Pontus Olofsson *(olofsson@bu.edu)*
Earth & Environment, Boston University

1. Introduction

Training module: Use of global tree cover and change datasets in REDD+ Measuring, Reporting and Verifying. Boston, 2015.

1.1 Overview

The main goal of these tutorials is to estimate areas of forest cover change from a sample of reference observations stratified by data from a global map of tree cover change (Hansen et al., 2013). The workflow includes downloading the data and creating a discrete stratification, selecting a stratified sample, interpreting the sample and analyzing the sample for inference of area and confidence intervals of area. While different estimators are available, this training material if focused on design-based inference of area using stratified estimation (Cochran, 1977; Olofsson, Foody, Stehman, & Woodcock, 2013) which has proven useful for estimating areas of discrete map categories (Stehman, 2013).

Before making use of this material it is strongly recommended that users read the module on the use of global datasets, titled *Use of global tree cover and change datasets in REDD+ Measuring, Reporting and Verifying (MRV)* and included in this set of training materials, which is part of the *Methods & Guidance* from the Global Forest Observations Initiative (GFOI, 2014).

The tools for completing this work will be done in a Virtual Machine (VM) that hosts a suite of open-source tools and that can be installed on Windows, OS X, Solaris and Linux operating systems. The VM is a VirtualBox disk image of the 14.04 LTS release of Ubuntu Mate distribution (https://ubuntu-mate.org/) as of June 10, 2015. This image is stored as an archived Open Virtualization Format (see: http://en.wikipedia.org/wiki/Open_Virtualization_Format for reference). In case shared folders, USB devices, etc. don't mount automatically, you may need to install the "Guest Additions" which is very simple and full instructions are available here: https://www.virtualbox.org/manual/ch04.html. In addition to the default Ubuntu applications, a suite of useful software is pre-installed on the VM, including:

Git
GDAL
GRASS GIS
Orfeo ToolBox
QGIS
Python (including many scientific Python libraries)
R
RStudio

1.2 Install

1. Download a binary of *Oracle VirtualBox* compatible for your operating system from https://www.virtualbox.org/wiki/Downloads; follow the instructions to install.
2. Download the latest version of the VM from http://earth.bu.edu/public/ceholden/VM/ (as of June 10, 2015 the latest version is "Ubuntu_Mate_14.04_20150610.ova")
3. Open *VirtualBox* and click *File* > *Install Appliance* and browse to the directory containing the downloaded ".ova" file. Once installed the ".ova" file can be deleted.
4. Before launching the VM, it is necessary to pay attention to the memory allocation: In *VirtualBox*, highlight the VM in the left-hand pane and click *Settings* (yellow cogwheel) > *System*> *Motherboard* tab > increase *Base Memory* on the scale line; the amount of allocated memory will depend on the host computer but it is recommended to use slightly more than half of the memory of the host computer (upper part of the green part of the *Base Memory* scale line).
5. Another useful feature is *Shared folders* which are directories on the host computer or mounted devices that can be accessed from within the VM (although not technically not the same thing, in this and other modules the terms *folder* and *directory* means the same thing). You can add shared directories from *Settings* > *Shared Folders*.
6. In *VirtualBox*, highlight the VM in the left-hand pane and click *Start* (green arrow). This will launch the VM; the username and password are both "opengeo-vm" (without quotation marks).
7. To view in fullscreen mode click *View* > *Switch to Fullscreen* in the VM top menu.
8. The interface of the VM can be changed to style of Windows and OS X and various Linux distributions by the *Menu* in the upper left corner > *Preferences* > *MATE Tweak* > *Interface* (the Windows-style interface is referred to as Redmond and the OS X interface as Cupertino according to the locations of the headquarters of Microsoft and Apple).
9. Some of the more useful applications are the *Caja* file browser (similar to the file browsers in Windows and OS X), the *MATE Terminal* which provides text-based access to the operating system, *QGIS* which is the main graphical user interface used in this training material, the *LibreOffice* suite which is similar to *Microsoft Office*, and *Atom* or *Pluma* both ASCII text editors. You can add these software to the quick launch panel by right clicking the name in the menu > *Add to panel.*
10. Make yourself acquainted to the VM and these software.

