Converting US Topo GeoPDF Layers to GeoTIFF

Larry Moore
U.S. Geological Survey
February, 2013
Version 1

Introduction

Background

In 2009 the U.S. Geological Survey (USGS) began publishing digital quadrangle maps modeled on the legacy 7.5-minute topographic series (circa 1947-1992). The new maps, branded US Topo in 2010, are published in GeoPDF format – Portable Document Format (PDF) with geospatial extensions. PDF was chosen as the physical format because it is the only format in common use that

- Can carry vector, text, and raster data
- Can be displayed on a normal office computer without specialized software or expertise
- Allows a map to be printed at correct scale without specialized software or expertise

PDF geospatial extensions were implemented in US Topo because they add value "for free" – no cost to users, and do not conflict with base PDF.

US Topo was intended to be a map product, not a GIS product. US Topo maps are derived directly from national GIS databases and represent a repackaging of existing data, not creation of new data. Nevertheless, there is demand for the ability to load these symbolized maps into GIS software; the basic benefit of this capability is to supply the GIS user with a pre-built, symbolized background map.

There are several potential solutions to this GIS user problem. The USGS is working on ways to distribute the symbolized layers of US Topo in formats that are more GIS-friendly. It is also likely that the major GIS vendors will eventually implement GeoPDF import functions. In the interim, converting US Topo GeoPDFs to some other format is also an option, and is the subject of this paper.

A common question from GIS users is "how can I convert a US Topo to GeoTIFF?" GeoTIFF, though a simple raster image instead of an intelligent GIS format, provides an easy way to display map symbology and annotation. GeoPDF-to-GeoTIFF conversion would seem like a relatively easy way to import US Topos into GIS. Unfortunately, this reformat is not especially simple, mostly due to the lack of powerful software for manipulating layered GeoPDF files. PDF-to-TIFF conversion is not especially difficult or rare, but preserving georeferencing through this conversion is. Also, very little software is currently available for separating PDF layers, and even less for doing this while preserving georeferencing.

Scope

This paper presents software and procedures for separating the layers of a US Topo and converting the separated layers to GeoTIFFs. Two different methods are discussed:

1. Global Mapper commercial software. This is by far the easier of the two methods, but depends completely on proprietary software.
2. A collection of free and open-source software, including PDFCreator and the Geospatial Data Abstraction Library (GDAL) utilities. The required software is all no-cost and is mostly open source or public domain. This process is much more complex than the Global Mapper option.

Neither process is fully automated. Both produce results of high visual quality.

Option 1: Global Mapper

See http://www.bluemarblegeo.com/products/global-mapper.php for technical and purchasing information about Global Mapper.[1]

The following procedures are for Global Mapper v13.1. Other versions have not been tested by USGS, though Global Mapper documentation indicates that similar procedures should work for v11 and later.

Global Mapper can load layered GeoPDFs and preserve the georeferencing. However, Global Mapper PDF display behavior is different than Adobe Acrobat behavior. Global Mapper rasterizes the entire dataset, flattening all layers to a single raster image plane. Therefore, the Global Mapper display of a GeoPDF does not look like an Acrobat display of vectors, but rather like a file that has already been converted to a raster format such asTIFF or JPEG.

As part of the file load process, Global Mapper gives the user the option to select layers. All layers are ON by default; any layers clicked OFF by the user will not be loaded (Figure 1). This layer selection must be done when the file is loaded; once the map is displayed in Global Mapper, it has been flattened to one raster image plane and the layer characteristics of the original PDF are discarded.

When Global Mapper rasterizes a US Topo, it sets the resolution at about two ground meters per pixel, or about 300 lines per inch at map scale. If the data are exported to a raster file, this resolution is maintained by default, though it can be changed. Decreasing the resolution will reduce the file size and lower the image quality. Increasing the resolution will increase the file size, but will not noticeably improve image quality.

The displayed image can be exported to a GeoTIFF (or other raster formats) using the standard GM export functions.

