Description about Data for Japanese Encephalitis Burden Project

Duy M. Nguyen and Quan M. Tran

July 8, 2018

1 Introduction

Japanese Encephalitis (JE) is a vector-borne disease whose transmission is found across temperate regions of Asia. There are more than 3 billion people living in endemic areas [1]. We decided to choose endemic countries from both CDC [2] and WHO [3] (excluding Guam and Saipan) those are: Australia, Bangladesh, Bhutan, Brunei, Burma (Myanmar), Cambodia, China (including Macao and Hong Kong), India, Indonesia, Japan, Laos, Malaysia, Nepal, North Korea, Pakistan, Papua New Guinea, Philippines, Russia, Singapore, South Korea, Sri Lanka, Taiwan, Thailand, Timor- Leste and Vietnam. Figure 1 illustrates the current endemic regions that we are working on.

Figure 1: Current regions of interest

It is noted that the collected spatial information described in following sections was created at different time or period. It also has distinct resolution and was based on either consensus or modelled data. Therefore, some of spatial maps can be very detailed and intensive in download size, while others are of small size and have missing information in several areas.

2 Data Description

2.1 Administrative Layer

One of our datasets is the administrative layers which are collected from GADM website [4]. The current version of data when it is downloaded is 3.6, which was released on 2018 May 6th. One of the advantages of the website is the data is always updated frequently, every 3-6 months. The data is stored in Vector Layer format.

2.2 World Climate

WorldClim[5] provides various information about climate such as temperature, precipitaion or wind speed collected from 1970 to 2000. However, we only use bioclimatic variables derived from rainfall and temperature on a monthly basis. These variables can be used to generate more biologically meaningful data and often utilized in species distribution modelling or related ecological modelling techniques. These data consist of 19 spatial maps for 19 bioclimatic variables covering the entire world with the resolution of 30 seconds $\times$ 30 seconds. Figure 2 depicts the global annual mean temperature in the period between 1970 and 2000, in which red color means high temperature while blue represents low temperature.

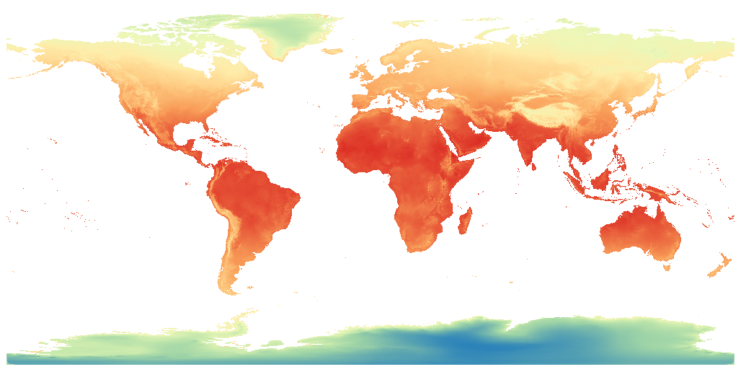

Figure 2: Annual Mean Temperature

2.3 Japanese Encephalitis Vector Distribution

Longbottom et. al. [1] collated a comprehensive database of presence records of *Culex tritaeniorhynchus*, one of the vectors of JE, within endemic regions. The data was collected from various formal resources spanning from 1928 to 2014. They then generated the map of environmental suitability for these mosquitoes. By using an ensemble of boosted regression tree models and combining occurrence dataset with environmental and socio-economic variables, the authors produced the fine-scale map of suitability of this vector ranging from 0 to 1 within risk areas with the resolution of 5km $\times$ 5km. This map offers insights into the spatial epidemiology of JE. Figure 3 shows the result of [1] in which areas of interest are quite different to ours. Our endemic regions include the West of China while Longbottom et. al. did not take the area into account.

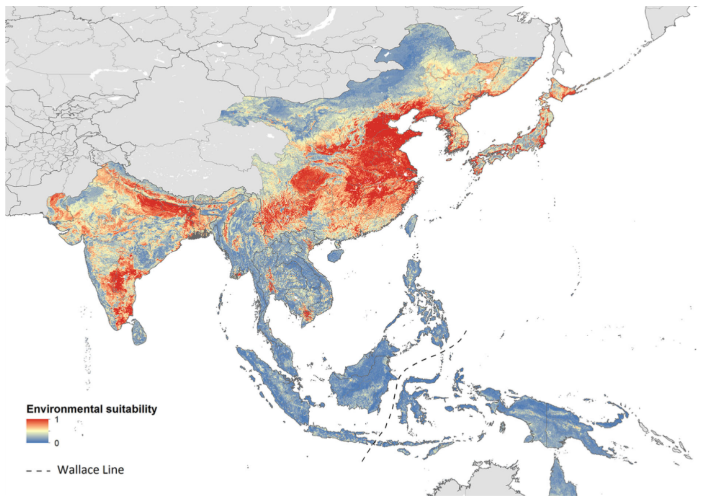

Figure 3: Projected environmental suitability for Culex vector within areas at risk of JE transmission

