

*



* Photoshop not covered in OSS 2014

Quantifying the Ecohydrological Impacts of Forest Disturbance

(Doing things the hard way)

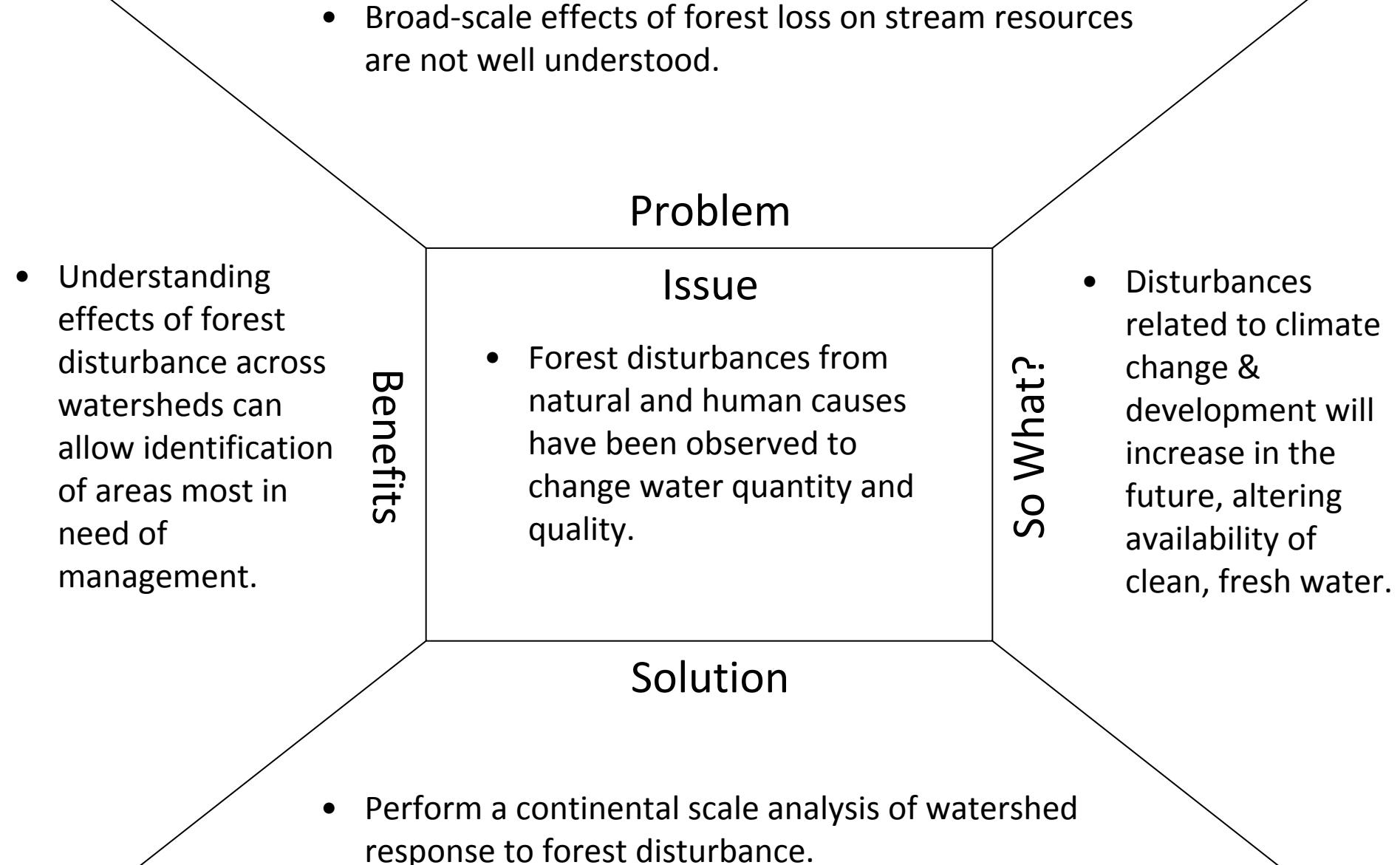
Tony Chang, Montana State

Jon Duncan, UNC-CH

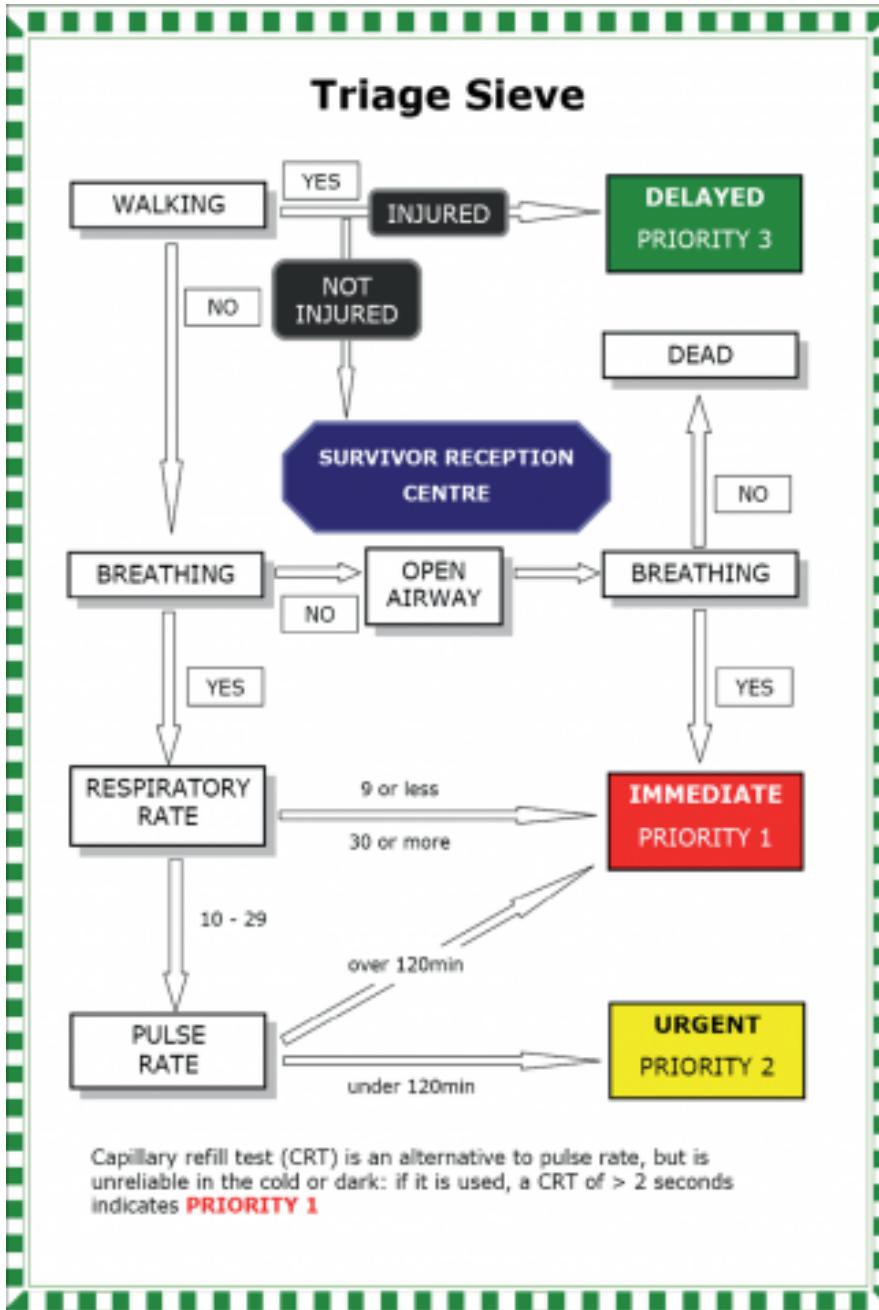
Gan Tian, Utah State

John Lovette, UNC- CH

Mike Treglia, Texas A&M / U Tulsa



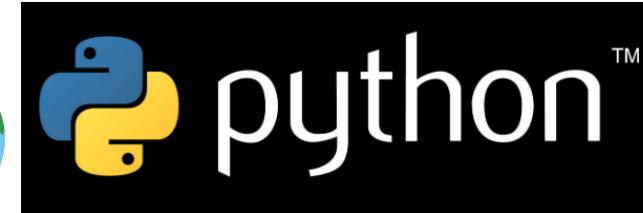
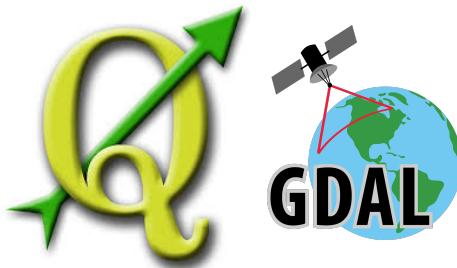
Medical Triage Analogy?



-
- Imagine a landscape where a great number of watersheds are injured. That injury being a type of disturbance.
 - This facilitates dedicating triage and long-term management efforts to most sensitive watersheds.
 - We don't have the time to examine each watershed closely to know which one to address first.
 - Analyze effects of disturbance on watersheds (figure out which watersheds are injured, and more resilient to injury)
- **Forest disturbance from natural and human causes have been observed to change water quantity and quality.**

Our Team's Open Source Technical Skills

Some Software:



Technical Expertise:

- Process-based Hydrologic Models and Hydrology(Jon, John & Tian)
- Algorithm and Model-based Statistics (Tony and Mike)
- Data Acquisition/Data Sources (Everybody)
- GIS Data Manipulation (Everybody - in varied forms)

Some Initial Steps

Important Discussions:

- Authorship
- Collaboration
- (What else?)

RE: Authorship of manuscript titled: CONUS Ecohydrologic Response to Forest Disturbances

Together our team working on project "CONUS Ecohydrologic Response to Forest Disturbances" has drafted a list of responsibilities for manuscript co-authors. The goal is to have an open dialog regarding authorship and to clearly define each co-author's responsibilities and accomplishments throughout the effort. Below are common contributions made by co-authors. We feel that co-authorship requires roughly 2-3 contributions below, the appropriate minimum depending on individual circumstances. Note that some contributions often relegated to the acknowledgments section are included, but that those who contribute in only one category would be placed in the acknowledgments. We highlight one view on this distinction below.

