

Neural Network for Acreage Estimation Using Precision Agriculture and Survey Data

Sean Rhodes¹, Luca Sartore^{1,2}, Jake Abernethy¹, Chad Garber¹, Darcy Miller¹, Denise A. Abreu¹

¹ National Agricultural Statistics Service, 1400 Independence Ave SW 20250 Washington D.C., USA

² National Institute of Statistical Sciences, P.O. Box 33762, Washington D.C., USA 20033

Background

Historically, the United States (US) Department of Agriculture's National Agricultural Statistics Service (NASS) conducts the June Area Survey (JAS) based on an area frame, which has complete coverage of all land in the contiguous US. Response rates have been declining in many federal surveys, including the JAS, leading to increased item and unit nonresponse. NASS has explored utilizing automatic imputation for the JAS using Artificial Intelligence (AI) models. Previous research has found that NASS's Predictive Cropland Data Layer (PCDL) and Crop Sequence Boundaries (CSBs) have good predictive power for corn and soybeans (two major US crop commodities). NASS also recognizes the benefits of for imputation purposes of machine-logged data (MLD) collected onboard farm machinery, such as planters. This poster presents the structure of customized neural network architectures utilized to leverage MLD, administrative data, PCDL, CSBs, and survey data to estimate planted acreage, and its uncertainty at the field level.

Methods

Over 7,000 JAS tracts, each representing a unique farm operation, were used to organize and design the customized neural network developed for acreage estimation and uncertainty quantification. To estimate imputed acreage values, define:

- A_i : Ground-truth Farm Service Agency (FSA) acreage of a specific commodity planted within digitized tract i
- W_i : Total size of JAS digitized tract i expressed in acres
- \hat{A}_{iC} : Remote-sensing-derived acreage estimate (from CSB, PCDL, MLD) within tract i for crop C , where the crop categories are Corn, Soybeans, and Other. The remote-sensing proportions satisfy:

$$W_i = \hat{A}_{i,Corn} + \hat{A}_{i,Soybeans} + \hat{A}_{i,Other}$$

Response vectors were normalized by the total tract size and are assumed to follow a Dirichlet composition distribution that accommodates the presence of zero components. (Tang 2022).

Let $X_i = (x_{i1}, \dots, x_{im})^T$ denote an m -dimensional compositional random vector. The distribution of X_i is defined through

$$X_i \stackrel{d}{=} \frac{Z \circ Y}{Z^T Y}$$

where Z is a vector of Bernoulli indicators for structural zeros, Y follows a Dirichlet distribution, and \circ denotes the Hadamard (elementwise) product.

The probability density function of X_i is modeled as

$$f(x_i) = \frac{\prod_{j=1}^m p_{ij}^{1-z_{ij}} (1-p_{ij})^{z_{ij}}}{1-p_{i1} \cdots p_{im}} \left[\frac{\Gamma(\alpha_{i\mathbb{J}}^*)}{\prod_{j \in \mathbb{J}} \Gamma(\alpha_{ij})} \prod_{j \in \mathbb{J}} x_{ij}^{\alpha_{ij}-1} \right]$$

where:

- p_{ij} : Probability that component j of tract i is structurally zero
- z_{ij} : Zero-indicator for component j
- \mathbb{J} : Set of indices corresponding to non-zero components
- α_{ij} : Dirichlet parameters with $\alpha_{i\mathbb{J}}^* = \sum_{j \in \mathbb{J}} \alpha_{ij}$