1.3 Using the terminal

1. QGIS provides a graphical interface for executing many useful programs and tools but there are a few programs, including some very useful GDAL programs, that need accessing via the terminal. For these reasons and many others, learning to use the terminal is important. For users without Linux and Unix experience this can be a bit intimidating at first but it's not hard.
2. Open *MATE terminal*: you will see black window saying

```
opengeo-vm@opengeo-vm:~$
```

This means username "opengeo-vm" at the host of the server with the same name; "~" means you are currently in your home directory; "$" is the prompt symbol. There are many good terminal tutorials online, like this one for example: https://www.digitalocean.com/community/tutorials/an-introduction-to-the-linux-terminal. Spend an half an hour to complete this or another online terminal tutorial if needed.

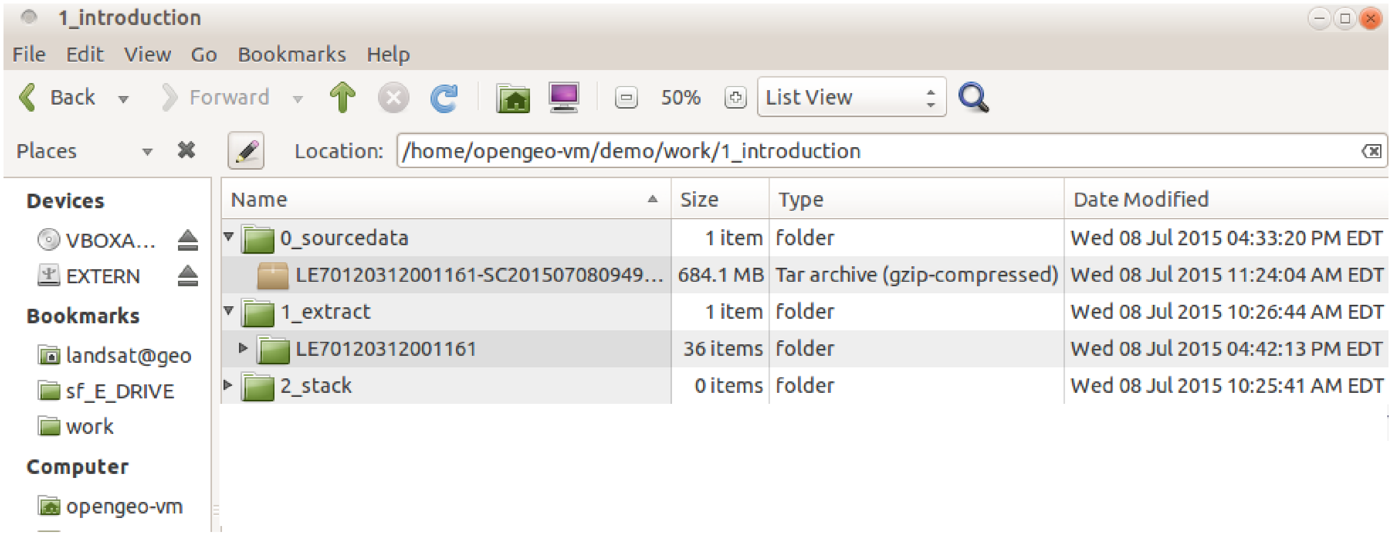
1.4 Data organization

1. Keeping your data organized will greatly facilitate your work. We suggest the following structure: in your home directory, which might be a shared folder, a mobile disk or a directory on the virtual machine, create three directories: one for scratch files (temporary files, test files, etc.), one for the datasets that you will use for interpretation (Landsat data in WRS path 12, row 31 in the example below) and one where you store your work. These directories will be referred to as your scratch, data and work directories. For the data directory, it is recommended to create subdirectories for each image path and row, with each image in separate subdirectories. Each section of the workflow will have its own work subdirectory ("1_introduction", "2_stratification", "3_sample", "4_interpretation", "5_analysis"). This gives structure as depicted below:

2. Create the structure. In this example, my home directory is /home/opengeo-vm/demo; navigate to your home directory and create a directory called "datasets" in the *MATE Terminal*. If you didn't learn how to do it in Subsection 1.3, type:
```
opengeo-vm@opengeo-vm:~/demo$ cd ~
opengeo-vm@opengeo-vm:~/demo$ mkdir datasets
opengeo-vm@opengeo-vm:~/demo$ cd datasets
opengeo-vm@opengeo-vm:~/demo/datasets$
```
3. Create your working directory. If you didn't learn how to do it in Subsection 1.3, type:
```
opengeo-vm@opengeo-vm:~/demo/datasets$ cd ..
opengeo-vm@opengeo-vm:~/demo$ mkdir work
opengeo-vm@opengeo-vm:~/demo$ cd work
opengeo-vm@opengeo-vm:~/demo/work$
```
4. Create a scratch directory. If you now navigate to your home directory and type `ls -l` to list the directory content, you should see the following:
```
opengeo-vm@opengeo-vm:~/demo$ ls -l
drwxrwxr-x 2 opengeo-vm opengeo-vm 4096 Jun 24 11:24 datasets
drwxrwxr-x 2 opengeo-vm opengeo-vm 4096 Jun 24 14:05 scratch
```
5. Create your working directory by opening the MATE Terminal:
```
opengeo-vm@opengeo-vm:~$ mkdir work
opengeo-vm@opengeo-vm:~$ cd work
opengeo-vm@opengeo-vm:~/work$
```
6. Create the following structure in your working directory:
```
opengeo-vm@opengeo-vm:~/work$ ls
1_introduction  2_stratification  3_sample  4_interpretation
5_analysis
opengeo-vm@opengeo-vm:~/work$
```
7. Alternatively, you can create the whole structure by typing:
```
opengeo-vm@opengeo-vm:~$ mkdir -p
demo/{work/{1_introduction,2_stratification,3_sample,4_interpret
ation,5_analysis},datasets,scratch}
```

1.5 Extract Landsat data

1. In "1_introduction", create six subdirectories: "0_sourcedata", "1_extract" and "2_stack" .
2. Copy a Landsat tarball (a compressed Landsat image with extension "tar.gz") to "0_sourcedata".
3. You can extract the data using either the *Caja* file browser or *MATE Terminal*: to use *Caja*, right click the tarball > *Extract to…*; this will open the Extract dialog window > navigate to "1_extract" > click *Create Folder* to create a directory with the name of the Landsat tarball > click *Extract*. It should look like this in Caja:



with the contents of the subdirectory in “1_extract” (in this example “LE70120312001161”) containing all of individual Landsat bands. In MATE terminal, you will need navigate to “1_extract”, create the image directory and type `tar -xvf [tarball]` to extract the tarball (a very useful trick when using the terminal is to type the first letter of directories or files and hit the tab key to autocomplete, this is especially helpful when completing long filenames like that of the tarball):

```
opengeo-vm@opengeo-vm:~$ cd work/1_introduction/1_extract/
opengeo-vm@opengeo-vm:~/work/1_introduction/1_extract$ mkdir
LE70120312005204
opengeo-vm@opengeo-vm:~/work/1_introduction/1_extract$ tar -xvf
/home/opengeo-
vm/work/1_introduction/0_sourcedata/LE70120312001161-
SC20150708094915.tar.gz
```

