# ACI Yield per Pixel Computation and Normalization

Script: aci\_yield\_per\_pixel.py

## Purpose

This stage converts municipality- and RM-level acreage and yield information into per-pixel yield estimates that can be rasterized into annual biomass surfaces.

## Methodology

The script merges reallocated ACI acreage data (label\_area\_deltas\_<year>.csv) with MASC-derived yield data (aci\_masc\_merged\_<year>.csv). For each RM × Label, ACI acreage is distributed across all contributing MUNI\_NAMEs in proportion to their share of total ACI pixels (rm\_label\_pct). This produces a consistent acreage and pixel count for every MUNI\_NAME × Label combination.

Because some crop Labels have no direct ground-truth yield per acre (for example, rare or seed crops absent from MASC records in certain years), the script imputes their yields using the median yield of all valid, non-zero records for that same Label. Using the median rather than the mean avoids bias from extreme values or unusually high-yield RMs. This ensures that all Labels carry plausible, data-driven yield estimates without artificially inflating provincial totals. A global median acts as a fallback if a Label has no valid entries at all.

After imputation, total yield per record (masc\_yield\_tonnes\_total) and per-pixel yield (aci\_yield\_tonnes\_per\_pixel) are computed. The provincial sum is then normalized to the official MASC ground-truth total (masc\_summary.csv) through a single scalar factor. This guarantees that the sum of modeled yields equals the verified total for each year while preserving within-province spatial proportions.

## Outcome

The output file data/processed/<year>/aci\_yield\_per\_pixel\_<year>.csv contains:  
- Corrected ACI and MASC acreage values  
- Median-imputed yields for missing crops  
- Ground-truth–normalized total and per-pixel yields (gt\_ fields)

Across all years (2017–2024), normalization factors remained within ±4 %, confirming internal consistency and statistical defensibility. The resulting dataset is fully aligned with provincial ground-truth totals and ready for pixel-based biomass raster generation.