NOMADS User Guide
V1.0

Table of Contents

Introduction

The model data repository at NCEI contains both deep archived (offline) and online model data. We provide a variety of ways to access our weather and climate model data. You can access the online data using traditional access methods (web-based or FTP), or you can use open and distributed access methods promoted under the collaborative approach called the NOAA National Operational Model Archive and Distribution System (NOMADS).

On the Data Products page you are presented with a table that contains basic information about each dataset, as well as links to the various services available for each dataset. The Plot/FTP option takes you into the "NOMADS Web Interface" which allows you to browse the inventories, retrieve model data, create plots, download plots, and FTP data. This page describes how to use the the different access methods and features provided by the NOMADS interface.

NCEI maintains a repository of weather model data sets as well as distributed access servers to these data sets. We provide access to weather model forecast data in near real-time as well as access to historical data in the online and offline (tape archive) repositories.

To view more detailed information regarding what datasets we have available, first visit the NCEI Model Data Inventories page to browse our inventories and locate datasets of interest. After you determine which datasets you want to further explore, visit the Data Products page to review the methods and features available to access the data. To directly read and analyze model data or retrieve model data, follow the instructions provided there.

Explanation of "Online" and "Offline" Data

Due to the vast amount of model data that NOMADS archives, it is not possible to keep it all online, all the time. Eventually, online data needs to be pushed back into NCEI's High Performance Storage System (HPSS).

Upon entering a dataset's entry point from the NOMADS Data Products pages, the following captions will be displayed under "Data Availability:"

- Online
- Archived Offline

Both captions are followed by their corresponding date ranges. "Online" data is stored directly on the back-end data server and is readily accessible using the NOMADS web interface: HTTP, FTP, GDS, and other NOMADS services. Offline data is stored in NCEI's deep archive and must first be retrieved from the HPSS then placed into a public staging area to make it accessible. The data will remain staged for a specified, limited amount of time. The following two sections explain the offline order interface in more detail.

Obtaining Offline Data

Because data are listed "Archived Offline," this does not indicate that you cannot access these data. However, more steps and a waiting period are required to obtain it. To submit an order for offline data, enter the date range of interest and click "Build Order." The web interface will recognize when a request includes an offline date range, and require you to enter a valid email address. This email address will be used only for notifying your of the order's status. The system does not check for invalid email addresses.

To begin the offline data request process, enter a valid email address in the text box then click the "Confirm Offline Order" button to the left. Doing so will initiate automated processing on our server to pull the data out of NCEI's deep archive and place it in the staging area set aside for users. When the order is picked up by our system, an email will be sent to the email address you entered echoing details about the order and providing a URL to check on the order and eventually retrieve the data when the process completes. Another email message will be sent upon completion, notifying you that the order is complete. Occasionally missing or corrupt files will fail to be retrieved from the deep archive: a list of such files is appended to the completion email.

The system processes requests in the order they are received. The time it takes to complete the order depends on how many other orders are in front of yours, how large your order is, the HPSS processing load, and to a lesser extent, the NOMADS server load.

Offline Order Limitations

Staging

When offline data has been staged, it will appear the same as online data for a period of five days (from the time the order was completed). If you cannot or do not access and retrieve the data within this allotted time, the staging area will be purged and you will need to resubmit the offline data request. A limited amount of space is allocated to all users; when this system is in demand, the space can fill up quickly. A low disk space warning has been put in place to notify users before they enter their email to begin an offline order. Orders placed when this warning is displayed cannot be guaranteed to complete successfully. Please be considerate of other users, all orders larger than 50% of the total staging area volume may be terminated without notice. If you have an urgent need for a large volume of offline data, please contact nomads.ncdc@noaa.gov (link sends e-mail). In some cases, we will not be able to service large requests.

Throttles

We have put the following measures in place to prevent overload of the offline order system, and prevent orders from failing to complete, or "completing" unsuccessfully:

- Orders of high-resolution numerical weather prediction (NWP) data which span more than 15 days will be automatically truncated to 15 days starting with the data request start date.

- When there are 40 offline orders in the order queue, the system will block order submission until orders can complete.
- If staging area disk space is critically low, submission of orders will be blocked until the space frees up. In some cases, this can take up to five days for enough staged data to expire.
- If staging area disk space becomes critical with orders still queued, processing of orders will halt until disk space frees up.
- Lastly, if the system is experiencing technical difficulties for any reason, or we feel the system is being overloaded (too many orders in a short period of time), submission of orders will be blocked until the issues are resolved.

These throttles have worked well, however there are still instances where orders may fail. Most of the time this is due to something going wrong in the automated process which retrieves the data, or the HPSS being reset when an order is processing. If you receive a notification that your order was completed but data is missing, the best course of action in most cases is to resubmit the order. If they continue to fail, contact nomads.ncdc@noaa.gov (link sends e-mail). Certain files will fail repeatedly because they are missing from our archive.

