Pipeline for Agile Estimation of Land Accounts (PAELA)

Component 1

The purpose of Component 1 is to prepare a dataset of Regions of Interest (ROIs) for the specified year. The inputs for this component include:

- KML files[1] containing ROI polygons,
- A reference stacked image using Sentinel-2 data (generated in Step 1.1),
- A land cover classification table.

The final output of this component is a merged and refined ROIs shapefile, which includes all ROIs for different land cover classes and updated for the specified year. This dataset is then used in Component 2 to train the land cover classification model.

In addition, spectral signature plots are generated to visualize and assess the reflectance behaviour of the different land cover classes. These include plots showing the mean reflectance values with plus or minus one standard deviation, and comparisons across groups of spectrally similar land cover types, providing information on their spectral separability.

Step 1.1: Run the script to create the stack

Purpose: This step creates a stacked image for a specified year using Sentinel-2 L2A data accessed from Google Earth Engine using Google Colab. The stack will be used in Step 1.4 to add ROIs showing areas of cloud that may remain in the imagery despite applying a cloud mask.

Preparation:

1) Download the Colab notebook:

- Locate the Colab notebook and the instructions text file 'Colab_COMP1_notebook_and_instructions_SENT.txt' in the directory '\Z_Visit_Vanuatu_April2025\Component_1\code\'.
- Follow the instructions in the text file to download and upload the notebook to your own Google Drive and then open it in Google Colab.

Process:

1) Start running the notebook in Google Colab, one cell at a time and following all instructions given in the notebook. The initial cells of code install the necessary Python packages, initialise Google Earth Engine, and mount your Google Drive.
2) Generate the stack:

- Specify the target year for the stack in the relevant cell of the notebook.
- The code will connect to the Sentinel L2A collection, select the required bands for the stack, and calculate the median composite for the specified year after applying a cloud mask.

3) Save the output:

- The code saves the stack to your Google Drive under folder 'GEE_exports_COMPONENT_1'.

[1] KML (Keyhole Markup Language) files are used by Google Earth and other applications to store and display geographic data

- Manually copy this file to your computer in folder '\Z_Visit_Vanuatu_April2025\Component_1\output\theYear'.

- After saving the output, the remaining sections of code in the notebook are used for Step 1.5, below.

Output:

- The output is a GeoTIFF file of the stacked image for the specified year.

Step 1.2: Refine ROIs in Google Earth Pro

Purpose: This step ensures that the Regions of Interest (ROIs) are updated and refined based on visual interpretation in Google Earth Pro. The updated ROIs can then be used in subsequent steps for model training. The process involves first creating a new version of the existing ROIs, renaming the files for consistency, and then editing the polygons as needed.

Preparation:

1) Create a new folder in which to place the updated ROIs:

- Copy all KML files from the folder '\Z_Visit_Vanuatu_April2025\Component_1\output**ZZZZ**\vROIs**X**\' to the new folder '\Z_Visit_Vanuatu_April2025\Component_1\output**ZZZZ**\vROIs**Y**\'.
- Copy all components of the shapefile 'Cloud_**ZZZZ**_vROIs**X**' (.prj, .shp, .dbf, .shx) into the same new folder '\Z_Visit_Vanuatu_April2025\Component_1\output**ZZZZ**\vROIs**Y**\'.

**Note:*

- **X**: The version of the ROIs being updated.
- **Y**: The new version of the ROIs.
- **ZZZZ**: The year of the ROIs.

For example, if you want to create a new version **3** (Y) of the existing version **2** in 2020 (X) of ROIs for the year **2020** (ZZZZ), then Step 1.2 involves:

- Copying all KML files from the folder '\Z_Visit_Vanuatu_April2025\Component_1\output**2020**\vROIs**2**\' to the new folder '\Z_Visit_Vanuatu_April2025\Component_1\output**2020**\vROIs**3**\'.
- Copying all components of the shapefile 'Cloud_**2020**_vROIs**2**' (.prj, .shp, .dbf, .shx) into the same new folder: '\Z_Visit_Vanuatu_April2025\Component_1\output**2020**\vROIs**3**\'.

As a further example, if you want to create a version **1** (Y) of the ROIs for 2023, using the existing version **2** (X) ROIs for the year 2020, then Step 1.2 involves:

- Copying all KML files from the folder '\Z_Visit_Vanuatu_April2025\Component_1\output**2020**\vROIs**2**\' to the new folder '\Z_Visit_Vanuatu_April2025\Component_1\output**2023**\vROIs**1**\'.

- Copying all components of the shapefile 'Cloud_**2020**_vROIs**2**' (.prj, .shp, .dbf, .shx) into the same new folder: '\Z_Visit_Vanuatu_April2025\Component_1\output**2023**\vROIs**1**\'.

2) Prepare the script for automating the renaming of the files:

- Locate the PyQGIS script 'Renaming_kml_shp.py' in the folder '\Z_Visit_Vanuatu_April2025\Component_1\Scripts\'.

Process:

1) Set parameters and run the PyQGIS script:

- Open QGIS and load the script by navigating to Plugins → Python Console → Show Editor → Open Script…

- Open the script and update the 'input_folder' variable to: "…/Z_Visit_Vanuatu_April2025/Component_1/output/**ZZZZ**/vROIs**Y**", using the file path for the local drive where this folder is saved on your computer.

  Note: When entering file paths, always use ***forward slashes*** *(/) instead of backslashes (\). This will allow compatibility across operating systems (Windows and Linux) and avoids errors in Python.*

  For example, if you are using your D: drive, working with the 2020 ROIs and creating version 3, set:

  input_folder = "D:/Z_Visit_Vanuatu_April2025/Component_1/output/**2020**/vROIs**3**"

  Similarly, if you are working with the 2023 ROIs and creating version 1, set:

  input_folder = "D:/Z_Visit_Vanuatu_April2025/Component_1/output/**2023**/vROIs**1**"

- Run the script to:
  - Rename all KML and shapefile components in vROIsY to include the year (ZZZZ) and folder name (vROIsY).
  - Update the <name> tags within the KML files for consistent naming.

2) Open the renamed KML files in Google Earth Pro:

- Open Goolge Earth Pro and then drag all KML files from the vROIs**Y** folder into the Google Earth Pro window.

  For example, if you are refining the 2020 ROIs version 3, drag the KML files from the vROIs3 folder into Google Earth Pro.

  Similarly, if you are refining the 2023 ROIs version 1, drag the KML files from the vROIs1 folder into Google Earth Pro.

3) View imagery for the relevant reference period:

   Click on the 'View Historical Imagery' icon on the toolbar. Use the slider that appears to select imagery for a date during the relevant reference period (e.g., 2020).

4) Edit the polygons:

- Delete or add new polygons as needed for each KML file in order to refine the ROIs.
    - To delete: Right-click the polygon and then select 'Delete'.
    - To add: Use the "Add Polygon" tool (toolbar icon with a shape), draw the polygon, and then click on OK.

5) Save the updated KML files:

- For each modified file:
    - Right-click the file in Google Earth Pro and select 'Save Place As....'
    - Save the file to the 'vROIs**Y**' folder, overwriting the existing file when prompted.

    For example, if you are refining the 2020 ROIs version 3, save the file to the 'vROIs3' folder.

    Similarly, if you are refining the 2023 ROIs version 1, save the file to the 'vROIs1' folder.

Output:

The folder 'vROIsY' will contain:

- Updated and renamed KML files reflecting refined ROIs.
- The shapefile Cloud_ZZZZ_vROIsY (.prj, .shp, .dbf, .shx) with consistent names.

Step 1.3: Convert kml files to shapefiles and merge them in QGIS

Purpose: This step converts the KML files containing the refined ROIs (produced in Step 1.2) into shapefiles and then merges them into a single shapefile. This shapefile will subsequently be used in Step 1.5 to generate spectral signature plots to visualize and assess the reflectance behaviour of the different land cover ROIs. In addition, the output shapefile will be used for model training in Component 2.

Preparation:

1) Verify Required Files:

- Ensure the following files are in place:
    - Updated KML files in the folder '\Z_Visit_Vanuatu_April2025\Component_1\output\ZZZZ\vROIsY\' from Step 1.2.
    - Stacked image for the specified year (from Step 1.1) for Coordinate Reference System (CRS) alignment in the folder 'Z_Visit_Vanuatu_April2025\Component_1\output\ZZZZ'.
    - CSV file with land cover classification details, located at '\Z_Visit_Vanuatu_April2025\Component_1\LC_classes_table.csv'.

2) Prepare the script:

- Locate the PyQGIS script 'Component_1B_20250331_ROIs_kml_to_shp.py' in the folder '\Z_Visit_Vanuatu_April2025\Component_1\Scripts\'.

Process:

1) Set parameters in the PyQGIS script:

- Open QGIS and load the script by navigating to Plugins → Python Console → Show Editor → Open Script…
- Update the following variables in the script:
Note: When entering file paths, always use ***forward slashes*** *(/) instead of backslashes (\). This will allow compatibility across operating systems (Windows and Linux) and avoids errors in Python.*
    - 'inputOutput_path': Path to the folder containing renamed KML files: "…/Z_Visit_Vanuatu_April2025/Component_1/output/ZZZZ/vROIsY", using the file path for the local drive where this folder is saved on your computer.
    For example, if you are working with the 2020 version 3 ROIs, and you are using your D: drive, set:
    inputOutput_path =
    "D:/Z_Visit_Vanuatu_April2025/Component_1/output/2020/vROIs3"
    - ref_layer_path: Path to the reference raster file:
    "…/Z_Visit_Vanuatu_April2025/Component_1/output/ZZZZ/stacked_image23481112_ZZZZ_20m.tif".
    For example, for year 2020 and if you are using your D: drive, set:

      ref_layer_path =
      "D:/Z_Visit_Vanuatu_April2025/Component_1/output/2020/stacked_image23481112_2020_20m.tif"

      Similarly, for year 2023 and if you are using your D: drive, set:

      ref_layer_path =
      "D:/Z_Visit_Vanuatu_April2025/Component_1/output/2023/stacked_image23481112_2023_20m.tif"
    - LCtable_path: Path to the CSV file:
    "D:/Z_Visit_Vanuatu_April2025/Component_1/LC_classes_table.csv". This path only needs to be updated if you are using a drive other than your D: drive.

