StatEO 2026
5–7 May 2026
ESA-ESRIN, Frascati

# Optimizing Crowdsourced Training Samples for Large-Scale Crop Mapping

Qingying Wu[1,2], Qiangyi Yu[1], Hendrik Boogaard[2], Arun Pratihast[2]

[1] The Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, China;

[2] Wageningen Environmental Research, Wageningen University and Research, Wageningen, The Netherlands

wuqingying@caas.cn

## Background

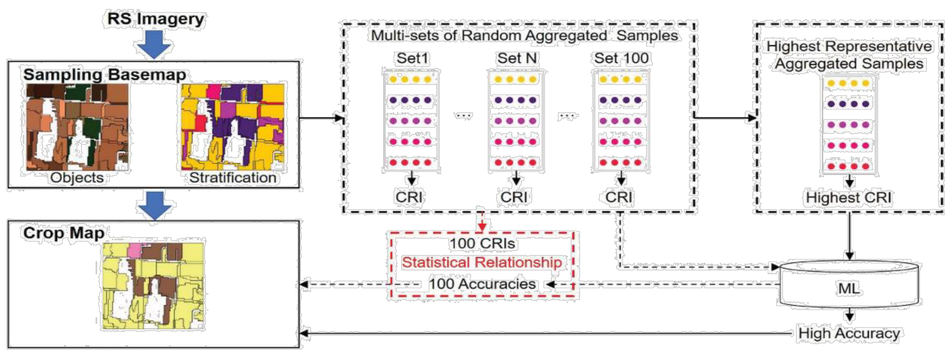
With advances in machine learning, crop mapping performance is now constrained more by training sample quality than by algorithmic complexity.

- **Collecting new samples.** Fieldwork is costly and irreversible, requiring sampling guidance to determine optimal sample size and distribution.
- **Reusing existing samples.** Open repositories aggregate millions of samples, yet label and temporal inconsistencies undermine their fitness, requiring systematic quality evaluation.

## Method

### RESSOS: Representativeness Evaluation of Sample Sets with Object Stratification

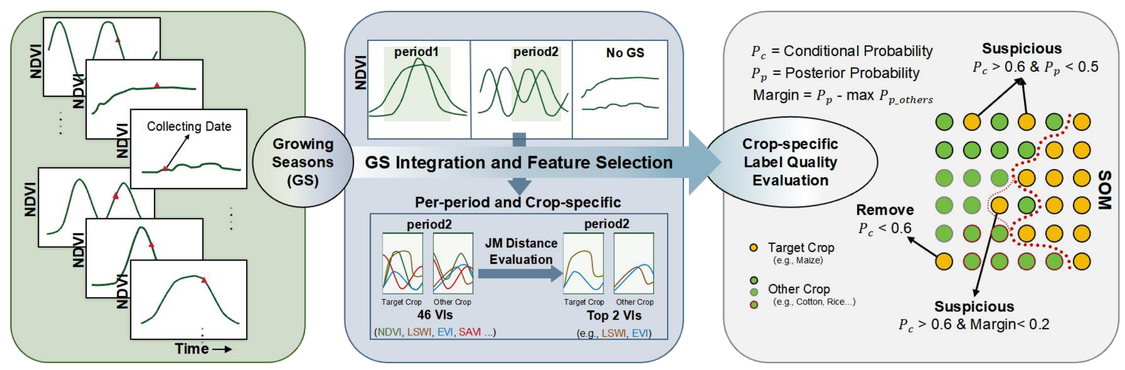
RESSOS introduces the Comprehensive Representativeness Indicator (CRI) to guide representative sample collection. It generates object-based stratifications through unsupervised clustering, produces multiple random aggregated sample sets, and selects the set with the highest CRI as the RESSOS samples.




