Appendix B: 2019 sex ratio imputation

Data set and imputation procedure overview

The experiment design was randomized complete block design with four blocks, nine levels of main plot (Crop ID) and two levels of split-plot (Corn weed management) effects. There were 72 experimental units (eu), 8 observational units (quadrats) per eu. Within each quadrat, six cohorts of plants were sexed. By the time that the plants were sexed, a few plants' sexes were not observable, and thus, they were marked as Unknown in the data set. This caused Total > Male + Female in some eu's.

Complete case analysis, in which missing observations are removed from the data set, is acceptable when less than 5% of the data is missing at random because removing the incomplete cases does not significantly change the outcome (**?**). The 2019 sex ratio is imputed because 22% of the data (755 `NA`s out of 3456 entries (6 cohorts/quadrat x 8 quadrats/eu x 72 eu)) was missing at random (MAR). We considered our data MAR even though the number of missing values were higher in soybean plots than in the other three crops because 1) the quadrats were placed randomly in the field before weed emergence, 2) waterhemp and other weed seedlings were present in the quadrats at the beginning of the season, and 3) waterhemp and other weed species were found in soybean plots by the time we sexed the waterhemp plants, but not in the eight fixed quadrats.

The MICE imputation method assumes that after controlling for all the observations in the original data set, all the missingness occurs at random (**?**). A good imputation model is numerically recognized by low fraction of information missing due to nonresponse (`fmi`), small proportion of total variance that is attributable to the missing data (λ) values; and visually recognized by the similarity in kernel density estimates and data points distribution between the observed and imputed data sets (**?**). The number of imputations (m) should be chosen such that the loss of efficiency, $le = \frac{fmi}{m} \leq 0.05$ (**?**). To give the imputation model more information, the data was imputed using 3456 data points. The analysis model on the imputed data sets was run using 72 data points.

The recommended specifications of the imputation model for this data set are: 1) uses at least 24 imputations (24 was selected because it is divisible by 4 cores in the computer processor), 2) includes all the covariates (Number of emerged seedlings, Male, Female, and Total) and predictors in the analysis model, and 3) uses an overdispersed Poisson regression model (**?**; **?**; and **?**). While the first two specifications were met, overdispersed Poisson regression model could not be specified. The extensions for count data mentioned in Chapter 7 (**?**) do not fit this data set well. Both `micemd`'s `mice.impute.2l.glm.pois` and `mice.impute.2l.2stage.pois` functions impute the numbers of Male, Female, and Total separately (**?**). The predictive mean matching (`pmm`) method in the `mice` package can handle counts (**?**) and is the optimal solution for this data set at this writing (Figure **??**).

Diagnosis of the imputation model with m = 24

The imputation code with predicted mean matching (`pmm`) method is provided in (**?**). Each round of imputation is distinguished by the number of imputations (m). The loss of efficiency (le) are all below the recommended value of 0.05 (**?**) for the analysis model terms using the imputed data sets, under three m values and the imputation performance improved as m increased (Table **??**). m was capped at 2400 imputations for this manuscript because of limited computational power. The diagnosis of data imputation is demonstrated here with m = 24 for ease of view. The distribution of the imputed data sets match that of the observation (Figure **??**) so the outputs (imputed data sets) were used for further regression analyses that involved sex ratio.

Diagnosis of the analysis model with m = 24

The similarity in the magnitude of reduction of deviance and dispersion values suggests comparable, small impacts of population aboveground mass and population stand density covariates on the improvement of the

goodness of fit for comparing sex ratio using 2019 imputed data (**?**). In addition, using either of them as a covariate in the analysis models resulted in nonestimable corn weed management effects on sex ratios in some treatments, so 2019 sex ratios were averaged over corn weed management regimes and compared without any covariates for simplicity.

