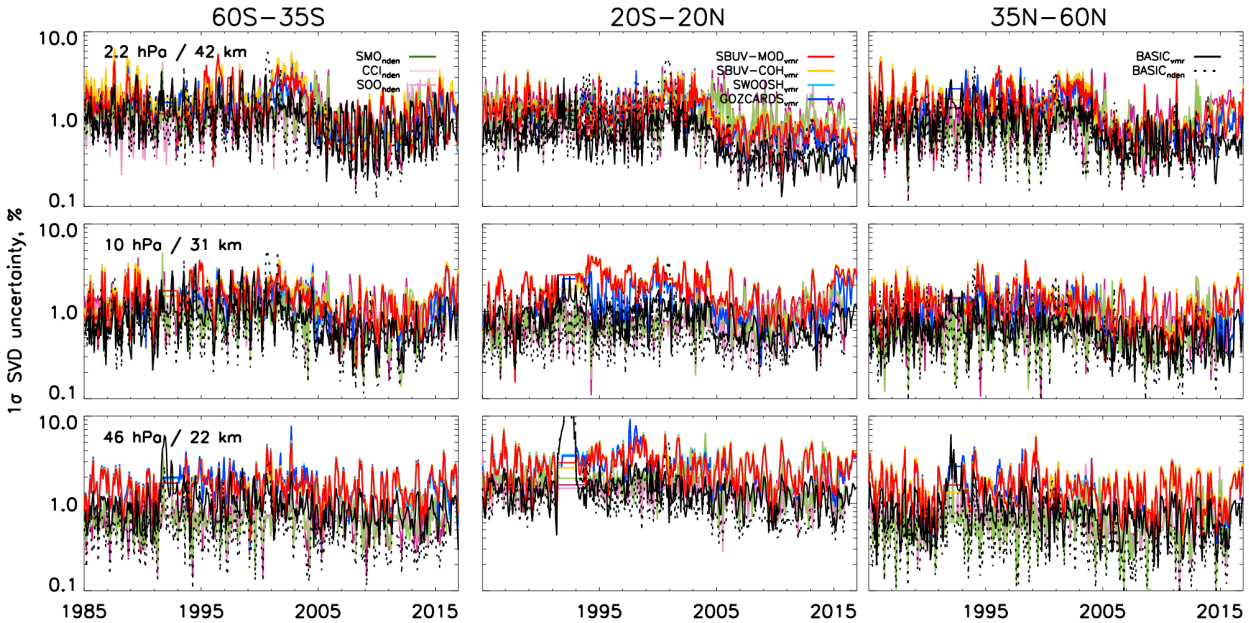


Figure 3.11: *Latitude weighted mean 1-sigma errors (%) estimated from the application of SVD for three number density composites (SAGE-MIPAS-OMPS (SMO), SAGE-CCI-OMPS (CCI), and SAGE-OSIRIS-OMPS (SOO)) and the 1-sigma uncertainty in the $BASIC_{NDEN}$ composite (dotted black line) derived from these, and for four VMR-based composites (SWOOSH, GOZCARDS, SBUV MOD, and SBUV COH) and the $BASIC_{VMR}$ composite (solid black line). Note that number density is on altitude, and VMR on pressure level, so comparing between the VMR and number density data sets is only indicative.*

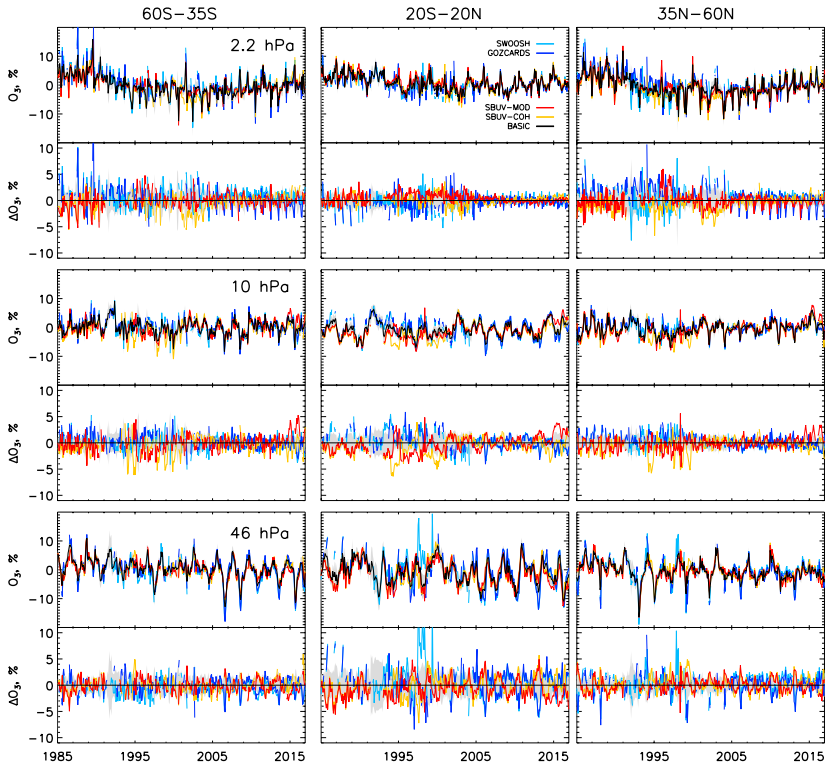


Figure 3.12: *Selected pressure levels in three latitude bands for the four VMR composites and the BASIC$_{VMR}$ composite. Each pair of plots show the relative (%) deseasonalised time series bias-shifted to agree with SWOOSH for the July 2005 to December 2013 period (upper half) and anomalies relative to the BASIC$_{VMR}$ composite (lower half). The 2-sigma uncertainty on the BASIC$_{VMR}$ is shown with grey shading.*

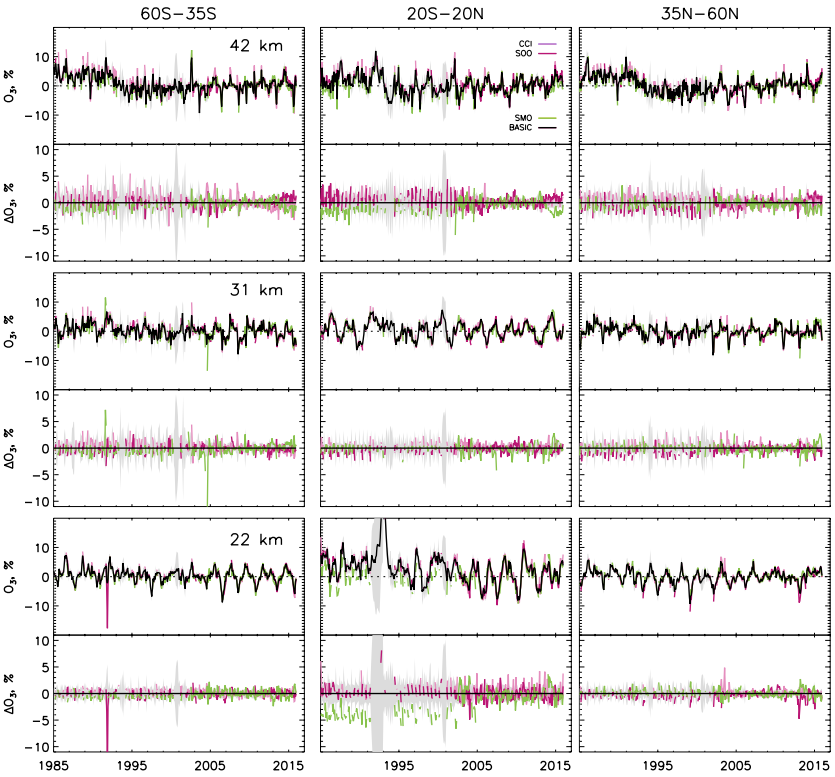


Figure 3.13: *As for **Figure 3.12** but for the number density composites and BASIC$_{NDEN}$ derived from these.*

Examples of the SVD-estimated uncertainties are shown in **Figure 3.11**. These uncertainties are typically at least two times smaller than those provided with composites (where available), which might suggest that provided uncertainties are conservative given that the SVD-uncertainties represent a deviation of the estimated mean uncertainty from the group. Here we have separately estimated uncertainties in four VMR-based composites and three number density-based composites; the uncertainties are not comparable between the sets of composites, only within. Nevertheless, it is interesting to compare all seven in the same sub-figures. For example, in the VMR-based composites it is clear that the uncertainties in the 20°S–20°N region at 10 hPa and 46 hPa are larger between 1994 and 2001; at 10 hPa this is mainly due to the SBUV MOD and SBUV COH composites and reflects a period of known drift in the SBUV-based composites (see *Section 3.1.4*). The number-density uncertainties are typically lower than the VMR-composites, but this reflects the fact that the number density composites are based on very similar underlying data with similar vertical resolutions. The VMR-based uncertainties integrate information from the lower resolution SBUV-based composites, with those at a similar resolution to the number density composites and, as such, the uncertainties are larger to accommodate the higher uncertainty in the absolute level. Nevertheless, the uncertainties in BASIC$_{VMR}$ (based on the four VMR-composites: GOZCARDS, SWOOSH, SBUV MOD, and SBUV COH; solid, black line) and BASIC$_{NDEN}$ (based on the three number density composites: SAGE-CCI-OMPS, SAGE-OSIRIS-OMPS, and SAGE-MIPAS-OMPS; dotted, black) in **Figure 3.11** are similar. Note that sometimes the BASIC$_{NDEN}$ uncertainty becomes temporarily large, which occurs because of missing data in all the underlying composites; this can also be seen during the Mt. Pinatubo eruption, particularly over the Equator at 46 hPa. What appears to be common to all regions presented is that uncertainties prior to 2005 are larger than after this time, especially between 1995 and 2000, which may have a significant effect on the estimated decadal

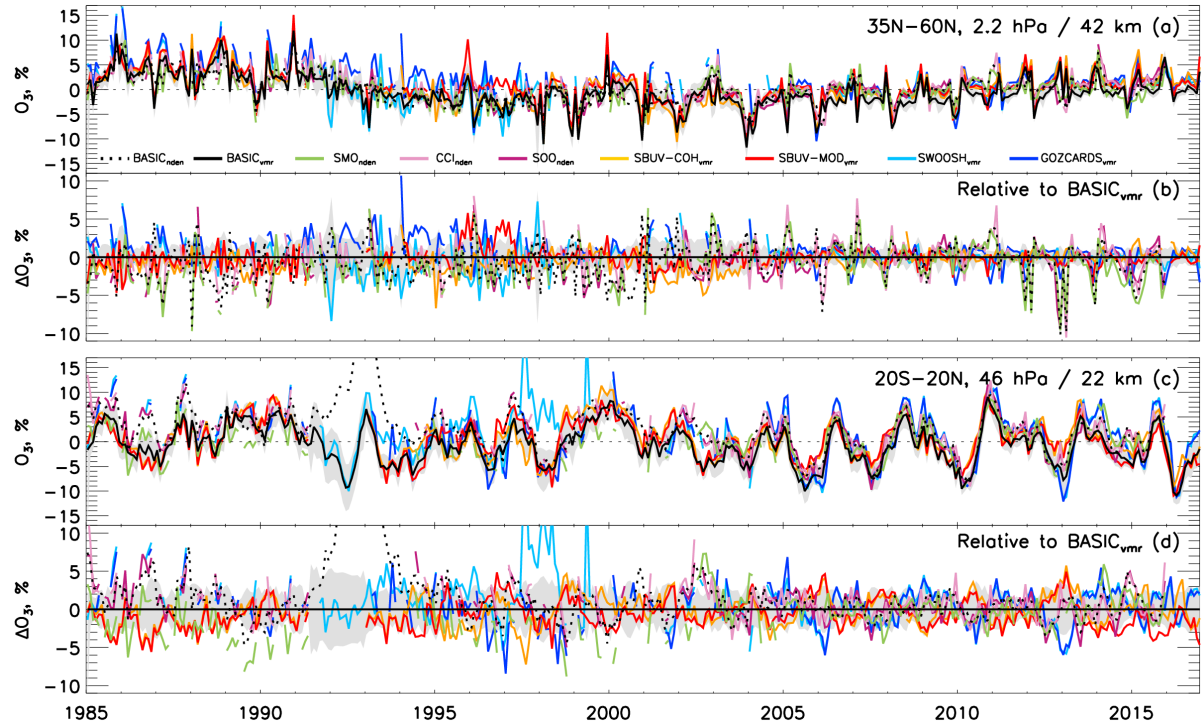


Figure 3.14: *Two levels from* ***Figures 3.12*** *and* ***3.13*** *overlaying the VMR and number density time series for comparative purposes. Each pair of plots show the relative (%) change compared to the July 2005 – December 2012 mean (a & c) and the change relative to BASIC$_{VMR}$ (b & d). Note that while number density and VMR time series shown correspond to approximately the same region in the atmosphere, they are not exact and should be considered only indicative.*

trend (see below) especially since this larger uncertainty results directly from different offsets in the composites. Another interesting feature, apparent in many regions, is that the uncertainties begin to increase again after 2014, which suggests composites are diverging again. This divergence may lead to an inflation in uncertainties in trend estimates even though more data are being accumulated. Understanding why data are diverging again and which are most likely correct will be an important issue to follow up in future work.

Figure 3.12 presents the four VMR-based composites; (upper) relative to the mean of July 2005 to December 2012 and (lower) relative to BASIC$_{VMR}$ for the same regions shown in **Figure 3.11**. Similarly, **Figure 3.13** provides the same for the three number-density composites and BASIC$_{NDEN}$. Panels in **Figures 3.12** and **3.13** represent approximately the same pressure/altitude region in the stratosphere, and the similar short- and long-term variability reflects this. As discussed above, the relative differences between composites (lower panels) are typically larger for the VMR composites than the number density and are larger at higher pressures (lower altitudes), which reflects the decreasing vertical resolution in SBUV. The number density composites contain mostly similar data, so while there are clearly periods of drift and rapid divergence (**Figure 3.12**) between the SWOOSH/GOZCARDS composites (constructed using similar instruments) and the SBUV-composites (especially for 1994–2004 and prior to 1990), the offsets between composites are clearer in the number density composites (lower panels of **Figure 3.13**) because the underlying data are usually the same (SAGE II until 2005, OMPS from 2011). So, if drifts exist in these number density-based data they will not become apparent, and differences will mainly reflect the differences in composite merging or data-screening prior to the merge. This is especially apparent in the SAGE-MIPAS-OMPS data set where, at 22 km in the equatorial region, this data set is offset by 4 % from the other two number density composites, and the BASIC$_{NDEN}$ composite rejects this as an unlikely level of ozone prior to 2004. This is a good example of how the BASIC method goes beyond a simple composite average. Offsets are seen in other panels mainly associated with the SAGE-MIPAS-OMPS composite, and it is these offsets that likely contribute to the larger positive post-2000 decadal trends in SAGE II-MIPAS-OMPS presented in the tropical lower stratosphere presented in *Steinbrecht et al.* (2017) and in *Section 5.1* of this Report (**Figure 5.2**). Once again, the divergence between composites after 2014 appears in several panels in **Figures 3.12** and **3.13**.

We note that the BASIC composites can only estimate ozone as accurately as the information available within the considered composites can permit, that is if all the data are wrong in the same way, at the same time, BASIC cannot estimate the true state of ozone at that time. Therefore, in principle, the more composites that can be incorporated in the analysis, the more useful and robust the result should be. However, the BASIC$_{VMR}$ and BASIC$_{NDEN}$ composites show different variability on monthly and approximately two-yearly timescales because they use different data-types (sources, resolution, vertical grid, units, *etc.*).

In **Figure 3.14**, we make this clear by showing all seven composites and both BASIC composites from two of the panels presented in the preceding figures (35°N–60°N at 2.2 hPa/42 km and 20°S–20°N at 46 hPa/22 km); the lower panels are all relative to the $BASIC_{VMR}$ composite, and the difference between $BASIC_{VMR}$ and $BASIC_{NDEN}$ is clear (solid and dotted black lines, respectively). There is some sensitivity to the exact altitude/pressure level compared in these plots, but choosing an altitude above or below the ones presented leads to broadly similar results. It is important to note that prior to 2004, SAGE II data are in all composites except SBUV COH and SBUV MOD. Therefore, applying the BASIC method to all seven composites treated independently would lead to a bias in the composites containing SAGE II, a concern raised by *Harris et al.* (2015). It would also require a transfer function through a model (*e.g.*, European Center for Medium-Range Weather Forecast Re-Analysis (ERA-Interim)) to put all the composites on the same coordinate system and therefore applying such a coordinate change introduces additional uncertainties (*e.g.*, due to unknown uncertainties in the parameters of the reanalysis models) not considered explicitly in the uncertainty estimate. As such, because $BASIC_{NDEN}$ is based only on the SAGE II composites, and the $BASIC_{VMR}$ equally between those with and without SAGE II data, it is unsurprising that the two BASIC composites show differences prior to 2004, with $BASIC_{NDEN}$ lining up closer to the GOZCARDS and SWOOSH composites that contain SAGE II.

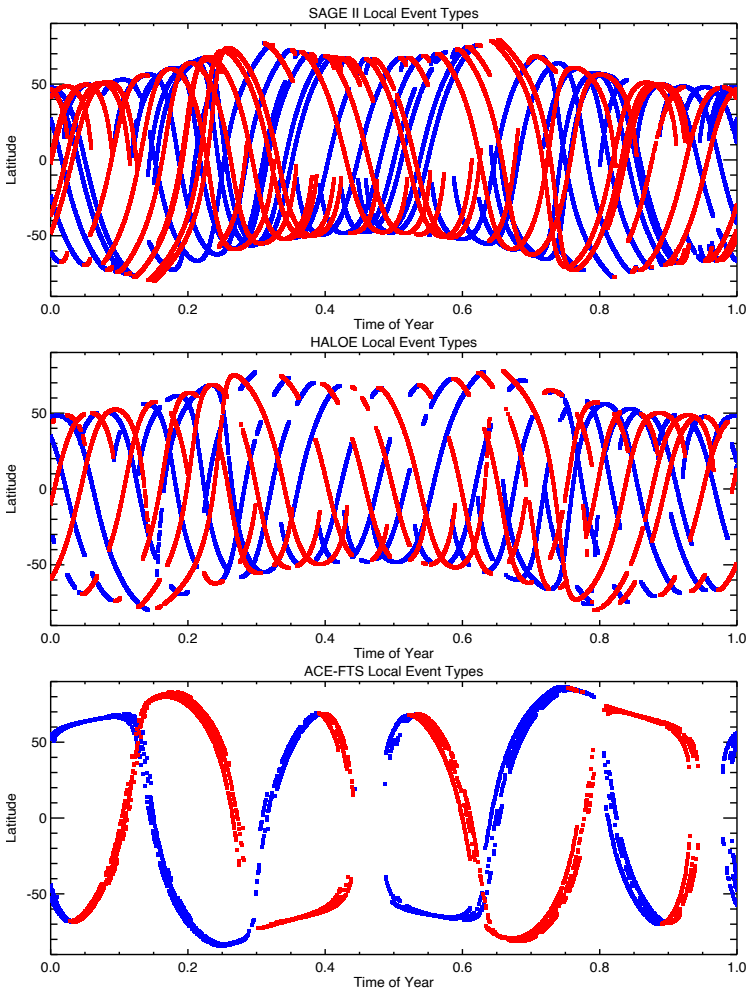


Figure 3.15: *Latitude and time of year of all events for SAGE II, HALOE, and ACE-FTS separated by local event type (blue for sunrise and red for sunset) plotted every 3 years (to reduce clutter) illustrating the drifting sampling patterns over time. Sampling patterns can systematically shift several weeks over a few years for instruments like SAGE II (in its later years) or HALOE (continuously) while ACE-FTS is essentially constant. Time of year is expressed as the modulus of the year fraction.*

A more thorough and detailed analysis of the $BASIC_{VMR}$ composite presented in *Ball et al.* (2017, 2018), along with a trend analysis, reveals that after accounting for many of the artefacts and drifts in the data, differences in post-1997 ozone change profile shapes that have been presented in previous studies (*e.g.*, *Tummon et al.*, 2015) disappear, and the trends in both hemispheres look similar, suggesting that artefacts are indeed important in biasing trend estimates. The enhanced uncertainty between 1995 and 2005, also presented here, shows that trends may also be sensitive to the inflection date used in piecewise linear trend (PWLT) multiple linear regression (MLR, see *Section 4.3.4*). Additionally, *Ball et al.* (2017) found that the use of MLR led to a large, post-1997 negative trend in the tropics (20°S–20°N) peaking at 7 hPa (also reported elsewhere), which disappeared in BASIC when applying dynamical linear modelling (DLM) to estimate the change.

The use of the BASIC method and composites, and comparisons similar to the limited set presented here, should aid composite teams in understanding artefacts in the composites and improving the merging procedure. Revealing artefacts allows for evidence of the reliability of composites to inform our understanding of why composites show diverging decadal trend estimates and to make informed decisions about how these artefacts might be addressed in the future. Ultimately, a more advanced approach is to merge the instrument data underlying each composite only once, using a methodology adapted from that developed in BASIC, which will lead to a single composite that provides the best estimate of ozone given all the satellite data available.

3.2 Sampling bias and uncertainty correction characterisation

WMO (2014) identified three factors that were not accounted for in trend analyses with a potential major impact on resulting trends: Diurnal variability of ozone, biases between data sets, and long-term drifts between data sets. However, there is an additional complication that is intricately tied to these three factors in trend analyses, namely the non-uniform temporal, spatial, and diurnal sampling

of the different instruments used for those analyses. This non-uniform sampling can have a detrimental impact not only on the regression techniques used to derive long-term trends in ozone but also on other analyses performed to determine diurnal variability or the magnitude of potential biases and drifts between data sets.

In order to perform regression analyses to determine long-term ozone trends, data sets are first typically reduced to monthly zonal mean (MZM) values that are utilised as though they are representative of the centre of the month and the centre of the latitude bin. While this assumption is reasonable for highly sampled data sets (*e.g.*, nadir and limb scatter measurements) it generally breaks down when applied to sparsely sampled data sets (*e.g.*, ground or occultation measurements), though even highly sampled data sets are susceptible to changes in the local solar time of observations that can be problematic in the presence of diurnal variability (*Bhartia et al.*, 2013, *Frith et al.*, 2017). This is not a new concept; T*oohey et al.* (2013) and *Sofieva et al.* (2014) both investigated non-uniform temporal sampling as an added source of noise and uncertainty that could be characterised and included in trend analyses. However, orbital drift can lead to a systematic drift in sampling patterns over time, making the standard practice of using deseasonalised anomalies for trend analysis insufficient to remove potential sampling biases. *Millán et al.* (2016) investigated the impacts of non-uniform sampling biases on resulting trends from different instruments by repeating a "representative year" of sampling for each data set and running a model through it over ~30 years to analyse the effect on trends. While illustrative, this did not account for the actual sampling bias as it changed from year to year for those instruments. As such, it is still necessary to consider the non-uniform sampling of different satellite data sets and how representative the derived MZM ozone values are of the actual month and zonal band. Data from ground-based observations can exhibit similar problems and have the additional complication of making measurements from only a single latitude and longitude such that one must also consider their representativeness of the zonal band itself.

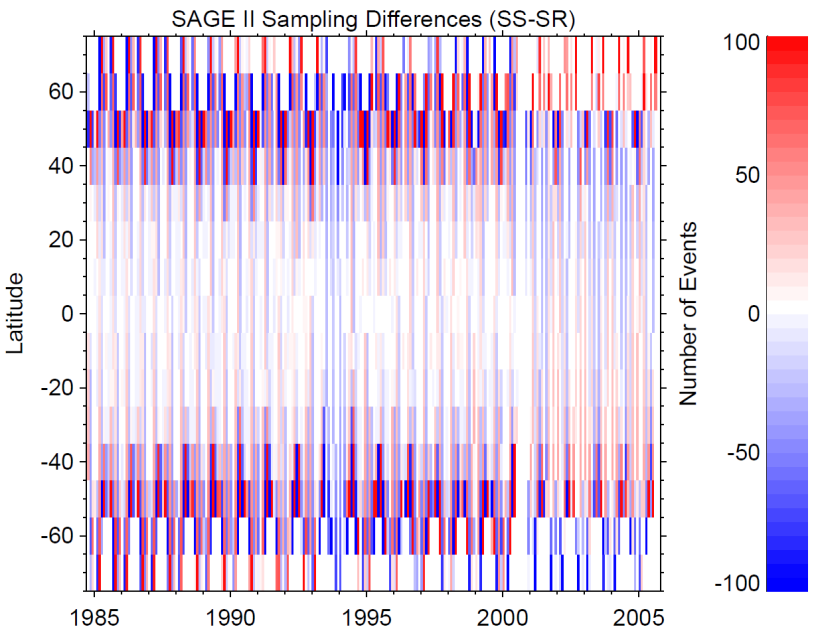


Figure 3.16: *The difference in the total number of sunset (SS) and sunrise (SR) events in each month and 10 degree latitude bin from SAGE II. In addition to the rapid oscillation between SR and SS dominated months, instrument anomalies resulted in large periods and locations of SR/SS dominated sampling (bottom panel of* ***Figure 8*** *of Damadeo et al., 2018).*

3.2.1 Sampling bias for occultation instruments estimated using simultaneous temporal and spatial (STS) analysis

Occultation instruments are classic examples of non-uniform temporal sampling drift. **Figure 3.15** illustrates the drifting sampling patterns of three occultation instruments (*i.e.*, SAGE II, HALOE and ACE-FTS) over their mission lifetimes. SAGE II and HALOE exhibit a systematic drift in sampling towards earlier times of year at every latitude over their mission lifetimes. ACE-FTS has a much slower drift but has a non-uniform distribution of sunrise and sunset measurements as a function of time of year for every latitude, making the separation of the seasonal cycle and diurnal variability impossible when computing monthly zonal means. In addition to non-uniform seasonal sampling, occultation instruments can also exhibit non-uniform diurnal sampling over the mission lifetime. SAGE II is an extreme case of this, where instrument anomalies caused long periods of sunrise or sunset dominated sampling (**Figure 3.16**). In the presence of diurnal variability (*e.g.*, in the upper stratosphere where trends are of the largest magnitude) these diurnal sampling biases can detrimentally impact trend analyses.

Damadeo et al. (2018) discusses the non-uniform temporal, spatial, and diurnal sampling of occultation instruments in great detail and how the use of MZM values can create sampling-induced biases that alias into long-duration variability (*i.e.*, solar cycle and/or long-term trends). Ultimately where (*i.e.*, at what latitudes and altitudes) the sampling biases alias into trend and/or solar cycle results is somewhat "random" as it is dependent upon the chance combination of drifting sampling patterns, spatially-varying seasonal gradients, and frequency of interannual variability. That work also utilises a simultaneous temporal and spatial (STS) regression technique that properly accounts for the non-uniform sampling patterns of occultation instruments (*Damadeo et al.*, 2014) and applies it to the SAGE II, HALOE, and ACE-FTS data sets simultaneously to derive trend results unaffected by sampling biases.

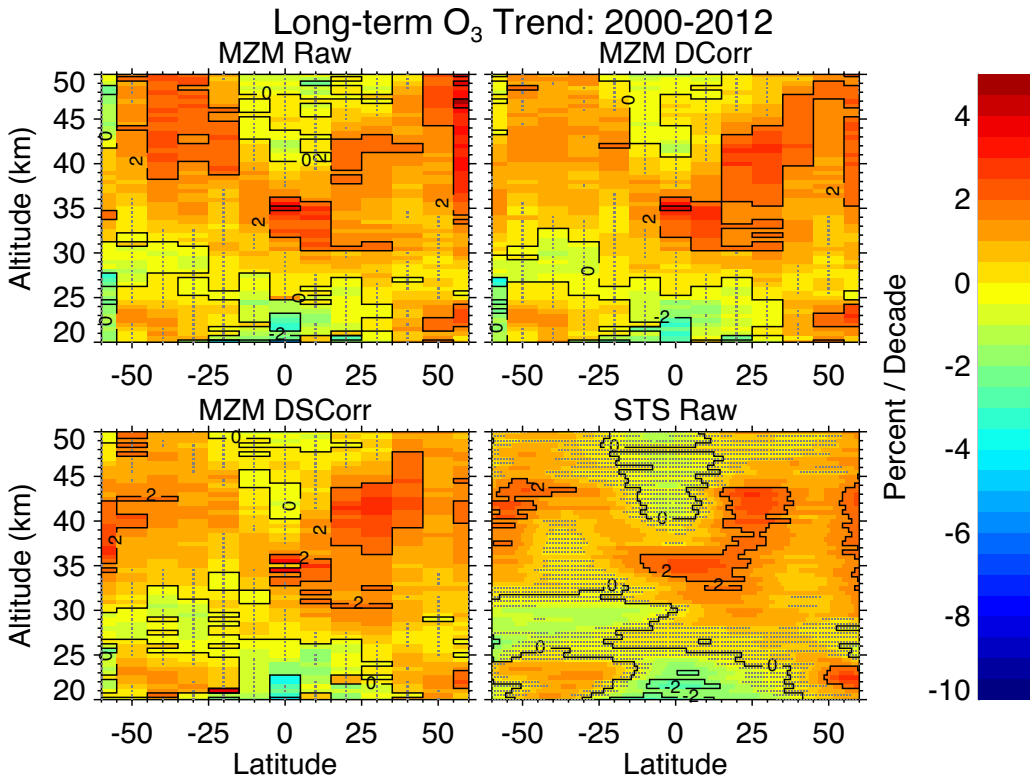


Figure 3.17: *Long-term trends derived from both the MZM and the STS regressions during the potential recovery period. Results are also shown when using the STS regression results to create a diurnally corrected (DCorr) and a diurnally & seasonally corrected (DSCorr) data set for use with the MZM regression. The diurnal correction has the greatest influence on the upper stratosphere while the seasonal correction has the greatest influence at higher latitudes. Stippling denotes areas where the trend results are not significant at the 2σ level. Contour lines are plotted at 2 % intervals. (**Figure 11** from Damadeo et al., 2018).*

Lastly, in an effort to quantify the impact of non-uniform sampling on derived trends when using MZM methodology (*i.e.*, regressing to MZM values separately for each latitude bin), *Damadeo et al.* (2018) uses the results of the STS analysis to create diurnally as well as seasonally corrected versions of these data sets for use with MZM analysis. Each version (*i.e.*, raw, diurnally corrected, and diurnally plus seasonally corrected) is then run through an MZM regression model to derive long-term ozone trends. **Figure 3.17** illustrates the difference in trend results derived between the different "corrected" data sets. The diurnal correction exhibits the largest influence, showing differences in trend results of about 1–2 % per decade in the upper stratosphere at mid-latitudes (*i.e.*, where typical positive trends are largest). The seasonal correction has the largest influence at high latitudes and at the tropical middle stratosphere although at a reduced magnitude of about 0.5–1 % per decade.

Since typically derived "recovery" trends are only about 2–3 % per decade, the influence of non-uniform sampling patterns on derived trends can be significant and is strongly dependent upon what data sets are used and how they are incorporated into the analysis.

3.2.2 Station means versus zonal means

This section focuses on the question of whether monthly averaged ozone partial columns in the middle and upper stratosphere at single lidar stations are representative of the monthly zonal mean. The motivation for this dedicated analysis arises from the wide use of zonal means in calculating trends and in studies related to the interannual ozone variability in cross sections of the middle and upper stratosphere.

