## Appendix

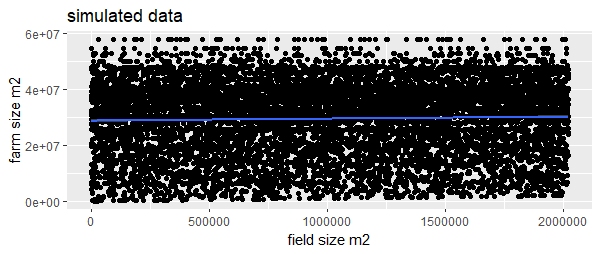
### Appendix 1 – Simulation experiments

As farm size is the summed area of all fields of one farm, we wanted to test if there is a deterministic association between field and farm size, and if this association depends on the number of fields in a farm. Therefore, we generated random field size data, assigned each field a random farm ID, and then calculated the farm size by aggregating field sizes by farm ID.

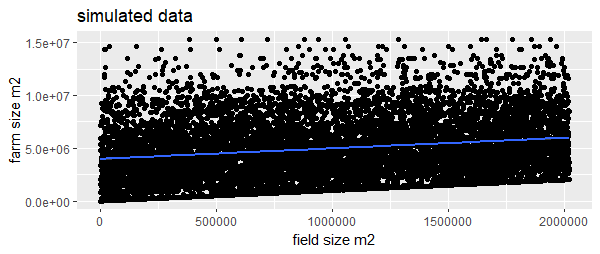
In the first experiment, we allocated an average of 30 fields to each farm. We did not observe an obvious trend between field and farm size (Fig. A1.1)

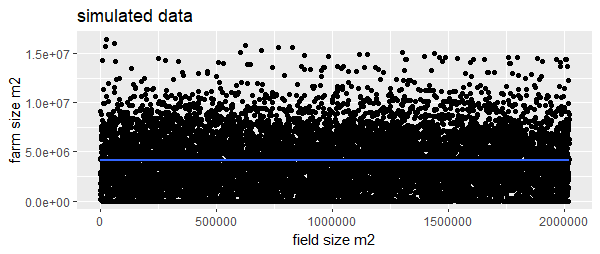
In the second experiment, we allocated an average of 5 fields per farm. Now, we can observe a positive relationship between field size and farm size. The slope would further increase with a decreasing number of fields per farm (Fig. A1.2).

To avoid potential biases from this effect in our model, and because in a practical setting, the number of fields operated by a farm (as well as the total size) is unknown, we removed this trend by subtracting the field size from the farm size of each field. The regression model then predicts the summed area of all remaining fields of a farm (Fig. A1.3).



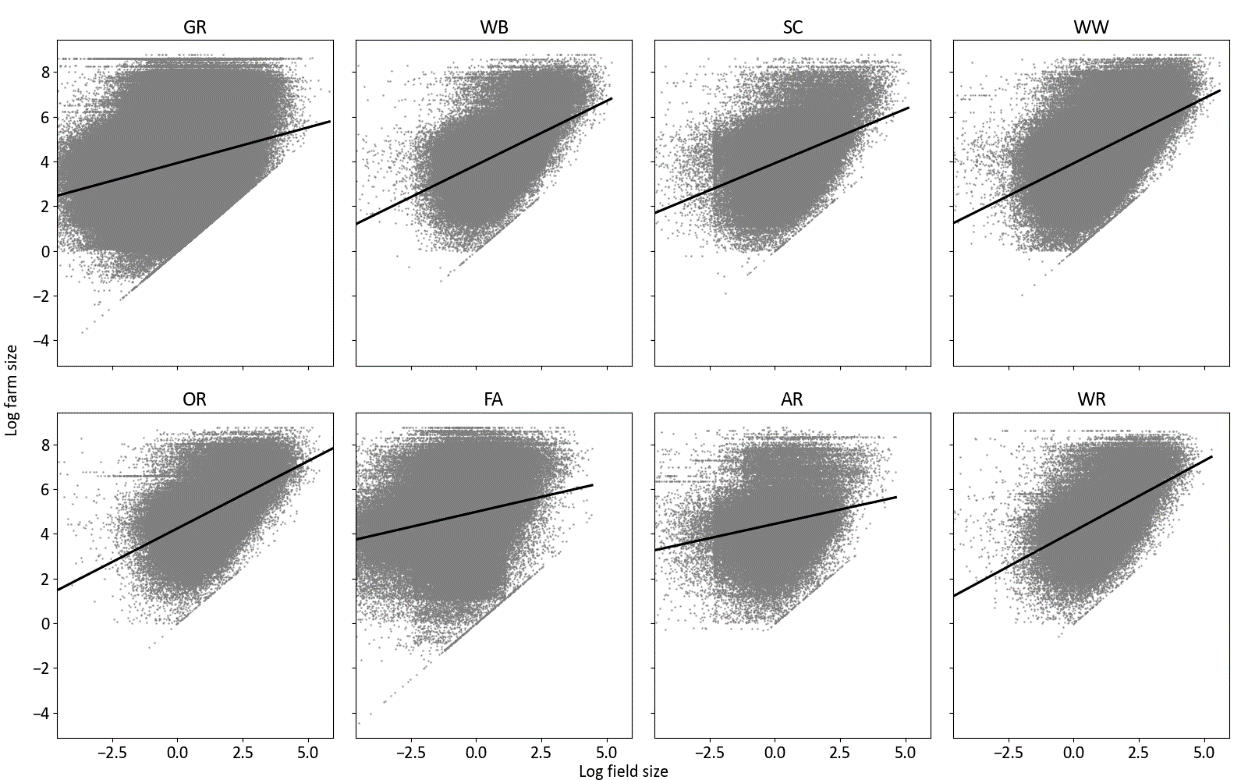
*Figure Appendix. 1.1: simulated field size (x-axis) vs simulated farm size (y-axis), each point is a single agricultural field. Farms operate about 30 fields on average. No obvious trend is observable.*

