Filtering NetCDF Files by Using the EverVIEW Slice and Dice Tool

Background

Network Common Data Form (NetCDF) is a self-describing, machine-independent file format for storing array-oriented scientific data. It was created by the Unidata Program Center at the University Corporation for Atmospheric Research (UCAR), in conjunction with the National Aeronautics and Space Administration (NASA) and other organizations, to provide a common interface between applications and real-time meteorological and other scientific data (Unidata Program Center, 2009).

Over the past few years, there has been a growing movement within the community of natural resource managers in The Everglades, Fla., to use NetCDF as the standard data container for datasets based on multidimensional arrays. As a consequence, a need surfaced for additional tools to view and manipulate NetCDF datasets, specifically to filter the files by creating subsets of large NetCDF files. The U.S. Geological Survey (USGS) and the Joint Ecosystem Modeling (JEM) group, an ecological modeling community of practice, are working to address these needs with applications like the EverVIEW Slice and Dice Tool, which allows users to filter

grid-based NetCDF files. The major functions of this tool are as follows: (1) to create subsets of NetCDF files temporally, spatially, and by data value; (2) to view the NetCDF data in table form; and (3) to export the filtered data to a comma-separated value (CSV) file format.
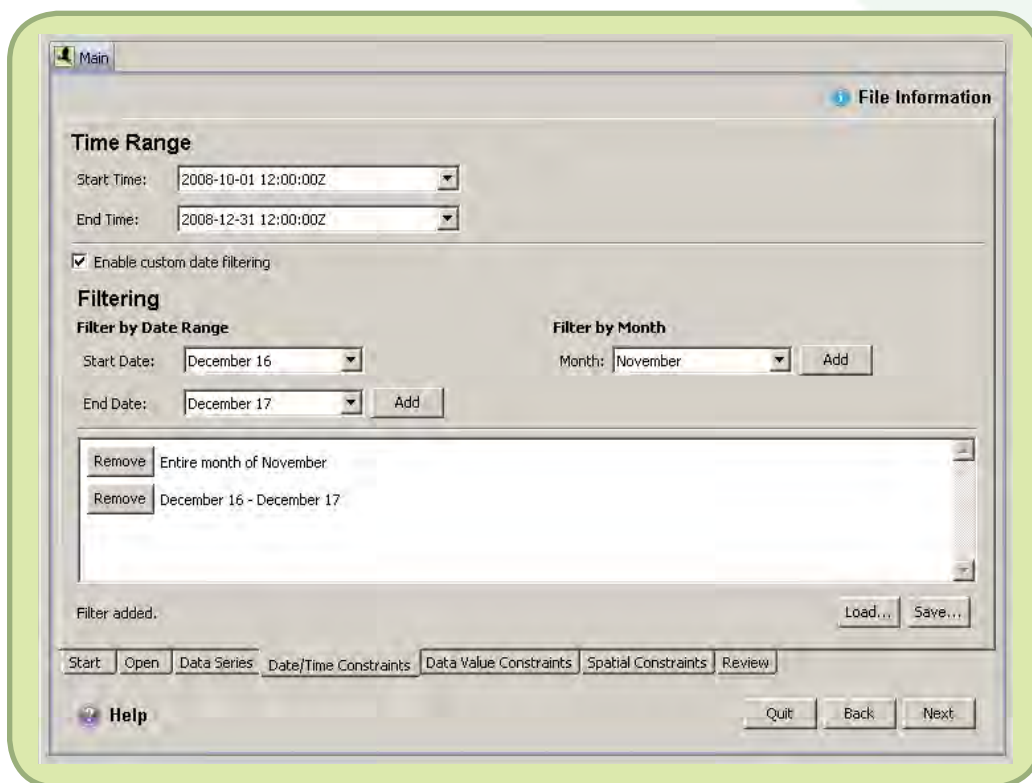
Using the Tool

Temporal Filtering

Ecologic and hydrologic models often produce large output files spanning large time intervals. The EverVIEW Slice and Dice Tool allows the user to create a new NetCDF file containing data matching their temporal template. The template definition could be as simple as a start and end timestamp. If the user needs to extract specific time ranges within a year, such as seasons, the template interface supports defining more complex ranges such as start and end month and day or specific months, which will repeat for every year of data in the source NetCDF files. Since defining complex temporal templates can be time consuming, the tool allows the user to save and load templates to an extensible markup language (XML) file.

Spatial Filtering

The spatial filtering functionality of the EverVIEW Slice and Dice Tool is separated into the following two parts: (1) filtering based on an Environmental Systems Research Institute (ESRI)[1] shapefile or (2) filtering based on coordinates. The shapefile-based filtering allows the user to select cells from a NetCDF grid on the basis of whether they are contained within the geometry of a polygon. This filtering uses the center of each grid cell as the reference point for the cell, so if the center of a grid cell is contained within the bounds of the polygon, then the cell is included in the output file. The polygons in a shapefile are sorted on the basis of a user selection of an attribute from the shapefile, and multiple polygons can be

decisionmaking or analysis. For this reason, the EverVIEW Slice and Dice Tool allows the user to export a NetCDF grid to the CSV format. The exported file contains the raw data values from the grid at a specific time period and the coordinates of the grid cell, which allows the user to access this data through other software packages.

Summary

Natural resource managers across The Everglades need access to information contained in large hydrologic and ecologic modeling outputs to make informed restoration decisions. Applications such as the EverVIEW Slice and Dice Tool allow those managers to filter data by creating subsets of large, complex model outputs temporally, spatially, or by data value, thus targeting those data most important to them. The USGS and JEM will continue to work with scientists and natural resource managers across The Everglades to solve complex restoration problems through technological advances.

selected. The coordinate-based filtering allows the user to enter four coordinates that constitute a bounding box. If the center of a grid cell is contained within this bounding box, then the grid cell is included in the output file.

Data-Value Filtering

There are many cases in which the user needs to focus only on cells whose data value exceeds a certain threshold or falls within a certain range; therefore, the EverVIEW Slice and Dice Tool allows filtering based on the value of each grid cell. Similar to the temporal template, multiple data-value ranges can be defined in the data-value template and can be saved to and loaded from an XML file. Since the data-value template XML file is different from the temporal template XML file, the end user can mix and match previously defined filters when creating subsets of the new NetCDF files or can apply identical filters to multiple files.

Table Viewing

The Table Viewer function of the EverVIEW Slice and Dice Tool displays the data values of the NetCDF grid in a spreadsheet format to allow the user to view the extracted grid cell values. This spreadsheet format is scrollable so that NetCDF grids of any dimension can be accommodated. The Table Viewer also allows the user to choose any grid and any time period from the NetCDF file to view. The Table Viewer can be used to inspect the chosen input file and also automatically opens after the output file is created so that the user can verify the output.

Exporting to CSV Format

Few users have software on their desktop computers that is capable of displaying the data contained within NetCDF files, but those users still need access to the data values for

most important to them. The USGS and JEM will continue to work with scientists and natural resource managers across The Everglades to solve complex restoration problems through technological advances.

For More Information

Craig Conzelmann
USGS National Wetlands Research Center
Tel: 337–266–8842
Email: conzelmannc@usgs.gov

Stephanie S. Romañach, Ph.D
USGS Fort Lauderdale Research and Education Center
Tel: 954–577–6341
Email: sromanach@usgs.gov

Reference

Unidata Program Center, 2009, NetCDF (network Common Data Form): Boulder, Colo., Unidata Program Center, accessed March 3, 2010, at http://www.unidata.ucar.edu/software/netcdf/.

[1]Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.