2) Convert KML to Shapefiles:

- Run the script.

It will process each KML file in the specified folder by:

- Checking if the base name of each KML file matches an entry in the CSV file.
- Adding fields defined in the CSV (e.g., MC_ID, MC_name, C_ID) to the shapefiles.
- Reprojecting geometries to match the CRS of the reference raster file.
- Saving each processed KML file as a shapefile in the same folder.

It will also merge the separate shapefiles for each land cover class and clouds into a single shapefile:

- After conversion, the script merges all shapefiles in the folder into a single shapefile named 'ROIs_ZZZZ_vROIsY'.
- The merged shapefile is saved in the same folder '\Z_Visit_Vanuatu_April2025\Component_1\output\ZZZZ\vROIsY\'.

For example, if you are working with the 2020 ROIs and creating version 3, the merged shapefile will be saved as: '\Z_Visit_Vanuatu_April2025\Component_1\output\2020\vROIs3\ROIs_2020_vROIs3.shp'

Similarly, if you are working with the 2023 ROIs and creating version 1, the merged shapefile will be saved as: '\Z_Visit_Vanuatu_April2025\Component_1\output\2023\vROIs1\ROIs_2023_vROIs1.shp'

Output:

The folder '\Z_Visit_Vanuatu_April2025\Component_1\output\ZZZZ\vROIsY\' will now also contain:

- The merged shapefile 'ROIs_ZZZZ_vROIsY'.

This shapefile contains all the refined and processed ROIs, with land cover classification attributes added from the CSV file.

Step 1.4: Edit Cloud ROIs in QGIS, if needed

Purpose: This step ensures that the cloud ROIs in the merged shapefile ROIs_ZZZZ_vROIsY correctly represent the cloud conditions for the target year (ZZZZ). This step is only necessary if in Step 1.2 you created ROIs for a given year using data from a different year as the base (e.g., creating the 2023 ROIs based on the 2020 ROIs). In such cases, the cloud ROIs from the base year (e.g., 2020) will not match the actual cloud conditions of the target year (e.g., 2023), and should therefore be removed and recreated.

You do **not** need to carry out Step 1.4 when the ROIs for a particular year (e.g., 2020) are being refined or updated using data from the same year (e.g., creating version 3 for 2020). In those cases, the existing cloud ROIs already reflect the correct year.

Preparation:

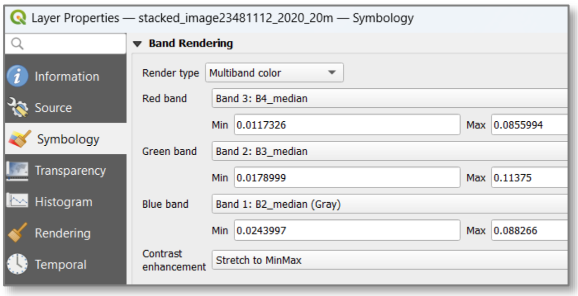
1) Add the required Layers to QGIS:

- Load the merged shapefile 'ROIs_ZZZZ_vROIsY' into QGIS.
- Load the stacked image for the same year (from Step 1.1).

For example, load the merged shapefile 'ROIs_2023_vROIs1.shp' and the stacked image 'stacked_image23481112_2023_20m.tif'.

2) Set the Symbology for the stack:

- For example, apply a true-colour composite symbology (Sentinel bands 4-3-2) to the stacked image to visually identify remaining cloud pixels.
    - Right click on the stacked image, then select Properties → Symbology
    - Select 'Multiband color' as the render type.
    - Specify the Red Band as 'B4_median', Green band as 'B3_median' and Blue band as 'B2_median' respectively.



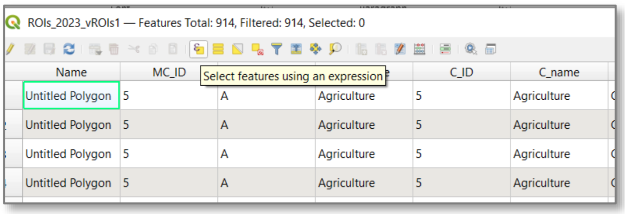
Process:

1) Start an Edit Session:

- Right-click the ROIs_ZZZZ_vROIsY layer in the Layers Panel and select ‘Toggle Editing’.

2) Select and Remove Existing Cloud Polygons:

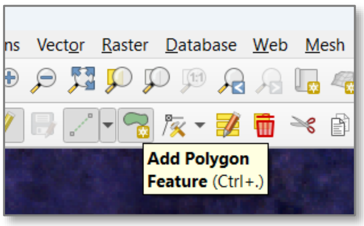
- Right-click the ROIs_ZZZZ_vROIsY layer and select ‘Open Attribute Table’.
- Click on the button ‘Select features using an expression’ then use the window on the left to select all polygons where MC_name = 'Cloud'. Click on ‘Select Features’, then on ‘Close’.



- Delete the selected polygons by clicking on the button ‘Delete selected features’.

3) Create new Cloud ROIs:

- Based on visual interpretation of the true-colour composite, create polygons for new cloud ROIs, using the ‘Add Polygon Feature’ button. Left click to create the vertices of the polygon, creating the final vertex by right clicking.



- For each new polygon, assign the following attributes:
  - MC_ID = 12
  - MC_name = Cloud

4) Stop the Edit Session:

- Save the changes by toggling off editing and confirming to save edits.

Output:

An updated shapefile, ‘ROIs_ZZZZ_vROIsY’, with cloud ROIs that reflect the actual conditions of the specific year (ZZZZ). For example, ‘ROIs_2023_vROIs1’.

Step 1.5: Spectral signature analysis of ROIs

Purpose: This step extracts pixel-level reflectance values from the stacked Sentinel-2 image for each ROI polygon and visualizes their spectral signatures. This is used to assess the spectral separability of land cover classes.

Preparation:

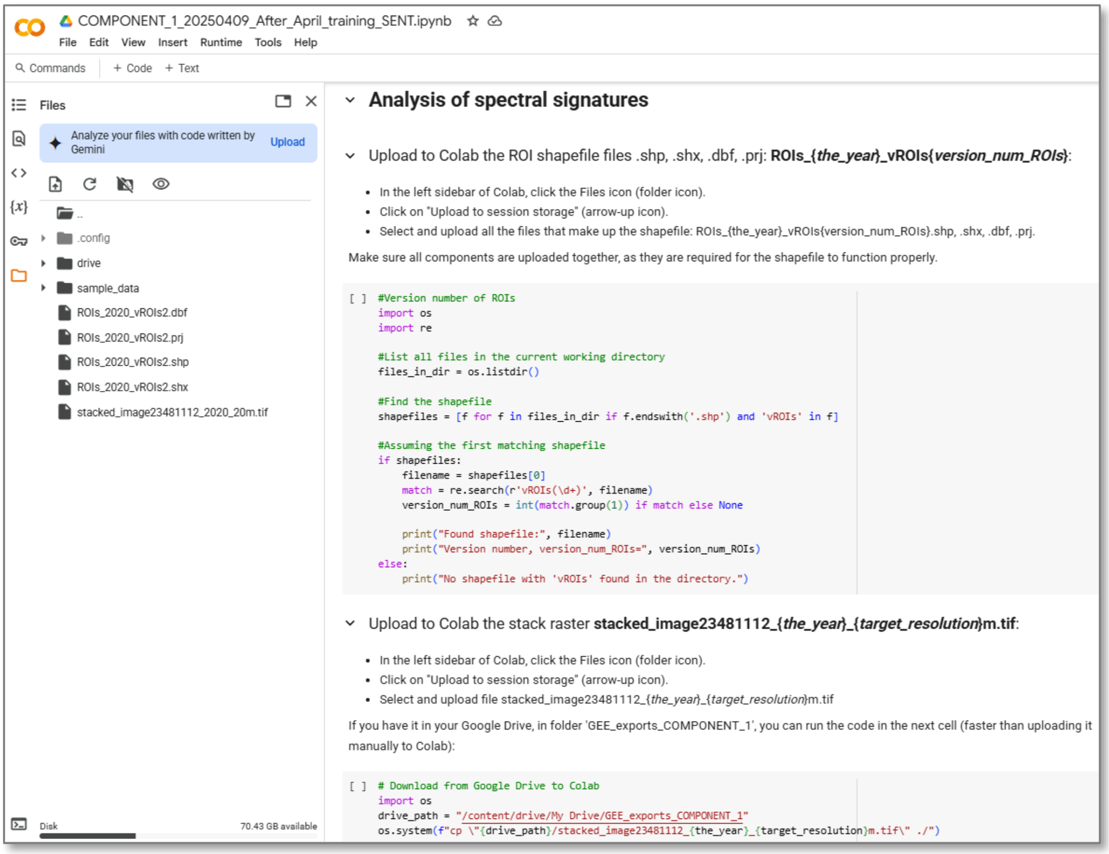
1) When requested in the code, upload the required files to Colab:

   - ROI shapefile: ROIs_{the_year}_vROIs{version_num_ROIs}.shp, .prj, .dbf, .shx.
   - The raster stack: stacked_image23481112_{the_year}_{target_resolution}m.tif.

     If you have completed Step 1.1 for the relevant year, the raster stack should be already in your Google Drive under 'GEE_exports_COMPONENT_1'. In this case, you can run the provided code to copy it to Colab directly.

   For example, if you want to visualize the spectral signatures for ROIs in 2020 (version 2), you will need to upload shapefile 'ROIs_2020_vROIs2.shp' and 'stacked_image23481112_2020_20m.tif'.

   Similarly, if you want to visualize the spectral signatures for ROIs in 2023 (version 1), you will need to upload shapefile 'ROIs_2023_vROIs1.shp' and 'stacked_image23481112_2023_20m.tif'.



Process:

1) Continue from Step 1.1 in the same notebook:
   Use the Colab notebook already loaded in Step 1.1.

2) Extract pixel values from the stacked raster:

- Raster values for all valid pixels within each ROI polygon are extracted and combined into a reference dataset.

