National Aeronautics and Space Administration
Goddard Earth Science Data Information and Services Center (GES DISC)

README Document for Nitrogen Dioxide Surface-Level Annual Average Concentrations

Last Revised January 30, 2023

Goddard Earth Sciences Data and Information Services Center (GES DISC)
http://disc.gsfc.nasa.gov
NASA Goddard Space Flight Center
Code 619
Greenbelt, MD 20771 USA

Prepared By:

Gaige Kerr	**Susan Anenberg**
Name The George Washington University Department of Environmental and Occupational Health	Name The George Washington University Department of Environmental and Occupational Health
January 30, 2023	
Date	

Reviewed By:

Reviewer Name	Date
Reviewer Name GES DISC GSFC Code 619	Date

Goddard Space Flight Center
Greenbelt, Maryland

Revision History

Revision Date	Changes	Author
August 22, 2022	Changed the GES DISC organization code from 610.2 to 619	Lena Iredell

Table of Contents

1.0 Introduction

This document provides basic information for using nitrogen dioxide (NO_2) surface-level annual average products.

The NO_2 surface-level annual average concentrations consist of products generated for the focus of understanding the global distribution, trends, and health impacts of this health-harming air pollutant at a high spatial resolution.

1.1 Dataset Description

An existing global NO_2 concentration dataset, representing a 2010-2012 average, at 100 m resolution and produced from a land use regression model (Larkin *et al.*, 2017; https://doi.org/10.1021/acs.est.7b01148) was adjusted to correct for a high bias in rural areas, and then further adapted to other years in the 1990s-2010s using Earth Observations from the Ozone Monitoring Instrument (OMI) aboard NASA's Aura satellite. The output dataset represents surface-level annual average NO_2 concentrations at a spatial resolution of 0.0083˚ x 0.0083˚ (~1 km^2).

1.2 Data Disclaimer

Data should be used with care and proper citations. Additional details beyond those provided in this document regarding how NO_2 concentrations were estimated, quality assurance/quality control to characterize the uncertainties in these estimates, and examples of their use in applications related to health and air quality can be found in the citation in Section 1.2.1.

1.2.1 Data Citation and Acknowledgment

Cite the following reference when using the dataset: Anenberg, S. C., Mohegh, A., Goldberg, D. L., Kerr, G. H., Brauer, M., Burkart, K., et al. (2022). Long-term trends in urban NO_2 concentrations and associated paediatric asthma incidence: estimates from global datasets. *Lancet Planetary Health.* 6(1): e49-58. https://doi.org/10.1016/s2542-5196(21)00255-2.

1.2.2 Contact Information

For more information on the dataset, contact Susan Anenberg (sanenberg@gwu.edu).

2.0 Data Organization

The data consist of annual averages compiled into annual files.

2.1 File Naming Convention

Global annual average NO_2 concentrations data files are named in accordance with the following convention:

SurfaceNO2_0.0083deg_YYYY.nc

where "YYYY" is the four-digit year [1990, 1995, 2000, 2005-2020].

Filename example: SurfaceNO2_0.0083deg_2020.nc

2.2 File Format and Structure

The data files are in NetCDF-4 format. NetCDF is a set of software libraries and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data that was developed by UCAR/Unidata (http://doi.org/10.5065/D6H70CW6) https://www.unidata.ucar.edu/software/netcdf/. These files follow the CF 1.5 conventions.

2.3 Key Science Data Fields

The key science data field is "*SFC_NITROGEN_DIOXIDE_CONC*" which provides estimates of global annual average surface-level concentrations of NO_2 at ~0.01 ˚ x 0.01˚ (~1 km^2) resolution.

3.0 Data Contents

3.1 Data Set Attributes (File Metadata)

In addition to SDS arrays containing variables and dimension scales, global metadata is also stored in the files. Some metadata are required by standard conventions, some are present to meet data provenance requirements and others as a convenience to users of Nitrogen Dioxide Surface-Level Annual Average Concentrations products. A summary of global attributes present in all files is shown in Table 1.

Global Attribute	Description	Type
GDAL_AREA_OR_POINT	indicates whether a pixel value should be assumed to represent sampling over the region of the pixel or a point sample at the center of the pixel	string
Conventions	Climate and Forecast (CF) metadata conventions	string
GDAL	Geospatial Data Abstraction Library	string

	(GDAL) version	
title	description of dataset	string
NCO	netCDF4 Operations version information	string
institution	specifies where the original data were produced	string
source	the method of production of the original data including the model used to generate the data	string
references	published references describing the data	string
ShortName	short name of the data type (collection)	string
LongName	long name of the collection	string
DataSetQuality	information on QA/QC underlying dataset	string
IdentifierProductDOI	Digital Object Identifier	string
ProcessingLevel	degree of data processing applied to Earth Observations data	string
RangeBeginningDate	begin data of data	string
RangeBeginningTime	begin time of data	string
RangeEndingDate	end date of data	string
RangeEndingTime	end time of data	string
SouthBoundingCoordinate	southbound latitude	string
NorthBoundingCoordinate	northbound latitude	string
WestBoundingCoordinate	westbound longitude	string
EastBoundingCoordinate	eastbound longitude	string
history	audit trail for modifications to the original data	string
VersionID	version number	string

Table 1. Global metadata attributes associated with each SDS.