Weltzin et al. 2006: "*Drawing the line between acknowledgements and co-authorship can be challenging and one way of thinking about the differences may be to consider whether or not the participant is responsible and accountable for the article. A contributor receiving credit for the article should be willing to be held accountable for its contents and not be just responsible for a portion of work involved. In contrast, an acknowledgement may contribute formal or informal ideas to ongoing projects, collect enormous amounts of data, and develop and/or conduct statistical analyses, but may not be accountable for the final contents of all or even portions of the final manuscript.*"

We ask that you fill in the table below with the description of your contributions for each activity that applies. These documents will be archived on our team's collaborative website with each manuscript.

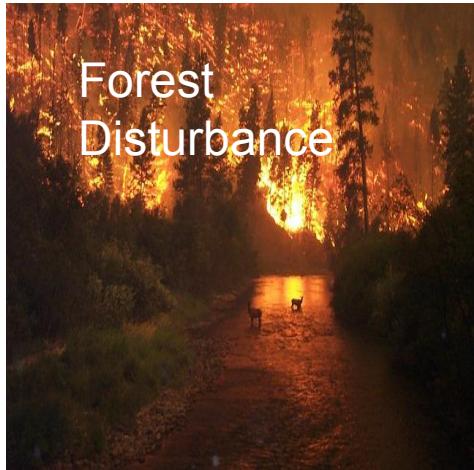
Authorship order to be determined by the agreed upon sum of task points within the weighted scoring system denoted by numbers by activity type

Author contributions by: TC-Tony Chang, JD-Jon Duncan, TG-Tian Gan, JL-John Lovette, ML-Michael Treglia

Activity	Your specific contributions
3 Conceived of the idea for the MS – e.g., framing the question or objectives, overall idea	TC; JD; MLT
3 Designed the MS – help determine structure of the MS, the intellectual content and scope, etc.	TC; JPL; MLT; JD; TG
1 Supervised – oversight and responsibility for the organization of the project and manuscript	TC
2 Performed research - Data collection	TC; TG; JD; JPL
2 Performed research - Data analysis or modeling	TC; MLT; JD; JPL; TG
2 Contributed new methods or models	MLT
1 Drafted figures and tables and MS formatting	MLT; JPL; TC; TG
3 Wrote parts of the MS -- selected portions of the manuscript (points for each portion)	TC; JPL; MLT; TG