*Figure Appendix.* *1.2: simulated field size (x-axis) vs simulated farm size (y-axis). Farms operate about 5 fields on average; strong positive trend.*



*Figure Appendix. 1.3: simulated field size (x-axis) vs simulated remaining farm size (y-axis), which means that for each field we subtracted the field size from the farm size. Farms operate about 5 fields on the average; no obvious trend.*

### Appendix 2 – Field size to farm size relationships for different crop groups



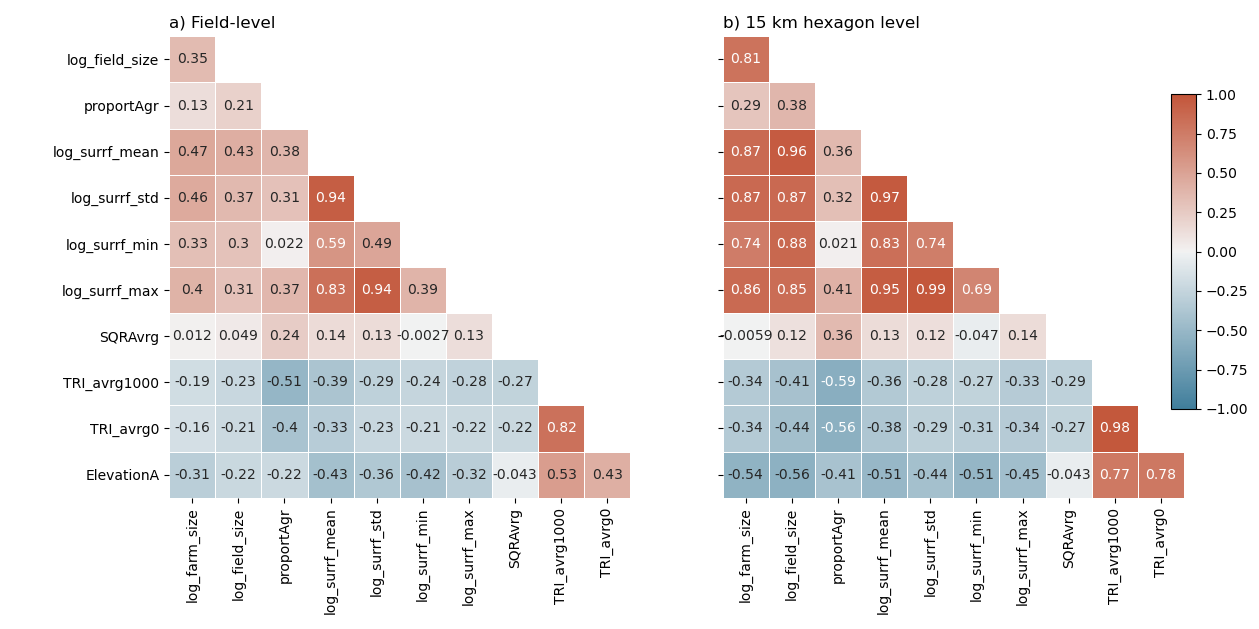
*Figure Appendix. 2.1: Field size to farm size relationships for different crop groups; x-axis is the log field size and y-axis represents the log farm size*

