**ACI Pixel Reallocation Workflow Report**

**Overview**

The purpose of the ACI reallocation process was to adjust the Annual Crop Inventory (ACI) dataset to better align with the Manitoba Agricultural Services Corporation (MASC) ground-truth acreage values on a per–Rural Municipality (RM) and crop basis. This ensures that downstream yield raster generation reflects realistic production patterns and addresses known classification issues in the ACI data, without forcing provincial totals to match MASC.  
  
**Input Data**

- masc\_imputed\_<year>.csv: MASC-imputed ground truth (rm, crop, acres, yield metrics).  
- aci\_summary\_<year>.csv: Output from aci\_prepare.py (RM × Label) summarizing ACI raster by pixel\_count, hectares, acres.  
- crop\_label\_lut.csv: Lookup mapping MASC “crop” names to ACI “Label” strings.  
- muni\_rm\_lut.csv: Lookup mapping municipality names in ACI polygons (“MUNI\_NAME”) to the RM names used in MASC (“Risk Area / R.M.”).  
All data follows the project structure under data/reference and data/interim/<year>/, with outputs in outputs/reports/.  
  
**Workflow**

Step A: Build ACI RM × Label summaries  
- We created aci\_prepare.py to read the ACI raster (aci\_<year>\_mb\_v\*.tif), perform categorical zonal statistics against municipalities.geojson, and output aci\_summary\_<year>.csv.  
- Key settings: all\_touched=False so each pixel is counted toward exactly one polygon; agricultural labels only (label\_set) to exclude non-ag land covers.  
  
Step B: Load and align tabular sources  
- aci\_reallocate\_pixels.py loads aci\_summary\_<year>.csv, masc\_imputed\_<year>.csv, crop\_label\_lut.csv, and muni\_rm\_lut.csv.  
- ACI table is merged with muni\_rm\_lut.csv on MUNI\_NAME to add the column rm (Risk Area / R.M.).  
- MASC table is merged with crop\_label\_lut.csv to map MASC crop → ACI Label.  
- Both are aggregated to RM × Label to form a unified comparison table with acres\_aci, acres\_masc, and acres\_diff.  
  
Step C: Identify correction cases per RM  
- Case 1: acres\_aci > 0 and acres\_masc = 0 → ACI-only crops (false positives). Set these ACI acreages to 0 within that RM.  
- Case 2: acres\_aci = 0 and acres\_masc > 0 → Missing crops (false negatives). Introduce these by reallocating acres from crops with the largest positive surplus (acres\_aci − acres\_masc) within the same RM, conserving total acres inside the RM.  
  
Step D: Reallocation and outputs  
- Reallocation runs independently for each RM, never creating or destroying acres globally, only moving them between labels when needed for Case 2, and zeroing them in Case 1.  
- The corrected table is saved to data/interim/<year>/aci\_summary\_reallocated\_<year>.csv.  
- A run log is appended to outputs/reports/aci\_reallocation\_stats.csv with: year, aci\_total\_before, masc\_total, diff\_before, aci\_total\_after, diff\_after, case1\_removed, case2\_added.  
- A before/after change report is written to outputs/reports/aci\_reallocation\_changes\_<year>\_detail.csv and \_summary.csv highlighting only rows that changed and summary counts by RM.  
  
**Problems Discovered and Fixes**

*Problem 1: Counting non-ag classes inflated totals*  
- Symptom: provincial totals exceeded expected cropland by a large margin.  
- Fix: filter aci\_summary to agricultural labels only (Agriculture undifferentiated, pastures/forages, cereals, oilseeds, pulses, vegetables, fruits, corn, etc.).  
  
*Problem 2: Boundary counting method*  
- Clarification: all\_touched=True would double-count boundary pixels. Requirement was one-to-one assignment. We set all\_touched=False to ensure unique pixel assignment; adjacent RMs jointly cover edges without gaps.  
  