Global Mapper includes a powerful scripting language, and this process can be scripted to convert any number of maps at one time. Such a script has been written by Stephen Aichele (saichele@usgs.gov). A listing of the script is included in Appendix A of this document. The script and a map cell Shapefile (needed for clipping collars; see the inline documentation of the script listing) are also included in a tar file of software at ftp://ftpext.usgs.gov/pub/cr/co/lakewood/Larry_Moore/ustopo2tiff / (described more fully in the next section).

As explained in the inline documentation, changing the behavior of this script (to alter the layers selected for export, for example), requires changes to the code. More complex versions of this script that would present run-time options to the user could be written.

[1] The USGS has no business relationship with Global Mapper or Blue Marble Geographic, Inc., and does not endorse or sponsor this commercial product. Global Mapper procedures are described here because, as far as we know, it provides unique capabilities for separating GeoPDF layers and converting them to GeoTIFF. Any reader who knows of other software with these capabilities is encouraged to notify lmoore@usgs.gov.

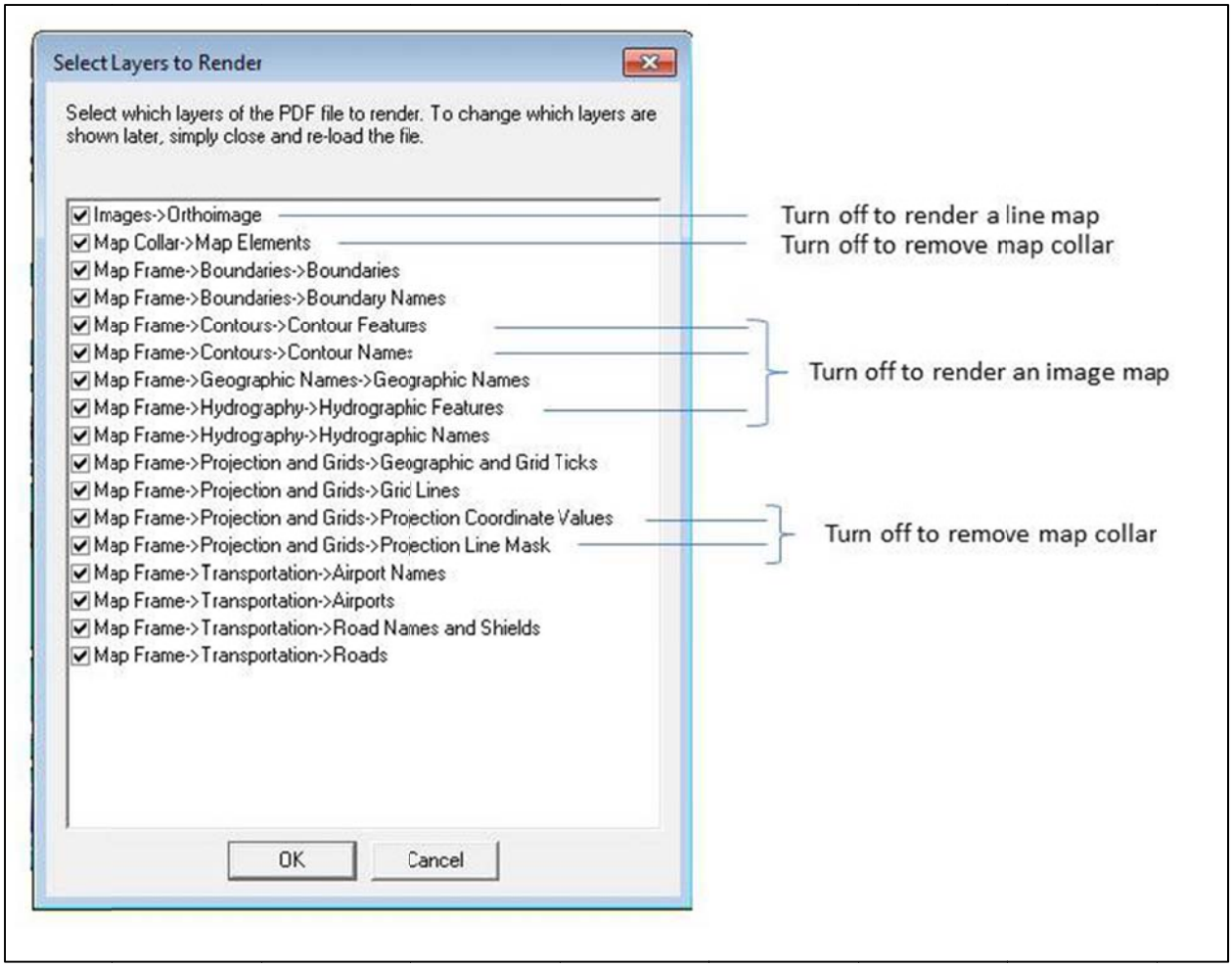

Figure 1. Screen shot of Global Mapper layer selection box. Layers may be "turned off" when a US Topo GeoPDF is opened. The notes on the right side are the current author's suggestions for useful combinations of layers, not requirements. This layer selection must be done as part of the file open process; layers cannot be selected later. The exact layer list may vary with the vintage of the US Topo.

Option 2: Free and Open Source Software

This option has a few advantages over the Global Mapper option: it uses software that is all free, and mostly open source or public domain, and the user has more control over the resolution of the output TIFF. However, this is a much more complex procedure. The initial software setup is especially difficult and tedious, and the reformat cannot yet be automated to the extent allowed by Global Mapper scripting.

Process Overview

1. Use PDFCreator as a print driver to separate layers of a US Topo. Save each layer or layer combination as a non-geospatial PDF.
2. Use *gdalinfo* to extract georeferencing parameters from the original US Topo; save this information for use in step 3c.
3. For each of the non-geopatial PDFs produced by step 1
    a. Convert the PDF to an RGB TIFF with *gdal_translate*.
    b. Convert the RGB TIFF to an 8-bit palette-color TIFF with *cmrdc*.