### Appendix 3 – Hypothesis on the effect of proportion of agriculture and soil quality on farm sizes

In contexts where agriculture has expanded since long time periods, like Europe, agriculture tends to adjust its location to areas that have higher soil quality and agronomic value (Mather & Needle, 1998). A hypothesis would be that farms in more productive areas would tend to be more profitable and would tend to grow over time and consolidate into larger farms. However, on the other hand, prime agricultural lands tend to attract larger population densities, which may result in dense smaller farms landscapes. Over the long term, standard theory would say that these farms would tend to consolidate as the economy develops and labour force goes out of agriculture especially in regions where the vicinity to urban agglomerations supports off-farm economic development. But it might also depend on the types of crops (horticulture can and will often prevail on smaller farms in densely populated peri urban areas, possibly with pluri-activity) and in contrast to prime agricultural regions that favour field crops like cereals and consequentially larger farms.

Regions with a large share of agriculture are, at least in quite consolidated regions such as Europe, likely to be regions that are more favorable to agriculture and thus have been settled first. If this is related to a long history of staple crops production, it could result in larger farms because of progressive consolidation, with capitalization, mechanization (e.g. the Parisian Basin in France). In contrast, regions with a smaller proportion of agriculture, or perhaps regions with a smaller proportion of crops (in possibly larger agricultural landscapes dominated by grazing lands) are likely to be more marginal regions for agriculture (e.g. mountainous regions, cold northern Europe, dry regions in the Mediterranean). Marginal agriculture would be associated with less capitalization, lower incomes, etc. and thus smaller farms.

### Appendix 4 – Correlation matrix



*Figure Appendix. 4.1: Correlation matrix of all input variables at a) the field level and b) the 15km hexagons.*

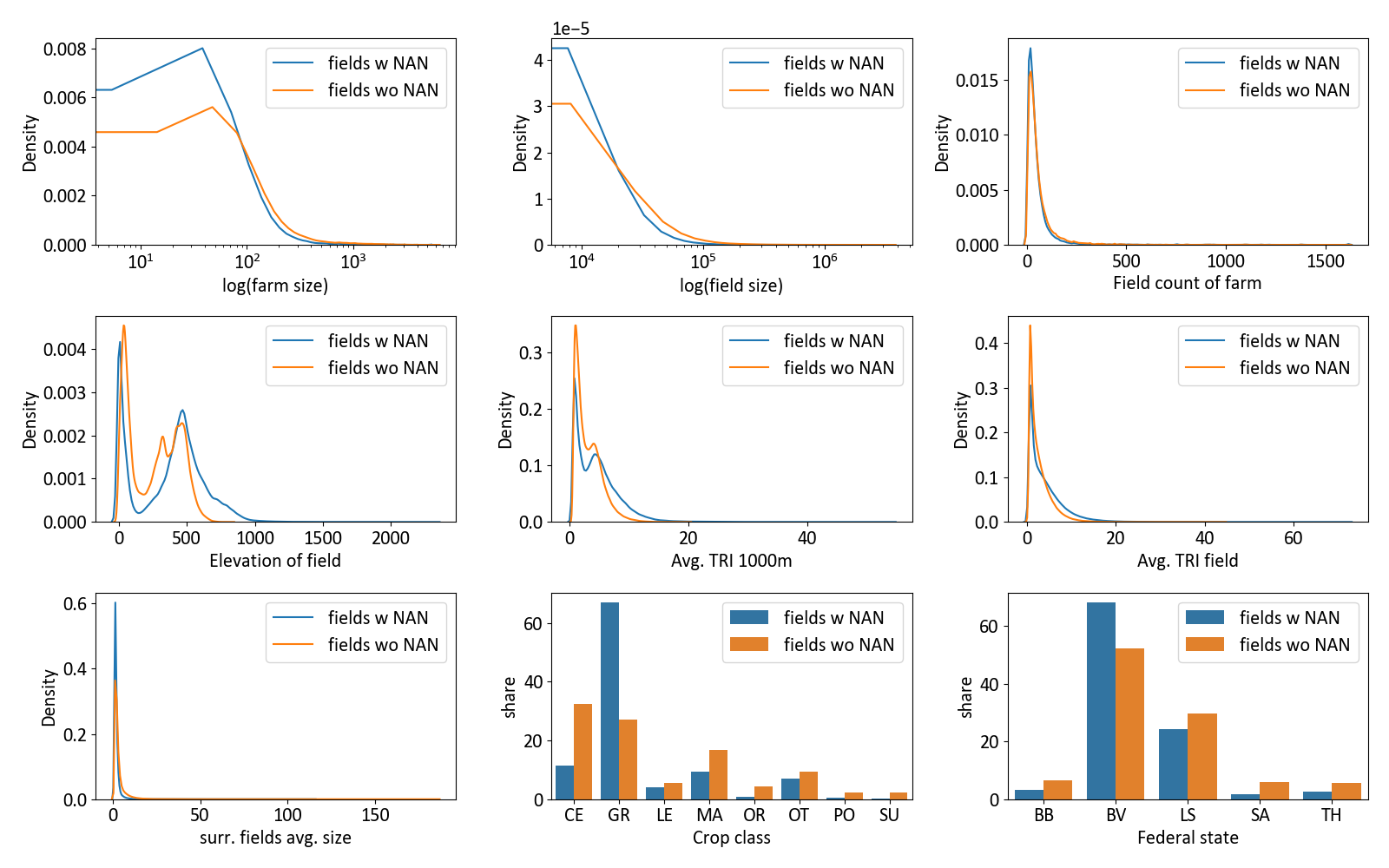
### Appendix 5 – Propensity score matching