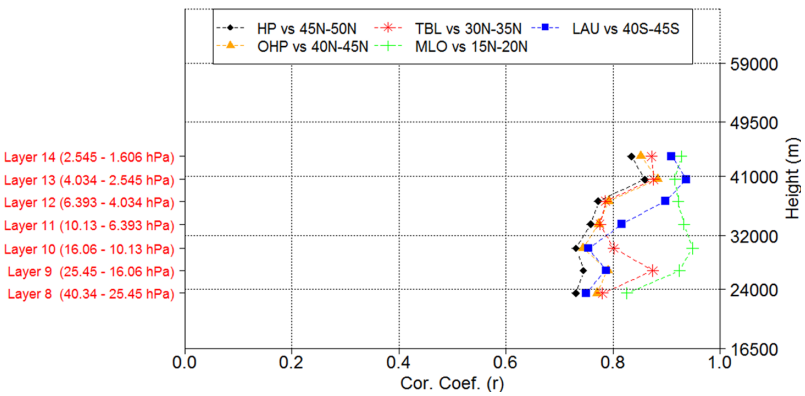


Figure 3.18: *Correlation between monthly mean SBUV overpass data at five lidar stations versus the corresponding 5° monthly zonal mean SBUV data of each site.*

We performed a comparison of lidar station overpass SBUV MOD data and zonal mean SBUV MOD data. A comparison of zonal mean SBUV MOD data to the station mean data by the lidar instrument itself would introduce uncertainty due to the use of different instruments. Five lidar stations with long-term ozone profile data records were chosen: Hohenpeissenberg (47.8°N, 11.0°E), Haute Provence (43.9°N, 5.7°E), and Table Mountain (34.4°N, 117.7°W) in the northern mid-latitudes; MLO (19.5°N, 155.6°W) in the tropics; and Lauder (45.0°S, 169.7°E) in the southern mid-latitudes. Furthermore, the analysis is confined to SBUV layers 8 (40–25 hPa) up to 14 (2.5–1.6 hPa), because the accuracy of the lidar data is limited in the upper stratosphere and that of the SBUV data is limited below the 30 hPa level.

Figure 3.18 shows that SBUV overpass data at the five selected lidar locations are highly correlated with the respective SBUV zonal mean data. Natural oscillations (seasonal, QBO, *etc.*) were removed prior to computing the correlation but not when computing the long-term trends. Correlation coefficients increase with altitude from about 0.75 to 0.9 for all sites and these values are statistically significant. This finding is of particular importance, especially when it comes to the calculation of long-term trends. It suggests that the variability at a single point in the middle and upper stratosphere is comparable to that found in the 5-degree zonal mean data encompassing the lidar station location. This implies that higher frequency spatial variability has little impact at these altitudes, making the derived trends from station data and satellite zonal mean data more directly comparable.

Although the level of agreement between ozone variability at single stations and from zonal means encompassing the stations has yet to be quantified (*WMO,* 2014; *Frith et al.,* 2017; *Zerefos et al.,* 2018), **Figure 3.19** shows an example of the spatial distribution of the correlation coefficients between SBUV overpass data at Hohenpeissenberg and at 633 station locations around the globe. The SBUV MOD data at station locations were downloaded from **https://acd-ext.gsfc.nasa.gov/anonftp/toms/sbuv/MERGED**. The "zonality representativeness" is obvious in the chromatic scale of **Figure 3.19** as well as the finding that as we move higher in altitude, higher correlations are found even at distances exceeding 1000 km. The results are similar when the calculations are repeated between overpasses over the other four available lidar stations and all 633 locations of SBUV overpasses (not shown here, see *Zerefos et al,* 2018). Overall, stations correlate well and are representative over a fairly wide range of longitudes and latitudes. These findings are also true for MLO but, as mentioned in *Section 5.4*, MLO Umkehr cannot represent the tropical belt between 20°S and 20°N. Instead, according to these findings, MLO represents the northern zone well between 15–20°N.

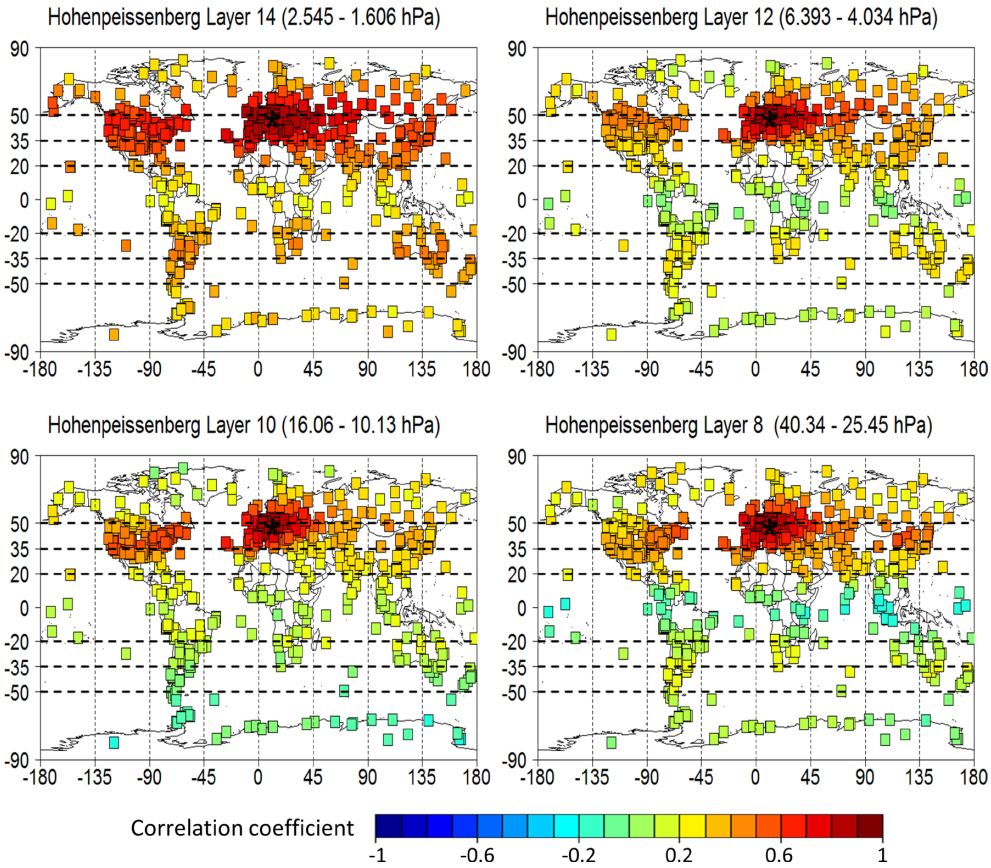


Figure 3.19: *Correlation between the time series, previously deseasonalised and known variability removed, of layered ozone monthly SBUV MOD overpasses at the Hohenpeissenberg station and the SBUV MOD overpasses at various other locations around the globe. Four layers are shown in the panels. The black star indicates the location of Hohenpeissenberg.*

3.3 Summary

Any measurement process unavoidably brings about uncertainties, which ultimately propagate into ozone profile trend uncertainties. Some sources of uncertainty can be directly estimated by the regression algorithm from the time series, others have to be quantified by independent means. For instance, a constant drift in ozone levels over time can be, to a large degree, collinear with the trend proxy term in the regression. It will therefore be absorbed in the trend estimate but not in the trend uncertainty estimate.

In this particular case, as in others, there is a clear benefit of having several complementary contemporary data records since none of the individual satellite or ground-based records provide superior stability over the entire spatio-temporal domain of interest. Intercomparisons of ozone time series of various kinds (single profile measurements, local and monthly zonal means or monthly deseasonalised anomalies, single-sensor or merged records) have revealed measurement-related artefacts, such as drifts, discontinuities, and spikes. For some artefacts the evidence was comprehensive enough to exclude (part of) the data record from further analyses. Other issues were not, or could not be, removed, but they have been taken into consideration in the interpretation of the trend results in *Chapter 5*. These include the drift in a few satellite data records in part of the stratosphere, most notably for OSIRIS and OMPS-LP. Improvements are required, especially for the OMPS-LP data record as it drifts by 5–10 % per decade, most likely as a result of unstable altitude registration. Most ground-based station records exhibit anomalous behaviour during some periods in time. Although the anomalies are broadly consistent with reported systematic errors of 5–10 %, they are episodic rather than systematic in nature. Despite these residual artefacts, the agreement between observational records has generally been improved when compared to the consistency found for earlier data versions used by previous assessments (*e.g.*, *WMO,* 2014; *Tummon et al.,* 2015; *Harris et al.,* 2015; and references therein).

Complementary analysis methods and tools are an asset as well. Comprehensive approaches that intercompare not one or two but many data sets in a coherent way are key in attributing issues to a particular data record. The Bayesian algorithm BASIC proves more robust against outliers than traditional methods to infer the underlying ozone time series from a set of (imperfect) data records. This recent development has shown clear potential in providing insights in more subtle uncertainty patterns relevant for trend studies. MC simulations have proven useful in estimating the additional trend uncertainty related to remaining potential artefacts that cannot be cleanly identified and removed as well as how the merging process deals with these artefacts. For example, seemingly statistically significant discrepancies between trends derived from two SBUV-based records are found to overlap within uncertainty estimates when those estimates include the uncertainty of the individual SBUV data records propagated through the merging process using MC simulations.

The impact of sampling uncertainty on trends is now much better understood. This source of uncertainty is unrelated to the performance of the instrument and becomes only important if the data are analysed at an aggregate level sufficiently far away from that of the original individual profile measurements. Studies using SBUV data showed high correlations between time series at individual sites and those averaged in corresponding 5° latitude belts. The impact of sampling uncertainty is a more important issue for the analysis of monthly zonal mean ozone values by the sparsely-sampled occultation sounders. The interplay of changes in the measurement pattern and diurnal and seasonal gradients lead to systematic changes in derived trends by up to 1–2 % per decade in parts of the stratosphere.

The results described in this chapter are further considered in the interpretation of the trend results in *Chapter 5.*

Chapter 4: The LOTUS regression model

One of the primary motivations of the LOTUS effort is to attempt to reconcile the discrepancies in ozone trend results from the wealth of literature on the subject. Doing so requires investigating the various methodologies employed to derive long-term trends in ozone as well as to examine the large array of possible variables that feed into those methodologies and analyse their impacts on potential trend results. Given the limited amount of time, the LOTUS group focused on the most common methodology of multiple linear regression and performed a number of sensitivity tests with the goal of trying to establish best practices and come to a consensus on a single regression model to use for this study. This chapter discusses the details and results of the sensitivity tests before describing the components of the final single model that was chosen and the reasons for that choice.

4.1 Regression methodology

MLR methods have been used for trend detection in ozone time series for decades. They evolved into the most commonly used approach in the community, with many smaller or more substantial variations of a baseline method having been developed. For the LOTUS project, we decided to base our sensitivity tests on a MLR with an iterative lag-1 autocorrelation correction (see *Appendix B* of *Damadeo et al.,* 2014).

In general, the regression problem can be written as,

$$y = X\beta + \varepsilon \qquad (4.1),$$

where y is the length n vector of observations, β is the length m vector of proxy coefficients, X is the $n \times m$ matrix of proxies, and ε are the fit residuals. The goal of the regression procedure is to find the values of β which minimise the quantity

$$(y - X\beta)^T \Omega^{-1} (y - X\beta) \qquad (4.2),$$

where Ω is the covariance matrix of the observations. The problem admits a direct solution,

$$\beta = (X^T \Omega^{-1} X)^{-1} \Omega^{-1} y \qquad (4.3),$$

which can also be used to obtain an error estimate for the proxy coefficients assuming the covariance matrix is correctly specified.

The regression is performed in an iterative procedure (*Cochrane and Orcutt*, 1949) with Ω set to unity for the first iteration. The first iteration is equivalent to an unweighted least squares fit. After the first iteration, the autocorrelation coefficient, ρ, is calculated through,

$$\rho = \frac{\sum_{i=2}^{N} (\varepsilon_i - \overline{\varepsilon})(\varepsilon_{i-1} - \overline{\varepsilon})}{\sum_{i=1}^{N} (\varepsilon_i - \overline{\varepsilon})^2} \qquad (4.4),$$

where $\overline{\varepsilon}$ is the mean value of the residuals. Typically, the autocorrelation coefficient is on the order of 0.2–0.3. For the next iteration, the covariance matrix is updated taking into account the observed autocorrelation (*Prais and Winsten,* 1954) with modifications by *Savin and White* (1978) if gaps are present in the data. The procedure is repeated until the autocorrelation coefficient has converged within a tolerance level of 0.01. The final error estimate is calculated by scaling the covariance matrix to match the observed variance of the residuals.

This baseline MLR was used for sensitivity tests to decide which proxies to use in the final "LOTUS regression" model, for evaluations of possible lags for proxies, and for the evaluation of weighted or unweighted regressor data. The final set-up of the "LOTUS regression" model is described in more detail in *Section 4.5*.

4.2 Proxies

Proxies are used in multiple regression analyses to represent the observed variability in the parameter being modeled, in this case ozone. There is a wealth of literature concerning the viability of various proxies to represent dynamical and chemical processes that affect ozone (*e.g.*, *WMO,* 2011; *WMO,* 2014; and references therein).

We briefly describe the most common proxies for ozone trend analyses below and provide information on where these proxies may be found. Our focus is therefore not to provide detailed studies about the effects of these proxies on ozone distribution but rather a short estimate about their influence mechanism (dynamical or chemical) and a description on how the proxy has been implemented in regression models before. The listed links for the proxies are not exhaustive and should only be seen as a subset of all available possible sources.

4.2.1 Non-trend proxies

11-year solar cycle

The 11-year solar cycle has different effects on ozone in the different regions of the atmosphere.

The solar ultraviolet spectral irradiance reaching the Earth's atmosphere changes over the course of the cycle. In the upper stratosphere this leads to changes in radiative heating and photochemistry (production rate of ozone) which then affects the ozone distribution. In the lower stratosphere the changes in ozone are thought to occur mainly through a dynamical response to solar ultraviolet variations. The exact mechanisms of this dynamical response are not yet fully understood (*WMO,* 2014), but the 11-year solar cycle proxy is important for all latitudes. Effects of the solar cycle on ozone are described, for example, in *Lee and Smith* (2003).

There are different possibilities to describe the 11-year solar cycle as a proxy. The most common ones are: 10.7 cm solar radio flux, sunspot number, Mg II core-to-wing ratio, and as a more recent alternative the 30.0 cm solar radio flux (as suggested by *Dudok de Wit et al.,* 2014). Note, all of these time series are highly intercorrelated, so only one is chosen for a solar cycle representation in a regression model. Additionally, while these proxies could theoretically be phase shifted to account for any potential lagged response (*e.g.*, in dynamical forcings in the lower stratosphere), in practice adding this additional degree of freedom can cause aliasing due to correlations with volcanic effects (*Chiodo et al.,* 2014; *Damadeo et al.,* 2014) that can negatively impact trend analyses.

Time series of these proxies can be found here:

i. Solar flux (10.7 cm):
   ftp://ftp.geolab.nrcan.gc.ca/data/solar_flux
ii. Solar flux (30.0 cm):
   ftp://ftpsedr.cls.fr/pub/previsol/solarflux/observation
iii. MG II index:
   http://www.iup.uni-bremen.de/UVSAT/Datasets/mgii
iv. Sunspot number:
   http://sidc.oma.be/silso/datafiles

QBO

The quasi-biennial oscillation (QBO) is a modulation of the zonal wind and temperature in the tropical stratosphere over time and pressure region, measured by radiosondes. These changes in wind and temperature affect ozone in this region, as well as ozone outside the tropics. The QBO time series are given as wind measurements at several different stratospheric pressure levels. Effects of QBO on ozone are described, for example, in *Baldwin et al.* (2001).

Due to the oscillating nature of the QBO, and its region of influence beyond the tropics, it is important to account for phase shifts when it is used as a proxy. There are different ways to accomplish this: (1) at least two (or more) measured QBO time series from different pressure levels that are mostly orthogonal are used simultaneously (as done, for example, in *Steinbrecht et al.*, 2017), (2) one measured QBO time series is chosen and an orthogonal time series to this QBO time series is artificially created (as done, for example, in *Harris et al.,* 2015), or (3) QBO time series at multiple pressure levels are taken and a principle component (CP) analysis is performed with them to get to orthogonal QBO time series (as done, for example, in *Damadeo et al.,* 2014).

Time series of QBO values at the different pressure levels can be found here:

http://www.geo.fu-berlin.de/met/ag/strat/produkte/qbo/qbo.dat

ENSO

The El Niño–Southern Oscillation (ENSO) is an important mode of interannual variability in wind and sea surface temperatures over the tropical Pacific Ocean. These variations cause variability in tropical upwelling and therefore changes in lower stratospheric temperature and water vapor. These then affect the ozone concentration in the tropics chemically and dynamically. Through atmospheric teleconnections, ENSO also affects ozone distributions in regions beyond the tropics (*WMO,* 2014; *Oman et al.,* 2013). ENSO time series can be given as sea level pressure difference between Darwin, Australia, and Tahiti (Southern Oscillation Index), as differences in sea surface temperatures, or as a combination of several different indices.

ENSO as a proxy is often used only as a single time series. However, it has been shown that ENSO effects outside the tropics can be delayed by some time compared to the original signal. Therefore ENSO proxies are either lagged (by a variable number of months) or an orthogonal ENSO time series is created, and the original and orthogonal ENSO proxy are used in combination to account for the time lags.

Time series of ENSO values can be found here:

https://www.esrl.noaa.gov/psd/enso/mei
(*Wolter and Timlin*, 2011)

AO

The Arctic Oscillation (AO), also known as the Northern Annular Mode (NAM), is a description of North-South

movement of the westerly winds that circle the Arctic associated with pressure anomalies of one sign in the Arctic with the opposite anomalies centred about 37°N–45°N. The effect of the AO on ozone is discussed, for example, by *Thompson and Wallace* (2000).

Time series of AO values can be found here:

https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/ao.shtml

AAO

The Antarctic Oscillation (AAO), also known as the Southern Annular Mode (SAM), is a description of North-South movement of the westerly winds that circle Antarctica associated with pressure anomalies of one sign centred in the Antarctic and anomalies of the opposite sign centred about 40°S–50°S. It can have a clear influence on ozone in the Southern Hemisphere polar regions as has been shown by *Thompson and Solomon* (2002).

Time series of AAO values can be found here:

https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/aao/aao.shtml

NAO

The North Atlantic Oscillation (NAO) index is based on the pressure difference at sea level between the Icelandic low (Subpolar Low) and the Azores high (Subtropical High). Its variations impact the strength and direction of westerly winds across the North Atlantic, which can then affect ozone distribution, mainly in the lower stratosphere. The effects of NAO on ozone is described, for example, by *Weiss et al.* (2001).

NAO can be given as a time series of air pressure differences between a location in Iceland and a location in the Azores or Portugal, or as empirical orthogonal functions (EOFs) of surface pressure defined regionally about these locations.

Time series of NAO values can be found here:

https://climatedataguide.ucar.edu/climate-data/hurrell-north-atlantic-oscillation-nao-index-station-based

EHF

The Eddy Heat Flux (EHF) is a metric used to describe the stratospheric meridional circulation, based on wave energy moving from the troposphere upward into the stratosphere. It is calculated as a product of North-South (meridional) wind and temperature departures from their respective zonal-mean values, and it is mainly important in the winter hemisphere in the polar regions. More information about the influence of EHF on ozone can be found, for example, in *Gabriel and Schmitz* (2003).

Eddy Heat Flux time series are calculated from reanalysis data, for example ERA-Interim.

Tropopause pressure

Long-term changes in tropopause pressure can be used as a proxy for tropospheric expansion and therefore as an indicator for the influence of climate change on the atmosphere. The connections between ozone and tropopause height has been discussed, for example, by *Steinbrecht et al.* (1998). The average tropopause pressure for use as a proxy in the regression has to be calculated specifically, depending on the analysed data: For station data (*e.g.*, a sonde station), normally the tropopause pressure that is recorded directly in the data files is used; for satellite data analyses, tropopause pressures are normally retrieved from reanalysis data.

Aerosol

Stratospheric sulfate aerosol concentrations can have an impact on ozone concentrations, especially in the polar regions. With increasing sulfate concentrations the surface area of atmospheric particles increases and therefore offers more opportunity for heterogeneous chemical processes and the effects on stratospheric temperature and transport changes. A major source of sulfate aerosols comes from volcanic eruptions. Two major eruptions took place in the last 40 years (*i.e.*, El Chichón in 1982 and Pinatubo in 1991) and have to be considered with ozone trend analyses. The peak of the atmospheric aerosol concentration varies regionally due to the transport times of the aerosols. Some satellite and ground-based instruments had substantial problems measuring ozone during periods of very elevated stratospheric aerosol, which is normally accounted for by removing several years of data after major volcanic eruptions from these data sets. More information about the effects of stratospheric sulfate aerosols on ozone can be found, for example, in *Solomon et al.* (1998).

There are different ways to describe aerosol as a proxy: (1) a theoretical functional form that describes the injection and loss of aerosol in the atmosphere, or (2) time series based on aerosol-related measurements. In both cases lags for the time series have to be considered to take into account the transport times.

Time series of aerosol optical depth (AOD) values can be found here:

> Mean AOD at 550 nm:
> **https://data.giss.nasa.gov/modelforce/strataer/tau.line_2012.12.txt**
> (Note, this time series is only available until mid-2012; if it was to be used as a proxy in a regression model, it would have to be extended. This is normally done by extending the average value of the last few years or the last value of the time series, assuming that during those years the mean AOD was representative of background values.)

4.2.2 Trend proxies

In addition to the periodic or punctuated influences of geophysical variability detailed above in *Section 4.2.1*, studies of long-term ozone levels also reveal an additional trend-like behavior. These long-term changes stem primarily from long-term variability in chemically reactive halogens (*e.g.*, *Molina and Rowland,* 1974*; WMO,* 2014 and references therein) and the effects of steadily increasing GHG on stratospheric temperatures and dynamical transport processes (*e.g.*, *Shepherd et al.,* 2008; *Li et al.,* 2009). Halide compounds will photodissociate, releasing halogens (*e.g.*, chlorine or bromine) that chemically interfere in the Chapman cycle resulting in catalytic ozone loss and a subsequent inverse relationship between halide loading and ozone levels that is most prominent in the upper stratosphere. The continued injection of GHGs (*e.g.*, carbon dioxide or methane) into the atmosphere results in stratospheric cooling, which slows the reaction rates of chemical ozone loss, and tropospheric heating that have a combined effect on dynamical transport mechanisms such as increasing the Brewer-Dobson circulation. This increased tropical upwelling is predicted to result in the decrease of ozone in the tropical UTLS and increase in ozone at mid-latitudes in the lower stratosphere and these dynamical influences can potentially be more influential than chemical forcing in the lower stratosphere. However, while changes in greenhouse gases exhibit a linear behavior, halogen concentration peaked in the mid-1990s and then began to decline. This combination of long-term effects complicates the ability of regression models to accurately derive long-term trends in ozone. Whereas early (*i.e.*, before the mid-1990s) works could make use of a simple linear trend to model long-term changes in ozone, studies thereafter have utilised more complicated trend proxies in regression analyses that are detailed below.

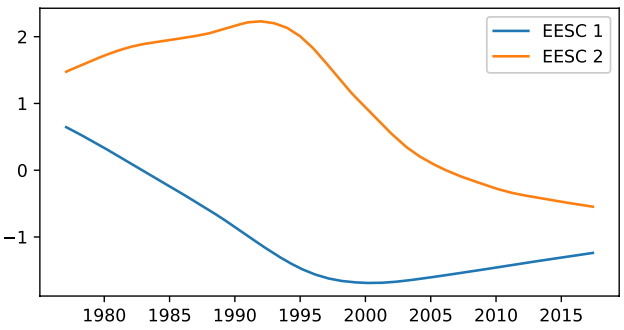


Figure 4.1: *The leading two EESC EOF terms derived from multiple mean age-of-air EESC time series proxies.*

EESC

The equivalent effective stratospheric chlorine (EESC) proxy describes the total halogen loading (chlorine and bromine) of the stratosphere that contributes to ozone depletion (*Newman et al.,* 2007). The shape and timing of the peak of the EESC time series depend on the strength of the Brewer-Dobson circulation and are therefore different for different locations in the atmosphere. To account for this in ozone regression analyses, two different methods have been used: (1) creation of individual EESC time series depending on the age of air at each chosen location (see for example *Bodeker et al.,* 2013), (2) creation of two orthogonal functions that allows the regression to determine the shape that best fits the data (see **Figure 4.1** and *Damadeo et al.,* 2014). Note that the combination of EESC-based orthogonal functions allows for an EESC-like function but with maximum halogen loading occurring at any point in the time series, or not at all (*i.e.*, monotonic trend), depending on the best fit to the data. That is, the fit is not constrained to the range of classically-defined EESC curves. Significant differences between actual and EOF-based EESC proxy fits indicate the data are diverging from a linear fit to actual EESC or responding to other forcing not explicitly represented in the regression model.

Time series of EESC values can be found here:

> **https://acd-ext.gsfc.nasa.gov/Data_services/automailer/index.html**

PWLT

A piecewise linear trend proxy (PWLT) is a combination of two linear trend terms. The first is a regular linear trend term, while the second is a linear trend term that is set to 0 until a specific time (inflection point) and is a simple linear trend afterwards. The two lines of the trend proxies are forced to meet at the inflection point. The trends in the two periods (before and after the inflection point) are therefore linked. The inflection point is chosen to coincide with the peak concentration of ODSs in the atmosphere. Since the timing of this peak changes depending on the location in the atmosphere, ideally the PWLT proxy takes this variability into account. Often, however, PWLT is applied with the same inflection point (end of 1997) at every location. Both trend terms of PWLT are fit simultaneously with the other proxies. A PWLT was used, for example by *Harris et al.* (2015).

ILT

The independent linear trend (ILT) method is also based on two different trends (like PWLT) to describe the ozone decrease in the 1980s and 1990s and the slow ozone increase since the early 2000s. Unlike the PWLT, ILT results are not linked but fit independently for the two chosen time periods. The analyses with ILT can be done in one or two steps. For the two step approach all available proxies for the ozone regression are used in a first regression fit. Then the contributions of the proxies beside the trend proxies are removed from the original data, which leaves only the contribution of the trend proxies and the residuals. In the last step the trend proxies (two different proxies for two different time periods) are fit to these already modified data to determine the trends. A regression using ILT was used, for example, by *Steinbrecht et al.* (2017). In a one step approach the two steps explained above are combined but with one additional regression term representing a constant for describing the period between the two trend terms.

4.3 Sensitivity tests

Part of the difficulty in deriving long-term trends in stratospheric ozone is the large variety of choices to be made during the process. There are a large number of data sets from different instruments that have been combined in different ways and with different merging techniques (see *Chapter 2*). There are also a large number of potential proxies to choose from to use in regression analyses, each with some apparent merit for use in different temporal and/or spatial regimes. Indeed, the SI2N effort saw a great number of different analyses applied to various different combinations of data sets performed using different combinations of proxies and different regression methodologies by different groups (*Harris et al.,* 2016). In order to attempt to disentangle the effects of these variables (*i.e.*, data sets used, proxies used, and analysis technique), a number of sensitivity tests were performed. These sensitivity tests were designed to determine what variables are influential versus non-influential and to try to establish best practices if possible.

4.3.1 Survey of existing regression models

Perhaps the main difference between recent ozone trend studies has been the varied use of regression-based models and data sets. As such, a logical first step in performing sensitivity tests would be to apply the same set of regression-based methodologies to a single data set. This serves two purposes: (1) to validate the consistency of execution against previous studies and (2) to probe the sensitivity of resulting trends to different combinations of proxies and methods. In considering this, groups involved in past ozone trends studies were asked to run their regression models on a single data set, SBUV MOD, provided by the LOTUS working group. This data set was recomputed over large latitude bins (*i.e.*, 50°S–35°S, 20°S–20°N, and 35°N–50°N) and nine different pressure levels between ~40 hPa and ~0.6 hPa (modified from *Frith et al.,* 2017) and was chosen for its ease of use for this particular sensitivity test. The data were given to each of 15 different groups to apply their regression methodology on and provide the LOTUS group with all of the relevant results (*i.e.*, coefficient values and uncertainties and proxies used). Each separate regression analysis was applied as the group has done in the past (*i.e.*, with regard to regression methodology and choice of proxies) with the exception of restricting the long-term trend term to a piecewise linear trend with a turnaround time at the end of 1997 (*Jeannet et al.,* 2007; *Zerefos et al.,* 2012; *Coldewey-Egbers et al.,* 2014; *Damadeo et al.,* 2014; *Fragkos et al.,* 2016; *Misios et al.,* 2016; *Ball et al.,* 2017; *Sofieva et al.,* 2017; *Steinbrecht et al., 2017; Weber et al.,* 2018).

Figure 4.2 shows ozone trend results for each of the 15 different regression models in two separate latitude bands (50°S–35°S and 35°N–50°N) for the time periods before/after 1997 with dashed/solid lines. Since this sensitivity test was applied to only a single data set, the emphasis here is not on the trend values themselves but rather how they compare.

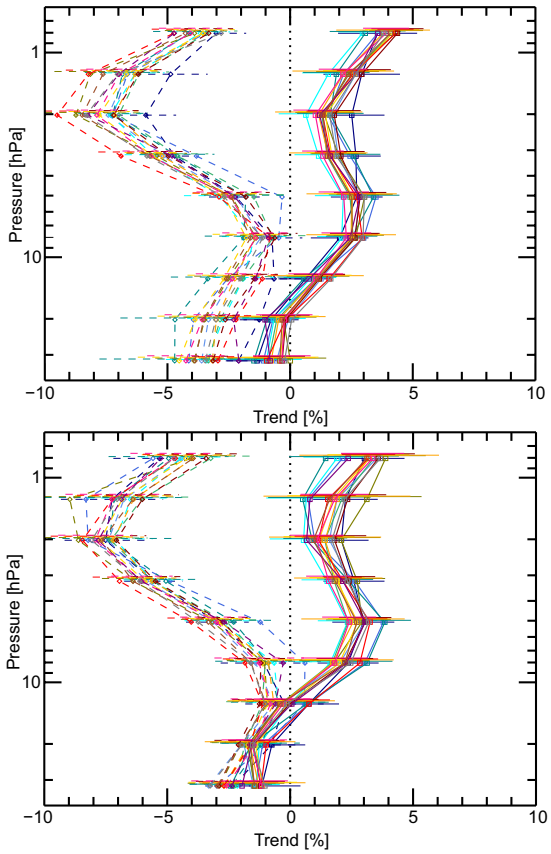


Figure 4.2: *Derived ozone trends in percent per decade from the 15 different regressions applied to the same SBUV MOD data set between 35°N–50°N (top) and 50°S–35°S (bottom). In each plot, the dashed/solid lines represent trend values before/after 1997 at each of the 9 pressure levels.*

The different regression models are mostly in agreement (*i.e.*, within their respective error estimates) at all latitudes, pressure levels, and time periods showing how the resulting trends are fairly robust with respect to the choice of regression methodology and non-trend proxies. However, there is still a spread in the trend values of about 1.5 % per decade in the lower to mid-stratosphere to 3 % per decade in the upper stratosphere, particularly in the Southern Hemisphere. With current estimates of "recovery" trends in the upper stratosphere at mid-latitudes of about 2–3 % per decade, this spread is a potentially large source of uncertainty. Results for the latitude bands 20°S–20°N and 50°S–50°N (not shown here) are very similar in their spread of results between the different regression models. As such, it was deemed prudent to perform more extensive sensitivity tests to see how trend results would change given different regression methodologies and proxies.

4.3.2 Weighted versus unweighted regression

With a few exceptions, nearly every regression analysis of long-term trends in ozone have involved unweighted regressions. The downside to this, of course, is that the uncertainties in the observations (or the resulting mean values) are never taken into consideration for the calculation of regression coefficients or the resulting uncertainties. As such, we applied both unweighted and weighted regression techniques to the data using the standard error of the mean as a weighting factor. The weighted regression and associated heteroscedasticity correction were applied as detailed in *Appendix B* of *Damadeo et al.* (2014).