    c. Populate appropriate GeoKeys to create a GeoTIFF with *gtagrd* and *gtagwr*.
    d. Optional – change pixels outside the projection line to background white with *clip_ustopo*.

Step 1 must be performed interactively for each input file and each extracted layer or layer combination. Steps 2 and 3 are automated in a Perl script, but only for one quadrangle at a time.

PDFCreator (step 1) is a Windows program. The rest of the software in this process runs from a command line environment within Windows.

Detailed Instructions

Environment setup and software configuration

PDF Creator

Download and install PDFCreator (http://www.pdfforge.org/)

Set up PDFCreator [2]

- Open Start / Settings / Printers and Faxes (Win XP) or Start / Devices and Printers (Win 7)
- Right-click on PDFCreator, choose Printing Preferences
- Click Advanced button
- Set the Paper Size to PostScript Custom Page Size, make it 22.75” x 29” for USGS topos
- OK / OK and dismiss the Printers and Faxes dialog
- These settings are now in effect every time you use this “printer”

Command-line environment

The software for steps 2 and 3 (“Process Overview” above) are all command-line utilities. It is possible to use a Windows command prompt window, but I strongly recommend installing MinGW (Minimalist GNU for Windows), with its MSYS (Minimal System) Bash Shell (http://www.mingw.org/). The Cygwin environment (http://www.cygwin.com/) should also work, but has not been tested. A Perl interpreter is required regardless of the environment used, but this software is readily available for all environments and operating systems.

[2] Thanks to Michael Smith of the Maine Office of GIS for this procedure

The home page for MinGW is http://www.mingw.org/, and the download page is http://sourceforge.net/projects/mingw/files/. Download the installation file by clicking on the link labeled *Looking for the latest version? Download Mingw-get-install-xxxxxx.exe*. Install procedures are the same as any Windows software. When the installer program displays the "select components" window, select the optional "MinGW Developer Toolkit" in addition to the defaults. MinGW does not require admin privileges to install.

The default install directory for the MinGW system is `C:\MinGW`. If the system is installed to this location, then the Windows path to your MinGW/MSYS home directory will be `C:\MinGW\msys\1.0\`*`username`*. It is useful to place a shortcut for this directory on your desktop.

The file `C:\MinGW\msys\1.0\msys.bat` is a batch file that opens a command window that behaves like a Bash shell Unix terminal window. It is useful to place a shortcut to this file on your desktop.