Analysis-Only Requests

Often times there is a need for just the initial forecast time from a model. We have received so many of these requests we have decided, for the most popular datasets, to create a separate dataset comprised of just the 000 hour forecast files. Some of the 000 to 003-hour averaged fields from the 003 and 006-hours are also included. Please check the data products page carefully for these datasets before placing an offline order for just the 000 hour forecasts! Excessive 000 hour forecast requests from users for datasets containing an analysis only subset may be terminated.

Distributed Data Access

Functional Overview

The NOMADS framework is a distributed data system that promotes the combining of data sets between distant participants using open and common server software and methodologies. Users effectively access model and observational data and products in a flexible and efficient manner from archives or in real-time through existing Internet infrastructure.

The framework of the server and software used to manipulate the archive database and distribute the real-time data is the Open source Project for a Network Data Access Protocol (OPeNDAP) formally known as the Distributed Ocean Data System (DODS). OPeNDAP is a binary-level protocol designed for the transport of scientific data subsets over the Internet. Through OPeNDAP, applications can open and read subsets from remote data sets.

The Grid Analysis and Display System (GrADS) is an open source desktop tool that integrates the analysis and display of a wide variety geophysical data. OPeNDAP is an open source software project that provides a way to share data over the Internet. NOMADS combines these

technologies to greatly expand the utility of the client and server concept. OPeNDAP-enabled clients such as GrADS, Ferret, IDL, Matlab, and end-user programs linked with various OPeNDAP client-libraries, can access datasets and subsets of datasets over the network from OPeNDAP servers. The server can perform calculations over large amounts of data and only transmit (fairly small) results back to the client. This avoids transmitting large amounts of data to the client.

"Downloading data from the OPeNDAP server to your local system is generally a misuse of the OPeNDAP server. The OPeNDAP data server is like a library in the purist sense of the term. When you check out a book from the library, do you photocopy the entire book, return it, and then read your copy? It is difficult to overcome the habit of working with data that you don't actually have on your own disk. The capability to access and use gigabytes of binary data over the Internet has not existed before. It does not make sense for terabytes of disk across many institutes to be redundantly occupied by the same data that are freely available over the net."

Getting Started with OpeNDAP

An OPeNDAP Quick Start Guide (link is external)

Using a Web Browser with OPeNDAP (link is external)

Using a Spreadsheet Application with OPeNDAP (link is external)

Desktop Services ("Clients")

For data users, the currently available OPeNDAP servers include data formats for many desktop applications, and the user base for OPeNDAP enabled clients is expanding. NOMADS users have multiple options to access data from their desktops. The OPeNDAP web site (link is external) maintains a list of OPeNDAP-enabled analysis applications and libraries.

With a certain level of understanding of the OPeNDAP protocol, a user can also use a web browser or even a spreadsheet to access OPeNDAP data.

OPeNDAP Servers

A data provider can range from a single scientist and a laptop computer, to a major data center. To make data available under the NOMADS framework, a provider must install an OPeNDAP Server as described on the OPeNDAP web site (link is external).

Quick Start to Retrieve or Plot Data

Getting Started

For new users, we recommend starting with the NOMADS Data Products page, also referred to as the NOMADS Web Interface. From the Web Interface, you can sort products either by data provider or data type using the various links. The Interface allows you to

1. Use FTP and HTTP to retrieve data
2. Remotely create and download graphics
3. Utilize distributed access to create products locally on your desktop

The sub-sections of this user guide below tell you how to quickly begin using these features.

Also, we have developed additional resources and links to help you quickly learn and efficiently use distributed access methods within the NOMADS OPeNDAP framework. These references include getting started guides, tutorials, and information about available OPeNDAP-enabled clients and servers. Please see the Introduction to Distributed Data Access section of this guide.

Using FTP to Retrieve Data

Your files are prepared for retrieval in two stages. First they are copied to a private FTP server and then to the public FTP server at ftp.ncdc.noaa.gov. Depending on the volume of data, it may take several minutes to complete the transfer. They will be deposited in directory /pub/has/nomads. All files from your dataset collection will have an extension which is your unique four digit order number. They will remain on our anonymous server for three days. You can use the instructions below to retrieve your files.

1. Connect to the FTP server
2. Use anonymous as the user name
3. Use your email address as the password
4. Turn off interactive mode: prompt off
5. Change to binary mode: binary on
6. Change directory to the NOMADS order data directory. The order number is sent to you in an email and is four digits: cd pub/has/nomads/model-<your order number>
7. Then use multi-get to grab all of your files: mget *

Here is some example output of what you may see during a command line FTP connection to the NCEI FTP server using Windows or DOS with an order number of 1337.

```
C:\> ftp ftp.ncdc.noaa.gov
User (nomads.ncdc.noaa.gov:(none)): anonymous
331 Anonymous login ok, send your complete email address as your password
Password: username@hostname.tld
230 Anonymous access granted, restrictions apply
ftp> prompt
```

```
Interactive mode Off .
ftp> binary
200 Type set to I
ftp> cd /pub/has/nomads/model-1337
250 CWD command successful
ftp> mget *
200 Type set to I
200 PORT command successful
150 Opening BINARY mode data connection for gfs_3_20080830_1200_000.grb2
(14400513 bytes)
226 Transfer complete
ftp: 14400513 bytes received in 4.56 Seconds 3155.93Kbytes/sec.
ftp>
```

To determine when your files are available for retrieval, you can monitor our ftp site.

