

Joseph Wielgosz and Brian Doty*
Center for Ocean-Land-Atmosphere Studies

James Gallagher and Daniel Holloway
The University of Rhode Island (URI)

Summary

The technologies of GrADS (the Grid Analysis and Display System) and DODS (the Distributed Ocean Data System) have been merged on both the client and server sides. GrADS, when linked with the DODS-enabled netCDF library, can access datasets and subsets of datasets over the network from DODS servers. Metadata retrieved from the servers is used by GrADS to facilitate comparison of different data, either by overlaying displays of data from different sources or via analytical techniques.

In addition, the GrADS and DODS technologies have been combined on the server side. The GrADS-DODS server allows DODS-enabled clients to retrieve arbitrary subsets of all the data formats that GrADS can handle, including GRIB. In addition, the GrADS-DODS server extends the capabilities of DODS by providing analysis capability on the server side. This greatly expands the utility of the client-server concept. For example, if one wants to average a large amount of data, all that data must be moved to the client in the case where the server only provides the ability to do simple subsetting. With the GrADS-DODS analysis server, the calculation can be performed at the server, with only the (fairly small) result being transmitted back to the client.

GrADS as a DODS Client

The Distributed Ocean Data System (DODS) provides a way to share data over the Internet. DODS is an open source software project for distributed data access originally developed by oceanographers and computer scientists at the University of Rhode Island (Gallagher and Milkowski 1995; Davis and Gallagher 1999).

DODS-enabled clients, such as GrADS, Ferret, IDL, Matlab, and end-user programs linked with various DODS client libraries, can retrieve data from DODS servers. The system is designed such that this data retrieval happens transparently. Libraries that are already in common use for retrieving data, such as the netCDF library, are modified to allow data to be obtained from a server over the network in addition to obtaining data from a file on local disk. Re-linking a program or package that already uses this library, using a DODS-enabled version of the same library, makes the client DODS-enabled. To then retrieve data from a DODS server, the end user simply opens the data set using the URL of the data set, rather than the local filename. Once the data set has been opened, the data is worked with as though it were local.

More information about DODS can be found at <http://www.unidata.ucar.edu/packages/dods>

The Grid Analysis and Display System (GrADS) is an open source desktop tool that integrates the analysis and display of a wide variety of geophysical data (Doty and Kinter, 1995). GrADS provides an interface to many data types including binary, GRIB, HDF, and netCDF (Hooper and Searight 1996). When GrADS is linked with the DODS-enabled netCDF library, data residing on DODS servers can be accessed via the netCDF interface in GrADS. GrADS also retrieves metadata from the server, and uses it to facilitate comparison of data sets from different servers with data sets from local disk.

More information on GrADS can be found at <http://grads.iges.org/grads>

The following examples shows the GrADS commands that would need to be entered to overlay similar data from two different DODS servers:

```
sdfopen http://cola8.iges.org:9090/
      dods/saprec
set gxout shaded
```

*Brian Doty, COLA, 4041 Powder Mill Road,
Suite 302, Calverton, MD 20705
e-mail: doty@cola.iges.org

```

set time jun1980
set grid off
display p
run cbar.gs
sdfopen http://www.cdc.noaa.gov/cgi-bin
      /nph-nc/Datasets/ncep_reanalysis.derived
      /surface-gauss/prate.mon.mean.nc
set gxout contour
display prate.2*86400*30

```

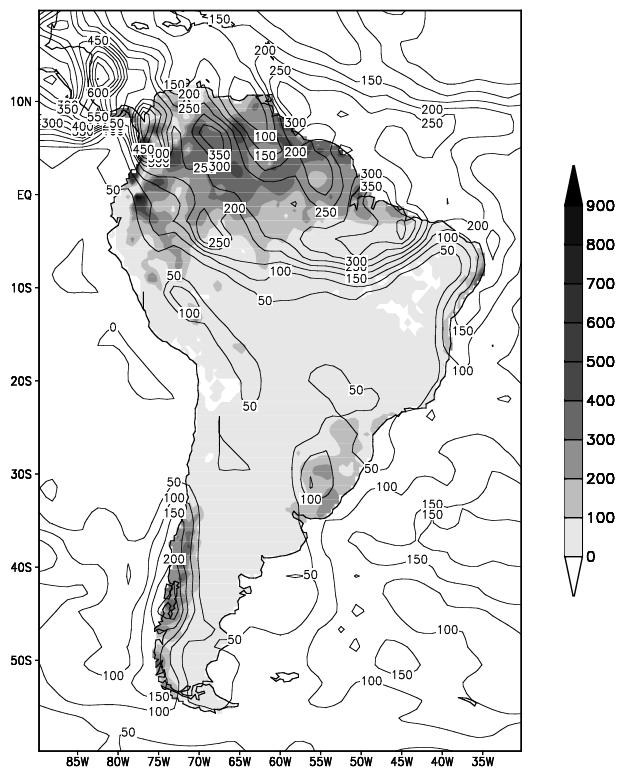


Figure 1: Observed South American precipitation (shaded) and NCEP reanalysis precipitation (contoured) for June 1980.

