Computers & Geosciences 48 (2012) 57–66

Contents lists available at SciVerse ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

# AMFIC-WSDB: A web database for hosting and easy retrieval of atmospheric data from satellites

A.K. Georgoulias [a,*], K.A. Kourtidis [a], E. Kosmidis [b], T. Despotakis [b], P. Symeonidis [c]

[a] Laboratory of Atmospheric Pollution and Pollution Control Engineering of Atmospheric Pollutants, Department of Environmental Engineering, Democritus University of Thrace, 67100 Xanthi, Greece
[b] DOTSOFT S.A., 3 Kountourioti street, 54625 Thessaloniki, Greece
[c] Draxis Environmental Technologies Ltd., 63 Mitropoleos street, 54623 Thessaloniki, Greece

ARTICLE INFO

*Article history:*
Received 12 January 2012
Received in revised form
1 May 2012
Accepted 2 May 2012
Available online 15 May 2012

*Keywords:*
Web database
AMFIC
Satellite observations
Methane
SCIAMACHY
MODIS

ABSTRACT

This paper presents the unique features and functionality of the AMFIC Web Satellite Database (AMFIC-WSDB) which was developed within the Air quality Monitoring and Forecasting In China (AMFIC) European Union (EU) project. The method used for the analysis of the SCIAMACHY ENVISAT $CH_4$, CO, $NO_2$ and $SO_2$ satellite data that emerged from the AMFIC work packages, in order to be hosted by the AMFIC-WSBD, is described. Today, there is an increased demand for satellite data not only from specialized users but also from researchers and public institutions for research, management, information and education purposes. The AMFIC-WSDB has been developed to address the need of those users for fast and easy access to atmospheric composition data. Together with the web interface, it constitutes an integrated system, hosting spatially and temporally homogenized satellite products. The main idea behind the AMFIC-WSDB (http://www.amfic.satellite-earth-simulator.com/) is to offer timeseries ascii files through a set of interactive maps. This makes the database ideal not only for those interested in global datasets but also for users interested in specific geographical spots. Data from other state-of-the-art environmental satellite sensors (e.g. MODIS, OMI and GOME-2) could be easily integrated in the database in the future.

© 2012 Elsevier Ltd. All rights reserved.

## 1. Data formats, datasets and methodology

### 1.1. HDF and netCDF files

The increasing demand for satellite and ground-based atmospheric composition data has led to a significant number of simple web databases of which satellite datasets are available. Most of them offer data in binary formats (HDF, HDF-EOS, netCDF, etc.) which makes their use possible only by a specialized group of the scientific community. The availability of data is in this form because the amount of information is enormous and a way must be found so that the files have an acceptable size. With the constant improvement of the resolution of the various satellite instruments the availability of data in this form is now well established. Hierarchical Data Format (HDF) is a datafile format designed by the National Center for Supercomputing Applications (NCSA) to assist users in the storage and manipulation of scientific data across diverse operating systems and machines. A library of routines and a set of utility programs and tools for creating and using HDF files were developed. HDF is the most popular file type used for the storage of both gridded (level-3) and single pixel swath satellite data (level-2). In 1993, NASA chose the Hierarchical Data Format Version 4 (HDF4) (HDF Reference Manual, 2012) to be the official format for all data products derived by the Earth Observing System (EOS). The newer version of HDF files is HDF5 (HDF5 User's Guide, 2011). With the launch of TERRA satellite in 1999, HDF4 became a critical part of the EOS Data and Information System (EOSDIS) for storing, and distributing a huge amount of data. NASA developed a special set of datafiles based on HDF, called HDF-EOS, and created an HDF-EOS library, which reads and writes these datatypes in HDF4 (HDF-EOS4) or HDF5 (HDF-EOS5).