Note for Advanced Weather Interactive Processing System (AWIPS) Workstation Users:

You can request model subsets or complete model runs via the NOMADS Web Interface. Since the current NCEI holdings contain the NOAAPort Broadcast model output, we expect that NWS and NOAAPort users will use this service for papers, review of synoptic conditions against model analysis and forecast products, and other analysis tasks. Users may elect to use the NOMADS distributed remote access methods using OpeNDAP and their selected desktop client (see Distributed Data Access); browse and plot selected output for retrieval; or FTP the raw GRIB using our FTP4u service.

Once you obtain the raw GRIB for display on an AWIPS workstation, you must convert the GRIB to the AWIPS style NetCDF format. Unidata (Kambic, Oram, et al.), and FSL (Ramer, et al.), developed tables for use with the gribtonc routines to convert GRIB to NetCDF. Please see Unidata NetCDF Libraries and Decoders for more information.

Using HTTP to Retrieve Data

NCEI provides a web server to allow you to browse our datasets, select files, and then use your web browser to download the files to your host computer. On NOMADS data products pages, select the "HTTP" link for the desired dataset. Navigate to the desired file. Do not left-click on the ".grb" or ".idx" files--these are binary files, and your browser might attempt to display them. To download the files, right-click on the filename. This will bring up a menu where you can choose to save the file. Typical menu choices are "*Save Link As*" or "*Save Target As*."

Using Wget to Retrieve Data

GNU Wget is a freely available network utility to retrieve files from the World Wide Web, using HTTP (Hyper Text Transfer Protocol) and FTP (File Transfer Protocol), the two most widely used Internet protocols. Since Wget is a non-interactive downloading utility, you can execute scripts containing Wget commands. For detailed instructions, search online for the GNU Wget Manual. On NOMADS data products pages, select the "http" link for the desired dataset. Navigate to the desired file. Do not left-click on the ".grb" or ".idx" files--these are binary files,

and your browser might attempt to display them. To download the files, use the following instructions.

1. In your browser, right-click on the filename you wish to download--this will bring up a menu where you can copy the link's name, or URL
2. Paste the URL into your Wget command string

Example Wget command string:

wget -O OUTPUT_LOCAL_FILEPATH
http://nomads.ncdc.noaa.gov/data/eta/200305/20030501/early-eta_212_20030501_0000_012.grb

Note: Some operating systems or shells require you to place Wget URL's within single quote characters.

Using GDS Clients to Plot Data

NCEI provides a GrADS OPeNDAP Server (GDS) to access model data. From the NOMADS data products pages, select the "gds" link for the desired dataset. Use the GDS to navigate to the GRIB file you want to access. Select the "info" link to find the name and URL of the data on our server you desire to access with your client running on your desktop.

When you give a URL to an OPeNDAP client, it actually sends a URL with one of the following six file extensions: .das, .dds, .dods, .info, .ver or .help. The extensions tell the OPeNDAP server which object to return. No extension defaults to OPeNDAP, and retrieves the data subset.

Each example below shows you how to access a single dataset from the Eta model.

Grid Analysis and Display System (GrADS) Example

gradsdods
sdfopen
http://nomads.ncdc.noaa.gov:9090/dods/NCDC_NOAAPort_ETA/200305/20030501/early-eta_212_20030501_0000_fff
display t2m

Unidata's Integrated Data Viewer (IDV) Example

Just paste the selected NOMADS URL into the IDV data loader.

http://nomads.ncdc.noaa.gov:9090/dods/NCDC_NOAAPort_ETA/200305/20030501/early-eta_212_20030501_0000_fff

MATLAB Example

To load data served by NOMADS GDS into MATLAB one needs the MATLAB-OPeNDAP Command Line Tool. Follow the instructions found at this link (link is external) to download and install the tool for the platform you are using. Make sure that the place in which the command line tool is put is included in both the system (unix/linux) path as well as the MATLAB path. The easiest way to do this is to simply startup MATLAB from the directory with the tool in it.

A MATLAB OPeNDAP Oceanographic Toolbox is also available. This toolbox provides Graphical User Interfaces (GUIs) to access a number of ocean data sets via OPeNDAP as well as tools to plot the acquired data. To download, see this site (link is external). To use some of the GUIs you will also need the "timeplt" package developed by Rich Signell of the USGS. This package is available at this site.

