Supplement of "BIS-4D: Mapping soil properties and their uncertainties at 25 m resolution in the Netherlands"

Anatol Helfenstein[1,2], Vera L. Mulder[1], Mirjam J.D. Hack-ten Broeke[2], Maarten van Doorn[3,4], Kees Teuling[2], Dennis J.J. Walvoort[2], and Gerard B.M. Heuvelink[1,5]

[1]Soil Geography and Landscape Group, Wageningen University, PO Box 47, 6700 AA Wageningen, The Netherlands
[2]Soil, Water and Land Use Team, Wageningen Environmental Research, Droevendaalsesteeg 3, 6708 RC Wageningen, The Netherlands
[3]Nutriënten Management Instituut, Nieuwe Kanaal 7C, 6709 PA, Wageningen, the Netherlands
[4]Environmental Systems Analysis Group, Wageningen University, PO Box 47, 6700 AA, Wageningen, the Netherlands
[5]ISRIC – World Soil Information, PO Box 353, 6700 AJ, Wageningen, the Netherlands

Correspondence: Anatol Helfenstein (anatol.helfenstein@wur.nl)

Contents

S1 Clay **S3**
S1.1 Prediction maps . . . S3
S1.2 Accuracy assessment . . . S6
S1.3 Variable importance . . . S9

S2 Silt **S10**
S2.1 Prediction maps . . . S10
S2.2 Accuracy assessment . . . S13
S2.3 Variable importance . . . S16

S3 Sand **S17**
S3.1 Prediction maps . . . S17
S3.2 Accuracy assessment . . . S20
S3.3 Variable importance . . . S23

S4 Bulk density (BD) **S24**
S4.1 Prediction maps . . . S24
S4.2 Accuracy assessment . . . S27
S4.3 Variable importance . . . S30

S5 Soil organic matter (SOM) **S31**

S5.1 Prediction maps . . . S31
S5.2 Accuracy assessment . . . S34
S5.3 Variable importance . . . S37

S6 pH **S39**
S6.1 Prediction maps . . . S39
S6.2 Accuracy assessment . . . S42
S6.3 Variable importance . . . S45

S7 Total N (N_{tot}) **S46**
S7.1 Prediction maps . . . S46
S7.2 Accuracy assessment . . . S49
S7.3 Variable importance . . . S52

S8 Oxalate-extractable P (P_{ox}) **S53**
S8.1 Prediction maps . . . S53
S8.2 Accuracy assessment . . . S56
S8.3 Variable importance . . . S59

S9 Cation exchange capacity (CEC) **S60**
S9.1 Prediction maps . . . S60
S9.2 Accuracy assessment . . . S63
S9.3 Variable importance . . . S66

S1 Clay

S1.1 Prediction maps

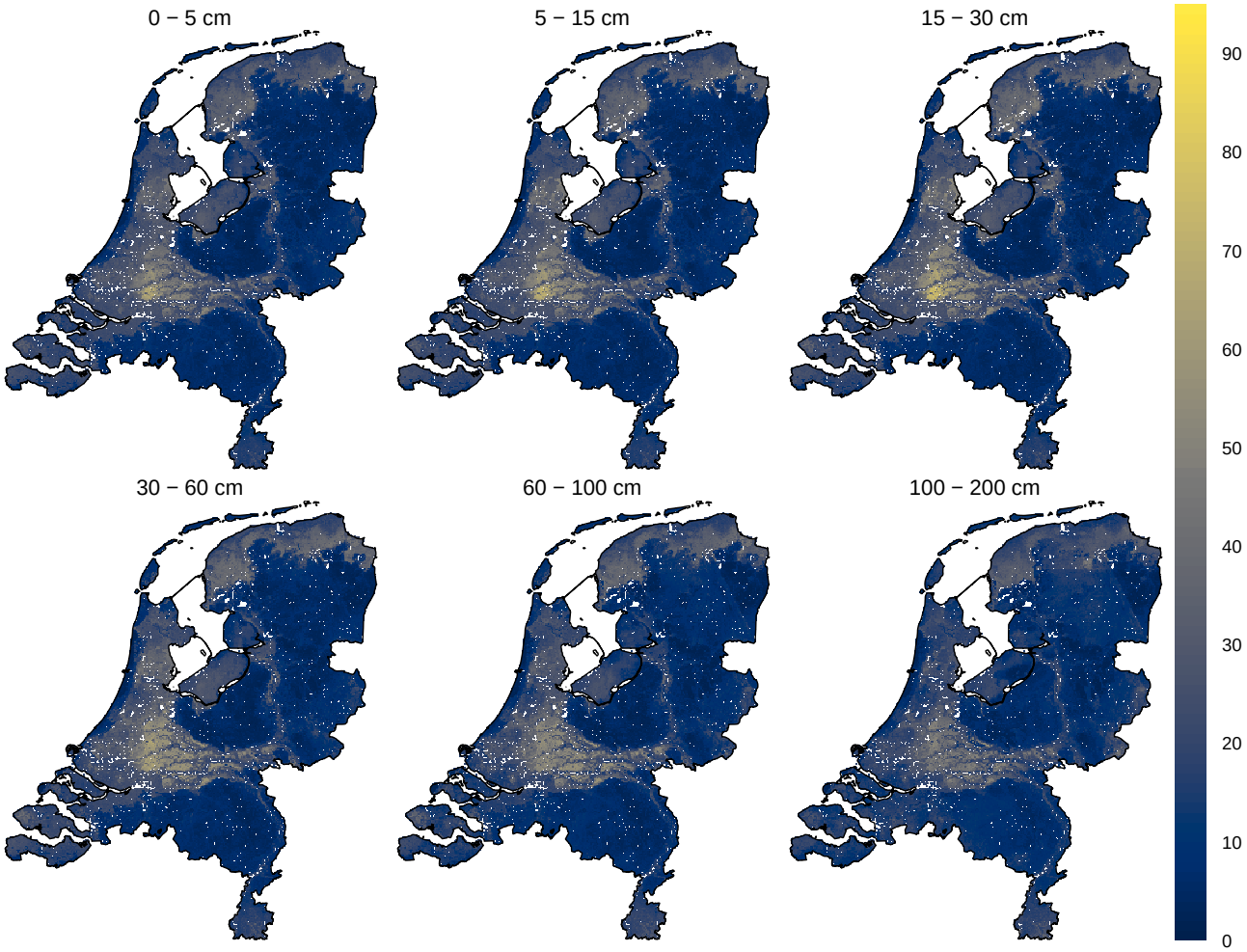

Figure S1. Predicted mean clay content [%] for every standard depth layer defined by *GlobalSoilMap* (GSM; Arrouays et al., 2015).

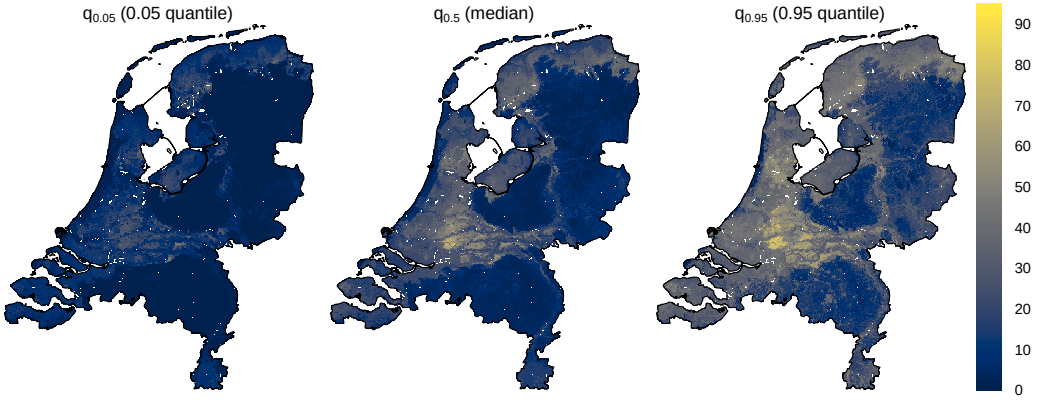

Figure S2. Predicted 5th, 50th (median) and 95th quantile for clay content [%] from 0 - 5 cm depth.

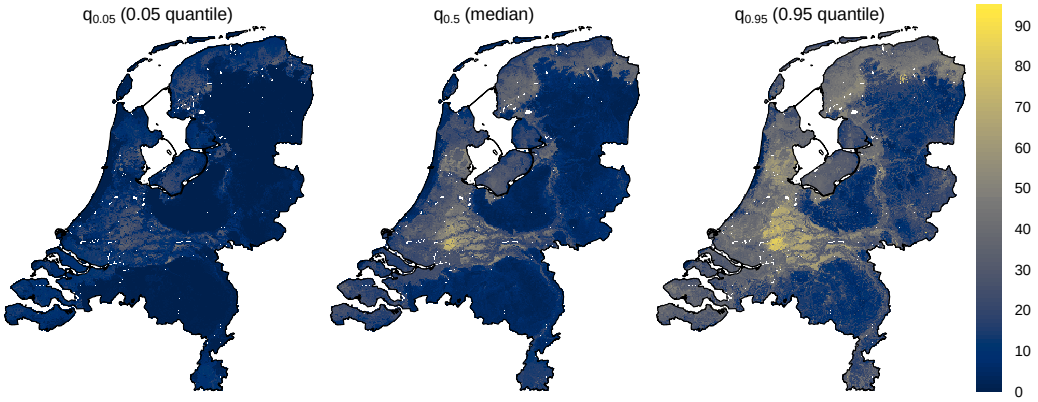

Figure S3. Predicted 5th, 50th (median) and 95th quantile for clay content [%] from 5 - 15 cm depth.

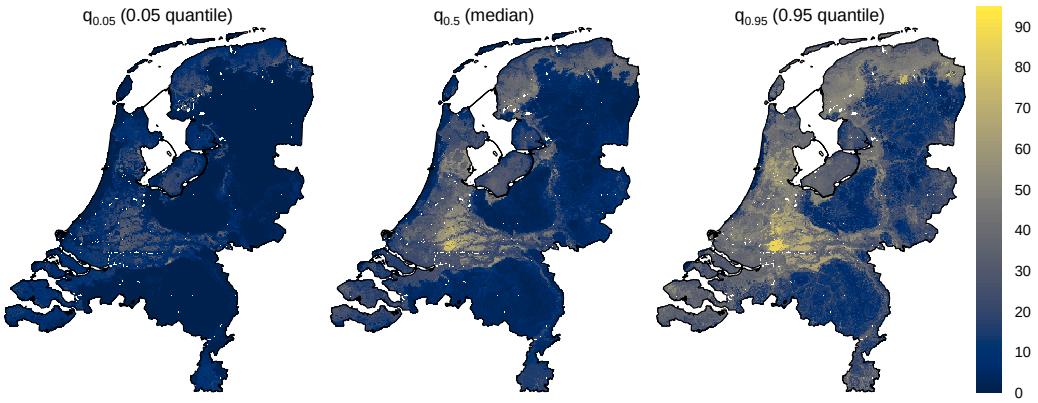

Figure S4. Predicted 5th, 50th (median) and 95th quantile for clay content [%] from 15 - 30 cm depth.

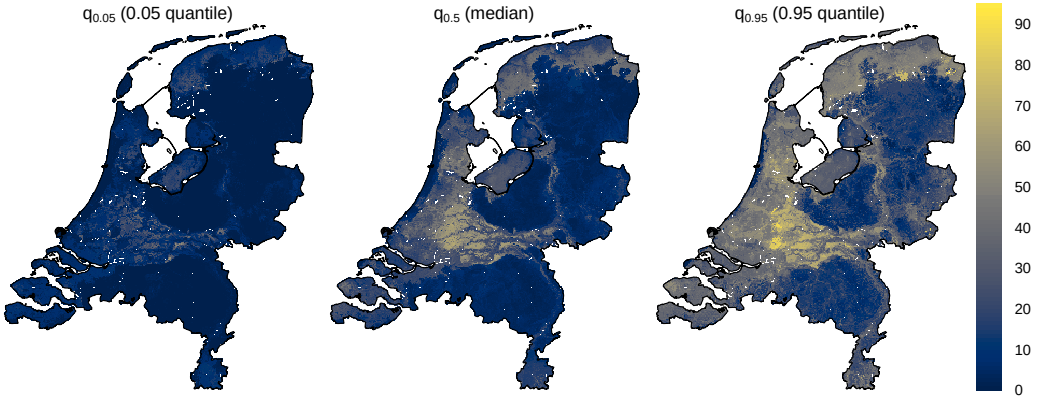

Figure S5. Predicted 5th, 50th (median) and 95th quantile for clay content [%] from 30 - 60 cm depth.

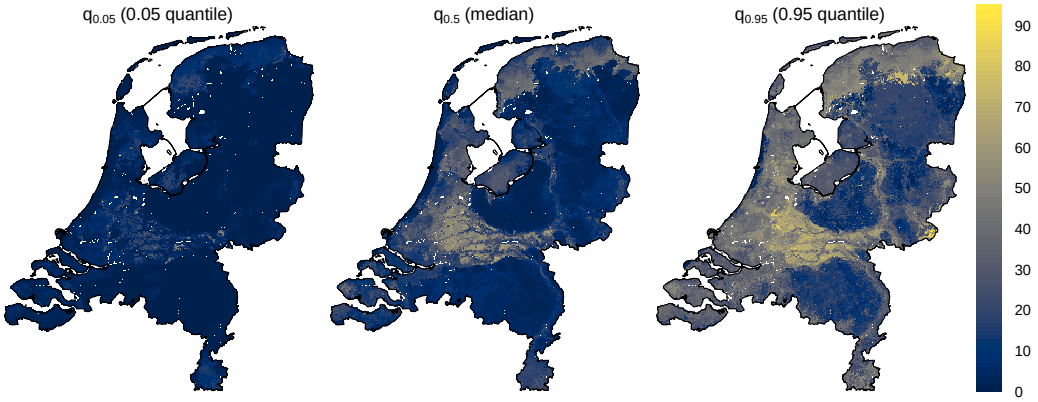

Figure S6. Predicted 5th, 50th (median) and 95th quantile for clay content [%] from 60 - 100 cm depth.

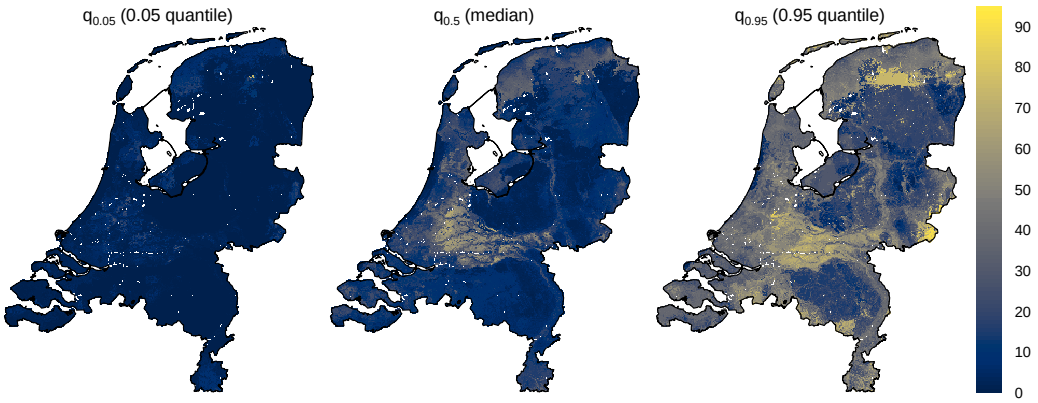

Figure S7. Predicted 5th, 50th (median) and 95th quantile for clay content [%] from 100 - 200 cm depth.

