Covariates for mapping and modelling arboviruses in Brazil

Short description and relevance: This is a guidebook to all covariate data, download links and references, to be collected and processed so they can be integrated readily into the analysis of epidemiological and genomic data in Sao Paulo and Brazil. We aim to match covariate data to the whole time frame we have access to epidemiological data (i.e., 01/01/2007 - real-time). Similarly, this document should become the reference for any analysis done and should make it easier when compiling relevant materials for research. All information in here should be updated as new data comes in and recorded in a format that all participants can easily understand. Information that is redundant or irrelevant should be deleted. Covariates are grouped into the following five categories:

i. Geographic data (administrative boundaries),
ii. Climatic and environmental data,
iii. Socio-economic and demographic data,
iv. Health system and health spending data,
v. Disease intervention and prevention data.

At the end of this document there are short descriptions of the files listed here (e.g., what is a shapefile, what is a raster file etc.).

File location: Links to files can be stored and directly uploaded at this link. The folder structure should be identical to the structure proposed in the google doc: https://drive.google.com/drive/folders/1JXDYRfnYiOlfJy0zgmkssdw11nFwJxR-?usp=sharing

i. Geographic data. Geographic data, like state and administrative boundaries, are important so all data can be matched to the same spatial resolution and identified across datasets.

	Name of covariate	Spatial resolution	Reference	Download link and status (processed yes/no)	Visualisation link	Type of data (shapefile, raster, csv, other)
1	State boundaries	Admin 1		ftp://geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_		Shapefile

				municipais/municipio_2015/Brasil/BR/ Download file “br_unidades_da_federacao.zip”		
2	Municipality boundaries	Admin 2		ftp://geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_municipais/municipio_2015/UFs/SP/ Download file “sp_municipios.zip” For the whole country: ftp://geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_municipais/municipio_2015/Brasil/BR/BR.zip		Shapefile
3	Sectors used in Census 2010	Sectors		ftp://geoftp.ibge.gov.br/organizacao_do_territorio/malha		Shapefile

				s_territoriais/malhas_de_setores_censitarios__divisoes_intramunicipais/censo_2010/setores_censitarios_shp/ Choose the state initials (e.g. “sp”), then download “sp_setores_censitarios.zip” for sectors used in census 2010 (higher resolution than subdistricts)		
4	Districts used in Census 2010 Other boundaries??	Districts		ftp://geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_de_setores_censitarios__divisoes_intramunicipais/censo_2010/setores_censitarios_shp/ Choose the state initials (e.g. “sp”), then		Shapefile

				download “sp_distritos.zip” for districts		
5	Subdistricts used in Census 2010	Subdistricts		ftp://geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_de_setores_censitarios__divisoes_intramunicipais/censo_2010/setores_censitarios_shp/ Choose the state initials (e.g. “sp”), then download “sp_subdistritos.zip” for subdistricts		Shapefile
6	CEP post codes	CEP				Shapefile

ii. Climatic and environmental data. Arbovirus transmission is strongly influenced and determined by climatic factors. We here collate all available climatic data from both satellite imagery and ground weather stations relevant for the analysis of all CADDE projects.

	Name of covariate	Spatial resolution	Temporal resolution	Reference with online link	Download link and status (processed yes/no)	Type of data (shapefile, raster, csv, other)
1	Precipitation (min,	Meteorolog	Ideally			

	mean, max)	ical station	same as case data, weekly?			
2	Vegetation coverage (EVI)	maximum available from remotely sensed images				
3	Temperature (daily minimum, mean and maximum)	Meteorological station or maximum available from remotely sensed images	Daily			
4	Aedes aegypti suitability	5km^2 global	Monthly, 2015	Kraemer et al. 2015 eLife https://elifesciences.org/articles/08347		raster
5	Aedes albopictus suitability	5km^2 global	Monthly, 2015	Kraemer et al. 2015 eLife https://elifesciences.org/articles/08347		raster
6	Aedes transmission potential	Same resolution as WorldClim V2 dataset (5, 10, 18, km^2), Brazil	Monthly, referring to period 1970-2000 as in climate data (WorldClim V2)	Ubolski et al. 2019 https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13205		raster
7	Haemagogus					

8	Sabethes					
9	Aloutta clamitans					
10	Callithrix jacchus					
11	Rivers in Brazil	1km^2	Giachetta et al. 2018 https://www.nature.com/articles/sdata2018127			shapefile
12	Relative humidity					

iii. Socio-economic data. Economic opportunities and factors influence the severity and transmission intensity of arboviruses. We here aim to collect information about for example income and quality of housing. Some materials will come from satellite imagery, others may come from government reports.