2.4 Rice distribution

The website MapSPAM [6], which is a platform for sharing results from the Spatial Production Allocation Model (SPAM) by HarvestChoice, provides several cropland distribution map including rice. SPAM is an effective approach to map intricate crop production patterns by applying cross-entropy method but does not require a huge amount of input data. The outcomes of SPAM are the estimation of 42 different crop distributions that are global level spatial maps with the resolution of each cell is 5 minutes $\times$ 5 minutes. The predicted values were adjusted to match with the consensus carried out by FAO for the average from 2004 to 2006.

An output of SPAM consists of *physical area* from which *harvested area*, *yield* and *production* can be calculated. Since the *Culex tritaeniorhynchus* usually breed in rice paddies [7], *physical area* of rice is taken into account (as shown in Figure 4). This variable represents the actual area where rice is grown in a pixel while *harvested area* multiplies a physical area in a pixel with harvests of a crop on the same plot. Both outcomes are measured in hectare.

Figure 4: Distribution of rice paddies area (cropped within ASEAN region) in which green stands for large area of rice fields in that cell

2.5 Livestock distribution

Because pigs play a critical role in the enzootic cycle of JEV [7], the distribution of this livestock also needs to be considered. For the pigs distribution, we use data from the Gridded Livestock of the World (GLW) v2.0 [8, 9]. The Geo-Wiki provides a revised and updated global map of livestock distributions, comprising cattle, chicken, duck, pig, sheep and goat, and production systems in a multi-partner collaboration centered on ILRI, FAO and ULB-LUBIES. The GLW produces a modelled global density of pigs which was adjusted to match FAOSTAT values in 2006. The result map has a resolution of 30 seconds $\times$ 30 seconds and it is depicted in Figure 5.

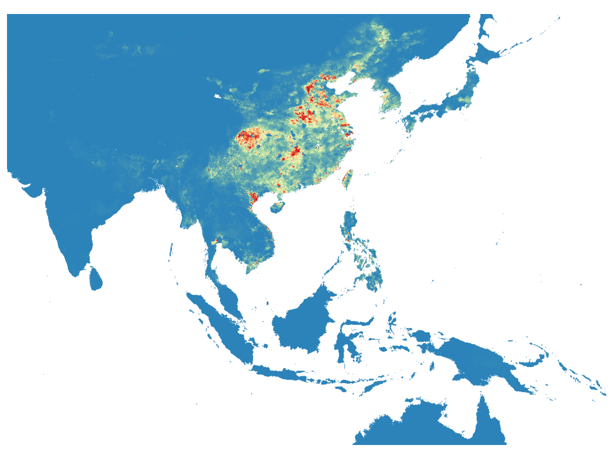

Figure 5: Spatial map of pig density (cropped within endemic regions)

2.6 Population density

The Socioeconomic Data and Applications Center (SEDAC) [10] is one of archive centers of the U.S. National Aeronautics and Space Administration. The purpose of SEDAC is to create and employ applications supporting the integration of socioeconomic and earth science data by focusing on interactions of humankind with the environment. SEDAC provides estimations of population density from 2000 to 2020 [11] based on counts consistent with national censuses. The projected population values were refined to match the 2015 Revision of the United Nation's World Population Prospects country total. Ultimately, the density was calculated by dividing population by the corresponding land area and provided as a global spatial raster at 30 seconds $\times$ 30 seconds as shown in Figure 6.