S1.2 Accuracy assessment

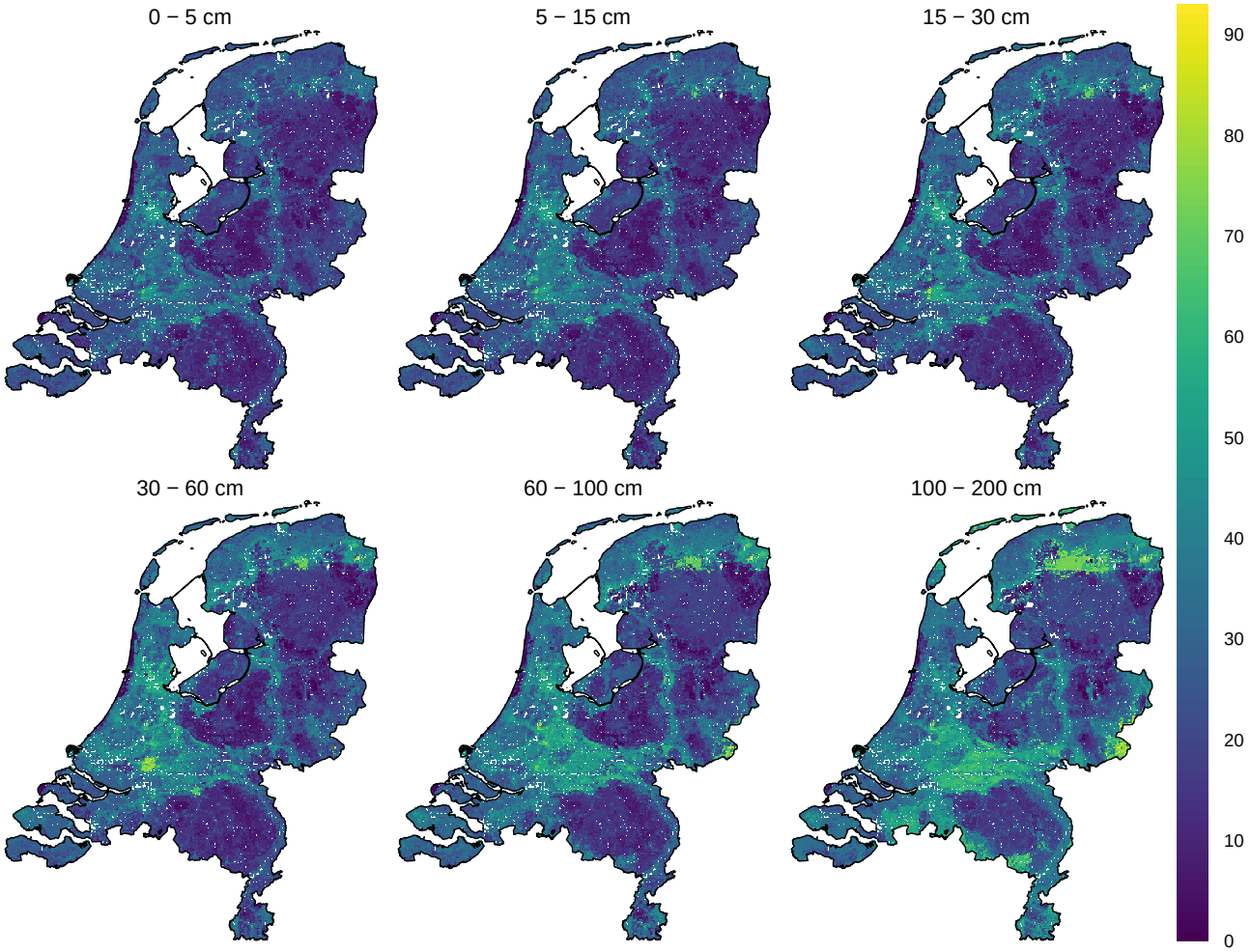

Figure S8. Maps of the 90th prediction interval width (PI90) as a measure of prediction uncertainty for clay content [%] for every GSM depth layer.

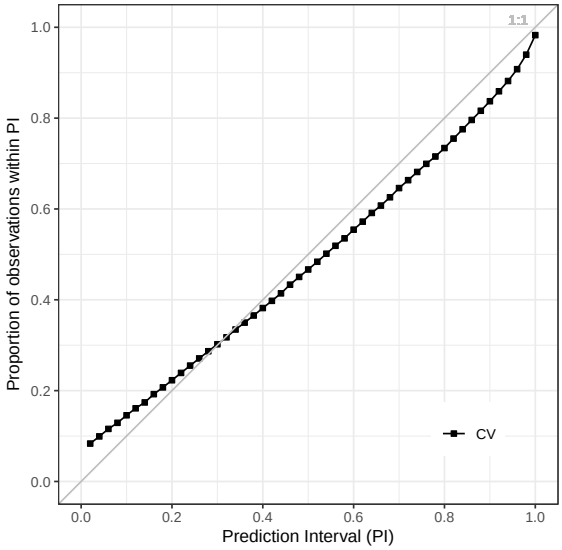

Figure S9. Prediction interval coverage probability (PICP) for prediction intervals between 0.02 and 1 of the PFB laboratory measurements used for 10-fold cross-validation (CV). The closer the points are to the 1:1 line, the more accurate the prediction uncertainty. Lines connecting the points do not represent actual data and are only for visual guidance.

Table S1. Model accuracy metrics of mean and median clay content [%] predictions using 10-fold cross-validation of PFB laboratory measurements (Sect. 2.6). Note that clay content was not measured in LSK or CCNL so design-based inference and computing 95 % confidence intervals (CI95) was not possible (Sect. 2.1.2). Here and in all other tables in the SI, we combined 0 - 5 cm and 5 - 15 cm depth layers and thereby deviate slightly from the GSM standard depth layers (Arrouays et al., 2015) because for several soil properties there were not enough observations available to statistically validate the 0 - 5 cm depth layer by itself. ME = mean error; RMSE = root mean squared error; MEC = model efficiency coefficient; PICP90 = prediction interval coverage probability of the PI90.

Prediction	Depth (cm)	n	**ME**	CI95 of ME	**RMSE**	CI95 of RMSE	**MEC**	CI95 of MEC	**PICP90**
	0 - 15	2986	**0.07**	-	**6.25**	-	**0.84**	-	**0.86**
	15 - 30	1794	**0.76**	-	**7.5**	-	**0.84**	-	**0.83**
Mean	30 - 60	3626	**0.27**	-	**8.43**	-	**0.77**	-	**0.82**
	60 - 100	3241	**-1.12**	-	**8.79**	-	**0.69**	-	**0.84**
	100 - 200	1244	**-1.56**	-	**9.31**	-	**0.6**	-	**0.83**
	0 - 15	2986	**0.54**	-	**6.15**	-	**0.84**	-	**0.86**
	15 - 30	1794	**1.1**	-	**7.22**	-	**0.85**	-	**0.83**
Median	30 - 60	3626	**0.87**	-	**8.04**	-	**0.79**	-	**0.82**
	60 - 100	3241	**-0.31**	-	**8.32**	-	**0.72**	-	**0.84**
	100 - 200	1244	**-0.39**	-	**8.8**	-	**0.64**	-	**0.83**

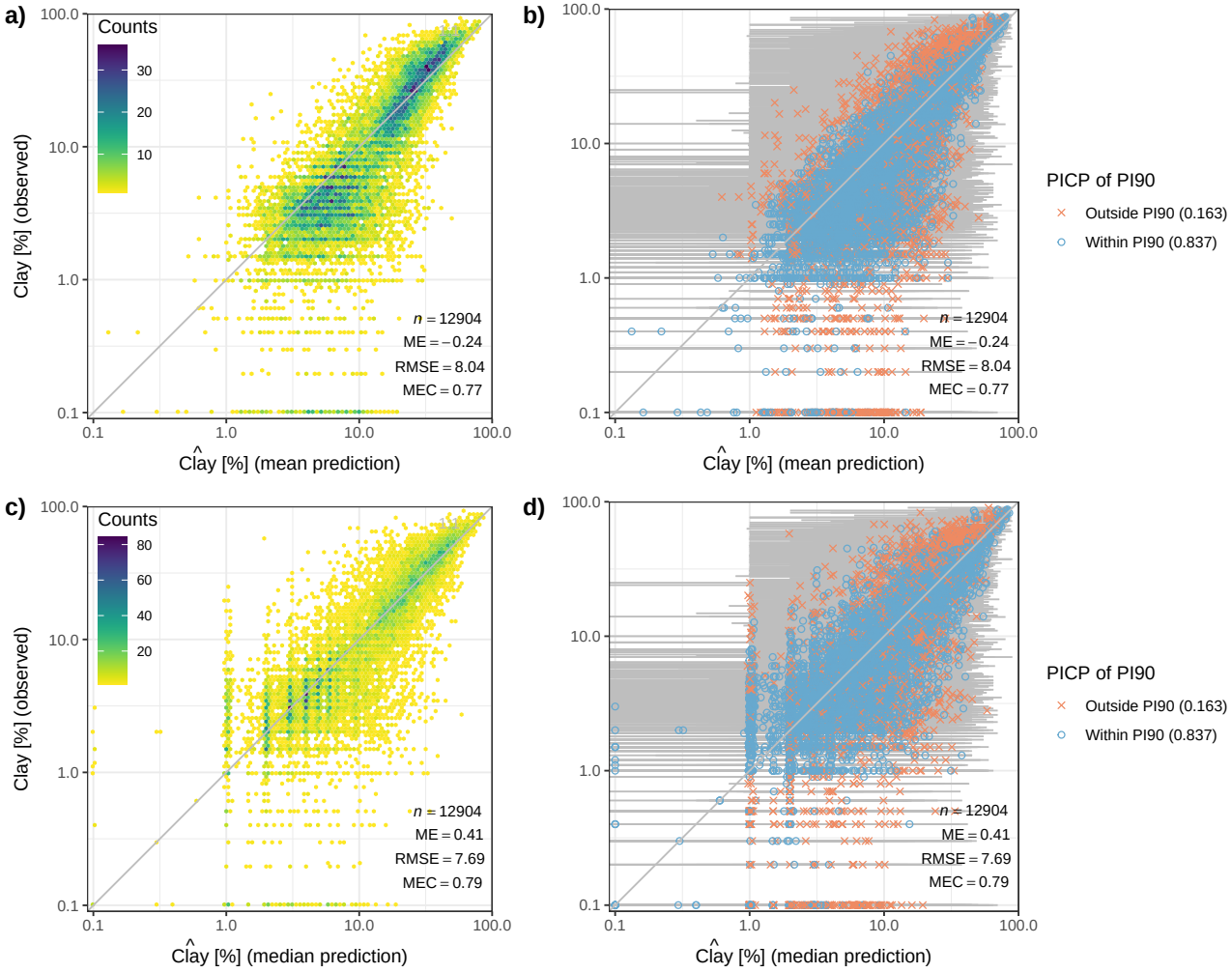

Figure S10. Predicted mean (a & b) and median (c & d) clay content [%] on the x-axis vs. measured clay content [%] on the y-axis (log-scale). Accuracy plots and metrics (ME, RMSE and MEC) were computed using 10-fold cross-validation of PFB laboratory measurements (Sect. 2.6). Plots a & c emphasize point density whereas plots b & d visualize prediction uncertainty (PI90 as error bars) and the PICP90 in the figure legends.

S1.3 Variable importance

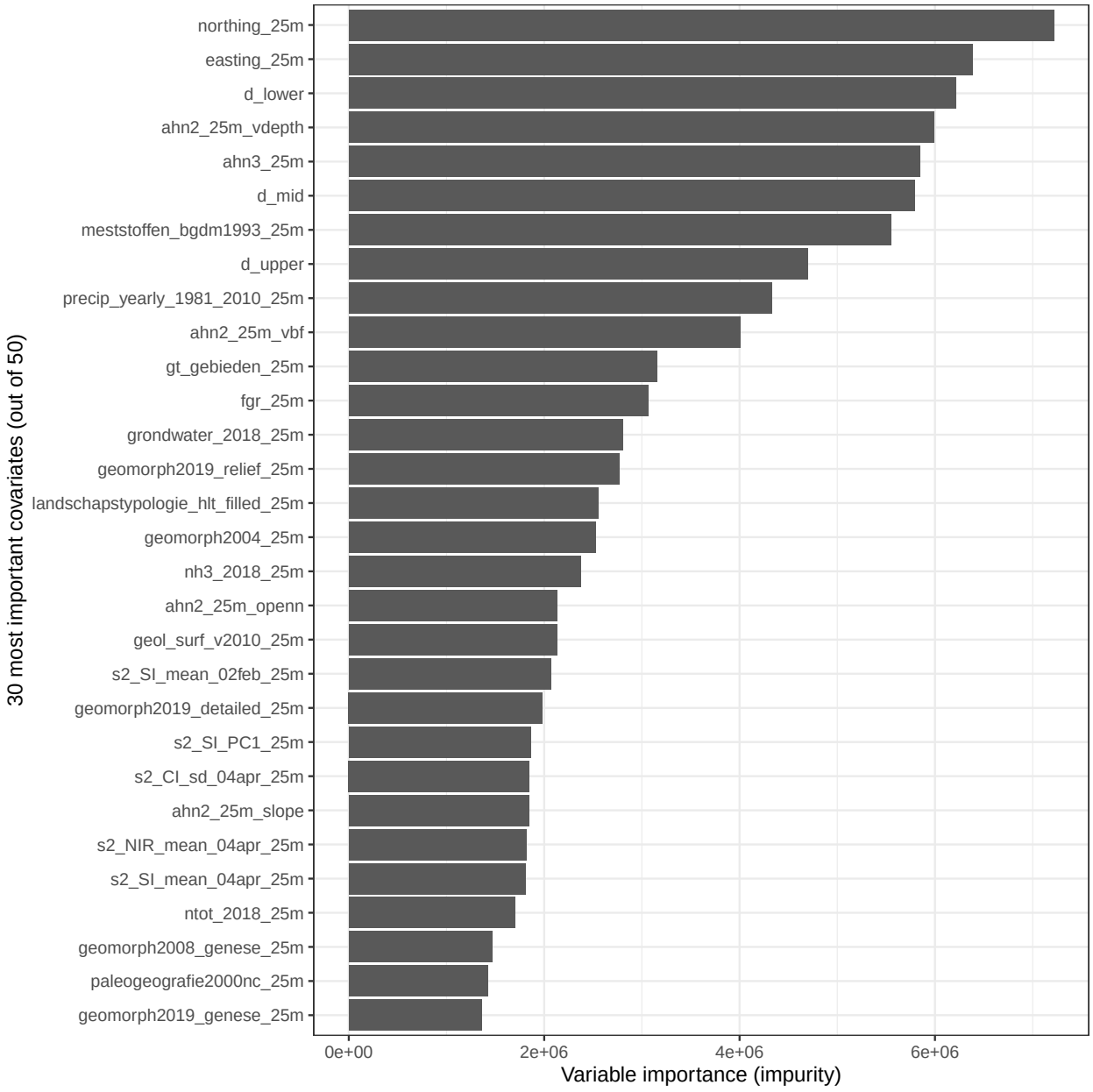

Figure S11. Variable importance for predicting clay content, assessed using the impurity method (Breiman, 2002) (Sect. 2.4). "d_upper", "d_mid" and "d_lower" denote the upper, midpoint and lower boundary of a sampled soil horizon during calibration and target depth layer during prediction (Sect. 2.2, Table 5). Variable importance using impurity favors covariates with more distinct values and is biased against categorical covariates because these have a finite number of binary splits due to limited number of classes (Sandri and Zuccolotto, 2008, 2010).

S2 Silt

S2.1 Prediction maps

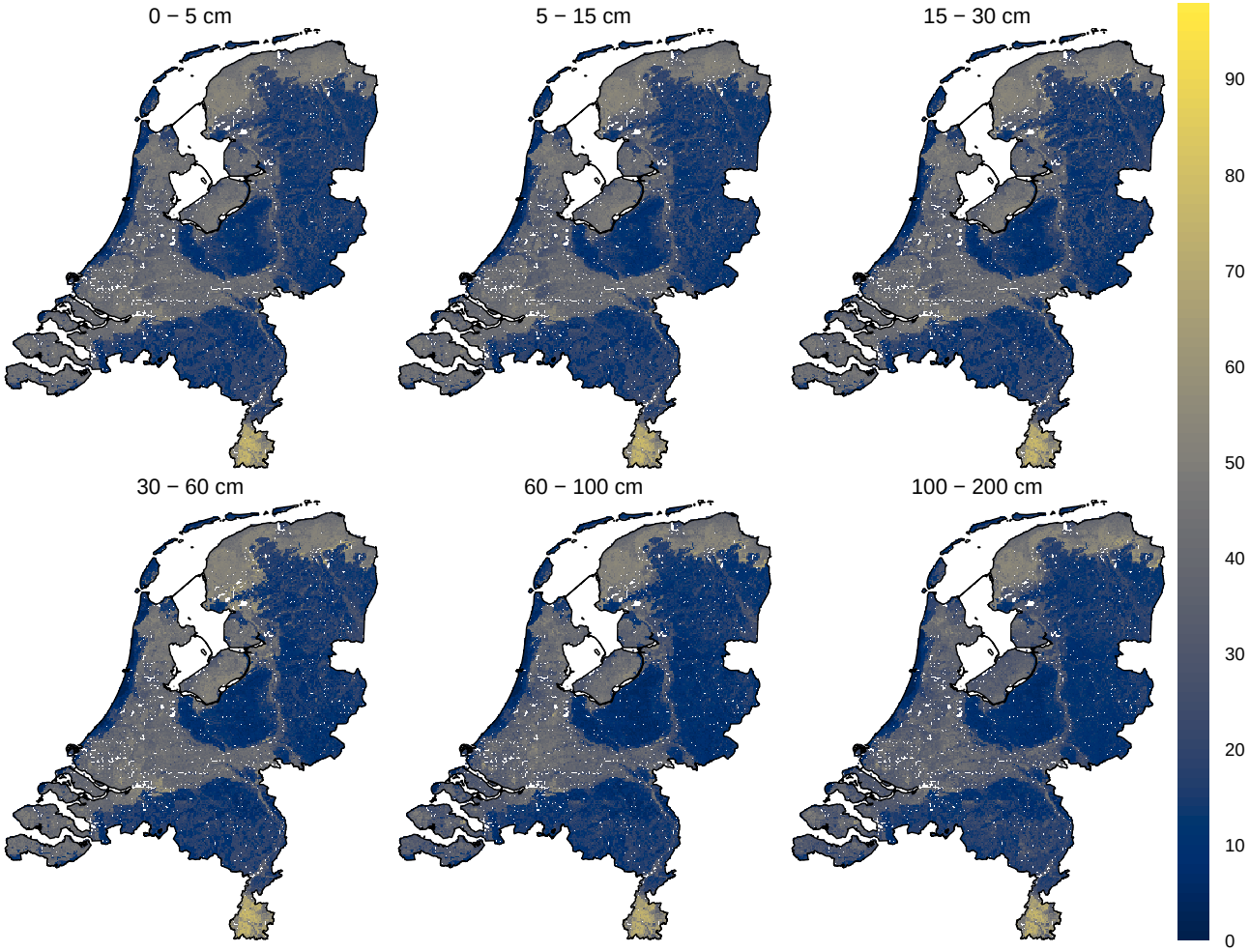

Figure S12. Predicted mean silt content [%] for every GSM depth layer.