Below is an example MATLAB session using the MATLAB-OPeNDAP Command Line Tool to get a subset of data and plot them.

```
>> path(path,'/Applications/DODS/bin/ml-toolbox/')
>>loaddods('http://nomads.ncdc.noaa.gov:9091/dods/NCEP_NARR_DAILY/197912/1979
1210/narr-a_221_19791210_0000_000?clwmrprs[0:0][0:28:28][0:4:239][0:4:585]')
>> whos
Name Size Bytes Class
clwmrprs 60x147x2 141120 double array
lat 60x1 480 double array
lev 2x1 16 double array
lon 147x1 1176 double array
time 1x1 8 double array
Grand total is 17850 elements using 142800 bytes
>> imagesc(lon,lat,squeeze(clwmrprs(:,:,1)));set(gca,'ydir','normal')
>> max(clwmrprs(:))
ans =
9.9990e+20
>> min(clwmrprs(:))
ans =
1.0000e-12
>> nn=find(clwmrprs>1e+20);
>> clwmrprs(nn)=nan;
>> imagesc(lon,lat,squeeze(clwmrprs(:,:,1)));set(gca,'ydir','normal')
>> load coast
>> hold on
>> plot(long,lat,'w')
>> print -dpng /users/petercornillon/desktop/figure.png
<< END EXAMPLE >>
```

Ferret Example

```
set data
"http://nomads.ncdc.noaa.gov:9090/dods/NCDC_NOAAPort_ETA/200305/20030501/earl
y-eta_212_20030501_0000_fff"
```

Using the NOMADS Server to Create Plots

This web site provides a built in GrADS front-end application to allow users to select data, specify numerous plot options, and then use GrADS on our server to create plots. You can create contour maps, timeseries plots, latitude-longitude versus time plots, and animations. After viewing plots you can download the plot in either postscript format or GrADS metadata format. Use this procedure if you are unfamiliar with GrADS, but want to quickly create some plots:

1. Point your browser to the NOMADS data products pages.
2. Select the "plot/ftp" link for the desired dataset.
3. Wait for the system to build you collection of data (build time can vary immensely depending on which dataset you are requesting--anywhere from seconds to a few hours).
4. When the resultant page fully loads, click the "Plot" button in the lower right.

Advanced Data Access Methods

The NOMADS web site provides a variety of options to access our weather and climate model data. You can access the online data using traditional access methods (web-based or FTP), or you can use open and distributed access methods under the NOMADS collaborative approach. This section of the User Guide provides a description of the different functions you can use to get or view data. Some of the access methods described in this section are accessible through the NOMADS data products pages. However, not all access methods are available for all datasets.

The following access methods involve the use of scripts and common programs to automatically pull large amounts of data from the NOMADS data directory structure. It is necessary to have a working knowledge of the Practical Extraction and Report Language (Perl), and have access to a command-line computer system. While it is possible to convert these scripts for use on machines running Microsoft Windows, the current scripts and guide have not been designed with this operating system in mind.

Subsetting Parameters and Levels Using FTP4u

The NOMADS Web Interface provides a utility which allows users to select GRIB files from an order and extract parameters and vertical levels of interest. Then, our server pushes the data to NCEI's anonymous FTP server where you can retrieve your subsetted data. You can easily extract individual parameters across multiple days and model runs. If the data in your order is GRIB format (.grb), use this function to get small subsets of the data, rather than transferring the entire file.

To begin, go to a NOMADS data products page, select the "plot" or "ftp" link for the desired dataset, build a collection for the desired date range (maximum possible length of the collection depends on the dataset), then choose "Select files for FTP" at the bottom of the next page. If your order requires an offline data request, you will need to wait until this request is completed.

The FTP4u program allows you to quickly select the entire set of files, or a particular type of file, by entering "*" or "*.grb" respectively, into the GRIB Filter text box and clicking the button

beside it, then use the list of check boxes below to specify the parameters and levels you want. You can also subset out a spatial region of interest. Please note that subsetting Lambert Conformal grids will cause the GrADS control and index files that come with the order to become invalid. After the above options are set, enter your email address in the form at the bottom of the page and click "Start FTP." The script will then subset your files and inform you where and how to retrieve your data. Please be patient and wait until the entire page fully loads.

Obtain ASCII Data Using GDS

A web browser or "Wget" can be used to obtain ASCII data directly from the NCEI NOMADS GrADS Data Server (GDS). Both require you to construct an OPeNDAP URL with an OPeNDAP "constraint," which allows for both spatial or temporal subsetting.

The general form of an OPeNDAP constraint is below.

.ascii?VAR[t1:tInc:t2][z1:zInc:z2][Lat1:LatInc:Lat2][Lon1:LonInc:Lon2]

To demonstrate, refer to the following example.

Example: http://nomads.ncdc.noaa.gov/dods/NCEP_NARR_DAILY/199601/19960101/narr-a_221_19960101_0000_000.info

At the top of this ".info" page, you will find the OPeNDAP URL. The OPeNDAP constraint is to be appended to the end of the URL for spatial or temporal subsetting.

The OPeNDAP constraint dimensions (t|z|lat|lon) are all grid relative numbers. You need to reference the time, altitude, latitude,and longitude from the ".info" page to determine the numeric values you need to use. Translated values are given at the bottom of the output page. The time value in this section is the absolute number of days since year zero. The "VAR" must be one of the variables listed on the ".info" page (immediately below the dimension definition section). If the chosen variable is defined only on one level ("tmp2m", for example), the [z1:z2] must be excluded. The colon and second value in brackets may be excluded to hold that dimension constant. If the brackets (for any axis) contains three entries with two colons, the ":*Inc:" is the increment, or step, for the variable. For example, [0:2:10] is equivalent to the sequence (0,2,4,6,8,10).