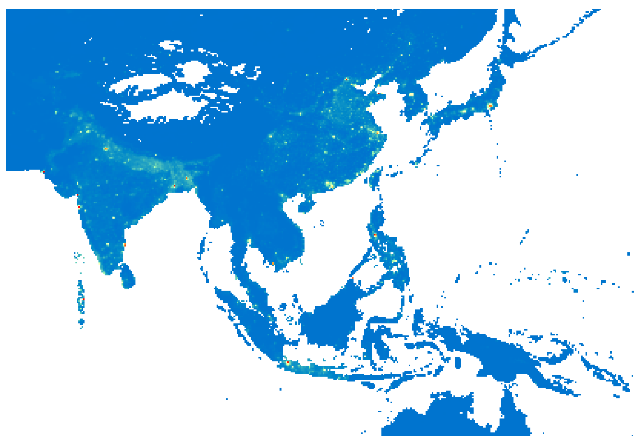

Figure 6: Map of adjusted population density in 2015 (cropped within endemic regions)

2.7 Specific age group density

SEDAC also supports projection of population counts as well as densities by age and sex for the year 2010 which are consistent with official censuses and population registers at national level [12]. They calculated the portions of males and females for each specific age bracket from age 0 to over 85 based on the given census, then applied these numerical results to the estimation of the total population in the year 2010 to achieve the population by sex and age. The outcomes provided are in the format of global rasters at 30 seconds $\times$ 30 seconds. Here we downloaded the data of each 5 - year age group ranging from 0 - 4 to 85+. Figure 7 illustrates the density of children age group (0 - 14) within the JE endemic areas.

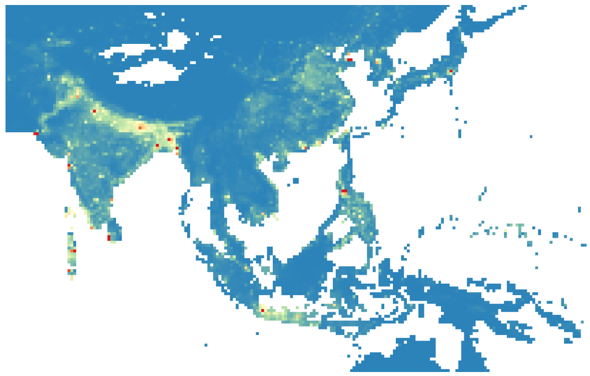

Figure 7: Density of children who are below 14 year old in the endemic regions

2.8 Urban/Rural category

Since the risk of JE transmission has a tendancy to be higher in rural areas where pigs are abundant and agricultural practices are dominant [13], the category of urban and rural in each cell is expected to contain key information about disease transmission. The Urban Extents Grid v1 (1995) [14, 15], can classify rural and urban areas by analysing the total population, settlement points, and the presence of Nighttime Lights. There are 2 conditions for a pixel to be classified as urban in GRUMP. The total population at that pixel has to be higher than 5000 individuals. The second condition is that the pixel is identified as urban extent, which is estimated based on settlement points, or is contiguous to lighted cells from the Nighttime Lights map. The result is a raster representation of urban mask by 30 seconds grid as shown in Figure 8 (cropped within endemic regions)

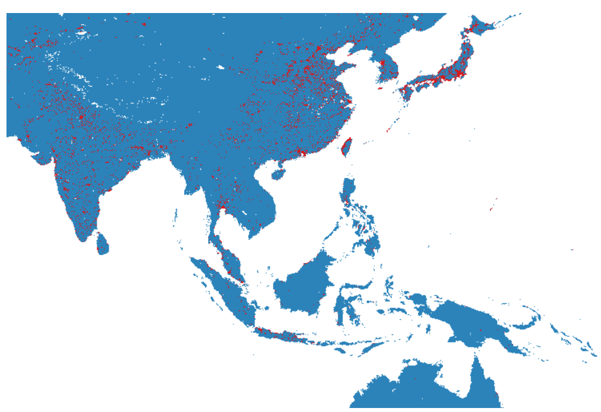

Figure 8: Urban areas which are classified as red cells in the endemic regions

2.9 Elevation

Elevation noticeably affects the JEV transmission in the light of the negative correlation between elevation and the presence of potential JE vectors; moreover, almost no JE cases were recorded above 1200m elevation [16]. Therefore the altitude data should be taken into account for the evaluation of the environmental suitability. We use the data produced by the NASA Shuttle Radar Topographic Mission (SRTM) [17]. The latest version, which is SRTM version 4.1, is considered to be a significant improvement from previous versions and is the highest quality SRTM dataset available. The original data is available as global raster at 3 seconds $\times$ 3 seconds but to enable faster processing it is aggregated to 7.5 seconds $\times$ 7.5 seconds, which is the format we downloaded to use.