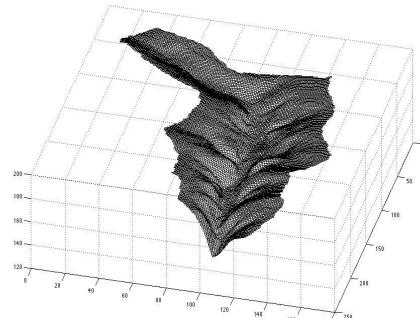
Conceptual Model

Drivers

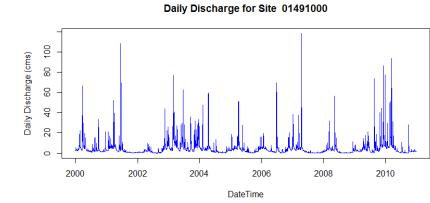


Fire
Logging
Insect/pathogen
Urbanization

Controls



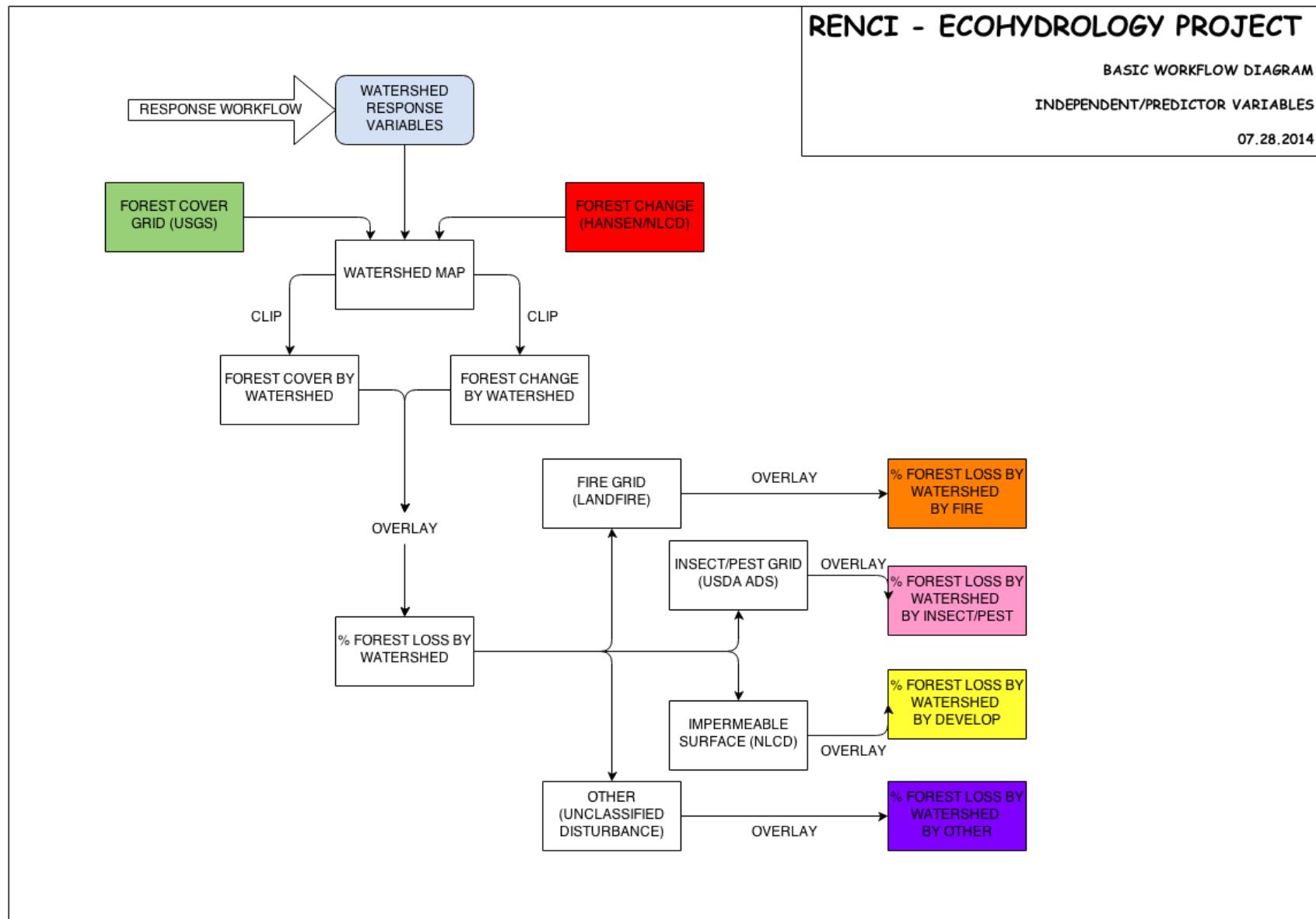
Responses



Streamflow
Water Quality

Size
Soils
Forest Type
Climate

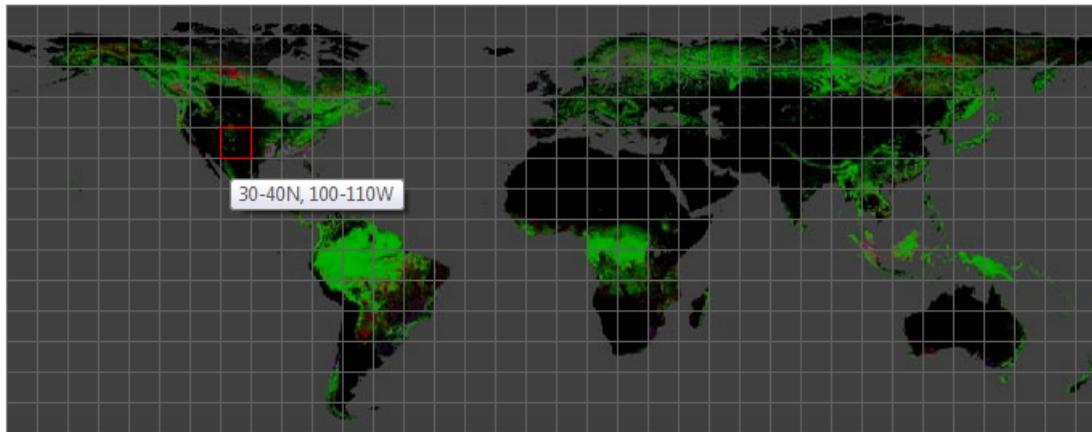
The Workflow



First Dataset...

Download Instructions

To download individual 10x10 degree granules, click on a region on the map below and then click on the URLs underneath it.



Granule with top-left corner at 50N, 130W:

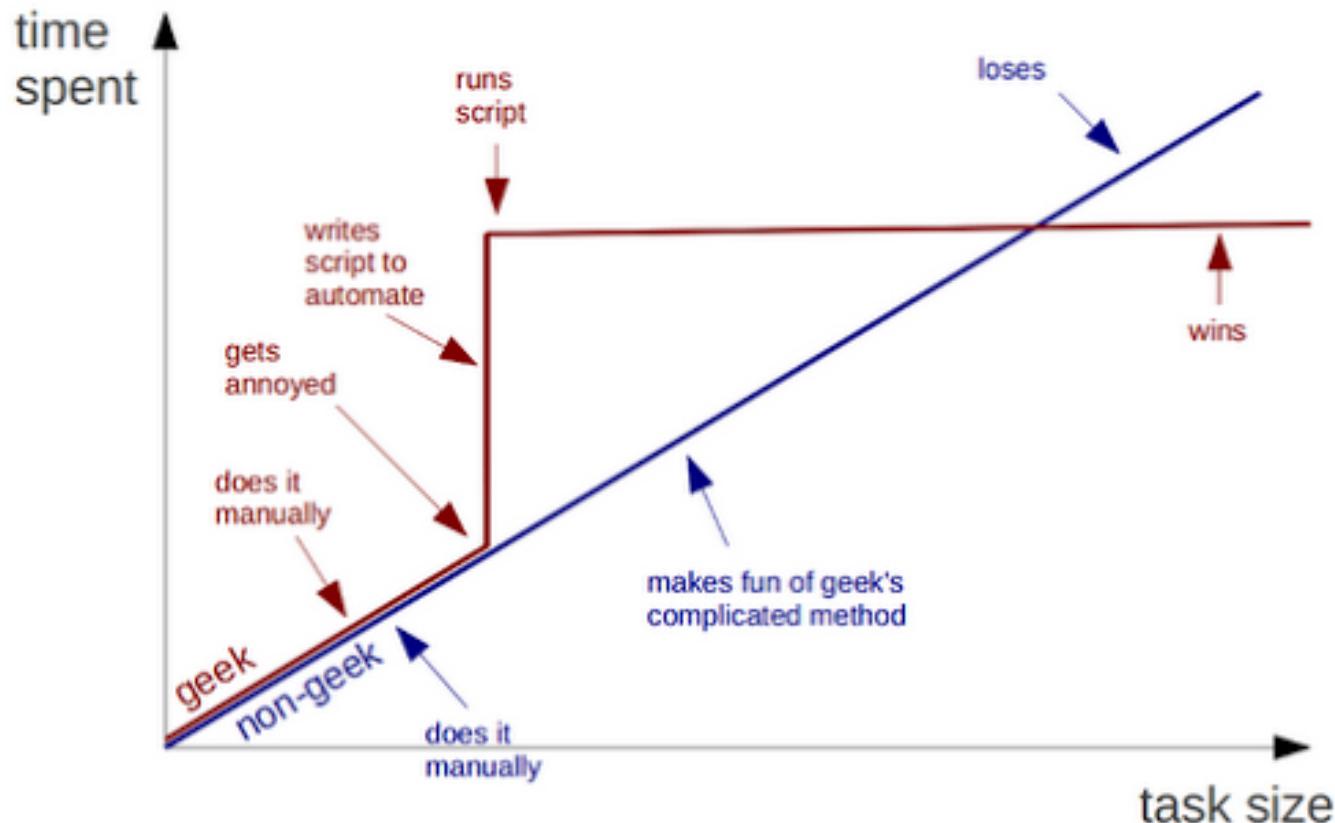
http://commandastorage.googleapis.com/earthenginepartners-hansen/GFC2013/Hansen_GFC2013_treecover2000_50N_130W.tif
http://commandastorage.googleapis.com/earthenginepartners-hansen/GFC2013/Hansen_GFC2013_loss_50N_130W.tif
http://commandastorage.googleapis.com/earthenginepartners-hansen/GFC2013/Hansen_GFC2013_gain_50N_130W.tif
http://commandastorage.googleapis.com/earthenginepartners-hansen/GFC2013/Hansen_GFC2013_lossyear_50N_130W.tif
http://commandastorage.googleapis.com/earthenginepartners-hansen/GFC2013/Hansen_GFC2013_datamask_50N_130W.tif
http://commandastorage.googleapis.com/earthenginepartners-hansen/GFC2013/Hansen_GFC2013_first_50N_130W.tif
http://commandastorage.googleapis.com/earthenginepartners-hansen/GFC2013/Hansen_GFC2013_last_50N_130W.tif

We have provided a complete set of granules spanning the range 180W–180E and 80N–60S, but the granules over the ocean are provided for completeness only and do not contain any meaningful data. Should you wish to download a complete layer, you may download a text file containing the complete list of URLs for each layer: [treecover2000](#), [loss](#), [gain](#), [lossyear](#), [datamask](#), [first](#), or [last](#).

<http://earthenginepartners.appspot.com/science-2013-global-forest/download.html>

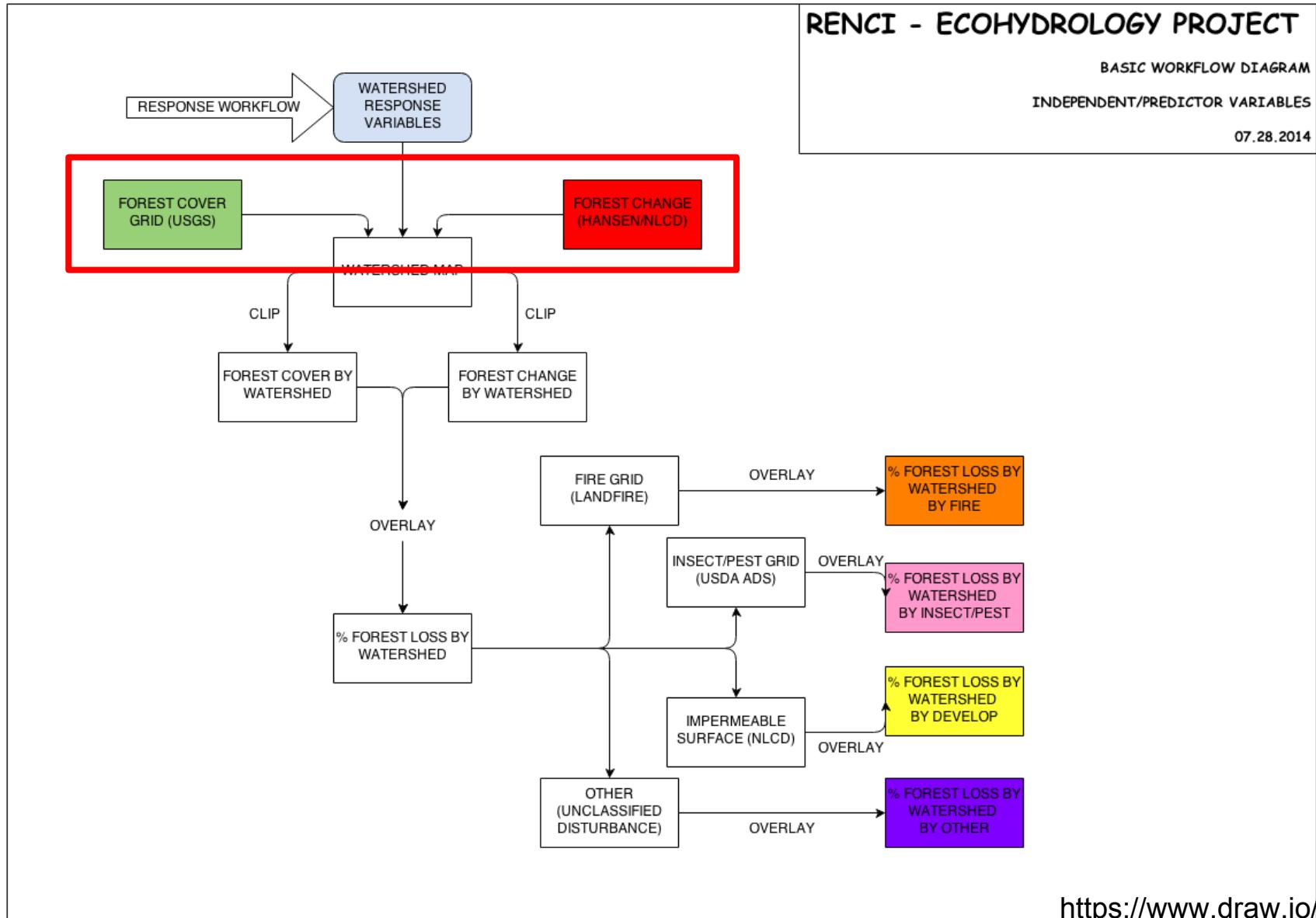
The Strategy

Geeks and repetitive tasks



Division of Labor

(Kind of... Lots of Helping Each Other Out)