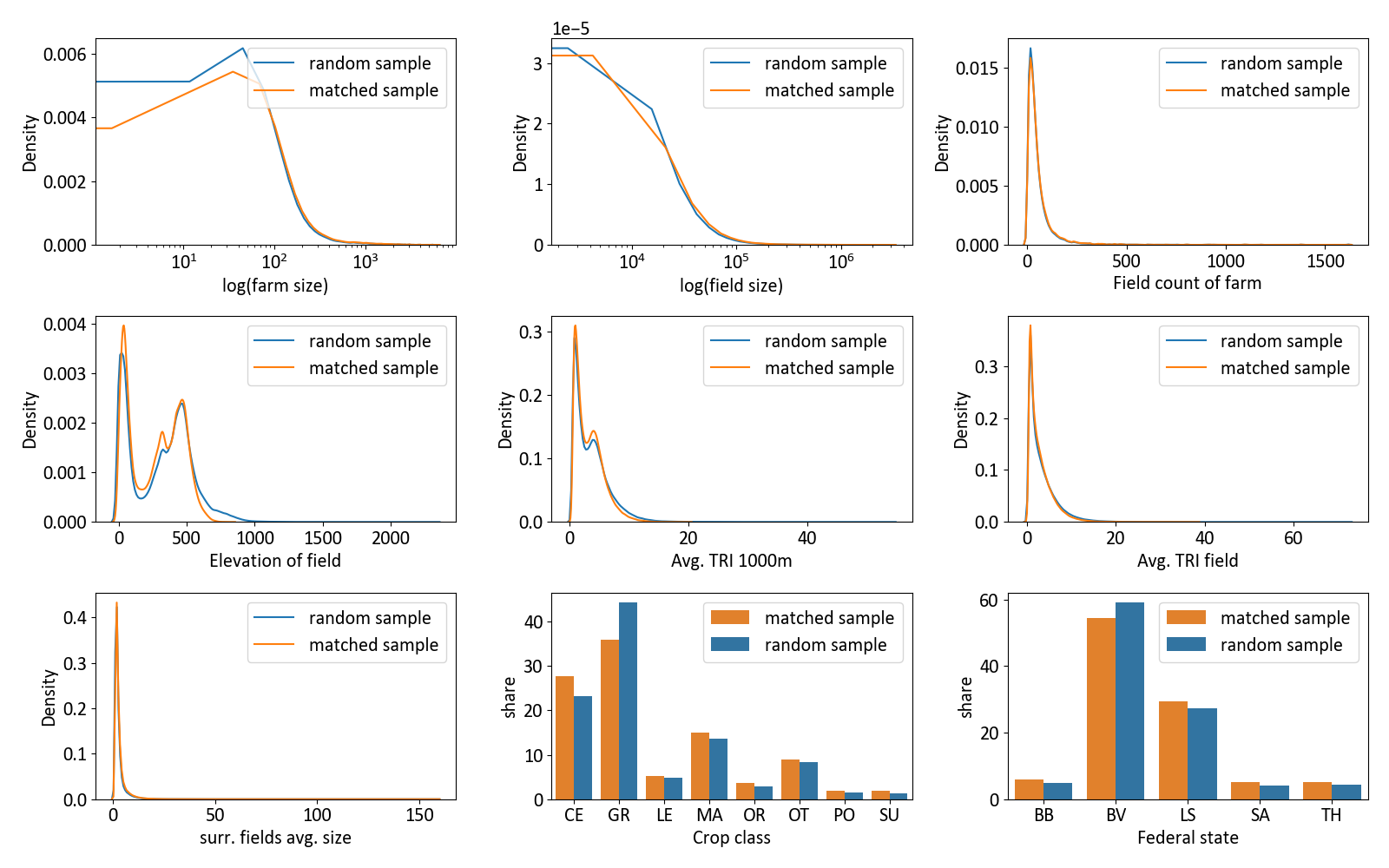
In our data set, fields with soil quality rating values and fields with NoData values were not randomly distributed, as seen from the data distributions over other explanatory variables (see Figure A4.1). NoData values were observed more frequently for fields on smaller farms, for smaller fields, in areas with higher altitudes, and for fields located in Bavaria. .Due to this bias, it was not possible to draw a random sample from all fields with soil quality rating values without causing a bias in the sample. We applied propensity score matching to generate a sample very similar to a pure random sample. The workflow included the following steps:

1. We took a 25% random sample from the population of all fields, including those with NoData values (hereafter called random sample, n=825,339).
2. We excluded all fields with NoData values from the total population (hereafter called reduced population, n=1,878,126).
3. fields. To find similar fields, we used propensity score matching between the random sample and the reduced population.

Using a logistic classifier, we calculated the propensity scores of all fields by predicting each field's membership in the sample or the reduced population. For the predictions, we used the variables crop type group, field size, number of fields on the farm, , , altitude, , average terrain ruggedness index (TRI) in a 1000 m radius around the field center, average TRI of the field, the average surrounding field sizes in a 1000 m radius around the field, and federal state. Propensity scores were derived by calculating the logit of the predicted values. The final matching was done using nearest neighbor matching and a caliper of 0.3 times the standard deviation of the propensity scores.

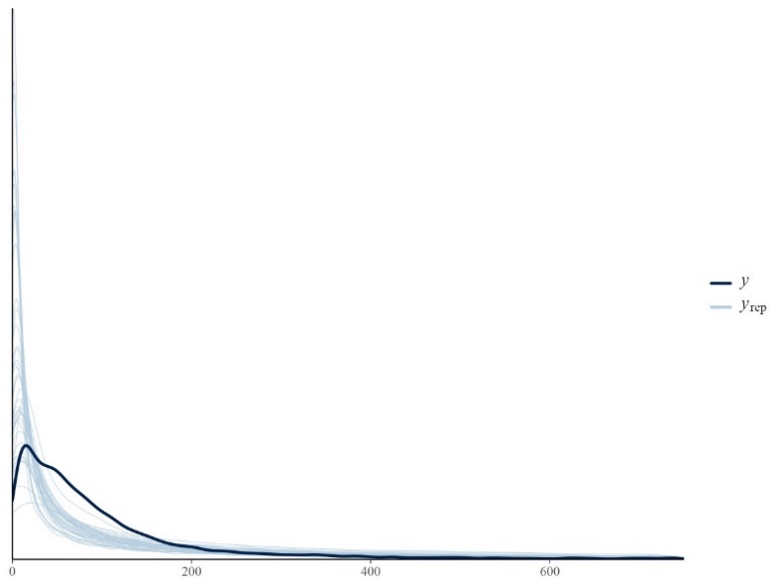
The matching procedure matched 801,207 fields from the reduced dataset with fields from the random sample. The distribution of the matched sample was very similar to the random sample, as shown by the data distribution of the other explanatory variables (see the second figure below figure 4.2).