The graphic that GrADS would be displaying at the end of this series of commands is shown in Figure 1. (If you have a DODS-enabled version of GrADS, the above series of commands will work from your workstation – give it a try!) The gridding of the two data sets is different, and the data sets cover different time periods. Nevertheless, the graphics from the two different data displays are automatically overlaid with a background map. The necessary metadata to properly geo-register the data is obtained from the DODS servers. In addition, the date/time of the data to be retrieved was only specified once. GrADS automatically does the necessary work to retrieve the correct date/time and spatial subset from the different servers.

GrADS as a DODS Server

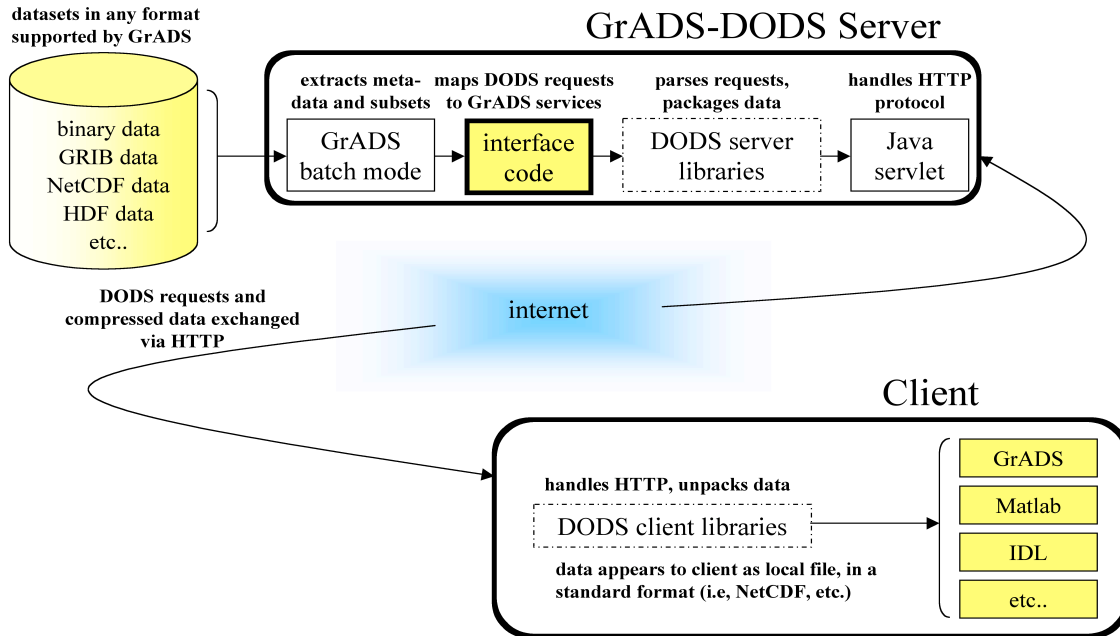
A DODS server handles data requests from DODS-enabled clients. The communication between the client and server uses a highly-abstracted and generalized data model, which allows a very wide variety of data to be potentially served by DODS. In practice, various DODS servers have been developed to serve particular data formats, such as netCDF, HDF, JGOFS, and others.

GrADS also has an abstracted data model of its own. Various data types, such as binary, GRIB, HDF, and netCDF, are handled in a way that is transparent to the user. The GrADS-DODS server consists of custom-developed code to translate between the GrADS data model and the DODS data model – in addition to the DODS code to handle the communications with the client; and the GrADS package itself, which performs the data retrieval and subsetting.

With the GrADS-DODS server, any data set that GrADS can handle can also be served over the network. Of particular note is the full support for GRIB data. GrADS allows GRIB data to be handled as a five-dimensional data set (lon, lat, lev, time, and variable) and provides the ability to efficiently retrieve arbitrary subsets. For example, a time series at a particular point in space can be directly retrieved from a GRIB data set. This capability is now available via the GrADS-DODS server. *A DODS client can access subsets of GRIB data without even being aware that the data is in that format.*

GrADS also provides extensive support for treating multiple disk files as a single data set. This allows the data administrator to split files into conveniently sized chunks, but still treat them as one data set from the users perspective. For example, the GrADS-DODS server running at COLA serves the real time NCEP MRF model output. The model output is in GRIB, and each 12-hour time step is in a different disk file. However, the user sees a single data set for the entire model run, and can work with the data via the netCDF API library.

The GrADS-DODS server has been designed to be extremely easy to install and to require no system privileges to run. Anyone who wishes to distribute data via this server may do so easily and quickly. The server also maintains a listing of online



datasets, in both HTML and XML, that can be browsed so that the available holdings on a given server can be easily determined.