3) Export the reference dataset:

- The reference dataset is saved as a CSV file ('data_extrVal_geom_crs_{the_year}_vROIs{version_num_ROIs}.csv') to Google Drive.

4) Visualize spectral signatures

- Line plots of mean reflectance values are created for each land cover class.
- Standard deviation bands are plotted around the means to show within-class variability.
- Separate plots group 'similar' classes (e.g., tree-dominated, non-vegetated) for easier comparison.

Output:

- A CSV file containing extracted spectral data for each ROI (reference dataset) saved as 'data_extrVal_geom_crs_{the_year}_vROIs{version_num_ROIs}.csv' in your Google Drive under 'GEE_exports_COMPONENT_1'.
- Visual plots of spectral signatures by land cover class.

Component 2

The purpose of Component 2 is to train a machine learning model for land cover classification in Step 2.2 using the Regions of Interest (ROIs) from Component 1 and the stacked image generated in Step 2.1. Its output is a trained Random Forest model, which is used for land cover classification in Step 2.3.

In this section, we will use the year 2020 and ROIs version 2 as an example (i.e., the shapefile named 'ROIs_2020_vROIs2'). Throughout the steps, generic names will be used (e.g., 'ROIs_{the_year}_vROIs{version_num_ROIs}') to help you apply the workflow to any year or version. Once Steps 2.1 through 2.3 are completed for 2020, the same process can be repeated for other years, such as 2023, using ROIs version 1.

Note that during the training, it was also discussed that Steps 2.2 and 2.3 can alternatively be carried out using the R notebook developed by Christophe. The R code also produces a land cover classified map for the selected year, using the outputs from previous PAELA steps as inputs.

Step 2.1: Create and Export the Stack

Purpose:
This step creates a stack for a specified year using Sentinel-2 L2A data. This stack contains more layers than the one used in Component 1, and includes median composites of Sentinel-2 bands and calculated indices both for the entire year and for the dry and wet seasons. The stacked image will be used in subsequent steps for model training and land cover classification.

Preparation:

1) Download the Colab notebook:

- Locate the instructions file 'Colab_COMP2_notebook_and_instructions_SENT.txt' in the directory:
  '\Z_Visit_Vanuatu_April2025\Component_2\code\'.
- Follow the instructions in the text file to download and upload the notebook to your Google Drive and open it in Google Colab.

Process:

1) Start running the notebook in Google Colab, one cell at a time and following all instructions given in the notebook:

- The initial cells in the notebook will install necessary Python packages, initialize Google Earth Engine, and mount your Google Drive.
- Don't forget to manually restart the runtime at the point specified in the notebook – if you do not do this, you will encounter errors.

2) Generate the stack:

- Specify the target year for the stack (the_year) and the desired spatial resolution (target_resolution) in the relevant cells of the notebook.
  For example, set the_year = 2020 and target_resolution = 20.
- The notebook will:
  - Load Sentinel-2 L2A data for the specified year within the region of interest ('bbox_small'). Selected bands are Blue(2), Green(3), Red(4), Near Infrared (NIR), and Shortwave Infrared 16 (SWIR16).
  - Calculate the following indices:
    - NDVI (Normalized Difference Vegetation Index)
    - MNDWI (Modified Normalized Difference Water Index)
    - SAVI (Soil-Adjusted Vegetation Index)
    - BSI (Bare Soil Index)
    - VCP (Vegetation Condition Proportion)
- Compute median composites for each band and index across the entire year, as well as for the dry (May–October) and wet (November–April) seasons.
- Combine the indices and bands into a single multi-band stack.

3) Save the output:

- The code saves the stack to Google Drive under folder 'GEE_exports_COMPONENT_2'.
- Manually copy this file from Google Drive to your computer under:
  '…\Z_Visit_Vanuatu_April2025\Component_2\input\ZZZZ'.

  For example, if the working folder is in your D: drive, manually copy 'stacked_image_2020_20m.tif' from your Google Drive to 'D:\Z_Visit_Vanuatu_April2025\Component_2\input\2020'.

Output:

- The output is a GeoTIFF file of the stacked image for the specified year.
- The stack contains the following bands:
  - Median values of Sentinel-2 bands (B2, B3, B4, B8, B11).
  - Annual indices: NDVI, MNDWI, SAVI, BSI, VCP.
  - Seasonal indices for the dry and wet seasons: NDVI, MNDWI, BSI.

Step 2.2: Model training & evaluation

Purpose:

This step trains a Random Forest model using the output ROIs from Component 1 and the stacked raster created in Step 2.1. The trained model will be used for land cover classification in Step 2.3.

Preparation:

1) When requested in the code, upload the required files to Colab:

- ROI shapefile files: ROIs_{the_year}_vROIs{version_num_ROIs}.shp, .prj, .dbf, .shx, from the folder '\Z_Visit_Vanuatu_April2025\Component_1\output\ZZZZ\vROIsY'.
- The stacked raster: stacked_image_{the_year}_{target_resolution}m.tif.

If the raster stack is already in your Google Drive under 'GEE_exports_COMPONENT_2', run the provided code to copy it to Colab directly.

For example, to train the model for 2020 using the ROIs version 2, you will need to upload the shapefile components for 'ROIs_2020_vROIs2' and the stacked raster 'stacked_image_2020_20m.tif'.

Process:

1) Continue from Step 2.1 in the same notebook:
   Use the Colab notebook already loaded in Step 2.1, as all steps in Component 2 are executed within the same notebook.
2) Extract pixel values from the stacked raster:

- Raster values for all valid pixels within each ROI polygon are extracted and combined into a reference dataset.

3) Export the reference dataset:

- The reference dataset is saved as a CSV file ('data_extrVal_geom_crs_{the_year}_vROIs{version_num_ROIs}.csv') to Google Drive.

4) Train the Random Forest model:

- The dataset is split into training and testing sets.
- A Random Forest model is trained using grid search for hyperparameter tuning.
- The best-performing model is selected for evaluation.

5) Evaluate the model:

- The model's accuracy is calculated, and a classification report is generated. A confusion matrix is also generated.

6) Save the trained model:

- The trained model, along with its metadata (feature names, statistics, etc.), is saved as a .joblib file in Colab (e.g., 'VANmodel_2020_vROIs2.joblib').
- The model file is automatically copied to Google Drive under the directory 'GEE_exports_COMPONENT_2'.
- Any existing model files with the same name are deleted before saving the new one.

7) Visualize model performance:

- Feature importance is visualized as a bar chart.
- The confusion matrix is visualized using a heatmap for better interpretation of the classification results.

Output:

- A trained Random Forest model saved as 'VANmodel_{the_year}_vROIs{version_num_ROIs}.joblib' in your Google Drive under 'GEE_exports_COMPONENT_2'.

Step 2.3: Land cover classification

Purpose:
This step applies the trained Random Forest model from Step 2.2 to predict land cover types for the entire study area using the stack generated in Step 2.1. It classifies the land cover based on the model's prediction and saves the results as a GeoTIFF file. The land cover classes are based on the classification table defined in Step 1.3.

Process:

1) Continue from Step 2.2 in the same notebook:

- Use the same Colab notebook previously loaded in Step 2.1, as all steps in Component 2 are executed within the same notebook.

2) Raster stack for prediction:

- The script begins by loading the raster stack from the specified year.

3) Make predictions:

- The model predicts land cover classes for each pixel in the raster stack.

4) Save predictions as GeoTIFF:

- The predictions are saved as a GeoTIFF file 'pred_landcover_{the_year}_vROIs{version_num_ROIs}.tif' in your Google Drive under 'GEE_exports_COMPONENT_2'.

  For example, the classified land cover rater is saved as 'pred_landcover_2020_vROIs2.tif'.

- Manually save the classified land cover map, the model, and the reference dataset csv from your Google Drive to your computer under '\Z_Visit_Vanuatu_April2025\Component_2\output\theYear\vMX'.
  For example, if this is the first version of the model, save the files to '\Z_Visit_Vanuatu_April2025\Component_2\output\2020\vM1'.

5) Plot the classification map:

- A map of predicted land cover classes is generated, with each class assigned a unique color.

Output:

- The output is a GeoTIFF file representing the predicted land cover classification map, where each pixel is assigned a land cover class based on the trained Random Forest model.
- The file is saved as pred_landcover_{the_year}_vROIs{version_num_ROIs}.tif in your Google Drive folder 'GEE_exports_COMPONENT_2' as well as in your computer.

Component 3

The purpose of Component 3 is to refine the outputs of the land cover classification by applying post-classification tools to the classified raster maps for both opening and closing years. This includes smoothing the classified raster maps, converting them to vector format, identifying and adjusting geometry and topology issues (overlaps and gaps), performing an initial analysis of improbable land cover transitions, and adjusting the base vector land cover maps as needed. The result is a set of clean, adjusted base land cover maps in both vector and raster formats, and the (adjusted) land cover change matrix that is used as input for Component 4.

In this section, we use the year 2020 as an example to demonstrate the post-processing steps using the classified raster 'pred_landcover_2020_vROIs2.tif'. Once Steps 3.1 and 3.2 are completed for 2020, the same steps can be applied to other years, using the classified raster created in Component 2, such as 'pred_landcover_2023_vROIs1.tif'.

Step 3.1: Post-Processing Land Cover Classification

Purpose: The purpose of Step 3.1 is to post-process the land cover classification rasters by applying a majority filter and a sieve operation to remove isolated or noisy pixels. The smoothed raster is then vectorized, and the LC codes are joined with descriptive class names using a reference table.

Note that the post-processing tools—particularly the majority filter—can simplify the map by removing small, isolated patches. This may unintentionally eliminate important features such as small urban areas. If retaining these is important for your analysis, consider skipping the filtering step by commenting out the relevant sections of the PyQGIS script and vectorizing the raster output from Component 2 directly. However, it is worth noting that this may result in a more detailed and heavier dataset.

Preparation:

1) Prepare the Script:

- Locate the PyQGIS script 'Component_3a_20250331_Post_processing.py' in the folder 'D:\Z_Visit_Vanuatu_April2025\Component_3\Scripts\'.
- This script will process the LC raster, vectorize it, join the CSV file for LC descriptions, and save the post-processed shapefile.