*Figure Appendix. 5.1: Comparison of fields without (fields wo NAN) and with (fields w NAN) soil quality rating.*

*Figure Appendix. 5.2: Comparison of a random sample of fields that are allowed to have no information on soil quality and the matched sample of fields with soil quality information and similar distributions of the other predictors.*

Appendix 6 – Prior predictive checks

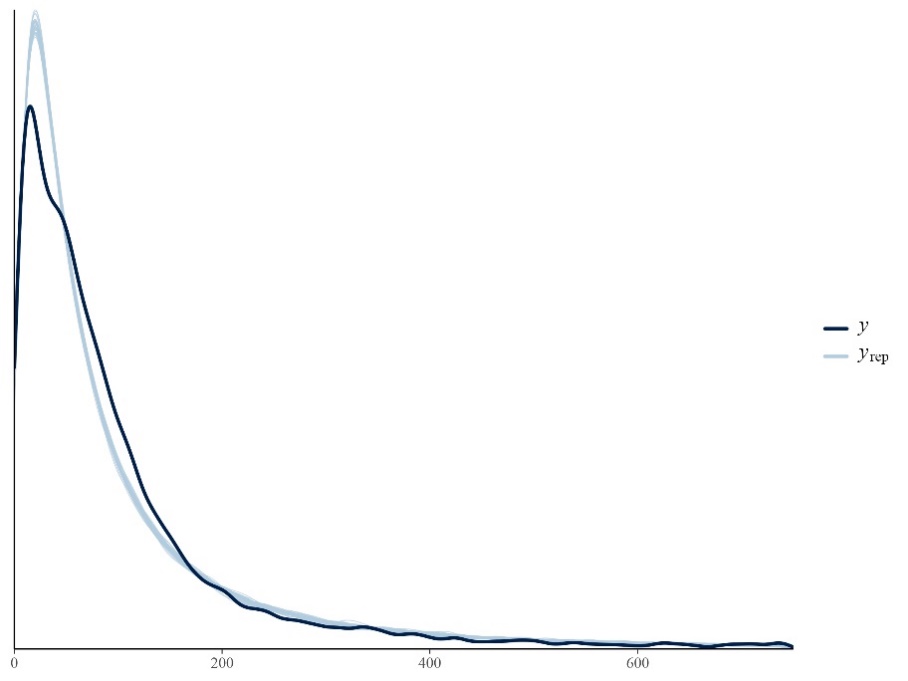
The prior distributions for the unknown coefficients were optimized in an iterative process of prior predictive checks, i.e., predicting the data only based on the chosen priors, and subsequently adjusting those prior distributions to yield realistic predictions based on information obtained from sampling diagnostics and predictive checks. In this way, we derived weakly informed priors that were on one hand regularizing enough to facilitate model convergence, and on the other hand resulted in plausible predictive simulations while not restricting and biasing the outcome distribution (Appendix 5.1).



*Figure Appendix. 6.1: Prior predictive checks for our chosen priors based on predictions without seeing the data to test how the priors interact with each other.*

### Appendix 7 – Posterior predictive checks

We evaluated the model fit based on posterior predictive checks, that is, predicting new hypothetical data sampled from the posterior predictive distribution and comparing it to a random draw of observed data (Appendix 6.1). We found a good overlap of both the posterior predictive distribution and the random draw of observed data.



*Figure Appendix. 7.1: Posterior predictive checks for our final model. The dark blue represents the random draw of observed data, and the light blue lines represent samples from the posterior distribution*

### Appendix 8 – Group level effects of the Bayesian model

Regarding the varying intercepts for the scale parameter “sigma”, which controls the spread of the predicted distribution, we can see that the largest positive deviations can be found in East Germany (Appendix 7.1). Negative deviations mean that the geometric standard deviation (i.e., the multiplicative error) decreases. Grassland has higher intercepts throughout, except for Bavaria, translating into wider distributions for this group, which means higher uncertainties.

### 

Figure Appendix. 8.1: Varying intercepts for the scale parameter “sigma”, which are added to the global intercept for the scale model. The red dotted lines represent federal states in the former GDR.

