# ACI Biomass per Pixel Computation and Normalization

Script: aci\_biomass\_per\_pixel.py

## Purpose

This stage converts municipality- and RM-level acreage, yield, and biomass information into per-pixel biomass estimates that can be rasterized into annual biomass yield surfaces. It extends the earlier yield-based process by integrating residue-to-product (RPR) and sustainable-availability (SAF) factors to derive ground-truth biomass directly from MASC-imputed crop data.

## Methodology

The script merges reallocated ACI acreage data (label\_area\_deltas\_<year>.csv) with MASC-derived yield and biomass 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), producing consistent acreage and pixel counts for every MUNI\_NAME × Label combination.

To establish a defensible provincial reference, “ground-truth biomass” is calculated directly from the MASC-imputed crop dataset (masc\_imputed\_<year>.csv). Each crop record is merged with its residue and sustainability factors from rpr\_saf\_masc\_crop.csv, and total biomass is computed as:

* biomass\_tonnes = yield × RPR × SAF

This sum across all crops defines the provincial ground-truth biomass total for the year—analogous to how MASC yield totals define ground-truth yield.

Because certain crop Labels have missing or zero biomass values (for example, minor forage or seed crops not present in MASC records), the script imputes biomass using the median of all valid, non-zero biomass values within the same Label. The median minimizes distortion from atypical high-biomass outliers, ensuring that each Label carries a realistic, data-driven biomass estimate. A global fallback median is applied if no valid records exist for a given Label.

After imputation, the total and per-pixel biomass values are computed (masc\_biomass\_tonnes\_total, aci\_biomass\_tonnes\_per\_pixel). Provincial totals are normalized so that:

* Σ gt\_masc\_biomass\_tonnes\_total = biomass ground truth
* Σ gt\_masc\_yield\_tonnes\_total = MASC yield ground truth

This dual normalization guarantees that modeled yield and biomass remain internally consistent and fully reconciled with empirical totals derived from MASC data.

## Outcome

The output file data/processed/<year>/aci\_biomass\_per\_pixel\_<year>.csv contains:

- Corrected ACI and MASC acreage and pixel counts  
- Median-imputed yield and biomass values for all Labels  
- Ground-truth–normalized total and per-pixel yield and biomass (gt\_ fields)

Across all years (2017–2024), yield normalization factors remained within ±4 %, while biomass factors ranged from approximately 0.93 – 0.96, confirming high consistency between modeled and ground-truth totals. The resulting dataset provides a complete, spatially resolved, and fully defensible basis for rasterized biomass yield mapping.

A log of all instances of this script being run can be found under outputs/reports/biomass\_normalization\_log.csv and outlines the factors. This log, along with aci\_masc\_merged\_<year>.csv document calculated biomass and how changes in the factors of RPR and SAF alter biomass totals.