3.2 Variable Data Attributes

Global Attribute	Description	Type	Extra Dimensions
SurfaceNO2	Annual average, surface-level concentrations of NO2, units of parts per billion by volume	32-bit floating-point	None

3.3 Geolocation Fields

Global Attribute	Description	Type
lat	latitude of the center of the cell, units of degrees north	32-bit floating-point
lon	longitude of the center of the cell, units of degrees east	32-bit floating-point

A list of key metadata fields can be found in Table 1. Global attributes in a Nitrogen Dioxide Surface-Level Annual Average Concentrations file can be viewed with *ncdump* software:
ncdump –h -c < filename>

3.4 Dimensions

Global Attribute	Description	Dimensions
lat	latitude of the center of the cell, starting at -60.0462, by 0.0083 degree increments	16200
lon	longitude of the center of the cell, latitude starting at -179.5471, by 0.0083 degree increments	43080

4.0 Products/Parameters

4.1 Data Fields

Data Field Name	Long_Name/Description	Type	Dimensions	Undefined Value	Units
SurfaceNO2	Annual average, surface-level nitrogen dioxide concentrations	32-bit floating-point	(lat, lon)	-999.0	ppbv

4.2 Fill Values

Variable Type	Fill Value
float	-999.0

5.0 Options for Reading the Data

5.1 Command Line Utilities

5.1.1 ncdump

The ncdump tool can be used as a simple browser for HDF data files, to display the dimension names and sizes; variable names, types, and shapes; attribute names and values; and optionally, the values of data for all variables or selected variables in a netCDF file. The most common use of ncdump is with the –h option, in which only the header information is displayed.

ncdump [-c|-h] [-v ...] [[-b|-f] [c|f]] [-l len] [-n name] [-d n[,n]] filename
Options/Arguments:
[-c] Coordinate variable data and header information
[-h] Header information only, no data
[-v var1[,...]] Data for variable(s) <var1>,... only data
[-f [c|f]] Full annotations for C or Fortran indices in data
[-l len] Line length maximum in data section (default 80)
[-n name] Name for netCDF (default derived from file name)
[-d n[,n]] Approximate floating-point values with less precision filename File name of input netCDF file
(https://www.unidata.ucar.edu/software/netcdf/workshops/2011/utilities/Ncdump.html)

5.1.2 HDFView

HDFView is a Java based graphical user interface created by the HDF Group which can be used to browse HDF files. The utility allows users to view all objects in an HDF file hierarchy which is represented as a tree structure. Additional information about HDFView can be found at https://support.hdfgroup.org/products/java/hdfview/ and for HDF at https://portal.hdfgroup.org/display/support

5.2 Tools/Programming

The product files can be read and queried using the NetCDF4 library and tools maintained by Unidata (http://www.unidata.ucar.edu/software/netcdf/). Support for reading NetCDF is offered in many programming languages, including Python, Matlab, IDL, C/C++ and Fortran. NetCDF4 files are legal HDF5 files with additional bookkeeping information managed by the NetCDF4 library. It is therefore possible to inspect and copy data out of the NetCDF4 files by using the HDF5 utilities and libraries maintained by the HDF Group (https://www.hdfgroup.org/products/hdf5_tools/index.html) or by using the HDF5 interface in your favorite programming language. However, the two libraries should not be considered fully interchangeable.

Matlab users should note that the Matlab NetCDF4 interface is currently (as of version R2017a) not able to read attributes that are string arrays, and will throw an exception if that is attempted.

5.2.1 Python

The following code snippet shows how to read the variable lat, lon, and no2 from the dataset with the name “filename”. Also shown are some basic information about the size of the variables arrays.

```
import netCDF4 as nc

nc_fid = nc.Dataset(filename ,mode='r',format='NETCDF4')

# Read in the variables
lat = nc_fid.variables['lat'][:]
lon = nc_fid.variables['lon'][:]
no2 = nc_fid.variables['SurfaceNO2'][:]

# Print out the minimum, maximum, and dimensions for the three variables
print("-- lat  Min/Max values", lat[:].min(), lat[:].max())
print("lat.shape:", lat.shape)
print("-- lon Min/Max values:", lon[:].min(), lon[:].max())
print("lon.shape:", lon.shape)
```

```
print("—no2 Min/Max values:", no2 [:].min(), no2 [:].max())
print("no2.shape:", no2.shape)
```

6.0 GES DISC Data Services

If you need assistance or wish to report a problem:
Email: gsfc-dl-help-disc@mail.nasa.gov
Voice: 301-614-5224
Fax: 301-614-5268
Address:
Goddard Earth Sciences Data and Information Services Center NASA Goddard Space Flight Center Code 619 Greenbelt, MD 20771 USA

6.1 How To Articles

The GESDISC web site contains many informative articles under the "How To Section", "FAQ" (frequently asked questions), "News", "Glossary", and "Help" . A sample of these articles includes:

Earthdata Login for Data Access

How to Download Data Files from HTTPS Service with wget

How to Obtain Data in NetCDF Format via OpeNDAP

Quick View Data with Panoply

How to Read Data in NetCDF Format with R

How to Read Data in HDF-5 or netCDF Format with GrADS

How to read and plot NetCDF MERRA-2 data in Python

How to Subset Level-2 Data

How to use the Level 3 and 4 Subsetter and Regridder

7.0 Acknowledgments

The creation of this dataset was funded by NASA and the Health Effects Institute.

8.0 References

Anenberg, S. C., Mohegh, A., Goldberg, D. L., Kerr, G. H., Brauer, M., Burkart, K., et al. (2022). Long-term trends in urban NO2 concentrations and associated paediatric asthma incidence: estimates from global datasets. *Lancet Planetary Health*. 6(1): e49-58. https://doi.org/10.1016/s2542-5196(21)00255-2

Larkin, A., Geddes, J. A., Martin, R. V., Xiao Q., Liu, Y., Marshall, J. D., Brauer, M., Hystad, P. (2017). Global land use regression model for nitrogen dioxide air pollution. *Environmental Science & Technology*. 51(12): 6957-6964. https://doi.org/10.1021/acs.est.7b01148