Automated Downloaders

The image shows two windows side-by-side. The left window is a Notepad++ editor titled 'C:\Users\tony\OSS\Git\OSS-2014-Project\Scripts\forest_cover_downloader.py - Notepad++'. It contains Python code for a script named 'forest_cover_downloader.py'. The right window is a terminal window titled 'Console2S-2014-Project\Scripts - VIM' showing a bash script named 'MCD45_wget_2.sh'.

```
C:\Users\tony\OSS\Git\OSS-2014-Project\Scripts\forest_cover_downloader.py - Notepad++
File Edit Search View Encoding Language Settings Macro Run Plugins Window ?
geotool.py shape_extract_climate.py new 1 forest_cover_downloader.py

1 #web download of data
2 #author: Tony Chang
3 #abstract: This is a short script to automate the download the Hansen et al #2013 Forest
4 #date: 07.23.2014
5 #organization: OSS 2014
6 #location: RENCI, Chapel Hill, NC
7 #written for Python 3.3

8
9 import urllib
10 import numpy as np
11 from matplotlib import pyplot as plt
12
13 import gdal
14 from gdalconst import *
15 import osr
16 import os
17
18 def get_url(var, ulc): #need to specify the variable and the upper left corner
19     workspace = 'http://commondatastorage.googleapis.com/earthenginepartners-hansen/GFC20'
20     filename = "%s_%s.tif" %(var,ulc)
21     url = "%s%s" %(workspace,filename)
22     return(url,filename)
23
24 def gen_ulc(lat,lon):
25     if (lon<100):
26         slon = '0'+ str(lon)
27     else:
28         slon = str(lon)

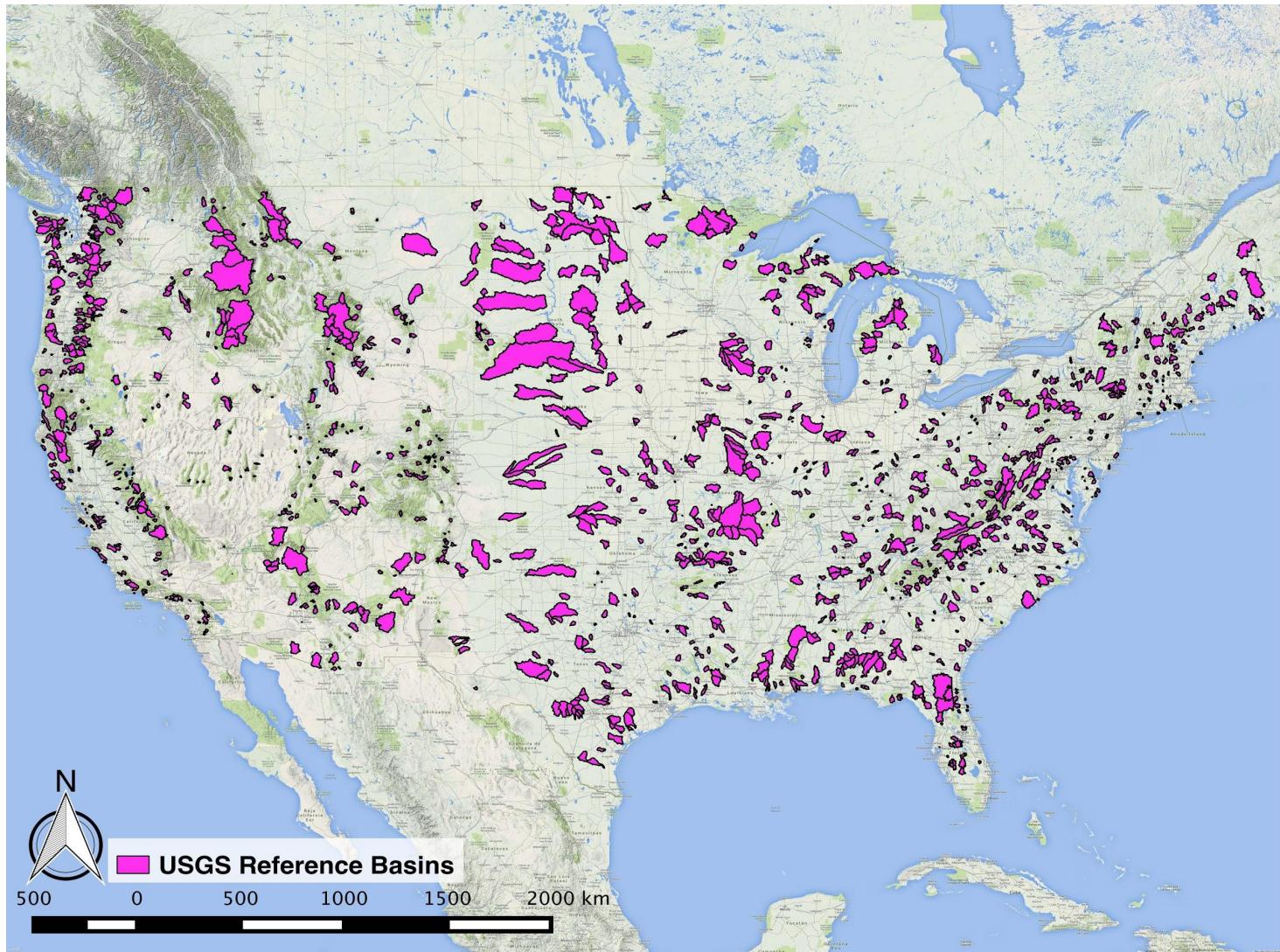
#!/bin/bash
## This script downloads monthly MCD45 modis landfire data from Modis-fire.umd.edu
## if mac does not have wget, use http://osxdaily.com/2012/05/22/install-wget-mac-os
## to activate

## Define years to download MCD45 fire data
#First Dataset...
YEAR=
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014

## Use wget to grab data from FTP. User and Password must be defined to access page
## User = 'user' Password = 'burnt_data'
## -nc tells wget to not grab anything that is already downloaded
```