Weighted regressions use the inverse of the variance in the data as the ideal weights for the regression, which is typically substituted with the inverse of the square of the uncertainties (*i.e.*, standard errors). However, an often forgotten assumption of this technique is that the variances in the data are known precisely. Since, in practice, this assumption almost never holds, a heteroscedasticity correction is necessary to attempt to modify the uncertainties using the nature of the residuals. The form of this correction, however, can be more complicated as it requires some a priori knowledge of the nature of the modification. Given that the standard errors change with both geophysical variability and the number of samples, the heteroscedasticity correction was assumed to have a seasonal form and allowed to vary for time periods containing different collections of data sets (see *Appendix B* of *Damadeo et al.*, 2014).

As an example we investigated the impacts of using weighted regressions on the merged SAGE-OSIRIS-OMPS ozone data set. **Figure 4.3** illustrates the importance of the heteroscedasticity correction. Because the standard error has a strong dependence on the sampling frequency, the weights increase significantly as the data set moves from lower samples (*i.e.*, SAGE II) to higher samples (*i.e.*, OSIRIS and OMPS). This creates mismatched weights over different time periods and results in trends that are dictated entirely by the latter data sets. Incorporating the heteroscedasticity correction helps to "even out" the weights to a more reasonable representation. Analysis of the residuals (**Figure 4.4**) reveals that the heteroscedasticity also reduces the variance in the residuals, resulting in a more robust fit. The standard deviation of the residuals is both generally reduced and more uniform over time after the heteroscedasticity correction. However, **Figure 4.4** also reveals the limitations of this correction. As previously mentioned, the heteroscedasticity correction can be complicated as it requires some a priori knowledge of the nature of the uncertainty modification. Since the standard error is strongly dependent on the sampling, it can change not only as new instruments are added but from month to month even within a single latitude bin, particularly during periods of overlap between instruments with very different sampling. As such, the true nature of the heteroscedasticity is much more complicated than can be simply modeled and it becomes extremely difficult to accurately correct for data sets that are already merged. Instead, this correction should be performed on each individual data set prior to merging. Since the analyses within the LOTUS investigation only use pre-merged data sets, an accurate heteroscedasticity correction cannot be computed and thus a weighted regression technique is not included in the final results.

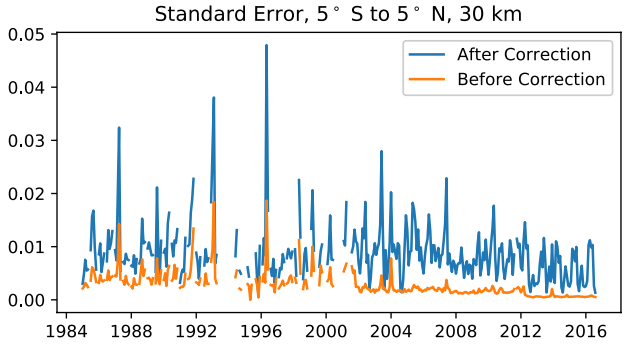


Figure 4.3: *Standard errors between 5°S–5°N at 30 km used for the weighted regression applied to the merged SAGE-OSIRIS-OMPS ozone data set both before and after the heteroscedasticity correction. Note how the errors/weights before the correction drop/rise dramatically once OSIRIS (added in 2001) and OMPS (added in 2012) are added as a result of increased sampling of the instruments.*

4.3.3 Non-trend proxy sensitivity

When performing MLR of a dependent variable (*e.g.*, ozone) to a set of proxies (*e.g.*, sources of geophysical variability), the primary motivation is to determine attribution (*e.g.*, how much variability in ozone is caused by solar flux). Precisely determining attribution requires that each proxy is orthogonal to (*i.e.*, has zero correlation with) every other proxy. Unfortunately this is almost never the case and the proxies used for the regression often have some degree of multicollinearity (*i.e.*, an individual proxy or combination of proxies is somewhat correlated to another individual or

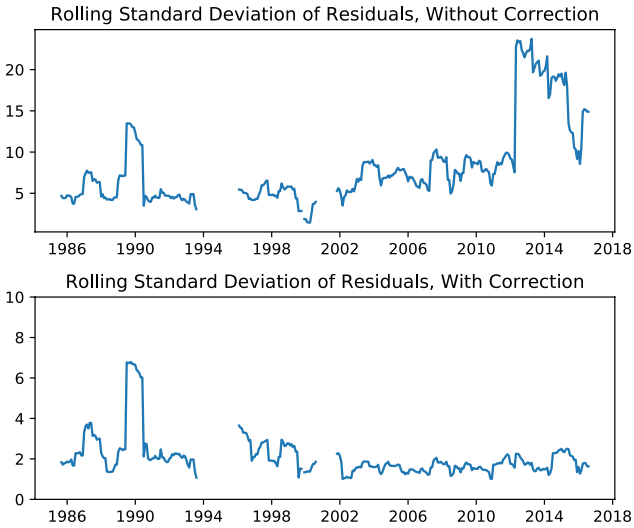


Figure 4.4: *Rolling (running average) standard deviation of the residuals of the weighted regression between 5°S–5°N at 30 km both with and without the heteroscedasticity correction. In addition to generally reducing the spread of the residuals of the fit, the heteroscedasticity correction also makes the rolling standard deviation more uniform throughout the data. Prior to the correction, the rolling standard deviation (not the standard error) increased when moving from SAGE II data to OSIRIS and later OMPS data.*

combination of proxies). While multicollinearity can affect the regression results, this information is captured in the covariance matrix and the final uncertainty estimates of the regression coefficients. However, our focus is on the impact of variability that, if not properly accounted for, may alias into the long-term trend estimate. As such, proxies that do not impact the long-term trend estimate or uncertainty only serve to complicate the statistical model and possibly alter the fits to other regression terms. To test the impact of various proxies on the trend, in the following sections we compare trend results from regressions run with and without the proxy included.

Many different MLR analyses of ozone have been performed over the years using many different combinations of proxies. To test the sensitivity of the regression to these various proxies, a simple model consisting of the leading two QBO EOFs, ENSO with zero lag, solar f10.7, and a PWLT with inflection point in 1997 was applied to each of three different data sets (*i.e.*, SAGE-OSIRIS-OMPS, GOZCARDS, and SBUV MOD) and then a single proxy was either added, subtracted, or substituted from this model. Since the aim of LOTUS is to investigate trends, the following figures compare the derived trend results (**Figures 4.5** and top row of **4.6**) and their uncertainties (**Figure 4.6,** bottom row) before and after each proxy change. Trend differences are shown as differences after the proxy change minus before the proxy change with units of percent per decade. Uncertainty / significance differences are calculated as differences after the change minus before the change with units of percent with the calculated value being the ratio of the trend to its uncertainty. Thus, for example, if initially a trend value was significant to the 2-sigma level and then afterwards it was significant to the 1-sigma level, the significance difference would be -50 %.

First it is worth noting what proxies had a very small impact on the trends differences and their uncertainties. Including any of the AO, AAO, NAO, or EHF proxies had a negligible impact on both the pre-1997 and post-1997 trends value for all three satellite data sets. Also, while some small differences are apparent in the resulting significances, none of the changes were sufficient to make trends that were not significant become so. It is also worth noting that each test that removed the QBO, ENSO, or solar proxies had a significant impact on the trends (1–2% per decade depending upon altitude and latitude) and a general decrease in overall significance levels, though there was some dependence on which data set was used.

The QBO proxy is often taken as the first two EOFs derived from the Singapore zonal winds. Even though higher order terms could be used, adding a third QBO EOF into the regression had negligible impact on the trend and uncertainty results. The ENSO proxy was applied without any lag to it, though often some lag between one to several months is used. Applying the regression with an ENSO proxy lagged anywhere between 1 and 5 months had negligible impacts on the determined trend values but significant impacts on the uncertainties. Including a lag (any lag) generally increases the overall significance, but the results are sporadic in terms of the location in the atmosphere and the degree to which the significance increases. Additionally, the results are not uniform across different data sets and so this analysis does not reveal any optimal lag. As such, the final "LOTUS regression" retains zero lag (see *Section 4.5*).

The two most commonly used solar proxies are the f10.7 and the Mg II proxies, with the f10.7 proxy being the most common. These two proxies yield nearly identical trend results but different uncertainty results.

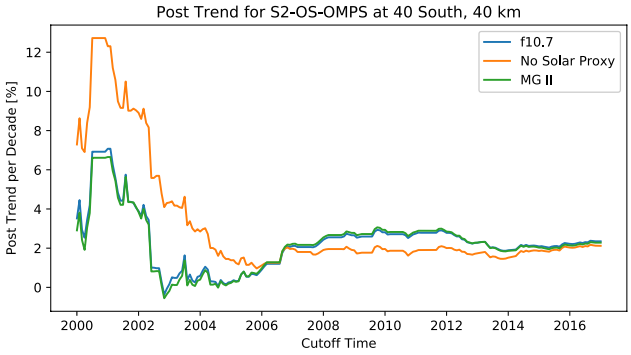


Figure 4.5: *Impact on potential recovery trends depending on the cutoff time of the regression for different solar proxies when applied to the SAGE-OSIRIS-OMPS data set centred at 40°S at 40 km. The impact on resulting trends of ignoring the solar proxy is evident for all but the longest data records while it is also apparent that the f10.7 and Mg II proxies have negligibly different impacts relative to each other.*

However, the changes in trend significance do not clearly indicate one proxy as better than the other and so the f10.7 proxy is retained as the baseline. Even though the different proxies do not yield different results, where any solar proxy shows its real influence is in the low frequency nature of the proxy. The solar cycle has a period of ~11 years, with an amplitude of influence on ozone variability of about 2 %. There is an expectation that proxies with longer periods can potentially have a greater influence on long-term trends in the regression process. This stems from the fact that resulting trends can be subject to endpoint anomalies if fitted over sufficiently short durations of data, particularly in the presence of other sources of variability with periods similar to the fitted duration. As such, the length of data used for the regression and the corresponding phase of the solar cycle at the endpoints can impact the trend results greatly. **Figure 4.5** shows the impact on potential recovery trends for different stopping dates. It is clear that not including the solar proxy is significantly different from including either the f10.7 or the Mg II proxies, though results start to converge with sufficiently long data records. It is also apparent that the length of the data record as it relates to the phase of the solar cycle is important, though the influence starts to level off to within 1 % per decade uncertainty after 2008 as the number of total solar cycles captured during the post-trend analysis increases.

The use of an aerosol proxy can be contentious. While it is well known that sulfate aerosols released from volcanic eruptions can influence stratospheric ozone through both chemical and dynamical effects, the exact relationship between volcanic aerosol and stratospheric ozone levels is not well characterised. Most regression analyses acknowledge the need to account for the El Chichón and Mount Pinatubo eruptions but do not agree on how this should be done. Some analyses simply ignore data immediately after the eruption (*e.g., Wang et al.,* 1996; *Randel and Wu,* 2007; *Harris et al.,* 2015), while others include a regression proxy of some form in an attempt to model the impact (*e.g., Bodeker et al.,* 1998; *Stolarski et al.,* 2006; *Bodeker et al.,* 2013; *Tummon et al.,* 2015). The net ozone response to aerosols depends on the ambient abundance of chlorine and dynamical conditions (*e.g., Tie and Brasseur,* 1995; *Aquila et al.,* 2013). To account for this, some studies include separate regression terms for eruptions of El Chichón and Mt. Pinatubo (*e.g., Stolarski et al.,* 2006; *Frith et al.,* 2014, *Weber et al.,* 2018). However, few aerosol proxies exist, though the most commonly used one is the NASA Goddard Institute for Space Studies (GISS) AOD proxy, which is what is tested here (**Figure 4.6**). The use of an aerosol proxy primarily influences the trend results in the lower stratosphere but only in the SAGE-OSIRIS-OMPS data set, suggesting this data may be more heavily influenced by aerosol interference. Some smaller, coherent patterns do appear in the middle to upper stratosphere, indicating a potential signal in these regions. The influence of adding an aerosol proxy on uncertainties is somewhat mixed (strong positive and negative deviations without a consistent pattern for all three analysed data sets), indicating that the use of this proxy needs further consideration to understand its impact. However, given the need to account for aerosol in some way and the desire not to simply omit data (as the period to omit is a question in itself), the GISS AOD proxy is included in the baseline (see *Section 4.2.1* for an explanation on how the GISS AOD proxy was extended beyond 2012, the last values reported for this proxy). We use a single AOD proxy as Mt. Pinatubo is the only significant eruption in the time period considered.

In general, varying the proxies applied in a regression model can affect the derived trends, though the effect can be mitigated by using data with a sufficiently long record. Additionally, the sensitivity of the trend to other proxies may vary with the resolution of the analysed data set and extent of spatial averaging. A recent study (*Zerefos et al.,* 2018) used 35 years of ozone data from the SBUV MOD data set evaluated both as zonal means and at select lidar station overpasses. As part of that study, the authors applied a similar regression model as that in LOTUS to

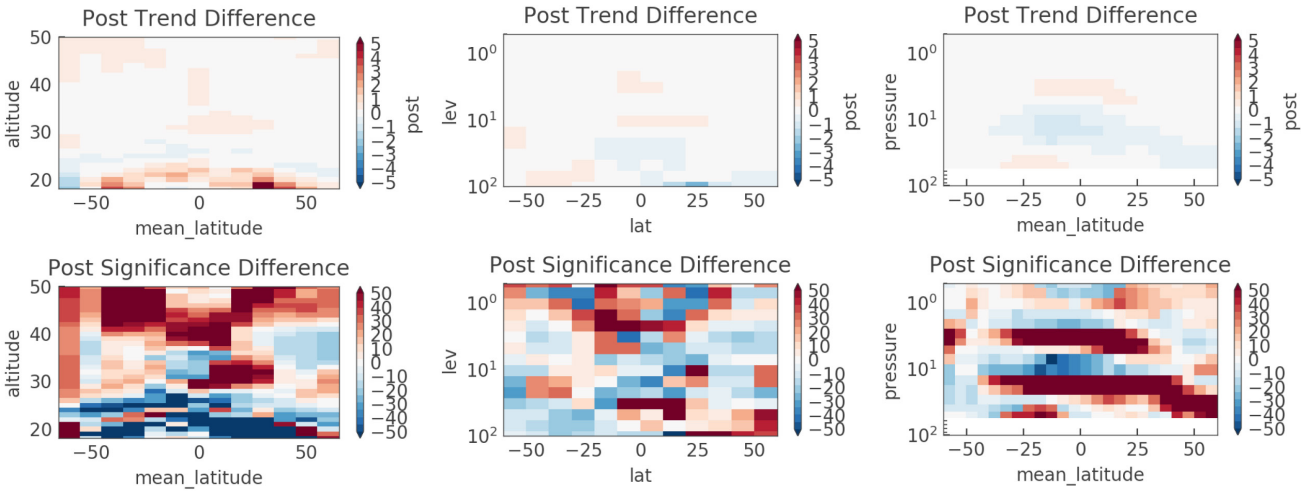


Figure 4.6: *Influence on the "Post-2000" trends (top row) and significances (bottom row) when adding the GISS aerosol proxy for the SAGE-OSIRIS-OMPS (left column), GOZCARDS (middle), and SBUV COH (right) data sets.*

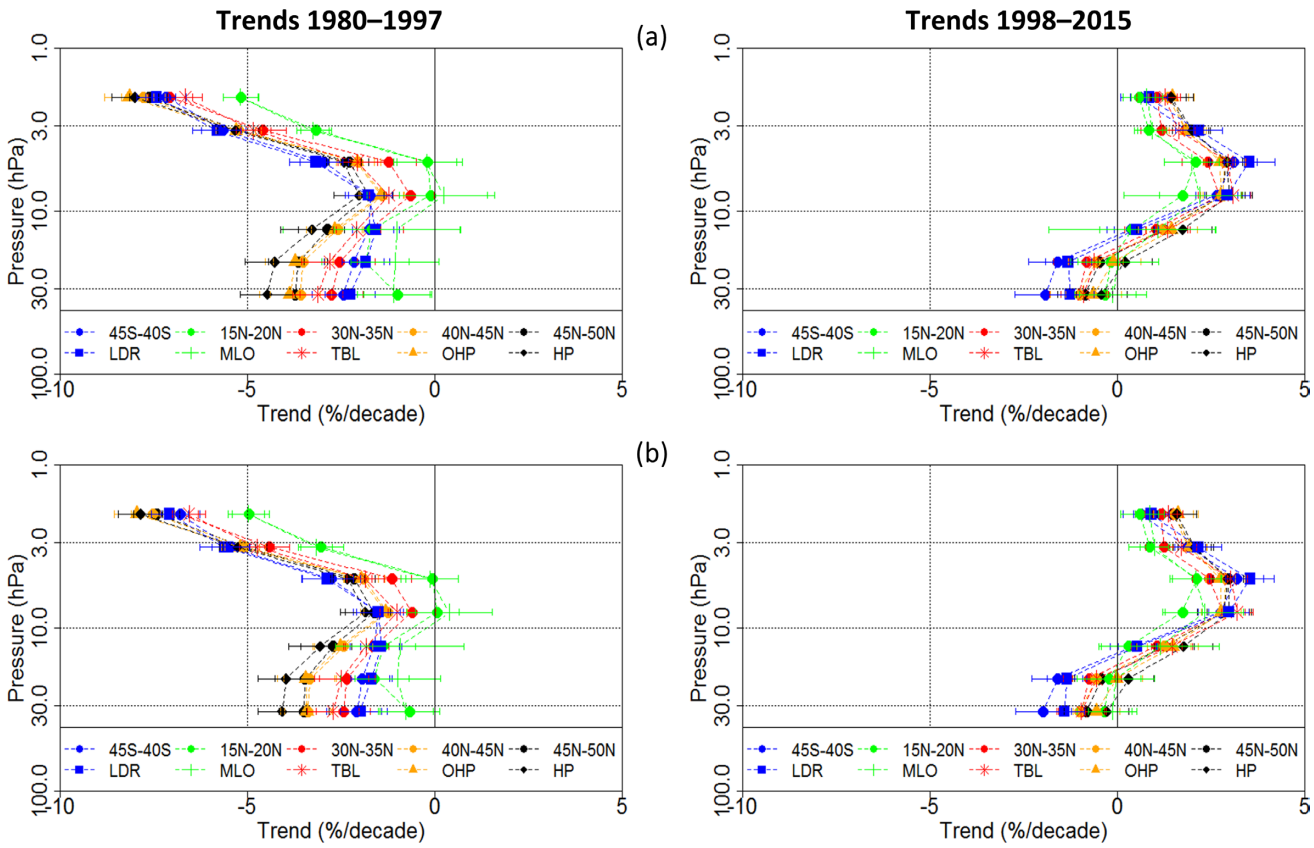


Figure 4.7: *Adapted from* ***Figure 7*** *of Zerefos et al. (2018) showing trends in the vertical distribution of ozone for the pre-1998 and post-1998 period, using (a) two linear trend terms (PWLT method) and volcanic effects and (b) the PWLT method including all proxies. The results are based on SBUV zonal means and SBUV overpasses over five lidar stations (LDR=Lauder, MLO=Mauna Loa Observatory, TBL=Table Mountain, OHP=Haute Provence, HP=Hohenpeissenberg).*

derive ozone trends and tested the effect of including and excluding almost all of the non-trend proxies at once. They found that a model using just the aerosol proxy and PWLT terms and one that also included all of the other non-trend proxies (*i.e.*, QBO, ENSO, AO/AAO, solar, and tropopause pressure) produced little difference in the resulting trend values and uncertainties when applied to SBUV MOD data with a 35 year duration (**Figure 4.7**). This is similar to the results of the sensitivity tests shown here, though we note the sensitivity of the trend is likely less in SBUV MOD due to the reduced vertical resolution of the data.

4.3.4 Trend proxy sensitivity

As detailed in *Section 4.2.2*, there have been four different proxies used to model the long-term trend in ozone for MLR analyses: A PWLT proxy, an ILT proxy, a single EESC proxy, and two EESC EOFs. The single EESC proxy represents the expected linear response of ozone to long-term variability in chemically reactive halogens. The two EESC EOFs are also meant to simulate the chemical forcing of ozone, but the extra degree of freedom allows for non-linearity in the ozone variations due to an imperfectly prescribed EESC shape (*i.e.*, incorrect age of air). The PWLT and ILT are less constrained and structured to better conform to the mean changes in the data from all long-term effects. However, the observational data are not yet sufficient to distinguish changes due to halogen chemistry from those due to other long-term variations induced by increasing GHGs. This means that in regions where chemical forcing is dominant and ozone responds directly to halogen levels (*e.g.*, the upper stratosphere), the EESC-based proxies are a better choice, but in regions where the effects of GHGs are dominant and the correlation between EESC and ozone degrades (*e.g.*, apparent monotonically decreasing ozone trends in the lower stratosphere despite decreasing EESC), the PWLT and ILT proxies are a better choice. While the focus of this Report is on the net changes in ozone from all long-term forcings, analyses of differences among the various trend proxies, in conjunction with longer data records, should allow for better attribution in future studies.

To determine the proxy that best represents observed ozone changes it is necessary to explore the strengths and weaknesses of each by focusing on their respective impacts on derived trends and uncertainties. Most notably, variability in the potential turnaround time can be problematic particularly when combined with nonlinear ozone changes. Stratospheric ozone levels decreased from the earliest satellite observations and this decrease appeared to abate over time.

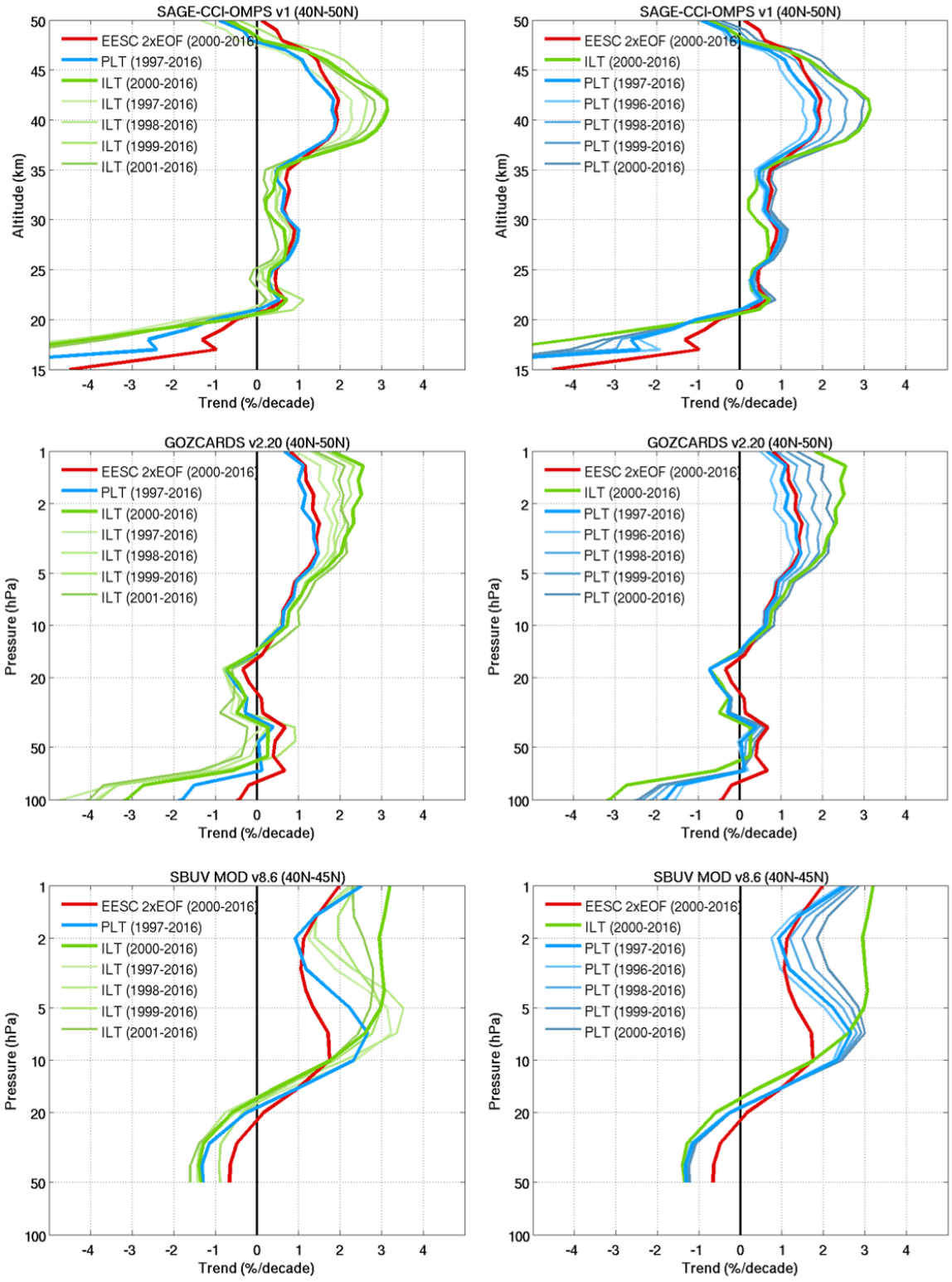


Figure 4.8: *Effect of time period start and end points on ozone trends in the past two decades obtained by regressions with PWLT and ILT.*

MLR Proxy Term	Allows for Curvature?	Allows for Variable Tornaround Time?	Allows for Monotonic Trends?
PWLT	No	No	Yes
ILT	No	Yes	Yes
Single EECS	Yes	No	No
Two EESC EOFs	Yes	Yes	Yes

Table 4.1: *Summary of the pros and cons of the different-long-term-ozone-trend proxies.*

Sensitivity tests have shown an optimal turnaround time of 1997 for the upper stratosphere (*Kyrölä et al.,* 2013; *Laine et al.,* 2014), but this likely changes with latitude and altitude and can be difficult to determine. However, the need for a predetermined turnaround time is the primary problem with using a PWLT proxy for linear regression models as choosing the wrong turnaround time will cause endpoint anomalies in the trend results, particularly in the presence of significant curvature in the actual long-term variability. **Figure 4.8** shows trend profiles derived from GOZCARDS and SBUV MOD where the end points for the PWLT and ILT proxies have been varied. The changes in trends that can be seen with changing time periods are a result of the endpoint anomalies due to the curvature of the long-term ozone variability and unexplained variability in the record. For example, the larger variations in ILT trends in SBUV MOD likely result from a known discontinuity in the data in late 2000. The ILT fits are more sensitive to this discontinuity than the PWLT fits because the endpoints of the PWLT fits are more constrained. Similarly, a single EESC proxy has the same problem as the PWLT (*i.e.*, the turnaround time varies with the mean age-of-air and must be predetermined), except it also cannot account for monotonic trends (*i.e.*, no turnaround in ozone) resulting from radiative and dynamical forcings. Instead, using a single EESC proxy in the presence of monotonic trends (*e.g.*, those seemingly present in the tropical lower stratosphere) will yield biased trend results (*Kuttippurath et al.,* 2015). Thus, a single EESC should never be used to represent the net long-term variability in ozone but only as a tool for determining chemical attribution (provided the correct age of air is known). Avoiding this particular pitfall, the ILT and two EESC EOFs allow for a variable turnaround time albeit in different ways. The ILT accomplishes this by avoiding fitting any trend term during a particular time period (*e.g.*, between 1997 and 2000), instead assuming the data to be constant during this period but allowing for shifts in the regression. This allows for two separate trends to be fit while attempting to avoid endpoint anomalies near the turnaround. The EESC EOFs can actually recreate the variable turnaround time (or lack thereof) and variations in curvature near the turnaround and can potentially allow for an independent assessment of what that turnaround time is. All told, the strengths and weaknesses of the different long-term variability proxies are summarised in **Table 4.1**.

Given the potential problems highlighted above, three different trend proxies were tested as part of a sensitivity study to examine their influence on derived trend results. For the sake of historical comparison, the PWLT proxy with a global turnaround at the beginning of 1997 was used for the baseline model. For the sake of comparison with the WMO Ozone Assessment (*WMO,* 2014), the ILT proxy with the declining period ending at the end of 1996 (hereafter called 'pre-1997' period) and the potential recovery period starting at the beginning of 2000 (hereafter called 'post-2000' period) were used. Lastly, the two EESC EOFs were also used and the resulting trends were determined as described in *Damadeo et al.* (2018) by extracting the combined EESC component of the fit and thereafter fitting a straight line to it over the given time periods to determine the mean trend during those times. Trends were evaluated pre-1997 and post-2000 for each of the three tests and the results are shown for three different data sets in **Figure 4.9**, **Figure 4.10**, and **Figure 4.11**. Not surprisingly, the general pattern of trend results is not that different between the three proxies. The pre-1997 trends are all about -8 % per decade in the upper stratosphere extra-tropics and the post-2000 trends are all about +2–3 % per decade in the same region. Trends in the middle stratosphere are generally about -2 % per decade in the pre-1997 period and about +0–1 % per decade in the post-2000 period. Additionally, all of the trends are noisy (*i.e.*, statistically insignificant) in the UTLS due to lack of high precision data in that region. There are, however, some subtle differences between the proxies. In the pre-1997 trends, upper stratospheric values are smallest (*i.e.*, least negative) in the ILT case and largest (*i.e.*, most negative) in the EESC EOFs case. Trend values are opposite in the post-2000 time frame; they are largest in the ILT case and smallest in the EESC EOFs case.

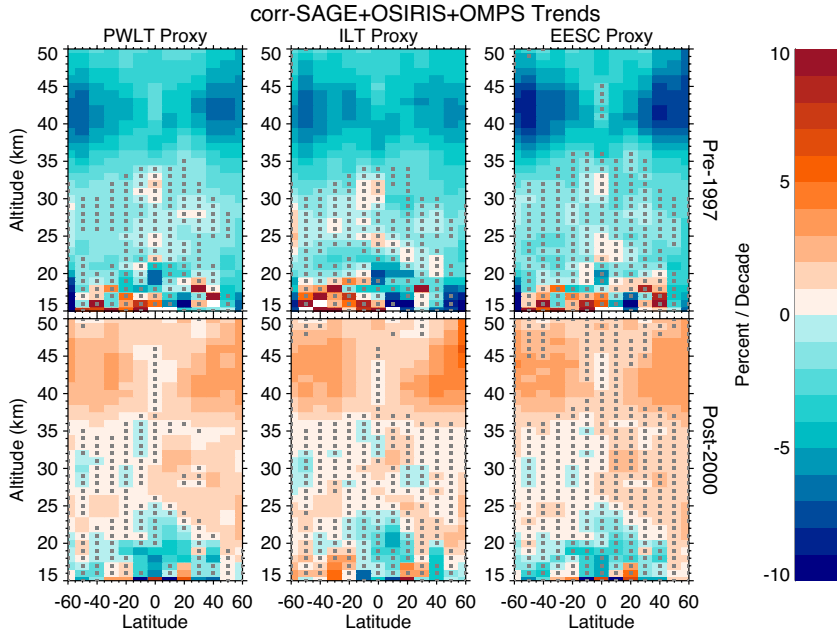


Figure 4.9: *Derived trends in ozone in percent per decade for the SAGE II-OSIRIS-OMPS data set (using the sampling bias adjusted SAGE II data from Damadeo et al., 2018) for both the pre-1997 (start of 1985 to end of 1996, top row) and post-2000 (start of 2000 to end of 2016, bottom row) time periods. Results are shown for each of the three trend proxies: The PWLT (left), ILT (middle), and EESC EOFs (right) proxies. Stippling denotes results that are not statistically significant at the 2-sigma level.*

In the later analyses, where trends are derived in broadband latitude ranges and merged, it made sense for the sake of brevity to pick a single MLR trend proxy term to use. The inability of the single EESC proxy term to capture variability beyond halogen chemistry automatically disqualifies it for use. Similarly, the sensitivity of the PWLT proxy term to the turnaround time suggests it is not ideal for use either. While both the ILT and EESC EOFs are acceptable, the desire to both investigate only the mean trends (*i.e.*, ignore the potential lack of direct correlation between the actual long-term ozone variability and ODSs) and have a more direct analog to compare with results from the last ozone assessment led us to choose the ILT proxy term for the work performed in *Chapter 5*.