Analysis at the Server

The GrADS-DODS server goes beyond the typical DODS server in a very powerful way – it supports general analysis capability on the server. In a research environment, it is difficult to predict what a scientist might want to do with the data. Even simple operations on the data (such as averaging) might require bringing the entire data set to the scientist’s desktop. This defeats some of the purpose of DODS. With the analysis capability of the GrADS-DODS server, it is much less likely that the scientist will feel compelled to transfer the entire data set to local disk.

The GrADS-DODS server uses the powerful ability of GrADS to encapsulate complex analysis tasks. Three things are needed for GrADS to do an analysis task:

- 1) the list of data sets to work with
- 2) the GrADS expression that defines the analysis task
- 3) the dimension environment that the expression operates within

For example, the following information is needed for GrADS to calculate the five-day mean forecast anomaly for 500mb heights using an MRF model forecast and NCEP reanalysis:

- 1) data sets: mrf0930, rean4xci
- 2) expression: ave(z-z.2,time=30sep2000, time=4oct2000)
- 3) dimensions: lon: -180 to 0 lat: 0 to 90 lev: 500 time: 30sep2000

For this calculation to be done via subsetting alone, about 1.3MB of data must be retrieved from the server. With the analysis being done at the server, only 65KB of data is sent back to the client.

To send this analysis task to the GrADS-DODS server, the three necessary pieces of information above are encoded into a special dataset name. Our example above would be encoded as:

```
_expr_{mrf0930, rean4xci}{ave(z-z.2,
time=30sep2000, time=4oct2000)}
{-180:0, 0:90, 500:500, 30sep2000:30sep2000}
```

The first time the server receives a request containing this dataset name, it performs the encoded task, and creates the result dataset. Further requests are served directly from the new dataset. This extension to DODS makes performing analysis

as easy as opening a dataset. Using DODS-enabled GrADS, we could view our results as follows:

```
sdfopen http://cola8.iges.org:9090/dods/  
_expr_{mrf0930, rean4xci}{ave(z-z.2,  
time=30sep2000,time=4oct2000)}  
{-180:0,0:90,500:500,30sep2000:30sep2000}  
display result
```

(These commands will also work from your workstation – give it a try!) In two commands, we have remotely performed analysis on several datasets, none of whose formats is known, and displayed the results as easily as if they were on local disk. Furthermore, the total amount of data transferred is 5% of what would have been required to do the task locally. Thus GrADS-DODS technology can dramatically increase the convenience of working with large datasets.

Server Implementation

The server and all its components are written in Java (except for GrADS itself, which is ANSI C). As distributed, it is essentially ready to run. The central component is a Java servlet, which connects two pieces of pre-existing software: GrADS, which handles the actual data operations; and the DODS core, which implements the protocol by which information is exchanged with the client across the internet. The Jakarta Project's Tomcat package (see <http://jakarta.apache.org/tomcat>) is used to provide an execution environment for the servlet.

All server settings are contained in a single configuration file, including the list of datasets to serve. Adding a new dataset is as simple as typing an online name that clients will use, and a filename for the data on disk. Upon starting, the server automatically extracts the necessary metadata and updates its directory listings.

The configuration file also gives control over resource usage. The administrator can limit CPU and disk usage for analysis requests, as well as the number of simultaneous users, maximum subset size, and total disk usage for cached data.

Conclusion

The GrADS-DODS server supplies a powerful engine for managing data in a distributed environment. The server may be used to quickly and easily share large data sets with colleagues. The server can also be used internally within large

enterprises to more effectively manage data and reduce network loads. For example, at COLA, a GrADS-DODS server is being used to front-end the four-times daily NCEP reanalysis. To share this 500GB data set over the network via NFS would be a substantial network load; with the GrADS-DODS server, users can easily do complex interactive analysis tasks on the data without loading the network and impacting other work. The server thus provides a highly scalable way to distribute data and analysis capability throughout the enterprise.

The server may be downloaded from: <ftp://grads.iges.org/grads/gds>. Complete installation instructions are included.

Acknowledgements: The development of the GrADS-DODS server was supported under a cooperative agreement with the National Aeronautics and Space Administration as part of the Seasonal to Interannual Earth Science Information Partnership (SIESIP) project being conducted under the direction of Prof. M. Kafatos of George Mason University.

References

- Davis, E. R., and J. Gallagher, 1999: Using DODS to Access and Deliver Remote Data, *Fifteenth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, Dallas, TX.
- Doty, B., and J. L. Kinter III, 1995: Geophysical Data Analysis and Visualization Using GrADS. *Visualization Techniques in Space and Atmospheric Sciences*, eds. E. P. Szuszcwicz and J. H. Bredekamp. (NASA, Washington, D.C.), 209-219.
- Gallagher, J., and G. Milkowski, 1995: Data Transport within the Distributed Oceanographic Data System, *World Wide Web Journal, Fourth International World Wide Web Proceedings*, 691-702.
- Hooper, D. K., and K. R. Searight, 1996: Adding netCDF Support to the Grid Analysis and Display System (GrADS). *Twelfth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, Atlanta, GA, 410-412.

