

Discovering New Global Climate Patterns: Curating a 21-Year High Temporal (Hourly) and Spatial (40km) Resolution Reanalysis Dataset

Chung-Yi Hou*, Robert Dattore**, Grace Peng**

*Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, Champaign, IL, United States. <hou@illinois.edu>

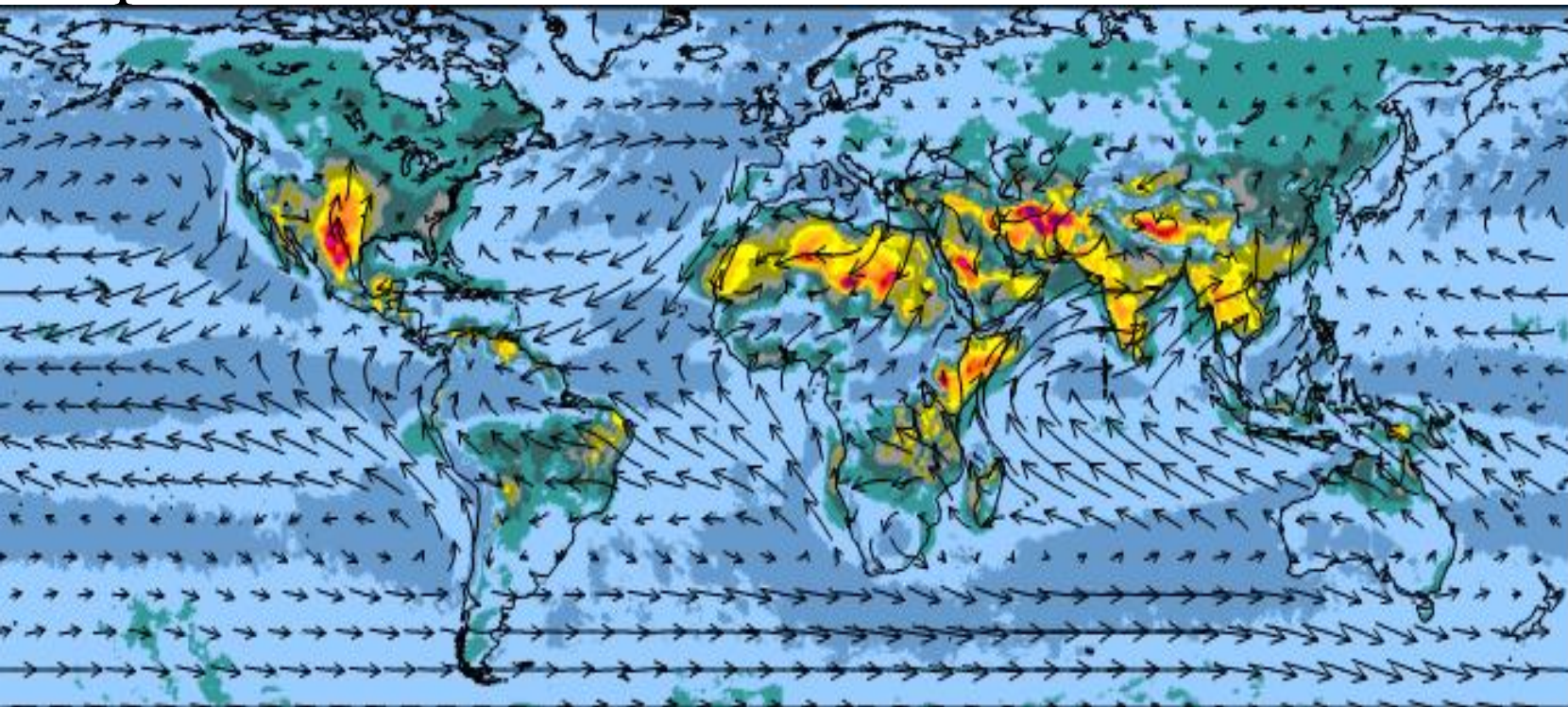
**Computational & Information Systems Laboratory – Data Support Section, National Center for Atmospheric Research, Boulder, CO, United States. <dattore@ucar.edu>, <grace@ucar.edu>

Introduction

The National Center for Atmospheric Research (NCAR) Global Climate Four-Dimensional Data Assimilation (CFDDA) Hourly 40km Reanalysis dataset is a dynamically downscaled dataset with high temporal and spatial resolution that was created using NCAR's CFDDA system.