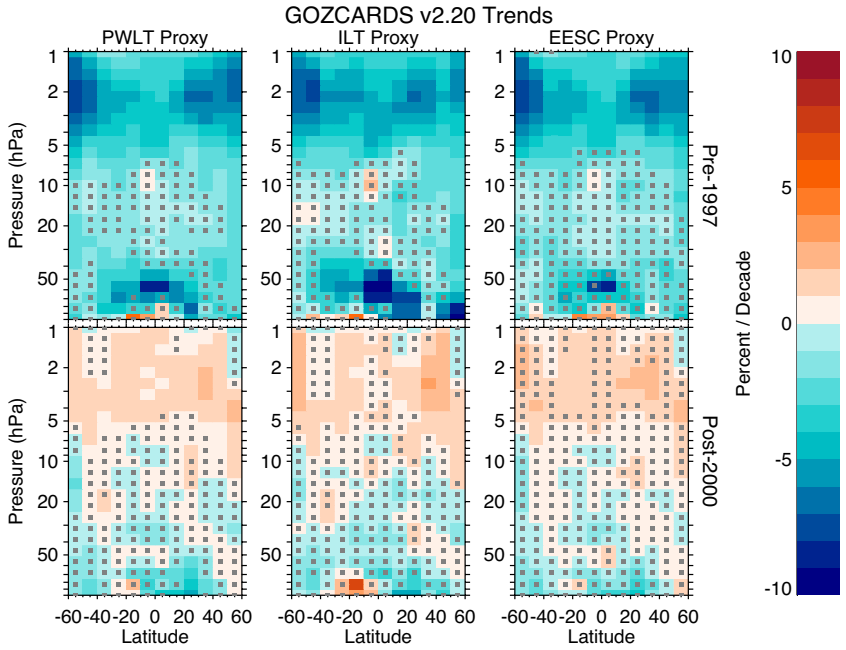


Figure 4.10: *As Figure 4.9 but for the GOZCARDS data set.*

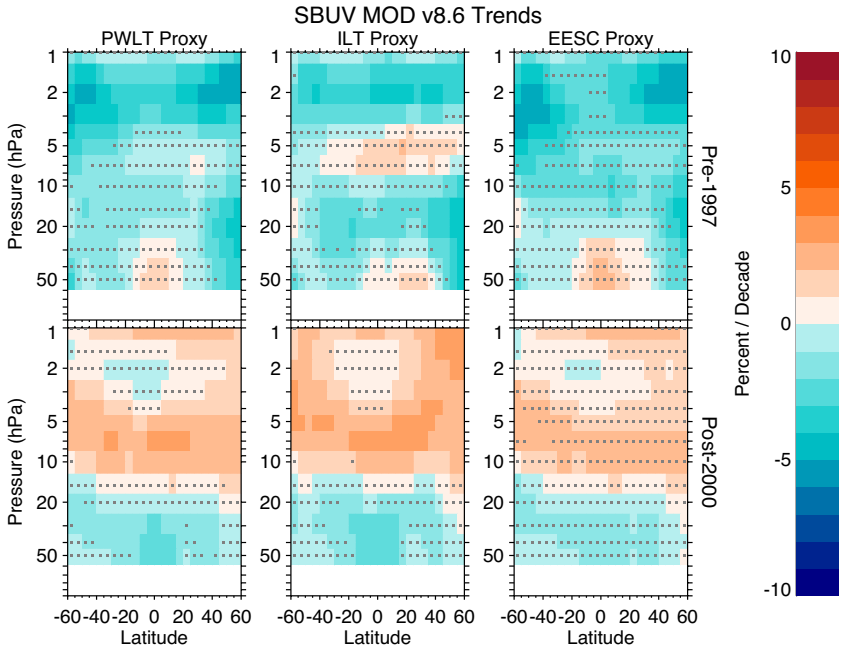


Figure 4.11: *As Figure 4.9 but for the SBUV MOD data set.*

4.4 Alternative approaches

Approaches other than MLR have been used in the community to quantify ozone changes over time. One of them is DLM (*Laine et al.,* 2014*; Ball et al.,* 2017, 2018). The regressors used for DLM are similar, though not identical, to those used in the "LOTUS regression" model but were kept identical to the analysis of *Ball et al.* (2017, 2018). The regressors include: A solar proxy (30 cm radio flux), a volcanic proxy (latitude dependent surface area density (SAD), based on *Thomason et al.*, 2018), two QBO proxies (30 hPa and 50 hPa wind fields as provided by the Freie University Berlin), and an ENSO proxy (Nino 3.4 HadSST). Seasonal cycle components, AR2 processes, and residuals are estimated together with these regressors, as well as the non-linear background trend. This non-linear background trend replaces the use of ILT, PWLT, or EESC and does not require an assumption about inflection dates, only a prior assumption about the smoothness of the non-linear background changes being estimated, which is determined from the data itself (see *Laine et al.* (2014); for further details, and *Ball et al.* (2017) for minor changes to the DLM algorithm used here). Because the background changes are non-linear, quoting a percent per decade trend is not appropriate with DLM, so nominally the overall change between two chosen dates is quoted (see for example **Figures S4.1**, **S4.2**, and **S4.3**), although the inferred non-linear background trends provide richer information about the long-term changes than a linear trend or a net change between two dates.

Another alternative to MLR is the application of a wavelet transform (WT). This method is widely used to analyse time series that contain non-stationary power at different frequencies and has been used more and more in geophysical and climatological studies (*Zitto et al.,* 2016). It allows an analysis which provides information not only on the frequencies present in the time series but also the times when the different frequency ranges are present in the sample. WT shows less sensitivity than PWLT to the choice of inflection point and therefore represents a promising alternative to MLR.

The method of empirical mode decomposition (EMD) is also well suited to deal with evolving trends over time (*Bai et al.,* 2017) and therefore as an alternative to the MLR approach. EMD decomposes any given signal into a finite number of intrinsic mode functions (IMFs), that represent simple oscillatory modes with varying frequency and amplitude along the time horizon, and a residual. This residual is then a monotonic or curved time series out of which the "trend" can be extracted. EMD is therefore not dependent on the length of data records for determining trends and is less vulnerable to outliers in the time series than MLR (*Bai et al.,* 2017).

4.5 The "LOTUS regression" model

4.5.1 General description

Based on the findings of the sensitivity tests presented in the earlier sections of this chapter, the LOTUS community agreed on one common regression model (the "LOTUS regression" model) to be used for all analyses that are presented in *Chapter 5*. The final choice of proxies and possible lags of proxies was based on finding the optimal regression for global analysis of satellite data and broad latitude band analyses. Therefore, proxies describing rather local or small-scale phenomena might not have been included in the general "LOTUS regression" model. To facilitate the comparison between satellite-based and ground-based ozone trends, the same regression model was applied to the ground-based (station) data, although proxies describing local and small-scale ozone variability might have improved the overall regression performance. Effects of this limitation in proxies for station data still need to be investigated further.

The "LOTUS regression" is applied to the ozone values without weights and so a correction for heteroscedasticity (*i.e.*, the non-constant variance in the data; see *Section 4.3.2*) is not applied. For data sets that are not already deseasonalised, Fourier components representing the seasonal cycle are also included (four sine and cosine pairs, representing the 12, 6, 4, and 3 month periodicity). No seasonal cross-terms are included in the "LOTUS regression" model as to mitigate the introduction of multicollinearity and to avoid inconsistencies between the treatment of data sets that are and are not intrinsically deseasonalised. A lag-1 autocorrelation correction was included in the regression model.

The "LOTUS regression" model uses the ILT proxy as a trend term (see *Section 4.2.2*). Additionally, it includes two orthogonal components of the QBO, the solar 10.7 cm flux, ENSO without any lag applied, and the GISS AOD. This aerosol data set was extended past 2012 by repeating the final available value from 2012 as the background AOD. To perform the ILT in a single step, the trend proxies included are:

- A linear increase until January 1997 and zero afterwards
- Zero until January 2000 and a linear increase afterwards
- Constant until January 1997 and zero afterwards
- Zero until January 2000 and a constant afterwards
- Constant between January 1997 and January 2000 and zero elsewhere

Including these proxies allows the ILT to be performed in a single step rather than the two step procedure used in *Steinbrecht et al.* (2017). The result of the regression is to obtain the coefficients A-J that correspond to the equation:

$$\begin{aligned} y(t) = {} & A \cdot QBO_1(t) + B \cdot QBO_2(t) + C \cdot ENSO(t) \\ & + D \cdot AOD(t) + E \cdot Solar(t) + F \cdot Linear_{pre}(t) \\ & + G \cdot Linear_{post}(t) + H \cdot C_1(t) + I \cdot C_2(t) + J \cdot C_3(t) \\ & + \varepsilon(t) \end{aligned} \quad (4.5),$$

where C_1 to C_3 are the three constant terms described above.

The "LOTUS regression" model has been implemented in the Python programming language. It is designed to be a flexible software package to both perform the sensitivity tests of *Section 4.3* and to run the final chosen models on the wide variety of data sets present within the LOTUS initiative. The software package and up to date documentation are available at **https://arg.usask.ca/docs/LOTUS_regression**.

4.5.2 Application to model simulations

In order to maintain comparability in the interpretation of results, we performed the analysis of trends in the vertical distribution of ozone from the CCMI-1 REF-C2 simulations (see *Chapter 2*, *Section 2.3*) using the same ILT method as for the observations. Linear trends for the pre-1997 (Jan 1985 – Dec 1996) and post-2000 (Jan 2000 – Dec 2016) periods were calculated at each grid point (*i.e.*, latitude and pressure level) of the models (and the separate ensemble simulation members). Since the models' simulated atmospheric conditions (and composition) differ from the observations, we calculated the appropriate proxies (predictors) that are included in the statistical trend analysis using model parameters. Thus, for each model/ensemble member we first calculated the QBO index, performing an EOF analysis on the simulated zonal winds at the equatorial region. Then we used the first two EOF terms as QBO1 and QBO2 indices. The ENSO index was calculated from the simulations' SSTs over the tropical Pacific, over the exact same area where the Nino3.4 index is calculated. As before, the data were deseasonalised over the period of 1998–2008.

Finally, the regression analysis was performed at the given pressure levels, using the following proxies: (1) two trend terms (identical to the method described earlier as the ILT method, *Section 4.2.2* and *Section 4.3.4)*; (2) two QBO terms (calculated as described above), in the case of models not simulating the QBO this proxy was not used; (3) one term for the ENSO effect (as described above); (4) one term for the solar forcing (we used the forcing as it was provided to the modelling groups; it should be noted that in our case the last five years of the solar forcing data are slightly different from the observations, not in terms of phase but in magnitude); and (5) one term for the volcanic effect (AOD; the same basis function that was used for the analysis of the satellite- and ground-based measurements was also used for the models). Normally the CCMI-1 REF-C2 model simulations do not include volcanic eruptions, but the effects can be present via different routes, for example SSTs or winds. All model results that are shown in *Chapter 5* (*Sections 5.2* and *5.5*) are shown as percent changes over the base period 1998–2008.

4.6 Summary

This chapter discusses several sources of uncertainties and sensitivities for trend analyses with MLR. First we tested existing regression models within the community by applying them to a common data record and comparing the resulting trend estimates. We found a general spread in derived trend values of 1–2 % per decade, with some differences as high as 3 % per decade. Next, we completed a series of sensitivity tests in an effort to identify the proxies that have the largest effect on the derived trend, leading to the observed spread in trend results across different regression models. Several sensitivity tests for proxy selection, proxy combination, and unweighted/weighted regression approaches were performed to understand their effects on the derived trend values and trend uncertainties in multiple merged satellite ozone records. We found the proxies AO, AAO, NAO, and EHF have only negligible effects on trends and significances, but excluding the QBO, solar, or ENSO proxies from the regression model had significant effects on the trend (1–2 % per decade difference) and uncertainty (around 1 % per decade) estimates. The three different trend proxies (PWLT, ILT, and EESC-based EOFs) produce generally very similar trend estimates. However, in the sensitivity tests performed here, subtle differences for the results of the trend proxies were found; for the pre-1997 trend estimates ILT produces the smallest (least negative) trend and EESC the largest trend, whereas for the post-2000 trend estimates the trend proxies behave exactly opposite. PWLT trends were shown to be affected most by end point problems caused by the chosen inflection point (and therefore the length of the analysed time series).

Based on these sensitivity tests, a "LOTUS regression" model was developed that includes two QBO proxies, a solar proxy, an ENSO proxy without any time lag applied, a stratospheric aerosol proxy, and the ILT as the trend proxy. Four Fourier components representing the seasonal cycle are also included. The "LOTUS regression" model is unweighted, and it includes a lag-1 autocorrelation correction. A detailed description of this regression model and its source code is publicly available on the LOTUS website, **https://arg.usask.ca/docs/LOTUS_regression**.

Chapter 5: Time series and trend results

The ultimate goal of LOTUS is to improve confidence in calculated ozone trend values via an improved understanding of the uncertainties. *Chapter 3* highlighted many of the challenges facing analyses of long-term ozone time series, and despite the fact that many of those challenges still need to be addressed, it is worthwhile to assess the trend results from this work in such a way as to be able to place those in the context of previous work. This chapter highlights the results of taking the "LOTUS regression" model from *Chapter 4* and applying it to the different data sets (*i.e.*, satellite, ground, and model) at different resolutions comparable to those in previous ozone assessments and comprehensive studies (*e.g.*, *WMO,* 2014; *Harris et al.,* 2015; *Steinbrecht et al.,* 2017). The individual satellite-based trend results are then combined to obtain a single mean ozone trend profile with respective uncertainty estimates. This important yet challenging final step in the assessment has been the cause of debate in the community in recent years. Different methods for combining the individual trend results are discussed and explained, and the final trend profile estimates are analysed for significance.

5.1 Satellite trends at native resolution

The regression model was applied to all satellite data sets described in *Chapter 2*, for all latitude bands and all vertical levels. In this section, only eight of the ten satellite data sets are discussed. The SAGE-OSIRIS-OMPS data set without the sampling-corrected SAGE data is excluded for reasons discussed in *Section 3.2.1.* Also, for reasons summarised in *Section 4.5.1*, *Chapter 5* only discusses results for the ILT "LOTUS regression" model. However, for the sake of completeness, the trend results for all ten data sets and with each trend proxy (*i.e.*, PWLT, ILT, and EESC EOFs) are shown in the *Supplement* (see **Figures S5.1** through **S5.6**).

5.1.1 Trend results

Figure 5.1 shows the trends derived for the pre-1997 period (this period covers January 1985 to December 1996; for simplicity it is called 'pre-1997'), for all latitude bands, all vertical levels, and all satellite data sets as well as for the CCMI-1 REF-C2 model results for comparison purposes. The general pattern of negative (4–9 % per decade) trends in the upper stratosphere (above 5 hPa / 35 km) is present in most of the satellite records (with SBUV MOD showing the smallest and SWOOSH showing the largest trends) and is consistent with previous findings (*i.e.*, *WMO,* 2014; *Harris et al.,* 2015; *Steinbrecht et al.,* 2017; and references therein). These upper stratospheric trends show a minor hemispheric asymmetry in several of the data sets, with larger (*i.e.*, more negative) trends in the Southern Hemisphere (SH) mid-latitudes than the Northern Hemisphere (NH) mid-latitudes. Pre-1997 trends in most records show positive values at 7–10 hPa (or 30–33 km) in the tropics, with slight differences between the trends derived in the SH and the NH for some data sets, though these are not statistically significant. Also, slightly positive trends, which are also not statistically significant, are found in all data sets in the lower stratosphere in NH subtropics (30–50 hPa or 20–25 km). SBUV COH is the only data set with a significant positive trend in the SH at mid-latitudes around 20 hPa (this result and the less negative SBUV MOD trend noted above are discussed in *Section 5.1.2*; see also **Figure S5.8**).

The time period of most interest is after 2000, when an ozone recovery is expected at upper stratospheric levels. **Figure 5.2** summarises the analyses for the post-2000 period (this period covers January 2000 to December 2016, or in the case of BASIC January 2000 to December 2015 (see *Chapter S4.1* in the Supplementary Material for more detail), and it is called 'post-2000' for simplicity). Overall, the results from the different data sets seem to agree on positive trends (~2–3 % per decade on average) in the mid-latitudes of both hemispheres between 5 hPa and 2 hPa (around 37 km to 45 km). As in the pre-1997 period, most of these upper stratospheric trends show a minor hemispheric asymmetry, with more positive trends in the NH than in the SH. Some asymmetry is also found in middle and lower stratospheric trends across the 20°S–20°N tropical band, including a more pronounced difference between the subtropical SH and NH. This trend asymmetry and its impact on the combined broad-band trends is noted in *Sofieva et al.* (2017), *Steinbrecht et al.* (2017), and *Zerefos et al.* (2018), and is further discussed in *Section 5.4* with respect to the representativeness of the ground-based records. However, the magnitude of these trends and their significances vary between the data sets. Out of the eight data sets, three have OMPS measurements contributing to the merged record beyond 2012 (**Table 2.2**), though the magnitude of the trend results is influenced largely by the record that is used between 2000 and 2005. The OSIRIS-based data sets produce larger recovery trends than the MIPAS-based data set, which may be related to the positive drift of OSIRIS in the upper stratosphere (see discussion of the record stability in *Section 3.1.2*). SBUV COH also appears to produce more significant recovery estimates for the upper stratosphere as compared to the SBUV MOD record (see also next section). In the middle to the lower stratosphere, SBUV-based data sets and the BASIC data set show more negative trends across all latitudes compared to GOZCARDS or SWOOSH, though the vertical resolution of SBUV is significantly reduced in this altitude range.

Additionally, SWOOSH limits negative trends to the tropical region only. OSIRIS-, CCI-, and MIPAS-based data sets show smaller negative trends in the lower stratosphere that are not clearly confined to a specific region and, in most cases, are not significant.

5.1.2 Discussion of differences

There are several possible reasons for the different trend patterns of the eight satellite data sets shown in **Figure 5.1** and **Figure 5.2**. In the following sections we discuss some of the reasons that could contribute to the trend differences based on findings from *Chapter 3* and relevant literature, such as the merging method of the data sets, application of a sampling bias correction, and conversions between different unit and grid systems to allow the construction of merged data sets.

Merging method

As detailed in *Sections 2.2.1* and *3.1.4*, although the SBUV MOD and SBUV COH are constructed from the same suite of instruments, different choices of instruments and different merging techniques allow for different, but equally valid, merged records. In the pre-1997 period, the nadir-based trends vary significantly from the limb-based trends and from each other. Data quality of SBUV instruments operating in the mid-1990s is negatively affected by on-orbit instrument problems and drifting orbits, resulting in larger intra-satellite biases and drifts during this period. SBUV MOD and SBUV COH both rely on the same instrument during the 1980s but diverge in the 1990s (see for example **Figure 5.5**, 20 hPa). The sensitivity to the mid-1990s data is enhanced when fitting over the shorter 1985–1996 time period. Furthermore, data after the eruption of Mt. Pinatubo (mid-1991–1993) are treated differently in the two merged records (*Section 2.2.1*). Estimates by *Frith et al.* (2017) suggest uncertainties of 10–15 % (2-sigma) from 1995–2000 in the merged records, suggesting trends fit over this time period for either SBUV merged record are highly uncertain. However, trends from SBUV COH and SBUV MOD computed using the "LOTUS regression" model over the longer 1979–1996 time period compare notably better than those computed over the 1985–1996 period (**Figure S5.8** in the

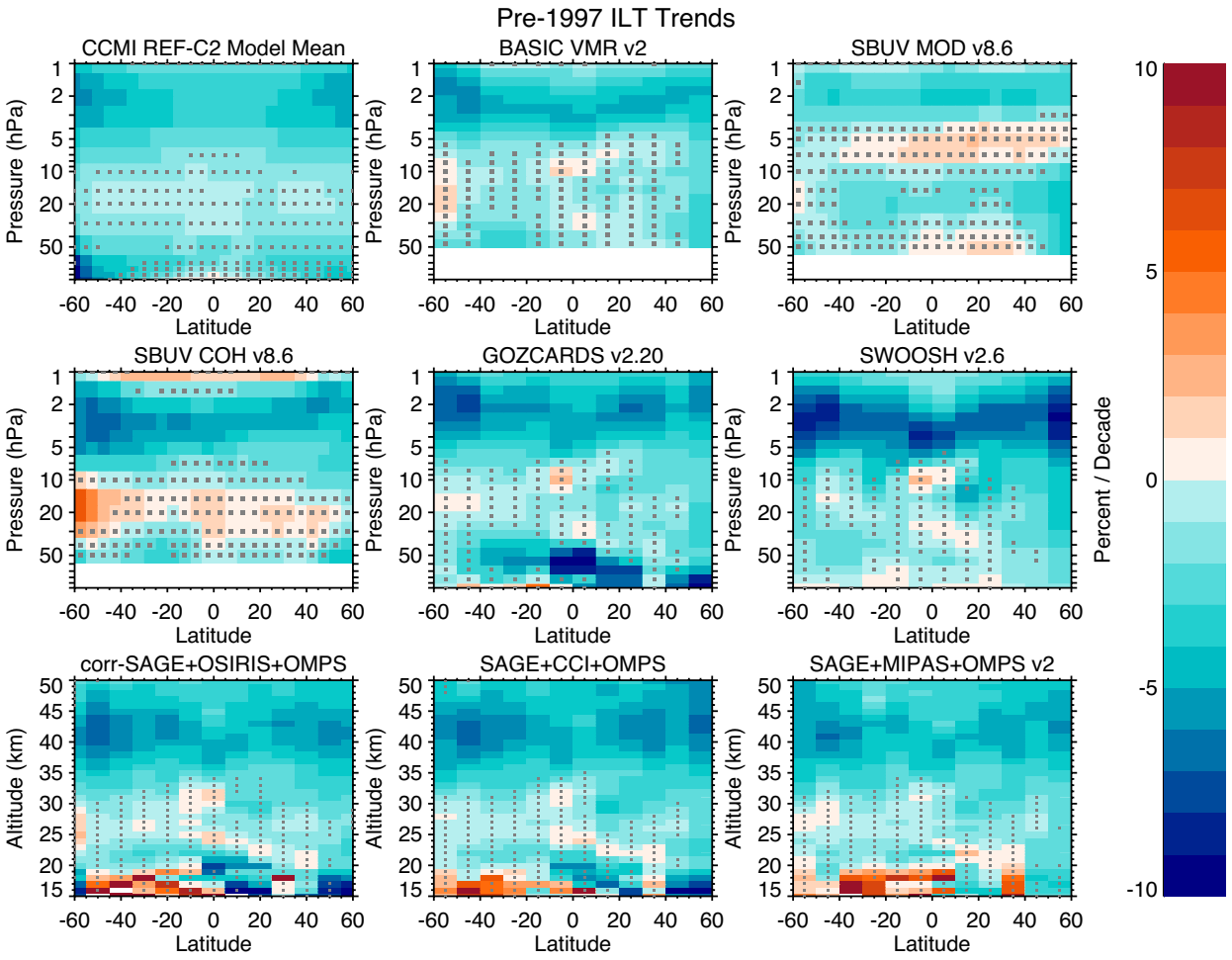


Figure 5.1: *Derived trends in satellite ozone in percent per decade for the pre-1997 period (Jan 1985 – Dec 1996) for each of the satellite data sets, using the ILT trend proxy in a regression analysis. Grey stippling denotes results that are not significant at the 2-sigma level. Data are presented on their natural latitudinal grid and vertical coordinate. For comparison, the mean of trends derived from CCMI participating models is included in the upper left panel. Results for other trend proxies can be found in the Supplement.*

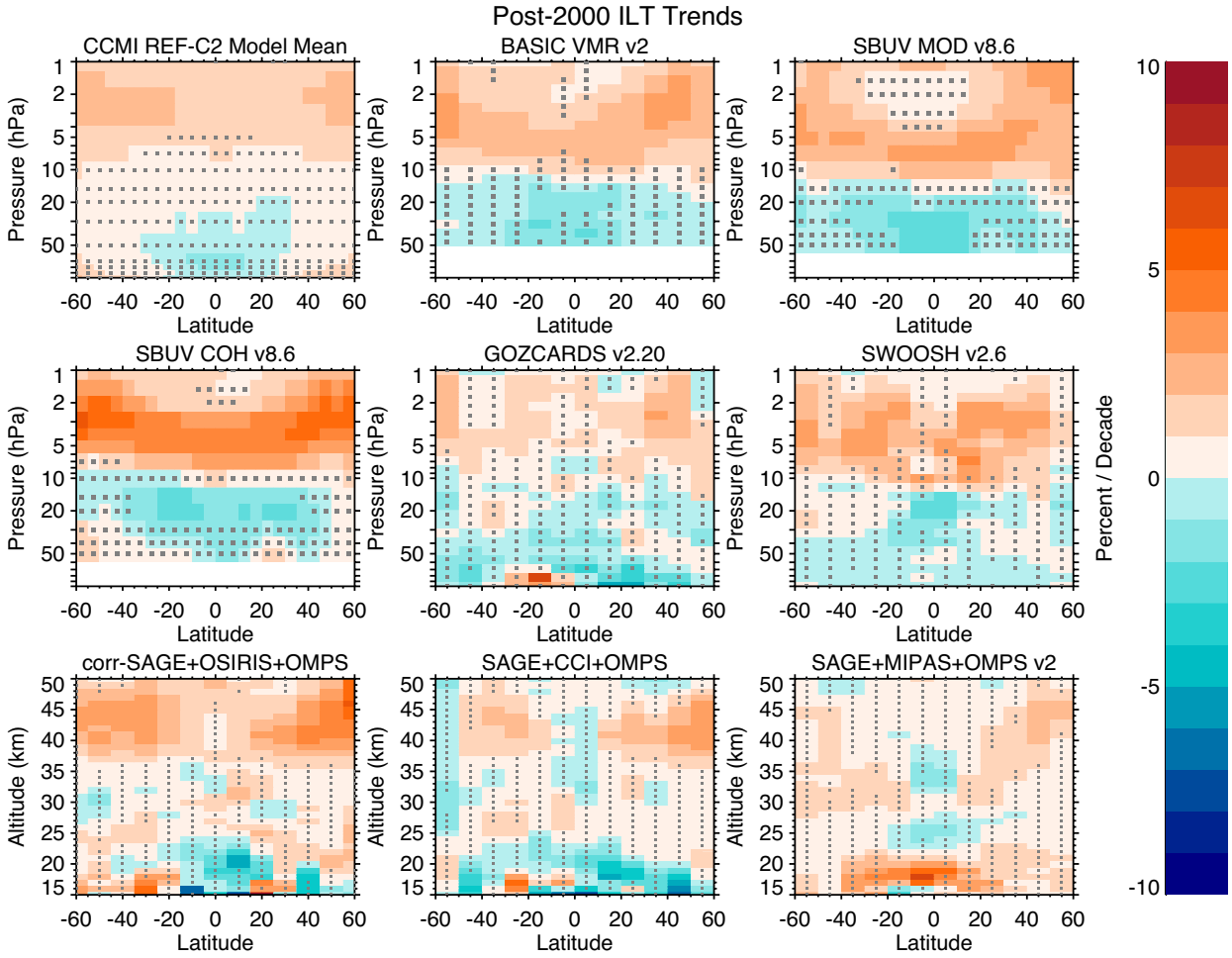


Figure 5.2: *Derived trends in satellite ozone in percent per decade for the post-2000 period (Jan 2000 – Dec 2016) for each of the satellite data sets, using the ILT trend proxy in a regression analysis. Grey stippling denotes results that are not significant at the 2-sigma level. Data are presented on their natural latitudinal grid and vertical coordinate. For comparison, the mean of trends derived from CCMI participating models is included in the upper left panel. Results for other trend proxies can be found in the Supplement.*

Supplement). Trends from both records also compare well with trends from GOZCARDS fit over the same time period (**Figure S5.8**, see also *Harris et al.*, 2015). Trends over the post-2000 period tend to be positive at NH and SH mid-latitudes between 5 hPa and 2 hPa. SBUV MOD trends are more positive at altitudes just above the 10 hPa level, particularly in the tropics, while the SBUV COH trends are more positive between 5–3 hPa at all latitudes. Because of their respective merging techniques, SBUV MOD is sensitive to successively increasing or decreasing biases in the instrument series that might alias into a trend while SBUV COH is sensitive to drifts in the reference instruments that might be propagated to other periods in the record. As discussed in *Section 3.1.4*, the higher trends in SBUV MOD are related to a positive bias in NOAA-19 data at the end of the record, while the higher trends in SBUV COH above 5 hPa are partially caused by a small drift in the NOAA-18 data used as a reference for the merged data set. *Frith et al.* (2017) used a MC approach to estimate a merging uncertainty for SBUV MOD and showed that SBUV MOD and SBUV COH trends agree within the estimated 2-sigma merging uncertainty (see **Figure 3.9**).

When combining measurements from a multitude of different instrument types (such as in GOZCARDS and SWOOSH), sources of potential differences in trends derived from these data sets increase even more. As shown by *Tummon et al.* (2015), although GOZCARDS and SWOOSH are based on almost the same data sources, differences exist for example in the annual cycle of the tropical stratosphere and in monthly mean anomalies of the SH mid-latitude stratosphere. These differences are caused by different merging approaches for the selected data sets and can have an impact on the estimated trends (for more information see *Tummon et al.*, 2015; *Harris et al.*, 2015; and *Section 2.3.3*). Both data sets, GOZCARDS and SWOOSH, have been updated since the analyses described in *Tummon et al.* (2015) and *Harris et al.* (2015), but the differences in merging technique remain and with that the possible effects on trend estimates.

Small differences in pre-1997 trends from the OSIRIS-based data set in comparison to the CCI-based data set are largely consistent with expected differences due to the use of sampling corrected SAGE data in the OSIRIS-based record (see next section).

However, small remaining differences, as well as larger differences with the MIPAS-based record (which also includes only SAGE data in the pre-1997 period), suggest differences in the latter part of the record are due to the merging technique, and data set selection can impact the pre-1997 trends through fits to other large-scale proxies such as the solar cycle. Differences up to 2 % per decade found in the post-2000 trends in the upper stratosphere are likely influenced by differences in the CCI and OSIRIS records and the respective merging methods, as the same version of OMPS-LP data is used at the end of the combined records. Upper stratospheric ozone from the OMPS-LP record compares well with Aura MLS over mid-latitudes (see **Figure 3.3** in *Chapter 3*) but shows differences over tropical regions (see **Figure 3.4** in *Chapter 3*). Also OMPS-LP and Aura MLS differences are found over the middle stratosphere in the tropics, which likely explains the difference in tropical trends derived from SAGE-CCI-OMPS and SAGE-OSIRIS-OMPS as compared to GOZCARDS and SWOOSH. Drift in the OMPS-LP record relative to lidar, microwave, and ozonesonde records (see **Figure 3.5** in *Chapter 3*) can also explain the stronger positive trends in the upper stratosphere in the combined datasets that incorporate the OMPS-LP record.

The significant differences in derived trends between SBUV and BASIC despite the inclusion of SBUV data in BASIC is likely a result of the merging technique applied to create the BASIC data set, which reduces the weight of SBUV measurements over this time period due to known instrument issues.

Sampling bias correction

The difference in evaluated trends due to the sampling bias correction of SAGE II data can be as large as 2 % per decade for the pre-1997 period and up to 1 % per decade for the post-2000 period, as illustrated in **Figure 5.3**. While the absolute differences in trends are larger for the pre-1997 period, which are derived from satellite data with coarse spatio-temporal sampling, the influence is more pronounced for the post-2000 trends as the trend differences represent a larger fraction of the trends themselves. For example, the largest impact in pre-1997 trends is found in the upper stratosphere, with typical differences of ~1 % per decade over the mid-latitudes (compared to the -6 % per decade trend), with stronger increases in trends over the NH. However, the post-2000 trends show mixed/positive differences up to 1 % per decade in the SH/NH respectively that represent as much as half of the trends there.

