

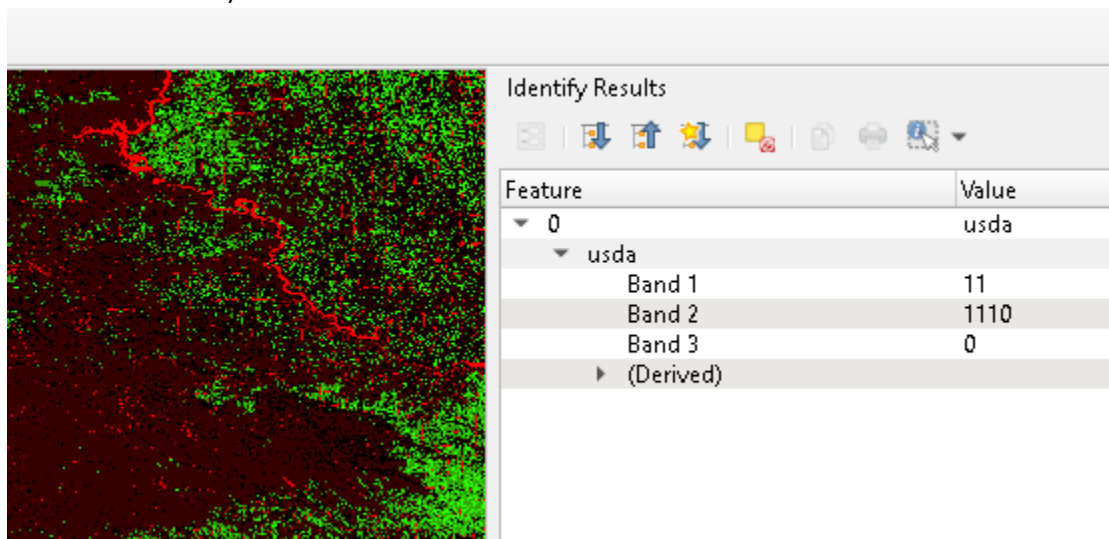
Raster Harmonization Instructions

Raster datasets need additional preprocessing compared to vector datasets

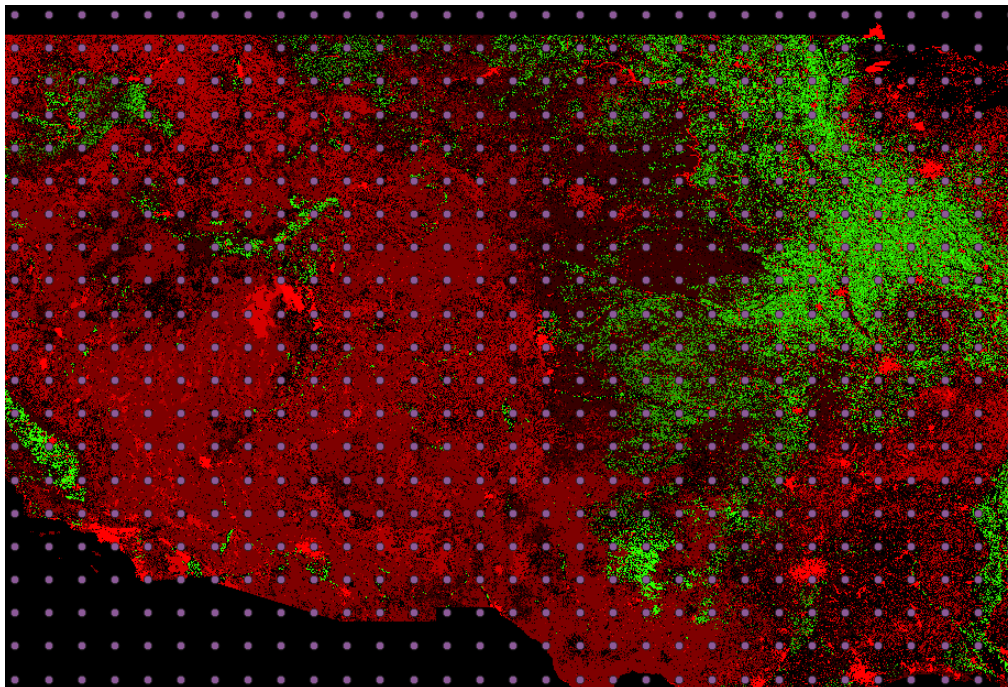
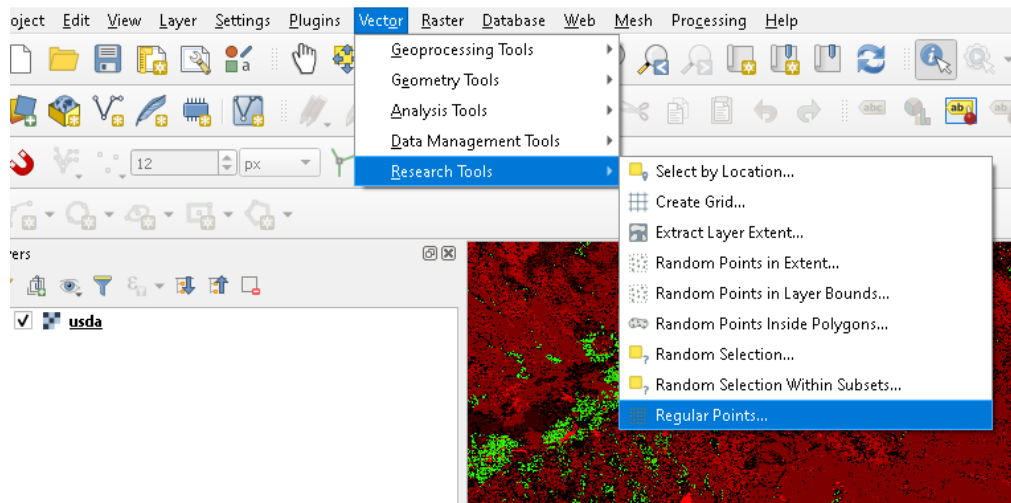
1. To start with, make sure the Raster (geotiffs etc) files are of 10m spatial resolution. Raster datasets can be resampled to match this resolution using standard GIS tools. An example of raster sampling using QGIS can be seen [here](#)
2. The raster file needs to be prepared according to the format required by the WorldCereal system (see [here](#)) to get a harmonized file. Each pixel needs to have 3 bands: Land cover (LC), crop type (CT), and irrigation (IRR). Band values can be zero if data is not available.
3. To use the data in the processing module, points need to be extracted from the raster using sampling techniques. This sampling should make sure that there is no spatial autocorrelation between pixels. An example using QGIS is shown below, although the procedure can be done in any GIS tool.