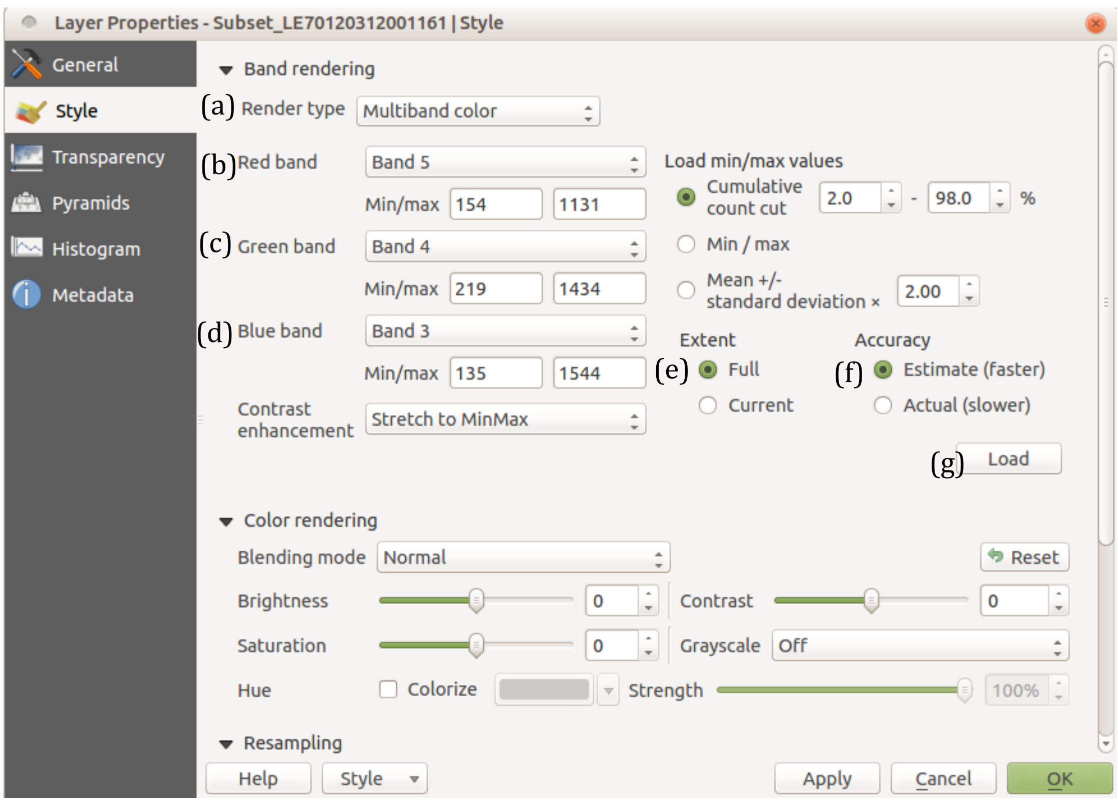
1.6 Stack Landsat bands

1. The main interface for analyzing geographical data in this training material is *QGIS*. It is similar to proprietary software like *ArcGIS* but it’s free and open source and allows for integration of user-contributed plugins, *GDAL*, *Orfeo Toolbox*, etc
2. To create a layer stack of your images, first open *QGIS*. After the application opens, click *Raster* > *Miscellaneous* > *Merge*; this will open the *Merge* dialog > click *Input files* to select Landsat bands 1-5 and 7 in “1_extract”..
3. Click Output files and navigate to “2_stack” and provide an appropriate file name like “LE70120312001161_stack” for example.
4. In the *Merge* dialog, leave all options unchecked except for *Layer stack* > *OK*. This will add the stacked image to the QGIS canvas.
5. Note: This is also possible via the command line using gdal_merge.py. When doing this, the ‘-separate’ option must be specified to create a layer stack (see http://www.gdal.org/gdal_merge.html for documentation).

1.7 Display image

1. If not already added to the QGIS canvas, open the Landsat image located in "~/work/1_introduction/2_extract" from *Layer > Add Layer > Add Raster Layer* or click  to the left of the *Layer* pane.
2. To change the color composite and stretch the image, right click the stacked image name in the layer pane > *Properties* > *Style*; the window below should appear; set *Render type* (a) to multicolor and the *Red*, *Green* and *Blue* band to bands 5, 4 and 3 (b-d), set *Extent* to *Full* (e), *Accuracy* to *Estimate* (f) and click *Load* (g) > *OK* to display a stretched 5-4-3-false color composite of the Landsat image.

   NOTE: if you're displaying an atmospherically corrected image downloaded from the USGS the no data value is -9999 which you will need to specify in *Properties > Transparency > Additional no data value* before stretching the image.



References

Cochran, W. G. (1977). *Sampling Techniques*. New York, NY: Wiley.

GFOI. (2014). *Integrating remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and Guidance from the Global Forest Observations Initiative* (1st ed.). Geneva, Switzerland: Group on Earth Observation.

Hansen, M. C., Potapov, P. V, Moore, R., Hancher, M., Turubanova, S. a, Tyukavina, A., … Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850–3. http://doi.org/10.1126/science.1244693

Olofsson, P., Foody, G. M., Stehman, S. V, & Woodcock, C. E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, *129*, 122–131. http://doi.org/10.1016/j.rse.2012.10.031

Stehman, S. V. (2013). Estimating area from an accuracy assessment error matrix. *Remote Sensing of Environment*, *132*, 202–211. http://doi.org/10.1016/j.rse.2013.01.016