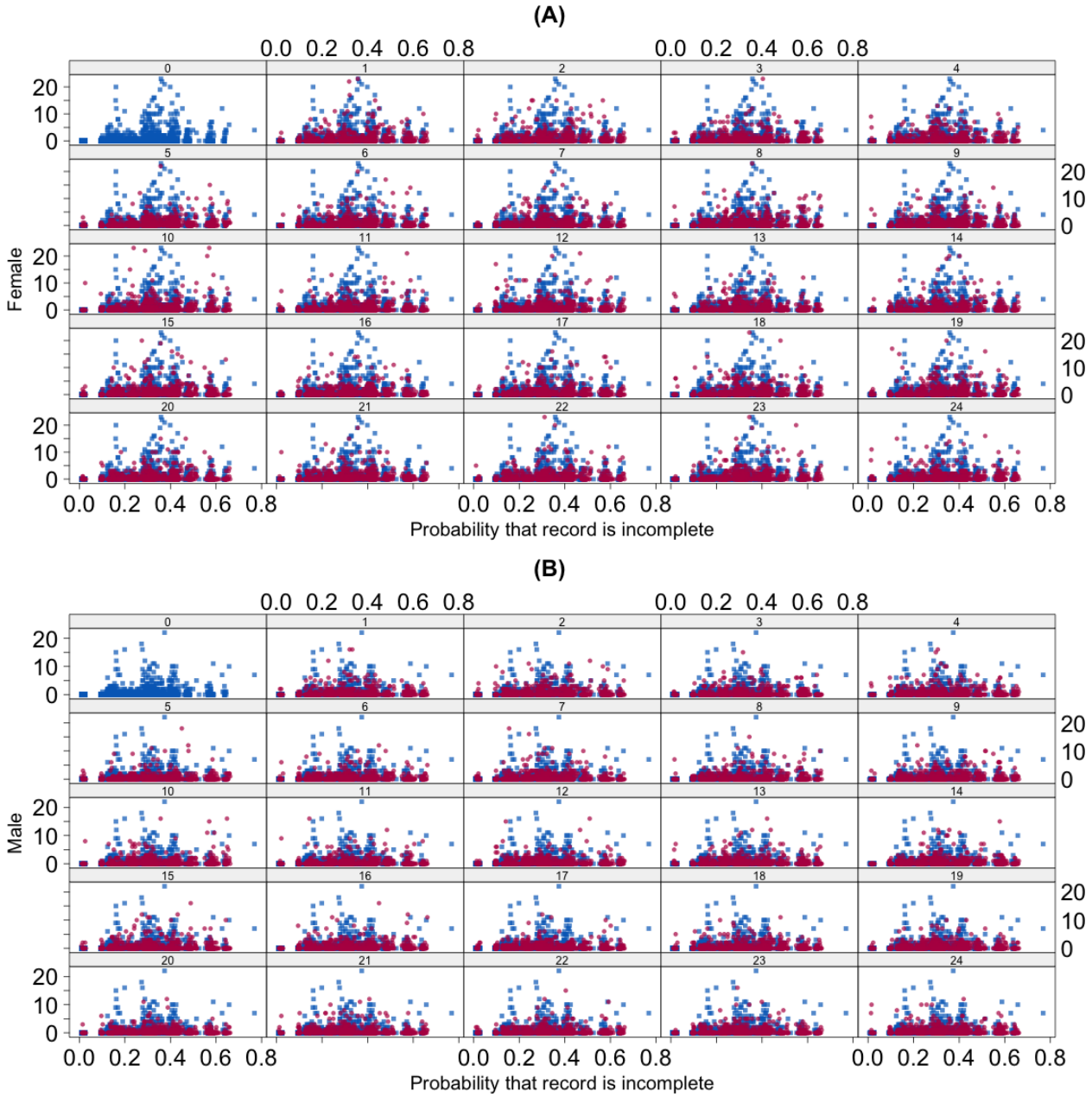

Figure 1: Number of female (A) and male (B) against the missingness probability for observed (data set numbered 0) and imputed data (data set numbered 1 to 24).

Table 1: Imputation model indices with different values of m. NAs resulted from nonestimable number of male and female in a treatment.