After installation, open a Bash (command line) window. Create these two directories:

```
$ mkdir /usr/local

$ mkdir /usr/local/bin
```

Check that `/usr/local/bin` is in your path:

```
$ echo $PATH
```

If it is not, then create a file named `.profile` in your home directory and add this line to it:

```
PATH=.:/usr/local/bin:$PATH
```

Then exit and restart the Bash window, then recheck the path.

Geospatial Data Abstraction Library utilities

GDAL (Geospatial Data Abstraction Library) source code and pre-compiled utilities can be retrieved from http://www.gdal.org/. However, Windows binaries of the two GDAL programs relevant to this paper are included in the .tar bundle described in the next section. GDAL utilities must be from version 1.9.1 or later.

GeoTIFF utiliites

The remaining programs are works by the current author. Windows binaries (including the two necessary GDAL utilities) can be retrieved from ftp://ftpext.usgs.gov/pub/cr/co/lakewood/Larry_Moore/ustopo2tiff , file *ustopo2tiff.tar.gz*. Retrieve this file and put it in your MSYS home directory. Unpack it in the Bash shell with

```
$ gunzip ustopo2tiff.tar.gz
$ tar –xvf ustopo2tiff.tar
```

This will create a directory `distribution` with three subdirectories: `bin` , `doc,` and `global_mapper`. Copy program and .dll files to `/usr/local/bin`:

```
$ cd distribution/bin
$ cp * /usr/local/bin
```

The copied files include a number of .dll files that will not be present on most Windows systems. If, when running the software, you get error messages about other missing DLLs, please send the file names to me so I can include them in this distribution.

The subdirectory `doc` contains this document. The subdirectory `global_mapper` contains a Global Mapper script and a Shapefile, as described in the Global Mapper section of this document.

The FTP directory also contains a tar file of sample data: ftp://ftpext.usgs.gov/pub/cr/co/lakewood/Larry_Moore/ustopo2tiff/sampledata.tar. Retrieving this is optional.

To test that everything is in place, run these programs from the command line. Each should return a short summary of the program's purpose, syntax, and options:

```
$ gtagrd
$ gtagwr
$ gdalinfo
$ gdal_translate
$ cmrdc
$ clip_ustopo
```

Processing steps

The example used in these instructions is the 2012 US Topo for Half Dome, CA (in Yosemite National Park). The file name of the file as downloaded from the USGS Map Store is CA_Half_Dome_20120404_TM_geo.pdf. The inputs and outputs for this example are in the sample data tar file referenced in the previous section.

1. Put the original US Topo file in a subdirectory by itself.
2. Open the US Topo for display in Acrobat Reader
3. Click off all layers except Contour Names and Contour Features
4. "Print" the resulting display, using PDFCreator as the print driver, to a PDF file. This file can be named anything, but must have a .pdf extension. PDFCreator will output to other formats; be sure to print the file to a new PDF.
5. Repeat steps 3 and 4 for any layer or combination of layers for which a GeoTIFF is desired. For this example, hydrography (names and features) and transportation (names and features) will all be saved to one file, in addition to the contour file already created.

   At this point, the directory contains three PDF files: the original US Topo GeoPDF and two files of selected layers that are plain PDFs without georeferencing:

   ```
   CA_Half_Dome_20120404_TM_geo.pdf
   contours_Half_Dome.pdf
   hydro_trans_Half_Dome.pdf
   ```

6. In a MSYS command window, cd to the directory containing these PDF files, and run the Perl script *ustopo2tif.prl* with the input US Topo file name as an argument:

   ```
   $ ustopo2tif.prl CA_Half_Dome_20120404_TM_geo.pdf
   ```

This creates two GeoTIFF files with the same base names as the files saved from PDFCreator :

```
contours_Half_Dome.tif
hydro_trans_Half_Dome.tif
```

The *ustopo2tif* Perl script

This script is designed to extract and convert cartographic linework and annotation. Though the process will work for the image layer (at least at moderate output resolutions), using it to convert the image is not recommended. There are better ways to obtain NAIP imagery than extracting from a US Topo GeoPDF.Several of the utilities in this script assume specific characteristics of US Topo data, and will not work correctly with other types of GeoPDFs.

