A. Model details and baseline

SVR is a regularized, data-efficient method that works well in high-dimensional settings [8, 33]. The RBF kernel is flexible and broadly competitive [9, 29, 34], with mature implementations available [31]. Together, these make RBF-SVR a simple, strong choice here without further ablations.

Hyperparameters were tuned once via grid search on training years and then fixed for all LOYO evaluations. The selected configuration is $C{\approx}52$, $\gamma{\approx}0.0070$ and $\epsilon{\approx}0.99$. Generalization across time is evaluated with LOYO over 2017-2024. We report per-year R^2, Root Mean Square Error (RMSE), and averaged metrics over all year leave-out predictions as overall indicator of model performance for comparison with literature reports.

As a compact phenology comparator, we implemented a classical baseline, including: MODIS NDVI seasonal summaries over DOY 100–300 (mean, max, integral, and DOY of max) from MOD13Q1 (250 m, 16-day), PRISM monthly precipitation (12 bands, mm), and growing/cooling degree-day totals (GDD base 10 °C capped at 30 °C; CDD base 18 °C) using PRISM daily aggregates. All layers are crop-masked and averaged to county following the same LOYO protocol as AEF, we train Linear, Ridge, Random Forest, Gradient Boosting, and RBF-SVR; per-year performance is reported for the best baseline model. In addition, we include county-mean baseline: predict each test county's yield by the training-years mean yield of that county. Counties lacking a training mean in a given fold are omitted from that fold's test set.

B. Soybean Supporting Results

This appendix provides diagnostics for soybean that parallel the corn analysis while avoiding repetition in the main text. We report spatial error structure, local clustering, unsupervised variance geometry, and supervised feature usage, complementing the main soybean metrics in Table 2.

B.1. Spatial Pattern of Errors

Figure 9a maps county-level relative errors (%) for each LOYO split. As in corn, errors are geographically coherent rather than speckled, with contiguous bands of over- and under-prediction that track agro-ecological gradients. Year-level mean bias is small in most years; 2023 shows a larger positive uptick ($+6.6\%$). Across 2017–2024 the average percent RMSE is 9.34% (Table 2), indicating modest, agronomically reasonable uncertainty at county scale.

B.2. Global and Local Spatial Autocorrelation

Soybean residuals are more clustered than corn on average. Using the same $k{=}8$ nearest-neighbor weights, global Moran's I ranges from 0.2639 (2020) to 0.5807 (2017), with mean $\bar{I}{\approx}0.464$ and $p_{\text{perm}}{=}0.001$ in all years (Table 5). Figure 9b shows LISA clusters (HH/LL/HL/LH) at $\alpha{=}0.05$: repeated HH zones appear along parts of the northern tier in several years, while LL patches recur in portions of the central Corn Belt, with HL/LH edges at their boundaries. The consistency of these patterns across independent LOYO folds suggests missing slow-varying drivers (e.g., rotation/management regimes or soils) rather than idiosyncratic noise.

B.3. Unsupervised Embedding Geometry vs. Error

The soybean embeddings exhibit a moderately concentrated variance structure: PC1 explains $\sim 23.3\%$ of variance; PCs 1–3 explain $\sim 54.6\%$; and 90% is reached by PC 12 (Fig. 7a, left). However, these dominant directions explain little of the supervised miss. Predicting absolute error from the first k PCs yields low test R^2 throughout (maximum $\sim 4.05\%$ at $k{=}24$; Fig. 7a, right). Thus, the geometry relevant for generalization lives in lower-variance subspaces rather than along variance-dominant axes, mirroring the corn finding.

B.4. Test-Fold Feature Importance and Stability

Permutation importance computed *only* on held-out data (20 shuffles per feature per fold) reveals a concentrated but not single-feature-dominated profile (Fig. 7). Three of the most influential directions (A49, A17, A28) overlap with corn, with A49 ranking first ($\Delta R^2{\approx}7.2\%$) and A17 second ($\sim 5.6\%$). Soybean elevates additional axes (A51, A03, A41, A05) that were not top-ranked for corn. As a set, the top-12 features account for $\sim 45\%$ of the summed test-fold importance, indicating that a relatively small subset of the 64 channels carries most of the out-of-sample leverage. Year-wise stability heatmaps (Fig. 7) show A49 informative in most splits with low dispersion; A17 is especially strong from 2018–2020; and A42 tends to be weak early and more informative later.

Figure 8 reports the Pearson correlation between each embedding dimension and the *out-of-sample* SVR predictions pooled across folds (sign only; not causal). Signs largely agree with corn for shared leaders (A49 positive; A17 and A28 negative), while A24 flips sign (negative in corn, positive in soybean), suggesting crop-specific semantics along that axis.