The dataset contains three-dimensional hourly analyses in netCDF format for the global atmospheric state from 1985 to 2005 (a total of 184,080 files) on a 40km horizontal grid (0.4° grid increment) with 28 vertical levels, providing good representation of local forcing and the diurnal variation of processes in the planetary boundary layer.

Making the dataset publicly available, accessible, and usable will provide a significant resource with greater diurnal cycle details to allow and promote studies of new climate characteristics.



Global Wind Pattern Generated by CFDDA¹.

Methods

The project aimed to make the entire dataset available, accessible, and usable and focused on the following three areas of the data curation process:

Verify Data Quality

- Consistency:** Comparison of file content and structure between the first data file and the last data file.
- Validation:** Comparison of the content and descriptions provided by the data files and the dataset's user documentation.
- Compliance:** Confirmation of data file format and its conformance to Climate and Forecast (CF) Metadata Conventions.



Harvest Metadata Descriptions

- Tool:** Learning NCAR Computational & Information Systems Laboratory (CISL) Research Data Archive's (RDA) software and processes for automatically gathering metadata from the data files.
- Format:** Understanding the rationale for CISL RDA's usage of Global Change Master Directory (GCMD) scientific keywords and customized metadata format over standards, such as ISO 19115.
- Content:** Creating the metadata content through CISL RDA's Metadata Manager.



Document Provenance Information

- The Data Curation Profiles²:** Utilize the instructions and worksheets provided by the toolkit to interview the scientists in order to organize and document the provenance relating to the dataset.

Results

The dataset is currently available for access and use at National Center for Atmospheric Research (NCAR) Computational & Information systems Laboratory's (CISL) Research Data Archive (RDA - <http://rda.ucar.edu/datasets/ds604.0/>).

The dataset's landing page provides three unique tabs with each tab featuring specific information regarding the dataset.

Dataset Home Page "Description" Tab

CISL Research Data Archive
Managed by NCAR's Data Support Section
Data for Atmospheric and Geosciences Research

Go to Dataset: [ds604.0](#)

Home Find Data Ancillary Services About/Contact Data Citation Web Services For Staff

NCAR Global Climate Four-Dimensional Data Assimilation (CFDDA) Hourly 40 km Reanalysis
ds604.0

For assistance, contact Grace Peng (303-497-1218).

Description Data Access Documentation

Abstract: Please note that this is a preliminary release of the dataset. The data page is still under construction.

NCAR Global Climate Four-Dimensional Data Assimilation (CFDDA) Hourly 40 km downscaled dataset with high temporal and spatial resolution that was created the CFDDA system is the 5th generation Pennsylvania State University / NCAR.

The dataset contains three-dimensional hourly analyses in netCDF format for the on a 40 km horizontal grid (0.4 degree grid increment) with 28 vertical levels. A hourly, meso-beta scale reanalyses files for all the days of the months within the representation of local forcing and the diurnal variation of processes in the planetary boundary layer. This was the only reanalysis dataset with full three-dimensional fields.

The dataset was generated by continuously assimilating standard surface and hourly and 6-hourly while upper-air measurements consisted primarily of standard hourly intervals typically. Initial land surface conditions were based on the NAS (GLDAS) on a 1 degree by 1 degree latitude-longitude grid, compiled using the values for substrate soil moisture and temperature, ground skin temperature, at surface temperatures (SSTs) were specified by the National Centers for Environmental Prediction and daily SST dataset defined on a 0.25 degree by 0.25 degree grid.

Temporal Range: 1985-01-01 00:00 +0000 to 2005-12-31 23:00 +0000

Usage Restrictions: RELEASE POLICY: All uses of this dataset must be properly cited. See "How to Cite This Dataset" for more information. Additionally, notify Grace Peng (303-497-1218) of tech use the NCAR Global Climate Four-Dimensional Data Assimilation (CFDDA) Hourly 40 km Reanalysis dataset.

