

## SOP 4: Corrected Observations (KM\_D Creation)

### *Version 1.0*

#### Revision History Log:

| Previous Version # | Revision Date | Author                                                    | Changes Made    | Reason for Change | New Version # |
|--------------------|---------------|-----------------------------------------------------------|-----------------|-------------------|---------------|
|                    | August 2013   | S. Hoekman,<br>W. Johnson,<br>B. Moynahan,<br>C. Sergeant | Initial version |                   | 1.0           |

#### Summary

This deliverable is based on the raw observations collected by field laptops. A copy is made of the KM\_C file geodatabase. To that basis are added individual corrections that were noted on the field data forms (KM\_B) but could not be entered in the field. Some corrections may also require data from waypoints, either directly off the field GPS unit designated for recording waypoints, or from the waypoints file (KM\_I) if it is available.

An ASCII CSV file is generated from the corrected geodatabase by an application written for the purpose. Additional attributes are added for data quality codes and data quality comments. Both the final geodatabase and CSV are submitted through the validation / quality assurance iterations until certified. Be aware the validation criteria here are much stricter than those for KM\_C and several iterations may be required to attain validation. The certified deliverable is stored in the repository and disseminated over the web.

It is recommended KM\_D be generated post-season after the KM\_C deliverable has been field corrected and certified. If the KM\_C product is ever determined to be flawed, then KM\_D will have to be regenerated from the corrected geodatabase.

SOP 4 requires users to work in the internal NPS network under valid Active Directory accounts. Network access questions should be directed to the Data Manager.

#### Detailed Steps

##### A. Project Leader (or Delegate) Tasks

1. Edit corrections into the raw observation geodatabase
  - a. Verify these precursors are available:
    - i. Either the certified KM\_C geodatabase or the final C:\KIMU\Murrelets.gdb folder from the field laptop.
    - ii. Either the certified KM\_I waypoints file or the original field GPS.
    - iii. Either the certified KM\_B field data observer form images or the complete set of physical forms.
  - b. Download the completed KM\_C deliverable MURRELETS.GDB directory from the production server if available to

[\\INPGLBAFS03\Data\SEAN\\_Data\Work\\_Zone\KM\KM\\_D\yyyy\](#) and then unzip it. Alternatively, obtain C:\KIMU\Murrelets.gdb folder from the field laptop if a certified KM\_C is not yet available.

- c. Copy the latest MURRELETS.MXD file available into  
[\\INPGLBAFS03\Data\SEAN\\_Data\Work\\_Zone\KM\KM\\_D\yyyy\](#)
- d. Apply corrections to the geodatabase.
  - i. Start the ESRI ArcMAP application on a workstation.
  - ii. Open the MURRELETS.MXD map file
  - iii. Be sure the following features are properly resolved in the map (if not, the source in the layer properties must be adjusted to point to the geodatabase object):
    1. Bird groups
    2. GPS points
    3. Observations
    4. Tracks
    5. Transects
  - iv. Be sure there is adequate background imagery available to clearly visualize the features at high magnification.
    1. If the workstation is connected to the NPS network, it will often find the imagery on the X: drive.
    2. If X: drive imagery is not available, copy  
C:\KIMU\UserData\GLBA\_Cache \ into  
[\\INPGLBAFS03\Data\SEAN\\_Data\Work\\_Zone\KM\KM\\_D\yyyy\](#)  
and see that the source is linked to the map feature named  
“Background Imagery”.
    3. It is also possible to employ different imagery found on, say, the ESRI World Imagery Service.
  - v. Edit the file geodatabase in ArcMAP
    1. Apply all corrections from the KM\_B field logs that were not already taken care of in the field.
    2. Verify tracklogs are attributed to the correct transect
      - i. Be sure the Transects and Tracks features are visible over the imagery.
      - ii. Open the Tracks attribute table and sort it by Transect ID
      - iii. Select one at a time each line of the attribute table
      - iv. Verify each selected row appears highlighted on the map as part of the correct transect. That is, the Transect ID in the table must match the Transect number on the map.
      - v. For any row that was physically recorded on a different transect, edit the proper Transect ID into the table entry. The visual location comes from the GPS (bulletproof) while the transect table entry was typed by hand in the field (fragile).