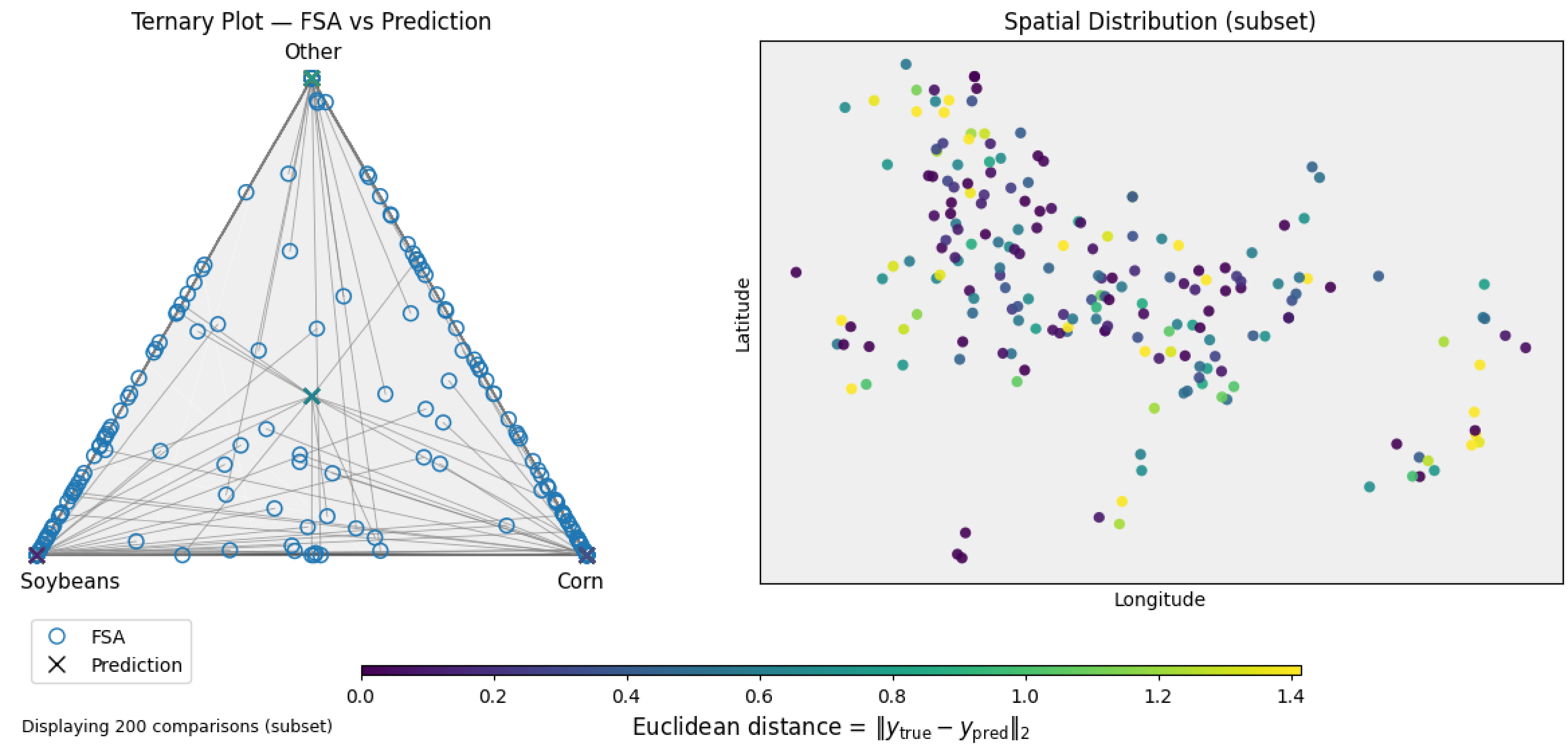
Using a Dirichlet composition distribution allows acreage compositions to include zeros while maintaining compositional validity and permitting tract-level standard-error estimation.



References

- Hunt, K., Abernethy, J., Beeson, P., Bowman, M., Wallander, S., Williams, R. (2023). Crop Sequence Boundaries (CSB): Delineated Fields Using Remotely Sensed Crop Rotations. In the Proceedings of the IX International Conference of Agricultural Statistics.
- Rhodes, S., Rosales, A., Sartore, L., & Murphy, T. (2023). Uncertainty Assessment for the Imputation of an Area Survey. In Proceedings of the Joint Statistical Meeting. <https://doi.org/10.5281/zenodo.8383528>
- Sartore, L., Boryan, C., Dau, A., & Willis, P. (2023). An Assessment of Crop-Specific Land Cover Predictions Using High-Order Markov Chains and Deep Neural Networks. Journal of Data Science, 21(2).
- US Department of Agriculture, Farm Service Agency. (2022) Common Land Units (CLUs).
- Tang, Man-Lai, et al. "Dirichlet Composition Distribution for Compositional Data with Zero Components: An Application to Fluorescence In Situ Hybridization (FISH) Detection of Chromosome." Biometrical Journal, vol. 64, 2022, pp. 714-732. Wiley Online Library, <https://doi.org/10.1002/bimj.202000334>.

Preliminary Testing Results



Epoch	Training Loss	Validation Loss
46	5.826	5.304
47	5.826	5.304
48	5.826	5.304
49	5.826	5.304
50	5.826	5.304

Crop	Mean Absolute Error	Total Absolute Error	Testing Loss
Soybeans	0.264	877.731	5.079
Corn	0.289	961.427	
Other	0.245	814.749	

Discussion

Focusing on the technical structure and workflows, the study provides an initial demonstration of how machine-logged data (MLD), remote sensing (PCDL/CSB), administrative records, and survey responses can be fused in a neural network framework to impute field-level acreage compositions. Training on historical reported survey responses and remote sensing data has provided preliminary network weights and structure needed to incorporate MLD into the network for acreage prediction. A major limitation is the small amount of available MLD, which restricts generalizability. Despite these limitations, having the structure of the network sets a foundation for continued development by testing on MLD.

Disclaimer

The findings and conclusions in this poster are those of the authors and should not be construed to represent any official USDA, US Government, or NISS determination or policy.