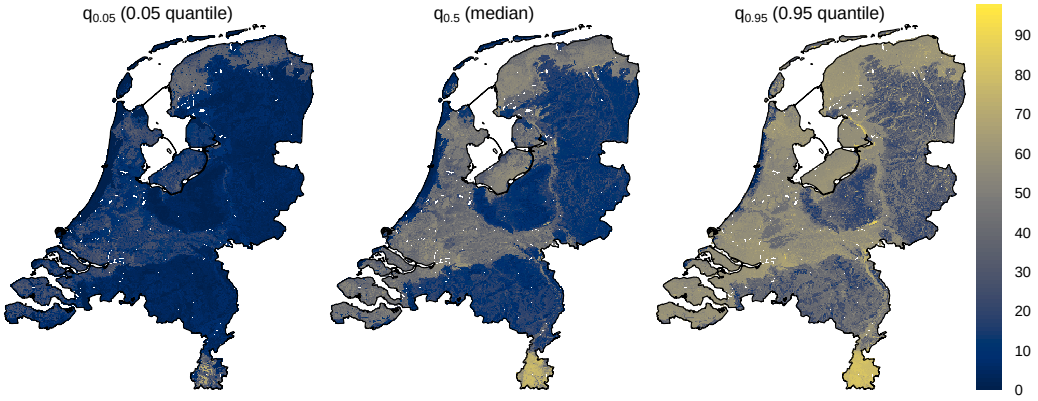

Figure S13. Predicted 5th, 50th (median) and 95th quantile for silt content [%] from 0 - 5 cm depth.

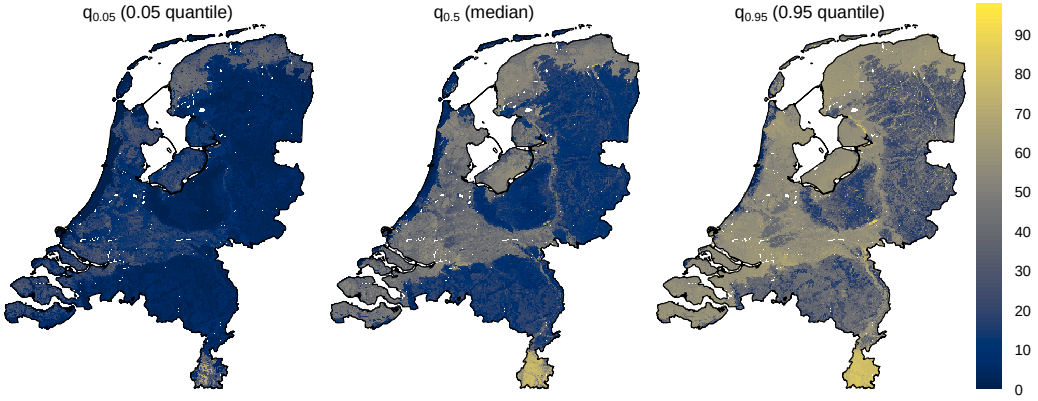

Figure S14. Predicted 5th, 50th (median) and 95th quantile for silt content [%] from 5 - 15 cm depth.

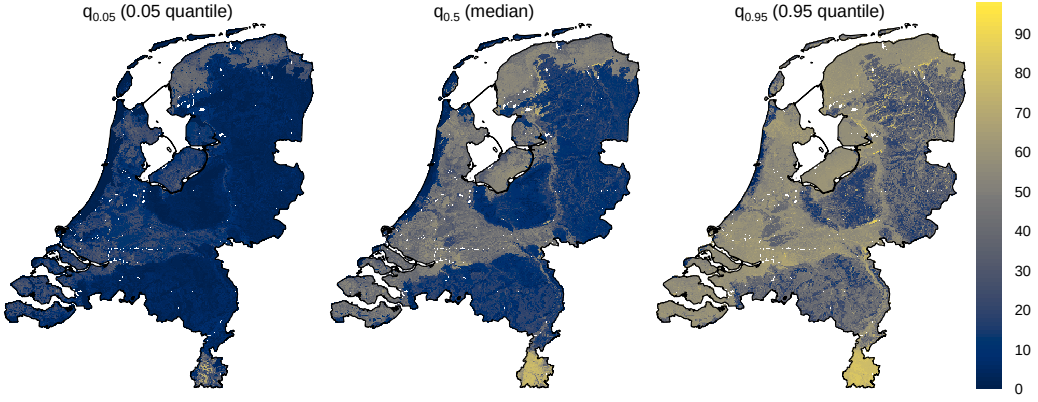

Figure S15. Predicted 5th, 50th (median) and 95th quantile for silt content [%] from 15 - 30 cm depth.

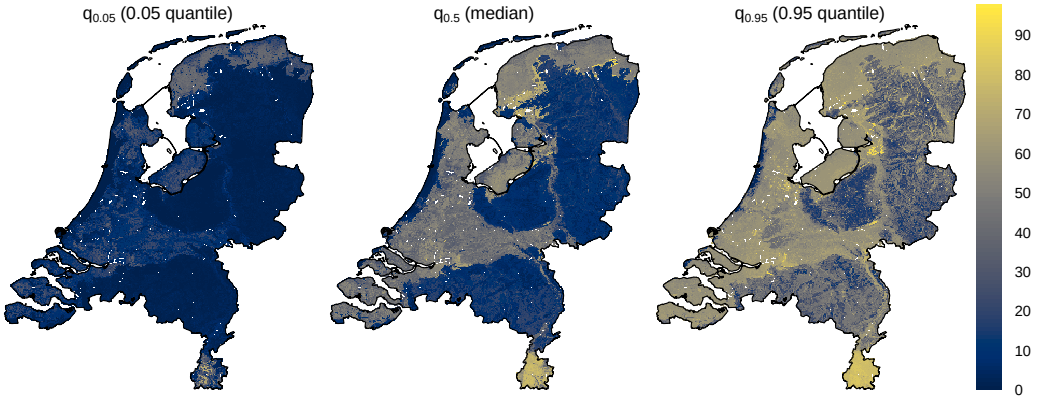

Figure S16. Predicted 5th, 50th (median) and 95th quantile for silt content [%] from 30 - 60 cm depth.

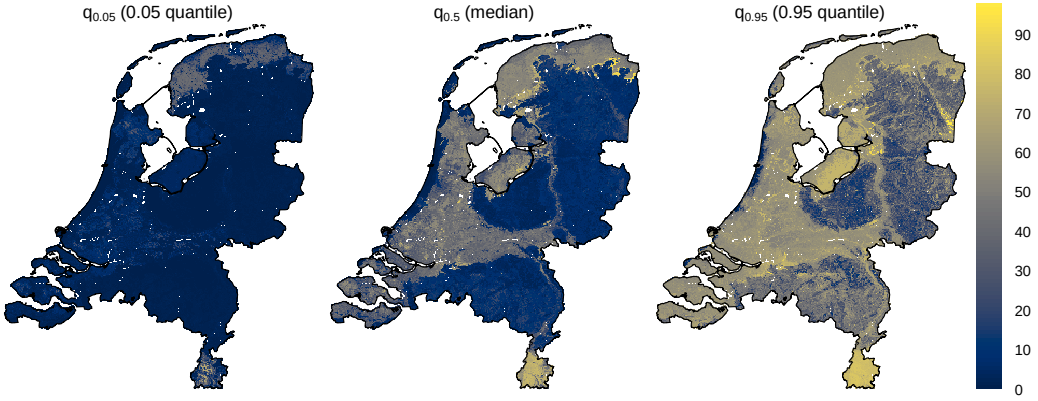

Figure S17. Predicted 5th, 50th (median) and 95th quantile for silt content [%] from 60 - 100 cm depth.

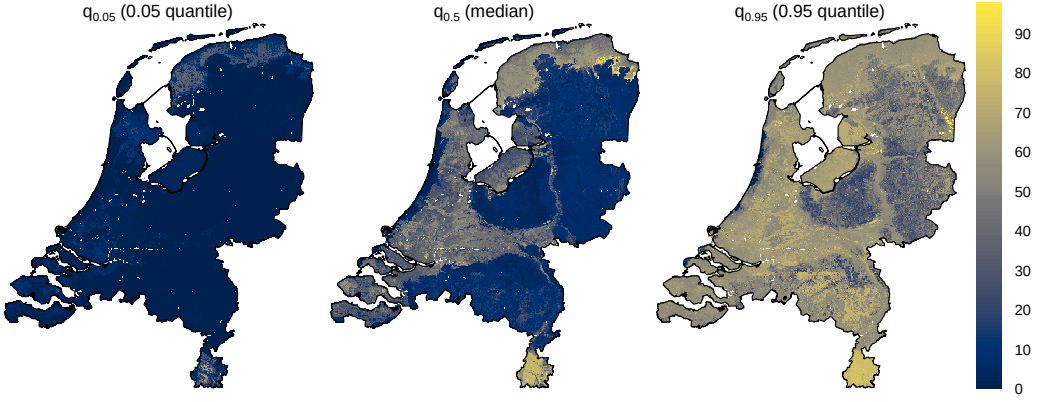

Figure S18. Predicted 5th, 50th (median) and 95th quantile for silt content [%] from 100 - 200 cm depth.

S2.2 Accuracy assessment

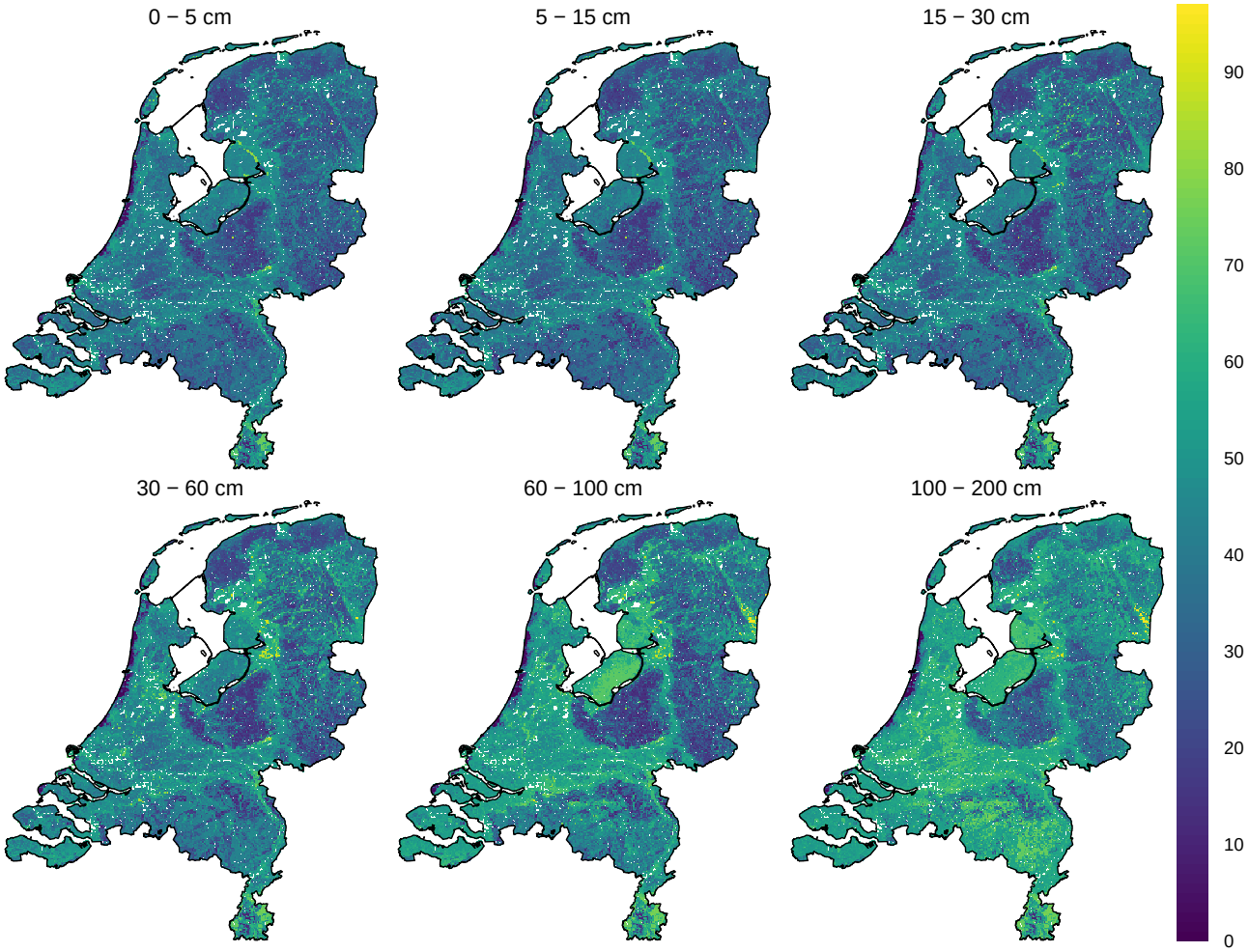

Figure S19. Maps of the PI90 as a measure of prediction uncertainty for silt content [%] for every GSM depth layer.

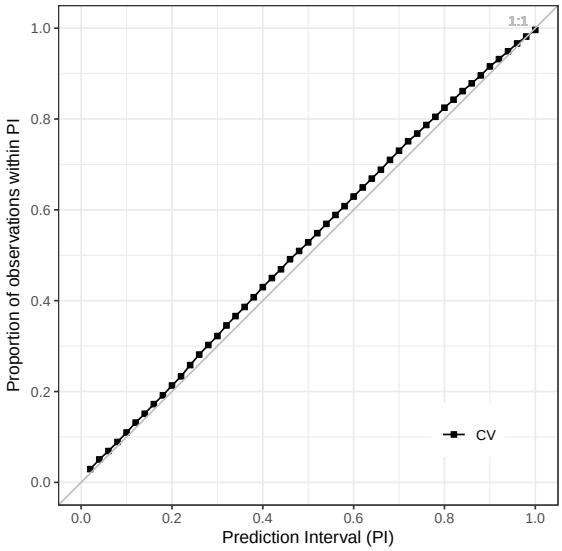

Figure S20. Prediction interval coverage probability (PICP) for prediction intervals between 0.02 and 1 of the PFB laboratory measurements used for 10-fold cross-validation (CV). The closer the points are to the 1:1 line, the more accurate the prediction uncertainty. Lines connecting the points do not represent actual data and are only for visual guidance.

Table S2. Model accuracy metrics of mean and median silt content [%] predictions using 10-fold cross-validation of PFB laboratory measurements (Sect. 2.6). Note that silt content was not measured in LSK or CCNL so design-based inference and computing 95 % confidence intervals (CI95) was not possible (Sect. 2.1.2). ME = mean error; RMSE = root mean squared error; MEC = model efficiency coefficient; PICP90 = prediction interval coverage probability of the PI90.