Variables:

Air Temperature	Cloud Liquid Water/Ice	Heat Flux
Hydrostatic Pressure	Incoming Solar Radiation	Land Surface Temperature
Planetary Boundary Layer Height	Precipitable Water	Precipitation
Sea Level Pressure	Snow Depth	Soil Moisture
Surface Air Temperature	Surface Winds	Upper Level Winds

Vertical Levels: See the detailed metadata for level information

Data Types: Grid

Spatial Coverage: Longitude Range: Westernmost=180W Easternmost=180E
Latitude Range: Southernmost=90S Northernmost=90N
Detailed coverage information

Data Contributors: [UCAR/NCAR/RAL](#) | [UCAR/NCAR/MMM](#) | [DOO/DTRA](#)

Related Resources: MMS: Mesoscale Model - Core of the CFDDA System
CDO: Collection of command line Operators to manipulate and analyse Climate and NWP model Data
NCO: Shell-command style, stand-alone programs that take netCDF, HDF, and/or DAP files as input and output the results in text, binary, or netCDF formats
Ncview: netCDF visual browser allowing users to visually inspect NetCDF data files
Panoply: Cross-platform application that plots geo-gridded and other arrays from netCDF, HDF, GRIB, and other datasets
IDV: Java-based software framework from Unidata for analysing and visualising geoscientific data

Publications: Please see Publications_References.pdf under the Documentation tab.

How to Cite This Dataset: [RIS](#)

Rife, D. L., J. O. Pinto, A. J. Monaghan, C. A. Davis, and J. R. Hannan. 2014. *NCAR Global Climate Four-Dimensional Data Assimilation (CFDDA) Hourly 40 km Reanalysis*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. <http://rda.ucar.edu/datasets/ds604.0/>. Accessed "dd mmm yyyy".
Please fill in the "Accessed" date with the day, month, and year (e.g., 5 Aug 2011) you last accessed the data from the RDA.

Bibliographic citation shown in (Federation of Earth Science Information Partners (ESIP)) style

Get a customized data citation

Total Volume: 300.71 MB

Data Formats: netCDF4

Related RDA Datasets: 091.0 - NCEP/DOE Reanalysis II
353.4 - NCEP ADP Operational Global Upper Air Observations, December 1972 - February 2007
464.0 - NCEP ADP Operational Global Surface Observations, February 1975 - February 2007

More Details: View more details for this dataset, including dataset citation, data contributors, and other detailed metadata

Data Access: Click the [Data Access](#) tab here or in the navigation bar near the top of the page

The Research Data Archive is managed by the Data Support Section of the Computational and Information Systems Laboratory at the National Center for Atmospheric Research in Boulder, Colorado. NCAR is sponsored by the National Science Foundation.

Current Status:

- The Home Page for the dataset has been completed and published with all the essential metadata.

Future Updates:

- Additional information will be added as needed to the sections, including Related Resources, Publications, and Related RDA Datasets, as the iterations of curation activities continue.

Dataset's Second Page "Data Access" Tab

To get data files, do one of the following:

- Select one or more files and [Create](#) a script to read them from the NCAR HPSS

[Reset](#) checkboxes Range selection is ☐ off ☐ on Total volume of selected files: 0.00 Mbytes

HPSS Filename	Data Format	File Contents	Valid Date Range	Size Mbytes	More Details
/FS/DSS/DS604.0/1985/19850101/CFDDA_19850101000000.mdv.nc	netCDF4	244 Grids	1985-01-01 00:00 1985-01-01 00:00	153.81	Details
/FS/DSS/DS604.0/2005/20051231/CFDDA_20051231230000.mdv.nc	netCDF4	216 Grids	2005-12-31 23:00 2005-12-31 23:00	146.90	Details

To get data files, do one of the following:

- Select one or more files and [Create](#) a script to read them from the NCAR HPSS

[Reset](#) checkboxes Range selection is ☐ off ☐ on Total volume of selected files: 0.00 Mbytes

Current Status:

- The first and the last file of the dataset have been archived successfully into the RDA.
- Archiving these two files helped test the data ingest and metadata harvest processes.