Once you have composed a URL, you can then copy it to your web browser or use it as an argument to Wget to download the information.

Example OPeNDAP Constraints

Click on an example URL in the table below to send a request for data to the NOMADS GDS which will then send the results back to your browser. North American Regional Reanalysis (NARR) data from January 1, 1996 is used in these examples. These examples are designed to demonstrate the versatility of this access method.

Example	Constraint	URL	Description
1	.ascii?tmpprs[0][10][100][340]	Example 1 URL	Point: single 750 millibar Air Temperature Point at Lat/Lon 37.5 °N / 92.5 °W at 00UTC
2	.ascii?tmpprs[0][6]	Example 2 URL	2D grid slice: the entire 850 millibar Air Temperature Grid at 00UTC
3	.ascii?tmp2m[3][130][330]	Example 3 URL	Point: for variable with no vertical coordinate - 2 meter Air Temperature for Lat/Lon 48.75 °N / 96.25 °W at 09UTC
4	.ascii?ugrdprs[0:7][16][90][320]	Example 4 URL	3-hour timeseries at a point: 500 mb U-Wind values for 33.75 °N / 100 °W at each 3-hour time step
5	.ascii?ugrdprs[0:2:7][16][90][320]	Example 5 URL	6-hour timeseries at point: same as example 4, however, every other time step is skipped
6	.ascii?ugrdprs[7][0:28][120][120]	Example 6 URL	Vertical sounding at a point: U-Wind component at all vertical levels for Lat/Lon 45.0 °N / 175.0 °W, time 21UTC
7	.ascii?rh2m[2][70:100][140:160]	Example 7 URL	Spatial region subset for variable with no vertical coordinate: 2 meter Relative Humidity for a Lat/Lon subregion (26.25 - 27.5) °N / (167.5 - 160.0) °W, time 06UTC

To access data for other datasets in this manner, start with the GDS link for your dataset from a NOMADS data products page, then drop down the directory structure until you find an "Info" link. Use the metadata on the ".info" page to form the OPeNDAP constraint.

Determining Grid Relative x/y-Values for the OPeNDAP Constraint

If you want a specific latitude and longitude from a grid, the following instructions might help you.

From a GDS ".info" page under longitude and latitude, you will obtain information you need.

Example: -220°E to -0.625°E (586 points, avg. res. 0.375°)

Generalized to: init to final (span points, avg. res. aRes)

So, x can range from zero to span -1. To determine the grid relative points x and y, you need to pinpoint a particular longitude and latitude location of interest (you need to use decimal units, so

convert from degree/min/sec if needed) use the following linear equations below. (Note: int() = truncate all decimals.)

lon = init + aRes * x

(lon - init) = aRes * x

x = int((lon - Init(Lon)) / aRes(Lon))

y = int((lat - Init(Lat)) / aRes(Lat))

GrADS interpolates all gridded data it knows how to serve with GDS into a lat/lon grid, so this linear equation will always work. There is a chance for small map-projection interpolation errors whenever using the GDS with a dataset that uses a non-rectangular native grid. The Latitude and Longitude entries on the info page are not representative of non-rectangular native grids!

Determining Grid Relative t-Values for the OPeNDAP Constraint

The GDS uses an epoch-based time reference for its underlying time axis. This epoch begins at "0000-00-00 00:00:00" and has units of decimal days. Mapping a calendar date to the t can be rather tricky for datasets with a long time series.

On the GDS ".info" page, the "Time:" follows the same format as latitude and longitude (see the section above). In this case a t-value of zero equates to exactly the init time, spanning linear aRes epoch time increments up to final time, which is span - 1.

Scripting Wget in a Time Loop

This is a technique best used to obtain large numbers of whole files quickly. The method works with datasets which are accessible via HTTP or FTP. The process involves setting up a script which will loop through a range of dates, creating a URL with the date variables, then passing this URL to Wget.

Sample URL

http://nomads.ncdc.noaa.gov/data/gfs/200306/20030607/gfs-avn_201_20030607_0000_000.grb

URL Generalization (for Perl Variables)

$URL =
"http://${SERVER}/data/${MODEL}/${YYYYMM}/${YYYYMMDD}/${MODELNAME}_${GRID_NUM}_${YYYYMMDD}_{CYCLE_hr}00.grb"

Wget Usage (System/Shell Command)

wget -O [Local output file] $URL

SERVER varies depending on the dataset. See model Data Products and click on HTTP links to determine this.

YYYYMM and YYYYMMDD are the variables you will need to loop. This can be done using various iteration structures (nested for, nested foreach, while, etc. loops) in any scripting language. This guide will not attempt to explain exactly how to script this.

MODEL is the top level directory for the corresponding model system. This is not equal to MODELNAME and GRID_NUM. For example, MODEL eta can contain files early-eta_211* and meso-eta_218*. To determine the appropriate MODELNAME and GRID_NUM, see the NOMADS data products pages.

CYCLE_hr may require another nested loop if the model is run more then once a day.