Summary. (i) Soybean achieves strong accuracy with the same embedding-only pipeline; (ii) residuals are geographically organized and *more* clustered than corn, pointing to slow-varying regional factors; (iii) dominant PCA directions do not explain error; and (iv) a compact subset of AEF channels repeatedly drives generalization, with a shared core across crops and a few crop-specific axes.

To quantify label efficiency, we held 2021 as the test year and randomly subsampled the training set to $\{1, 5, 10, 20, 30, 60, 90\}\%$ of available counties (across the remaining

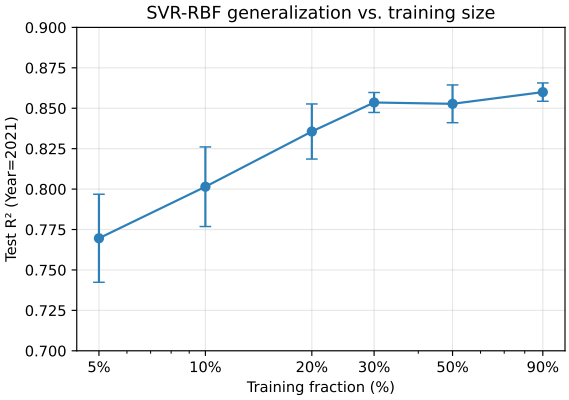


Figure 6. Corn, test year 2021: test R^2 as a function of training fraction.

years), repeating each fraction with 10 independent random draws. We report mean R^2 and its standard deviation over subsamples.

Performance rises sharply with a small number of labeled counties and then saturates (Fig. 6): at 5% of training data ($\sim$278 county–years) we already obtain R^2=0.762 $\pm$ 0.022; by 20% ($\sim$1,111) R^2 reaches 0.835 ± 0.010. Using only 30% of the training labels achieves R^2=0.854. The dispersion across random subsamples shrinks with sample size (Std drops from 0.0272 at 1% to 0.0057 at 90%), indicating stable generalization once a modest label budget is available. These results imply that the annual AEF embedding already carries most of the task-relevant variation, and the supervised layer mainly provides a light calibration.

