Enhancing infectious disease mapping with open access resources

D M Pigott[1], M U Kraemer (moritz.kraemer@zoo.ox.ac.uk)[1]
1.   Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, United Kingdom

The Spatial Ecology and Epidemiology Group (SEEG) at the University of Oxford is currently carrying out several disease and vector mapping projects. The source code for many of these projects have been and will continue to be made openly available through our GitHub account. This archive includes a description of the R software package 'seegSDM', that allows for species distribution modelling using a boosted regression tree approach and a walkthrough tutorial. This method has been successfully applied to diseases such as dengue, ebola virus, leishmaniasis, and anopheline vectors.

The group has collated a number of globally comprehensive and up-to-date databases from published and unpublished sources. The databases compile geographically located records of occurrence, from three sources: (i) comprehensive PubMed searches, (ii) information from unpublished health surveys and entomological field studies made available by collaborators, and (iii) internet disease surveillance systems such as HealthMap. These occurrences, defined as unique spatio-temporal records of pathogen transmission or vector presence, serve as standardised input data used to generate global risk maps. Prior to modelling these databases all go through a number of technical validation steps to ensure their spatio-temporal accuracy and overall validity. This is facilitated by overlaying these occurrence records with independently derived evidence consensus maps that summarise the quality and diversity of disease information.

Disease and vector specific databases are then made openly available through online depositories which cover the diseases mentioned above. Similarly, occurrence records of the dominant vector species of malaria and more comprehensive prevalence datasets of the global distribution of *Plasmodium falciparum, Plasmodium vivax* [1] and related inherited blood disorders are downloadable through the Malaria Atlas Project. Details about collection methods, design and validation of the respective datasets can be found either in published research articles or within data descriptor publications [2–4].

Future efforts include developing the Atlas of Baseline Risk Assessment for Infectious Diseases (ABRAID), an automated mapping platform which integrates the framework described above to generate spatially comprehensive, iteratively improving, evidence based maps of disease risk at the global level for a prioritised number of infectious diseases [5].

References

1.   Moyes CL, Temperley WH, Henry AJ, Burgert CR, Hay SI. Providing open access data online to advance malaria research and control. Malar J. Malaria Journal; 2013 Jan;12(1):161. http://dx.doi.org/10.1186/1475-2875-12-161

2.   Mylne A, Brady OJ, Huang Z, Pigott DM, Golding N, Kraemer MUG, et al. A comprehensive database of the geographic spread of past human Ebola outbreaks. Sci Data. 2014 Oct 23;1:140042. http://dx.doi.org/10.1038/sdata.2014.42

3.   Messina JP, Brady OJ, Pigott DM, Brownstein JS, Hoen AG, Hay SI. A global compendium of human dengue virus occurrence. Sci Data. 2014 May 27;1:1–6. http://dx.doi.org/10.1038/sdata.2014.4

4.   Pigott DM, Golding N, Messina JP, Battle KE, Duda K a, Balard Y, et al. Global database of leishmaniasis occurrence locations, 1960–2012. Sci Data. 2014 Sep 30;1:140036. http://dx.doi.org/10.1038/sdata.2014.36

5.   Hay SI, Battle KE, Pigott DM, Smith DL, Moyes CL, Bhatt S, et al. Global mapping of infectious disease. Philos Trans R Soc Lond B Biol Sci. 2013;368(20120250).