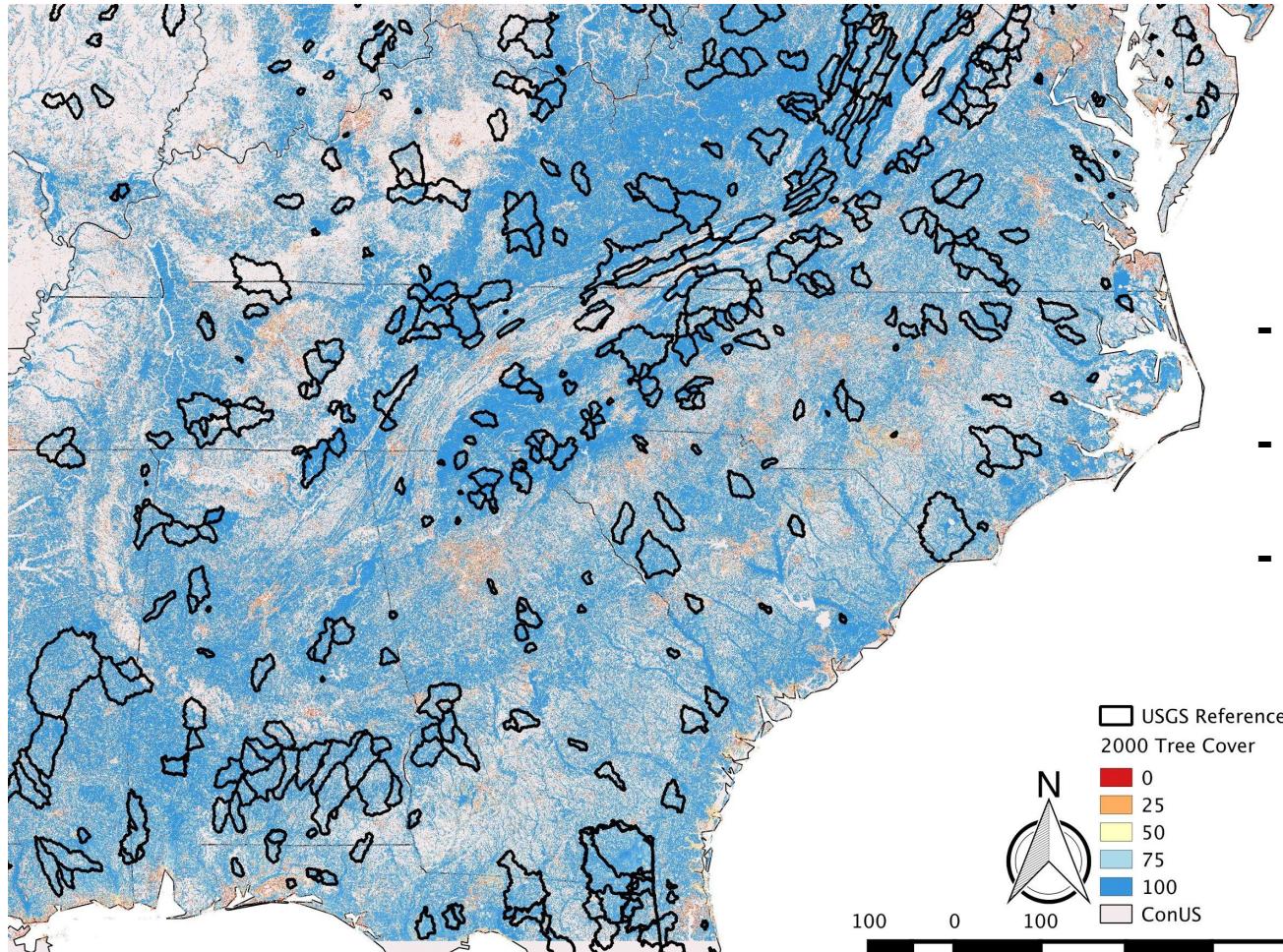
https://github.com/RENCI-Ecohydro/OSS-new/blob/master/Scripts/forest_cover_downloader.py
https://github.com/RENCI-Ecohydro/OSS-new/blob/master/Scripts/MCD45_wget_2.sh

The good...



http://water.usgs.gov/GIS/dsdl/gagesII_9322_point_shapefile.zip

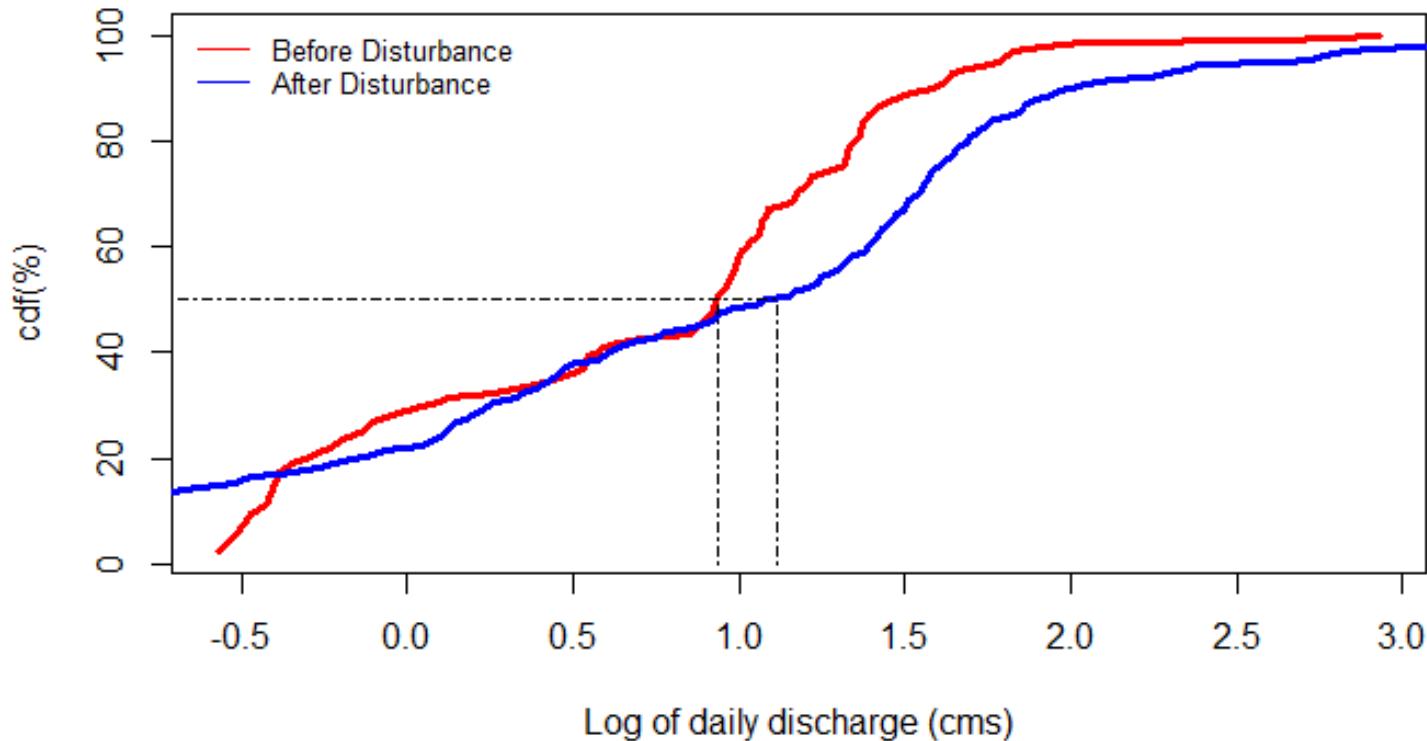
The bad...