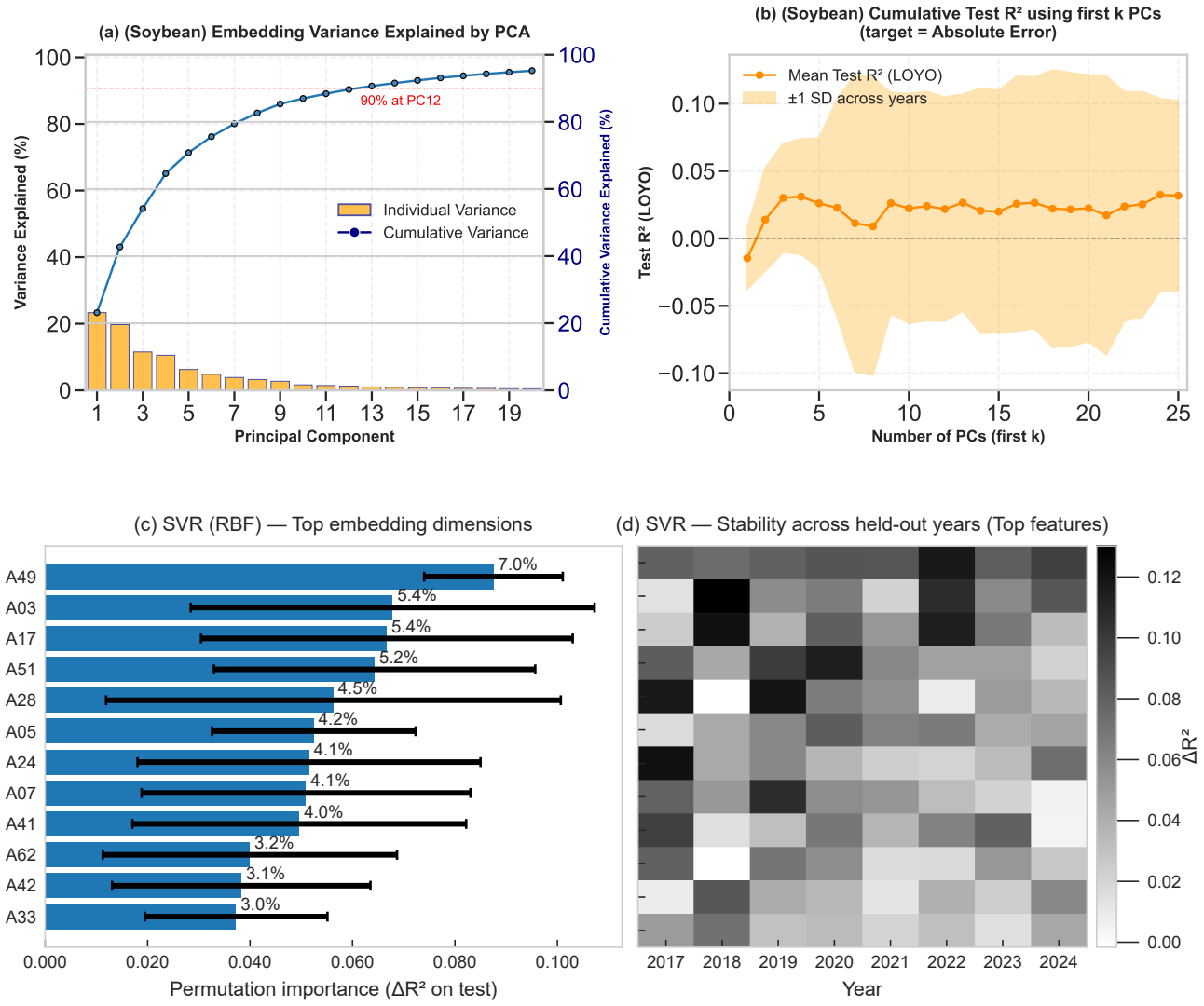


Figure 7. Soy: (a) Scree: variance explained by PCs of AEF embeddings (train-fold PCA, test projections). (b) Test R^2 for predicting absolute error from first k PCs—low throughout. (c) Top embedding dimensions by test-fold permutation importance (ΔR^2); error bars show inter-year SD. (d) Year-wise stability heatmap for the same features.

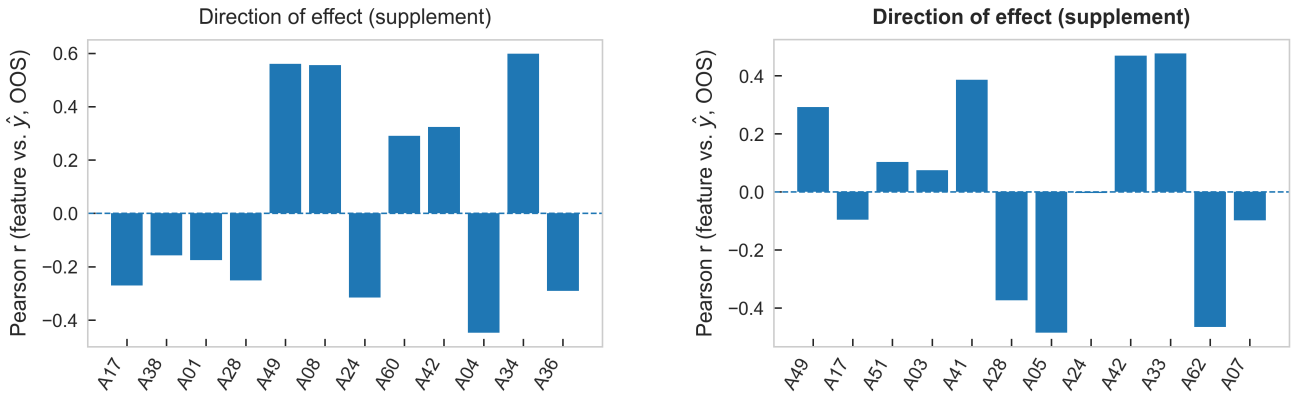
Table 5. Global Moran's I of county-level relative error under LOYO (county k=8 nearest neighbors; 999 permutations).

(a) Corn

Year	**N**	I	$E[I]$	z	$p_{\textbf{perm}}$
2017	836	0.4361	−0.0012	26.26	0.001
2018	737	0.4585	−0.0014	26.02	0.001
2019	681	0.3231	−0.0015	17.66	0.001
2020	932	0.3300	−0.0011	16.20	0.001
2021	794	0.4908	−0.0013	28.77	0.001
2022	859	0.3813	−0.0012	23.25	0.001
2023	811	0.3560	−0.0012	21.12	0.001
2024	700	0.3171	−0.0014	17.60	0.001

(b) Soybean

Year	**N**	I	$E[I]$	z	$p_{\textbf{perm}}$
2017	803	0.5807	−0.0012	25.56	0.001
2018	718	0.4852	−0.0014	27.12	0.001
2019	640	0.4232	−0.0016	22.44	0.001
2020	882	0.2639	−0.0011	16.32	0.001
2021	754	0.5755	−0.0013	32.96	0.001
2022	829	0.3994	−0.0012	23.93	0.001
2023	758	0.5753	−0.0013	33.04	0.001
2024	658	0.4053	−0.0015	21.90	0.001



(a) Corn

(b) Soybean

Figure 8. Direction of effect: Pearson correlation (r) between each AEF feature and out-of-sample SVR predictions, pooled across folds.