Generating point samples

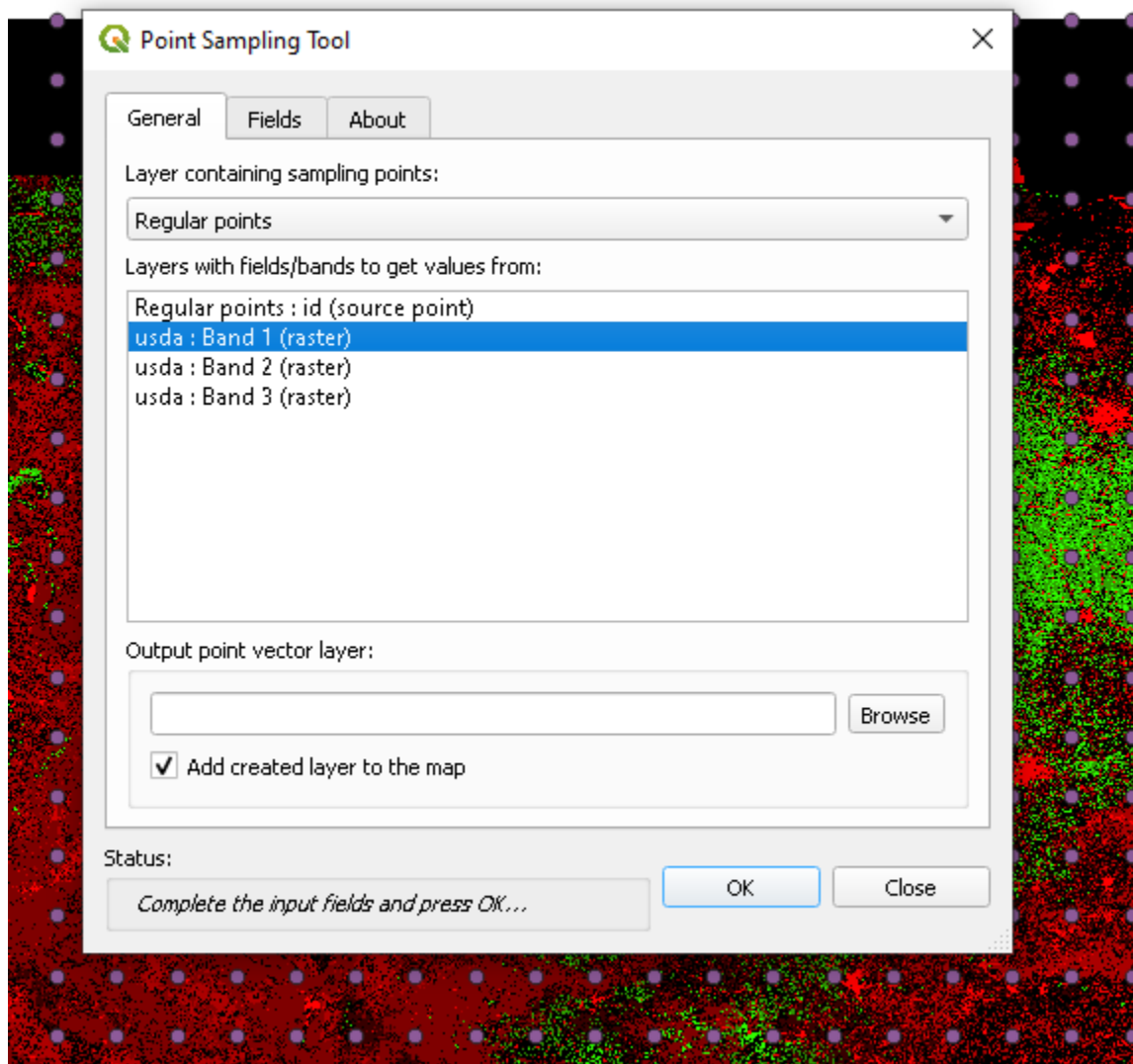
1. To reduce spatial autocorrelation amongst the samples, the points are selected from a harmonized raster file using the following procedure. Install the “Point Sampling Tool” plugin into QGIS so that samples can be generated.
2. The raster file needs to be already prepared according to the WorldCereal file format (see instructions above) and should contain 3 bands as shown below:



3. Create grid points on the raster using the Vector/research tools in QGIS as shown below. Make sure the samples are sufficiently apart to avoid spatial correlation.