To avoid overwriting output, [Local output file] should contain the YYYYMMDD and CYCLE_hr variables.

Are all the data you need in the same HTTP or FTP directory? If so, then you may use a single Wget command rather than writing a script. The general syntax is below.

wget -r [-lX] --no-parent -A[FILETYPE] -nd http://[SERVER NAME]/[PATH]

[SERVER NAME] and [PATH] define the URL to the desired directory.

[lX] Defines the directory level depth "X" you want Wget to scan. Example: [-l1]

The -A[FILETYPE] is optional, if omitted, Wget will attempt to download everything in the directory. If you use, for example, [-A.grb], then just the files ending with .grb will be downloaded.

Note that this may not work on directories forbidden by robots.txt

Mass GRIB Subsetting: Utilizing Partial-File HTTP Transfers

This recently implemented method allows you to install and run scripts on your local machine that will access and download parameter and vertical level subsets of certain NCEI-NOMADS datasets to your local disk. This method involves downloading three Perl scripts to your working area, then running a single command. Currently, this process is tested and works on UNIX/Linux systems only. A windows version is in the testing phase.

NCEI-NOMADS Datasets Currently Supported for Partial-File HTTP Transfers

- North American Regional Reanalysis (NARR)
- North American Mesoscale Model (NAM)
- Global Forecast System (GFS)
- Rapid Update Cycle (RUC) and Rapid Refresh (RAP)

Method

1. Visit and briefly review the following page: Fast Downloading of GRIB Files.
2. Ensure your system has the following dependencies installed: Perl and cURL.
3. Download the Perl scripts "get_inv.pl" and "get_grib.pl" from the web page referenced in step 1.
4. Add the two Perl scripts from step 3 to your ENV PATH.
5. Download the following Perl script get-httpsubset.pl.txt and rename it as "get-httpsubset.pl".
6. Run "get-httpsubset.pl" (see syntax and examples below).

Syntax

The "get-httpsubset.pl" Perl script can be run interactively (through a series of prompts) or by specifying a string of parameters with the command.

To run interactively use the following command.

- get-httpsubset.pl interactive

To run by specifying command-line parameters, use the following.

- get-httpsubset.pl <YYYYMMDDHH Start> <YYYYMMDDHH End> <GRIB Parameter names> <Levels> <output path> [Dataset] [OPTIONS]

YYYYMMDDHH Start and End: The date-time range for which you desire data. The program will exit if you get these incorrect.

GRIB Parameter Names: The list of parameters you wish to subset. These are the GRIB parameter codes. To select multiple parameters, just delimit them using the "-" character. Use "ALL" for all parameters.

Level: The vertical level(s) you wish to subset. Similar to the parameter names. Reference the third column of the snapshot Inventory. Substitute underscore for spaces (500_mb) for 500 millibars. "ALL" for all levels.

Output Path: This is the absolute path to the directory where you want the downloaded subset files to be saved. The program will automatically create a hierarchical date directory structure (YYYYMM/YYYYMMDD) under the given directory to ease storage of a large number of files.

Dataset (default = narr-a): Can be one of either narr-a, narr-b, narrmon-a, narrmon-b, narrmonhr-a, narrmonhr-b, gfs, nam, gfs-avn, or meso-eta. This will determine which dataset the script will download. The NARR, NAM, and GFS have different periods of record and the dates you select with YYYYMMDDHH1 and YYYYMMDDHH2 need to be consistent with the dataset you select here, otherwise no data will be found.

OPTIONS: [-nocheck] skips the dependency check--not recommended. [--output-name-scheme=complex] changes the naming scheme of the files the program creates upon successful downloads to BASE/YYYYMM/YYYYMMDD/dataset-grid-yyyymmdd-hh00-000-vars-levs.grb. (By default, the files are placed into a simpler scheme: BASE/YYYYMM/dataset-grid-yyyymmdd-hh00-000.grb)

Examples

- get-httpsubset.pl 2003010100 2003010121 ACPCP sfc .

A day (January 1st, 2003) of Surface Precipitation from the narr-a dataset downloaded to the current working directory

- get-httpsubset.pl 2003010100 2003123121 HGT-TMP 500_mb-850_mb-1000_mb /home/user/data/

A year (2003) of 1000, 850, and 500 millibar Temperature and Geopotential Height downloaded into /home/user/data/

- get-httpsubset.pl 1980010100 1980123121 UGRD-VGRD 250_mb . narrmon-a

12 monthly averaged 250 millibar U and V winds for a year (1980) into the current working directory (narrmon-a dataset)

- get-httpsubset.pl 2003010100 2003010121 GFLUX sfc . narr-b

A day (Jan 01 2003) of Surface Ground Heat Flux from the narr-b dataset

- get-httpsubset.pl 2005040100 2005040121 TMP 2_m /home/user gfs

A day (April 1st, 2005) of one degree GFS, 2-Meter Temperature data downloaded to user home directory