Prediction	Depth (cm)	n	**ME**	CI95 of ME	**RMSE**	CI95 of RMSE	**MEC**	CI95 of MEC	**PICP90**
	0 - 15	2986	**-0.55**	-	**10.17**	-	**0.7**	-	**0.93**
	15 - 30	1794	**0.05**	-	**10.64**	-	**0.68**	-	**0.92**
Mean	30 - 60	3626	**-0.39**	-	**12.6**	-	**0.62**	-	**0.91**
	60 - 100	3241	**-0.38**	-	**13.84**	-	**0.54**	-	**0.91**
	100 - 200	1244	**0.52**	-	**15.21**	-	**0.51**	-	**0.9**
	0 - 15	2986	**-0.13**	-	**10.61**	-	**0.67**	-	**0.93**
	15 - 30	1794	**0.67**	-	**11.02**	-	**0.66**	-	**0.92**
Median	30 - 60	3626	**0.21**	-	**13.46**	-	**0.57**	-	**0.91**
	60 - 100	3241	**0.83**	-	**14.75**	-	**0.48**	-	**0.91**
	100 - 200	1244	**2.71**	-	**16.21**	-	**0.44**	-	**0.9**

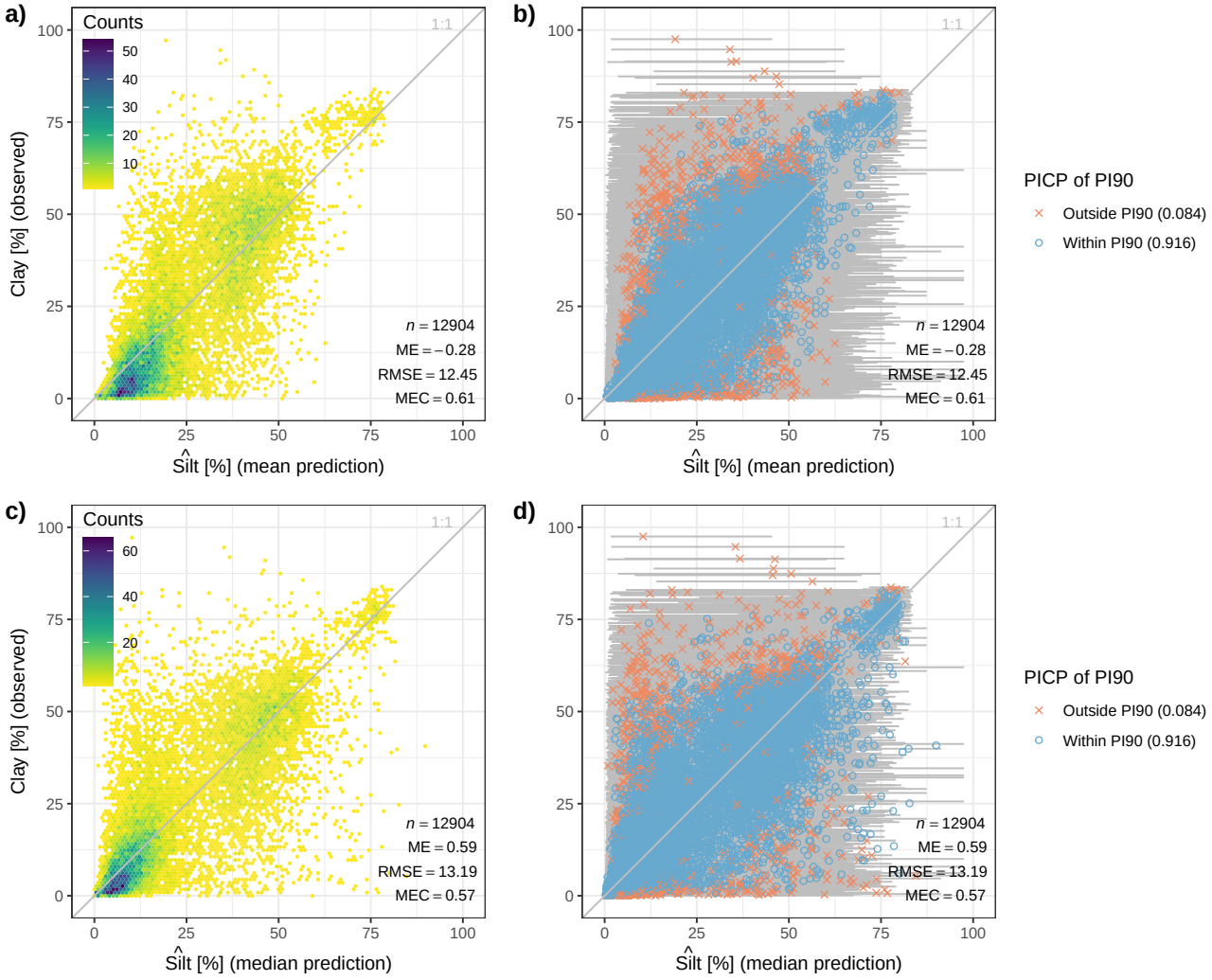

Figure S21. Predicted mean (a & b) and median (c & d) silt content [%] on the x-axis vs. measured silt content [%] on the y-axis. Accuracy plots and metrics (ME, RMSE and MEC) were computed using 10-fold cross-validation of PFB laboratory measurements (Sect. 2.6). Plots a & c emphasize point density whereas plots b & d visualize prediction uncertainty (PI90 as error bars) and the PICP90 in the figure legends.

S2.3 Variable importance

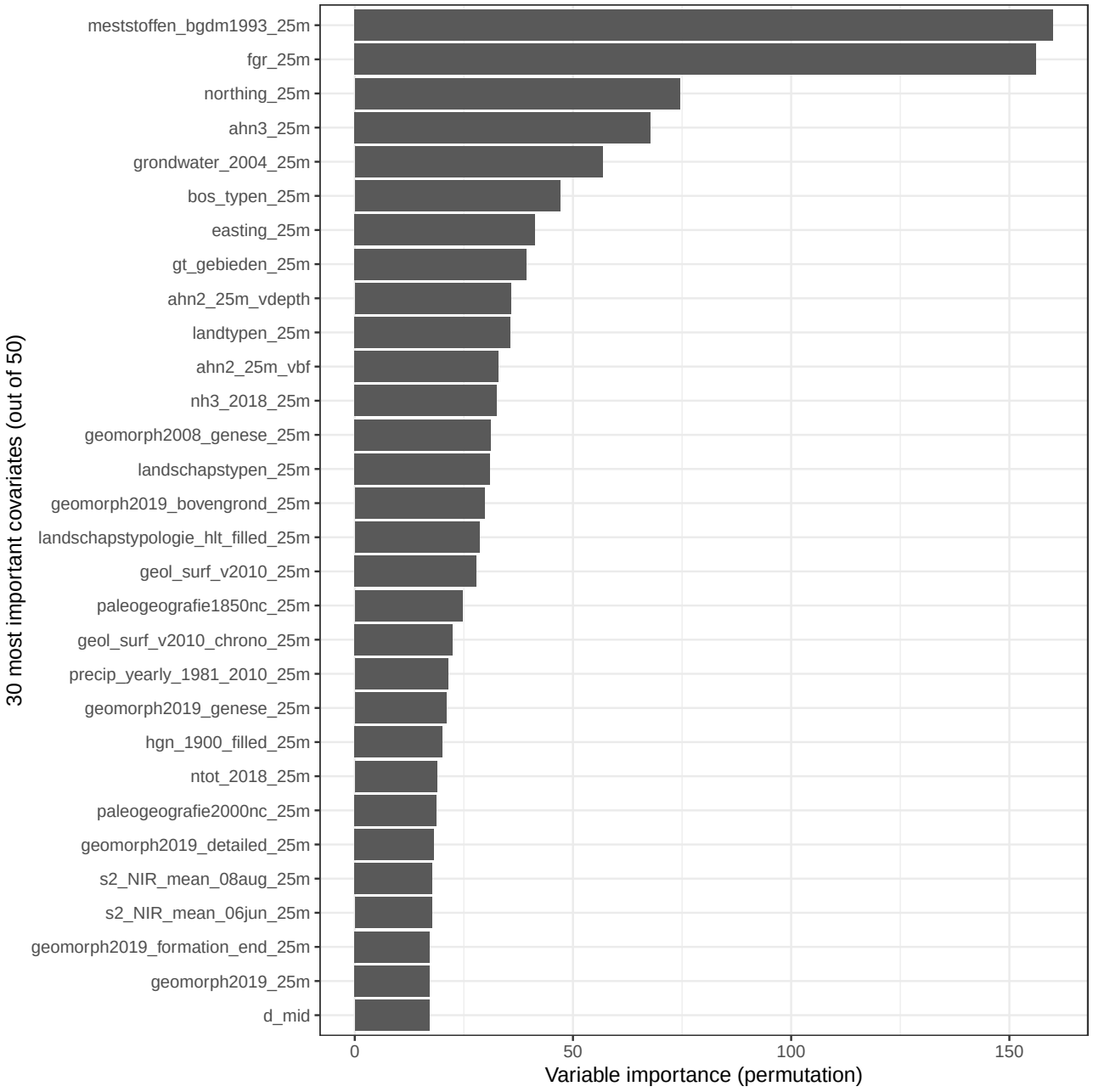

Figure S22. Variable importance for predicting silt content, assessed using the permutation method (Breiman, 2002) (Sect. 2.4). "d_upper", "d_mid" and "d_lower" denote the upper, midpoint and lower boundary of a sampled soil horizon during calibration and target depth layer during prediction (Sect. 2.2, Table 5).

S3 Sand

S3.1 Prediction maps

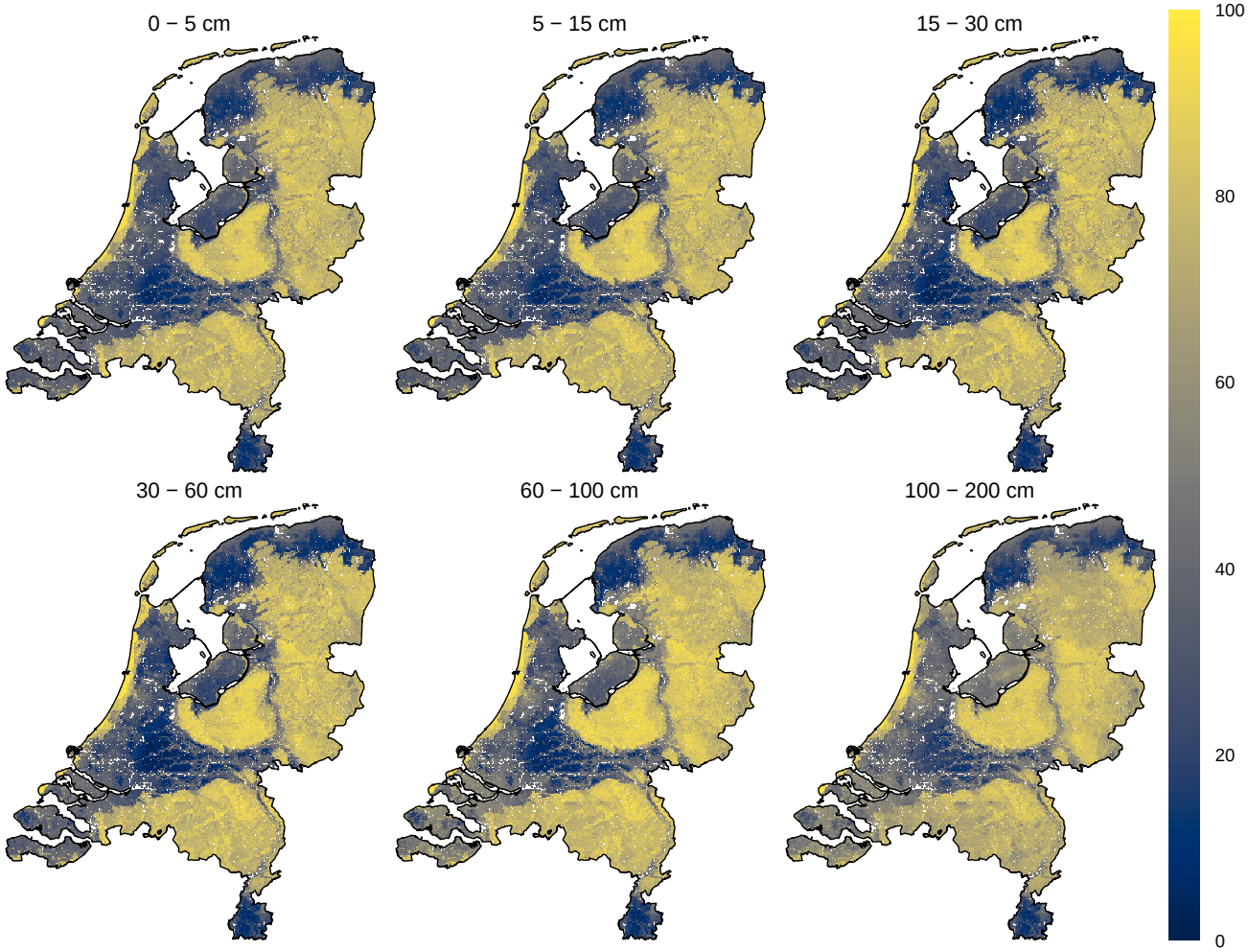

Figure S23. Predicted mean sand content [%] for every GSM depth layer.

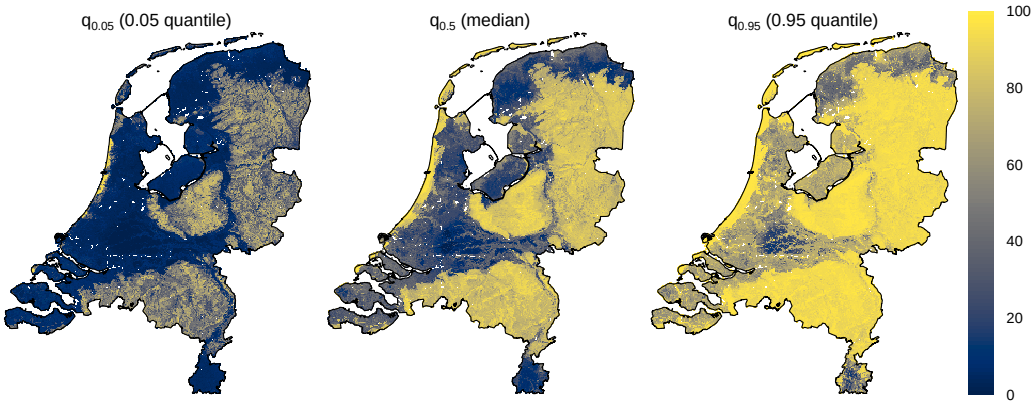

Figure S24. Predicted 5th, 50th (median) and 95th quantile for sand content [%] from 0 - 5 cm depth.

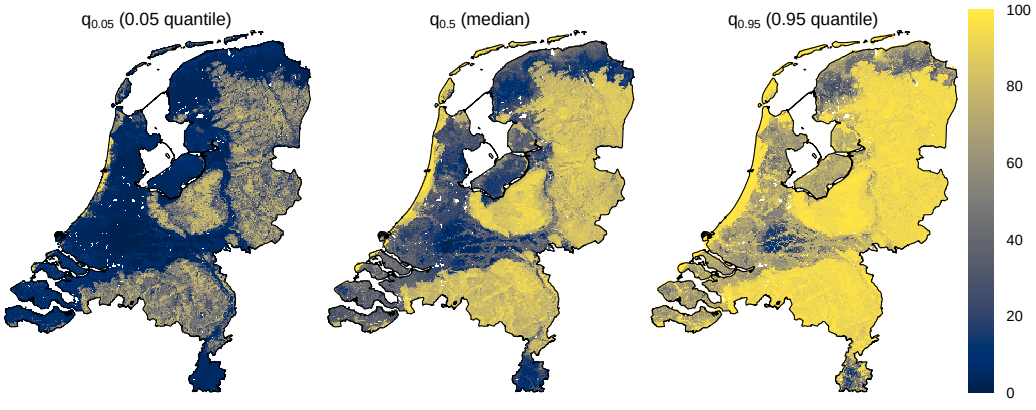

Figure S25. Predicted 5th, 50th (median) and 95th quantile for sand content [%] from 5 - 15 cm depth.

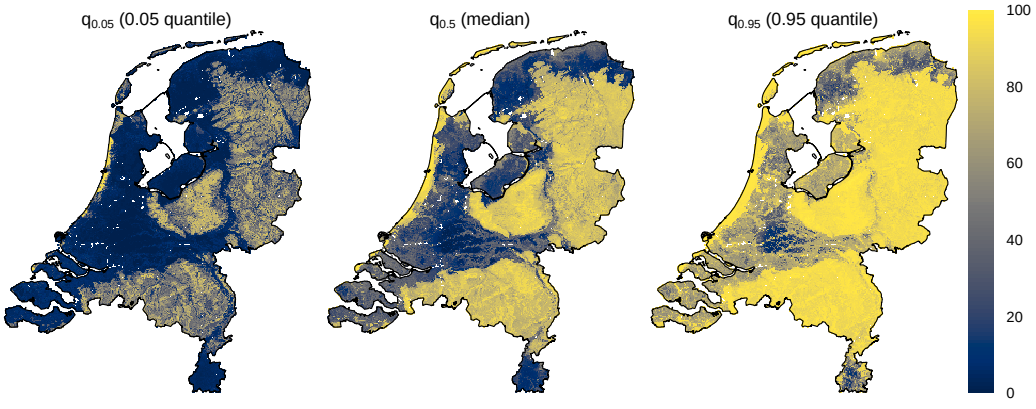

Figure S26. Predicted 5th, 50th (median) and 95th quantile for sand content [%] from 15 - 30 cm depth.