Process:

1) Set Parameters in the PyQGIS Script:

- Open QGIS and load the script by navigating to Plugins → Python Console → Show Editor → Open Script… Update the following variable in the script:
    - 'LC_Input_Raster': Path to the LC classified raster from Component 2 (e.g., 'D:/Z_Visit_Vanuatu_April2025/**Component_2**/output/2020/vM1/pred_landcover_2020_vROIs2.tif').
    - 'output_path': Path to save the output vector shapefile (e.g., 'D:/Z_Visit_Vanuatu_April2025/**Component_3**/output/2020').
    - 'LC_Raster_style', 'LC_Vector_style', and 'LCcodes_descriptions_txt': Only needs to be updated if you are using a drive other than your D: drive.

*Note: When entering file paths, always use **forward slashes** (/) instead of backslashes (\). This will allow compatibility across operating systems (Windows and Linux) and avoids errors in Python.*

2) Run the Script in QGIS:

- Execute the script to perform the following:
    - Post-process the classified land cover raster map.
    - Convert Raster to Vector.
    - Join LC Descriptions: The land cover classification codes from the vectorized data will be matched with their corresponding descriptions from the CSV file with land cover classification details (see Step 1.3) .
    - Save Output: The labeled shapefile will be saved to the specified output path.

Output:

The final output of this step is the post-processed land cover vector dataset in shapefile format (e.g., 'RF_Classification_2020_vROIs2_MS_VS')

Step 3.2: Overlaps and gaps

Purpose:

The purpose of Step 3.2 is to identify and address geometry and topology issues such as overlaps and gaps between polygons in the LC vector dataset from Step 3.1. The final output will be a cleaned and adjusted land cover shapefile.

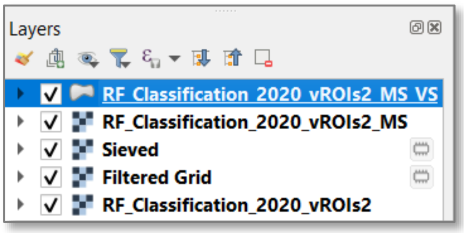
Preparation:

1) Prepare the Script:

- Locate the PyQGIS script ‘Component_3b_20250331_Post_processing.py’ in the folder ‘…\Z_Visit_Vanuatu_April2025\Component_3\Scripts\’.
  This script will identify and address overlaps and gaps, update the area calculations, and save the adjusted shapefile.

2) Add the required Layers to QGIS:

- Load the land cover shapefile from Step 3.1 (e.g., ‘RF_Classification_2020_vROIs2_MS_VS.shp’) into QGIS and make it the active layer by selecting it in the Layers panel. **This is essential for the following steps to work correctly.**



Process:

1) Set Parameters in the PyQGIS Script:

- Open QGIS and load the script by navigating to Plugins → Python Console → Show Editor → Open Script… Update the following variables in the script:
  - 'output_path': Path to save the datasets containing the overlaps and gaps, and the cleaned land cover shapefile (e.g., ‘D:/Z_Visit_Vanuatu_April2025/Component_3/output/2020’).

  *Note: When entering file paths, always use **forward slashes** (/) instead of backslashes (\). This will allow compatibility across operating systems (Windows and Linux) and avoids errors in Python.*

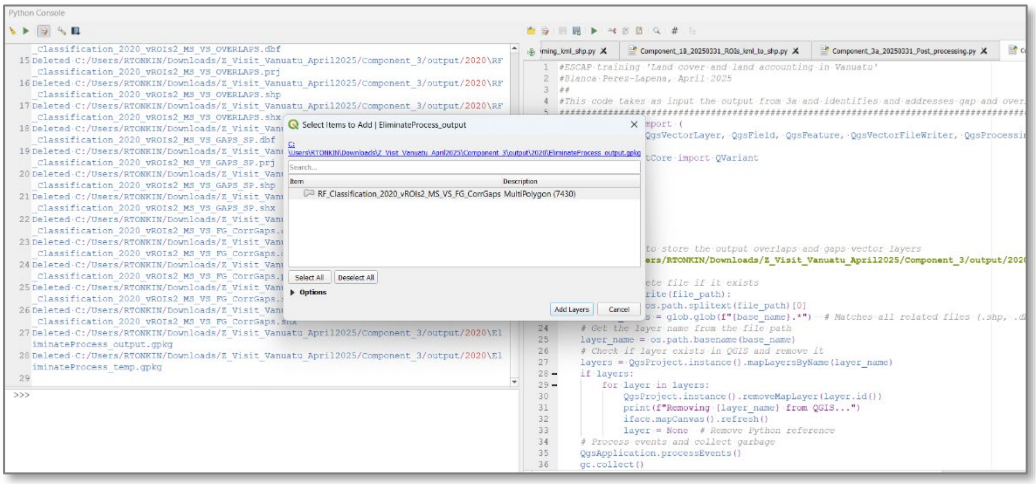
2) Run the Script in QGIS:

- Execute the script to perform the following:
  - Identify and automatically address overlap and gap issues.

    Note**:** When the code begins the automated elimination of gaps, a pop-up window “Select items to Add” may appear. This happens because the script is loading a GeoPackage that may contain multiple layers, and QGIS requires the user to confirm which layer to load.

In this window:

- Click on the listed item.
- Then click on “Add Layers” to continue the process.



  - Update the area for each feature.
  - Save the adjusted shapefile to the specified output path.

Output:

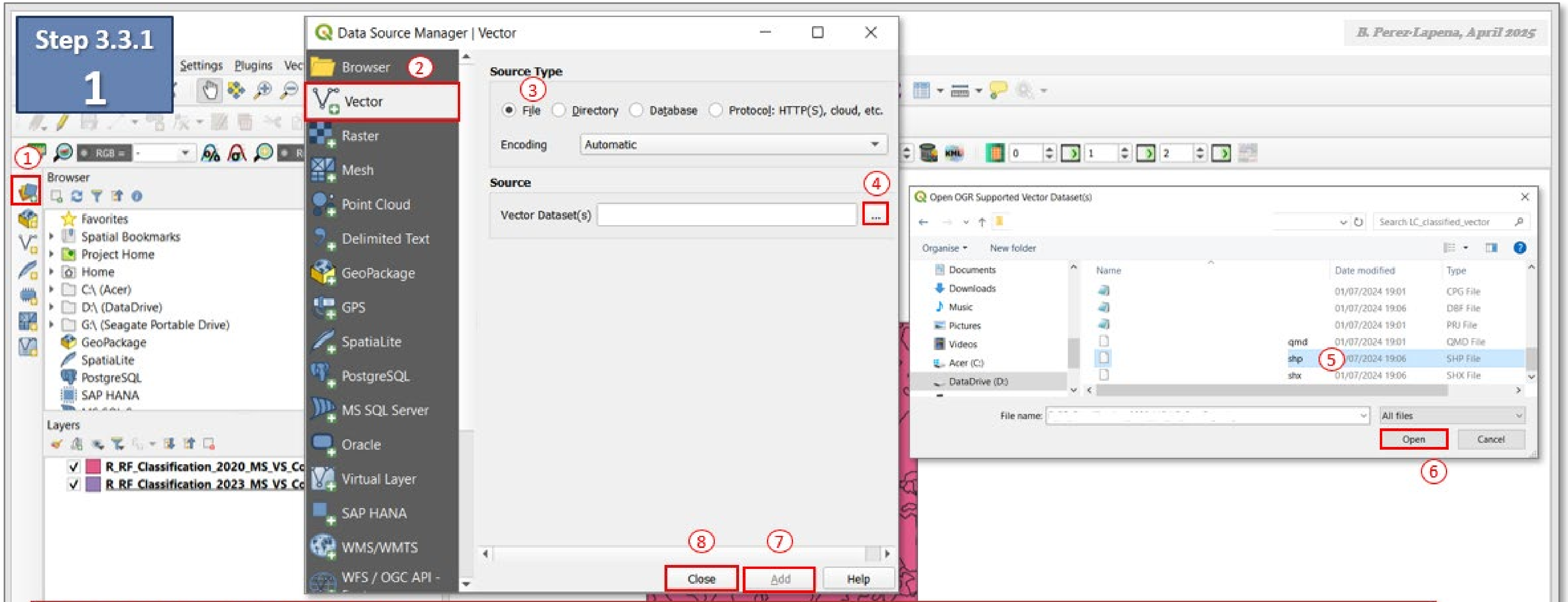
Adjusted land cover shapefile (e.g., ‘D:\Z_Visit_Vanuatu_April2025\Component_3\output\2020\RF_Classification_2020_vROIs2_MS_VS_FG_CorrGaps_I.shp’).

Step 3.3: Land Cover Change

In this step, the focus is on land cover change analysis. Please note that the slides you will be following include images from the previous training. The screenshots may contain outdated names and paths, which should be ignored. **The corresponding text boxes on the slides reflect the correct names and paths provided** you need to use for this analysis. For these sub-steps we will use 2020 as the opening year and 2023 as the closing year.