General Data Structure

Organization of Data Directories

Our repository contains a directory for each model system (NARR, GFS, RAP...). Each model directory contains one directory for each month. Each month directory contains a directory for each day in the month. The directory for each day contains all the data from model runs initialized during that day. These directories contain the GRIB files. We also have a parallel structure that contains the table of contents listing (".toc" files) of each GRIB file. A sample of our directory structure is shown below.

```
noaaport/
|-- merged
|   |-- eta
|   |   `-- 200305
|   |       `-- 20030505
|   |-- eta_toc
|   |   `-- 200305
|   |       `-- 20030505
|   |-- gfs
|   |   `-- 200305
|   |       `-- 20030505
|   |-- gfs_toc
|   |   `-- 200305
|   |       `-- 20030505
|   |-- ruc
|   |   `-- 200305
|   |       `-- 20030505
|   `-- ruc_toc
|       `-- 200305
|           `-- 20030505
```

Organization of Data Files

Our model data repository contains data from different climate and weather models. It has recently been expanded to place station datasets under the distributed data access framework. The data from the climate models are relatively static, while the data from the weather models are collected in near real-time. The weather models are run at different frequencies and at different spatial resolutions and domains. The frequency of a model is the number of times or cycles the model is run per day. A model is typically run at four cycles per day (usually 00UTC, 06UTC, 12UTC, and 18UTC). The resolution and domain for a model run is determined by its grid number. A model run is the single execution of a specific model using a specific grid initialized at a specific time and run for so many forecast hours. The same model may be run at different grids. Our weather model data is cataloged by specific combinations of model and grid.

Output from a model run contains data from the analysis field (the initialization field) and the forecast data at the end of each time step (forecast hour). Our repository for weather model data contains a separate file for each analysis field and each forecast hour. Thus, each file contains data from the gridded analysis field or data from one forecast hour from a single model run. All files contain gridded data primarily GRIB format. For example, a model run with 20 time steps will consist of 21 GRIB files: one analysis (hour 00) plus 20 forecast hours.

Files are named to identify the model run where the data originated. The filenames are divided into five sections, delimited by underscores. In certain occasions, the 4th and 5th section may be excluded (such as a climate reanalysis with monthly time steps). An example of this naming convention is below.

<model>_<grid>_<yyyymmdd>_<cycle>_<hour>

where

- model: name of the model
- grid: grid number (212, for example); this number will serve other purposes when non-gridded data is involved; 1-3 digits
- yyyymmdd: initialization year, month, day; 8 digits (20030501, for example)
- cycle: initialization time in hhmm; 4 digits (0000, for example)
- hour: forecast hour relative to initialization; 3 digits (006, for example)

In instances where hours and days have no meaning (as with daily and monthly datasets), the filenames take the following default values below.

<model>_<999>_<00000101>_<0000>_<000>

The following is a list of filename extensions (and their corresponding purpose) that you may encounter while browsing the NOMADS data structure. Please note that each file type may not exist for every dataset on our system.

Extension	**Description**	**Example**
.grb	GRIB data	meso-eta_218_20030501_1800_045.grb
.unf	GrADS unformatted binary data file	meso-eta_218_20030501_1800_000.unf
.idx	index file produced by GrADS gribmap utility	meso-eta_218_20030501_1800_fff.idx
.map	map file producted by GrADS stnmap utility	meso-eta_218_20030501_1800_fff.map
.ctl	GRADS data descriptor file	meso-eta_218_20030501_1800_fff.ctl
.toc	Basic table of contents dump produced by "wgrib" utility. ".toc" files exist for each individual file, as well as a merged file for each cycle (fff.toc)	meso-eta_218_20030501_1800_045.toc
.inv	"wgrib" grib inventory data, containing information which allows for partial HTTP file transfers to work. Similar, but more detailed then ".toc" files	meso-eta_218_20030501_1800_045.inv

Use of Templates in GrADS Data Descriptor Files

Our repository has a single data descriptor file for each model run. We use the GrADS templates feature to aggregate all the forecast hour files for the same run. Since these files represent multiple forecast hours, we identify these files by using "fff" in the forecast hour field in the file name.

Examples	
File	**Description**
meso-eta_218_20030501_1800_fff.ctl	GrADS descriptor file for run
meso-eta_218_20030501_1800_fff.idx	index file for run
meso-eta_218_20030501_1800_fff.toc	table of contents for run

MD5 Hash Files

MD5, short for Message-Digest algorithm 5, is a cryptographic hash function created by Ron Rivest for the MIT Laboratory for Computer Science and RSA Data Security Inc. in 1992. The MD5 message-digest algorithm produces a 128-bit "fingerprint" or "message digest" of the input. MD5 was created to be a digital signature for applications where large files can be verified by checking the MD5 hash output. The specific technical details of MD5 can be found in RFC1321 (link is external).

Purpose of MD5 Files

NOMADS uses data files as inputs to create a message digest or "signature" for each specific file. The MD5 checksum that is created can be used as a compact digital fingerprint for the associated file. This signature or fingerprint can be used to verify the file by using a piece of software like md5sum (see external links below). The MD5 file is to be used solely for the validation of the downloaded data from NOMADS and offers no security, other than to verify that the file was not corrupted during the download process.