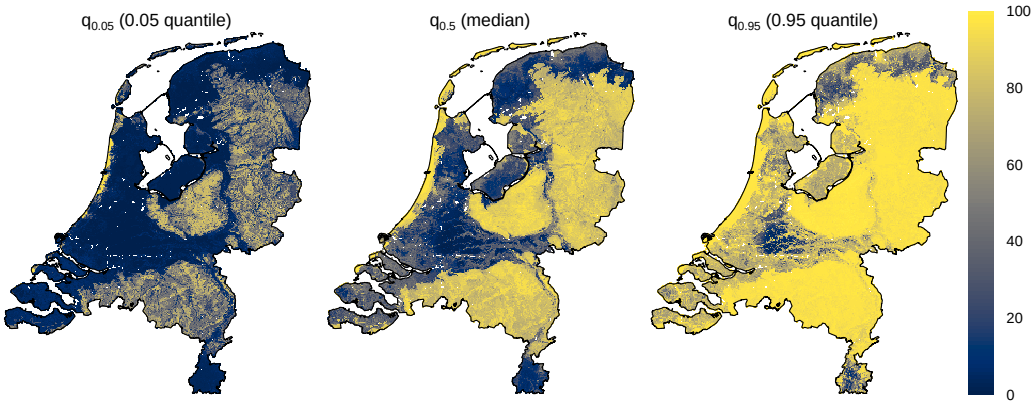

Figure S27. Predicted 5th, 50th (median) and 95th quantile for sand content [%] from 30 - 60 cm depth.

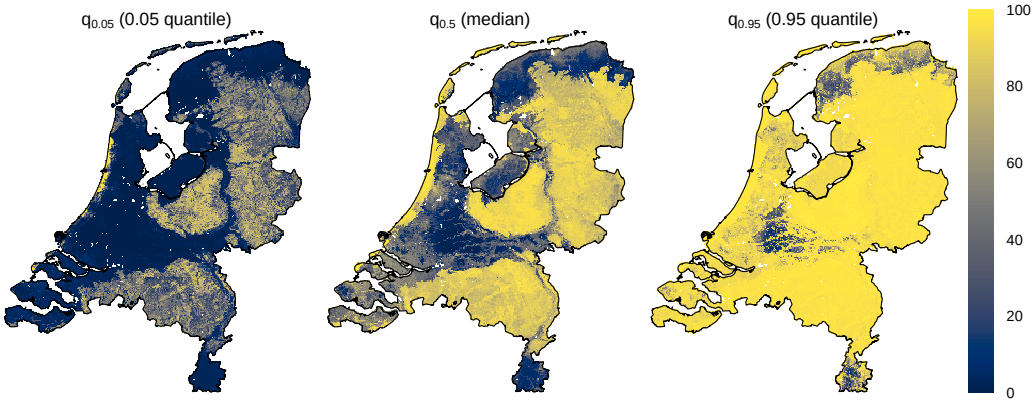

Figure S28. Predicted 5th, 50th (median) and 95th quantile for sand content [%] from 60 - 100 cm depth.

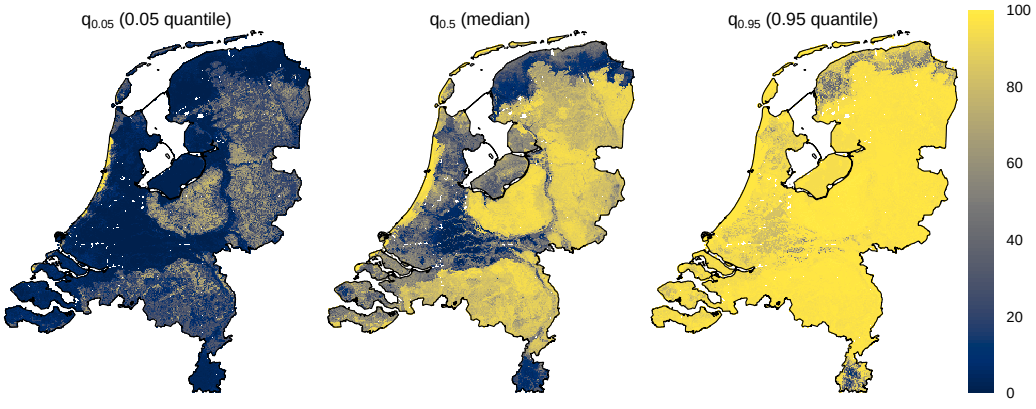

Figure S29. Predicted 5th, 50th (median) and 95th quantile for sand content [%] from 100 - 200 cm depth.

S3.2 Accuracy assessment

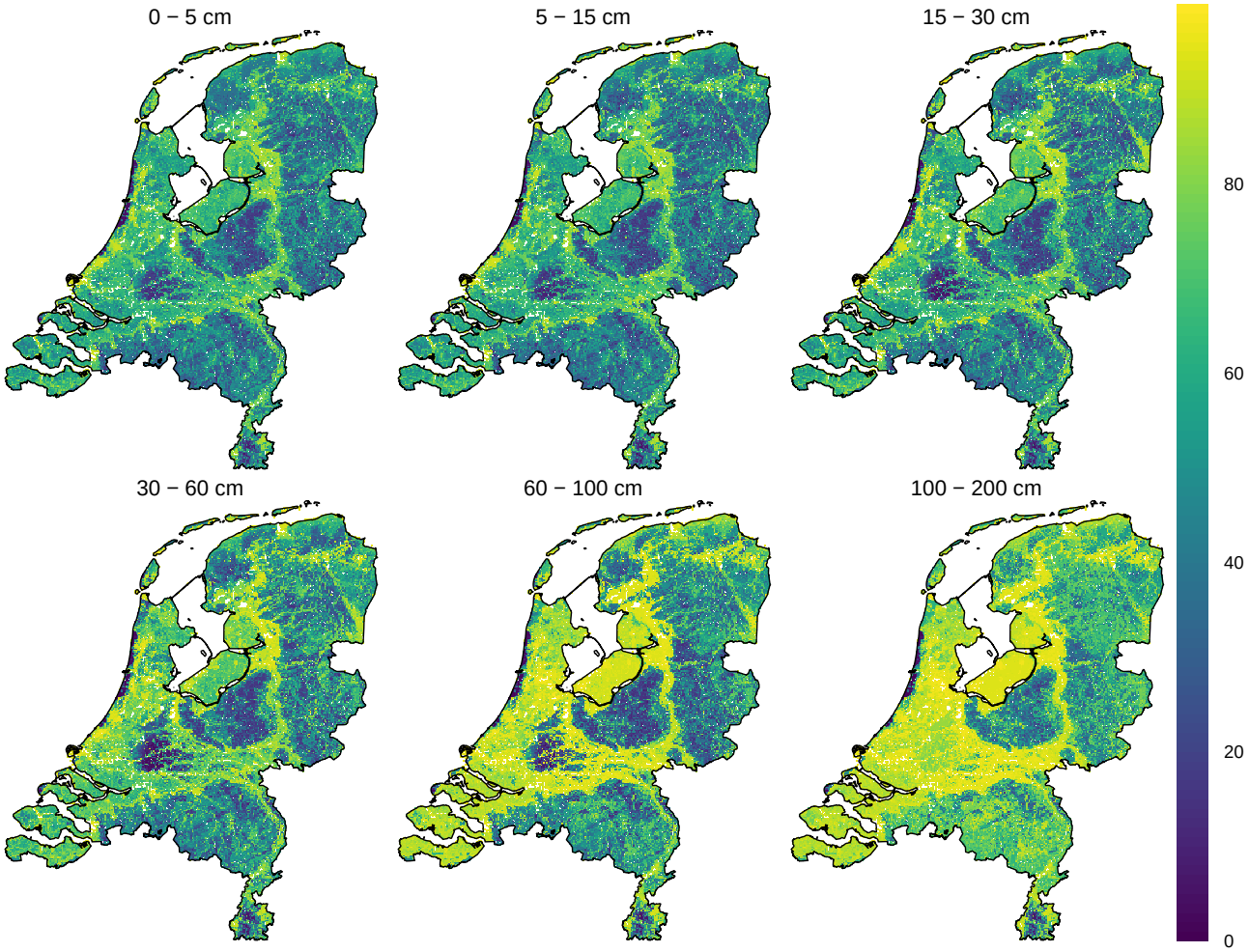

Figure S30. Maps of the PI90 as a measure of prediction uncertainty for sand content [%] for every GSM depth layer.

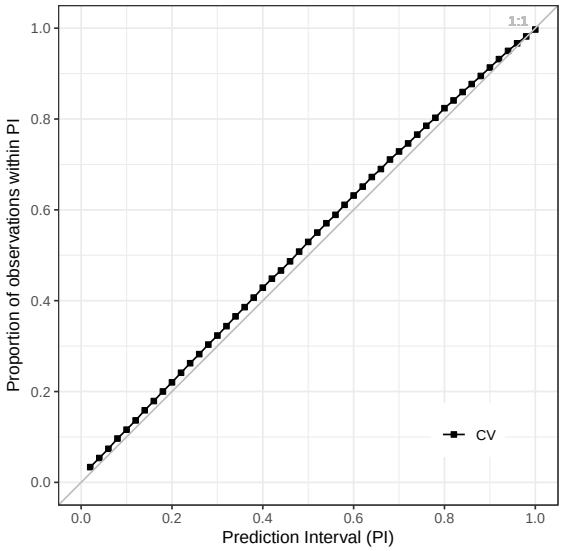

Figure S31. Prediction interval coverage probability (PICP) for prediction intervals between 0.02 and 1 of the PFB laboratory measurements used for 10-fold cross-validation (CV). The closer the points are to the 1:1 line, the more accurate the prediction uncertainty. Lines connecting the points do not represent actual data and are only for visual guidance.

Table S3. Model accuracy metrics of mean and median sand content [%] predictions using 10-fold cross-validation of PFB laboratory measurements (Sect. 2.6). Note that sand content was not measured in LSK or CCNL so design-based inference and computing 95 % confidence intervals (CI95) was not possible (Sect. 2.1.2). ME = mean error; RMSE = root mean squared error; MEC = model efficiency coefficient; PICP90 = prediction interval coverage probability of the PI90.

Prediction	Depth (cm)	n	**ME**	CI95 of ME	**RMSE**	CI95 of RMSE	**MEC**	CI95 of MEC	**PICP90**
	0 - 15	2986	**-0.35**	-	**13.66**	-	**0.8**	-	**0.93**
	15 - 30	1794	**-1.17**	-	**14.31**	-	**0.81**	-	**0.91**
Mean	30 - 60	3626	**0.26**	-	**17.19**	-	**0.75**	-	**0.91**
	60 - 100	3241	**1.61**	-	**18.99**	-	**0.67**	-	**0.91**
	100 - 200	1244	**1.16**	-	**20.75**	-	**0.61**	-	**0.9**
	0 - 15	2986	**-1.23**	-	**13.82**	-	**0.79**	-	**0.93**
	15 - 30	1794	**-2.13**	-	**14.23**	-	**0.81**	-	**0.91**
Median	30 - 60	3626	**-0.95**	-	**17.05**	-	**0.75**	-	**0.91**
	60 - 100	3241	**-0.42**	-	**18.97**	-	**0.67**	-	**0.91**
	100 - 200	1244	**-2.21**	-	**20.57**	-	**0.61**	-	**0.9**

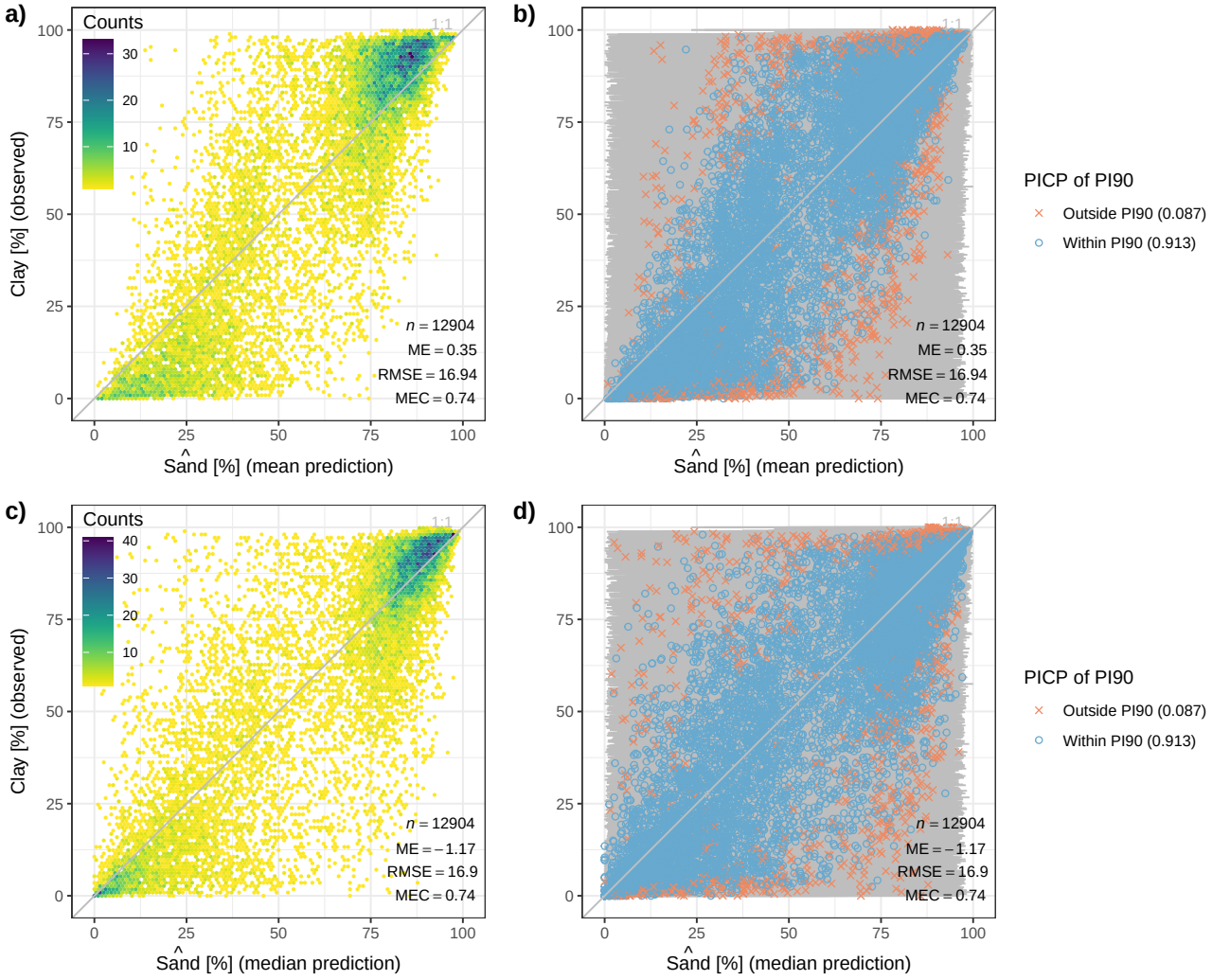

Figure S32. Predicted mean (a & b) and median (c & d) sand content [%] on the x-axis vs. measured sand content [%] on the y-axis. Accuracy plots and metrics (ME, RMSE and MEC) were computed using 10-fold cross-validation of PFB laboratory measurements (Sect. 2.6). Plots a & c emphasize point density whereas plots b & d visualize prediction uncertainty (PI90 as error bars) and the PICP90 in the figure legends.

S3.3 Variable importance

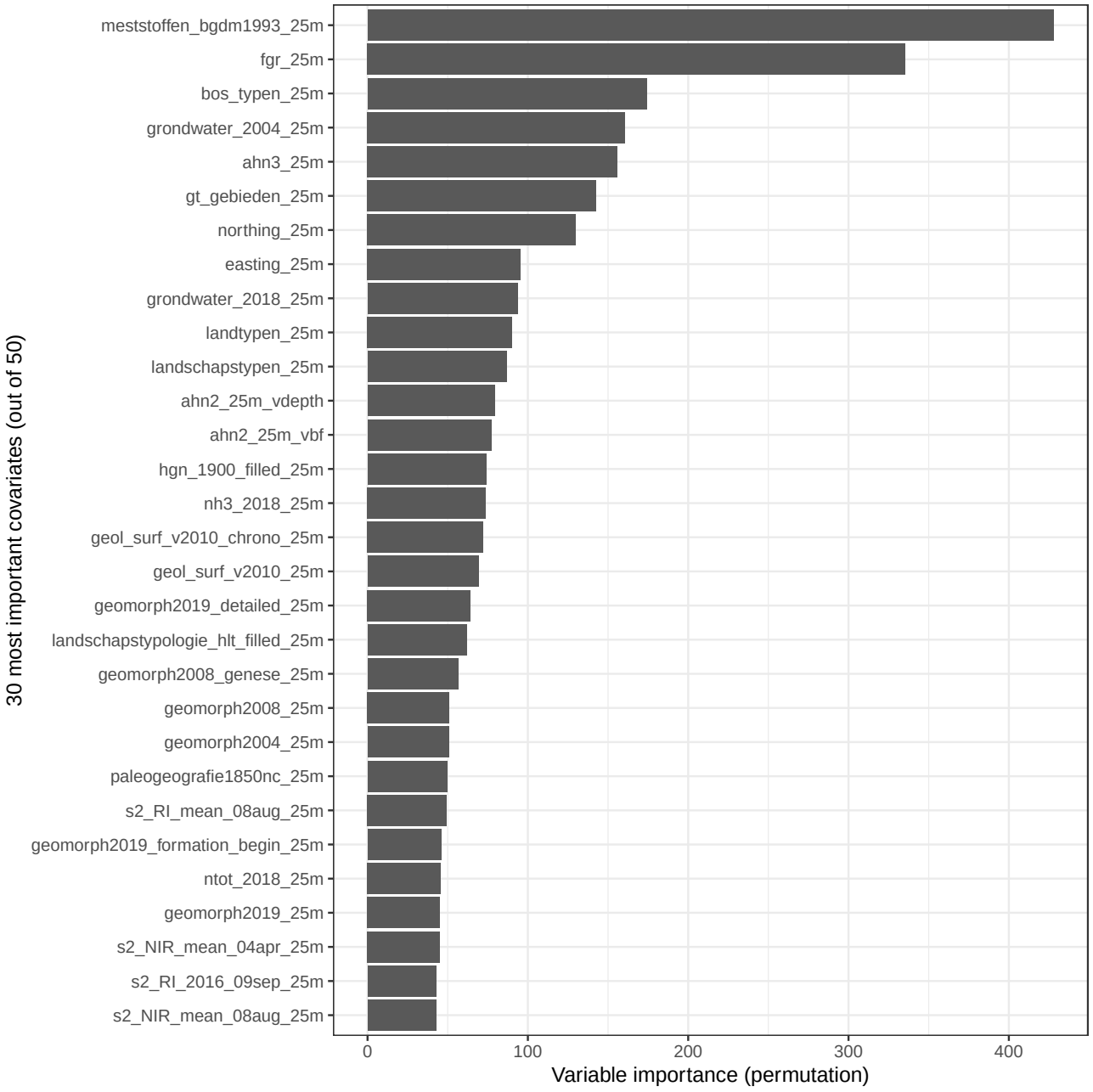

Figure S33. Variable importance for predicting sand content, assessed using the permutation method (Breiman, 2002) (Sect. 2.4). "d_upper", "d_mid" and "d_lower" denote the upper, midpoint and lower boundary of a sampled soil horizon during calibration and target depth layer during prediction (Sect. 2.2, Table 5).