Conversion between different unit and grid systems

Stratospheric cooling was observed in the upper stratosphere in the 1980s and beginning of the 1990s (*Steinbrecht et al.,* 2009; *Randel et al.,* 2009), which led to a difference of 2 % per decade in trends depending on the measurement unit and vertical scale (*McLinden et al.*, 2011; also see **Box 2.1** in *Chapter 3*, *WMO,* 2014). *McLinden et al.* (2011, **Figure 3**) showed that trends derived from ozone records prior to 1997 differed, such that VMR trends computed on pressure surfaces (*i.e.*, GOZCARDS, SWOOSH, SBUV, BASIC) were found to be less negative than number density trends computed on altitude surfaces (*i.e.*, CCI or OSIRIS-based combined records; see also *Harris et al.,* 2015). These differences exist only in the presence of temperature trends due to continuous changes in conversion between pressure and altitude scales. Since the late 1990s, only small cooling trends are observed in the upper stratosphere (*e.g.*, *Thompson et al.,* 2012; *Randel et al.,* 2016), and thus the influence of unit and vertical coordinate representation on post-2000 ozone trends is expected to be small. In addition, even though all the pre-1997 limb instrument trends are based primarily on SAGE II data, the differences in trends between the different coordinate representations in the merged limb records do not match expectation (**Figures 5.1** and **5.6**). For example, the pre-1997 trend in NH for GOZCARDS (VMR on pressure) is close to that of SAGE-CCI-OMPS and SAGE-OSIRIS-OMPS (number density on altitude) but differs significantly from the SWOOSH (VMR on pressure) trend. This suggests that the influence of temperature changes on the agreement of ozone trends derived in different unit/coordinate representations is either smaller than expected or obscured by other sources of uncertainty.

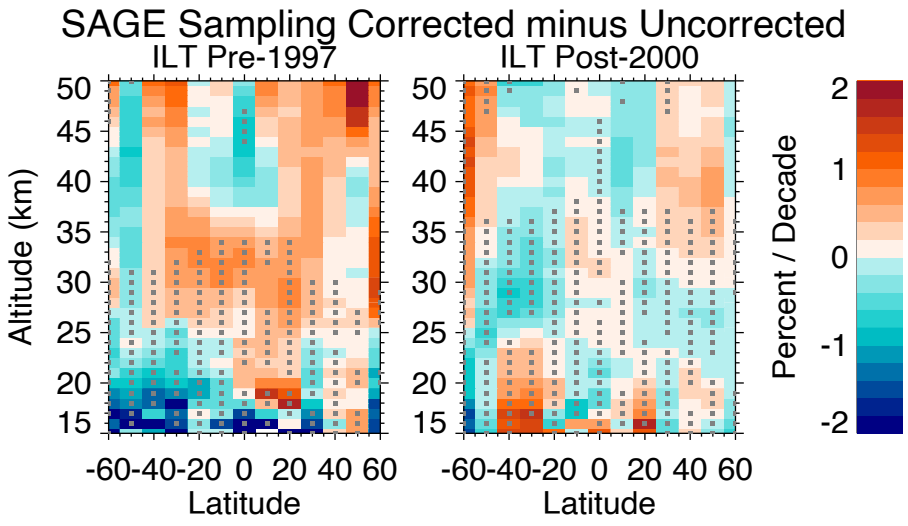


Figure 5.3: *Impact of sampling bias correction on ozone trends derived from the SAGE-ORIRIS-OMPS data set. Grey stippling denotes where the trends derived from either version of data (i.e., sampling-corrected or not) were not significant at the 2-sigmal level.*

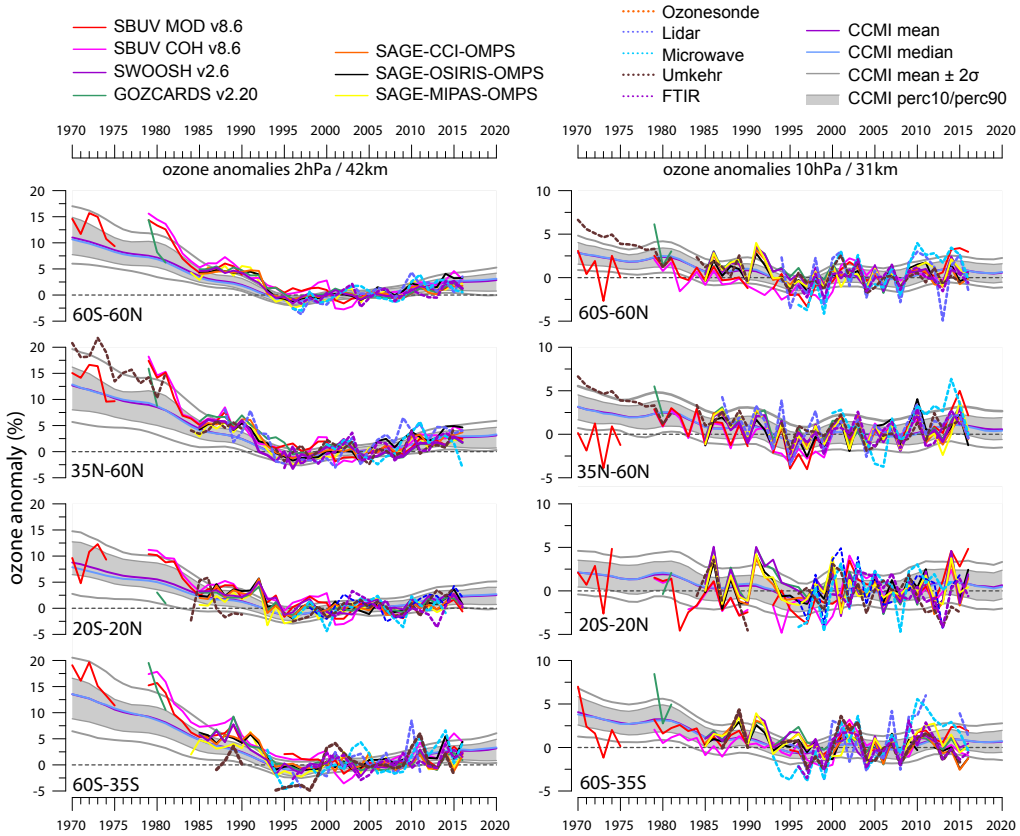


Figure 5.4: *The evolution of ozone changes as annual mean anomalies at the 2 hPa/42 km (left panel) and 10 hPa/31 km (right panel) levels. Three different latitude bands are shown. Satellite data are based on zonal means, and ground-based stations are averaged over the latitude bands. The grey "envelope" gives the CCMI-1 model results, based on the models' 10th and 90th percentile. The model mean and the median are also plotted together with the ±2 standard deviation range of the models. All anomalies are calculated over the base period 1998–2008, and the CCMI-1 models are shown as 1-2-1 year filtered averages (see text).*

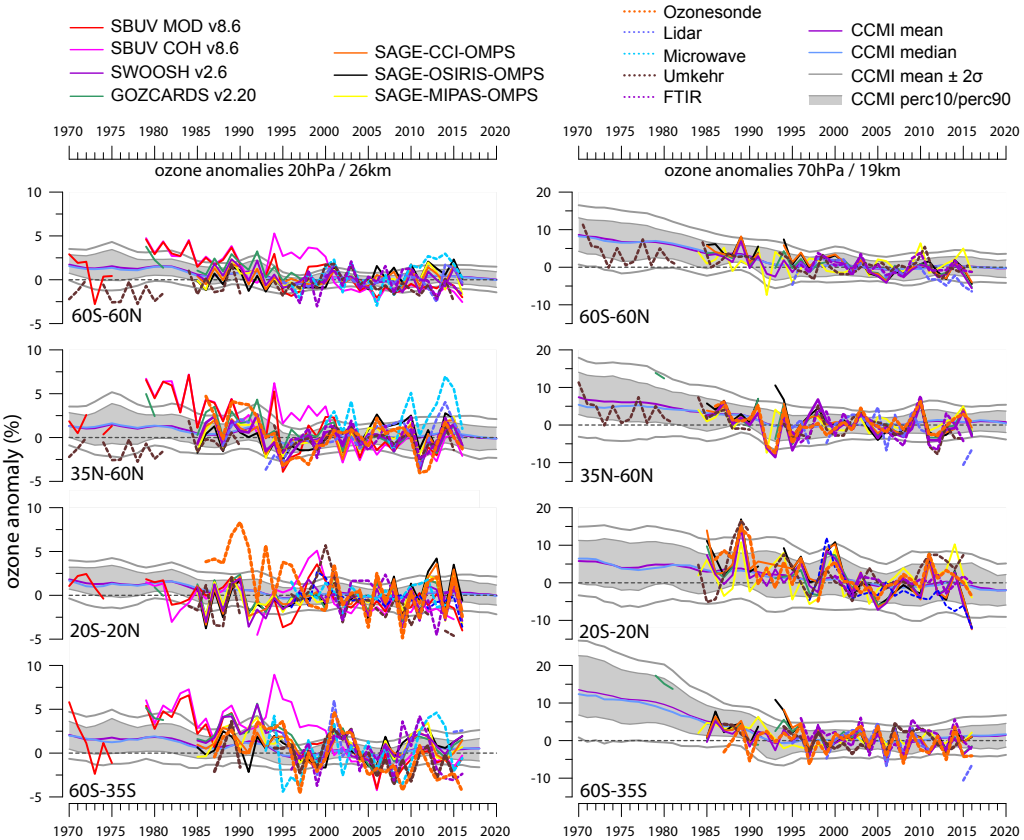


Figure 5.5: *Same as* ***Figure 5.4*** *but for the 20 hPa/26 km (left panel) and 70 hPa/19 km (right panel) levels.*

5.2 Time series in broad latitude bands

While much of the literature has analysed stratospheric ozone trends from individual data sets at their native resolution, the more expansive works (*e.g.*, *WMO,* 2014; *Harris et al.,* 2015; *Steinbrecht et al.,* 2017) have investigated trends in broad latitude bands. The three cited studies selected three broad bands: the SH at mid-latitudes (60°S–35°S), the tropics (20°S–20°N), and the NH at mid-latitudes (35°N–60°N). To place the results of LOTUS in the context of those works, the updated/modified data sets have been analysed in these broad latitude bands. An average of near global ozone profiles (60°S–60°N) have also been reported before; however, these are not the main focus here and will therefore only be shown in the *Supplement* (**Figure S5.7**).

The data sets are not natively provided for these broad latitude bands, so they must first be converted (for further details see *Sections 2.1.2*, *2.2.5*, and *2.3.2*). **Figures 5.4** and **5.5** illustrate annual mean deseasonalised anomaly time series for satellite, ground-based and model data for the three broad bands at four selected altitude or pressure levels (indicated at the top of each panel). These annual means are shown here only for illustration purposes; the regression results in *Section 5.3* are based on monthly mean anomaly data. The CCMI-1 model time series were further smoothed using a 1-2-1-year filter (described in *Section 2.3.2*). The model median (blue curve) and the model mean (dark pink) are shown with the lower 10th-percentile and the upper 90th-percentile forming the grey envelope of the CCMI-1 models' range. Grey curves surrounding the time series denote the range of CCMI-1 model mean ±2 standard deviations. Time series constructed from the satellite data and ground-based station data show interannual variability caused by natural variations such as volcanic eruptions and QBO (most pronounced in the plots for the middle and lower stratosphere), whereas in model results natural variability is smoothed out for plotting purposes (see *Section 2.3.2* for details).

Figure 5.4 shows the evolution of ozone changes in the upper stratosphere (2 hPa/42 km, left panel). The ozone decline due to the increase in ODSs up until the mid-1990s is evident, as expected. Since then, a leveling off can be seen, with the ozone values after about 2005 indicating an increase, which is most prominent in the NH mid-latitude band (35°N–60°N). Lower levels are presented in **Figure 5.4** (right panel) and **Figure 5.5**. In general, the ODS-related long-term evolution of ozone can be seen at these levels but is less pronounced than in the upper stratosphere.

The REF-C2 simulations by CCMI-1 models are capable of capturing the long-term variations in ozone, that is the solar cycle and the trends. The mean and median of the CCMI-1 REF-C2 multi-model simulations are smoother and their uncertainty ranges narrower than the anomalies calculated from observations. This is expected because the REF-C2 simulations do not include volcanic variability, and variability from forcings such as QBO and ENSO will tend to cancel in the average as their

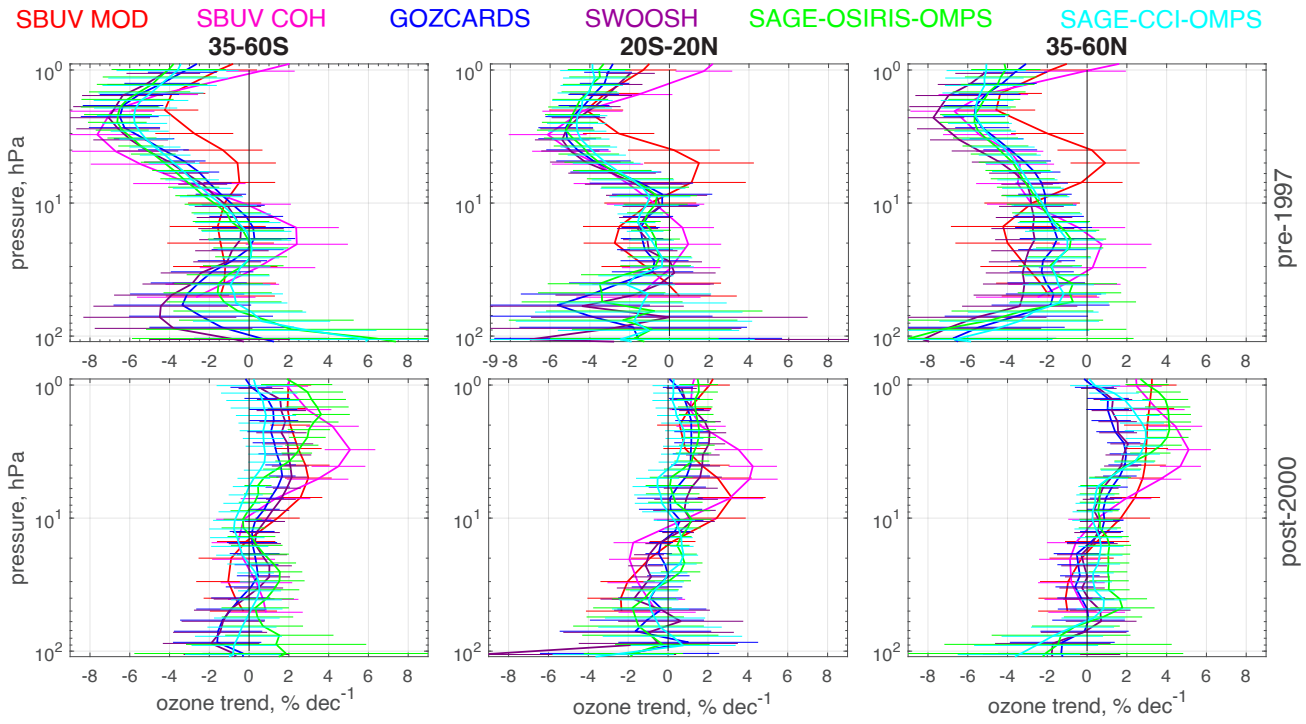


Figure 5.6: *Ozone trends with 2-sigma uncertainties for the pre-1997 (top) and post-2000 (bottom) period from the ILT regression for latitude bands 60°S–35°S (left), 20°S–20°N (centre), and 35°N–60°N. Coloured lines are the trend estimates from individual merged data sets on their original vertical grid.*

phases will vary among individual simulations. A year-to-year direct comparison between models and observational data with respect to the natural variability can therefore not be made here. Making such a comparison would require model simulations that are tied to real-world observations, such as simulations using specified dynamics or the REF-C1 runs, which were based on observed SSTs and aerosol loading. Though these runs exist, the output does not cover the entire time period analysed by LOTUS (1985–2016). The model mean only represents the range of ozone variability due to the longer term evolution of ODSs and GHGs. The lack of volcanic eruptions and other natural forcings such as solar and QBO is also the main difference between this work and the relevant figures in the 2014 WMO Ozone Assessment (*WMO,* 2014). The grey shading given by the CCMVal-2 models in the lower stratosphere shown in **Figures 2-7** and **2-8** of the 2014 WMO Ozone Assessment (*WMO*, 2014) is wider than the 10/90 percentile range presented in **Figure 5.4** and **Figure 5.5**, since the CCMVal2-based large model variability is caused mainly by volcanic eruptions.

Overall both models and observations (satellite and ground-based) follow the same evolution of ozone changes in time and at the various altitude/latitude bands. The relative differences between individual observational time series are larger during the earlier years and at the 20 hPa level. However, the ground-based measurements during the period prior to 1980 are represented only by a single Umkehr record at Arosa, Switzerland and thus may not be representative of the broad-band variability captured by the Nimbus-4 satellite. In addition, after mid-1972 the Nimbus-4 BUV data coverage was reduced due to instrument problems, adding significant noise to the record. The bi-annual variability in the observed ozone anomaly time series is associated with the QBO signal.

5.3 Combined satellite trends in broad latitude bands

5.3.1 Selection and preparation of data sets

For the evaluation of ozone trends in the three broad latitude bands (*i.e.*, 60°S–35°S, 20°S–20°N, and 35°N–60°N) we consider six of the eight merged data sets discussed in *Section 5.1*: SBUV MOD, SBUV COH, GOZCARDS v2.20, SWOOSH v2.6, corr-SAGE-OSIRIS-OMPS, and SAGE-CCI-OMPS. The SAGE-MIPAS-OMPS data set is now excluded because of concerns of larger discontinuities when switching between the instruments (see *Section 3.1.5*). This collection of data sets includes two merged data sets for each satellite data set group discussed in *Chapter 2* (*i.e.*, nadir mixing ratio profiles versus pressure, limb mixing ratio profiles versus pressure, and limb number density profiles versus altitude) and was chosen such that they are as independent as possible. For this reason, the BASIC data set, which combines four of the merged data sets in the above list (GOZCARDS, SWOOSH, SBUV MOD, and SBUV COH) is not used.

Broad latitude band trend results for individual merged data sets are derived by applying the ILT "LOTUS regression" to the relative deseasonalised ozone anomaly time series averaged over the three latitude bands as discussed in *Section 2.2.5*. However, these broad-band trends are in the native coordinate system and combining the trends requires the profiles to be expressed in the same vertical coordinate and at the same grid levels. The reference vertical scale used below is the pressure grid of GOZCARDS and SWOOSH. The SAGE-OSIRIS-OMPS and SAGE-CCI-OMPS altitude levels are first converted to pressure levels using the mean ERA-Interim altitude-pressure profile in the considered broad latitude bands and time period, and the resulting trend profiles are subsequently linearly interpolated to the reference grid. The SBUV data, on the other hand, are already on a pressure grid but one that is coarser than that of GOZCARDS, so the SBUV trend profiles are linearly interpolated to the finer reference grid.

Figure 5.6 shows the profile trends in broad latitude bands for each of the six selected data sets. Similar to **Figures 5.1** and **5.2**, the trends in **Figure 5.6** generally agree with each other with differences caused by the reasons discussed in *Section 5.1.2*. The 2-sigma error bars represent the trend uncertainty estimated by the regression model (using the fit residuals). These are very similar for limb data sets and are slightly larger for the nadir merged data sets.

5.3.2 Approach to combine trends

In order to facilitate comparison with prior studies or with model simulations, it is useful to combine the trend profiles, and their uncertainties, derived from the various observational records. Such an approach has been used in previous WMO Ozone Assessments (*e.g.*, *WMO,* 2014) and it provides an estimate of how confident we are in trend results from the global satellite observing system. Ideally, ozone time series are combined using the averaged anomaly from all available records such that a single trend can be derived (similar to, for example, the BASIC approach). However, this approach is time consuming and requires consideration of differences in the individual records pertaining to temporal and spatial sampling, stability, vertical coordinate system, different units, and vertical smoothing. Therefore, the most efficient approach for this Report was to calculate trends separately for each record and then combine those results. As mentioned in *Chapter 3*, tracking ozone recovery with multiple observations is important for redundancy, avoiding gaps in satellite operations, and avoiding impacts of drifts of individual records on the trend assessment.

While combining trends is somewhat straightforward (*i.e.*, usually computed as either an unweighted or a weighted mean), combining uncertainties is much more complicated.

One of the main challenges is to assess the independence of the trend results given that some data records (*e.g.*, SAGE II) are used multiple times and that one regression model and one set of proxies are used for all analyses. When considering the independence of the trends and their uncertainties, it is useful to consider the theoretical lower bound to the uncertainty of the regressed trend, which is defined not only by the length of the time series and the magnitude of the trend but also by the atmospheric variability not characterised by the regression model. Indeed, any variance that is not represented in the regression model will propagate into the fit residuals, which will, in turn, lead to larger uncertainties in the regression coefficients. Hence, even trends derived from an ideal data set (*i.e.*, one that is infinitely long with no sampling issues, drift, or need for merging) will have an uncertainty equal to this theoretical lower bound determined by the true, incompletely modeled atmospheric realisation. As a result, even the trend uncertainties of completely independent data sets are correlated when estimated with the same regression model.

We estimate the overall trend $\bar{t}$ as the unweighted mean of six trend estimates and evaluate its variance as the maximum of two variance terms:

$$\sigma_{mean}^2 = max\left(\frac{1}{N^2}\sum_{i,j} C_{ij}\sigma_i\sigma_j \;,\; \frac{1}{n_{eff}}\sum\frac{(x_i-\bar{x})^2}{N-1}\right) \quad (5.1),$$

where N is the number of observation records, C_{ij} are the correlation coefficients for the trend estimates x_i from data sets i and j , σ_i are the trend uncertainties estimated from the fit residuals for the individual data sets, and n_{eff} is the effective number of independent trend estimates. The first term in Equation (5.1), the quadratic form $\frac{1}{N^2}\Sigma_{i,j} C_{ij}\sigma_i\sigma_j$, is the variance of the mean of correlated values, obtained through traditional propagation of errors. The second term is the unbiased estimator of the standard error of the mean, where n_{eff} independent measurements are assumed from the N=6 different trend estimates, and can capture systematic biases in trend uncertainties between the different merged data sets that would not be captured by the first term (*e.g.*, as a result of random drifts between data sets or differing unit representations). The effective number of independent values n_{eff} in Equation (5.1) is approximated by

$$n_{eff} = \frac{N^2}{\sum_{i,j=1}^{N} C_{ij}} \quad (5.2).$$

The first term in Equation (5.1) serves as an approximation of the theoretical lower bound of trend uncertainty due to the actual realisation of the ozone time series. This special approach of using the maximum of both terms to estimate the combined trend uncertainty is done because the sample of trend estimates is small and because, in case the trend estimates within the sample coincide, the observed variance is not necessarily representative of the actual uncertainty of the combined trend. We do not use the sum of both terms in Equation (5.1) because the variance of the trend estimates in term 2 can be partly due to the uncertainties already represented in term 1, and thus using the sum would overestimate the uncertainty. However, in the case where some contribution to the standard error is independent of propagated uncertainties, the maximum of the two terms would underestimate the uncertainty. Realistically, separating out the overlap in the total uncertainty captured by each of the two terms in Equation (5.1) is impossible and, in practice, one of the two terms dominates such that the maximum is a reasonable approximation of the true uncertainty. **Figure S5.9** in the *Supplement* shows the uncertainty obtained by using both the max() and sum() approaches.

It is difficult to estimate the correlation of trend uncertainties as these correspond to the largest temporal scales in the ozone time series. In our analysis, we approximate the trend correlation by the correlation between the fit residual time series of the different regressions. These fit residuals include small-scale variations in the ozone time series from atmospheric variability that is not captured by the regression model (*e.g.*, from a sudden stratospheric warming event), which is likely to be highly correlated across all data sets, and any systematic biases in the data sets themselves (*e.g.*, from instrument behavioral anomalies), which can be highly correlated between merged data sets relying upon the same instrument. However, these correlations at short temporal scales do not necessarily translate to equivalent correlations at long temporal scales and so the correlations of the fit residuals likely overestimate the correlations of the trend uncertainties. Since the true correlations between the trend uncertainties do not have a straightforward solution, we use the correlations of the fit residuals as a conservative upper bound instead.

Detailed results of a correlation analysis of the fit residuals are presented in the *Supplement* (see *Appendix B* with **Figures S5.10** to **S5.15**). The matrices of the correlation coefficients and of the fit residuals for pre-1997 and post-2000 trend estimates are:

$$C_{pre}^{limb} = \begin{pmatrix} 1 & 0.8 & 0.4 & 0.4 & 0.4 & 0.4 \\ 0.8 & 1 & 0.4 & 0.4 & 0.4 & 0.4 \\ 0.4 & 0.4 & 1 & 0.9 & 0.6 & 0.7 \\ 0.4 & 0.4 & 0.9 & 1 & 0.6 & 0.6 \\ 0.4 & 0.4 & 0.6 & 0.6 & 1 & 0.8 \\ 0.4 & 0.4 & 0.7 & 0.6 & 0.8 & 1 \end{pmatrix}$$

$$C_{post}^{limb} = \begin{pmatrix} 1 & 0.9 & 0.6 & 0.6 & 0.6 & 0.6 \\ 0.9 & 1 & 0.6 & 0.6 & 0.6 & 0.6 \\ 0.6 & 0.6 & 1 & 0.95 & 0.6 & 0.65 \\ 0.6 & 0.6 & 0.95 & 1 & 0.65 & 0.7 \\ 0.6 & 0.6 & 0.6 & 0.65 & 1 & 0.82 \\ 0.6 & 0.6 & 0.65 & 0.7 & 0.82 & 1 \end{pmatrix} \quad (5.3).$$

The order of the data sets in these matrices is as follows: 1=SBUV MOD, 2=SBUV COH, 3=GOZCARDS, 4=SWOOSH, 5=SAGE-OSIRIS-OMPS, and 6=SAGE-CCI-OMPS. Using these correlation matrices and Equation (5.2) we obtain n_{eff}=1.6 for pre-1997 trends and n_{eff}=1.4 for post-2000 trends. Below ~50 hPa, where there are no SBUV data, n_{eff}=1.3 for both pre-1997 and post-2000 periods. The experimental correlation coefficients vary slightly with latitude and altitude (see *Supplement*) as a result of atmospheric variability, but these variations might not be

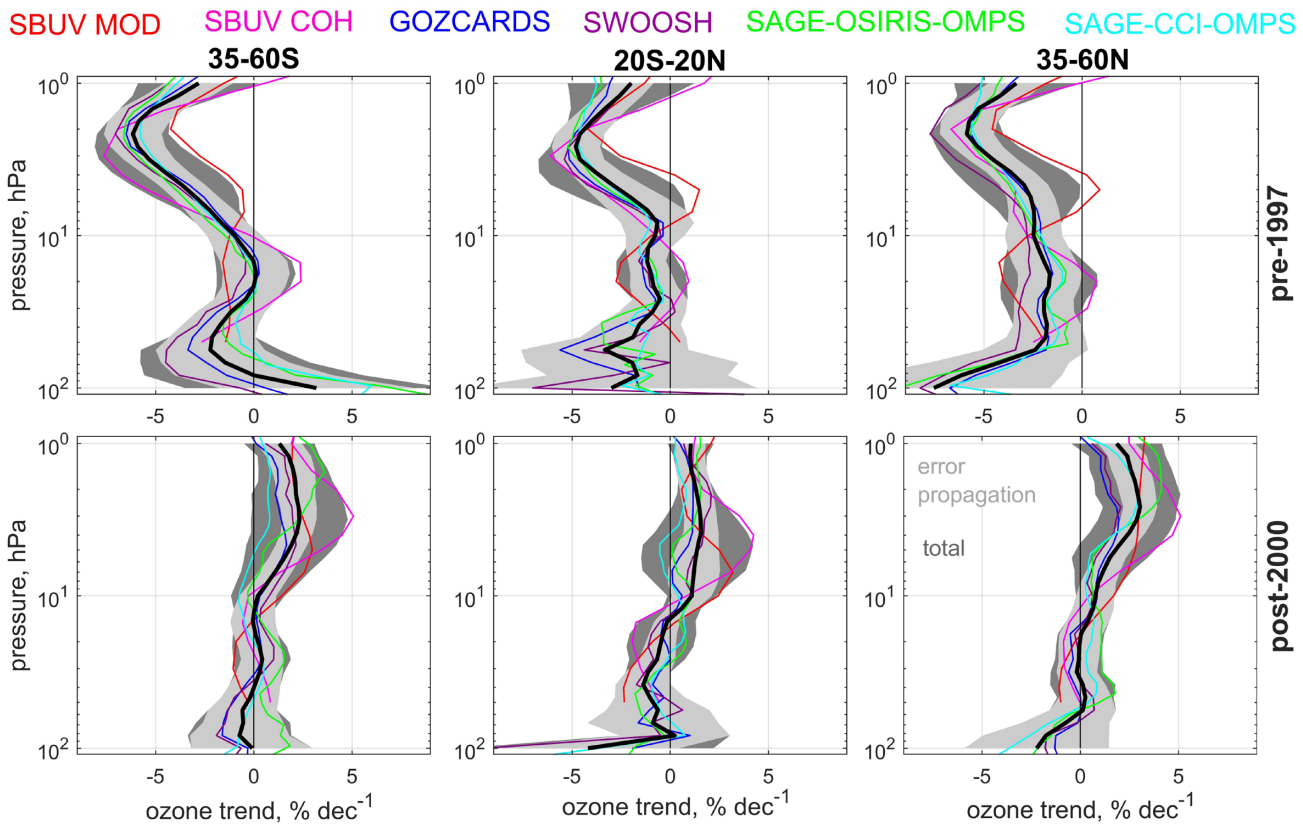


Figure 5.7: *Combining pre-1997 (top) and post-2000 (bottom) trend estimates and uncertainties (2-sigma) by Equation (5.1) from six limb profile data sets. Black solid line indicates the mean trend. The uncertainty component corresponding to error propagation (1st term in Equation (5.1)) is shown by light grey shading, while the total uncertainty is indicated by dark grey shading.*

related to the correlation of the trend estimates. Therefore, constant correlation coefficients are used for trend estimates. In any case, small variations in correlation coefficients result in insignificant changes in the effective number of independent trend values.

Figure 5.7 shows the mean overall trends and their uncertainties estimated using Equation (5.1). Light grey shading indicates the first term in Equation (5.1), which represents error propagation. In the lower stratosphere, this term generally dominates and represents the total uncertainty. In the upper stratosphere, the second term representing the variance of the sample mean dominates.

We think that Equation (5.1) overestimates the trend uncertainty for two reasons. First, the estimated trend correlations are influenced by substantial correlations at small scales, as discussed above. Second, the pre-1997 trends for SBUV would be in better agreement with limb data sets, and have smaller uncertainty estimates, if the time series from 1979 would be considered (see discussion in *Section 5.1.2* and **Figure S5.8** in the *Supplement*).

5.3.3 Alternative methods

In *WMO* (*2014*), the combined trend from several satellite records was evaluated by first adding an estimated mean drift of 2% per decade to each instrument's 2-sigma statistical uncertainty and then calculating the mean trend weighted by the total error of individual trends. The uncertainty of the combined trend was then computed by propagating these individual trend uncertainties through the weighted mean. This uncertainty estimate does not take into account the statistical dependence (*i.e.*, correlation) nor the spread between individual trend estimates. The latter is often considerable and larger than the systematic error assumed by *WMO* (2014), likely leading to an underestimate of the uncertainties.

Harris et al. (2015) used another method that combines the trend and uncertainty in a joint distribution (J-distribution; *SPARC,* 2013). The central value is computed as the arithmetic mean and its uncertainty is evaluated as

$$\sigma^2_{mean} = \frac{1}{N}\sum \sigma_i^2 + \sum \frac{(x_i - \bar{x})^2}{N-1} \qquad (5.4).$$

This J-distribution approach is appropriate to estimate the population variance based on several data sets (*e.g.*, *Sofieva et al.,* 2014). For the current application, the estimate by Equation (5.4) is quite conservative. First, it assumes that trend estimates from several merged data sets do not reduce the random uncertainty. Indeed, the first term in Equation (5.4) is the mean of individual uncertainty variances, which is equivalent to the (conservative) assumption that all individual trend estimates are fully correlated. Second, the difference between individual trend values is also due to their (random) uncertainties and thus the second term in Equation (5.4) also includes (at least partially) the first term.

