

Introductory information

- Data from observations of the role of water hyacinths in macroplastic transport and accumulation
- The dataset is in an Excel format and is comprised of three datasheet
- Contact: Louise Schreyers, Wageningen University and Research, louise.schreyers@wur.nl

Methodological information

- Method description:
 - (1) *Visual counting*. The data presented in the 'visual counting' datasheet was collected through a field measurement campaign in spring 2020 at Thu Thiem bridge (10°47'08.3"N, 106°43'06.2"E), Ho Chi Minh City, Vietnam. The data details the number of macroplastic items that were counted flowing downstream of the Saigon river, over a timeframe of 2 minutes. Items were distinguished between 'entangled' (entrapped within water hyacinths) and 'free-floating' (not entangled in the hyacinth patches, hence free-floating at the water surface). Three (3) observation points were selected on the bridge and are detailed in the column 'bridge location'.
 - (2) *Aerial survey – Manual labelling*. The data presented in the 'Aerial survey_manual labelling' datasheet was collected with a DJI Phantom 4 Pro Unmanned Aerial Vehicle (DJI, Shenzhen, China; <http://www.dji.com>) on 23 May 2020. 10 flights crossing the Saigon river were conducted, the four initial ones 100 m downstream of the Thu Thiem bridge, the 6 subsequent ones, 80 m upstream of the bridge. Each flight was scheduled automatically, with 22 waypoints along the river transect. Each waypoint is distant from each other by approximately 15 m. The waypoint and flight numbers are indicated in the datasheet. The drone imagery was analyzed through manual labelling in the open source Visual Geometry Group Image Annotator (VIA) tool (<https://www.robots.ox.ac.uk/~vgg/software/via/>). This enabled to detect and label macroplastic items that were identified on the selected images from the drone surveys.

Each row indicates one macroplastic item identified through manual labelling. Each item as a unique identifier. Rectangular shaped polygons were created around each identified items, and thus the width and length of each items is specified (in pixels). The manual labelling distinguished between 'entangled' (entrapped in floating water hyacinths patches) plastic items and 'free-floating' ones (not entangled, hence outside of water hyacinths patches).

In addition, the polymer category was also identified through manual labelling. The categorization to which the columns 'Polymer_category' refers to is the following: (1) EPS (expanded polystyrene), (2) PO_hard (hard polyolefins), (3) PO_soft (soft polyolefins), (4) PS (polystyrene), (5) PET (polyethylene terephthalate), (6) Multilayer, and (7) Rest. The items were categorized based on color, shape and other visual properties such as transparency.

The 'image_code' column presents the unique identifier created for each drone imagery in which the plastic item was detected. The 'GSD' column refers to the Ground Sampling Distance that was computed for each image, based on the flying elevation, the image width in pixels, sensor width and focal length of the camera. The GSD value ultimately allows to estimate the width and length of each plastic items in centimeters.
 - (3) *Aerial survey – color filtering*. The data presented in the 'Aerial_survey_color_filtering' details the results from a color segmentation (or filtering) process. The color segmentation was done using the open source Open CV library (<https://opencv.org/>), in Python. From the drone imagery collected close to Thu Thiem bridge on 23 May 2020, the aquatic vegetation (floating water hyacinths) areas were estimated by using color thresholds. 75 drone images were analyzed through color segmentation. The color segmentation allowed to estimate the area surface in pixels initially. With the GSD computed for each image, ultimately an estimated area in m² could be derived.

Data specific information

- (1) *Visual counting*. **Date**: date at which the visual counting of plastic items took place. The date format is Month/Day/Year. **Timestamp**: exact time at which the visual counting took place. **Bridge location**: bridge observation points at which the visual counting was done. **Duration**: specifies the duration (in minutes) for which the plastic items were counted. **Free-floating**: number of free-floating items counted. **Entangled**: number of entangled items counted.
- (2) *Aerial survey – Manual labelling*. **Item_ID**: plastic items unique identifier. **Height (pixels)**: height of the plastic items expressed in number of pixels. **Width (pixels)**: width of the plastic items identified expressed in pixels. **Polymer_type**: polymer category identified for each plastic items. Refer to the Methodological information section for the polymer category used.

Water_hyacinths: indicates whether the plastic item is entangled in water hyacinths or not. **Flight:** number of the drone flight, from 1 to 10. **Image_code:** unique identifier for each drone image on which the plastic item was detected. **Waypoint_number:** indicates at which waypoint was the drone image taken, from 1 to 22. **Timestamp:** starting time for the drone flight at which the drone imagery was captured. **GSD:** Ground Sampling Distance for each drone image, expressed in cm/pixel.

- (3) *Aerial survey – Color filtering.* **Image_code:** unique identifier for each drone image analyzed through color filtering. **GSD:** Ground Sampling Distance for each drone image, expressed in cm/pixel. **Vegetation_area (pixels):** estimated area of floating vegetation (water hyacinths) in pixels.

Sharing and access information

No restrictions are placed on the use of this data. License: CC0 (Creative Commons Zero)