	n/m			riv			fmi			lambda			le		
term	m = 24	m = 240	m = 2400	m = 24	m = 240	m = 2400	m = 24	m = 240	m = 2400	m = 24	m = 240	m = 2400	m = 24	m = 240	m = 2400
(Intercept)	0.2083	0.1167	0.1433	0.3739	0.3644	0.3386	0.2793	0.2699	0.2554	0.2722	0.2671	0.2529	0.0116	0.0011	0.0001
Block: 2	0.2083	0.1167	0.1433	0.5344	0.5998	0.5840	0.3575	0.3781	0.3712	0.3483	0.3749	0.3687	0.0149	0.0016	0.0002
Block: 3	0.2083	0.1167	0.1433	0.5804	0.5371	0.5117	0.3770	0.3525	0.3410	0.3672	0.3494	0.3385	0.0157	0.0015	0.0001
Block: 4	0.2083	0.1167	0.1433	0.9424	0.7040	0.6527	0.4978	0.4164	0.3975	0.4852	0.4131	0.3949	0.0207	0.0017	0.0002
Crop ID: C3	0.2083	0.1167	0.1433	0.3516	0.4493	0.4273	0.2669	0.3130	0.3019	0.2601	0.3100	0.2994	0.0111	0.0013	0.0001
Crop ID: C2	0.2083	0.1167	0.1433	0.1682	0.2340	0.2591	0.1480	0.1923	0.2083	0.1440	0.1896	0.2058	0.0062	0.0008	0.0001
Crop ID: C4	0.2083	0.1167	0.1433	0.8327	0.8853	0.8554	0.4663	0.4730	0.4636	0.4544	0.4696	0.4610	0.0194	0.0020	0.0002
Crop ID: O3	0.2083	0.1167	0.1433	0.7229	0.6626	0.6850	0.4307	0.4018	0.4091	0.4196	0.3985	0.4065	0.0179	0.0017	0.0002
Crop ID: A4	0.2083	0.1167	0.1433	0.3330	0.5269	0.5021	0.2563	0.3482	0.3368	0.2498	0.3451	0.3343	0.0107	0.0015	0.0001
Crop ID: S2	0.2083	0.1167	0.1433	1.5154	0.0006	0.0009	0.6167	0.0031	0.0034	0.6024	0.0006	0.0009	0.0257	0.0000	0.0000
Crop ID: S3	0.2083	0.1167	0.1433	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Crop ID: S4	0.2083	0.1167	0.1433	0.0006	NA	NA	0.0031	NA	NA	0.0006	NA	NA	0.0001	NA	NA
Corn weed managementlow	0.2083	0.1167	0.1433	0.4410	0.4470	0.4214	0.3140	0.3119	0.2990	0.3060	0.3089	0.2965	0.0131	0.0013	0.0001
Crop ID: C3 x Corn weed managementlow	0.2083	0.1167	0.1433	0.6788	0.6881	0.6047	0.4150	0.4109	0.3794	0.4043	0.4076	0.3768	0.0173	0.0017	0.0002
Crop ID: C2 x Corn weed managementlow	0.2083	0.1167	0.1433	0.4194	0.8696	0.7311	0.3032	0.4685	0.4249	0.2955	0.4651	0.4223	0.0126	0.0020	0.0002
Crop ID: C4 x Corn weed managementlow	0.2083	0.1167	0.1433	1.3038	1.0537	1.0291	0.5798	0.5166	0.5097	0.5659	0.5131	0.5072	0.0242	0.0022	0.0002
Crop ID: O3 x Corn weed managementlow	0.2083	0.1167	0.1433	0.6842	0.6712	0.8059	0.4170	0.4049	0.4488	0.4062	0.4016	0.4463	0.0174	0.0017	0.0002
Crop ID: A4 x Corn weed managementlow	0.2083	0.1167	0.1433	0.2450	0.4780	0.4938	0.2020	0.3264	0.3331	0.1968	0.3234	0.3306	0.0084	0.0014	0.0001
Crop ID: S2 x Corn weed managementlow	0.2083	0.1167	0.1433	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Crop ID: S3 x Corn weed managementlow	0.2083	0.1167	0.1433	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Crop ID: S4 x Corn weed managementlow	0.2083	0.1167	0.1433	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Note: Some zeros are due to rounding. C2: corn in the 2-year rotation, C3: corn in the 3-year rotation, C4: corn in the 4-year rotation, S2: soybean in the 2-year rotation, S3: soybean in the 3-year rotation, S4: soybean in the 4-year rotation, O3: oat in the 3-year rotation, and A4: oat in the 4-year rotation

Table 2: Numerical diagnosis of model's goodness of fit with and without covariates on 2019 imputed data sets.

(A) Population aboveground mass covariate																							
Imputation	1	2	3	4	**5**	6	**7**	8	9	10	**11**	12	**13**	14	**15**	16	17	18	19	20	21	22	23	24
Deviance	45.87	42.80	34.31	36.78	**40.89**	62.70	**52.69**	56.73	22.47	46.68	**39.74**	38.19	**26.05**	47.66	**28.19**	46.58	43.79	31.49	48.21	50.59	31.26	34.89	51.67	38.57
Dispersion	1.95	1.84	1.38	1.48	**1.54**	2.40	**2.14**	2.38	1.05	1.93	**1.75**	1.60	**1.06**	2.01	**1.22**	1.90	1.61	1.24	1.87	2.19	1.37	1.36	2.09	1.50
(B) Population stand density covariate																								
Imputation	1	2	3	4	**5**	6	**7**	8	9	10	**11**	12	**13**	14	**15**	16	17	18	19	20	21	22	23	24
Deviance	45.22	37.60	41.66	39.11	**52.50**	66.21	**61.79**	51.69	27.44	47.33	**36.24**	47.05	**28.19**	44.59	**24.53**	59.57	48.72	38.23	35.92	66.81	35.47	49.25	58.26	58.51
Dispersion	1.98	1.62	1.67	1.58	**2.00**	2.54	**2.55**	2.12	1.28	1.94	**1.59**	1.99	**1.16**	1.84	**1.05**	2.47	1.79	1.56	1.40	2.96	1.54	1.88	2.41	2.29
(C) No covariate																								
Imputation	1	2	3	4	**5**	6	**7**	8	9	10	**11**	12	**13**	14	**15**	16	17	18	19	20	21	22	23	24
Deviance	78.24	63.29	72.17	72.88	**90.87**	105.28	**91.76**	81.60	49.72	73.75	**63.36**	64.03	**45.61**	80.43	**47.17**	95.78	75.41	57.65	72.34	96.18	68.97	69.57	96.86	87.33
Dispersion	2.10	1.72	1.87	1.82	**2.12**	2.56	**2.36**	2.07	1.42	1.99	**1.70**	1.71	**1.19**	2.03	**1.23**	2.42	1.75	1.48	1.76	2.52	1.80	1.66	2.44	2.15

Note:

Bold columns are imputations that produced full sets of data (five among 24 imputed sets) that allow estimation of the effects of crop idenity, corn weed management, and their interaction.