S4 Bulk density (BD)

S4.1 Prediction maps

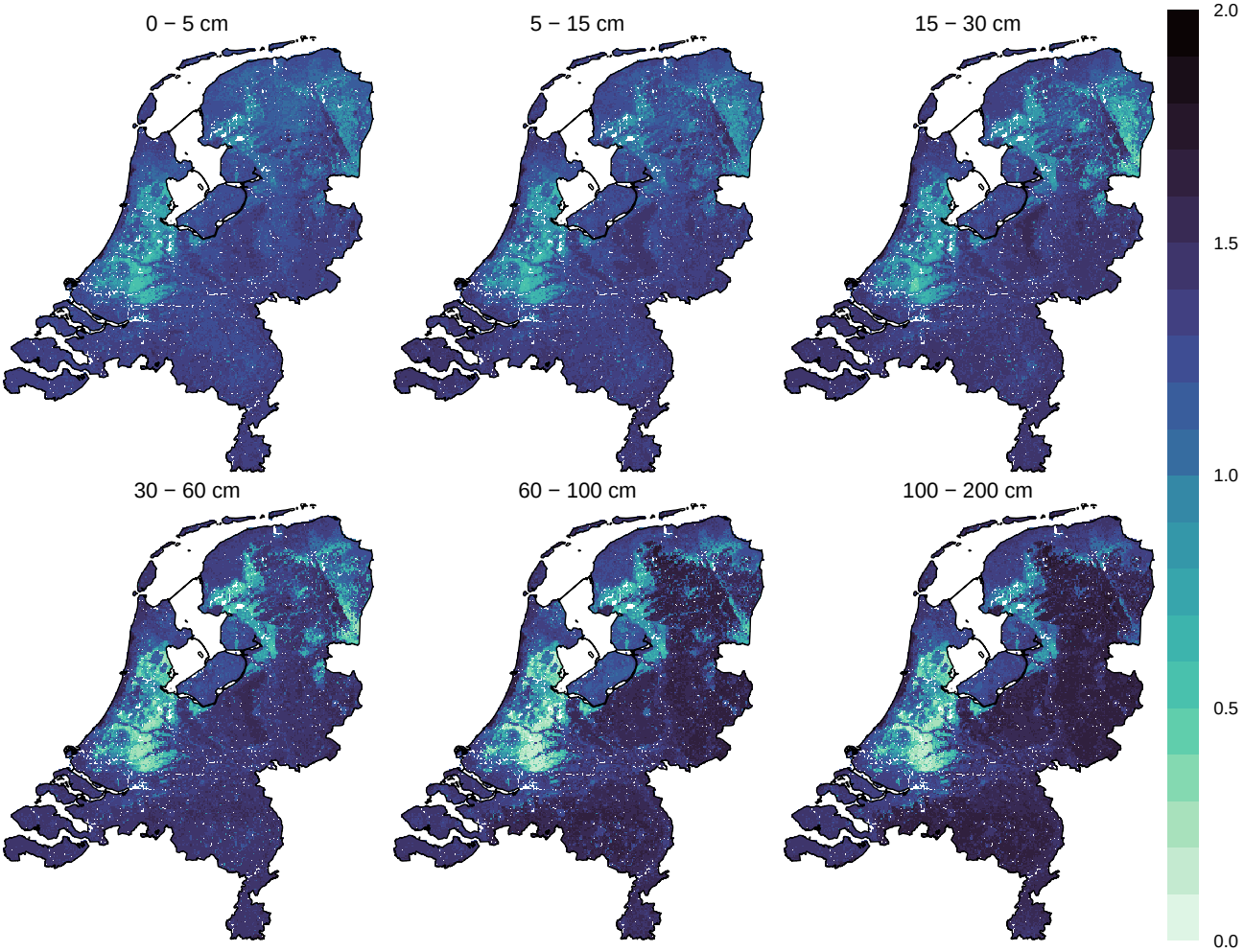

Figure S34. Predicted mean BD [g/cm^3] for every GSM depth layer.

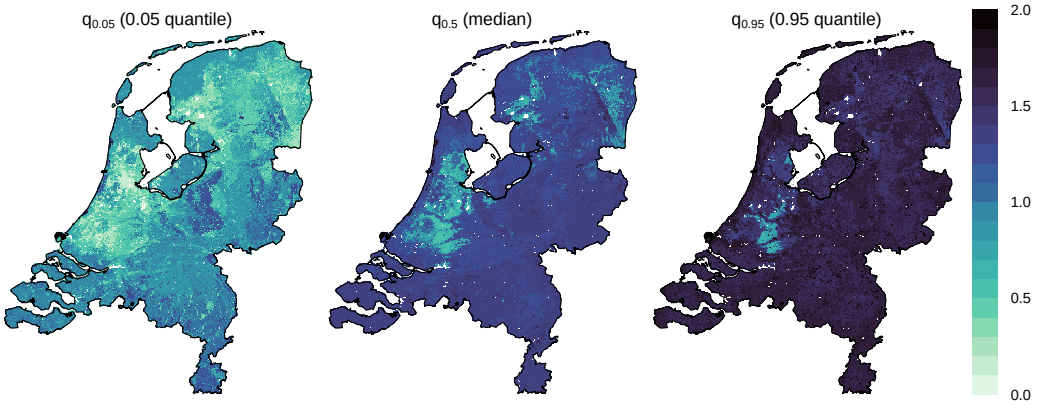

Figure S35. Predicted 5th, 50th (median) and 95th quantile for BD [g/cm^3] from 0 - 5 cm depth.

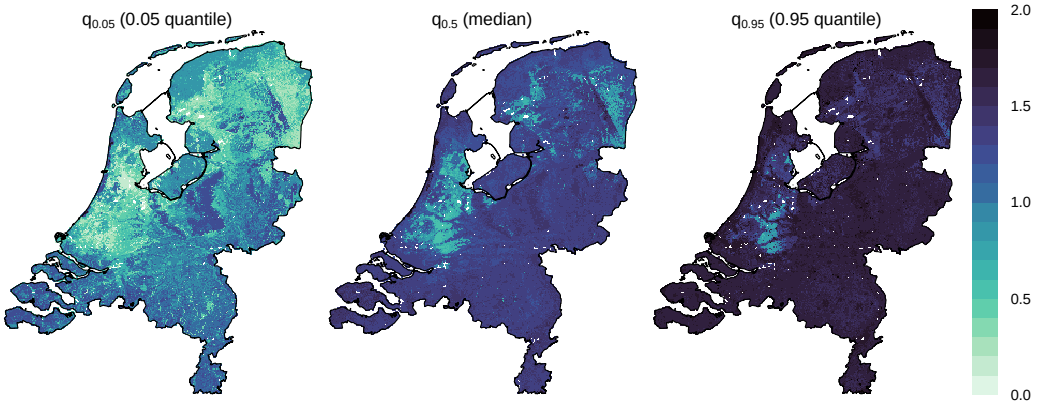

Figure S36. Predicted 5th, 50th (median) and 95th quantile for BD [g/cm^3] from 5 - 15 cm depth.

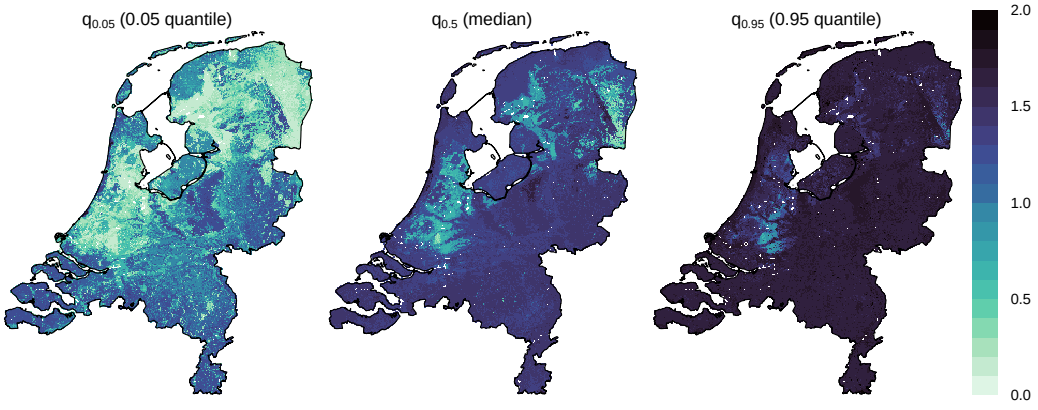

Figure S37. Predicted 5th, 50th (median) and 95th quantile for BD [g/cm^3] from 15 - 30 cm depth.

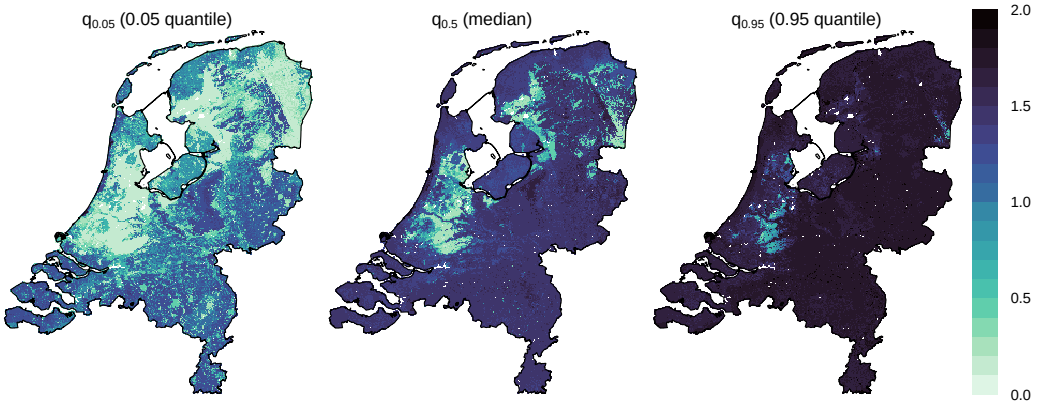

Figure S38. Predicted 5th, 50th (median) and 95th quantile for BD [g/cm^3] from 30 - 60 cm depth.

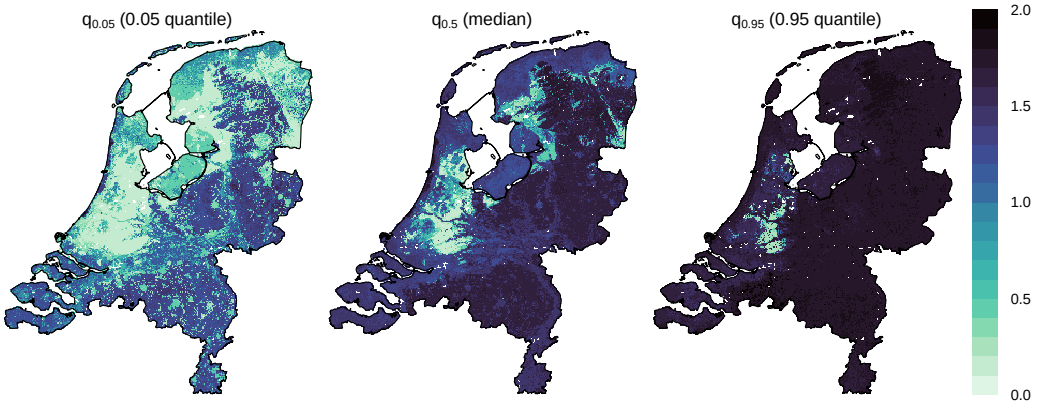

Figure S39. Predicted 5th, 50th (median) and 95th quantile for BD [g/cm^3] from 60 - 100 cm depth.

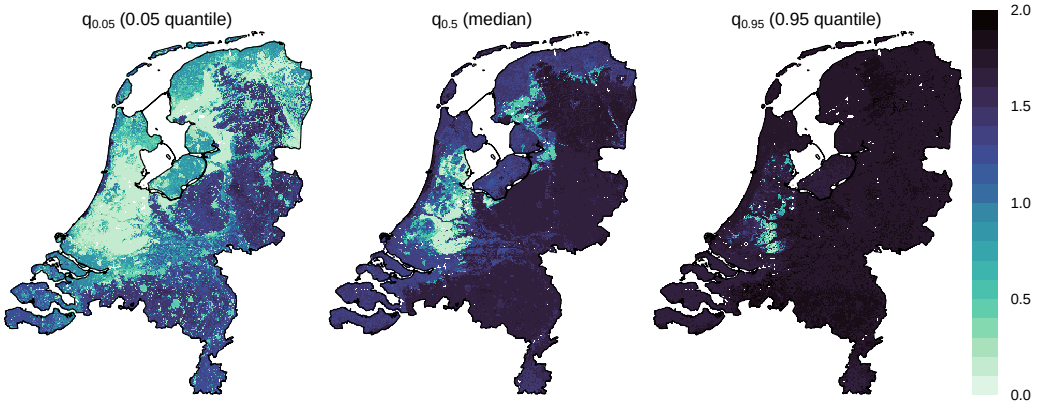

Figure S40. Predicted 5th, 50th (median) and 95th quantile for BD [g/cm^3] from 100 - 200 cm depth.

S4.2 Accuracy assessment

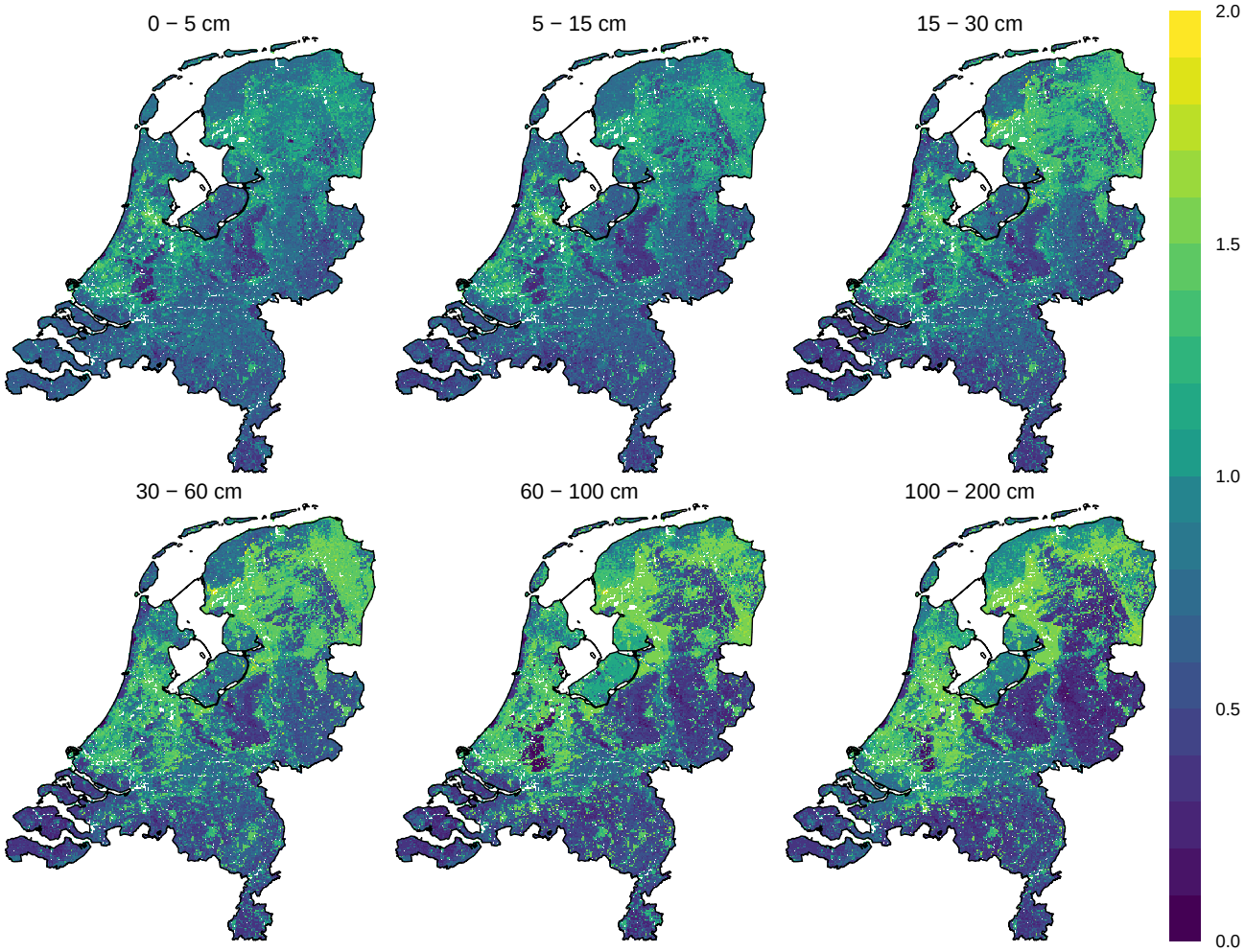

Figure S41. Maps of the PI90 as a measure of prediction uncertainty for BD [g/cm^3] for every GSM depth layer.

