

GLOBE Observer Mosquito Habitat Mapper Data 2017-2020

Three Cases: Metadata and Procedures

The data sets described here were used in an article submitted to the journal GeoHealth in 2021. The data files and further supplemental links (including general information about GLOBE data) can be accessed at <https://observer.globe.gov/get-data/mosquito-habitat-data>.

Case 1: Removal of records with suspect geolocation data. A Python script was applied to remove records where the measured position (in decimal degrees) was identical to the GLOBE MGRS site position. GPS-obtained latitude and longitude coordinates are reported in decimal degrees, so records identified by whole numbers were also removed. This procedure removed 5704 (23%) of the 24983 records in the Mosquito Habitat Mapper database, with 19,279 records remaining. The secondary data sets cleaned only for geolocation anomalies were labeled Case 1.

Case 2: Identifying suspected training events. For this test, we sought to identify groups of data that exceeded 10 records sharing these characteristics. Another Python script was employed to extract the photos for ease of visual inspection. Because we needed to manually review the photo records, we set the threshold for groups at >10, so that the analysis could be completed in the time allotted. Groups identified through this procedure were outputted as case 2: groups. The resulting data set cleaned of groups >10 was labeled Case 2. The resulting data set included 20,006 records and identified 2,447 records found in clusters we postulated were training events.

Case 3: The Case 3 secondary dataset result from applying the Python scripts used to create Cases 1 and 2. We used the Case 3 data sets, with improved geolocation and large groups eliminated, in the following analysis.

The information in this document was last updated 2021-04-12