Steinbrecht et al. (2017) used the arithmetic mean as the central value, like the J-distribution method. However, they suggest to estimate the variance of the combined trend as the variance of the sample mean of estimates that are not necessarily independent

$$\sigma^2_{mean} = \frac{1}{n_{eff}} \sum \frac{(x_i - \bar{x})^2}{N-1} \qquad (5.5),$$

where n_{eff} is the effective number of independent data sets. **Figure 5** of *Steinbrecht et al.* (2017) shows uncertainties for $n_{eff}=1$ (*i.e.*, all trend results assumed fully correlated), while their **Table 6** presents the uncertainty estimates for $n_{eff}=3$ at altitudes above the 50 hPa level and $n_{eff}=2$ at lower altitude levels. Only post-2000 trends are discussed in this work. This choice is motivated by the expectation that the following three groups of data sets are fairly independent in the past two decades: The SBUV records, those based on SAGE-MLS as the "backbone," and the rest relying on SAGE and another instrument. Pre-1997 trends in the limb group are highly correlated as they rely heavily on the SAGE II record (see correlations in Equation (5.3)), which would lower the effective degrees of freedom by one unit. **Figure 5.8** shows the uncertainties according to Equation (5.5) with $n_{eff}=2$ used for pre-1997 trends and $n_{eff}=3$ for post-2000 trends above the 50 hPa level (green line). Corresponding values at lower altitudes are reduced by one since no SBUV data is used in the lowermost stratosphere.

5.3.4 Discussion

Figure 5.8 compares the results from three methods to combine trend and trend uncertainty for the same six satellite trend profiles. The combined trend values, computed from the arithmetic mean, are by definition the same for each of the methods. However, differences in the combined trend uncertainties are considerable. Higher significances are found post-2000 with the method used by *Steinbrecht et al.* (2017) as a result of the assumption of 2–3 independent trend estimates (mainly affecting the upper stratosphere, see **Figure 5.7**) as well as the lack of a systematic lower bound to the uncertainties where the trend results coincide (*e.g.*, around 10 hPa and at lower altitudes). This is less of an issue in the pre-1997 period, when SBUV results diverge considerably, leading to uncertainties that are similar between the methods used by *Steinbrecht et al.* (2017) and LOTUS. The systematic inclusion of uncertainty from propagation of regression errors (J-distribution) likely overestimates the uncertainty during the ozone depletion period. As mentioned earlier, part of this error component is already present in the derived sample standard deviation.

The LOTUS and J-distribution methods have the uncertainty from error propagation as a lower bound. This limits the potential for large vertical variations in trend uncertainty, as regression errors usually have only a weak and smooth dependence with height. In most cases, however, it is the variance in the post-2000 trend sample that drives the uncertainty of the combined result, and this has a clear vertical structure as well. For all methods there is a localised large increase in uncertainty around 5 hPa in the tropics. We also note gradual increases in post-2000 trend uncertainty with decreasing pressure (higher altitudes) in the mid-latitude upper stratosphere. **Figure S5.9** in the *Supplement* shows 1-sigma uncertainty profiles of **Figure 5.8** but also adds one more curve using the sum() alternative of Equation 5.1 rather than the max(), for illustrative purposes.

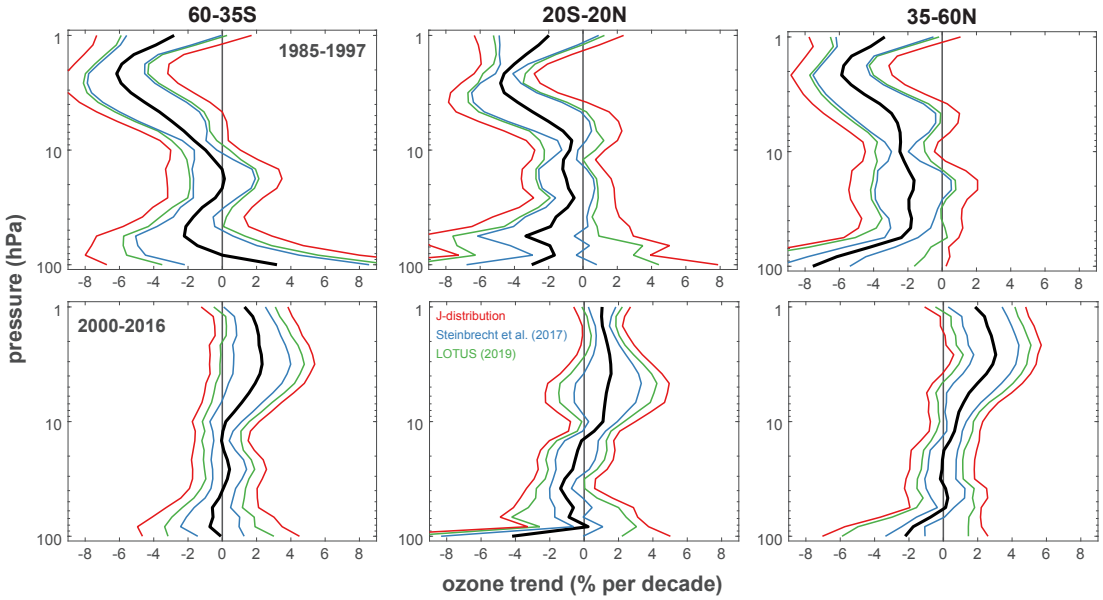


Figure 5.8: *Combined trends of six limb and nadir data records (thick black lines) with 2-sigma uncertainties (thin coloured lines) for the pre-1997 period (top) and the post-2000 period (bottom) from the ILT regression for latitude bands 60°S–35°S (left), 20°S–20°N (centre), and 35°N–60°N. Green lines denote results from the LOTUS method (Equation 5.1), blue lines denote results from the method of Steinbrecht et al. (2017; Equation 5.5) with $n_{eff}=2$ for pre-1997 and $n_{eff}=3$ for post-2000 trends, and red lines denote the results from the J-distribution method (Equation 5.4).*

The results presented here indicate that the uncertainty of the combined trend is sensitive to the method used. Each method is based on fair assumptions as the combined result is limited to the small number of trend realisations, and trend uncertainties derived with the LOTUS method lie in between those obtained according to *Steinbrecht et al.* (2017) (smaller values) and the J-distribution method (larger values), though only a limited set of methods to combine trend uncertainties were tested in detail. However, the LOTUS method not only incorporates (at least partially) the influence of systematic sources of uncertainty but also improves the calculation of uncertainties by considering the correlations between contributing trends. Ultimately this exercise in improving the trend uncertainties applies to the concept of merging trends derived from several different data sets and, while some assumptions are made about the nature of the correlations between the trend uncertainties, the most meaningful way to improve the uncertainties in future analyses would be to reconcile the discrepancies between the data sets themselves prior to the merging process.

It should be noted that the combined satellite trend uncertainties in **Figures 5.7** and **5.8** do not include contributions from the choice of regression model and proxies. Sensitivity tests in *Section 4.3.1* have shown that this error can be as large as ~1.5 % (lower/middle stratosphere) or ~3 % (upper stratosphere) per decade, although these numbers are likely a worst case scenario. Adding these in quadrature would reduce the significance of the reported trends in the upper stratosphere considerably and so it is a subject of further research to provide robust quantitative results.

5.4 Ground-based trends

This section compares pre-1997 (January 1985 – December 1996) and post-2000 trends (January 2000 – December 2016) from ground- and satellite-based ozone profiles, averaged over three broad latitude bands (**Figure 5.9**). The main purpose of these trend comparisons is to verify the robustness of the derived trends in the combined records (see discussion about the consistency of records in *Chapter 3*). The broad-band averaging smoothes the records and mitigates atmospheric noise that can be introduced in the station record by short-term meteorological variability and infrequent sampling methods. The disagreement between trends is used for the evaluation of the magnitude of uncorrected drifts in the records. Intercomparisons of multiple records and analyses of the fit residuals provide information that helps identify drifts and provide means for their correction in the future data reprocessing.

As mentioned in previous chapters, there are several ground-based measurement methods for tracking stratospheric ozone recovery, including lidars, Umkehr measurements performed with Dobson/Brewer photometers (hereafter called 'Umkehr'), Microwave radiometers

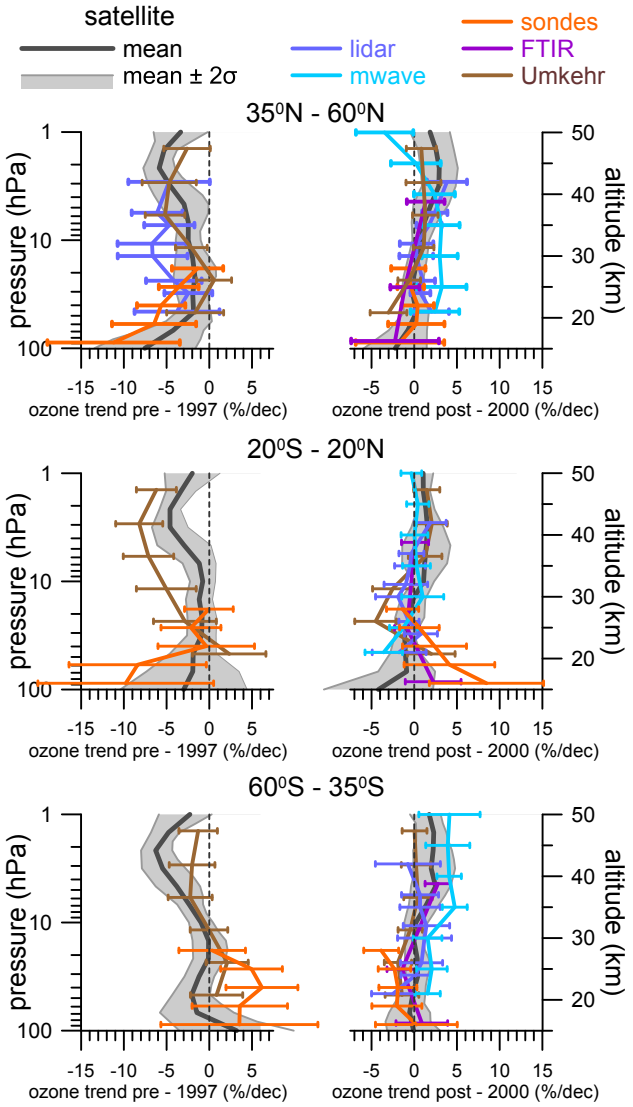


Figure 5.9: *Ozone trends for the pre-1997 and post-2000 periods from the ground-based stations, averaged over the available latitude bands. Upper panel shows 35°N–60°N, middle panel shows 20°S–20°N, and lower panel shows 60°S–35°S.*

(hereafter called 'MWR'), FTIR spectrometers (hereafter called 'FTIR'), and balloon-borne ozonesondes (hereafter called 'ozonesondes'). The length of the historical records, the temporal and vertical sampling of the different records, and the spatial distribution of the different stations variy between these methods (see *Chapter 2* for more details).

The ability of observations at several ground-based stations to capture the trends observed by satellites over the broad-band regions has been studied in this Report and is summarised in recent publications (*i.e.*, *Steinbrecht et al.*, 2009; *Zerefos et al.*, 2018; and reference therein). Results of the analyses performed by *Zerefos et al.* (2018) suggest agreement between trends derived from the subset of the SBUV MOD record selected to match the geolocation of lidar ground-based station and 5° zonally averaged SBUV-MOD records centred at the latitude of the ground-based station.

In the NH mid-latitudes, the ground-based networks are the most densely represented over Europe and North America. Moreover, according to the study of *Zerefos et al.* (2018) and the trend analyses of satellite records presented above (see **Figure 5.2**) ozone variability and trends at altitudes above 10 hPa are coherent over a wide range of latitudes (~20° or wider). Thus, several stations within that latitude range should be able to capture the trends representative of a broad-band average. However, ground-based records are generally shorter than the long SBUV records and some, such as sondes and lidars, are characterised by higher vertical resolutions. This makes trends derived from these records more sensitive to geophysical variability (*e.g.*, polar vortex influence in the winter at mid-latitudes) and incorrect evaluation of long-term atmospheric variability due to, for example, the 11-year solar cycle.

Figure 5.9 shows resulting trends for ground-based stations located within the NH mid-latitudes (top), tropics (middle), and SH mid-latitudes (bottom). Ground-based trends are compared to the satellite broad-band averages derived in *Section 5.3*. Satellite trends for the pre-1997 period (left) and post-2000 period (right) are shown with a grey envelope that represents the combined error (see *Section 5.3.2* for details). The broad-band trends for each ground-based record are calculated from the deseasonalised monthly mean anomalies averaged over the broad-band latitude ranges and at the vertical grid specific to each measurement technique and data processing method (see *Section 2.1* for further details). Deseasonalised anomaly data records are combined prior to the regression analysis in case multiple stations provide data for the broad latitude bands and for the considered measurement technique (see more details in *Section 2.1.2*). The uncertainties for trends obtained for both individual station and broad-band combined records are shown as standard errors of the ILT fit.

The trends for the pre-1997 period in the NH mid-latitudes (**Figure 5.9**, top panel) are represented by the combined Umkehr (brown) and lidar (blue) profiles (see **Table 2.1** in *Chapter 2* for the selection of the stations in the broad-band averages shown in this section). The mean trend pattern in the Umkehr trends is similar to the combined satellite and model trends (not shown). However, the error bars for combined Umkehr trends are larger than the satellite combined error bars, as would be expected. Comparisons between combined satellite and model-derived trends are discussed in the following section (see **Figure 5.11**). While not tested here, these differences are most likely based on limited temporal and spatial sampling within the 35°N–60°N latitude band that are not completely captured by three NH Umkehr stations. Lidar combined datasets tend to have stronger negative trends than found in the Umkehr, model, and satellite records between 20 hPa and 10 hPa pressure. The differences in trends can be explained by the shorter lidar records in the pre-1997 period. For lidar long-term records used in this study, a sufficient number of monthly observations was reached only by the end of the 1980s as indicated in **Figure 2.3**. Therefore, limited temporal and spatial sampling can influence the trends and their uncertainties during this early period. However, the error bars in Umkehr, lidar, and satellite trends overlap, thus indicating consistent estimates of the observed trends between different observing systems. Negative ozonesonde trends in the lower stratosphere tend to be on the far edge of the combined satellite uncertainty envelope, and they are also consistent with the lidar- and model-based trends (**Figure 5.11**).

For the post-2000 period, ground-based data are represented by FTIR, lidar, Umkehr, and MWR records. Some of these records (*i.e.*, FTIR and MWR) began past 1997. For these types of records the ILT method is essentially converted to the multiple linear regression with just a single trend term, but it still uses the same proxies. The exception is the aerosol proxy that becomes non-orthogonal to the trends itself and therefore would alias the recovery trend analyses. In these instances, the aerosol proxy was removed from the statistical MLR model. Results presented for ground-based trends in the NH for post-2000 period (right top panel of **Figure 5.9**) show general good agreement between the different instruments. A trend of -2 % per decade at 50 km (1998–2014) have been published for the Bern MWR (*Moreira et al.*, 2015); Payern MWR also shows a negative trend at 50 km (-0.5 % per decade). The possibility of an influence on the 50 km level ozone content variation of information coming from the levels above 50 km should not be neglected as the MWR averaging kernels are large at that altitude and because the measurement contribution is high for MWR at 50 km. The ground-based trends support the mean values of the recovery trends in the stratosphere detected by satellite observations and models (**Figure 5.11**). Similar to satellite records, they also suggest slightly more negative trends in the lower stratosphere (NH mid-latitudes), although error bars in both data sets are large.

Ground-based trends for tropical (**Figure 5.9,** middle panels) and SH mid-latitudes (**Figure 5.9**, bottom panels) are based on only a few station records (See **Table 2.1** in *Chapter 2*). Thus, it is expected that trends from the ground-based measurements can be biased due to their limited spatial coverage and have larger uncertainties due to limited sampling frequency. For example, the Umkehr record for the pre-1997 period at 10 hPa shows a more negative trend than the satellite (**Figure 5.9**, left middle panel) and model (**Figure 5.11**) estimates. However, there is only one Umkehr record available from the MLO in Hawaii (located at 19°N) and one ozonesonde record from Hilo station (near MLO), and thus these records are not fully representative of the trends derived from the broad 20°S–20°N latitude band. Middle and low stratospheric ozone variability is only weakly correlated between the NH and SH tropics (*i.e.*, *Zerefos et al.*, 2018; and references therein). Still, the subset of SBUV MOD, limited in space to the MLO station location, and 5° zonally averaged satellite measurements at 20°N describe

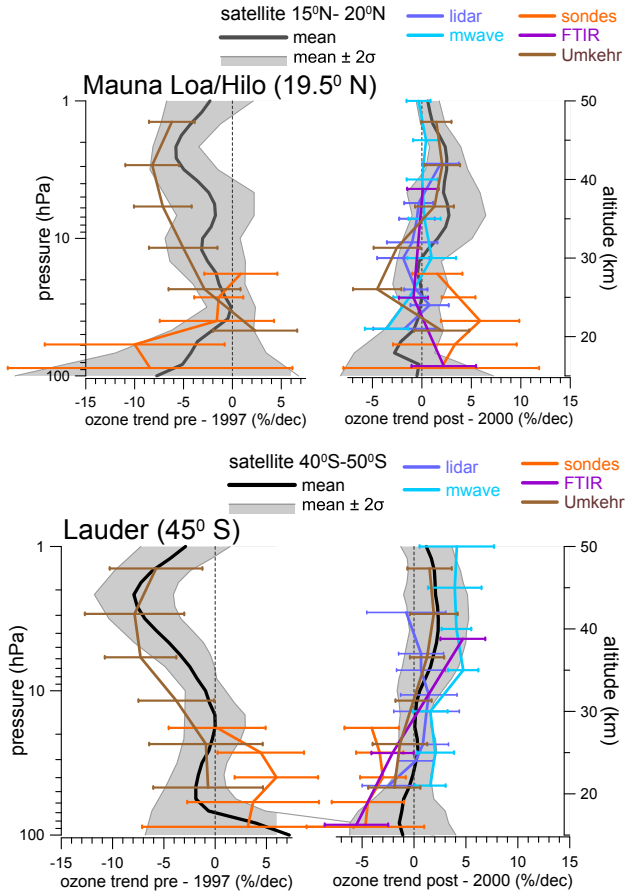


Figure 5.10: *Ozone profile trends from different ground-based data records for the pre-1997 (left) and post-2000 periods (right) at Hawaii, USA (19.5–19.7°N, 155.1–155.6°W; top row) and at Lauder, New Zealand (45.0°S, 169.7°W; bottom row). Error bars represent the 95 % confidence interval. The black line represents the multi-model mean for the latitude band 15°N–25°N and 50°S–40°S, respectively.*

very similar ozone variability (with correlations larger than 0.6 in the middle and upper stratosphere) and also comparable trends.

In order to reduce spatial sampling differences between broad-band satellite and ground-based data collected at MLO/Hilo station (and Lauder, New Zealand as discussed below) the combined satellite trends were recalculated for ~10° zonal bands. All satellite trends for this narrow zonal band are computed analogously to the reported trends in broad-band zones (*Section 5.3.2*). Since the combined datasets were provided with different zonal resolutions, the following combination of trend results were used: The SBUV MOD and SBUV COH trends from 15°N–20°N and 20°N–25°N were averaged whereas the mean of 10°N–20°N and 20°N–30°N is used for the GOZCARDS, SWOOSH, and SAGE-CCI-OMPS records. The SAGE-OSIRIS-OMPS trend is already derived for 15°N–25°N and was used as is.

When ground-based trends are compared to the satellite trends averaged over a narrower zonal band (**Figure 5.10** upper panel), uncertainties in the mean satellite trends increase and the agreement between satellite, ozonesonde, and Umkehr trends for the pre-1997 period at MLO improve. Ground-based trends prior to 1997 for SH mid-latitudes (bottom panel of **Figure 5.9**) are presented by two Umkehr records (Lauder, New Zealand and Perth, Australia) and by the combination of longer ozonesonde records from Lauder, New Zealand (begins in 1986) and shorter record from Macquarie Island (begins in 1994). The Umkehr combined trends show agreement with satellite trends averaged over the 60°S–35°S latitude band. The combined ozonesonde record detects statistically significant positive trends between 40 hPa and 30 hPa that are in agreement with Umkehr trends within their respective error bars but disagree with satellite broad-band estimates. The single station ozonesonde or Umkehr records are not expected to capture trends representative of the broad latitude range, while satellite-derived trends in 5–10° zonally averaged bands also show spatial variability in the middle stratosphere of SH mid-latitudes (see **Figure 5.2**). Variability in broad-band trends can be demonstrated by comparing combined Umkehr pre-1997 trends (bottom panel of **Figure 5.9**) against the larger negative trend derived from the Umkehr record at Lauder only (**Figure 5.10**, bottom panel). Additionally, the mean satellite trends and uncertainties used in **Figure 5.9** are changed from a broad band representation (60°S–35°S) to a 50°S–40°S degree latitude band centred at Lauder (45°S) (**Figure 5.10**). The SBUV MOD and SBUV COH trends from 50°S–45°S and 45°S–40°S were averaged. For the GOZCARDS, SWOOSH, and SAGE-CCI-OMPS records, the trends from 50°S–40°S are used, and for SAGE-OSIRIS-OMPS the average trends between 55°S–45°S and 45°S–35°S were used. The agreement between the 10-degree mean satellite, ozonesonde, and Umkehr trends at Lauder is improved as compared to the results shown in **Figure 5.9**. However, ozonesondes at Lauder at 40 hPa show trends that are different from the satellite averages and outside of the range of satellite trend uncertainties. Umkehr trends at Lauder are less negative than satellite mean trends and agree with ozonesonde trends within their respective uncertainties, thus pointing to either limited sampling at ground-based stations, shortness of the record (Umkehr and sonde records started after 1986), or inhomogeneities in the instrument record (*i.e.*, not all ozonesonde records used in LOTUS analyses were fully homogenised). The trend derived from the combined ozonesonde records is very similar to the trend derived at Lauder only. This similarity is expected because the Lauder record contributed most to the combined record before 1997, and only in the later part of time period is the data set (starting from 1994) a combination of two or three records (the third SH ozonesonde record is added in 1999).

For the post-2000 trends, in addition to Umkehr and ozonesonde MLO records, MWR and lidar records become available, but again these additional records are from MLO only. The trends from all instruments agree well and within the error bars.

There are multiple ozonesonde stations (including SHADOZ; see *Thompson et al.,* 2007) that are used to create the tropical broad-band record for post-2000 trends (see **Table 2.1** and *Section 2.1* for further details). **Figure 5.10** summarises trends derived from different instruments at MLO only (top panel); the ozonesonde trends are those at the neighbouring Hilo station. These single-station ozonesonde trends agree with the broad-band results within the uncertainties, although they indicate stronger trends between 20 km and 30 km and weaker trends at 15 km (100 hPa), which are in close agreement with FTIR measurements at the MLO station. However, ozonesonde records used in the LOTUS Report were provided before the process of homogenisation was finalised (*Sterling et al.,* 2018; *Witte et al.,* 2017; *Witte et al.,* 2018), such that these records can contain uncorrected step changes (see *Chapter 3* for discussion) that can potentially impact the derived trends. Therefore, single-station and broad-band latitude averaged trends will have to be re-evaluated after all ozonesonde homogenised records become available for trend analyses.

Ground-based trends derived from records available in the SH (bottom panel in **Figure 5.9**) indicate that ground-based trends tend to overlap with the satellite broad-band and model averaged trends (bottom panel in **Figure 5.11**) within their respective uncertainties. However, a wider range of ground-based trends in the middle and upper stratosphere is found, depending on the instrument, as compared to the combined satellite or model range of uncertainties. One reason for the instrument trend difference is the combination of two stations for FTIR and Umkehr instruments (see **Table 2.1** in *Chapter 2*), while MWR and lidar trends are derived from the Lauder record only. In order to prove the consistency in the trends derived from different instruments we plotted trends just for the Lauder station (bottom panel of **Figure 5.10**). Note that the combined satellite trend in **Figure 5.10** is based on a narrower latitude range than results shown in **Figure 5.9**. The agreement among trends derived from five instrument-specific datasets at Lauder is improved in comparison to the combined trends shown in **Figure 5.9**, thus highlighting the variability in trends at the ground-based stations away from Lauder (but still inside of the 60°S–35°S band). In addition, the subset of SBUV MOD data limited in space to the Lauder station location and 5° zonal averaged satellite measurements at 45°S captures very similar ozone variability (with correlations of 0.7-0.8 in the middle and 0.5-0.6 in the upper stratosphere; *Zerefos et al.*, 2018) and closely comparable trends (see **Figure 4.7** in *Chapter 4*). The post-2000 trends at altitudes below 20 hPa at Lauder are negative, and ozonesonde and FTIR records show statistically significant negative trends of ~5 % per decade in the lower stratosphere (below 60 hPa). The ozone trends in the lower stratosphere at Lauder are different from the near zero trends derived from the broad-band combined satellite record in the SH mid-latitudes. However, combined ozonesonde and FTIR records available within the 60°S–35°S latitude band also show near zero trends (bottom panel in **Figure 5.9**), thus indicating spatial variability of the trends detected in the SH mid-latitude lower stratosphere.

5.5 Comparison between combined satellite and CCMI model trends

Satellite ozone profile trends and uncertainties derived for the broad latitude bands at 60°S–35°S, 20°S–20°N, and 35°N–60°N are compared with CCMI-1 REF-C2 model trends in **Figure 5.11**. The satellite observed trends were combined based on the method described in *Section 5.3.2* so that only one average satellite-based trend profile and its uncertainties remain in the figure. The mean and median of trends calculated from CCMI-1 REF-C2 model simulations, averaged as described in *Section 2.3.2*, are shown in **Figure 5.11**. The model uncertainties are shown as the grey area, enveloping the model mean at ±2 standard deviations. Trends are fit to the CCMI-1 simulations over the same time periods as for the observations: Pre-1997 (Jan 1985 – Dec 1996) and Post-2000 (Jan 2000 – Dec 2016).

5.5.1 Pre-1997 period

As shown in previous sections, the left panels in **Figure 5.11** display a significant ozone decline in the pre-1997 period for all three broad latitude bands. The largest trends in satellite observations (and models) in the upper stratosphere are -5.9 % ± 1.9 % per decade (-5.4 % ± 2.9 % per decade) in the NH mid-latitudes, -4.8 % ± 1.3 % per decade (-3.8 % ± 2.6 % per decade) in the tropics, and -6.2 % ± 1.8 % per decade (-5.7 % ± 3.0 % per decade) in the SH mid-latitudes. However, smaller but nonetheless significant negative trends are present in pressure regions between 7 hPa and 1 hPa.

In the NH mid-latitudes the median trend from the CCMI-1 model simulations at 100 hPa is similar to that at 50 hPa; however, observations show an enhanced negative trend, up to -8 % per decade. At the same time, the error bars for both models and observations are large in the region just above the tropopause (~10 km/200 hPa in the extratropics and ~17 km/100 hPa in the tropics). This is due to low ozone values, large interannual ozone variability, and large vertical and horizontal ozone gradients in this region. The large error bars in the observational trends are also related to less reliable observations at these altitudes and increased variability that is not fit by the ILT regression model proxies. In the SH mid-latitudes below 20 km/50 hPa (UTLS), observations in the broad latitude bands show a small positive trend (although not statistically significant), while models show strong negative trends. The four limb profile trends exhibit coherent behaviour at altitudes below the 70 hPa level (**Figure 5.6**). The uncertainty envelope in the models tends to increase below the 50 hPa level, still the observation-based trend is found outside of the model 2-sigma uncertainty.

5.5.2 Post-2000 period

In the post-2000 period (right panels in **Figures 5.11**) satellite-based trend analyses show a statistically significant increase of 3.0 % ± 1.9 % per decade at the level of maximum response (around 3 hPa) in the NH mid-latitudes, which is identical to the 3 % per decade reported in *WMO* (2014). Even trends at 2 hPa, although not as strong as at 3 hPa, are statistically significant, indicating that observations are now covering a long enough period to detect the recovery signal on several levels of the NH mid-latitude upper stratosphere. Models suggest a slightly smaller but nonetheless statistically significant trend, 2.2 % ± 1.6 % per decade, for this latitude band in the upper stratosphere. In the lower stratosphere, models predict slightly positive trends for NH mid-latitudes, however, these have large uncertainties, indicating that the different models do not agree on the magnitude or sign of ozone evolution at this level. Satellite observations suggest small negative trends, but their uncertainties are also large, so that this trend value is not statistically significant.

In the tropical latitude band (20°S–20°N), the trends above 3 hPa or 43 km altitude are found to be positive (1.5 % ± 1.4 % per decade) in the satellite-based results. Model simulations suggest positive trends that are statistically significant for altitudes above 4 hPa with a maximum trend value of 1.8 % ± 1.2 % per decade. These satellite-based results support the trends and uncertainties reported in *WMO* (2014), while the uncertainties here are smaller than the uncertainties reported in *Harris et al.* (2015). The updated method of trend combination (see *Section 5.3.2*), quality-improved data sets (see *Chapter 2*), and four additional years of measurements help to isolate the trend signal from the background variability. In the middle stratosphere, a weak negative trend is detected in the observations that is just barely statistically significant at 30 hPa or 23 km, whereas the models suggest no trend at all at this level. In the lower stratosphere, satellite-based trends and models indicate even stronger negative trends but with larger uncertainties.

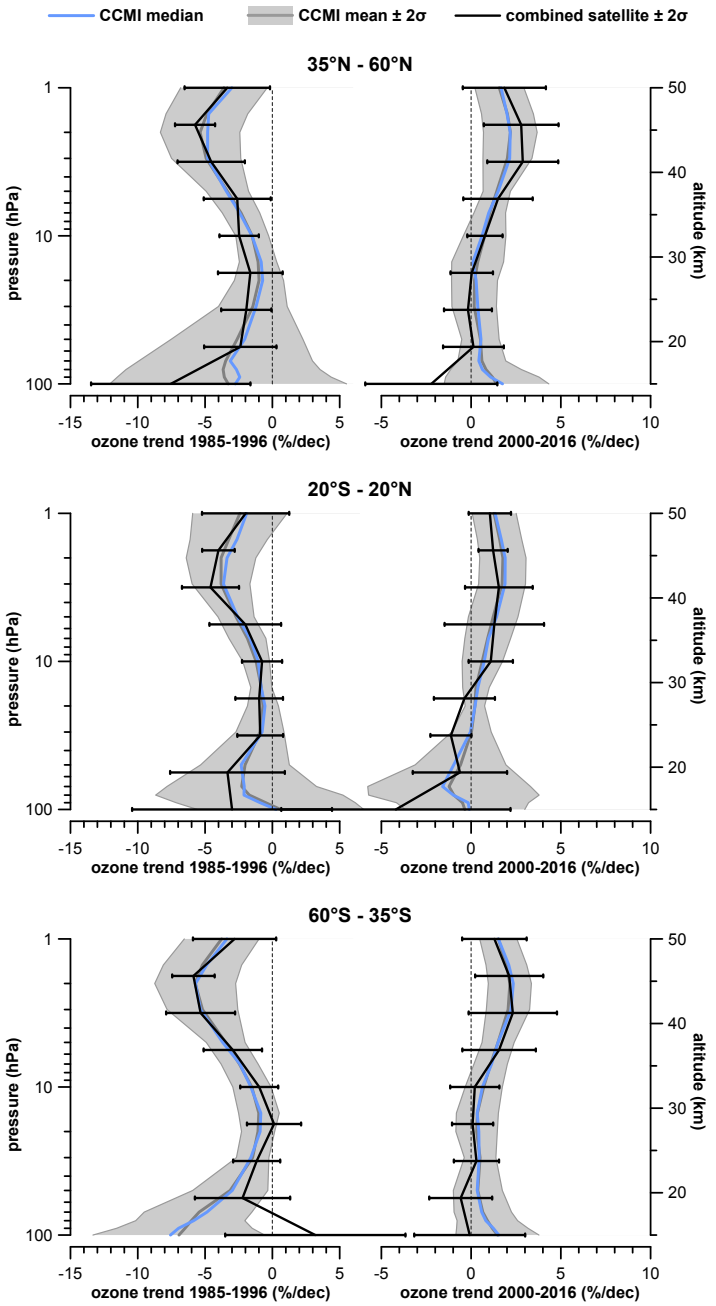


Figure 5.11: *Ozone trends for the pre-1997 and post-2000 periods from the CCMI REF-C2 models' simulation and broadband satellite data sets, averaged over the available latitude bands: 35°N–60°N (upper panel), 20°S–20°N (middle panel), and 60°S–35°S (lower panel).*

In the upper stratosphere at SH mid-latitudes, a positive trend of 2.1 % ± 1.9 % per decade that is just statistically significant is detectable at around 2 hPa/45 km in the satellite-based results given the four additional years of observations. Trends calculated in the upper stratosphere just above or below this level also indicate positive trends, but these are not statistically significant. Model results show statistically significant trends for the altitude region above 10 hPa that are of almost identical magnitude as the observational trends (2.2 % ± 1.1 % per decade at about 2 hPa/45 km). For the rest of the profile shown in **Figure 5.11**, trends calculated from satellite data are around zero, with only a small indication of negative trends at around 60 hPa or 20 km that are not statistically significant. Model trends are also close to zero between 40 hPa and 10 hPa (~23 km to ~33 km) and become slightly positive below these levels. However, the model trends are not statistically significant in the middle and lower stratosphere of the SH mid-latitudes.