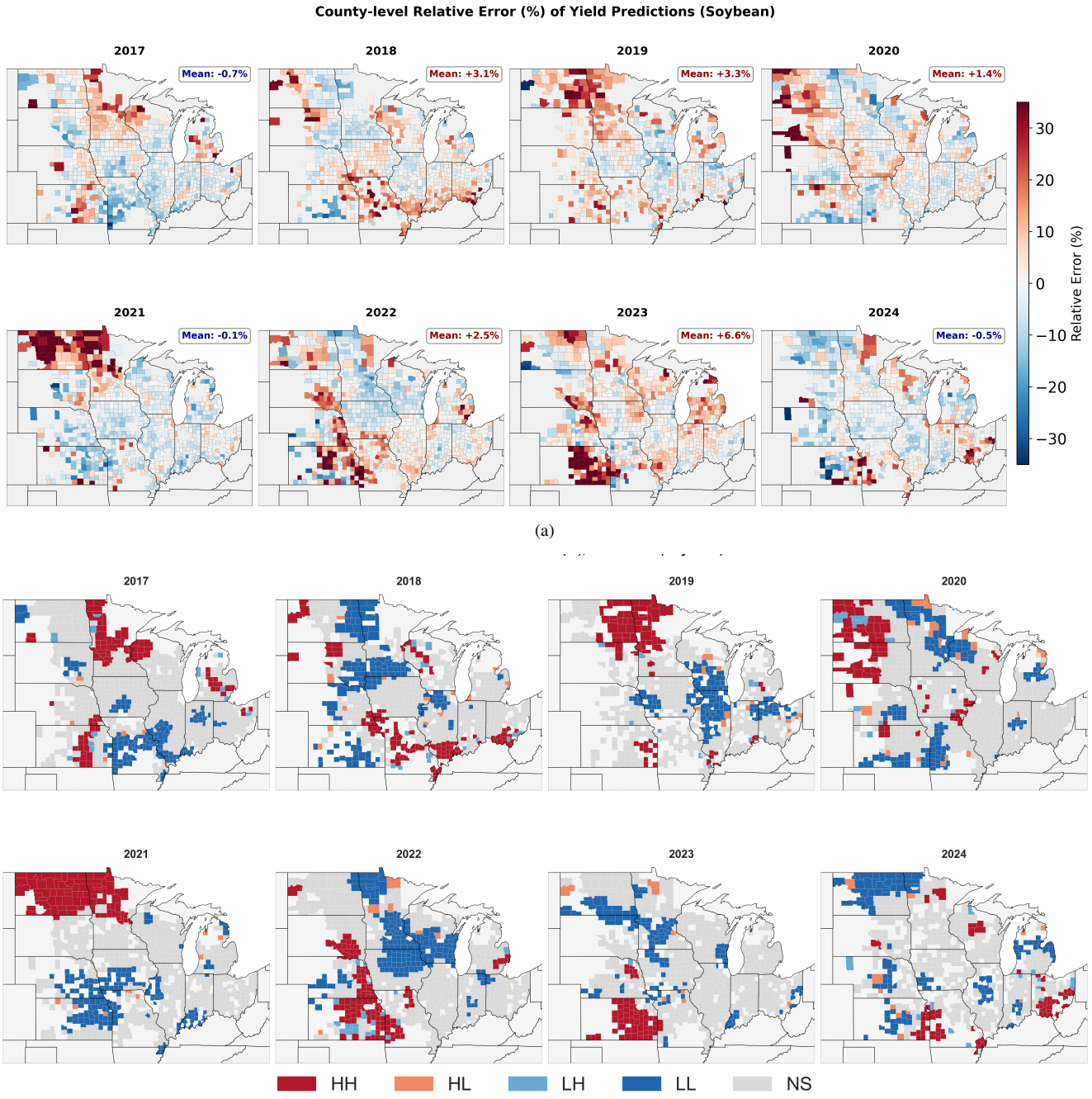


Figure 9. Soybean: (a) County-level relative error (%) under LOYO. Blue = under-prediction; red = over-prediction. (b) LISA clusters (HH, HL, LH, LL) of relative error (%); $\alpha = 0.05$; $k = 8$ nearest neighbors.