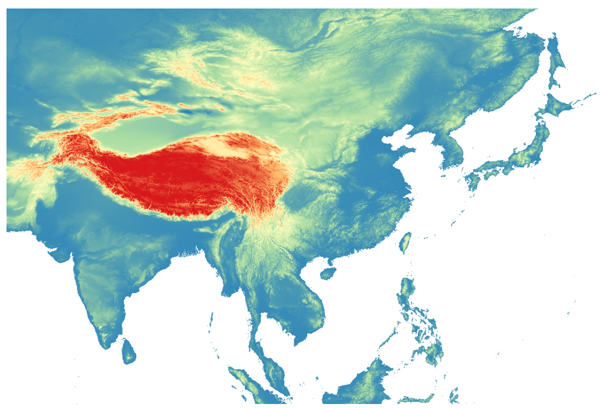

Figure 9: Map of Elevation in which red cells reflect areas at higher altitude

References

[1] J. Longbottom, A. J. Browne, D. M. Pigott, M. E. Sinka, N. Golding, S. I. Hay, C. L. Moyes, and F. M. Shearer, "Mapping the spatial distribution of the japanese encephalitis vector, culex tritaeniorhynchus giles, 1901 (diptera: Culicidae) within areas of japanese encephalitis risk," *Parasites & Vectors*, vol. 10, no. 1, mar 2017. [Online]. Available: https://doi.org/10.1186/s13071-017-2086-8

[2] "Japanese encephalitis - chapter 3 - 2018 - yellow book — travelers' health — cdc," https://wwwnc.cdc.gov/travel/yellowbook/2018/infectious-diseases-related-to-travel/japanese-encephalitis, accessed: 2018-05-17.

[3] "Who — japanese encephalitis," http://www.who.int/ith/diseases/japanese_encephalitis/en/, accessed: 2018-05-17.

[4] "Gadm maps and data," https://gadm.org/data.html, accessed: 2018-05-17.

[5] "Worldclim-global climate data," http://worldclim.org/version2, accessed: 2018-05-17.

[6] "Mapspam," http://mapspam.info/, accessed: 2018-05-17.

[7] T. Solomon, "Control of japanese encephalitis — within our grasp?" *New England Journal of Medicine*, vol. 355, no. 9, pp. 869–871, aug 2006. [Online]. Available: https://doi.org/10.1056/nejmp058263

[8] "Welcome to livestock geo-wiki project," https://livestock.geo-wiki.org/home-2/, accessed: 2018-05-17.

[9] T. P. Robinson, G. R. W. Wint, G. Conchedda, T. P. V. Boeckel, V. Ercoli, E. Palamara, G. Cinardi, L. D'Aietti, S. I. Hay, and M. Gilbert, "Mapping the global distribution of livestock," *PLoS ONE*, vol. 9, no. 5, p. e96084, may 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0096084

[10] "Socioeconomic data and application center," http://sedac.ciesin.columbia.edu/, accessed: 2018-05-20.

[11] Center For International Earth Science Information Network-CIESIN-Columbia University, "Gridded population of the world, version 4 (gpwv4): Population density adjusted to match 2015 revision of un wpp country totals, revision 10," 2017.

[12] ——, "Gridded population of the world, version 4 (gpwv4): Basic demographic characteristics, revision 10," 2017.

[13] "Japanese encephalitis," https://www.vaccineimpact.org/diseases/je/, accessed: 2018-05-21.

[14] Center For International Earth Science Information Network-CIESIN-Columbia University; International Food Policy Research Institute-IFPRI; The World Bank; Centro Internacional De Agricultura Tropical-CIAT, "Global rural-urban mapping project, version 1 (grumpv1): Urban extents grid," 2011. [Online]. Available: http://sedac.ciesin.columbia.edu/data/set/grump-v1-urban-extents

[15] D. Balk, U. Deichmann, G. Yetman, F. Pozzi, S. Hay, and A. Nelson, "Determining global population distribution: Methods, applications and data," in *Advances in Parasitology*. Elsevier, 2006, pp. 119–156. [Online]. Available: https://doi.org/10.1016/s0065-308x(05)62004-0

[16] J. Peiris, F. Amerasinghe, C. Arunagiri, L. Perera, S. Karunaratne, C. Ratnayake, T. Kulatilaka, and M. Abeysinghe, "Japanese encephalitis in sri lanka: comparison of vector and virus ecology in different agro-climatic areas," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 87, no. 5, pp. 541–548, sep 1993. [Online]. Available: https://doi.org/10.1016/0035-9203(93)90080-a

[17] J. A, R. HI, N. A, and G. E, "Hole-filled srtm for the globe version 4, available from the cgiar-csi srtm 90 m database (http://srtm.csi.cgiar.org)," 2008.