Step 3.3.1: Save Vector Files Under Another Name

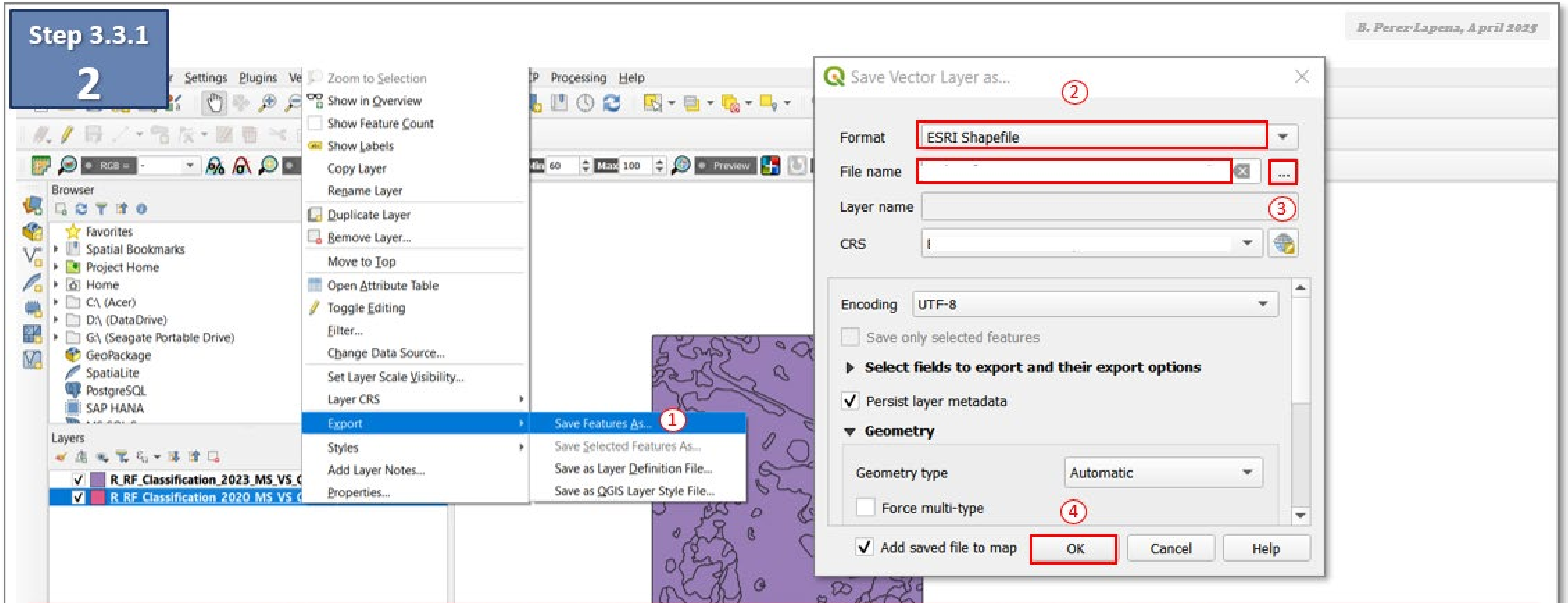
Under this sub-step, you will save your vector files under a new name to avoid overwriting your original data. This way any changes you make can be tracked and the original data remains intact.



1. Click on "Open Source Manager" icon
2. Click on the "Vector" tab
3. Enable the "File" source type
4. Press the "..." button
5. In folder "\Z_Visit_Vanuatu_April2025\Component_3\output\2020" select the shapefile "RF_Classification_2020_vROIs2_MS_VS_FG_CorrGaps_l.shp"
6. Click on "Open"
7. Click on "Add"

Repeat steps 1 through 7 to add the 2023 vector land cover layer "RF_Classification_2023_vROIs1_MS_VS_FG_CorrGaps_l.shp" in folder "\Z_Visit_Vanuatu_April2025\Component_3\output\2023\"

8. Click on "Close"



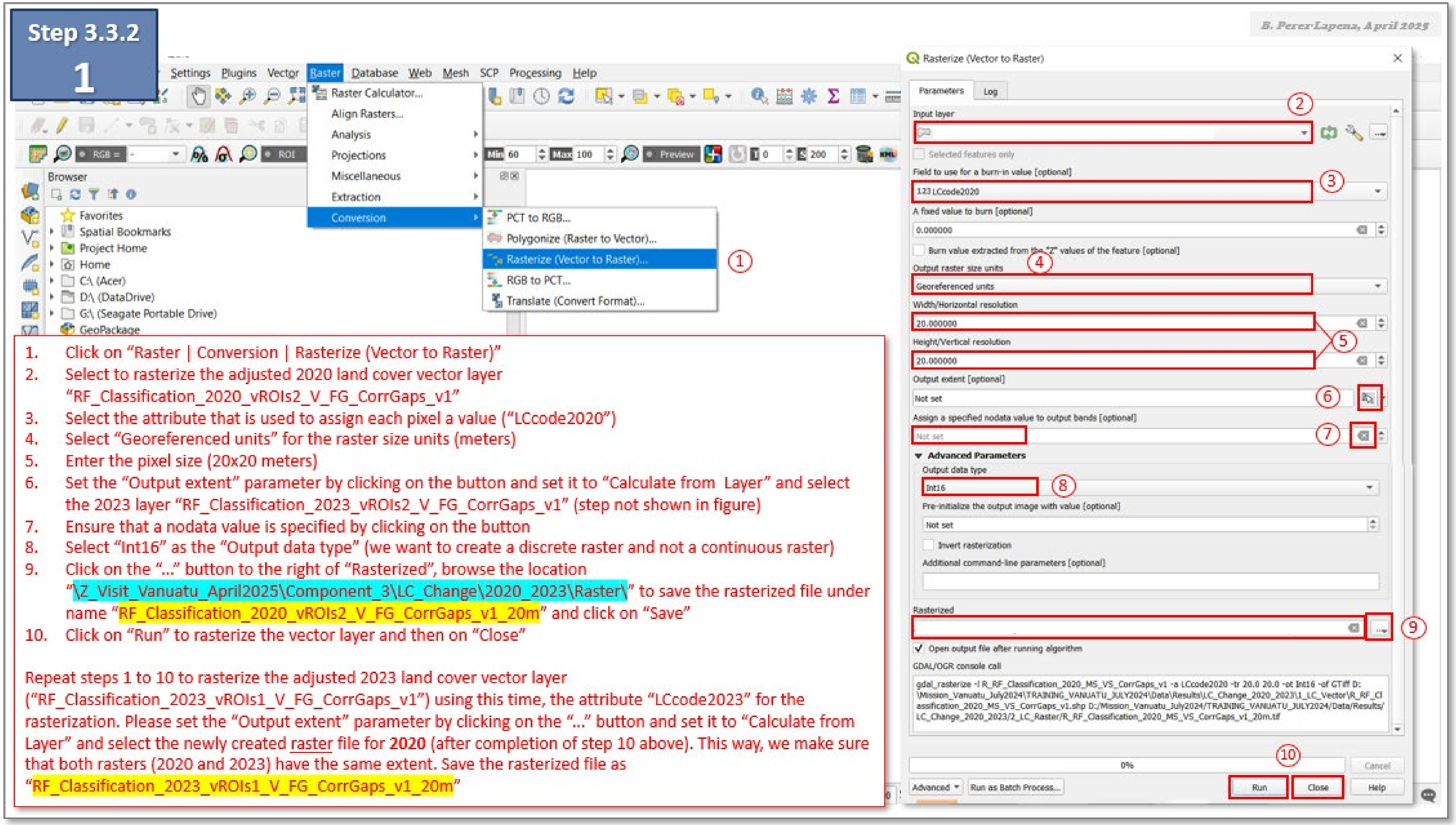
We are going to save each of the adjusted land cover layers you just added to the project as a new shapefile (ending with "_v1")

1. Right click on layer "RF_Classification_2020_vROIs2_MS_VS_FG_CorrGaps_I" and then on "Export | Save Features As"
2. Save the layer as a shapefile
3. Save the layer in folder "\Z_Visit_Vanuatu_April2025\Component_3\LC_Change\2020_2023\Vector" under name "RF_Classification_2020_vROIs2_V_FG_CorrGaps_v1.shp"
4. Click on "OK"
5. Repeat steps 1 to 4 to save layer in the same folder but this time for the year 2023 under name "RF_Classification_**2023**_vROIs1_V_FG_CorrGaps_v1.shp"

Remove the non-v1 layers from the Layers panel

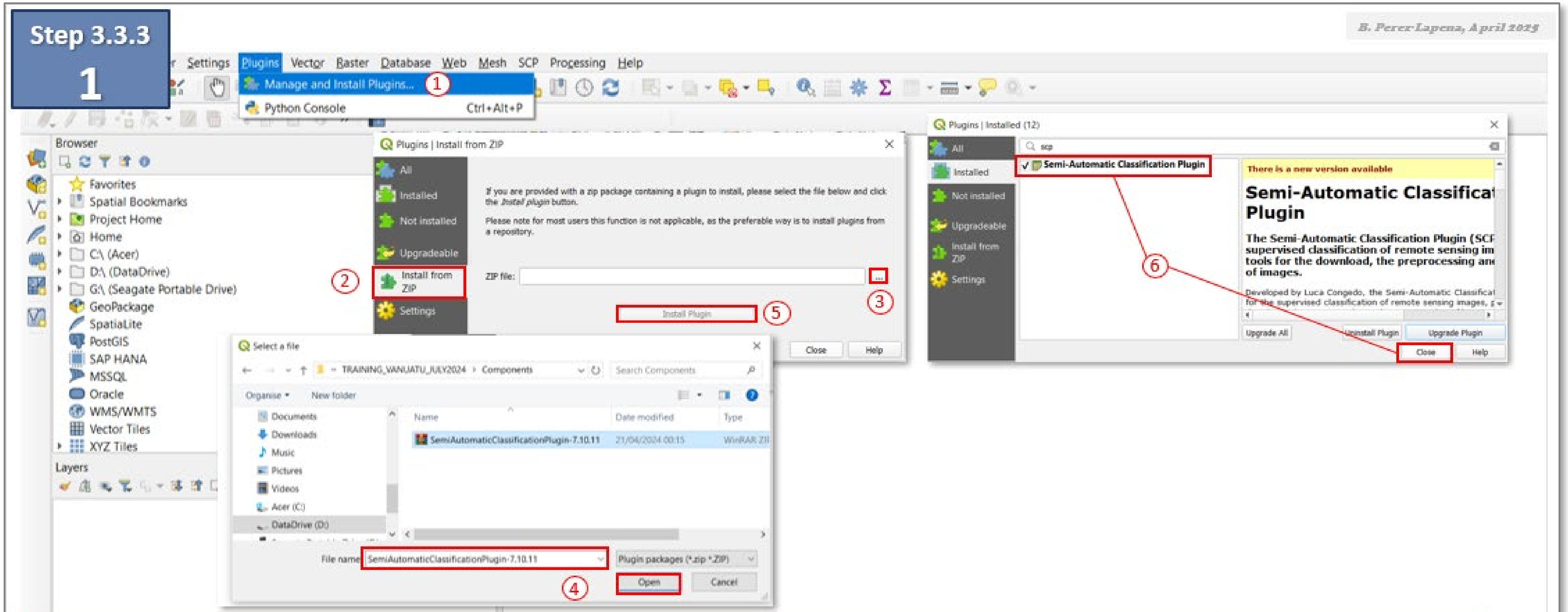
Step 3.3.2: Rasterize

In this sub-step we will convert vector data to raster format for further processing.



