Descriptives: SEA and MacArthur

February 9, 2016

Contents

1 Data

Short Name	Long Name
dari	dist_to_all_rivers
dbri	dist_to_big_rivers
droa	dist_to_roads
selv	srtm_elevation
sslp	srtm_slope
am50	accessibility_map
gpw3	gpw_v3
nm6k	ndvi_max_mask_lt6k
lnye	ltdr_yearly_ndvi_mean
lnyx	ltdr_yearly_ndvi_max
alp4	v4avg_lights_x_pct
nlc4	v4composites
ncc4	v4composites_calibrated
at41	terrestrial_air_temperature_v4.01
pc41	terrestrial_precipitation_v4.01

Table 1: Variables used in this analysis.

Variable names are defined in Table 1.

2 Descriptives

Table 2: Cell level descriptive statistics (N=6188)

Statistic	N	Mean	St. Dev.	Min	Max
tc00_e	6,188	38.763	33.873	0	97
per_loss_e	6,188	9.071	16.994	0	97
dari_e	6,129	2,164.268	1,515.136	137.573	11,064.130
droa_e	6,131	25,433.280	223,696.500	255.029	4,460,040.000
selv_e	6,184	125.572	173.278	0.000	1,313.829
sslp_e	6,184	1.373	2.148	0.000	17.241
am50_e	6,181	385.607	286.001	4.342	1,828.021
gpw3_1990e	6,156	79.584	1,312.301	0.100	90,440.270
gpw3_1995e	6,156	93.991	1,552.600	0.119	107,002.400
gpw3_2000e	6,156	107.989	1,785.916	0.137	123,083.300
lnyx_mean	6,188	6,976.404	813.780	1,686.406	8,880.642
ncc4_mean	6,188	0.144	1.317	0.000	56.921
at41_mean	6,188	23.624	1.283	18.130	26.440
pc41_mean	6,188	131.166	31.431	108.389	290.746

Variables with 'mean' appendix are summaries of the yearly mean values.

Table 3: ADM2 level descriptive statistics (N=176)

Statistic	N	Mean	St. Dev.	Min	Max
tc00_e	176	23.753	25.328	0.000	92.824
per_loss_e	176	6.740	10.081	0.000	50.667
dari_e	175	2,167.215	754.885	665.675	5,706.997
droa_e	176	36,945.490	172,528.300	1,227.143	1,777,850.000
selv_e	176	76.027	112.420	3.331	694.954
sslp_e	176	0.900	1.230	0.062	6.133
am50_e	176	274.112	180.083	21.460	869.140
gpw3_1990e	176	157.126	369.933	1.151	3,754.542
gpw3_1995e	176	185.625	437.672	1.361	4,442.101
gpw3_2000e	176	213.272	503.438	1.566	5,109.679
lnyx_mean	176	6,557.321	780.621	4,459.846	8,207.006
ncc4_mean	176	0.490	2.186	0.000	24.470
at41_mean	176	23.902	1.142	20.912	26.440
pc41_mean	176	124.622	26.203	108.389	290.746

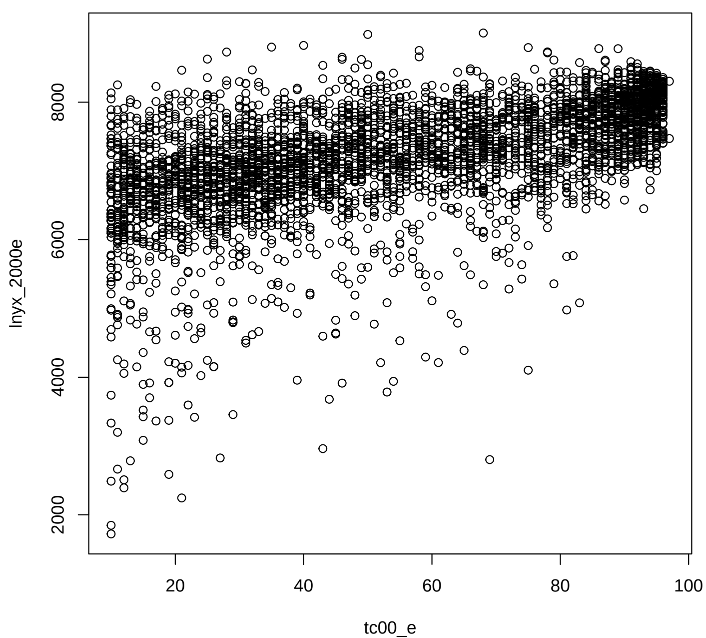
Variables with 'mean' appendix are summaries of the yearly mean values.
All variables represent the mean cell value within a ADM 2.