5.6 Summary of observed profile trends

5.6.1 Pre-1997 period

Upper stratosphere

For the pre-1997 period we find negative trends across nearly the entire stratosphere in nearly all satellite and ground-based data records (**Table ES.1**; **Figure ES.1**; also **Figures 5.1**, **5.5**, and **5.8**). Individual and combined satellite data show highly statistically significant evidence of declining ozone concentrations since the mid 1980s and well into the 1990s in the upper stratosphere (altitudes above the 10–5 hPa level), where ozone is in photochemical equilibrium (**Figure 5.6**). The depletion reaches a maximum rate of 5.9–6.2 % per decade at mid-latitudes (near 2 hPa, ~42 km) and of 4.8 % per decade in the tropics. Ground-based measurements are much more sparsely sampled in space and time, and, as a result, the significance of the trends is not as high and trend values differ. Nonetheless, both lidar and Umkehr data corroborate the satellite findings for broad-band regions in the upper stratosphere (see **Figure 5.8** and **Figure 5.9**).

Middle stratosphere

Ozone decline rates in the middle stratosphere down to the 50 hPa level are considerably smaller, 1–2 % per decade, than in the upper layers of the stratosphere. In fact, these values are too small to be statistically significant in most regions of the globe (see **Figure 5.1**). The combined satellite trends for the broad-latitude bands do not show any statistically significant trend in the middle stratosphere. A few individual satellite data sets (*e.g.*, SBUV MOD, GOZCARDS, and SWOOSH) show statistically significant trend values in the NH mid-latitudes at their native resolution as well as in the broader latitude bands (**Figure 5.6**). Ground-based measurements generally do not show statistically significant trends in the tropics and SH mid-latitudes. Only lidar measurements in the NH mid-latitudes display significant negative trends in the middle stratosphere, but these are based on shorter records in the 1985–1997 period (**Figure 5.9**).

Lower stratosphere

In the NH lower stratosphere, both satellite and ozonesonde data point to a negative trend of 5 % per decade and more, which is seemingly significant for the satellite analysis (**Figure 5.8**). Satellite trends close to 100 hPa in the tropics and in the SH are less than 3 % per decade, but the negative trends are reversed to positive values around the Equator (**Figure 5.1**). However, our confidence in the results in the lower stratosphere is not as high as that in the upper stratosphere. Estimating trend uncertainty, and therefore statistical significance, is inherently more complicated in the lower stratosphere because of large natural variability influenced by transport and mixing processes, low ozone concentrations, waning sensitivity of satellite observations, and the lack of independent measurements[1]. Additional research will be needed to put the trend results in the lower stratosphere on more solid ground.

5.6.2 Post-2000 period

Upper stratosphere

For the post-2000 period, we find positive trends in all satellite and most ground-based records in a large part of the upper stratosphere (**Table ES.1**; **Figure ES.1**; also **Figures 5.2**, **5.3**, **5.5**, and **5.8**). The statistical significance of the estimates varies between latitude bands and between individual data sets and the combined satellite result. However, the majority of individual data sets in the upper stratosphere show significant trends over the NH mid-latitudes, while several data sets also contain statistically significant trends over (part of the) tropics and SH mid-latitudes (**Figure 5.5** and **Figure 5.8**).

In the upper stratosphere of NH mid-latitudes positive trends range between 2.0 % and 3.1 % per decade when satellite results are combined. Lidar and Umkehr data support this finding, but MWR data do not (**Figure 5.9**). However, the anomaly time series clearly indicate aberrant values in the two-station combined MWR record in recent years, which leads to a negative trend (**Figure 3.4** and **Figure 5.4**). There is currently no clear understanding of this discrepancy and additional research is needed. Combined satellite trends at NH mid-latitudes are statistically significant between 3.8–1.2 hPa (**Figure 5.13**; **Table S5.1**). Statistical significance in ground-based trends is only found for combined lidar trends at altitudes above the 5 hPa level.

In the tropics, most individual satellite and ground-based records show small positive trends that lie close to or barely beyond the 2-sigma threshold for significance. The only exceptions are SAGE-CCI-OMPS over the full range and SBUV MOD at 3-2 hPa. The combined satellite trend in the upper stratosphere ranges between 1.0 % and 1.6 % per decade and is significant between 2.6–1.0 hPa. SBUV MOD and SBUV COH results diverge largely at 8–3 hPa (**Figure 5.6**). None of the ground-based trends in the tropics are significant. Ground-based observations of tropical upper stratospheric ozone were considered at just one location (MLO, Hawaii) which complicates a direct comparison to the satellite data over the broader band.

Upper stratospheric trends derived from combined satellite data over SH mid-latitudes range between 1.8 % and 2.3 % per decade. All analyses of individual satellite and ground-based records, except for the only lidar record, show positive or zero trends in this region. The combined satellite trend

[1] Below the 50 hPa level, the SBUV profile data are not used (*Section 2.2.2.*) and all merged limb profile records rely (mostly) on SAGE II data.

data are significant over the 1.8–1.2 hPa range. However, the confidence in the positive trend significance is less than in the NH since some satellite records show trends right at the 2-sigma threshold. The MWR and FTIR trends are significant at altitudes above the 5 hPa level. Similar to the tropical belt, only one or two ground-based sites were available to assess trends in the SH mid-latitudes. In general, satellite and ground-based trends agree in the upper stratosphere within their respective uncertainty bounds.

Middle stratosphere

Trends in the middle stratosphere and over mid-latitudes are found to be smaller than 0.5 % per decade and not significant. However, in the tropical region, trends derived from the SBUV-based data sets and SWOOSH are negative and statistically significant (**Figure 5.2**). Other satellite records and ground-based records in the tropics (mostly from MLO/Hilo) show trends close to zero that are not statistically significant (**Figure 5.6**, **Figure 5.9**, and **Figure 5.11**).

Lower stratosphere

In the lower stratosphere, individual and combined satellite trends are mostly negative but statistically insignificant at both their native resolution and in broad latitude bands (**Figure 5.2**, **Figure 5.6**, and **Figure 5.11**). Most ground-based measurements are also not statistically significant for all three latitude bands in the lower stratosphere; however, in the tropics, ozonesondes show a significant positive trend at around 100 hPa of 8 % per decade, and FTIR measurements also indicate a positive trend which is not significant.

5.6.3 Comparison of LOTUS trend results with previous assessments

Figures 5.12–**5.14** display comparisons of LOTUS broadband averaged trend results with those of the most recent assessments (*i.e.*, *WMO*, 2014; *Harris et al.*, 2015; and *Steinbrecht et al.*, 2017). For both periods, pre-1997 and post-2000, combined trends with their respective uncertainties are shown in **Figure 5.12**, while **Figure 5.13** illustrates the differences in the significance of trends between the various assessments. In addition, the trend uncertainties are directly compared in **Figure 5.14**. *Steinbrecht et al.* (*2017*; or S17 hereafter) found latitude-pressure patterns in post-2000 trends similar to those reported here (or L19 hereafter); magnitudes of the trends agree within 0.5% per decade. This is not surprising since both assessments determined trends for the combined satellite records as the unweighted mean. Furthermore, trends are estimated for the same analysis period (post-2000), from similar data records, using similar non-trend proxies and similar trend proxies.

However, the uncertainty of the combined trend is computed slightly differently between *S17* and *L19* (see also *Section 5.3.4*). Both rely on the concept of variance of the sample mean, which determines the total uncertainty as a sum of random and systematic effects, where drifts are considered as relative to the zero sample mean drift. *S17* uncertainty is computed as a biased estimator of the standard deviation of the sample mean and under the assumption that all combined records contain only three independent data records in the post-2000 period, such that $n_{eff}=3$ (or *2*, at altitudes below the 50 hPa level). The *L19* approach is based on $n_{eff}=1.4$ (or *1.3*, below the 50 hPa level) independent records (derived from correlation analyses of the trend fit residuals) and includes an additional term that represents a lower bound, which is equal to the propagated uncertainties from the regression coefficients. Note that for the comparison between *S17* and *L19* results, *S17* derived trend uncertainties were corrected by a factor *N/(N-1)* in order to consider the same unbiased definition of the sample variance. In the upper stratosphere, where the sample variance is large, the difference in n_{eff} plays a decisive role in testing the null hypothesis (*i.e.*, no-trend). *S17* trends are statistically significant across the entire upper stratosphere. In the *L19* trend analysis, high significance is found only at NH mid-latitudes, and less significant trends are found in the tropics and at SH mid-latitudes. At the moment, there is not sufficient information in the trend analyses to help determine which assumption is more realistic. However, it is not unreasonable to assume that in the upper stratosphere, *S17* and *L19* uncertainties represent, respectively, a lower and upper bound. Note that for the comparison between *S17* and *L19* results in **Figure 5.12**, *S17* derived trend uncertainties were corrected by a factor *N/(N-1)* in order to consider the same unbiased definition of the sample variance.

In the lower stratosphere, there are concerns that *S17* may underestimate true uncertainty. The trend sample has only a few members and their spread is fairly, perhaps unrealistically, small at these altitudes. Also, due to the small trend sample there is the risk of underestimating the true variance of the mean when trends coincide. Even the *L19* approach, which includes a term for error propagation from the regression coefficients, is not capable of capturing all sources of uncertainty (*e.g.*, most importantly measurement drift). Therefore, *L19* assumes that these errors constitute a lower bound, which is primarily reached in the lower stratosphere and also in part of the middle stratosphere. Pre-1997 trend results (not shown in the *S17* publication but provided by authors through personal communication) are also in good agreement with *L19* (**Figure 5.12**) but differ by 1–2 % per decade, most likely as a result of the earlier start in the time series used in the *S17* analysis (1979) as compared to *L19* (1985). All LOTUS trends for satellite, ground-based, and model data were computed starting in 1985 for ease of comparison. *Steinbrecht et al.* (2017) and earlier assessments by *WMO* (2014) and *Harris et al.* (2015) analysed time series starting in the early and/or late 1970s, which has an effect on the magnitude of the derived trends.

Harris et al. (2015; or *H15* hereafter) trends are comparable to LOTUS trends, though the study used a somewhat different set of satellite records; the length of the records was shorter (by about 4 years) and some records exhibited large drifts (*i.e.*, OSIRIS).

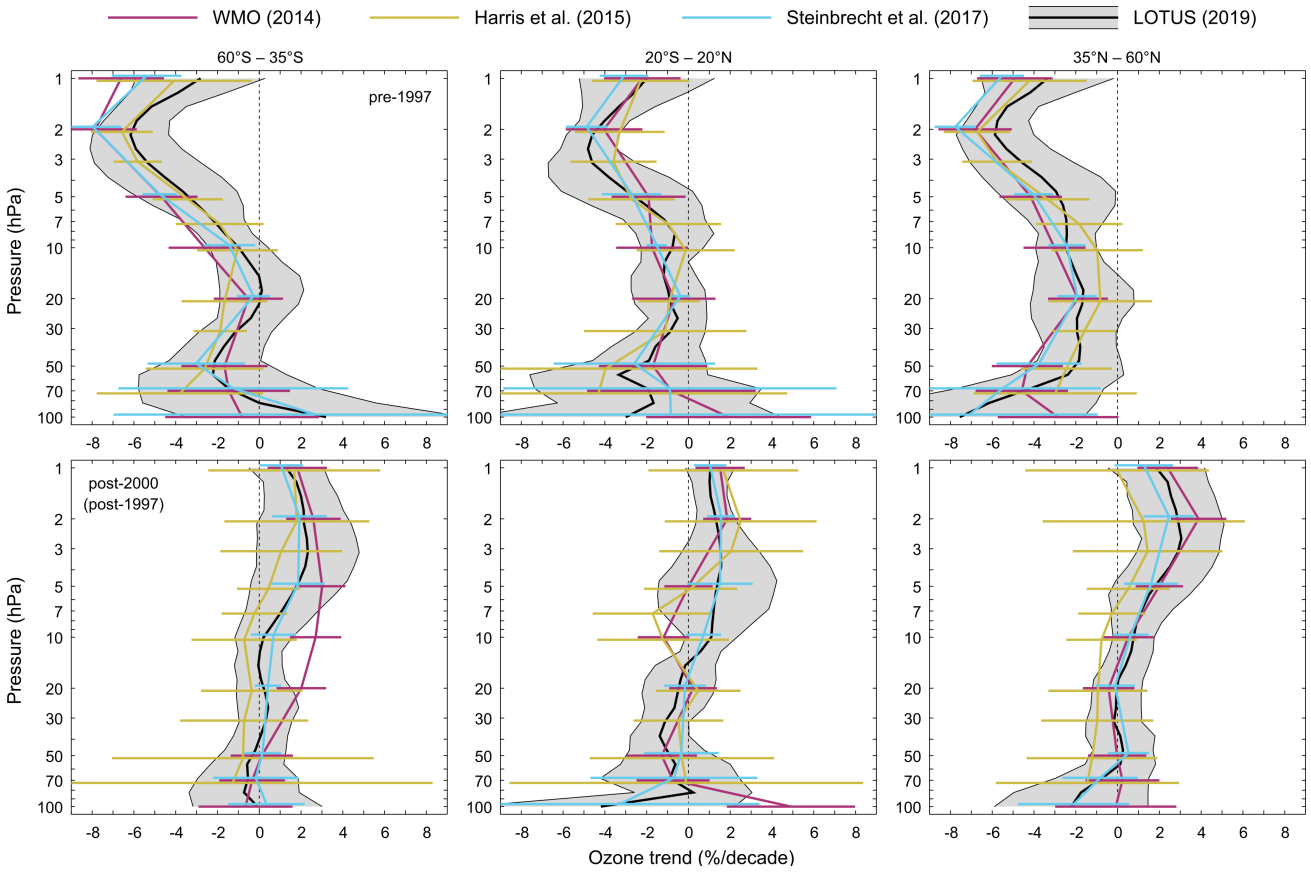


Figure 5.13: *Overview of ozone profile trends from past and recent assessments: LOTUS (this work), WMO (2014), Harris et al. (2015), and Steinbrecht et al. (2017) are shown in black, red, orange, and blue respectively. Top row shows trends before the turnaround of ODSs and bottom row since the turnaround (analysis time period differs by assessment). Shaded area and error bars represent the 95% confidence interval for the combined trend. Coloured profiles are slightly offset on the vertical axis for display purposes.*

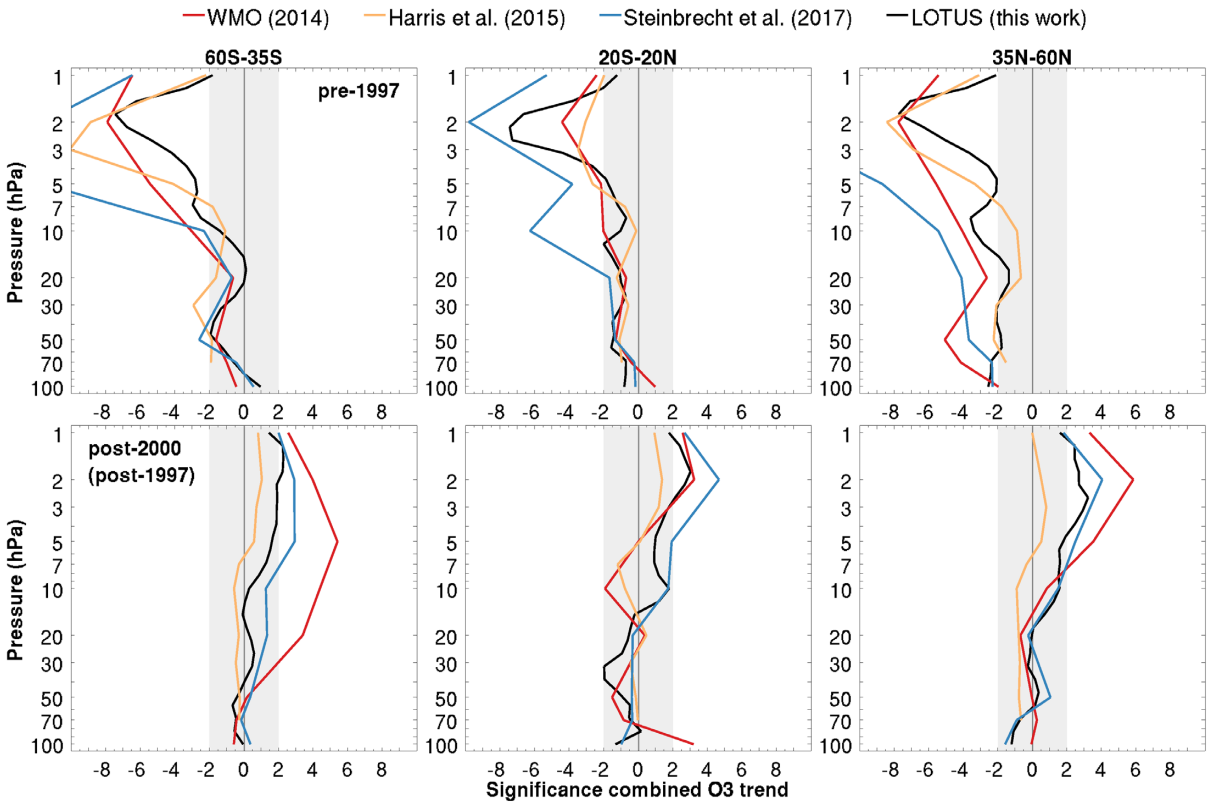


Figure 5.12: *As in* ***Figure 5.12*** *but for the significance of ozone profile trends from past and recent assessments.*

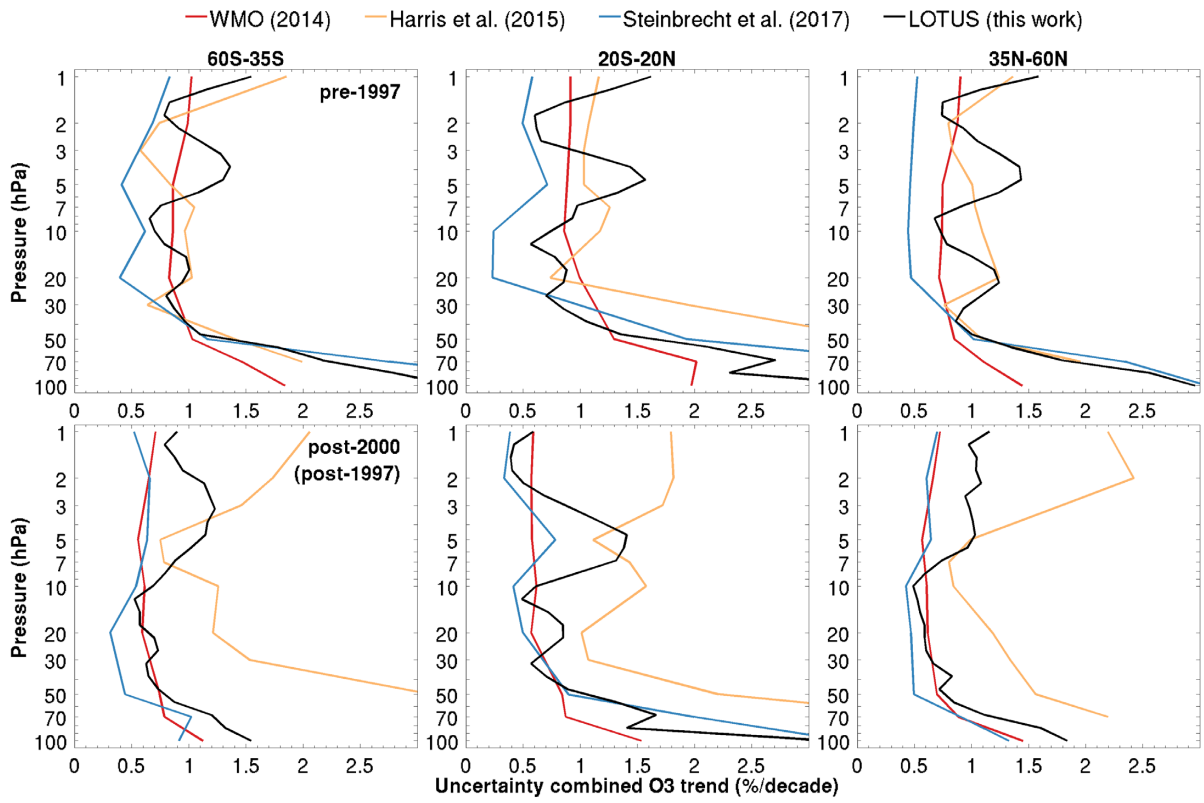


Figure 5.14: *As in* ***Figure 5.12*** *but for the uncertainty (1-sigma) of ozone profile trends from past and recent assessments.*

Nevertheless, *H15* trends in broad-band regions are consistent with LOTUS and other published analyses in the pre-1997 period, while detecting smaller trends in the upper stratosphere for NH and SH mid-latitudes, and stronger negative trends in the tropical mid-stratosphere over the post-1997 period. However, *H15* used the very conservative J-distribution approach (*Section 5.3.3*) to compute combined uncertainties and thus most of the post-2000 trend uncertainties are likely overestimated and consequently the trends are not statistically significant (*Section 5.3.4*).

Ball et al. (2017, 2018) have taken a less traditional approach for evaluation of long-term trends. The satellite datasets used in the LOTUS Report were re-combined using a Bayesian statistical approach to obtain a single BASIC time series. In this approach, common variability between multiple data sets is given greater weight than data sets that deviate from the group, thus reducing the influence of data with non-physical offsets and drifts. For trend determination, the authors used DLM (*Laine et al.*, 2014; *Ball et al.*, 2017, 2018) instead of the MLR used in the LOTUS Report. It is hard to compare the MLR and DLM results directly, because the background changes are nonlinear and the trend is represented as a change between the beginning and end of the period.

Ball et al. (2017, 2018) trends are similar but not identical to the LOTUS analyses and results are hard to compare directly given the different methodologies. The LOTUS MLR model applied to the BASIC data set shows broad agreement with trends derived from other datasets (*Section 5.1* and *Supplement*). Generally, DLM and LOTUS results agree in the upper stratosphere, while the DLM significance in the post-2000 period is lower (*i.e.*, has larger uncertainties). In addition, the NH and SH trends based on the DLM-BASIC analysis are approximately symmetric, while larger asymmetry in pre-1997 period is found in the LOTUS MLR analyses. Future versions of the BASIC dataset that use a Bayesian approach to combine data records will be more applicable to analysis of individual satellite records rather than previously merged records.

5.7 Summary

In this chapter we analysed and compared trend profiles from multiple satellite-based merged data sets and ground-based station data, as well as the estimated mean trend and uncertainties computed from the individual satellite data set trends. For comparison purposes, the multi-model mean/median of CCMI simulations were also shown.

The ozone trends derived from satellite measurements at their native resolutions show a very similar, though oppositely signed, pattern for the pre-1997 (1985–1996) and post-2000 (2000–2016) trends (*Section 5.1*). The pre-1997 results show a general pattern of negative (4–9 % per decade) trends in the upper stratosphere (above 5 hPa / 35 km), which is consistent with previous findings and post-2000 trends from the different satellite data sets show broadly positive trends (~2–3 % per decade on average) in the mid-latitudes of both hemispheres between 5 hPa and 2 hPa (around 37 km to 45 km).

Differences between results from the analysed data sets might be caused by differences in merging techniques, conversions between different native units, consideration of sampling biases, and inherent instrument measurement uncertainty.

Broad latitude band averages (60°S–35°S, 20°S–20°N, and 35°N–60°N) of data from the different satellite data sets were calculated and their trends analysed to provide information for different specific geographical regions (SH and NH mid-latitudes and tropics). The ozone trends in broad bands from individual merged datasets were inter-compared and the method for evaluation of overall combined trends and their uncertainty was proposed. The LOTUS method not only incorporates the influence of systematic sources of uncertainty but also improves the calculation of uncertainties by considering the correlations between contributing trends (*Section 5.3*).

The combined trend profile from satellite measurements (*Section 5.5*) shows significant ozone decline in the pre-1997 period for all three broad latitude bands of -5.9 % ± 1.9 % per decade (NH mid-latitudes), -4.8 % ± 1.3 % per decade (tropics), and -6.2 % ± 1.8 % per decade (SH mid-latitudes) in the upper stratosphere. Post-2000 trends are less clear, with only the upper stratosphere in the NH mid-latitudes showing a clear significant ozone increase (3.1 % ± 1.9 % per decade). This value is in line with previous studies. Trends in the upper stratosphere in the tropics and SH mid-latitudes are barely significant (1.5 % ± 1.4 % per decade and 2.1 % ± 1.9 % per decade, respectively).

Ground-based trends in broad bands support these satellite results. The agreement is improved when there are multiple stations that are used for averages and when ground-based trends are similar over the broad-band latitude and altitude regions. However, noticeable differences between Northern and Southern tropics suggest that ground-based and satellite-based trends should be compared over smaller latitude band widths. Although trends derived from satellite records limited to station overpass and averaged over 5° latitude bands agree well in the stratosphere and at different ground-based station locations, it is still advantageous to compare ground-based data with overpass satellite records to reduce spatial inhomogeneities.

Differences between instrument-based trends for co-located records need to be investigated further. Among potential causes that can impact trends and associated uncertainties are the limited or temporally inhomogeneous frequency of measurements, clear-sky sampling biases (which impact trends in lower stratosphere), differences in spectroscopic databases used to retrieve ozone profiles from initial raw measurements, differences between reanalyses-based temperature profiles used for altitude/pressure conversions and data processing, and other details of retrieval algorithms used for instrument-specific data processing. Analyses and comparisons done for this Report have brought attention to the opportunity to study instrumental artefacts at the "super stations" (*i.e.*, Lauder and Hilo) that have a wide range of co-located data sources.

References

Adams, C., *et al.*, 2013: Characterization of Odin-OSIRIS ozone profiles with the SAGE II dataset, *Atmos. Meas. Tech.*, **6**, 1447-1459, doi:10.5194/amt-6-1447-2013**.**

Adams, C., *et al.*, 2014: Assessment of Odin-OSIRIS ozone measurements from 2001 to the present using MLS, GOMOS, and ozonesondes, *Atmos. Meas. Tech.*, **7**, 49-64, doi:10.5194/amt-7-49-2014.

Aquila, V., *et al.*, 2013: The response of ozone and nitrogen dioxide to the eruption of Mt. Pinatubo at southern and northern midlatitudes, *J. Atmos. Sci.*, **70**, 894–900, doi:10.1175/JAS-D-12-0143.1.

Bai, K., *et al.*, 2017: An intercomparison of multidecadal observational and reanalysis data sets for global total ozone trends and variability analysis, *J. Geophys. Res.*, **122**, 7119–7139, doi:10.1002/2016JD025835.

Baldwin, M.P., *et al.*, 2001: The Quasi-Biennial Oscillation, *Rev. Geophys.*, **39**, 179–229, doi:10.1029/1999RG000073.

Ball, W.T., *et al.*, 2017: Reconciling differences in stratospheric ozone composites, *Atmos. Chem. Phys.*, **17**, 12269-12302, doi:10.5194/acp-17-12269-2017.

Ball, W.T., *et al.*, 2018: Evidence for a continuous decline in lower stratospheric ozone offsetting ozone layer recovery, *Atmos. Chem. Phys.*, **18**, 1379-1394, doi:10.5194/acp-18-1379-2018.

Bhartia, P.K., *et al.*, 2013: Solar Backscatter UV (SBUV) total ozone and profile algorithm, *Atmos. Meas. Tech.*, **6**, 2533–2548, doi:10.5194/amt-6-2533-2013.

Bodeker, G.E., Boyd, I.S., and Matthews, W.A., 1998: Trends and variability in vertical ozone and temperature profiles measured by ozonesondes at Lauder, New Zealand: 1986–1996, *J. Geophys. Res.,* **103**(D22), 28661–28681, doi:10.1029/98JD02581.

Bodeker, G.E., *et al.,* 2013: A vertically resolved, global, gap-free ozone database for assessing or constraining global climate model simulations, *Earth Syst. Sci. Data*, **5**, 31-43, doi:10.5194/essd-5-31-2013.

Bourassa, A.E., *et al.,* 2014: Trends in stratospheric ozone derived from merged SAGE II and Odin-OSIRIS satellite observations, *Atmos. Chem. Phys.*, **14**, 6983-6994, doi:10.5194/acp-14-6983-2014.

Bourassa, A.E., *et al.,* 2018: Drift-corrected Odin-OSIRIS ozone product: algorithm and updated stratospheric ozone trends, *Atmos. Meas. Tech.*, **11**, 489-498, doi:10.5194/amt-11-489-2018.

Carpenter, B., *et al.,* 2016: Stan: A probabilistic programming language, *J. Stat. Softw.*, **20**, 1–37, doi:10.18637/jss.v076.i01.

Chiodo, G., *et al.*, 2014: On the detection of the solar signal in the tropical stratosphere, *Atmos. Chem. Phys.*, **14**, 5251-5269, doi:10.5194/acp-14-5251-2014.

Cochrane, D. and Orcutt, G.H., 1949: Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American statistical association*, **44**(245), 32-61, doi: 10.2307/2280349.

Coldewey-Egbers, M., *et al.,* 2014: A new health check of the ozone layer at global and regional scales, *Geophys. Res. Lett.*, **41**, 4363–4372, doi:10.1002/2014GL060212.

Damadeo, R.P., *et al.*, 2013: SAGE version 7.0 algorithm: application to SAGE II, *Atmos. Meas. Tech.*, **6**, 3539-3561, doi:10.5194/amt-6-3539-2013.

Damadeo, R.P., Zawodny, J.M., and Thomason, L.W., 2014: Reevaluation of stratospheric ozone trends from SAGE II data using a simultaneous temporal and spatial analysis, *Atmos. Chem. Phys.*, **14**, 13455-13470, doi:10.5194/acp-14-13455-2014.

Damadeo, R.P., *et al.*, 2018: The impact of nonuniform sampling on stratospheric ozone trends derived from occultation instruments, *Atmos. Chem. Phys.*, **18**, 535-554,doi:10.5194/acp-18-535-2018.

