InvasivesBC
Combined Observation/Survey Report
"How-To Guide"

Background

InvasivesBC houses two separate data sets – one from the former provincial Invasive Alien Plant Program (IAPP) database (2005-June2023), and one with all InvasivesBC records (June 2023-present). Now that one season of InvasivesBC data has been entered (2023), this new Combined Observation/Survey Report has been launched to support InvasivesBC users who want to be able to identify where there is overlap of InvasivesBC records and previous IAPP sites (in excel without having to check individually on the map). This new report pulls records from BOTH datasets and will allow users to develop treatment lists and summarize all records in a location more easily when working with large numbers of records. This is the first in a series of "combined IAPP/InvasivesBC" reports that will be developed.

Source data

The Combined Observation/Survey Report contains everything from the IAPP site selection extract and the <u>most recent</u> terrestrial and aquatic observations from InvasivesBC.

<u>Most recent</u> observations are defined as the most recent record by date and time if multiple observations of the same plant overlap. Recency is compared between activity subtypes (ie terrestrial and aquatic records). For example, if there is a more recent *terrestrial* observation of common tansy that overlaps an older *aquatic* observation of common tansy, only the terrestrial observation will be included in the combined report. This can be thought of as the InvasivesBC version of the IAPP site selection extract (most recent survey of each plant on a site).

Any amount of overlap results in the removal of older observations. This means there is potential for large polygons covering very different areas to be excluded from the report due to a small amount of overlap. It is the observation or treatment persons responsibility to do an adequate survey of the area in the field.

Included columns

As this Report is a combination of IAPP and InvasivesBC data, the columns are mostly limited to those that are common to both data sets. Some columns are not 100% comparable between the two data sets but are still included and have been massaged/edited into a more useable format:

- ID column will be a Site ID for IAPP records and a Short ID/Observation ID for InvasivesBC records.
- Project code column is the Survey Paper file ID for IAPP records and the Project code for InvasivesBC records.
- Funding Agency column contains the IAPP Survey Agency for IAPP records and the Funding Agency for InvasivesBC records.
- Record Type column can be either IAPP Survey, Terrestrial Observation, or Aquatic Observation.
- Observation Type column will be either IAPP Survey, Positive Observation, or Negative Observation.
- Area sqm column is area in square meters for both datasets (IAPP records have been converted from hectares to square meters).
- Recent Treatments is a combination of Treatment Date and Treatment Type columns from the site selection extract for IAPP records and the Short ID, Treatment Type, and Date of any treatments of the same plant that overlap the observation and occurred on or after the observation date (this means IAPP records will only have the most recent treatment and InvasivesBC records may have a list of multiple treatments separated by commas).
- Invasive Plant Column: Where IAPP plant names are different than equivalent InvasivesBC plant names, the InvasivesBC plant name is used for IAPP records. This was done to eliminate duplicate plant names to make filtering easier and to make spatial comparisons more representative between the two datasets. There are some plant names in IAPP that do not have an equivalent plant name in InvasivesBC (i.e. Knapweed species), in these cases the IAPP plant name is used. The fact that there is no InvasivesBC equivalent plant name means that there will never be a nearest observation ID and any cluster will only contain IAPP records. Plant names are used for all spatial distance calculations and clustering of records.

Nearest ID and Distance to Nearest column

The Nearest ID column contains the ID of the nearest record from the sister dataset that has the same plant and is within 100 meters. The Distance to Nearest column is the distance in meters to the nearest record of the same plant from the sister database. This is only calculated between the two datasets and not within them. Distance is calculated from the outside edge of polygons.

For IAPP records the nearest ID will always be a Short ID/Observation ID (InvasivesBC record) and for InvasivesBC records the nearest ID will always be an IAPP Site ID. If the columns are blank, there is no record of the same plant within 100 meters in the sister dataset. If the Distance to Nearest column is 0, the nearest record overlaps.

These two columns are meant to make it easier to make site lists and not end up with duplicate/overlapping survey/observations between the two datasets.

Removing Duplicate/Overlapping Records

To remove duplicate IAPP surveys from this Report in excel:

1. Sort by Record Type and Nearest ID
2. Filter for Record Type "IAPP Survey"
3. Remove all the IAPP rows that have a Nearest ID value (leave all rows that are blank).
4. This will result in only IAPP surveys that are more than 100 meters away from any InvasivesBC observation of the same plant left in the excel file.
   a. If desired, users can review the IAPP rows that have a Nearest ID value before deleting, and consider viewing any records on a map for those that you may question whether or not they are worth keeping both the IAPP and the InvasivesBC record in the list (eg. if they are close to 100m away from each other or if you have local knowledge about the records).
5. Any IAPP surveys you have removed from your excel file that had a Nearest ID value will have a corresponding InvasivesBC observation with a Site ID in the Nearest ID column for future reference.

Clusters Columns (Cluster, IDs in Cluster, Total Cluster Area)

Another feature of the Combined Report is the Cluster Column which has been developed to support users who want to summarize records within a particular area (similar to the area formerly referred to as an IAPP site). The Cluster Column contains a plant name and number (i.e. Annual hawksbeard (Crepis tectorum)-24). If records from either/both datasets are within 100 meters and have the same plant name, they are assigned the same cluster name. A single record can be its own cluster and there is no limit to how many records can be in a single cluster. Clusters can be both between and within datasets, only the plant name and the distance are used.

The number in the cluster name is randomly assigned daily. If two users extract the combined report on the same day the clusters will be the same, but if the report is extracted on different days there is no guarantee that the cluster names will be the same. Because of this, when referencing a cluster to someone else it may be better to say "the cluster containing site 312556" instead of "cluster Knapweed-123".

Clusters will be updated as observation records are added and deleted. For example, if there are two clusters of spotted knapweed 150 meters apart and a new spotted knapweed observation is entered directly between them, they will become a single cluster the next time the report is pulled. Conversely if the observation between them is deleted they will then be two separate clusters. Negative observations and IAPP surveys with 0 area are also included in all clustering.

The **IDs** in Cluster Column is a list of all IDs in the cluster separated by commas. These IDs can be both IAPP Site IDs and/or InvasivesBC Short IDs/Observation IDs.

The **Total Cluster Area** column is the combined area in square meters of all records in the cluster. The area of negative observations is included in this number so it must be remembered that in InvasivesBC the area of a negative observation polygon is the area searched and the target invasive plant was not found, not the area covered by an infestation as it was in IAPP. This column is only intended to give a rough idea of the composition or overall size of the cluster. For example, if there are 15 record IDs in the cluster but the total cluster area is 25 square meters, the cluster is composed of many small patches. If there are 2 record IDs in the cluster and the cluster area is 9000 square meters, the cluster is composed of two very large polygons. The Total Cluster Area does not incorporate the same spatial math as the InvasivesBC Spatial Reports (ie. subtracting the area of more recent negative observations which overlap positive observations, removing double counting, etc.) and therefore does not generate as accurate an area total. It is only intended to indicate the composition of a cluster and should not be used to report on the current area of that species. Users can remove negative occurrences from combined report and use excel to calculate the total area (sum the area_sqm column) in a cluster to get the correct area if that is desired. The area number in the total_cluster_area column will still include the negative area though as it was calculated when the report is generated.

For questions or more information please contact InvasivesBC@gov.bc.ca or visit the InvasivesBC resource webpage.