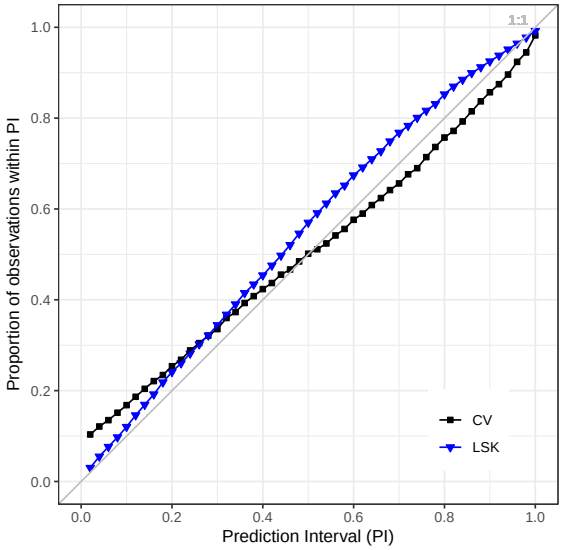

Figure S42. Prediction interval coverage probability (PICP) for prediction intervals between 0.02 and 1 of data used for different statistical validation strategies: 10-fold cross-validation of PFB laboratory measurements (CV) and LSK laboratory measurements (Sect. 2.6). The closer the points are to the 1:1 line, the more accurate the prediction uncertainty. Lines connecting the points do not represent actual data and are only for visual guidance.

Table S4. Model accuracy metrics of mean and median BD [g/cm^3] predictions using design-based inference of LSK laboratory measurements (Sect. 2.6). ME = mean error; RMSE = root mean squared error; MEC = model efficiency coefficient; PICP90 = prediction interval coverage probability of the PI90. The lower and upper 97.5% confidence limits of the accuracy metrics were computed using design-based inference according to de Gruijter et al. (2006) and Sect. 2.6.2 of Helfenstein et al. (2022). Together, the lower and upper 97.5% confidence limits give the 95% confidence interval (CI95).

Prediction	Depth (cm)	n	**ME**	CI95 of ME	**RMSE**	CI95 of RMSE	**MEC**	CI95 of MEC	**PICP90**
	0 - 15	1250	**0**	-0.01, 0.01	**0.21**	0.2, 0.22	**0.39**	0.34, 0.45	**0.94**
	15 - 30	245	**0.04**	0, 0.07	**0.24**	0.19, 0.28	**0.78**	0.74, 0.82	**0.91**
Mean	30 - 60	1321	**0.02**	0, 0.03	**0.28**	0.26, 0.29	**0.54**	0.49, 0.59	**0.95**
	60 - 100	1244	**0.01**	-0.01, 0.02	**0.31**	0.29, 0.33	**0.49**	0.43, 0.55	**0.92**
	100 - 200	843	**-0.03**	-0.05, 0	**0.36**	0.33, 0.38	**0.47**	0.41, 0.54	**0.88**
	0 - 15	1250	**-0.03**	-0.05, -0.02	**0.22**	0.2, 0.23	**0.34**	0.27, 0.41	**0.94**
	15 - 30	245	**-0.01**	-0.04, 0.03	**0.24**	0.17, 0.29	**0.78**	0.72, 0.83	**0.91**
Median	30 - 60	1321	**-0.01**	-0.03, 0	**0.28**	0.26, 0.29	**0.54**	0.48, 0.61	**0.95**
	60 - 100	1244	**-0.03**	-0.05, -0.01	**0.32**	0.3, 0.35	**0.44**	0.37, 0.53	**0.92**
	100 - 200	843	**-0.06**	-0.09, -0.04	**0.38**	0.34, 0.41	**0.41**	0.32, 0.51	**0.88**

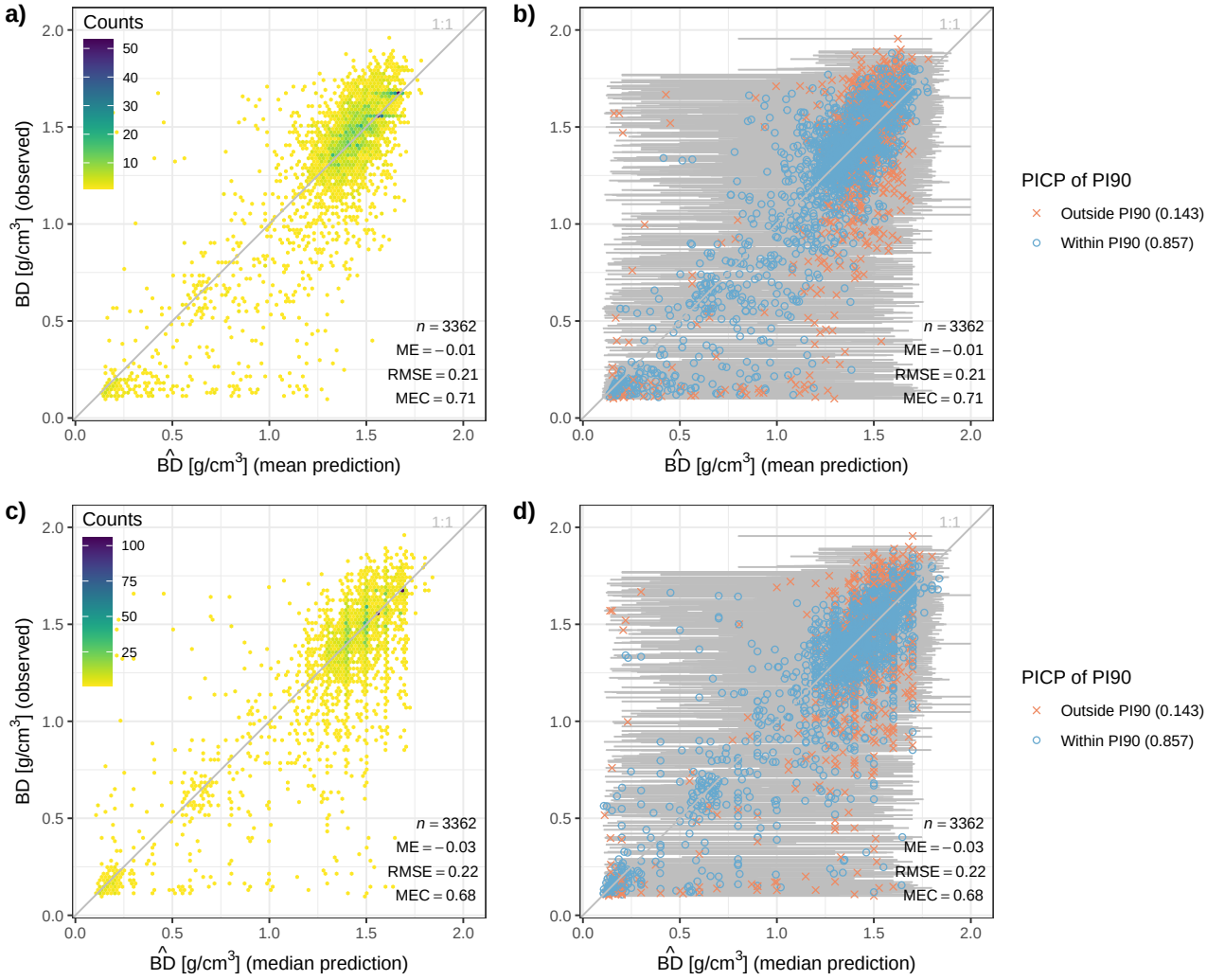

Figure S43. Predicted mean (a & b) and median (c & d) BD [g/cm^3] on the x-axis vs. measured BD [g/cm^3] on the y-axis. Accuracy plots and metrics (ME, RMSE and MEC) were computed using 10-fold cross-validation of PFB laboratory measurements (Sect. 2.6). Plots a & c emphasize point density whereas plots b & d visualize prediction uncertainty (PI90 as error bars) and the PICP90 in the figure legends.

S4.3 Variable importance

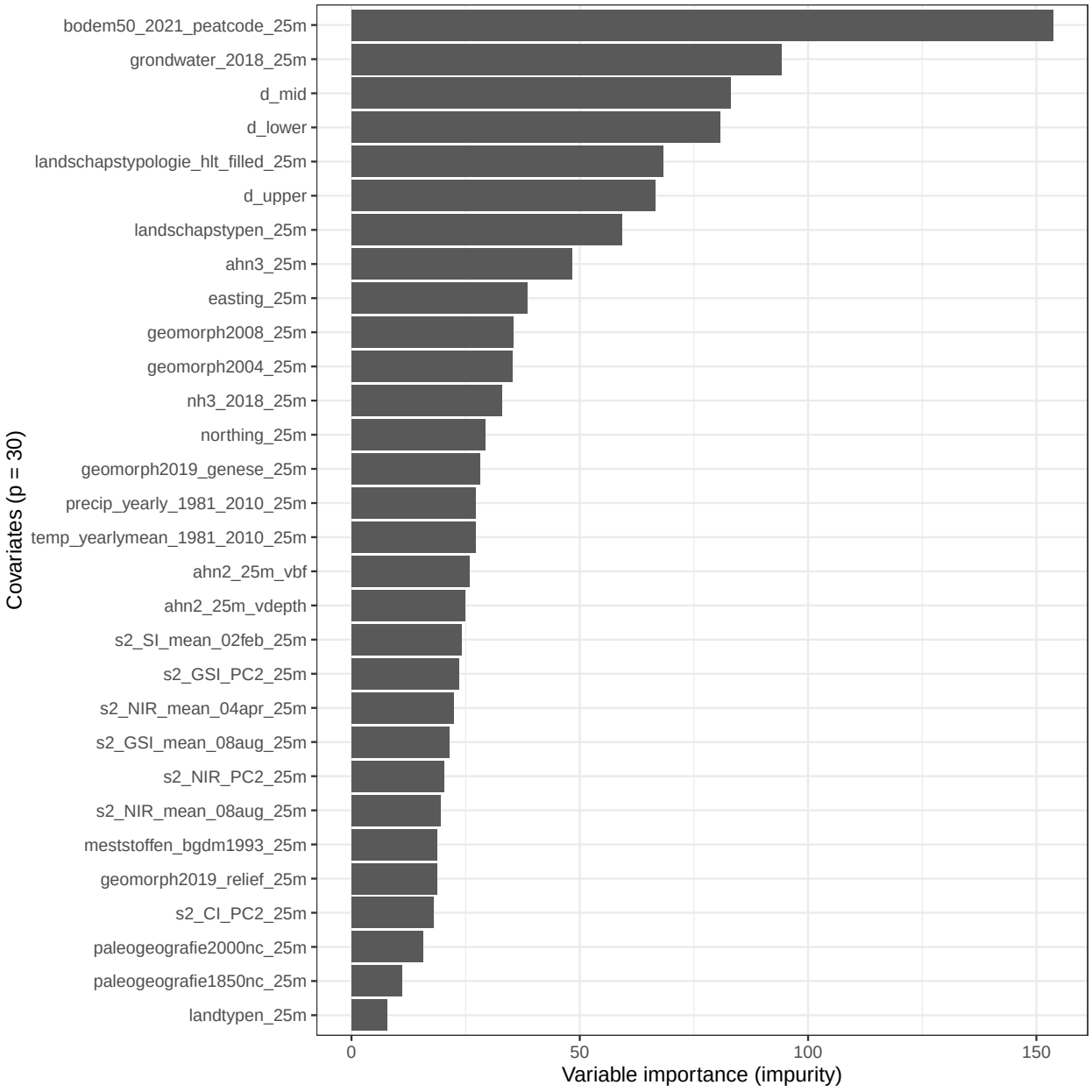

Figure S44. Variable importance for predicting BD, assessed using the impurity method (Breiman, 2002) (Sect. 2.4). "d_upper", "d_mid" and "d_lower" denote the upper, midpoint and lower boundary of a sampled soil horizon during calibration and target depth layer during prediction (Sect. 2.2, Table 5). Variable importance using impurity favors covariates with more distinct values and is biased against categorical covariates because these have a finite number of binary splits due to limited number of classes (Sandri and Zuccolotto, 2008, 2010).

S5 Soil organic matter (SOM)

S5.1 Prediction maps

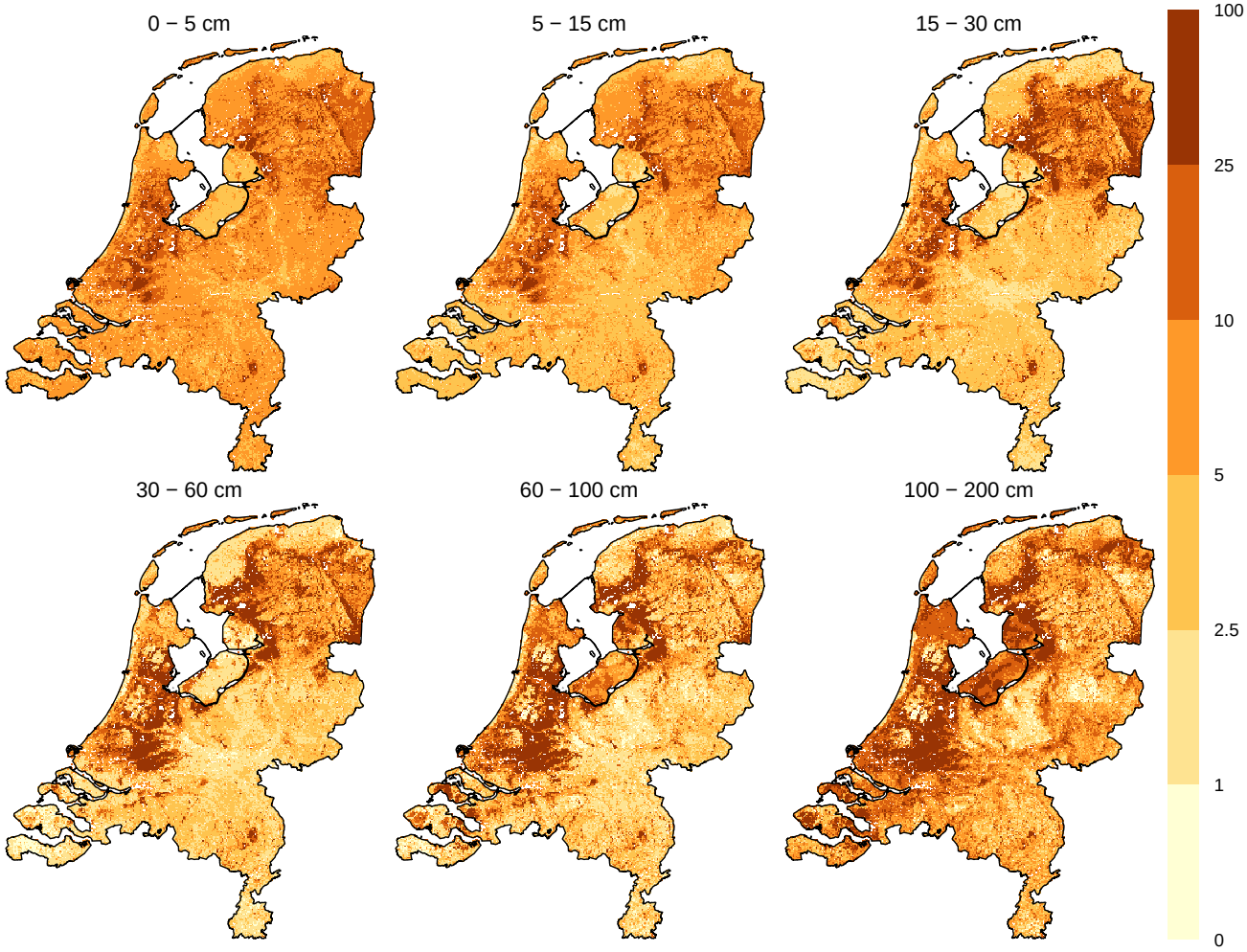

Figure S45. Predicted mean SOM [%] for every GSM depth layer in 2023.