The Perl script *ustopo2tif.prl* automates much of the tedious work in the GeoPDF-to-GeoTIFF reformat. The script accepts several options on the command line. For a summary of the syntax, type

```
$ ustopo2tif.prl -?
```

The default resolution of the output GeoTIFF is 300 lines per inch. To output a different resolution, use the `res` option. For example, to increase the output resolution to 500:

```
$ ustopo2tif.prl --res=500 CA_Half_Dome_20120404_TM_geo.pdf
```

The behavior of the software at resolutions above 500 is not well tested. High resolutions combined with dense data may cause crashes due to inadequate memory or non-robust memory management in some of the software.

US Topo data layers are created not to the exact extent of the quadrangle projection line, but to a minimum bounding UTM rectangle of the projection line. A US Topo data layer therefore has slivers of linework outside the projection line. In an Acrobat display of a US Topo GeoPDF, these slivers are covered by a mask layer that can be clicked off by the user. The default behavior of the *ustopo2tif* Perl script is to "clip" these slivers by turning all pixels outside the projection line to background white. To override this behavior and maintain the original UTM map rectangle, use the `noclip` option:

```
$ ustopo2tif.prl –-noclip CA_Half_Dome_20120404_TM_geo.pdf
```

Future Work and Support

GDAL 2.0 (release date currently uncertain) may contain support for GeoPDF layer operations (http://www.gdal.org/frmt_pdf.html). This may make it possible to eliminate PDFCreator from this process and fully automate layer separation and GeoTIFF conversion of US Topos.

As noted in other parts of this paper, some of the utilities used in this conversion process are special-purpose and may not be especially robust. It is not clear how useful this reformat is, or how much effort would be required to make it simpler and more general.

Send bug reports to Larry Moore, lmoore@usgs.gov. Please include the full file name of the US Topo (e.g., CA_Half_Dome_20120404_TM_geo.pdf) that caused the problem, the text of any error messages, a description the layers being converted, and any other information that might be relevant.

Finally, please be aware that this is an experimental process and is not a high priority within the US Topo program. User support and rapid turn-around on bug fixes are not promised. The author expects that the process described here will have a relatively short life, and will soon be replaced by commercial software and more powerful built-in capabilities of commercial GISs.

Key URLs referenced in this paper

Global Mapper: http://www.bluemarblegeo.com/products/global-mapper.php

PDFCreator: http://www.pdfforge.org/

MinGW: http://www.mingw.org (home) http://sourceforge.net/projects/mingw/files (download)

GDAL (Geospatial Data Abstraction Library): http://www.gdal.org

Perl script, geotiff utilities, the two GDAL utilities used for this process, DLL files, and this documentation: ftp://ftpext.usgs.gov/pub/cr/co/lakewood/Larry_Moore/ustopo2tiff /ustopo2tiff.tar.gz

Sample data: ftp://ftpext.usgs.gov/pub/cr/co/lakewood/Larry_Moore/ustopo2tiff/sampledata.tar