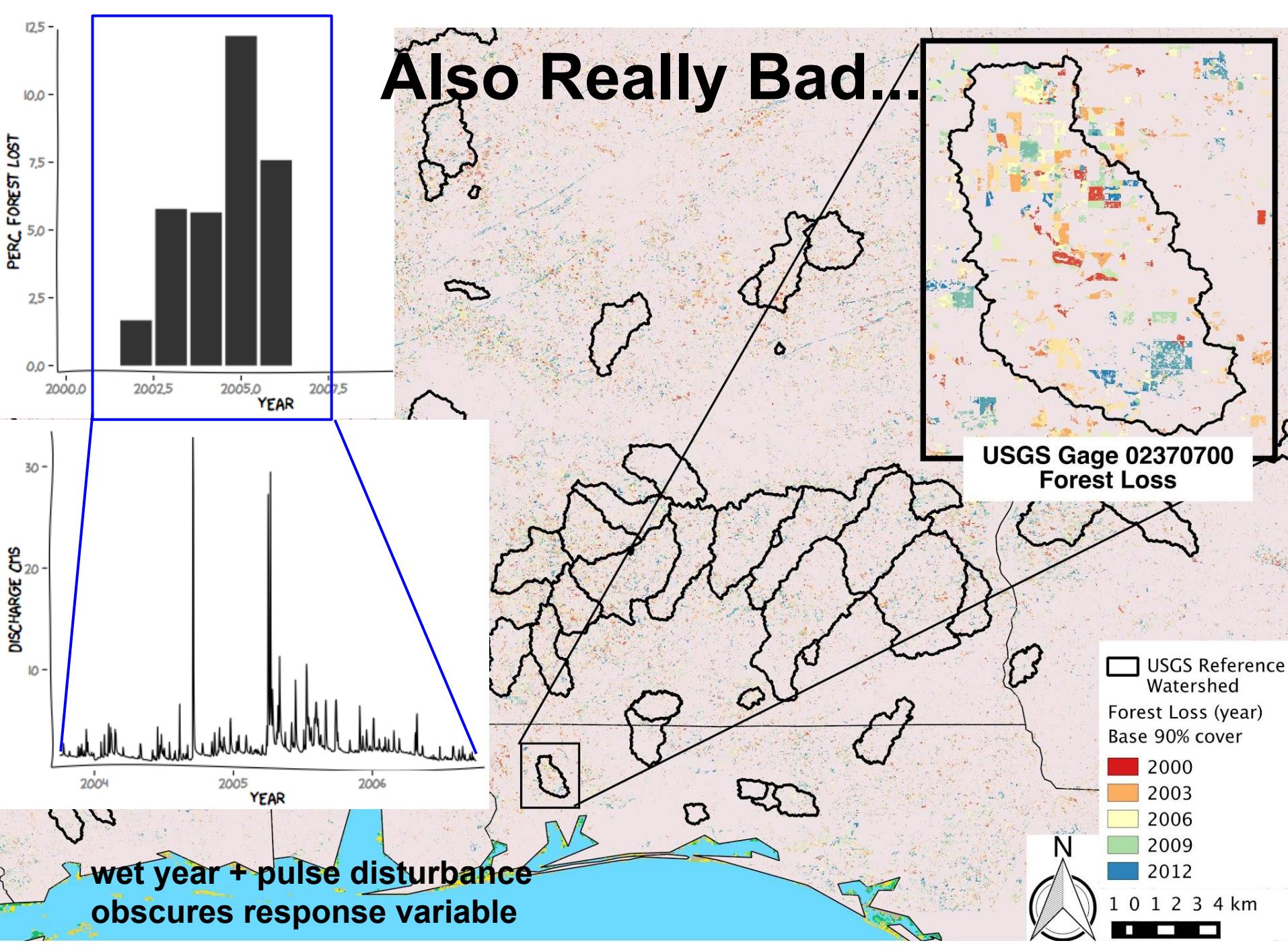


- ~ 200 GB of raster data
- 1.5 petapixels
- 1856 watersheds

Really bad...

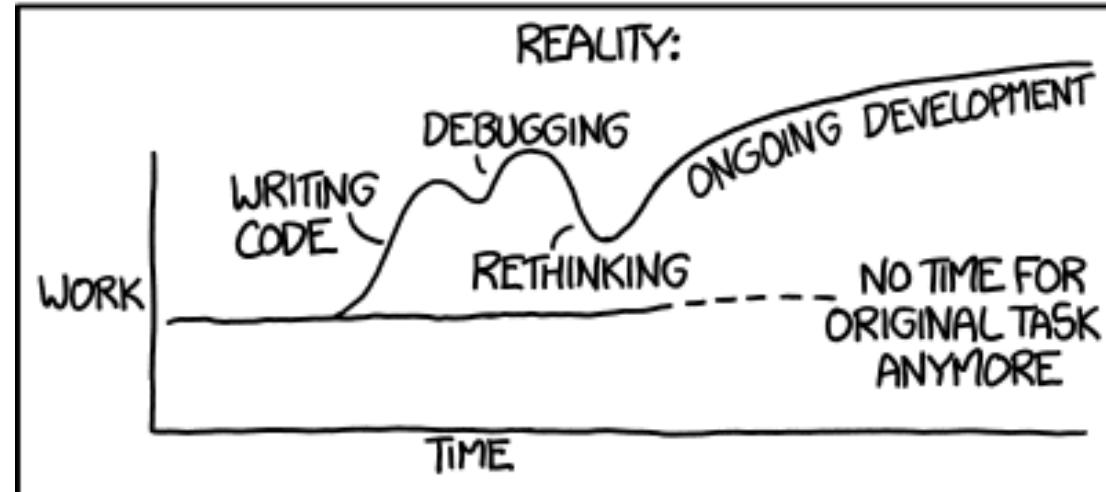
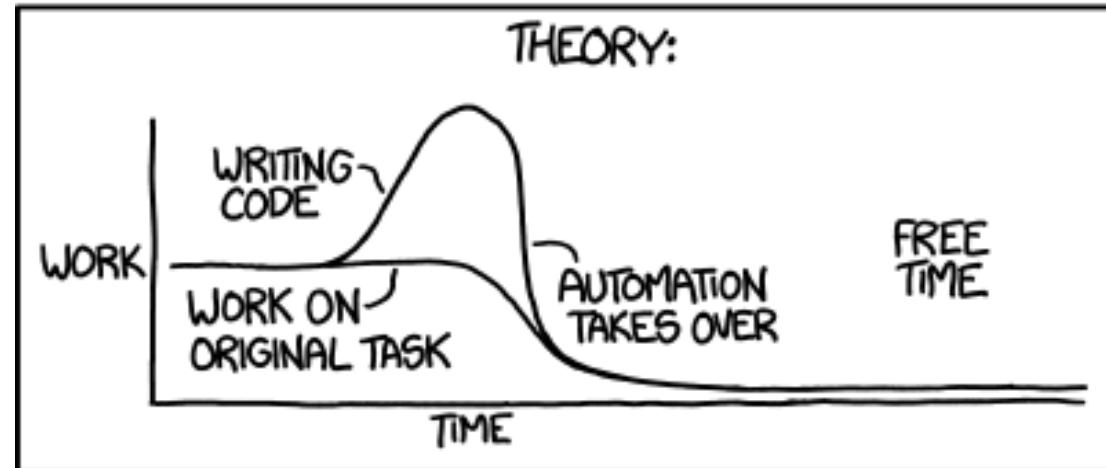
CDF Before and After Disturbance





The ugly...