Davis, S.M., *et al.*, 2016: The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database: a long-term database for climate studies, *Earth Syst. Sci. Data,* **8**, 461-490, doi:10.5194/essd-8-461-2016.

De Backer, H., De Muer, D., and De Sadelaer, G., 1998: Comparison of ozone profiles obtained with Brewer-Mast and Z-ECC sensors during simultaneous ascents, *J. Geoph. Res.*, **103**(D16), 19641-19648, doi:10.1029/98JD01711.

DeLand, M.T., *et al.*, 2012: Calibration of the SBUV version 8.6 ozone data product, *Atmos. Meas. Tech.*, **5**, 2951-2967, doi:10.5194/amt-5-2951-2012.

Deshler, T., *et al.*, 2008: Atmospheric comparison of electrochemical cell ozonesondes from different manufacturers, and with different cathode solution strengths: The Balloon Experiment on Standards for Ozonesondes, *J. Geophys. Res.*, **113,** D04307, doi:10.1029/2007JD008975.

Deshler, T., *et al.*, 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, doi:10.5194/amt-10-2021-2017.

Douglass, A.R., *et al.*, 2017: Multi-decadal records of stratospheric composition and their relationship to stratospheric circulation change, *Atmos. Chem. Phys.*, **17**, 12081–12096, doi:10.5194/acp-7-4935-2007.

Dudock de Wit, T., Bruinsma, S., and Shibasaki, K., 2014: Synoptic radio observations as proxies for upper atmosphere modelling, *J. Space Weather Space Clim.*, **4**, doi:10.1051/swsc/2014003.

Eyring, V., et al., 2013: Long-term ozone changes and associated climate impacts in CMIP5 simulations, *J. Geophys. Res. Atmos.*, **118**, 5029–5060, doi:10.1002/jgrd.50316.

Farman, J.C., Gardiner, B.G., and Shanklin, J.D., 1985: Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction, *Nature*, **315**, 207–210, doi: 10.1038/315207a0.

Fragkos, K., *et al.*, 2016: Extreme total column ozone events and effects on UV solar radiation at Thessaloniki, Greece, *Theoretical and Applied Climatology*, **126**, 3-4, 505-517, doi: 10.1007/s00704-015-1562-3.

Frith, S.M., *et al.*, 2014: Recent changes in total column ozone based on the SBUV Version 8.6 Merged Ozone Data Set, *J. Geophys. Res.-Atmos.*, **119**, 9735–9751, doi:10.1002/2014JD021889.

Frith, S.M., *et al.*, 2017: Estimating uncertainties in the SBUV Version 8.6 merged profile ozone data set, *Atmos. Chem. Phys.*, **17**, 14695-14707, doi:10.5194/acp-17-14695-2017.

Froidevaux, L., *et al.*, 2015: Global OZone Chemistry And Related trace gas Data records for the Stratosphere (GOZCARDS): methodology and sample results with a focus on HCl, H2O, and O3, *Atmos. Chem. Phys.*, **15**, 10471-10507, doi:/10.5194/acp-15-10471-2015.

Gabriel, A., and Schmitz, G., 2003: The Influence of Large-Scale Eddy Flux Variability on the Zonal Mean Ozone Distribution. *J. Climate*, **16**, 2615–2627, doi: 10.1175/1520-0442(2003)016<2615:TIOLEF>2.0.CO;2.

GCOS, 2011: Systematic Observation Requirements for Satellite-based Products for Climate, 2011 update, p. 127, available at **https://library.wmo.int/doc_num.php?explnum_id=3710**.

Godin S., *et al.*, 1999: Ozone Differential Absorption Lidar Algorithm Intercomparison, *Appl. Opt.*, Vol **38**, 30, 6225-6236, doi: 10.1364/AO.38.006225.

Godin-Beekmann, S., Porteneuve, J., and Garnier,A., 2003: Systematic DIAL lidar monitoring of the stratospheric ozone vertical distribution at Observatoire de Haute-Provence (43.92°N, 5.71°E), *J. Env. Monitoring*, **5**, 57-67, doi:10.1039/B205880D.

Götz, F.W.P., 1931: Zum Strahlungsklima des Spitzbergensommers. Strahlungs-und Ozonmessungen in der Königsbucht 1929, *Gerlands Beitr. Geophys.*, **31**, 119-154.

Harris, N.R.P., *et al.*, 2015: Past changes in the vertical distribution of ozone – Part 3: Analysis and interpretation of trends, *Atmos. Chem. Phys.*, **15**, 9965-9982, doi:10.5194/acp-15-9965-2015.

Harris, N.R.P., *et al.*, 2016: The SI2N Initiative: A better understanding of ozone profile trends, *SPARC Newsletter*, **47**, 4-8, available at: **https://www.sparc-climate.org/publications/newsletter/**.

Hase, F., Blumenstock, T., and Paton-Walsh, C., 1999: Analysis of the instrumental line shape of high-resolution Fourier transform IR spectrometers with gas cell measurements and new retrieval software, *Appl. Opt.*, **38**, 3417-3422, doi:10.1364/AO.38.003417.

Hase, F., 2000: Inversion von Spurengasprofilen aus hochaufgelösten bodengebundenen FTIR-Messungen in Absorption, Dissertation, *Wissenschaftliche Berichte Forschungszentrum Karlsruhe*, **FZKA 6512**; ISSN 0947–8620, Forschungszentrum Karlsruhe, Karlsruhe, Germany.

Hassler, B., *et al.*, 2014: Past changes in the vertical distribution of ozone – Part 1: Measurement techniques, uncertainties and availability, *Atmos. Meas. Tech.*, **7**, 1395-1427, doi:10.5194/amt-7-1395-2014.

Hegglin, M.I., 2016: Report on the IGAC/SPARC Chemistry-Climate Model Initiative (CCMI) 2015 science workshop, *SPARC Newsletter*, **46**, 37–42, available at: **https://www.sparc-climate.org/publications/newsletter/**.

Hocke, K., *et al.*, 2007: Comparison and synergy of stratospheric ozone measurements by satellite limb sounders and the ground-based microwave radiometer SOMORA, *Atmos. Chem. Phys.*, **7**, 4117–4131, doi:10.5194/acp-7-4117-2007.

Hubert, D., *et al.*, 2016: Ground-based assessment of the bias and long-term stability of 14 limb and occultation ozone profile data records, *Atmos. Meas. Tech.*, **9**, 2497-2534, doi:10.5194/amt-9-2497-2016.

Hubert, D. *et al.*, 2019: The temporal and spatial homogeneity of ozone profile data records obtained by ozonesonde, lidar and microwave radiometer networks, in preparation.

Jeannet, P., *et al.*, 2007: Ozone balloon soundings at Payerne (Switzerland): Reevaluation of the time series 1967–2002 and trend analysis, *J. Geophys. Res.*, **112**, D11302, doi:10.1029/2005JD006862.

Johnson, B.J., *et al.*, 2002: ECC Ozonesonde pump efficiency measurements and tests on the sensitivity to ozone of buffered and unbuffered ECC sensor cathode solutions, *J. Geophys. Res.*, **107**, D19 doi: 10.1029/2001JD000557.

Kramarova, N.A., *et al.*, 2013a: Validation of ozone monthly zonal mean profiles obtained from the version 8.6 Solar Backscatter Ultraviolet algorithm, *Atmos. Chem. Phys.*, **13**, 6887-6905, doi:10.5194/acp-13-6887-2013.

Kramarova, N.A., *et al.*, 2013b: Interpreting SBUV smoothing errors: an example using the quasi-biennial oscillation, *Atmos. Meas. Tech.*, **6**, 2089-2099, doi:10.5194/amt-6-2089-2013.

Kramarova, N.A., *et al.*, 2018: Validation of ozone profile retrievals derived from the OMPS LP version 2.5 algorithm against correlative satellite measurements, *Atmos. Meas. Tech.*, **11**, 2837-2861, doi:10.5194/amt-11-2837-2018.

Komhyr, W.D., 1969: Electrochemical concentration cells for gas analysis, *Ann.Geoph.*, **25**, 203-210.

Kuttippurath, J., *et al.*, 2015: A cautionary note on the use of EESC-based regression analysis for ozone trend studies, *Geophys. Res. Lett.*, **42**, 162–168, doi:10.1002/2014GL062142.

Kyrölä, E., *et al.*, 2013: Combined SAGE II–GOMOS ozone profile data set for 1984–2011 and trend analysis of the vertical distribution of ozone, *Atmos. Chem. Phys.*, **13**, 10645-10658, doi:10.5194/acp-13-10645-2013.

Laeng, A., *et al.*, 2014: Validation of MIPAS IMK/IAA V5R_O3_224 ozone profiles, *Atmos. Meas. Tech.*, **7**, 3971-3987, doi:10.5194/amt-7-3971-2014.

Laeng, A., *et al.*, 2019: Creating long-term climate data records using transfer functions: methodology and application to SAGE II, MIPAS and OMPS ozone profile datasts, in preparation.

Laine, M., Latva-Pukkila, N., and Kyrölä, E., 2014: Analysing time-varying trends in stratospheric ozone time series using the state space approach, *Atmos. Chem. Phys.*, **14**, 9707-9725, doi:10.5194/acp-14-9707-2014.

Leblanc, T., and McDermid, I.S., 2000: Stratospheric ozone climatology from lidar measurements at Table Mountain (34.4°N, 117.7°W) and Mauna Loa (19.5°N, 155.6°W), *J. Geophys. Res.*, **105**(D11), 14613–14623, doi:10.1029/2000JD900030.

Leblanc, T., *et al.*, 2016a: Proposed standardized definitions for vertical resolution and uncertainty in the NDACC lidar ozone and temperature algorithms – Part 1: Vertical resolution, *Atmos. Meas. Tech.*, **9**, 4029-4049, doi:10.5194/amt-9-4029-2016.

Leblanc, T., *et al.*, 2016b: Proposed standardized definitions for vertical resolution and uncertainty in the NDACC lidar ozone and temperature algorithms – Part 2: Ozone DIAL uncertainty budget, *Atmos. Meas. Tech.*, **9**, 4051-4078, doi:10.5194/amt-9-4051-2016.

Lee, H. and Smith, A.K. 2003: Simulation of the combined effects of solar cycle, quasi-biennial oscillation, and volcanic forcing on stratospheric ozone changes in recent decades, *J. Geophys. Res.*, **108**, ACH 4, doi:10.1029/2001JD001503.

Li, F., Stolarski, R.S., and Newman, P.A., 2009: Stratospheric ozone in the post-CFC era, *Atmos. Chem. Phys.*, **9**, 2207-2213, doi:10.5194/acp-9-2207-2009.

Lin, M., *et al.*, 2015: Climate variability modulates western US ozone air quality in spring via deep stratospheric intrusions, *Nature Communications*, **6**, 7105, doi:10.1038/ncomms8105.

Liu, G., *et al.*, 2009: Ozone correlation lengths and measurement uncertainties from analysis of historical ozonesonde data in North America and Europe, *J. Geophys. Res.*, **114**, D04112, doi:10.1029/2008JD010576.

Logan, J.A., 1994: Trends in the vertical distribution of ozone: An analysis of ozonesonde data, *J. Geophys. Res.*, **99**(D12), 25553–25585, doi:10.1029/94JD02333.

Logan, J.A. 1999a:, An analysis of ozonesonde data for the lower stratosphere: Recommendations for testing models, *J. Geophys. Res.*, **104**(D13), 16151–16170, doi:10.1029/1999JD900216.

Logan, J.A., *et al.*, 1999b: Trends in the vertical distribution of ozone: A comparison of two analyses of ozonesonde data, *J. Geophys. Res.*, **104**(D21), 26373–26399, doi:10.1029/1999JD900300.

Logan, J.A., *et al.*, 2012, Changes in ozone over Europe: Analysis of ozone measurements from sondes, regular aircraft (MOZAIC) and alpine surface sites, *J. Geophys. Res.*, **117**, D09301, doi:10.1029/2011JD016952.

Long, C.S., *et al.*, 2017: Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP), *Atmospheric Chemistry and Physics*, **17(23)**, 14593–14629, doi:10.5194/acp-17-14593-2017.

Maillard Barras, E., Ruffieux, D., and Hocke, K., 2009: Stratospheric ozone profiles over Switzerland measured by SOMORA, ozonesonde, and MLS/Aura satellite, *Int. J. Remote Sens.*, **30,** 4033–4041, doi:10.1080/01431160902821890

McDermid, I.S., *et al.*, 1998:, OPAL: Network for the Detection of Stratospheric Change ozone profiler assessment at Lauder, New Zealand 2. Intercomparison of revised results, *J. Geophys. Res.*, **103(D22)**, 28693–28699, doi:10.1029/98JD02707.

McLinden, C.A., and Fioletov, V., 2011: Quantifying stratospheric ozone trends: Complications due to stratospheric cooling, *Geophys. Res. Lett.*, **38,** L03808, doi:10.1029/2010GL046012.

McPeters, R.D., *et al.*, 1999: Results from the 1995 stratospheric ozone profile intercomparison at Mauna Loa, *J. Geophys. Res.*, **104**, 30505–30514.

McPeters, R.D., *et al.*, 2013: The version 8.6 SBUV ozone data record: An overview, *J. Geophys. Res. Atmos.*, **118**, 8032–8039, doi:10.1002/jgrd.50597.

Mégie, G., and Menzies, R.T., 1980: Complementarity of UV and IR differential absorption lidar for global measurements of atmospheric species, *Appl. Opt.*, **19**, 1173-1183 doi: 10.1364/AO.19.001173.

Millán, L.F., *et al.*, 2016: Case studies of the impact of orbital sampling on stratospheric trend detection and derivation of tropical vertical velocities: solar occultation vs. limb emission sounding, *Atmos. Chem. Phys.*, **16**, 11 521–11 534, doi:10.5194/acp-16-11521-2016.

Misios, S., *et al.*, 2016: Solar signals in CMIP-5 Simulations: Effects of Atmosphere-Ocean Coupling, *Q. J. R. Meteorol. Soc.*, QJ-15-0113, **142**(695), doi:10.1002/qj.2695, 928-941.

Molina, M.J. and Rowland, F.S., 1974: Stratospheric sink for chlorofluoromethanes: chlorine atom catalysed destruction of ozone. *Nature*, Vol. **249**(5460), p. 810-812, doi:10.1038/249810a0.

Moreira, L., *et al.*, 2015: Trend analysis of the 20-year time series of stratospheric ozone profiles observed by the GROMOS microwave radiometer at Bern, *Atmos. Chem. Phys.*, **15**, 10999-11009, doi:10.5194/acp-15-10999-2015.

Morgenstern, O., *et al.*, 2017: Review of the global models used within phase 1 of the Chemistry–Climate Model Initiative (CCMI), *Geosci. Model Dev.*, **10**, 639-671, doi:10.5194/gmd-10-639-2017.

Morris, G.A., *et al.*, 2013: On the use of the correction factor with Japanese ozonesonde data, *Atmos. Chem. Phys.*, **13**, 1243-1260, doi:10.5194/acp-13-1243-2013.

Moy, L., *et al.*, 2017: Altitude registration of limb-scattered radiation, *Atmos. Meas. Tech.*, **10**, 167-178, doi:10.5194/amt-10-167-2017.

Nair, P.J., *et al.*, 2012: Relative drifts and stability of satellite and ground-based stratospheric ozone profiles at NDACC lidar stations, *Atmos. Meas. Tech.*, **5**, 1301-1318, doi:10.5194/amt-5-1301-2012.

Nair, P.J., *et al.*, 2013: Ozone trends derived from the total column and vertical profiles at a northern mid-latitude station, *Atmos. Chem. Phys.*, **13**, 10373-10384, doi:10.5194/acp-13-10373-2013.

Nair, P.J., *et al.*, 2015: Subtropical and midlatitude ozone trends in the stratosphere: Implications for recovery. *J. Geophys. Res. Atmos.*, **120**, 7247–7257. doi: 10.1002/2014JD022371.

Neal, R.M. 1993: Probabilistic inference using Markov chain Monte Carlo methods, *Technical Report*, **CRG-TR-93-1**, University of Toronto, Toronto.

Nedoluha G.E., *et al.*, 2015: Unusual stratospheric ozone anomalies observed in 22 years of measurements from Lauder, New Zealand, *Atmos. Chem. Phys.*, **15**, 6817–6826, doi:10.5194/acp-15-6817-2015.

Newman, P.A., *et al.*, 2007: A new formulation of equivalent effective stratospheric chlorine (EESC), *Atmos. Chem. Phys.*, **7**, 4537-4552, doi:10.5194/acp-7-4537-2007.

Oman, L.D., *et al.*, 2013: The ozone response to ENSO in Aura satellite measurements and a chemistry-climate simulation, *J. Geophys. Res.*, **118**, 965–976, doi:10.1029/2012JD018546.

Pelon, J., Godin, S., and Mégie, G., 1986: Upper stratospheric (30–50 km) lidar observations of the ozone vertical distribution, *Journal of Geophysical Research: Atmospheres*, **91**, 8667-8671, 10.1029/JD091iD08p08667.

Petropavlovskikh, I., Bhartia, P.K., and DeLuisi, J., 2005: New Umkehr ozone profile retrieval algorithm optimized for climatological studies, *Geophys. Res. Lett.*, **32**, L16808, doi:10.1029/2005GL023323.

Petropavlovskikh, I., *et al.*, 2009: Effect of the out-of-band stray light on the retrieval of the Umkehr Dobson ozone profiles, *Int. J. of Remote Sens.*, **30 (24)**, 6461 – 6482, doi: 10.1080/01431160902865806.

Petropavlovskikh, I., *et al.*, 2011: Sensitivity of Dobson and Brewer Umkehr ozone profile retrievals to ozone cross-sections and stray light effects. *Atmos. Meas. Tech.*, **4,** 1-29, doi: 10.5194/amtd-4-1-2011.

Pougatchev, N.S., Connor, B.J., and Rinsland, C.P., 1995: Infrared measurements of the ozone vertical distribution above Kitt Peak, *J. Geophys. Res.*, **100**(D8), 16689–16697, doi:10.1029/95JD01296.

Prais, S.J. and Winsten, C.B., 1954: Trend estimators and serial correlation, *Cowles Commission discussion paper*, **383**, 1-26, Chicago.

Rahpoe, N., *et al.*, 2015: Relative drifts and biases between six ozone limb satellite measurements from the last decade, *Atmos. Meas. Tech.*, **8**, 4369-4381, doi:10.5194/amt-8-4369-2015.

Randel, W.J. and Wu, F., 2007: A stratospheric ozone profile data set for 1979–2005: Variability, trends, and comparisons with column ozone data, *J. Geophys. Res.*, **112**, D06313, doi:10.1029/2006JD007339.

Randel, W.J., *et al.*, 2009: An update of observed stratospheric temperature trends, *J. Geophys. Res.*, **114**, D02107, doi:10.1029/2008JD010421.

Randel, W.J., *et al.*, 2016: Stratospheric temperature trends over 1979–2015 derived from combined SSU, MLS, and SABER satellite observations, *J. Climate,* **29**, 4843–4859, doi:10.1175/JCLI-D-15-0629.1.

Rodgers, C.D., 2000: Inverse Methods for Atmospheric Sounding: Theory and Practice, Series on Atmospheric, *Oceanic and Planetary Physics*, Vol. **2**, World Scientific Publishing Co., Singapore.

Savin, N.E. and White, K.J., 1978: Estimation and testing for functional form and autocorrelation: A simultaneous approach, *Journal of Econometrics*, **8**(1), 1-12, doi:10.1016/0304-4076(78)90085-4.

Shepherd, T.G., 2008: Dynamics, stratospheric ozone, and climate change, *Atmosphere-Ocean,* **46:1**, 117-138, doi:10.3137/ao.460106.

Smit, H.G.J., *et al.*, 2007: Assessment of the performance of ECC-ozonesondes under quasi-flight conditions in the environmental simulation chamber: Insights from the Jülich Ozone Sonde Intercomparison Experiment (JOSIE), *J. Geophys. Res.*, **112**, D19306, doi:10.1029/2006JD007308.

Smit, H.G.J., and ASOPOS-panel, 2012a: Quality Assurance and Quality Control for Ozonesonde Measurements in GAW, WMO *Global Atmosphere Watch report series*, No. **201**, World Meteorological Organization, Geneva, available at **http://www.wmo.int/pages/prog/arep/gaw/documents/GAW_201_30_Sept.pdf**.

Smit, H.G.J., and the O3S-DQA-Panel (Ozone Sonde Data Quality Assessment), 2012b: 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/~deshler/NDACC_O3Sondes/O3s_DQA/O3S-DQA-Guidelines%20Homogenization-V2-19November2012.pdf**.

Sofieva, V.F., *et al.*, 2014: On sampling uncertainty of satellite ozone profile measurements, *Atmos. Meas. Tech.*, **7**, 1891–1900, doi:10.5194/amt-7-1891-2014.

Sofieva, V.F., *et al.*, 2017: Merged SAGE II, Ozone_cci and OMPS ozone profiles dataset and evaluation of ozone trends in the stratosphere, *Atmos. Chem. Phys.,* **17,** 12533-12552, doi:10.5194/acp-17-12533-2017.

Solomon, S., *et al.*, 1986: On the depletion of Antarctic ozone, *Nature*, **321**, 755–758, doi: 10.1038/321755a0.

Solomon, S., *et al.*, 1998: Ozone depletion at mid-latitudes: Coupling of volcanic aerosols and temperature variability to anthropogenic chlorine, *Geophys. Res. Lett.*, **25**, 11, 1944-8007, doi:10.1029/98GL01293.

SPARC, 1998: SPARC-IOC-GAW Assessment of Trends in the Vertical Distribution of Ozone. N. Harris, R. Hudson and C. Phillips (eds.), SPARC report No.**1**, WMO Global Ozone Research and Monitoring Project Report No. **43**, available at: **www.sparc-climate.org/publications/sparc-reports/**.

SPARC, 2013, SPARC Report on the Lifetimes of Stratospheric Ozone-Depleting Substances, Their Replacements, and Related Species. Ko, M.K.W., Newman, P. A., Reimann, S., and Strahan, S. E. (eds.), SPARC Report No. **6**, WCRP-15/2013, available at: **www.sparc-climate.org/publications/sparc-reports/**.

Steinbrecht, W., *et al.*, 1998: Correlations between tropopause height and total ozone: Implications for long-term changes, *J. Geophys. Res.*, **103**, 19 183–19, 192, doi:10.1029/98JD01929.

Steinbrecht, W., *et al.*, 2006: Long-term evolution of upper stratospheric ozone at selected stations of the Network for the Detection of Stratospheric Change (NDSC), *J. Geophys. Res.*, **111**, D10308, doi:10.1029/2005JD006454.

Steinbrecht, W., *et al.*, 2009: Ozone and temperature trends in the upper stratosphere at five stations of the Network for the Detection of Atmospheric Composition Change, *Int. J. of Remote Sens.*, **30**:15-16, 3875-3886, doi:10.1080/01431160902821841.

Steinbrecht, W., *et al.*, 2017: An update on ozone profile trends for the period 2000 to 2016, *Atmos. Chem. Phys.*, **17**, 10675-10690, doi:10.5194/acp-17-10675-2017.

Sterling, C.W., *et al.*, 2018: Homogenizing and estimating the uncertainty in NOAA's long-term vertical ozone profile records measured with the electrochemical concentration cell ozonesonde, *Atmos. Meas. Tech.* **11**, 3661-3687, doi:10.5194/amt-11-3661-2018.

Stolarski, R.S. and Frith, S.M., 2006: Search for evidence of trend slow-down in the long-term TOMS/SBUV total ozone data record: the importance of instrument drift uncertainty, *Atmos. Chem. Phys.*, **6**, 4057-4065, doi:10.5194/acp-6-4057-2006.

Stolarski, R., *et al.*, 2006: Trends in Stratospheric Ozone: Lessons Learned from a 3D Chemical Transport Model, *J. Atmos. Sci.*, **63**, 1028–1041, doi:10.1175/JAS3650.1.

Stübi R., *et al.*, 2008: In-flight comparison of Brewer-Mast and electrochemical concentration cell ozonesondes, *J. Geophys. Res.*, **113**, D13302, doi:10.1029/2007JD009091.

Studer, S., *et al.*, 2013: Intercomparison of stratospheric ozone profiles for the assessment of the upgraded GROMOS radiometer at Bern, *Atmos. Meas. Tech.*, **6,** 6097–6146, doi:10.5194/amtd-6-6097-2013.

Studer, S., *et al.*, 2014: A climatology of the diurnal variations in stratospheric and mesospheric ozone over Bern, Switzerland , *Atmos. Chem. Phys.*, **14,** 5905–5919, doi:10.5194/acp-14-5905-2014.

Tarasick, D.W., *et al.*, 2016: A re-evaluated Canadian ozonesonde record: measurements of the vertical distribution of ozone over Canada from 1966 to 2013, *Atmos. Meas. Tech.*, **9**, 195-214, doi:10.5194/amt-9-195-2016.

Tegtmeier, S., *et al.*, 2013: SPARC Data Initiative: A comparison of ozone climatologies from international satellite limb sounders, *J. Geophys. Res. Atmos.*, **118**, 12,229–12,247, doi:10.1002/2013JD019877.

Terao, Y., and Logan, J.A., 2007: Consistency of time series and trends of stratospheric ozone as seen by ozonesonde, SAGE II, HALOE, and SBUV(/2), *J. Geophys. Res.*, **112**, D06310, doi:10.1029/2006JD007667.

Thomason, L.W., *et al.*, 2018: A global space-based stratospheric aerosol climatology: 1979–2016, *Earth Syst. Sci. Data,* **10**, 469-492, doi:10.5194/essd-10-469-2018.

Thompson, A.M., *et al.*, 2007: Southern Hemisphere Additional Ozonesondes (SHADOZ) 1998-2004 tropical ozone climatology: 3. Instrumentation, station-to-station variability, and evaluation with simulated flight profiles, *J. Geophys. Res.*, **112,** D03304, doi:10.1029/2005JD007042.

Thompson, D.W.J. and Wallace, J.M., 2000: Annular Modes in the Extratropical Circulation. Part I: Month-to-Month Variability, *J. Climate*, **13,** 1000–1016, doi:10.1175/1520-0442.

Thompson, D.W.J. and Solomon, S., 2002: Interpretation of Recent Southern Hemisphere Climate Change, *Science*, **296**, 895–899, doi:10.1126/science.1069270.

Thompson, D.W.J., *et al.*, 2012: The mystery of recent stratospheric temperature trends, *Nature*, **491**(7426), 692–697, doi:10.1038/nature11579.

Tie, X., and Brasseur, G., 1995: The response of stratospheric ozone to volcanic eruptions: Sensitivity to atmospheric chlorine loading, *Geophys. Res. Lett.*, **22**, 3035–3038, doi:10.1029/95GL03057.

Toohey, M., *et al.*, 2013: Characterizing sampling biases in the trace gas climatologies of the SPARC Data Initiative, *J. Geophys. Res. Atmos.*, **118**, 11,847–11,862, doi:10.1002/jgrd.50874.

Tummon, F., *et al.*, 2015: Intercomparison of vertically resolved merged satellite ozone data sets: interannual variability and long-term trends, *Atmos. Chem. Phys.*, **15**, 3021-3043, doi:10.5194/acp-15-3021-2015.

Van Malderen, R., *et al.*, 2016: On instrumental errors and related correction strategies of ozonesondes: possible effect on calculated ozone trends for the nearby sites Uccle and De Bilt, A*tmos. Meas. Tech.*, **9**, 3793-3816, doi:10.5194/amt-9-3793-2016.

Verhoelst, T., *et al.*, 2015: Metrology of ground-based satellite validation: co-location mismatch and smoothing issues of total ozone comparisons, *Atmos. Meas. Tech.*, **8**, 5039-5062, doi:10.5194/amt-8-5039-2015.

Vigouroux, C., *et al.*, 2015: Trends of ozone total columns and vertical distribution from FTIR observations at eight NDACC stations around the globe, *Atmos. Chem. Phys.*, **15,** 2915-2933, doi:10.5194/acp-15-2915-2015.

Wang, H.J., Cunnold, D.M., and Bao, X., 1996: A critical analysis of Stratospheric Aerosol and Gas Experiment ozone trends, *J. Geophys. Res.*, **101**(D7), 12495–12514, doi:10.1029/96JD00581.

Weber, M., Rahpoe, N., and Burrows, J.P., 2016: Stability requirements on long-term (satellite) ozone observations and their implication for trend detection, *Quadrennial Ozone Symposium of the International Ozone Commision*, **QOS2016-280**.

Weber, M., *et al.*, 2018: Total ozone trends from 1979 to 2016 derived from five merged observational datasets – the emergence into ozone recovery, *Atmos. Chem. Phys.*, **18**, 2097–2117, doi:10.5194/acp-18-2097-2018.

Weiss, A.K., *et al.*, 2001: Chemical and dynamical contributions to ozone profile trends of the Payerne (Switzerland) balloon soundings, *J. Geophys. Res.*, **106**, D19, 2156-2202, doi:10.1029/2000JD000106.

Wild, J.D., *et al.*, 2019: A coherent ozone profile dataset from SBUV, SBUV/2: 1979 to 2017, in preparation.

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**, 6611–6636, doi:10.1002/2016JD026403.

Witte, J.C., *et al.*, 2018: First Reprocessing of Southern Hemisphere ADditional OZonesondes (SHADOZ) Profile Records: 3. Uncertainty in Ozone Profile and Total Column, *J. Geophys. Res. Atmos.,* **123**, doi:10.1002/2017JD027791.

WMO (World Meteorological Organization), 1986: Atmospheric Ozone 1985, Assessment of Our Understanding of the Processes Controlling Its Present Distribution and Change, Global Ozone Research and Monitoring Project Report No. **16**, Geneva, Switzerland, available at: **https://www.esrl.noaa.gov/csd/assessments/ozone/1985/report.html**.

WMO (World Meteorological Organization), 2007: Scientific Assessment of Ozone Depletion: 2006, Global Ozone Research and Monitoring Project - Report No. **50**, 572 pp., Geneva, Switzerland, available at: **https://www.esrl.noaa.gov/csd/assessments/ozone/2006/**.

WMO (World Meteorological Organization), 2011: Scientific Assessment of Ozone Depletion: 2010, Global Ozone Research and Monitoring Project - Report No. **52**, 516 pp., Geneva, Switzerland, available at: **https://www.esrl.noaa.gov/csd/assessments/ozone/2010/**.

WMO (World Meteorological Organization), 2014: Scientific Assessment of Ozone Depletion: 2014, Global Ozone Research and Monitoring Project-Report No. **55**, Geneva, Switzerland, available at: **https://www.esrl.noaa.gov/csd/assessments/ozone/2014/**.

WMO (World Meteorological Organization), 2018: Scientific Assessment of Ozone Depletion: 2018, Global Ozone Research and Monitoring Project-Report No. **58**, 588pp, Geneva, Switzerland, available at: **https://www.esrl.noaa.gov/csd/assessments/ozone/2018/**.

Wolter, K., and Timlin, M.S., 2011: El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext), *Intl. J. Climatology*, **31**, 1074-1087.

Zawada, D.J., *et al*., 2018: Tomographic retrievals of ozone with the OMPS Limb Profiler: algorithm description and preliminary results, *Atmos. Meas. Tech*., **11**, 2375-2393, doi:10.5194/amt-11-2375-2018.

Zerefos, C.S., *et al*., 2012: Evidence of a possible turning point in solar UV-B over Canada, Europe and Japan, *Atmos. Chem. Phys*., **12**, 2469-2477, doi:10.5194/acp-12-2469-2012.

Zerefos, C., *et al*., 2018: Representativeness of single lidar stations for zonally averaged ozone profiles, their trends and attribution to proxies, *Atmos. Chem. Phys*., **18**, 6427-6440, doi:10.5194/acp-18-6427-2018.

Zitto, M.E., *et al*., 2016: 110 years of temperature observations at Orcadas Antarctic Station: multidecadal variability, *Int. J. Climatol*., **36**, 809–823, doi:10.1002/joc.4384.