Appendix A. Global Mapper script for batch GeoPDF-to-GeoTIFF reformat

```
// Copying and pasting this text may require some format modifications…
//      better to download the script, included in a tar file at
//      ftp://ftpext.usgs.gov/pub/cr/co/lakewood/Larry_Moore/ustopo2tiff

// This Global Mapper script converts USGS GeoPDF US Topo maps to Geotiffs.
// The script supports two outputs: maps with collars, or with collars clipped.
//      Read inline documentation below.
//
// Author: Steve Aichele, saichele@usgs.gov
// Modified in Feb 2013 by Larry Moore, lmoore@usgs.gov
//
// Lots of stuff is hardcoded that may have to be modified depending on
//      1) vintage of US Topo and
//      2) desired content of output
//
// RUN INSTRUCTIONS:
// Place this script, and GeoPDF USTopos in a directory.
// Open GlobalMapper 13.1 or higher and run this script.
// If you want to clip map collars, the shapefile "All_7.5min_cells.shp"
//      must be in the same directory as the script and PDFs.


GLOBAL_MAPPER_SCRIPT VERSION="1.00"
UNLOAD_ALL

// set starting directory
DIR_LOOP_START DIRECTORY="." FILENAME_MASKS="*.pdf" RECURSE_DIR=NO


// Layer list follows.
// Comment out lines for layers you don't want in the output TIFF
// Notes:
//      There is variation in the actual content of different vintages of US Topos...
//      ...and also variation in layer names due to bugs and editorial design changes.
//      Having "extra" layers in this list does not hurt anything,
//      but having a layer missing means it cannot be exported to GeoTIFF.
//      The following list probably does not cover all layer name variations.
//
IMPORT FILENAME="%FNAME_W_DIR%" \
        TYPE="PDF" RECTIFY="Automatic" LOAD_FLAGS="2,400, \
                // collar -- name variations are due to a bug in some US Topos
                        Map Collar->Map Elements<SEP> \
                        Map Collar->Map elements<SEP> \
                // grid lines and grid annotation
                        Map Frame->Projection and Grids->Projection Coordinate Values<SEP> \
                        Map Frame->Projection and Grids->Geographic and Grid Tics<SEP> \
                        Map Frame->Projection and Grids->Grid Lines<SEP> \
                // mask of UTM rectangle data slivers outside projection line
                        Map Frame->Projection and Grids->Projection Line Mask<SEP> \
                // geographic names
                        Map Frame->Geographic Names->Geographic Names<SEP> \
                // structures
                        Map Frame->Structures->Structures Names<SEP> \
                        Map Frame->Structures->Structures<SEP> \
                // boundaries
                        Map Frame->Boundaries->Boundary Names<SEP> \
                        Map Frame->Boundaries->Boundaries<SEP> \
                // transportation -- name variations are due to a bug in some US Topos
                        Map Frame->Transportation->Road Names and Shields<SEP> \
                        Map Frame->Transportation->Roads<SEP> \
                        Map Frame->Transportation Names->Road Names and Shields<SEP> \
                        Map Frame->Transportation Names->Roads<SEP> \
                        Map Frame->Transportation->Airport Names<SEP> \
                        Map Frame->Transportation->Airports<SEP> \
                        Map Frame->Transportation Names->Airport Names<SEP> \
                        Map Frame->Transportation Names->Airports<SEP> \
```

```
                        Map Frame->Transportation->Railroad Names<SEP> \
                        Map Frame->Transportation->Railroads<SEP> \
                        Map Frame->Transportation Names->Railroad Names<SEP> \
                        Map Frame->Transportation Names->Railroads<SEP> \
                // at this writing, only Wyoming maps have a PLSS layer
                        Map Frame->Public Land Survey System->PLSS Names<SEP> \
                        Map Frame->Public Land Survey System->PLSS Grids<SEP> \
                // hydrography
                        Map Frame->Hydrography->Hydrographic Names<SEP> \
                        Map Frame->Hydrography->Hydrographic Features<SEP> \
                // contours -- name variations are due to a bug in some US Topos
                        Map Frame->Contour Names->Contour Names<SEP> \
                        Map Frame->Contour Names->Contour Features<SEP> \
                        Map Frame->Contours->Contour Features<SEP> \
                        Map Frame->Contours->Contour Names<SEP> \
                // timber cover
                        Map Frame->Land Cover->Woodland \
                // ortho image -- can be exported (uncomment), but not recommended
                        // Map Frame->Images->Orthoimage \
                        ,1" \
        SAMPLING_METHOD="NEAREST_NEIGHBOR" AUTO_CONTRAST="NO" CONTRAST_SHARED="YES" \
               CONTRAST_MODE="NONE" CLIP_COLLAR="NONE"  TEXTURE_MAP="NO"

// This command, if uncommented, creates GeoTiff's with map collars
EXPORT_RASTER FILENAME="%DIR%WC_%FNAME_WO_EXT%.tif" TYPE=GEOTIFF PALETTE=HALFTONE

// This command, if uncommented, outputs GeoTIFFs without map collars –
//      everything outside the projection line is set to white
// Clipping requires a shapefile of the standard 7.5-minute cells to define the clip area
////EXPORT_RASTER FILENAME="%DIR%%FNAME_WO_EXT%20.tif" TYPE=GEOTIFF PALETTE=HALFTONE \
////    POLYGON_CROP_FILE="All_7.5min_cells.shp"

UNLOAD_ALL

DIR_LOOP_END
```