**Region:** Shandong, China、111 km² 、Double cropping

### TISQE: Temporal-Integrated Sample Quality Evaluation

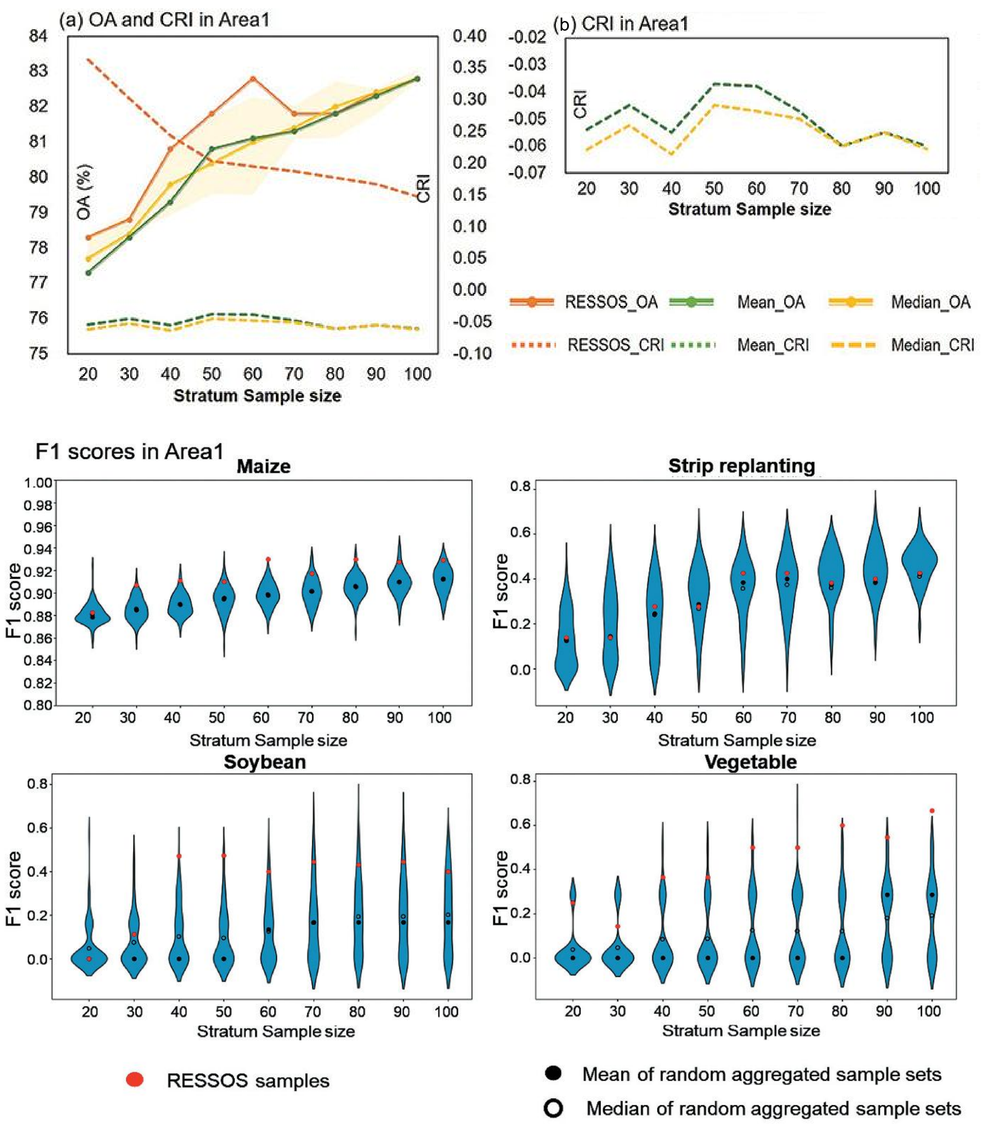
TISQE integrates temporal information into label evaluation. It identifies each sample's growing season, selects discriminative vegetation indices per crop per season, and applies a margin-aware binary self-organizing map (SOM) to classify samples as Keep, Suspect, Remove, or DataGap.



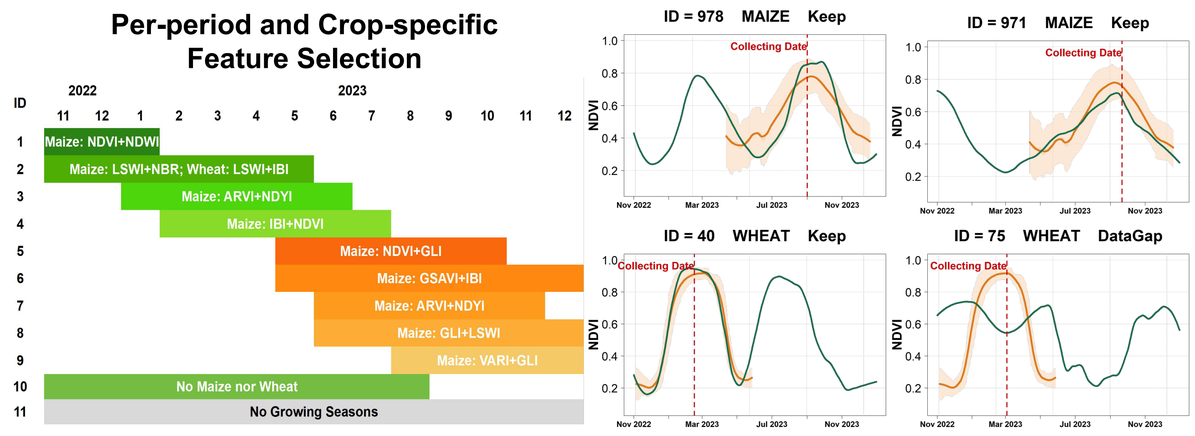

**Dateset:** WorldCereal CIMMYT Mexico 2023、1,546 samples (maize 902, wheat 402)、Confidence Score of 62.9, spatial evaluation only

## Result

- A positive correlation was found between CRI and overall accuracy (OA).
- RESSOS samples consistently achieved the highest OA and F1 scores below an identified sample size threshold, with the largest advantage for minor crop types.




- TISQE identified 11 growing seasons within the dataset and classified the 1,546 samples into Keep (58.8%), Remove (16.6%), Suspect (8.9%), and DataGap (10.9%) categories.
- With TISQE category weights and growing-season feature weights, the overall accuracy increased from 91.64% to 97.45%, with the largest gains for wheat.




### TISQE Quality Categories for Maize and Wheat

| | Keep | Remove | Suspect | DataGap | NA |
|---|---|---|---|---|---|
| MAIZE | 66.7% | 17.8% | 8.1% | 7.0% | 0.3% |
| WHEAT | 66.2% | 2.7% | 12.7% | 17.2% | 1.2% |

### Confusion Matrix and Accuracy Comparison

| | Raw | | | TISQE | | | Accuracy Δ |
|---|---|---|---|---|---|---|---|
| | MAIZE | WHEAT | F1 | MAIZE | WHEAT | F1 | |
| MAIZE | 172 | 5 | 93.73% | 185 | 2 | 98.14% | +4.41% |
| WHEAT | 18 | 80 | 87.43% | 5 | 83 | 95.95% | +8.52% |
| OA | 91.64% | | | 97.45% | | | +5.81% |

## Conclusion

RESSOS and TISQE form a complementary pipeline for high-quality training samples in crop mapping, covering both new sampling and existing crowdsourced collections. Sample quality, not algorithm complexity, is the next frontier for large-scale crop mapping.

Wu, Q., et al. (2025). "Enhancing machine learning-based crop mapping with high-quality training samples." Geo-spatial Information Science: 1–24.