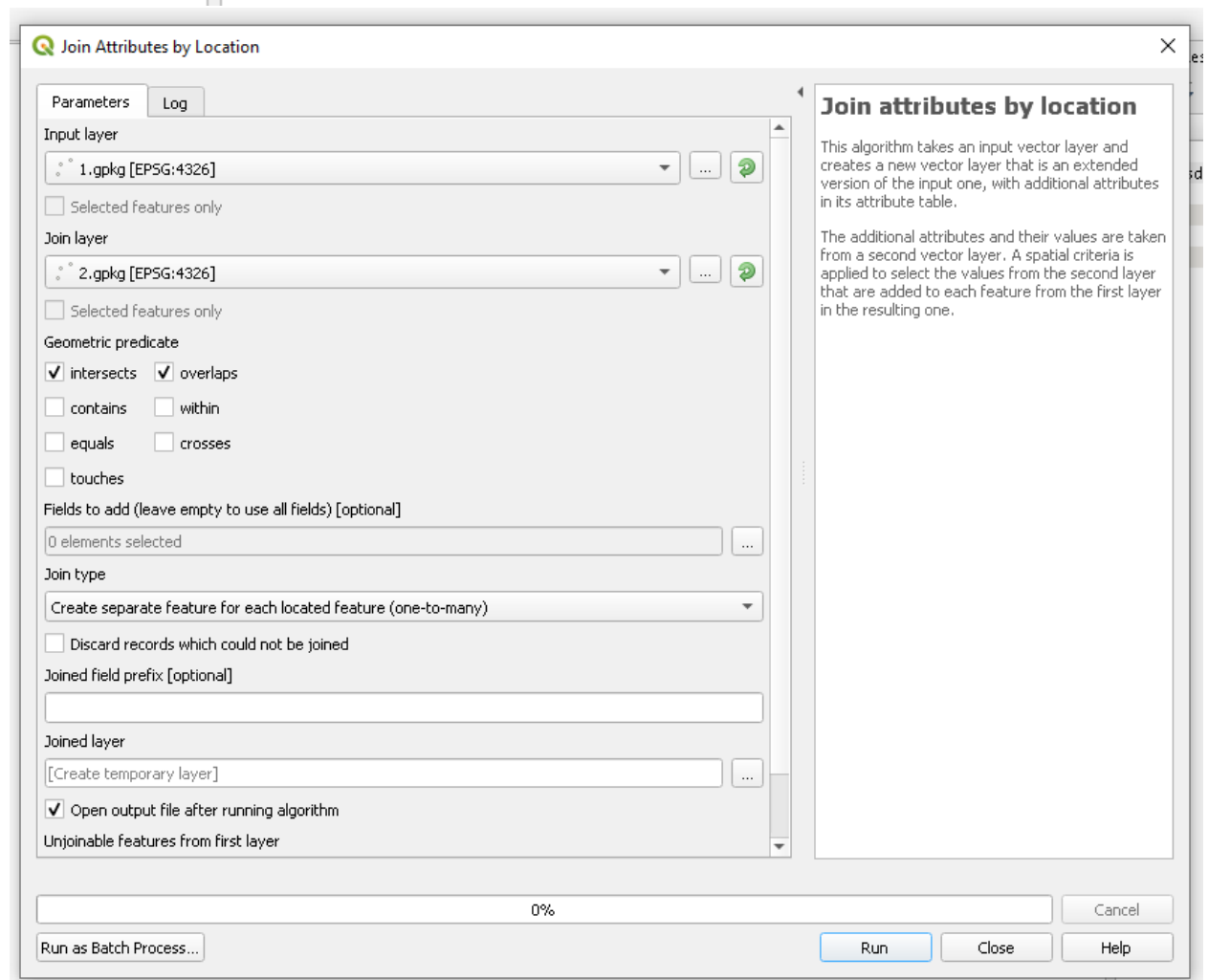
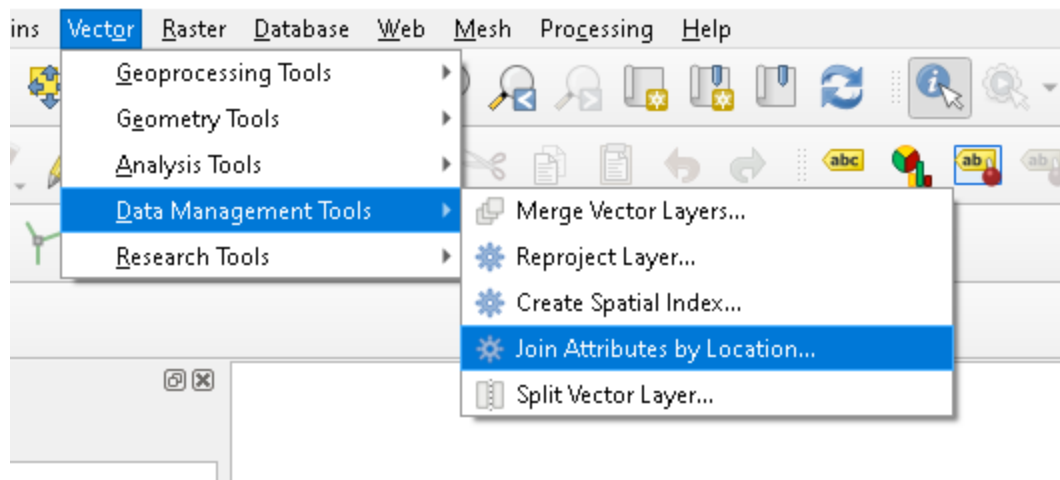