Identifying MD5 Files

An MD5 file has been added to many of the datasets to verify that the data have not been corrupted or changed during the download process. NOMADS may store these MD5 checksums in two different kinds of files.

1. A file that ends with ".md5" is the MD5 checksum for a file with the associated name. For example, xyz_218_19730830_000_034.grb.md5 contains the MD5 checksum for xyz_218_19730830_000_034.grb.
2. Files that start with "md5sum." contain a list of MD5 checksums for all the files in the associated directory. An example of this would be md5sum.19730830. This file contains all of the MD5 checksums for the 19730830 directory. These files can be used the same way by md5sum, and will report errors if some of the files are missing or corrupt.

Using md5sum on MD5 Files (Linux)

The md5sum program can be used to check single files or whole directories of files. md5sum will give an error for all the files that are corrupt or missing with a summary at the end.

Check a Single File

The individual .md5 files are validated by passing the MD5 file through md5sum with the -c switch.

md5sum -c xyz_218_19730830_000_034.grb.md5

Check the Current Directory

For checking files that start with "md5sum." you will perform the same operation. This will cause the md5sum program to run through the current directory and check the files against the "md5sum." file.

md5sum -c md5sum.19730830

Using gmd5sum on MD5 Files (FreeBSD)

On FreeBSD the md5sum command is called gmd5sum and is available through sysutils/coreutils in the ports collection. gmd5sum will give an error for any of the files that are corrupted or missing followed by a summary at the end.

Check a Single File

Individual files are validated using the corresponding .md5 file. The data files are validated by passing the MD5 file through gmd5sum with the -c switch.

gmd5sum -c xyz_218_19730830_000_034.grb.md5

Check the Current Directory

For checking files that start with "md5sum." you will perform the same operation. This will cause the gmd5sum program to run through the current directory and check the files.

gmd5sum -c md5sum.19730830

External Links

Here are some sites that offer free software to help you get started or better understand software that is installed on your system.

- GNU md5sum Manual Page (link is external): A Manual for both the Linux and FreeBSD version of the md5sum and gmd5sum utility.
- md5Summer (link is external): A GUI md5sum utility for Windows.
- HashCheck (link is external): Open-source Windows shell extension that adds MD5 checksum viewing, creation, and verification.

Miscellaneous Information

Building Dataset Collections

On the NOMADS data products pages you are given tables that contain one entry for each possible model-grid grid combination. When you choose either the "plot" or "ftp" functions, you first have to build a "dataset collection" for the model-grid you have chosen. Both functions take you to the same form. You build a dataset collection by selecting a time period, cycles, and forecast hours for your model-grid. Your collection contains files from all model runs initialized within the time period you specify.

After you build your dataset collection, you are given a page where you can review the collection. You can choose to revise your collection or continue to "plot" or "ftp" data. Both the "plot" and "ftp" function work only on the set of files in your collection. You will only have access to data contained within your dataset collection.

A dataset collection is a logical grouping of the files you have selected. When you build a collection, you are given a unique 4-digit order number. This order number is displayed on the page where you can review your collection. This order number is used to identify files if you later choose to put any data on our anonymous FTP server for retrieval. If you choose to revise your collection, you keep the same order number.

Working with Timeseries Data

The NOMADS Web Interface allows you to easily construct a timeseries across all the files in a collection by using the template function in GrADS. We dynamically build a GrADS control file (a ".ctl" file) for a timeseries of just the analysis fields (that is, forecast hour 000) from all the model runs in your collection and then use the control file and the gribmap utility to generate the necessary index file (a ".idx" file). This timeseries is built from all the "_000.grb" forecast hour files in your collection. Since this control file spans multiple cycles, we use "hh" in the filename to identify the file as follows:

meso-eta_218_20030501_hh00_000.ctl <-- aggregate of all 000 forecast hours

Use this ".ctl" file when you want to construct any kind of timeseries plots from analysis fields.

The first page of the web interface offers you the option to create the above file. The checkbox titled "Generate control file for analysis fields over all cycles" below the list of cycles and forecast hours toggles this feature. Turning this feature on requires the gribmap utility to be run

on the new template control file. Thus, on larger datasets this can take a very long time. The checkbox is unchecked by default for datasets in which we do not recommend its use.

Creating Plots with GrADS

After you build and review your dataset collection, you can use GrADS to generate plots. Press "Plot Data." First you select a control file (a ".ctl" file) from one of the model runs you have included in your collection. Next, you select a type of plot: map, timeseries, latitude/longitude versus time, or animation. If you want to construct a timeseries of analysis fields across multiple runs, then select the "hh" control file. (See the "Working with Timeseries Data" section above for more details.) Finally, you select a parameter, level, and other attributes for plotting. We use GrADS to generate the requested plot. You can download a postscript version of the plot or the GrADS meta file.

Scheduled Downtime Notification

Until we can get a redundant, load-balanced, front-end server system in place, there will be times when we need to take down the server for maintenance. Scheduled server and data outages will be posted on the NCEI banner. If you notice a dataset you need is unavailable, please check this notification panel before contacting us.