	Name of covariate	Spatial resolution	Temporal resolution	Reference	Download link and status (processed yes/no)	Type of data (shapefile, raster, csv, other)
1	Population density	100m	2019	https://data.humdata.org/dataset/brazil-high-resolution-population-density-maps-demographic-estimates		raster
2	Population count	100m	2019	https://data.humdata.org/dataset/brazil-high-		raster

				resolution-population-density-maps-demographic-estimates		
3	Population count by age class	100m	2019	https://data.humdata.org/dataset/brazil-high-resolution-population-density-maps-demographic-estimates		raster
4	Urban accessibility	1km^2	2016	Weiss et al. 2018 https://www.ncbi.nlm.nih.gov/pubmed/29320477	https://map.ox.ac.uk/research-project/accessibility_to_cities/	raster
5	Income (GDP per capita PPP)	30km^2	1990 - 2015 (every 5 years)	Kummu et al. 2018 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5800392/		raster
6	Conflict	Latitude and longitude	1980 - 2019	https://ucdp.uu.se/		csv
7	Location of river dams	Latitude and longitude	2011	https://sedac.ciesin.columbia.edu/data/set/grand-v1-dams-rev01/maps		shapefile
8	Nighttime lights					raster

	(proxy for income)					
9	Human mobility	Municipality level (admin 2)	TBC	TBC	TBC	TBC
10	Net migration	Municipality level (admin 2)	TBC Andreza CNM	TBC Andreza CNM	TBC Andreza CNM	TBC Andreza CNM
11	Deforestation	Cristina Banks-Leite (INPE)	X	X	X	X
12	Safe drinking water index	Municipality (admin 2)	Yearly	For methods: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002755#sec023	For data: https://figshare.com/articles/Microcephaly_and_potential_exposure_data_in_Brazil_2015-2017/7359197/1	csv
13	Living conditions (sanitation, housing characteristics, electricity, garbage collection, etc)	IBGE				

iv. Health system and spending data. Seeking care during outbreaks is important to reduce the burden of arboviruses. We here aim to collect information about locations of hospitals and other resources related to health care spending.

	Name of covariate	Spatial resolution	Temporal resolution	Reference	Download link and status (processed yes/no)	Type of data (shapefile, raster, csv, other)

1	Locations of hospitals	Latitude and longitude	?	?		csv
2	Health spending	?	?	?		
3	others?					
4	Number of health clinics per municipality	TBC Andreza CNM				
5	Tourism index (sylvatic transmission risk)					

http://saude.sp.gov.br/links/matriz
http://cnes2.datasus.gov.br/
http://www.atlasbrasil.org.br/2013/
ftp://ftp.ibge.gov.br/
http://www.imp.seade.gov.br/frontend/#/tabelas

v. Disease intervention and prevention data. Governments and local organizations as well as the public are performing interventions to curb transmission of arboviruses during outbreaks but also to prevent outbreaks in the future. We here aim to collect data in real-time on the interventions carried out.

	Name of covariate	Spatial resolution	Temporal resolution	Reference	Download link and status (processed yes/no)	Type of data (shapefile, raster, csv, other)
1	Yellow fever vaccination maps	Municipality level	2016	Shearer et al.		raster
2	Yellow fever vaccination doses during campaign?					
3	Dengue interventions?					
4	Mosquito					

	campaigns/ interventions					
5	Mosquito control efforts (no. people working per municipality)					
6	List of municipalities with surveillance of non-human primates					

Definitions and further information
This section is aimed at providing more information and links about the data-types that are used for this project.

Raster files: In its simplest form, a raster consists of a matrix of cells (or pixels) organized into rows and columns (or a grid) where each cell contains a value representing information, such as temperature. Rasters are digital aerial photographs, imagery from satellites, digital pictures, or even scanned maps. Please see this link for more information: http://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/what-is-raster-data.htm

Shapefiles: A shapefile is a simple, nontopological format for storing the geometric location and attribute information of geographic features. Geographic features in a shapefile can be represented by points, lines, or polygons (areas). The workspace containing shapefiles may also contain dBASE tables, which can store additional attributes that can be joined to a shapefile's features. Please see this link for more information: http://desktop.arcgis.com/en/arcmap/10.3/manage-data/shapefiles/what-is-a-shapefile.htm

CSV: Tables and excel files should contain the following information ranked according to desirability.

1. Latitude and Longitude of location of collection
2. Municipality (or other geographic indicator) of collection
3. Date of collection