Supplementary material - An aggregate learning approach for interpretable semi-supervised population prediction and disaggregation using ancillary data

April 6, 2019

1 Datasets and source code

Datasets, results and source code are available on Zenodo.

Note that some part of the dataset (the original, unprocessed data) cannot be distributed directly by the authors (regardless of their will), as they are owned by third parties who set licenses that prevent redistribution. However, all data is publicly available on the web (see below). The source and method to preprocess the unredistributable data is indicated in the public version of the dataset.

To reproduce our experiments, you will however only need the source code and the preprocessed dataset, which are publicly available. The output of our experiments, before the analysis made in the last section of the paper, is also available.

- `https://zenodo.org/record/1313600`

- `https://zenodo.org/record/1313627`

- `https://zenodo.org/record/1313643`

- `https://zenodo.org/record/1313613`

- `https://zenodo.org/record/1313635`

2 Sources for the dataset

Table 1: Features used.

Feature code	Source	Type	Description
protected_area_cls	[9]	vector	Presence of protected natural areas
protected_area_dst			Distance to nearest protected natural area
water_cls	NGA VMAP0 [6]	vector	Presence of inland water
water_dst			Distance to nearest inland water area
populated_places_cls			Indicates if pixel is in a known highly populated place
populated_places_dst			Distance to nearest populated place
rivers_cls			Presence of rivers
rivers_dst			Distance to nearest rivers
health_cls	[1]	vector	Presence of health facility
health_dst			Distance to nearest health facility
temperature_mean	WorldClim 1.4. [3]	raster	Mean temperature
precipitations			Mean annual precipitations
nightlights	VIIRS Day/Night Band Nighttime Lights. [2]	raster	Visible night lights
NPP_GPP	Terra/MODIS Net Primary Production (MOD17A3). [8]	raster	Net Primary Production
road_cls	OSM. [7]	vector	Presence of roads
road_dst			Distance to nearest road
cover_deciduous_forest_cls	MDA BaseVue 2013. [5]	raster	Indicates that the pixel has a specific class. See [5] for description of the classes.
cover_evergreen_forest_cls			
cover_shrub_scrub_cls			
cover_grassland_cls			
cover_minimal_veg_cls			
cover_agriculture_cls			
cover_agriculture_paddy_cls			
cover_wetland_cls			
cover_mangrove_cls			
cover_water_cls			
cover_high_density_urban_cls			
cover_mediumlow_density_urban_cls			
cover_deciduous_forest_dst			Distance to the nearest pixel that has a specific class. See [5] for description of the classes.
cover_evergreen_forest_dst			
cover_shrub_scrub_dst			
cover_grassland_dst			
cover_minimal_veg_dst			
cover_agriculture_dst			
cover_agriculture_paddy_dst			
cover_wetland_dst			
cover_mangrove_dst			
cover_water_dst			
cover_high_density_urban_dst			
cover_mediumlow_density_urban_dst			
elevation	HydroSHEDS. [4]	raster	Mean elevation
slope			Mean slope (derived from elevation)

References

[1] Cambodia Ministry of Health and OCHA ROAP: Location of health facilities. `https://data.humdata.org/dataset/cambodia-health` (jan 2010)

[2] Earth Observation Group, NOAA National Geophysical Data Center: Version 1 viirs day/night band nighttime lights; svdnb_npp_20150101-20151231_75n060e_vcm-orm-ntl_v10_c201701311200.avg_rade9. `https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html` (jan 2017)

[3] Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A.: Very high resolution interpolated climate surfaces for global land areas. International journal of climatology **25**(15), 1965–1978 (2005)

[4] Lehner, B., Verdin, K., Jarvis, A.: New global hydrography derived from spaceborne elevation data. Eos, Transactions American Geophysical Union **89**(10), 93–94 (2008)

[5] MDA: Basevue 2013 (jan 2017), `http://www.arcgis.com/home/item.html?id=1770449f11df418db482a14df4ac26eb`

[6] National Geospatial-Intelligence Agency (NGA): Vector map level 0 (vmap0). `http://geoengine.nga.mil/geospatial/SW_TOOLS/NIMAMUSE/webinter/rast_roam.html` (oct 2005)

[7] OpenStreetMap contributors: Dump retrieved from https://download.geofabrik.de. `https://download.geofabrik.de` (2017)

[8] Running, S.W., Zhao, M.: Daily gpp and annual npp (mod17a2/a3) products nasa earth observing system modis land algorithm (2015)

[9] UNEP-WCMC: Protected area profile for cambodia. World Database of Protected Areas. Available at: `www.protectedplanet.net` (oct 2017)