"I SPEND A LOT OF TIME ON THIS TASK.
I SHOULD WRITE A PROGRAM AUTOMATING IT!"



Challenges and Solutions

Big Data Problems

- Large remote-server for Python scripts
- PostGIS/Postgres and Spatial SQL



Dennis van Zuijlekom



<https://github.com/RENCI-Ecohydro/OSS-new/master/Scripts>

Python or SQL?

Python Code to Extract Forest Loss by Watershed:
Not efficient but readable...sort of.

```
#extent = source_layer.GetExtent() #use this to reference which tile will be used to gather data from
#print(fextent)
return(out)

#OS workshop
"GOOD GOD...."
#uses additional computing resources
#defines additional computing resources

def findSource(extent,var):
    logfile = '/oss/data/tiles/tile_extents_%s.csv' %(var)
    #log = pandas.io.parsers.read_csv(logfile)
    log = np.genfromtxt(logfile, delimiter=',', names = True, dtype=None)
    sxmin = extent[0]
    symin = extent[1]
    sxmax = extent[2]=====
    symax = extent[3]
    filter_ul = log[(log['minX']<sxmin) & (log['maxX']>sxmin) & (log['minY']<symin) & (log['maxY']>symin)]
    filter_lr = log[(log['minX']<sxmax) & (log['maxX']>sxmax) & (log['minY']<symin) & (log['maxY']>symax)]
    filter_ur = log[(log['minX']<sxmin) & (log['maxX']>sxmax) & (log['minY']<symin) & (log['maxY']>symax)]
    filter_ll = log[(log['minX']<sxmax) & (log['maxX']>sxmax) & (log['minY']<symin) & (log['maxY']>symin)]
    f_ul = str(filter_ul['filename'][0])
    f_lr = str(filter_lr['filename'][0])
    f_ur = str(filter_ur['filename'][0])=====
    f_ll = str(filter_ll['filename'][0])
    if (f_ul == f_lr): #upper corner matches lower corner=====
        fname_list = np.array([f_ul])
```

279 Lines: ~12 hours to run entire U.S.

PostGIS SQL Code to Extract Forest Loss by Watershed:
Efficient but a Steep Learning Curve

```
*new 2 - Notepad++
File Edit Search View Encoding Language Settings Macro Run Plugins Window ? X
File Edit Search View Encoding Language Settings Macro Run Plugins Window ? X
new 2 x
1 CREATE TABLE public.ForChYr_CountInt1 AS CRLF
2   SELECT CRLF
3     gage_id, (value_count).value, SUM((value_count).count) AS CRLF
4       count CRLF
5   FROM CRLF
6   (CRLF
7     SELECT CRLF
8       gage_id, CRLF
9       rid, CRLF
10      ST_ValueCount (CRLF
11        ST_Union(ST_Clip(rast, geom, TRUE)), 1, FALSE, ARRAY[0
12          , 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12] CRLF
13        ) value_count CRLF
14   FROM CRLF
15     (SELECT gage_id, geom FROM basinsse) v, CRLF
16     (SELECT rid, rast FROM forch_base90pct_int) r CRLF
17     WHERE ST_Intersects(rast, geom) CRLF
18     GROUP BY gage_id, rid, geom CRLF
19     ) i CRLF
20   GROUP BY gage_id, value CRLF
21   ORDER BY gage_id, value ;
```

19 Lines; ~5 minutes to run for the entire Southeastern U.S.
(Some post-processing of tables in R)

Worked well for QA/QC: EXACT SAME ANSWERS!!!

To do...

- Collect remaining response/predictor data
- Extract new predictor variables to watershed shapes
- Build statistical models
 - Time series analysis
 - Pre/Post treatment analysis
- Poster at AGU (Dec 2014)

Lessons Learned

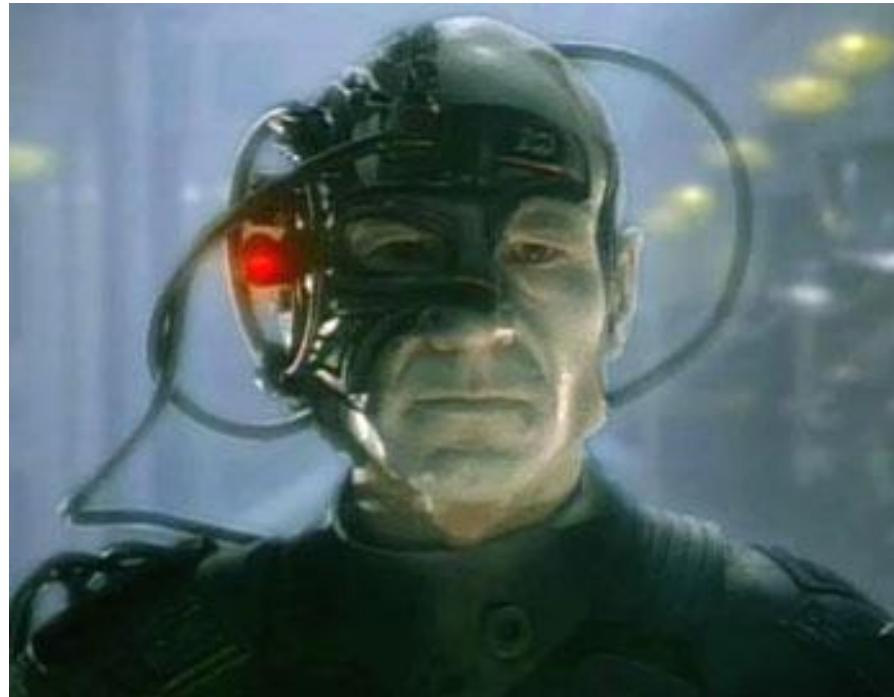
	COMMENT	DATE
O	CREATED MAIN LOOP & TIMING CONTROL	14 HOURS AGO
O	ENABLED CONFIG FILE PARSING	9 HOURS AGO
O	MISC BUGFIXES	5 HOURS AGO
O	CODE ADDITIONS/EDITS	4 HOURS AGO
O	MORE CODE	4 HOURS AGO
O	HERE HAVE CODE	4 HOURS AGO
O	AAAAAAA	3 HOURS AGO
O	ADKFJSLKDFJSOKLFJ	3 HOURS AGO
O	MY HANDS ARE TYPING WORDS	2 HOURS AGO
O	HAAAAAAAAANDS	2 HOURS AGO

AS A PROJECT DRAGS ON, MY GIT COMMIT MESSAGES GET LESS AND LESS INFORMATIVE.

(Try not to do this... and thanks to Terrell for helping us with GIT!)

Clone our repository!

<https://github.com/RENCI-Ecohydro/OSS-new>



You will be assimilated!