3. Extend tracklog segments to cover gaps that resulted from application or GPS interruptions.
  - i. For each transect used that year, magnify it so missing tracklog segments may be seen.
  - ii. Edit the Tracks feature where necessary to extend one of the tracklogs so it snaps with the start of the next tracklog along the transect, filling the gap.
  - iii. Never extend an “off transect” tracklog over an observation.
4. Audit the “On Transect” attribute of Tracks.
  - i. Visually determine locations where the boat is going along the transect correctly, but the tracklog has On Transect set to false. Be aware there are legitimate reasons the boat may be physically on the transect but has temporarily suspended operation, as when waiting for an obstruction to clear or a technical problem to be resolved: check KM\_B for explanations. A good way to see them is to open the Tracks table, sort by “On Transect” and zoom to each “False” track.
  - ii. Visually determine where the boat is not physically near the transect, but On Transect is set to true.
  - iii. Correct errors by editing individual On Transect values in the Tracks attribute table.
  - iv. Never assign a tracklog “off transect” when an observation is on it.
  - v. In the case where a single tracklog is set to one value but spans both on- and off-transect states, the editing software must be used to split the tracklog into two before assigning the correct state to the flawed section.
5. Correct any observations determined to be made off transect.
  - i. Sort the Tracks table by “On Transect.”
  - ii. Select the block having False for on transect.
  - iii. Use the “related tables” function to drill down to GPS Points, and then to Observations. The Observations should be empty.
  - iv. If observations are found related to off transect, make appropriate corrections.
6. Force consistent representation of Protocol ID, Vessel Name, Data Recorder, and Observers in the Tracks table. Be sure they all reflect upper case.
7. Edit the Bird Groups feature so that any remaining Species of “Pending” is set to “Unidentified”. (Locate by visually checking the Bird Groups layer for any red icons.)
8. Delete any test transect data. (The specific method, below, assures referential integrity will be maintained, so it is highly preferred.)

- i. Open the attribute table for TRANSECTS and select transect "TST".
    - ii. Cycle through the relations until the last object, BIRD GROUPS, shows on the screen. The table reflects the number of bird groups selected, which should be on the order of a few percent of all groups.
    - iii. Delete the selected.
    - iv. Back up to OBSERVATIONS. Verify a reasonable number are selected and delete them.
    - v. Back up to GPS POINTS. Verify a reasonable number are selected and delete them.
    - vi. Back up to TRACKS. Verify a reasonable number are selected and delete them.
  9. Be sure all edits have been saved.
  10. Verify the integrity of the FGDB
    - i. Open the Transect table and select all rows
    - ii. Next drill the "Related Table" Track and be sure all rows have been selected.
    - iii. Next drill the "Related Table" GPSPoint and be sure all rows have been selected.
    - iv. Next drill the "Related Table" Observation and be sure all rows have been selected.
    - v. Next drill the "Related Table" BirdGroups and be sure all rows have been selected.
    - vi. If any table shows more rows present than are selected, then those rows are orphans in the database structure and must be either attached to parents or deleted entirely.
  11. Close ArcMap.
2. Create the CSV file
  - a. Locate a machine that has the NPTransect application installed on it. NPTransect must always be installed on the root level of the C: drive.
  - b. Replace the application's geodatabase content at C:\KIMU\Murrelets.gdb\ with the corrected version from [\\INPGLBAFS03\Data\SEAN\\_Data\Work\\_Zone\KM\\_KM\\_D\yyyy\Murrelets.gdb\](#).
  - c. From a DOS command box on the application machine, execute the script C:\KIMU\Scripts\4)ExportToCSV.bat.
  - d. Verify a file was created in C:\KIMU\CSV\ with the current date and time.
  - e. Open the file in Excel and verify it has content of the expected form.
  - f. Copy the file from C:\KIMU\CSV\ to [\\INPGLBAFS03\Data\SEAN\\_Data\Work\\_Zone\KM\\_KM\\_D\yyyy\](#).
  - g. Rename the file to KM\_D\_yyyy.CSV.
  - h. Open the CSV file on \\INPGLBAFS03\ in Excel.

- i. Set the TRANSECT\_ID column's cell format to "custom 000" in order to preserve any leading zeroes in the transect name when it is resaved.
    - j. Correct the heading typo on the first row: DATA\_QUALITY\_CODE should be changed to DATA\_QUALITY\_COMMENT.
    - k. Fill the DATA\_QUALITY column with zero for every row (presume they're good).
    - l. Save the sheet as type CSV (comma delimited).
      - i. If Excel prompts whether you really want to save it as CSV, click Yes.
      - ii. Close Excel.
  3. Insert quality information into the CSV file
    - a. Reopen the CSV file on \\INPGLBAFS03\... in Excel.
    - b. Bring up the associated field data observer form, either as the images of KM\_B on the web site or as original paper forms.
    - c. Record any data quality issues indicated on the observer forms to the data file.
      - i. For any rows that are of questionable accuracy fill Data\_quality with the digit 4, and add Data\_quality\_comment text explaining why it is questionable. Questionable data include:
        1. Observations noted as exceptional on the field observer form.
        2. Observations reconstructed from paper backup used because the automated recording system failed.
        3. Distances that are not multiples of 10, which suggest typos.
        4. Angles that are not multiples of 5 suggesting typos – except 171 through 189, which should be resolved to the single digit.
      - ii. For rows that are corrupted or have lost required data, as may occur through hardware failure, mark Data\_quality with the digit 8, and add Data\_quality\_comment text explaining why it is unusable. Questionable data include:
        1. Observations noted as invalid on the field observer form.
        2. Angles of 0 or 360, which are not visible to observers.
        3. Note: Data\_quality 8 rows are not stored in the repository database (KM\_E).
      - iii. For rows that have no corresponding field log form, mark Data\_quality with the digit 1, and add Data\_quality\_comment text explaining why the quality is unknown.
      - iv. In the event of equipment failure that required paper recording of observations along with waypoint recording on the GPS:
        1. If no file with any content was captured at all, initiate a blank CSV file in Excel with the heading row copied from a prior year file.
        2. Regardless, insert rows into the spreadsheet reflecting as much data as was recorded on paper in chronological sequence.
        3. Use the waypoint identifiers on the form to look up the location and time on the GPS unit or KM\_I waypoint file – converting time and coordinates to match units specified in Appendix I if needed.

