Descriptives: SEA and MacArthur

February 22, 2016

Contents

WARNING: Currently, only 3000 cells (about one-half of all cells) are being used in this analysis, due to computational limitations. Please ignore the analysis below this line, for now.

1 Data

Short Name	Long Name
dari	dist_to_all_rivers
tc00	hansen200_tree_cover
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	39.216	33.918	0.000	97.590
per_loss_e	6,188	9.439	17.065	0.000	97.870
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	129.491	31.102	106.203	285.172

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	24.160	25.398	0.000	93.291
per_loss_e	176	7.046	10.188	0.000	51.251
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	122.580	26.040	106.203	285.172

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

Figure 1: Study area (ADM2) overlayed with Chinese development finance project locations.

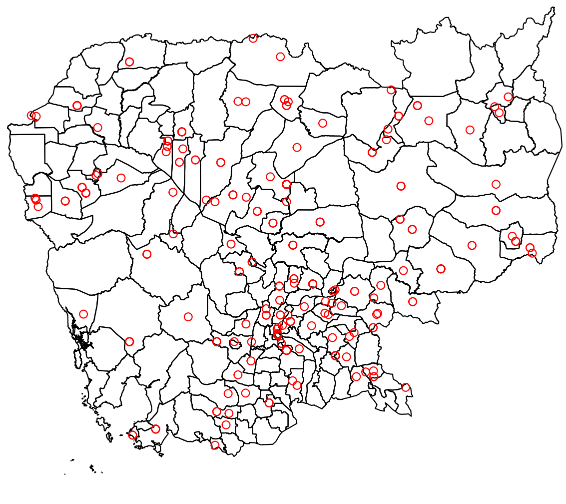
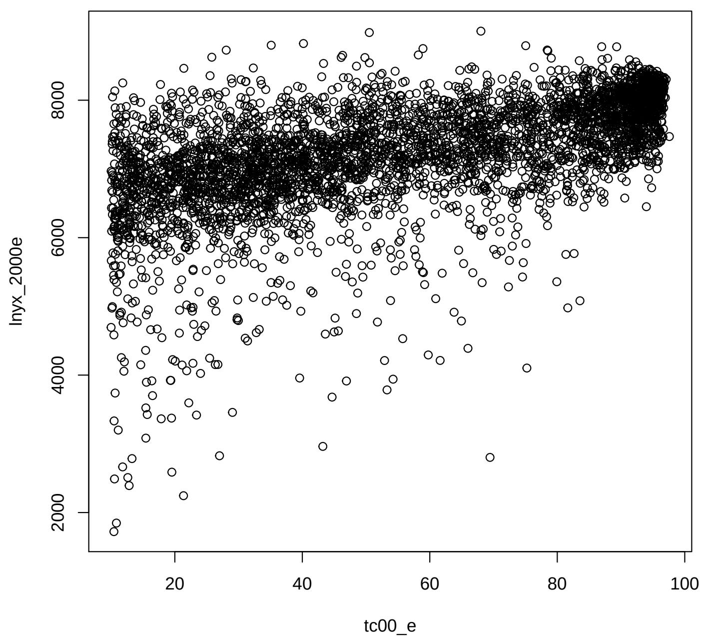
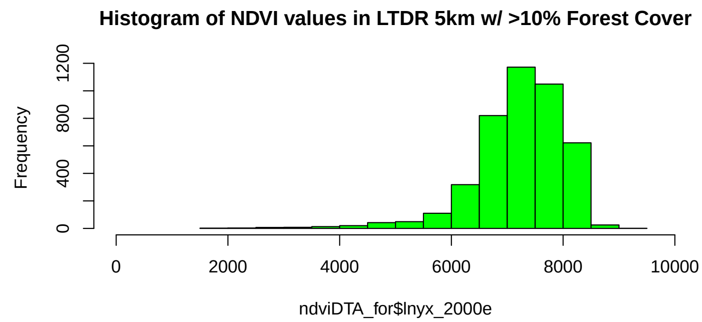


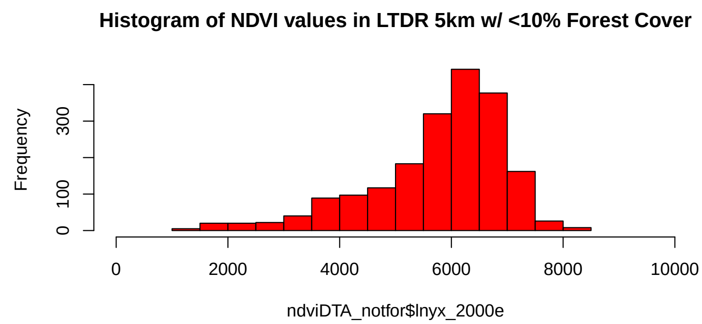
Figure 2: Relationship between LTDR 2000 NDVI and Hansen % Forest Cover 2000, of LTDR cells with at least 10% forest cover according to Hansen 2000.



The distribution of NDVI values for the remaining LTDR cells can be seen in figure 2. 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 historic LTDR trends can be relevant for establishing baselines when Hansen outcome variables are considered.

Figure 3: Histogram of LTDR NDVI Values in (a) locations with >10% forest cover in Hansen 2000, and (b) locations with <10% forest cover in Hansen 2000.





3 Methods

From each cell defined as forest (cells with > 10% forest cover according to Hansen 2000), the euclidean distance (Haversine) to each chinese investment site (N = 414) is calculated and recorded.