Appendix 9 – Error diagnosis measures per farm size and model

Table Appendix 9.1: Error diagnosis measures per farm size for the Bayesian model.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Farm size class | Number of fields | ME | P | RMSE | MAPE |
| (0, 5] | 39,373 | 63 | 70 | 94 | 3,994 |
| (5, 10] | 45,822 | 59 | 67 | 89 | 820 |
| (10, 20] | 98,854 | 52 | 67 | 85 | 364 |
| (20, 50] | 230,700 | 38 | 71 | 81 | 125 |
| (50, 100] | 287,574 | 6 | 78 | 78 | 50 |
| (100, 200] | 238,049 | -38 | 118 | 124 | 60 |
| (200, 500] | 141,405 | -101 | 223 | 245 | 70 |
| (500, 1000] | 57,460 | -290 | 328 | 438 | 52 |
| (1000, 7000] | 76,617 | -1401 | 949 | 1692 | 68 |

Table Appendix 9.2: Error diagnosis measures per farm size for the random forest regression with the same observations and variables for training as the Bayesian model.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Farm size class | Number of fields | ME | P | RMSE | MAPE |
| (0, 5] | 39,380 | 164 | 161 | 230 | 9,913 |
| (5, 10] | 45,829 | 148 | 145 | 207 | 2,073 |
| (10, 20] | 98,861 | 130 | 128 | 183 | 901 |
| (20, 50] | 230,700 | 110 | 121 | 164 | 340 |
| (50, 100] | 287,596 | 78 | 125 | 147 | 114 |
| (100, 200] | 238,056 | 45 | 174 | 179 | 60 |
| (200, 500] | 141,428 | 30 | 286 | 288 | 76 |
| (500, 1000] | 57,499 | -100 | 326 | 341 | 40 |
| (1000, 7000] | 76,665 | -1196 | 928 | 1514 | 55 |

Table Appendix 9.3: Error diagnosis measures per farm size for the random forest regression with the same observations but all available variables for training.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Farm size class | Number of fields | ME | P | RMSE | MAPE |
| (0, 5] | 39,380 | 167 | 186 | 250 | 10,027 |
| (5, 10] | 45,829 | 147 | 167 | 223 | 2,060 |
| (10, 20] | 98,861 | 125 | 147 | 193 | 866 |
| (20, 50] | 230,700 | 103 | 135 | 170 | 319 |
| (50, 100] | 287,596 | 70 | 136 | 153 | 103 |
| (100, 200] | 238,056 | 40 | 185 | 190 | 66 |
| (200, 500] | 141,428 | 39 | 301 | 303 | 81 |
| (500, 1000] | 57,499 | -70 | 328 | 335 | 40 |
| (1000, 7000] | 76,665 | -1156 | 920 | 1478 | 53 |

### Appendix 10 – Characteristics per farm size

Table Appendix 10.1: summary statistics stratified by farm size classes.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Farm size class | Average field size | Standard deviation field size | Average surrounding field size | Standard deviation surrounding field size | Average number of fields per farm | Average number of crops per farm |
| (0,5] | 0.73 | 0.73 | 1.85 | 1.52 | 3.72 | 1.43 |
| (5,10] | 1.07 | 1.15 | 1.86 | 1.48 | 6.79 | 1.97 |
| (10,20] | 1.37 | 1.49 | 1.86 | 1.34 | 11.00 | 2.66 |
| (20,50] | 1.78 | 2.02 | 2.01 | 1.31 | 18.84 | 3.32 |
| (50,100] | 2.16 | 2.49 | 2.16 | 1.43 | 32.38 | 3.95 |
| (100,200] | 2.66 | 3.38 | 2.5 | 1.91 | 50.83 | 4.43 |
| (200,500] | 4.31 | 6.73 | 3.89 | 3.63 | 68.27 | 4.84 |
| (500,1000] | 8.24 | 12.68 | 6.71 | 5.33 | 85.04 | 5.26 |
| (1000,7000] | 10.63 | 16.02 | 8.02 | 5.97 | 158.76 | 6.03 |

# References

Mather, A. S., & Needle, C. L. (1998). The forest transition: A theoretical basis. *Area*, *30*(2), 117–124. https://doi.org/10.1111/j.1475-4762.1998.tb00055.x