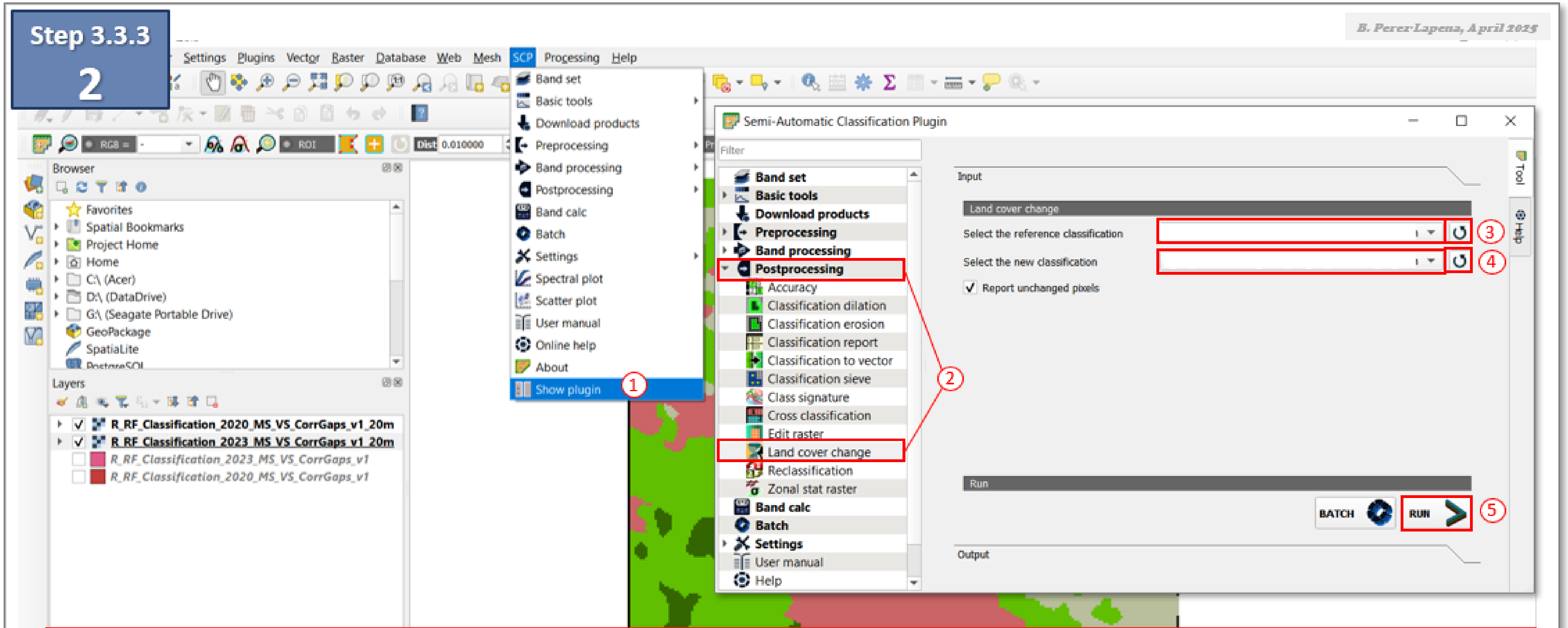
Step 3.3.3: Land cover change matrix/map

This sub-step creates a land cover change matrix and map to analyze the changes between two time periods.



1. Click on "Plugins | Manage and Install Plugins..."
2. Click on "Install from ZIP"
3. Click on the "..." button to the right of "ZIP file" and browse the location of the plugin "..\Components"
4. Select "SemiAutomaticClassificationPlugin-7.10.11" and click on "Open"
5. Click on "Install Plugin". In the security warning window, click on "Yes"
6. Click on "Installed" and activate the "Semi-Automatic Classification Plugin" and click on "Close". It will then add the "SCP" menu, the "SCP Working Toolbar" and the "SCP Edit Toolbar"

Congedo, Luca, (2021). Semi-Automatic Classification Plugin: A Python tool for the download and processing of remote sensing images in QGIS. Journal of Open Source Software, 6(64), 3172, https://doi.org/10.21105/joss.03172



1. Open the SCP plugin
2. Click on "Postprocessing | Land cover change"
3. Click on the circle arrow to load the layer for the reference classification and select "RF_Classification_2020_vROIs2_V_FG_CorrGaps_v1_20m"
4. Click on the circle arrow to load the layer for the new classification and select "RF_Classification_2023_vROIs1_V_FG_CorrGaps_v1_20m"
5. Click on "Run" and save the land cover change map as "Improbable_transitions20202023_v1_20m" under folder "\Z_Visit_Vanuatu_April2025\Component_3\LC_Change\2020_2023\Improbable_transitions\Output_SCP\" (not shown in the figure)

The land cover change map will be added to the Layers panel and the land cover matrix will also be calculated, shown under "Output", and the output will also be saved into a .csv file with the same name as the map.

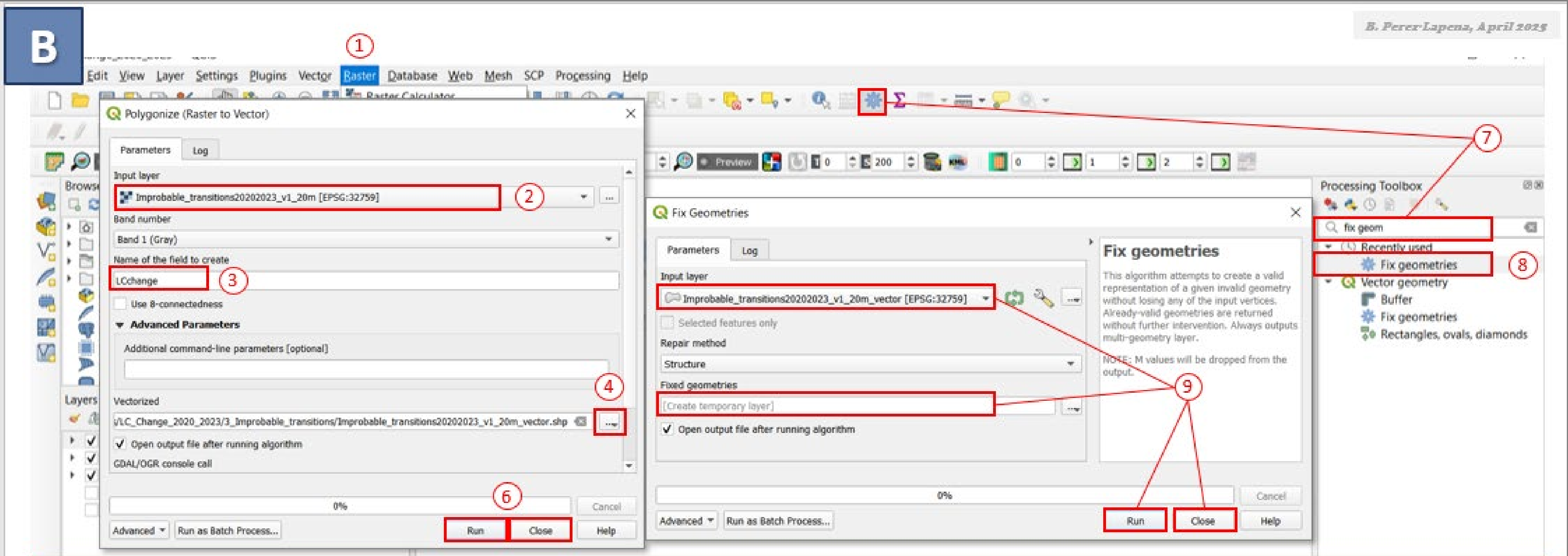
Step 3.3.4: Initial Improbable transitions analysis

In this step, we will focus on analyzing improbable transitions that may arise from the land cover change. This analysis involves several sub-steps:

A. <u>Fill in the Template</u>

1) Save the template "\Component_3\input\Improbable_transitions_template\ C3_TEMP_improbable_transitions_SENT.xlsx" to folder "\Component_3\LC_Change\2020_2023\Improbable_transitions\" under name "C3_TEMP_improbable_transitions20202023_v1_20m.xlsx"
2) Open the <u>destination</u> Excel file ("\Component_3\LC_Change\2020_2023\Improbable_transitions\ **C3_TEMP_improbable_transitions20202023_v1_20m.xlsx**" from step 1 above), select yellow cells in the top table, and "Delete" the contents of those yellow cells.
3) Open the <u>source</u> csv file ("\Component_3\LC_Change\2020_2023\Improbable_transitions\Output_SCP\ **Improbable_transitions20202023_v1_20m.csv**", output of the SCP plugin). Highlight column A and select Data → Text to Columns. Use a tab delimiter to format the data.
   Select the cell values from the top table (excluding the column names) | Copy
4) Go to the <u>destination</u> Excel file, click on cell A2 | Paste | Paste Values
5) In the <u>destination</u> Excel file, select cells in the top table (columns A through J) | Copy | Click to create a new workbook | Paste Values. Save this workbook as a .csv file "\Component_3\LC_Change\2020_2023\Improbable_transitions\Output_SCP\ LC_Change_20202023_Lookup.csv" (This csv file will be used in QGIS as a lookup table to understand the LC change codes)
6) In the <u>destination</u> Excel file, go to the land cover change matrix, select yellow cells starting on "V_ReferenceClass", and "Delete" the contents of those yellow cells
7) In the <u>source</u> csv file, select the cell values from the land cover change matrix (starting on "V_ReferenceClass" and excluding the row/column labelled "Total") | Copy
8) Go to the <u>destination</u> Excel file, click on cell B180 | Paste | Paste Values