Another very popular file format used for gridded data sets in environmental sciences is network Common Data Form (netCDF) (The NetCDF Users Guide, 2011). A set of interfaces offering data access and a freely available collection of libraries for C, FORTRAN, C++, Java, and other programming languages (developed within Unidata program at the University Corporation for Atmospheric Research-UCAR) support this type of data. The interfaces, libraries, and format support the creation, access, and sharing of scientific data. Recently, a netCDF4 format has been released which unlike the standard netCDF3 format is based on HDF5 format allowing for the use of multiple dimensions. The majority

* Corresponding author. Tel.: +30 25410 79383.
*E-mail address:* argeor@env.duth.gr (A.K. Georgoulias).

0098-3004/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2012.05.001

of the data that have been analyzed in order to be hosted from the AMFIC-WSDB base belong to the file formats presented above.

### 1.2. Analyzed datasets

The first data that were analyzed emerged from the various work packages of the EU project Air Quality Monitoring and Forecasting in China (AMFIC). The project lasted for 2 years (2008–2009) focusing on the development of an integrated information system for monitoring and forecasting tropospheric pollutants over China. The system uses satellite observations, in situ air quality measurements and modeling to generate consistent air quality information over China. The aims and results of the project can be found on http://www.amfic.eu. As part of this project the Laboratory of Atmospheric Pollution and Pollution Control Engineering of Atmospheric pollutants of the Democritus University of Thrace along with the companies of DOTSOFT S.A. and DRAXIS Environmental Technologies Ltd., developed and maintain a web database which hosts several AMFIC satellite data products. The original AMFIC datafiles had to be spatially and temporally homogenized and be brought to a common format that enables their inclusion in the database. A set of programming tools were developed within the framework of the project for the automatic conversion and analysis of the original data whose size was ~41 Gigabytes (GBs). Today, 4 different products from the Scanning Imaging Absorption Spectrometer for Atmospheric CHartographY (SCIAMACHY) satellite instrument aboard ENVISAT (Bovensmann et al., 1999) are available through the database at 3 different resolution modes ($10^\circ \times 10^\circ$, $5^\circ \times 5^\circ$ and $1^\circ \times 1^\circ$) and for 2 regions of interest (global and the greater China region) depending on the product. WFM-DOAS v1.0 $CH_4$ dry air columnar data ($XCH_4$) for 2003 and 2004 (Schneising et al., 2009), WFM-DOAS v0.6 CO total column data for 2003 and 2004 (Buchwitz et al., 2007), DOAS v1.0.3 $SO_2$ total column data for the period 2005–2007 (van Geffen, 2008) and TM4NO2A version 1.10 $NO_2$ tropospheric and total column data for the period 2003–2008 (Boersma et al., 2004) are available. $CO_2$ dry air columnar data ($XCO_2$) for 2003 and 2004 (Schneising et al., 2008) were also analyzed offline despite not being included in the database. The original $CH_4$, $CO_2$ and CO data were available in monthly ascii files (·was files) including level-2b single pixel swath measurements. The original $NO_2$ data were available in daily HDF4 files including level-2 single pixel swath measurements. For all these data the SCIAMACHY ground pixel typically covers an area of $30 \times 60$ km$^2$ except in the case of CO where the typical swath resolution is ($30 \times 120$ km$^2$). The original $SO_2$ data were available in daily HDF4 files including level-3 gridded ($0.25^\circ \times 0.25^\circ$) measurements.