4. Mark the Data\_quality as 4.
5. Mark Data\_quality\_comment “automated data collection failed”.
- d. Resave the corrected file as CSV and close Excel.
4. Create KM\_D by zipping together in place both the corrected GBD folder and the edited CSV file. Name the ZIP file “KM\_D\_yyyy.ZIP”.

#### **B. Project Leader Tasks**

1. Request validation by sending a notification email to the Data Manager, specifying in the message body that deliverable KM\_D, for year yyyy is ready; the Data Manager will know where to find the file named “KM\_D\_yyyy.ZIP”.

#### **C. Data Manager Tasks**

1. On receipt of the submission, assign the next formal Submission Number to this file, as found in the master Submission\_Log table.
  - a. Use the “Update Submission Log” web tool at [http://165.83.57.239/0\\_submission\\_update.aspx](http://165.83.57.239/0_submission_update.aspx).
  - b. Complete Submission\_Log details up through the Submission\_Date column.
2. Copy the file into the staging area for validation at a subdirectory named with the submission number in [\\inpglbafs03\data\SEAN\\_Data\Staging\KM\KM\\_D](\\inpglbafs03\data\SEAN_Data\Staging\KM\KM_D).
3. Validate the submission according to current criteria.
  - a. Invoke validation of the CSV file according to current criteria by using the web program at [http://165.83.57.239/KM\\_DM\\_validate\\_KMD.aspx](http://165.83.57.239/KM_DM_validate_KMD.aspx)
  - b. Validation summary results automatically are recorded in the Submission\_Log.
4. If submission fails mandatory criteria, reply with a “failure email” that includes:
  - a. The submission number assigned
  - b. The deliverable ID
  - c. The protocol ID
  - d. Documentation listing all the criteria failed
5. If submission passes mandatory criteria, reply with a “success email” that includes:
  - a. The submission number assigned
  - b. The deliverable ID
  - c. The protocol ID
  - d. Documentation listing any specific optional criteria failed
  - e. Request to certify deliverable as complete

#### **D. Project Leader Tasks**

1. On receipt of a failure email:
  - a. Review the validation report and remedy the mandatory errors.
    - i. In case it is impossible to correct a mandatory error on a particular row of the CSV file:
      1. Mark its Data\_quality with the digit 8.
      2. Record a Data\_quality\_comment explaining why it cannot be corrected.
      3. Be aware the cumulative database will not accept this row.
  - b. Initiate a new submission with the Data Manager.

2. On receipt of a success email, review any failed optional criteria:
  - a. If the criteria review indicates the validated data are acceptable:
    - i. Reply with a "certification email" stating the deliverable is certified and may be disseminated.
  - b. If the validated data are unacceptable:
    - i. Reply with a "withdrawal email", stating the deliverable is withdrawn.
    - ii. Take remedial action to obtain and resubmit a corrected deliverable, if possible.

#### **E. Data Manager Tasks**

1. On receipt of a withdrawal email:
  - a. Mark the withdrawal in the Submission\_Log's Status column using the web tool.
  - b. Terminate the process.
2. On receipt of a certification email:
  - a. Verify no sensitive information is in the deliverable. Products containing sensitive information cannot be disseminated. If sensitive:
    - i. Mark the sensitive designation in the log's submission status column using the web tool.
    - ii. Notify the submitter of the situation.
    - iii. Complete the deliverable tracking spreadsheet cells referring to dissemination with the word SENSITIVE.
    - iv. Do not attempt to install the sensitive item in the auxiliary repository.
  - b. Copy the submitted file to the development environment at AUXREP\KM\KM\_D\.
  - c. Create an appropriate web link accessible from the KM main page.
  - d. Test accessibility until successful, then propagate from development to integration and then to the production environment.
  - e. Verify deliverable is accessible from production web site.
  - f. Mark the certification in the Status column in Submission\_Log using the web tool. Mark any prior certified submission for this unit as decertified.
  - g. Update the deliverable tracking spreadsheet milestones with the date of completion for KM\_D.
3. Update the scope of the formal metadata so it includes this new date range.