3 Forest Thresholds

Because the Hansen dataset begins in the year 2000, it is not possible to conduct long-term analyses using this dataset alone. To mitigate this, we select to leverage the longer-term record (1982 to present) offered by the NASA Long Term Data Record (LTDR). However, the LTDR has both a coarser resolution (5km, as opposed to 30 meters) as well as a different unit of measurement (Normalized Difference Vegetation Index) than the Hansen product. To take advantage of the information in Hansen, as well as the long term record provided by the LTDR, we fuse the two data products together using the single year of forest cover data provided by Hansen (2000).

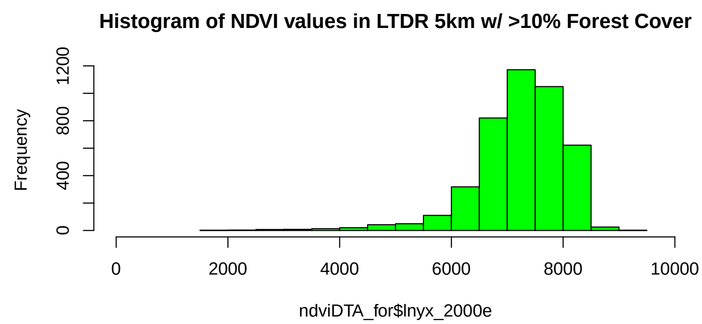
The fusion process aggregates the Hansen 2000 30 meter cells to calculate a mean % of forest cover estimated by Hansen within each 5km LTDR cell. We then contrast the mean annual values of NDVI present in LTDR to the forest cover values in Hansen, selecting only LTDR cells which have at least 10% forest cover according to the Hansen dataset.

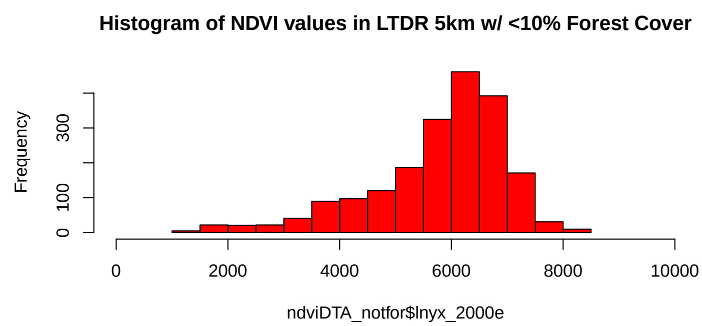
Figure 1: Relationship between LTDR 2000 NDVI and Hansen % Forest Cover 2000, of LTDR cells with at least 10% forest cover according to Hansen 2000.



The distribution of NDVI values for the remaining LTDR cells can be seen in figure 1. As expected, we see a strong correlation - areas with higher levels of NDVI also tend to have higher levels of forest cover as estimated by Hansen. Because LTDR and Hansen are not independent products (Hansen leverages the same input as LTDR in the product production), this analysis does not suggest that one product can be used to independently verify the other. Rather, here we illustrate that backcasting forest cover using LTDR can serve as a proxy to produce a longer time series of information roughly analogous to the higher-resolution approach to detecting forest cover Hansen employs.

Figure 2: Histogram of LTDR NDVI Values in (a) locations with >50% forest cover in Hansen 2000, and (b) locations with <10% forest cover in Hansen 2000.





Using this information, for any historic or future LTDR (5km) unit of observation, the probability each unit is forest or non-forest is calculated based on a maximum likelihood decision rule. The probability of each unit containing a forest is calculated, and all units which have a higher probability of inclusion into the ”forest” class than ”non-forest” class are used in later steps of the analysis. The estimated probability density function to estimate inclusion into the forested class is calculated using the following equation:

$$\rho(x|c_i) = \frac{1}{(2\pi)^{1/2} * \sigma_i} * exp(-\frac{1}{2} * \frac{(x-\mu_i)^2}{\sigma_i^2}) \quad (1)$$

where x is the LTDR value of the unit being classified as forest or non-forest (represented as classes c), μ_i is the estimated mean of all LTDR values determined to be included in class i and σ_i^2 is the estimated variance of each class, i.e. $x \in forest(c_1)$when:

$$\rho(x|c_1) * \rho(c_1) \geq \rho(x|c_0) * \rho(c_0) \tag{2}$$

where prior probabilities $\rho(c_i)$ are established based on the probability of a LTDR pixel being classified as forest or non-forest in the year 2000 (i.e., the percent of the landscape occupied by each class).