B. Polygonize: Convert the raster data to polygons

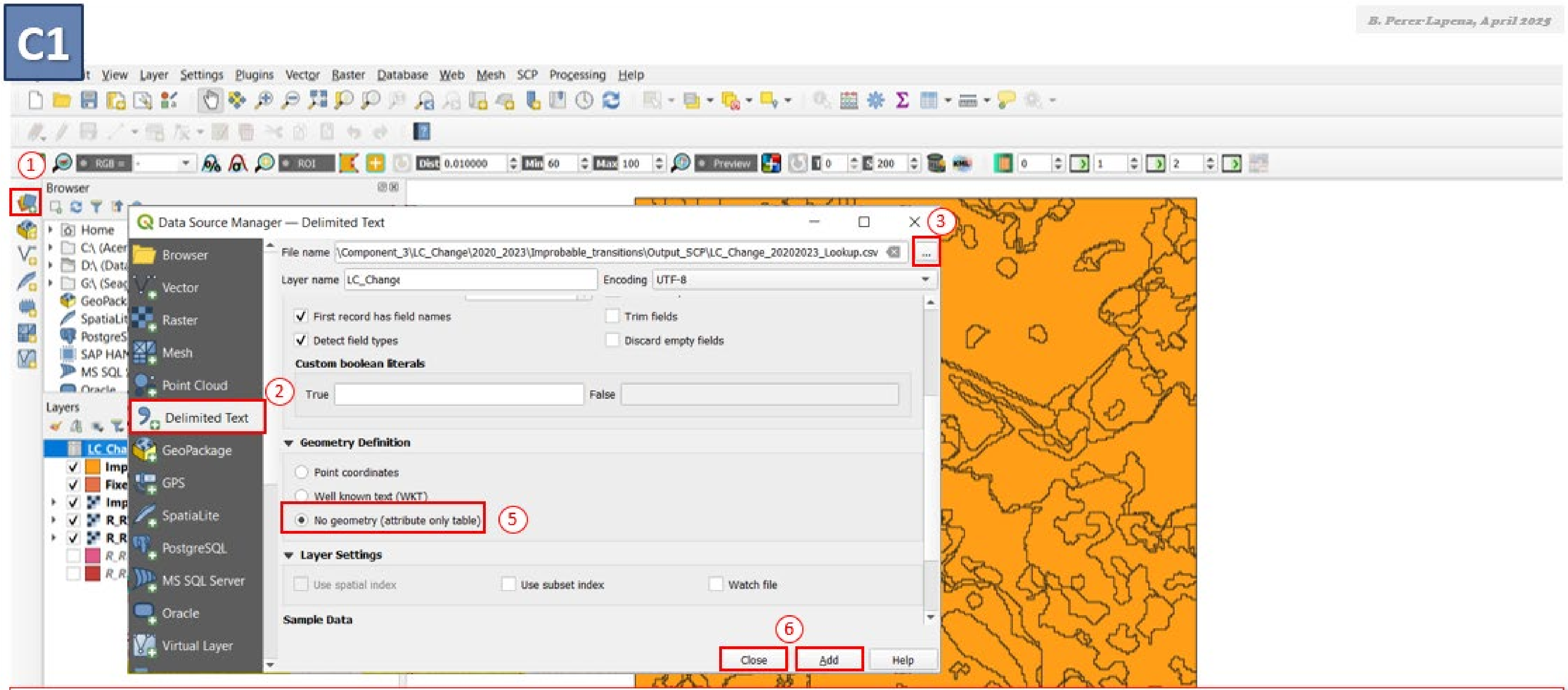


1. Click on the "Raster" menu, then on "Conversion" and select the "Polygonize" (Raster to Vector)" tool
2. Select the raster layer "Improbable_transitions20202023_v1_20m"
3. Type "LCchange" in the "Name of the field to create" box
4. Click on the "…" button and on "Save to file" (not shown in the figure)
5. Save it as a **shapefile** in the "\Component_3\LC_Change\2020_2023\Improbable_transitions" folder under name "Improbable_transitions20202023_v1_20m_vector" (step not shown in the figure)
6. Click on "Run" and then on "Close". The newly created layer will be added to the Layers panel

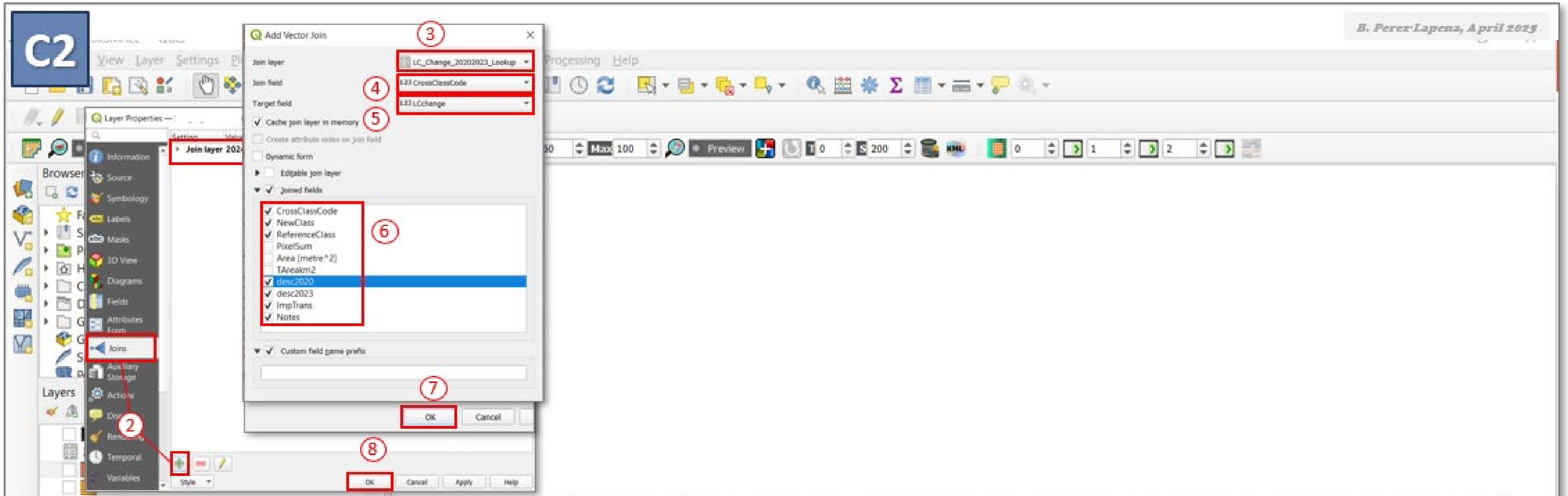
The polygonise operation may lead to geometry issues in the resulting layer. Fixing the geometries of layer "Improbable_transitions20202023_v1_20m_vector" can address these potential issues.

7. Click on the "Toolbox" icon to open the "Processing Toolbox". In the search box of the "Processing Toolbox" type "fix geom" to search for the tool
8. Double-click on the tool "Fix geometries"
9. Select layer "Improbable_transitions20202023_v1_20m_vector" and leave the output as a temporary layer. Click on "Run" and then on "Close"
10. Right-click on layer "Improbable_transitions20202023_v1_20m_vector" and then on "Remove Layer…" (we will remove the layer from the Layers panel because, in the next steps, we will overwrite the fixed geometries layer). This step is not shown in the figure
11. Right-click on the temporary "Fixed geometries" layer and on "Save Features As…", and save it in the same "Improbable_transitions" folder as "Improbable_transitions20202023_v1_20m_vector", overwriting the existing file with that name. (When asked to replace it, as it already exists, click on "Overwrite file"). This step is not shown in the figure

C. Join Table with LC Change Class Descriptions



1. Click on "Open Source Manager" icon
2. Click on the "Delimited Text" tab
3. Press the "..." button
4. Select "\Component_3\LC_Change\2020_2023\Improbable_transitions\Output_SCP\LC_Change_20202023_Lookup.csv". This file has been created using the output of the SCP plugin, associating to the codes the corresponding land cover class names.
5. Set the geometry definition as "No geometry (attribute only table)"
6. Click on "Add" and then on "Close"



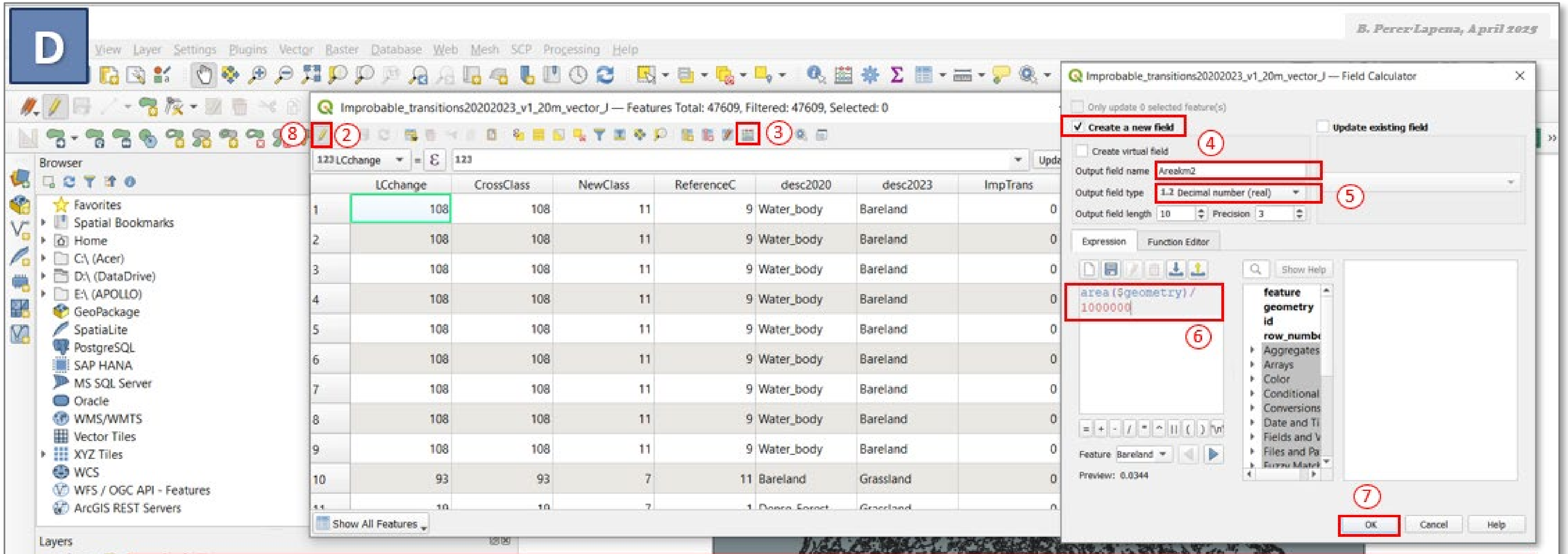
The attribute table of **"Improbable_transitions20202023_v1_20m_vector"** only contains the "LCchange" codes from applying the tool "polygonize". We would like to know what those codes refer to. This information is contained in the csv file we just added **"LC_Change_20202023_Lookup"** (see Slide C1). We are going to join several attribute values:

1. Right-click on the vector layer name "Improbable_transitions20202023_v1_20m_vector" and click on "Properties" (step not shown in figure)
2. Click on the "Joins" tab and on the button "Add new join"
3. Select layer "LC_Change_20202023_Lookup", which attribute table is the table to join (contains the descriptions for the codes)
4. Select "CrossClassCode" as the common attribute in file "LC_Change_20202023_Lookup" that will be used in the join
5. Select "LCchange" as the target attribute in layer "Improbable_transitions20202023_v1_20m_vector"
6. Tick to join fields and the attributes "CrossClassCode", "NewClass", "ReferenceClass", "desc2020", "desc2023", "ImpTrans", and "Notes". For each row, the values in these attributes will be brought to the attribute table of "Improbable_transitions20202023_v1_20m_vector" when there is a match in their change codes.
7. Click on "OK"
8. Click on "OK" to close the layer properties window

Joins are temporary. To make the join permanent, we need to save it as a new layer:

9. Right click on layer "Improbable_transitions20202023_v1_20m_vector" and then on "Export | Save Features As". Save the layer as **"Improbable_transitions20202023_v1_20m_vector_J"** in the same folder (this step is not shown in the figure)

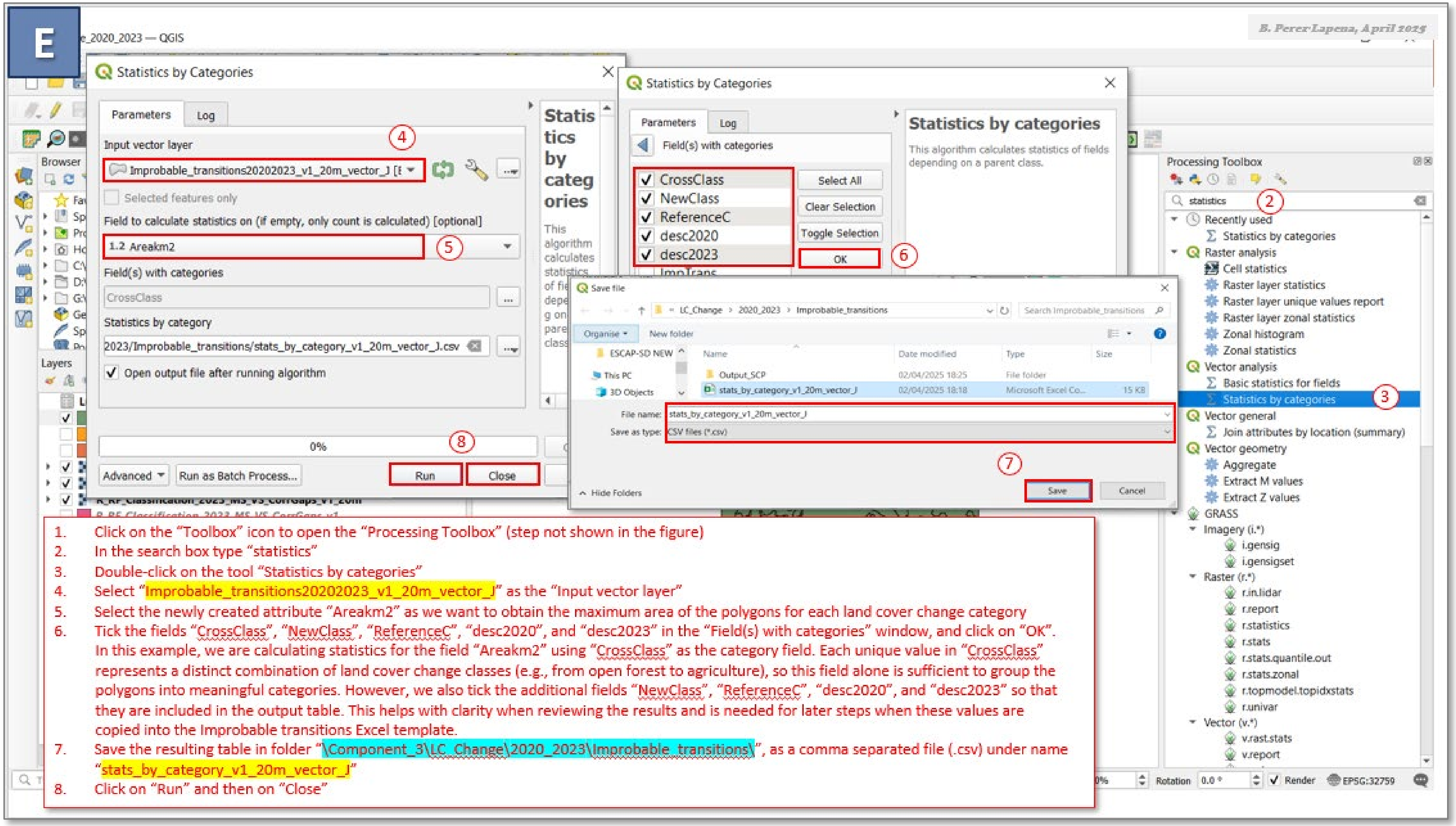
D. Calculate the area of each polygon in the dataset.



Add a new attribute containing the area of each polygon in "Improbable_transitions20202023_v1_20m_vector_J":

1. Right-click on layer "Improbable_transitions20202023_v1_20m_vector_J" and click on "Open Attribute Table" (step not shown in the figure)
2. Start an edit session by clicking on button "Toggle editing mode"
3. Click on the button "Open Field calculator"
4. Tick "Create a new field" and name the new field "Areakm2"
5. Select "Decimal number" as the data type
6. Type "area($geometry)/1000000" in the "Expression" box
7. Click on "OK" and the area, in square kilometers, will be calculated for each polygon in the layer and stored in the newly created attribute "Areakm2"
8. Stop the edit session by clicking on button "Toggle editing mode" and click on "Save" when prompt to save the changes

E. Perform a Statistics by categories operation



F. Fill in Template

1) In the destination Excel file ("\Component_3\LC_Change\2020_2023\Improbable_transitions\C3_TEMP_improbable_transitions20202023_v1_20m.xlsx", select yellow cells on the last table all the way below, and "Delete" the contents of those yellow cells.

2) Open the statistics by category table we just created "\Component_3\LC_Change\2020_2023\Improbable_transitions\stats_by_category_v1_20m_vector_J.csv"| Copy the desired data column by column and paste it in the corresponding columns of the last table all the way below in the destination Excel file (using Paste | Paste Values)

Step 3.3.5: Address Improbable Transition Polygons

After identifying improbable transition polygons, adjustments may be necessary. If so, make the required adjustments and then repeat Step 3.3.1 (saving the file as a new version, e.g. v2), Step 3.3.2 and Step 3.3.3. Finally, save the final vector and raster files in folders 'LC_vector_FINAL' and 'LC_raster_FINAL', respectively.

Component 4

Step 4.1: Land account

The purpose of this component is to compile the Land Account for the chosen accounting period, summarizing land cover changes based on the (adjusted) land cover change matrix produced in Component 3. This involves reorganizing the matrix into a format that presents the opening and closing areas for each land cover class, as well as the net changes, expansions, and regressions observed during the accounting period.

1) Open the <u>land account destination</u> Excel file ("\Component_4\Land_account\20250331_LC_accounts_VAN_bpl_SENT.xlsx"), select yellow cells in the bottom table, and "Delete" the contents of those yellow cells.

2) In the <u>source</u> csv file ("\Component_3\LC_Change\2020_2023\Improbable_transitions\Output_SCP\Improbable_transitions20202023_v1_20m.csv", output of the SCP plugin), select cell values from the land cover change matrix (starting on "V_ReferenceClass" and excluding row/column "Total") | Copy

3) Go to the <u>land account destination</u> Excel file, click on cell B50 | Paste | Paste Values

Important: The land cover change matrix at the bottom of the sheet that is used in the land account must include the same set of land cover classes for both the opening and closing years. This ensures the matrix is square (i.e., has the same number of rows and columns in the same order) and aligns correctly with the formulas in the land account Excel template.

This means:

- If a land cover class is present in the opening year but not in the closing year, you must still include it as a column in the matrix (with all values set to zero).
- If a class is only present in the closing year, add a corresponding row with zero values.

The land account template requires the land cover change matrix to follow a fixed structure. You <u>**must not insert or delete columns or rows**</u> in Excel, as this will break the formulas and links in the template. Instead, follow these steps:

a) Locate the land cover change matrix at the bottom of the sheet (the one in square meters) in the <u>land account destination</u> Excel file ('\Component_4\Land_account\20250331_LC_accounts_VAN_bpl_SENT.xlsx').

b) Identify the missing class:

- Compare the list of row labels (opening year) and column headers (closing year). If a class appears in one and not the other, it must be added to the side where it is missing.

c) Make space without inserting a new column or row:

- If the class is missing from the **columns** (closing year):
    - Find the correct position where the class should appear, based on its land cover class code.

    For example, if class 10 (Shrubs) is missing:

    - Cut the cell values starting from the column for class 11 and all columns to the right (i.e., the entire range of cells, not the full columns).
    - Paste those cells one column to the right to make space.

    - In the newly created empty column, enter the missing class code at the top (e.g., 10) and fill all cells below with 0s (indicating no transitions into that class).
- If the class is missing from the **rows** (opening year):
    - Locate the correct position where the class should be inserted.
    - Cut the cell values from that row and all rows below it, then paste one row down to make space.
    - In the newly created empty row, enter the missing class code on the left and fill all cells to the right with 0s (indicating no transitions from that class).