Future Updates:

- All the remaining data files will be ingested into the RDA and be made available, accessible, and usable.

Dataset's Third Page "Documentation" Tab

NCAR Global CFDDA (Climate Four-Dimensional Data Assimilation) 40 km Reanalysis
ds604.0

For assistance, contact Grace Peng (303-497-1218).

Description Data Access Documentation

[Web server holdings]

[View Selected Files/Get As a Tar File](#) [Perl Download Script](#) [Curl Download Script](#)

- Total 9 Files (10.5M) are listed below
- Click a file name to download a single file
- Currently 0 File selected (Clear Selection)

INDEX	File Name	Description	Size	Document Format	Date Archived
1	Publications_References.pdf	Related Publications and References for the CFDDA Dataset	65.3K	pdf	07/22/2014
2	CFDDA_Verification_and_Validation.pdf	Verification and Validation of the CFDDA Dataset	6.2M	PDF	07/16/2014
3	WRF_Obs_Nudging.pdf	Information about WRF Obs Nudging	9.5K	pdf	07/18/2014
4	CFDDA_Acknowledgement.pdf	Acknowledgment for the CFDDA's Data Creator, Sponsor, ...	140.5K	pdf	07/22/2014
5	File_Naming_Convention_and_Format.pdf	Explanation of the File Naming and Format	144.3K	pdf	07/22/2014
6	Use_of_MMS.pdf	Use of MMS in CFDDA generation	62.1K	pdf	07/22/2014
7	terrain.nc.nc	CFDDA Dataset's Terrain File	3.2M	netCDF	07/22/2014
8	CFDDA_User_Documentation_Rev2.pdf	User Documentation for the CFDDA Dataset	504.0K	PDF	07/22/2014
9	Initial_Input_Conditions_and_Data_Processing_Info.pdf	Additional Information regarding Input Datasets, Input ...	113.2K	pdf	07/22/2014

[View Selected Files/Get As a Tar File](#) [Perl Download Script](#) [Curl Download Script](#)

Current Status:

- Information considered to be most critical and helpful to the dataset's usability has been published.

Future Updates:

- Additional documents sharing the dataset team's project-specific knowledge and dataset users' experiences, such as Frequently Asked Questions, will also be made available.