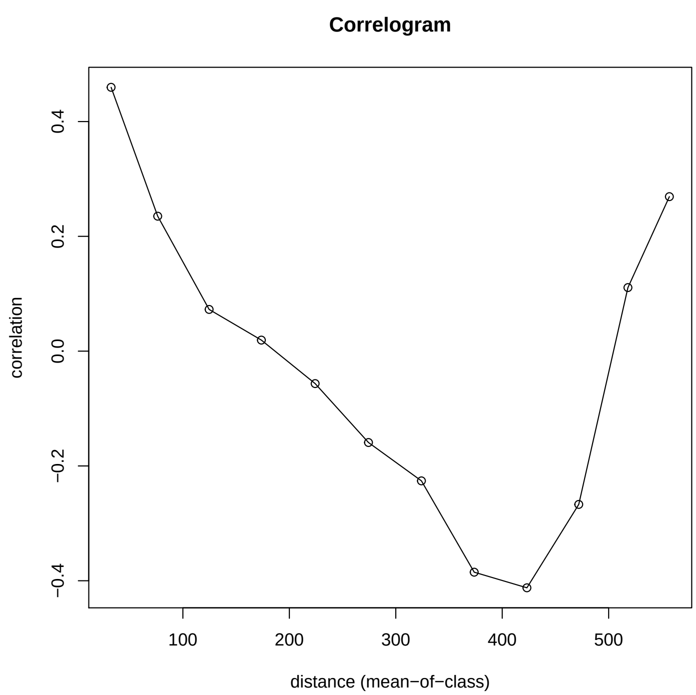
On average, the minimum distance between a unit of observation (the cell) and each chinese investment site is 26km, with an average distance of 219km.

DRAFT To establish the degree of impact Chinese investment had on each grid cell, a weighted distance-decay function is approximated. The spatial autocorrelation of forest tree cover is approximated using the 1999 LTDR dataset to avoid any potential confounds with treatments that start circa 2000. By examining the spatial autocorrelation of forest tree cover, we hope to establish the maximum distance at which spillover effects might feasibly be observed (as the spatial pattern of tree cover in 1999 is the produce of all preceeding impacts). Spatial autocorrelation is examined over 50 kilometer steps - for example, all cells within 50km are contrasted to one another, and the correlation of forest cover is recorded. This is then repeated in bands, i.e., 50-100km; 100-150km and so forth until the 1000km limit is reached. At each distance band, a summary measure of spatial autocorrelation - Morans I - is calculated, following:

$$I_h = (\frac{N}{\sum_i^N \sum_j^N w_{ij}}) * ((\sum_i^N \sum_j^N w_{ij} * (X_i - \bar{x}) * \frac{X_j - \bar{x}}{\sum_i^N (X_i - \bar{x})})^2) \qquad (1)$$

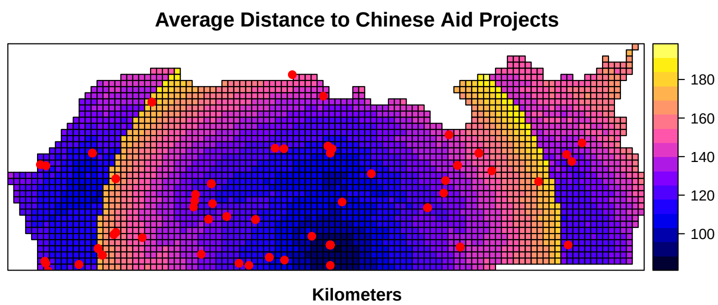
where h represents each spatial bin, N the number of spatial units, i and j are indexes for each unit, X is the variable of itnerest, and W_{ij} represents the weights matrix. In this application, the weights matrix is specified according to the bin (h) being analyzed. A best-fit gaussian spatial decay function is then found to approximate the distance decay observed on the Morans I measurements. Based on this distance-decay function, the estimated impact of chinese projects is weighted based on the cumulative distance-weighted average across all chinese projects.

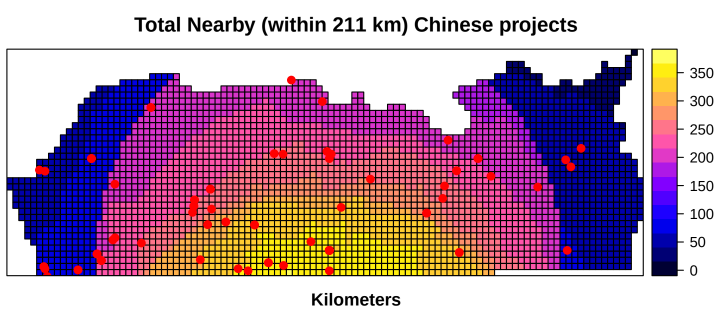
Figure 4: Distance decay of spatial autocorrelation observed in 1999 Forest Cover, as measured using the LTDR dataset.



As this correlogram illustrates, a lack of positive spatial autocorrelation among LTDR measurements in 1999 is first observed at approximately 211 km. We use this distance as a threshold to screen nearby Chinese aid projects - i.e., the average and sum distance of all Chinese aid projects within 211 km is taken for each cell, and projects beyond that distance are not included. These two values - the average and summed distance of Chinese aid projects - are used as our approximation of the strength of Chinese interventions within any given cell on the landscape.

Figure 5: Average Distance, Total Distance, and Total Count of Projects within 211 of each cell.





Functional Form of Analysis:

$$Hansen_{it} = \alpha + \theta * WeightedDistanceActiveChineseProject_{it} + \sum(\beta_j * X_j) + D_{region} + D_t + D_{region} * t + \epsilon_{it} \tag{2}$$