

Appendix

1 Data source and collection

Variables	Description
Forest Cover	NASA Long Term Data Record measurements of vegetative cover http://ltdr.nascom.nasa.gov/cgi-bin/ltdr/ltdrPage.cgi
World Bank Project Locations	Double-blind geocoded information on the geographic location of each World Bank project http://aiddata.org/level1/geocoded/worldbank
Distance to Rivers	The calculated average distance to all rivers http://hydrosheds.cr.usgs.gov/index.php
Distance to Commercial Rivers	Calculated average distance to all commercial rivers http://hydrosheds.cr.usgs.gov/index.php
Distance to Roads	Distance to nearest road http://sedac.ciesin.columbia.edu/data/set/groads-global-roads-open-access-v1
Elevation, Slope	Elevation and slope data measured from the Shuttle Radar Topography Mission http://www2.jpl.nasa.gov/srtm/
Accessibility to Urban Areas	European Commission Joint Research Centre estimation of urban travel times. http://forobs.jrc.ec.europa.eu/products/gam/download.php
Population Density	Center for International Earth Science estimation of population density, derived from Nighttime Lights http://sedac.ciesin.columbia.edu/data/collection/gpw-v3
Air Temperature, Precipitation	University of Delaware Long term, global temperature and precipitation data interpolated from weather station measurements. http://climate.geog.udel.edu/climate/

Table 1: Covariates of the data sets for World Bank projects

2 Variable importance and covariate correlation

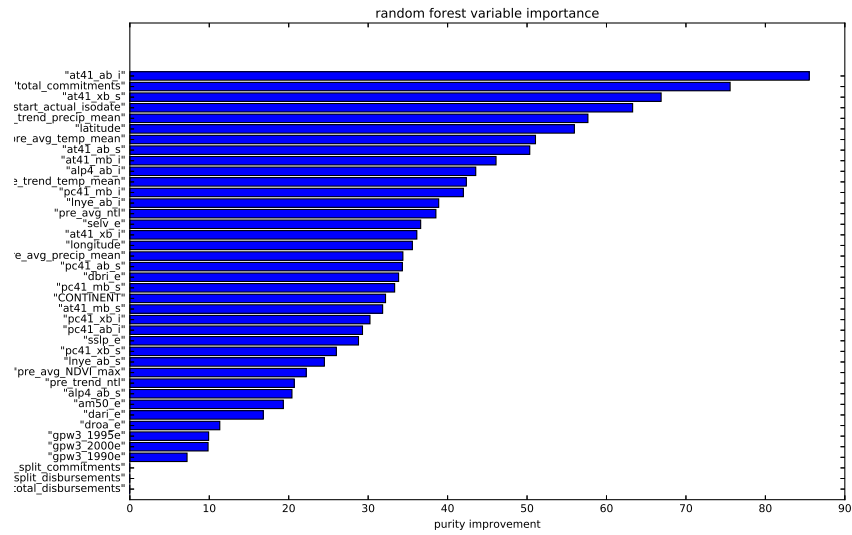


Fig. 1: Covariates ranked by the sum of purity improvements they achieve across all trees

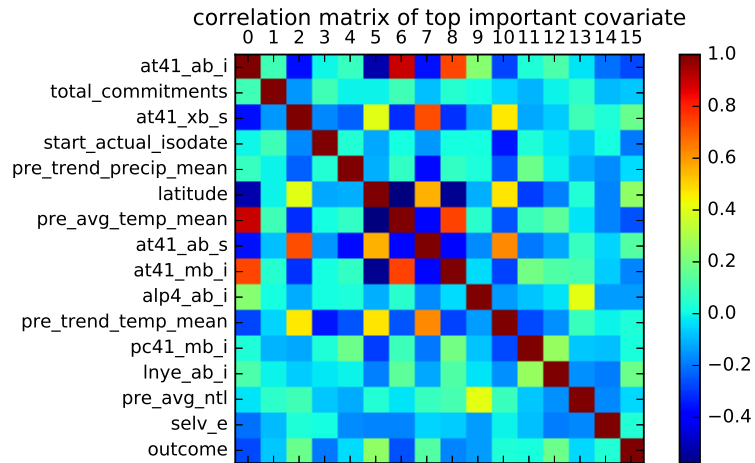


Fig. 2: Heat map of the correlation between the values of the 15 most highly ranked covariates