Acknowledgments

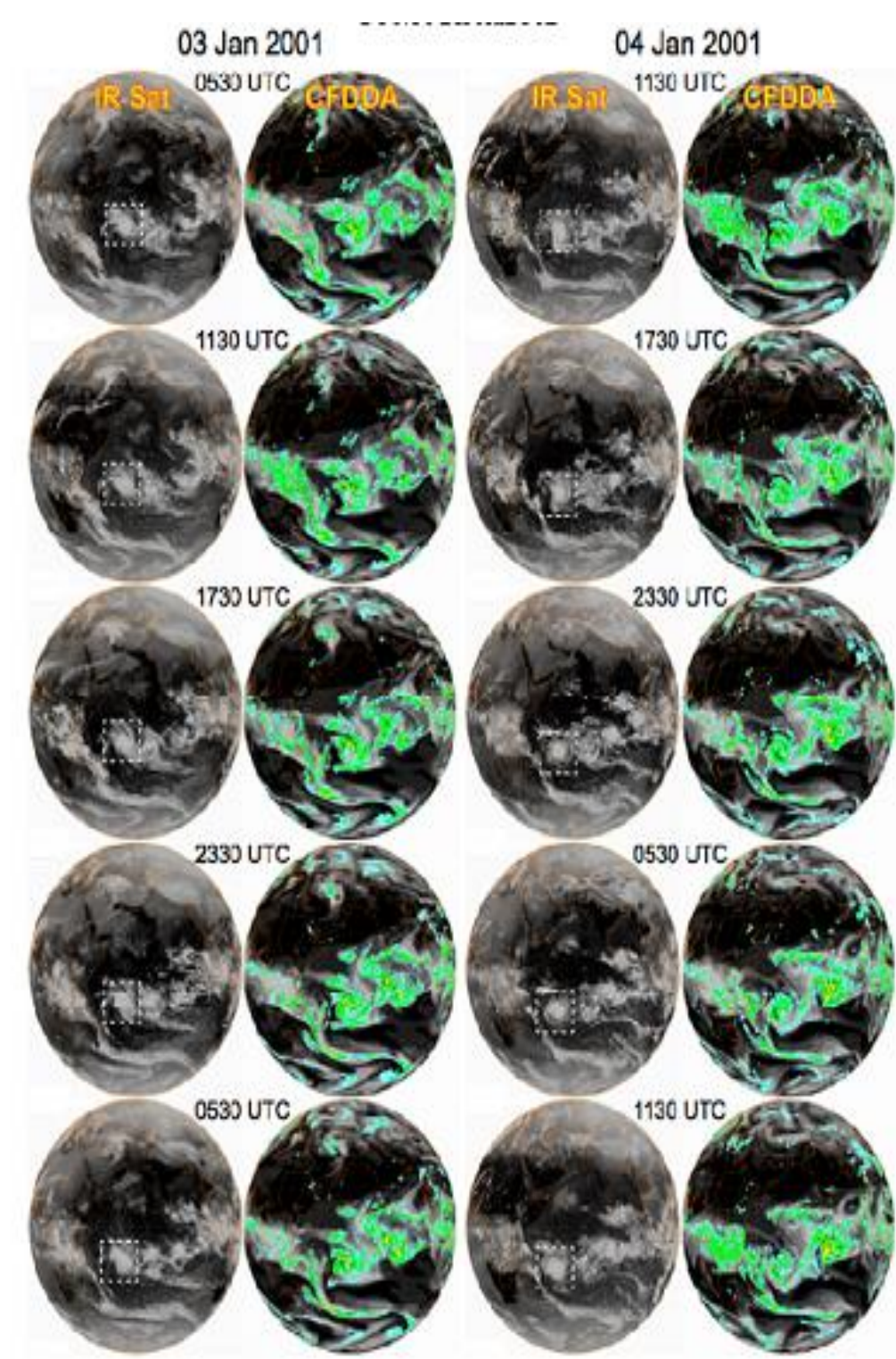
The author would like to thank her mentors, Steve Worley, Terri Betancourt, Andy Monaghan, Daran Rife, Matthew Mayernik, Chi-Fan Shih, Zaihua Ji, and Doug Schuster of National Center for Atmospheric Research (NCAR), Carole Palmer and Cheryl Thompson of Illinois, and the Data Curation Education in Research Centers (DCERC) project, funded by Institute of Museum and Library Services, for providing and supporting this learning opportunity.

Conclusions

By the time the data curation process started for the dataset, it had been five years since the data files were generated. In addition, at the end of the 18-month project duration, although the PI had generated a user document, the document had not been maintained or updated. Furthermore, not only had the Principal Investigator (PI) moved to a new research institution, but also the remaining team members had been reassigned to other projects. **As a result, the curation process showed that it was crucial to involve data curators during each stage of the data curation life cycle, such as the one proposed by the Digital Curation Center (DCC)³, instead of at the end.**

Particularly, in the areas of verifying data quality, harvesting metadata descriptions, and documenting provenance information, data curator's skill and knowledge could help the team make decisions, such as file format and structure and workflow documentation, that could have significant, positive impact on the likelihood and the ease of the dataset's management and long term preservation. For example, through the Data Curation Profiles Toolkit's guidelines, the discussions with the Project Manager revealed important information regarding the dataset that was helpful both to promote the data's usability and to enhance preservation planning.

With the proper resources invested in the dataset's curation process, especially with attention paid to the dataset's data quality, metadata description, and provenance information, this dataset is now prepared and can offer much potential to help with new climate pattern discovery.



Time series of regional to global cloud patterns over the eastern hemisphere for 0530 UTC 03 January 2001 to 1130 UTC 05 January 2001. IR satellite imagery from (first and second columns) Meteosat and (second and fourth columns) using the CFDDA dataset is shown. Relative humidity at 500 hPa (gray shades) is shown as a proxy for clouds, and surface rainfall accumulation (colors; green colors denote low intensity, and yellow and orange colors denote high intensity). Dashed white boxes track the visually defined position of tropical cyclone Ando over the Indian Ocean⁴.