4. Use the “Point sampling tool” plugin to extract the raster attributes to points creating a vector dataset.



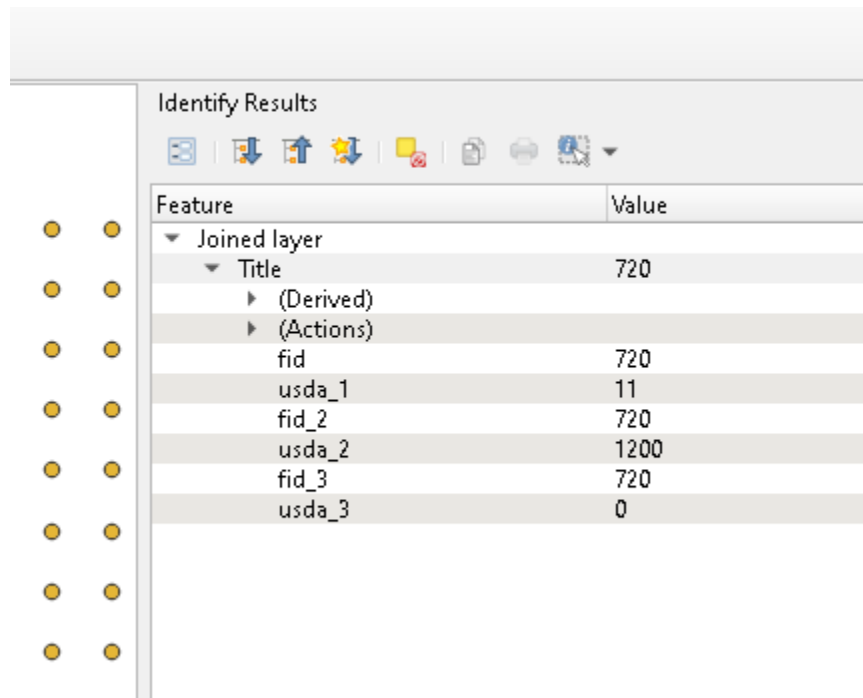
Repeat the same process for each of the 3 bands to get 3 vector datasets.

5. Join the 3 vector datasets one by one using the Vector/Data Management Tools/Join Attributes by Location option in QGIS



Repeat the step to join all the 3 vector datasets to get a point vector dataset with 3 attributes

for each of the 3 bands as shown below



Feature	Value
▼ Joined layer	
▼ Title	720
▶ (Derived)	
▶ (Actions)	
fid	720
usda_1	11
fid_2	720
usda_2	1200
fid_3	720
usda_3	0

In this example, “usda_1” contains LC, “usda_2” contains CT and “usda_3” contains IRR values.

6. Extract 640 x 640 m patches from the raster file where each sample point is used as a centroid. Contact the moderator if more help is required on this.
7. Harmonize the point vector dataset using the following [document](#).
8. Create a zip folder including the shape file and extracted patches. The zip file should be uploaded to the WorldCereal system.