The methodology developed for the analysis of the AMFIC datasets was also used for the analysis of data from the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard EOS TERRA and AQUA satellites (Salomonson et al., 1989) and meteorological data from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis project (Kalnay et al., 1996) in order to show that these datasets could easily be integrated in the AMFIC-WSDB in the future. This was done for 3 main reasons: (1) the method is suitable for the analysis of timeseries (Georgoulias et al., 2011; Georgoulias and Kourtidis, 2011, 2012), (2) the analyzed data can be easily integrated in an advanced version of the AMFIC web database which is scheduled for the future and (3) the data can be easily used for scientific as well as educational purposes. The combined size of the original data and the data that emerged from our analysis was ~2 Terabytes (TBs). The original MODIS TERRA (2/2000-2/2009) and AQUA (7/2002-12/2008) daily global HDF-EOS4 datafiles included joint aerosol and cloud properties level-3 gridded ($1^\circ \times 1^\circ$) data from the Collection 005. The HDF files for both TERRA (MOD08_D3) and AQUA (MYD08_D3) were acquired through Level 1 and Atmosphere Archive and Distribution System (LAADS; http://ladsweb.nascom.nasa.gov). The aerosol properties extracted from the original HDF files include the following: aerosol optical depth at 550 nm ($AOD_{550}$; land and ocean), standard deviation of $AOD_{550}$ (land and ocean), ratio of the $AOD_{550}$ attributed to fine mode (diameter $<1$ μm) aerosols (land and ocean), aerosol mass concentration (land and ocean), Angström exponent (0.47–0.66 μm land) and (0.55–0.865 μm ocean), cloud condensation nuclei (ocean only), effective particle radius (ocean only), aerosol asymmetry factor at 550 nm (ocean only) and aerosol cloud mask fraction (land and ocean). The cloud properties extracted from the original HDF files include the following: cloud optical depth (total=liquid, ice, mixed and undetermined phase), ice cloud optical depth, liquid cloud optical depth, cloud droplet effective radius (total), ice cloud droplet effective radius, liquid cloud droplet effective radius, cloud fraction (total), ice cloud fraction, liquid cloud fraction, cloud mask cloud fraction for daytime from the infrared algorithm (it takes into account all the pixels and not only those with cloud optical properties retrieval), cloud top pressure for daytime, cloud water path (total), ice cloud water path and liquid cloud water path. Details on the $1^\circ \times 1^\circ$ MODIS joint dataset can be found in Hubanks et al. (2008).

MODIS TERRA level-2 aerosol properties data (Levy et al., 2009), at a resolution of $10 \times 10$ km$^2$ pixel size at nadir, for the time period 2/2000–2/2010 were also acquired from LAADS to be used into a more focused regional study for the region of Central Europe (2.5°W–22.5°E, 42.5°N–55°N, Georgoulias and Kourtidis, 2012). The same properties as in the case of level-3 were extracted from the original MODIS TERRA (MOD04_L2) and AQUA (MYD04_L2) HDF files. 144 level-2 HDF files (granules) are available on a daily basis each one covering a different scene on the earth. Daily netCDF datafiles from the NCEP/NCAR reanalysis project were analyzed for the period 2000–2009 producing a full dataset of meteorological parameters (e.g. sea level pressure, wind speed and direction at the 850, 700 and 500 mbar pressure levels, surface wind speed and direction, atmospheric temperature and vertical air velocity at 10 pressure levels ranging from 1000 to 100 mbar and columnar precipitable water amount) on a $2.5^\circ \times 2.5^\circ$ resolution.

### 1.3. Methodology and tools

The main idea behind the method applied here was to use global daily data files of various formats in order to create timeseries columnar ascii files for every spot on a regular grid. First, a suite of scripts we developed in Interactive Data Language (IDL) allows for the automatic conversion of the binary formats into columnar latitude/longitude/parameter files. These files are thereafter read by FORTRAN scripts and the initial pixel values are assigned to geographical grid cells taking into account the center of each pixel. This procedure (hereafter denoted as gridding) was done for each satellite or reanalysis product separately. One columnar timeseries (e.g. Year, Month, Day, Parameter, etc) ascii file for each grid cell was created including all the pixels centered within the grid cell. As it was discussed in Section 1.2., the size of the grid used in the analysis of the greenhouse data from the AMFIC was $10^\circ \times 10^\circ$ and $5^\circ \times 5^\circ$ on a global scale and $1^\circ \times 1^\circ$ for the greater China region (20°–55°N, 75°–135°E). The corresponding grid resolution was $1^\circ \times 1^\circ$ for the $NO_2$ tropospheric (Fig. 1) and total column data and $0.25^\circ \times 0.25^\circ$ for the $SO_2$ total column data covering the region of China in both cases. All these files were then uploaded on the AMFIC-WSDB server becoming

Download English Version:

https://daneshyari.com/en/article/6923248

Download Persian Version:

https://daneshyari.com/article/6923248

Daneshyari.com