*Problem 3: Inconsistent RM naming in muni\_rm\_lut*  
- Symptom: duplicate RM groups such as CYPRESS-LANGFORD and NORTH CYPRESS-LANGFORD both appearing in results; also “ST.ANDREWS” vs “ST. ANDREWS”, and “STE.ANNE” vs “STE. ANNE” variants.  
- Root cause: internal inconsistencies inside muni\_rm\_lut.csv and a specific mis-map for MUNICIPALITY OF GLENBORO-SOUTH CYPRESS.  
- Fix: corrected muni\_rm\_lut.csv to map all intended municipalities to the correct RM names and harmonized internal forms (keeping your chosen capitalization/punctuation scheme). After fixes, spurious duplicate RM entries disappeared and case1\_removed counts dropped slightly across years.  
  
*Problem 4: String normalization side effects (removed)*  
- Early preprocessing applied lowercasing/stripping to keys. This hurt transparency and masked where mismatches occurred. We removed all “standardization” in aci\_reallocate\_pixels.py so names stay exactly as authored in the reference files.  
  
*Problem 5: Expectation about provincial totals*  
- Clarification: this reallocation process is not intended to make ACI provincial totals equal MASC. It corrects two logical cases at the RM × crop level while conserving total ACI acres within each RM. Residual provincial differences persist and reflect structural ACI bias, as documented in prior comparisons

*Problem 6: HILLSBURG–ROBLIN–SHELL RIVER used although outdated*  
During validation of the ACI–MASC reallocation workflow, a consistent set of unmatched records was traced to the name “HILLSBURG–ROBLIN–SHELL RIVER.” This municipality name does not appear in the post-2015 Manitoba municipal structure used in the ACI shapefile (municipalities.geojson) or in the current MASC RM list.

The discrepancy arises because the Municipality of Hillsburg–Roblin–Shell River was officially renamed to the Municipality of Roblin on March 30, 2015, following provincial amalgamation. Consequently, all ACI data reference the updated MUNICIPALITY OF ROBLIN name, while the MASC dataset still contains the legacy name in some records.

*Problem 7: UNORG TERRITORY has no ACI map*

A small number of records in the MASC dataset reference UNORG TERRITORY or UNORG TERRITORY 1/2, corresponding to lands in northern Manitoba that are not part of any organized municipality. These areas have no equivalent geometry in the ACI municipalities.geojson layer, which only includes incorporated municipalities and local government districts.

As a result, MASC data for unorganized territories remain unmatched in the reallocation process. The records are retained in the MASC dataset but excluded from pixel-based reallocation and subsequent raster generation steps.

**Results Snapshot: 2017–2024**

(example metrics recorded in outputs/reports/aci\_reallocation\_stats.csv**)**  
- ACI totals before reallocation exceeded MASC (diff\_before) each year.  
- After reallocation, aci\_total\_after decreased modestly but remained above MASC; diff\_after remained positive.  
- Case 1 removals: roughly 295–501 per year.  
- Case 2 additions: roughly 141–207 per year.  
These outcomes are consistent with expectations: large overrepresented crops donate acres to missing minority crops; totals are conserved per RM, not forced to equal MASC.  
  
**Files Written by the Workflow**

- data/interim/<year>/aci\_summary\_<year>.csv  
- data/interim/<year>/aci\_summary\_reallocated\_<year>.csv  
- outputs/reports/aci\_reallocation\_stats.csv  
- outputs/reports/aci\_reallocation\_changes\_<year>\_detail.csv  
- outputs/reports/aci\_reallocation\_changes\_<year>\_summary.csv  
  
**Conclusion**

The ACI pixel reallocation workflow produces a more complete RM × crop composition while preserving the spatial integrity and total acres within each RM. It also documents the issues encountered and corrections applied, providing a reproducible foundation for the next stage: computing yield-at-pixel without writing intermediate reallocated rasters.