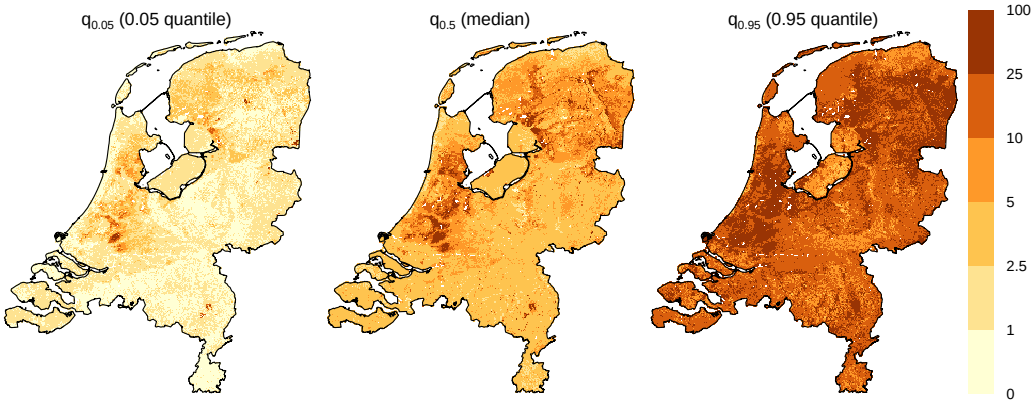

Figure S46. Predicted 5th, 50th (median) and 95th quantile for SOM [%] from 0 - 5 cm depth in 2023.

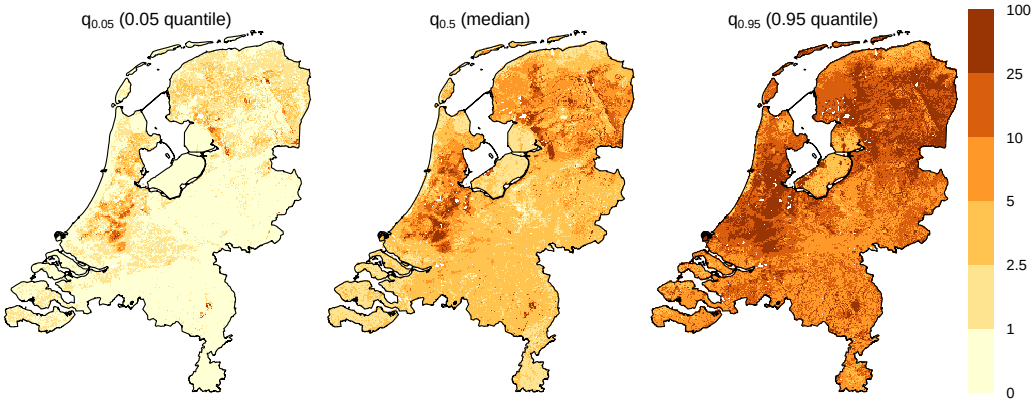

Figure S47. Predicted 5th, 50th (median) and 95th quantile for SOM [%] from 5 - 15 cm depth in 2023.

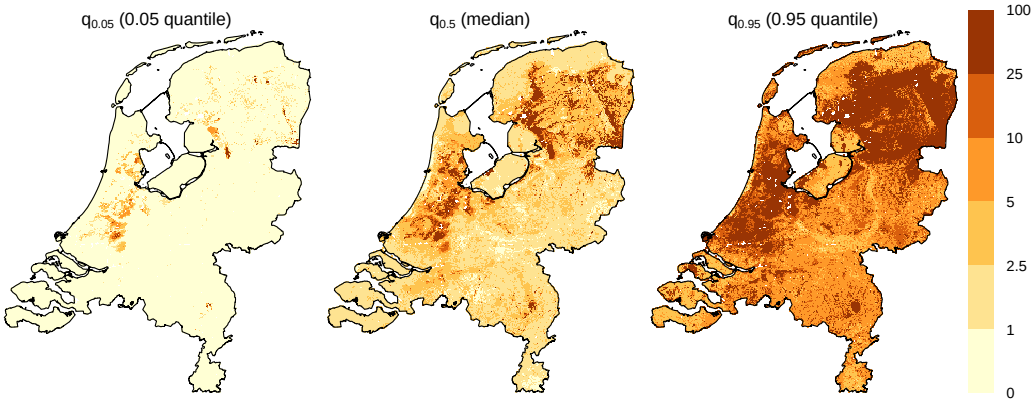

Figure S48. Predicted 5th, 50th (median) and 95th quantile for SOM [%] from 15 - 30 cm depth in 2023.

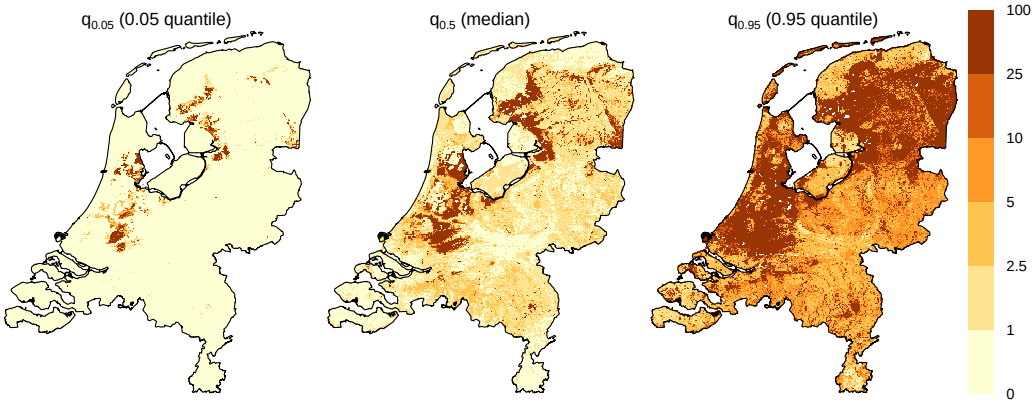

Figure S49. Predicted 5th, 50th (median) and 95th quantile for SOM [%] from 30 - 60 cm depth in 2023.

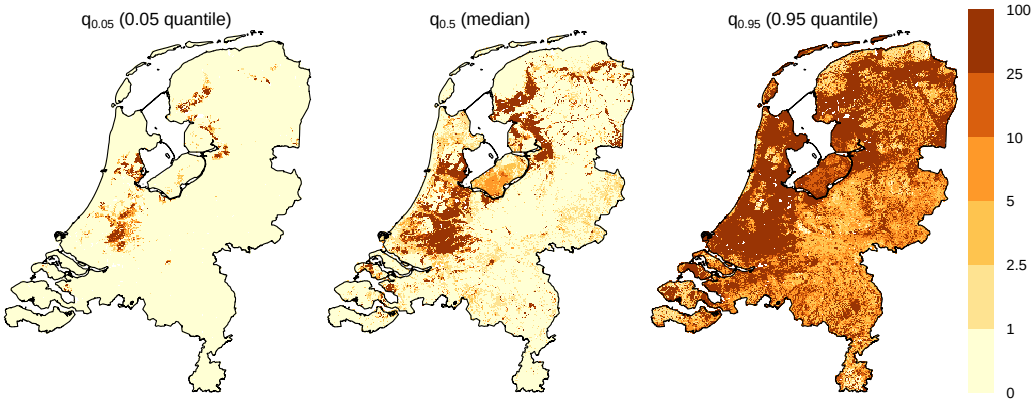

Figure S50. Predicted 5th, 50th (median) and 95th quantile for SOM [%] from 60 - 100 cm depth in 2023.

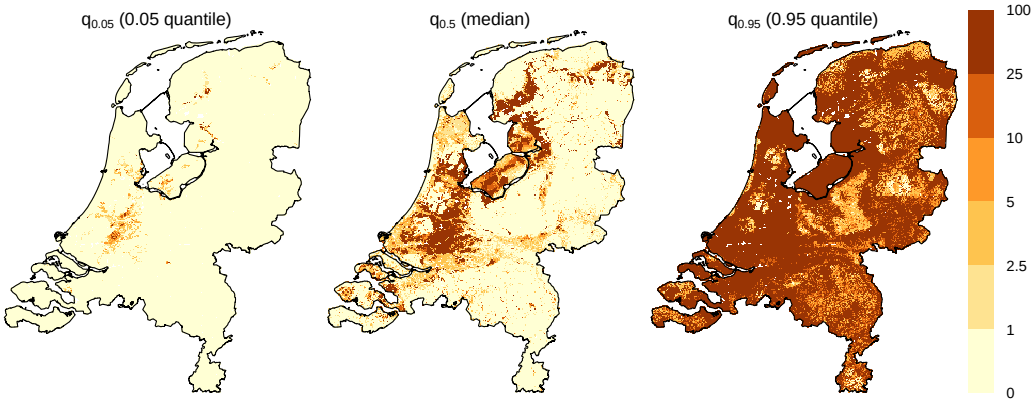

Figure S51. Predicted 5th, 50th (median) and 95th quantile for SOM [%] from 100 - 200 cm depth in 2023.

S5.2 Accuracy assessment

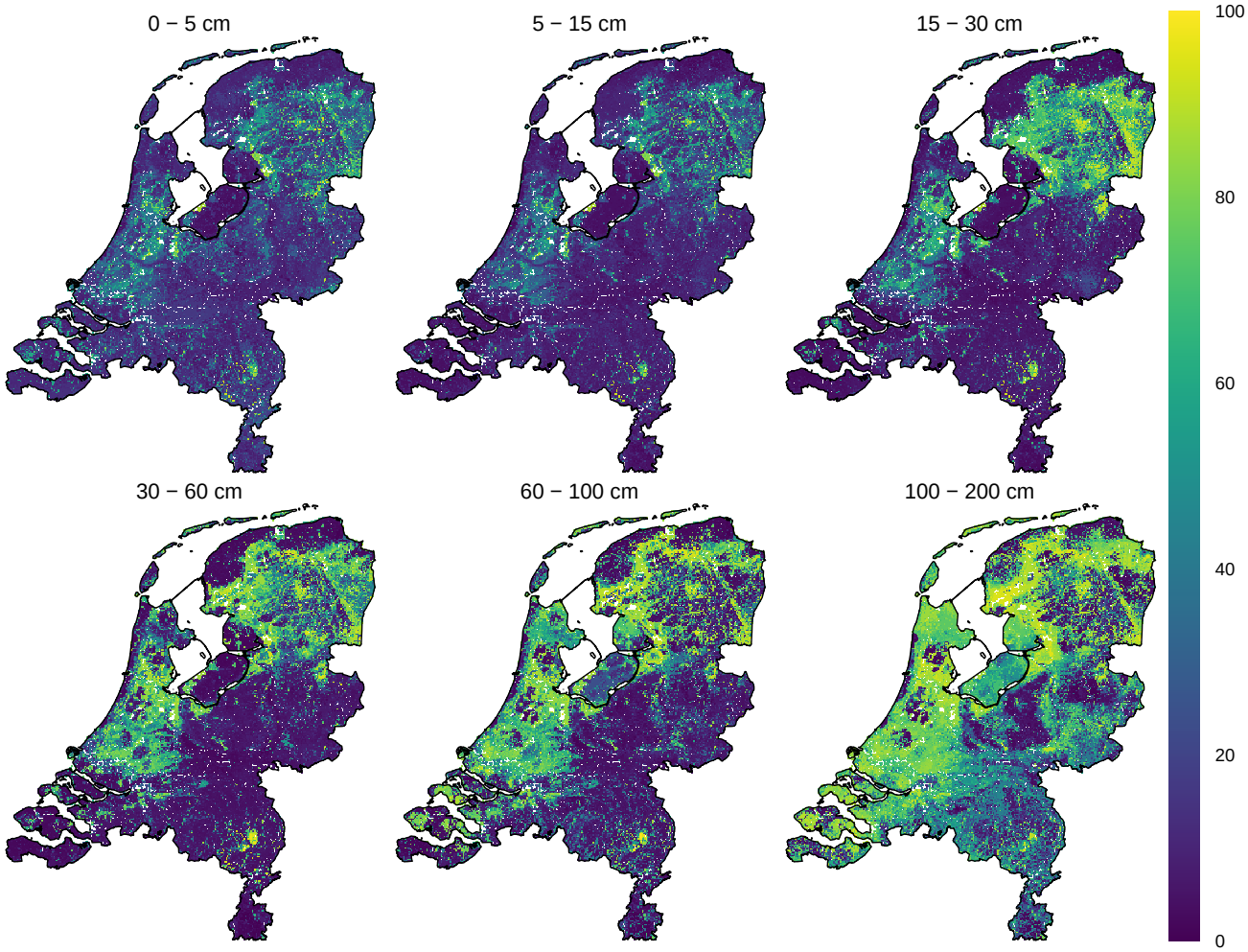

Figure S52. Maps of the PI90 as a measure of prediction uncertainty for SOM [%] for every GSM depth layer in 2023.

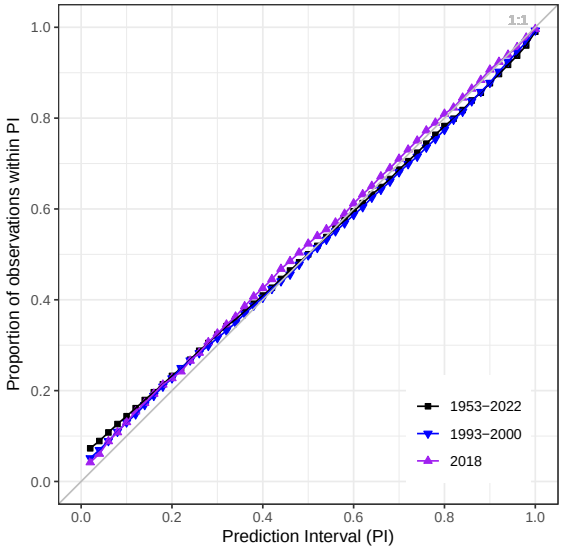

Figure S53. Prediction interval coverage probability (PICP) for prediction intervals between 0.02 and 1 of data used for different statistical validation strategies: 10-fold cross-validation of PFB laboratory measurements between 1953-2022, LSK laboratory measurements from 1993-2000 and CCNL laboratory measurements from 2018 (Sect. 2.1.2, 2.6 & Helfenstein et al., 2024). The closer the points are to the 1:1 line, the more accurate the prediction uncertainty. Lines connecting the points do not represent actual data and are only for visual guidance.

Table S5. Model accuracy metrics of mean and median SOM [%] predictions using design-based inference of LSK laboratory measurements from 1993-2000 (Sect. 2.6). ME = mean error; RMSE = root mean squared error; MEC = model efficiency coefficient; PICP90 = prediction interval coverage probability of the PI90. The lower and upper 97.5% confidence limits of the accuracy metrics were computed using design-based inference according to de Gruijter et al. (2006) and Sect. 2.6.2 of Helfenstein et al. (2022). Together, the lower and upper 97.5% confidence limits give the 95% confidence interval (CI95).

Prediction	Depth (cm)	n	**ME**	CI95 of ME	**RMSE**	CI95 of RMSE	**MEC**	CI95 of MEC	**PICP90**
	0 - 15	1085	**0.5**	0.22, 0.78	**4.84**	4.25, 5.37	**0.52**	0.38, 0.61	**0.75**
	15 - 30	205	**-0.64**	-1.78, 0.5	**5.66**	2.15, 7.71	**0.53**	0.07, 0.75	**0.85**
Mean	30 - 60	1145	**-2.1**	-2.63, -1.56	**9.2**	8.26, 10.06	**0.34**	0.11, 0.51	**0.87**
	60 - 100	1075	**-3.01**	-3.85, -2.17	**9.92**	8.87, 10.87	**0.46**	0.12, 0.61	**0.93**
	100 - 200	774	**-3.16**	-4, -2.31	**10.41**	8.4, 12.09	**0.44**	0.32, 0.54	**0.96**
	0 - 15	1085	**1.37**	1.11, 1.64	**5.08**	4.37, 5.7	**0.48**	0.35, 0.58	**0.75**
	15 - 30	205	**0.72**	-0.19, 1.64	**4.68**	NA, 6.74	**0.68**	0.4, 0.82	**0.85**
Median	30 - 60	1145	**0.46**	-0.05, 0.98	**9.72**	8.31, 10.95	**0.27**	-0.01, 0.48	**0.87**
	60 - 100	1075	**0.43**	-0.11, 0.97	**9.22**	7.64, 10.57	**0.53**	0.19, 0.7	**0.93**
	100 - 200	774	**0.92**	0.08, 1.76	**9.47**	6.27, 11.84	**0.54**	0.39, 0.69	**0.96**

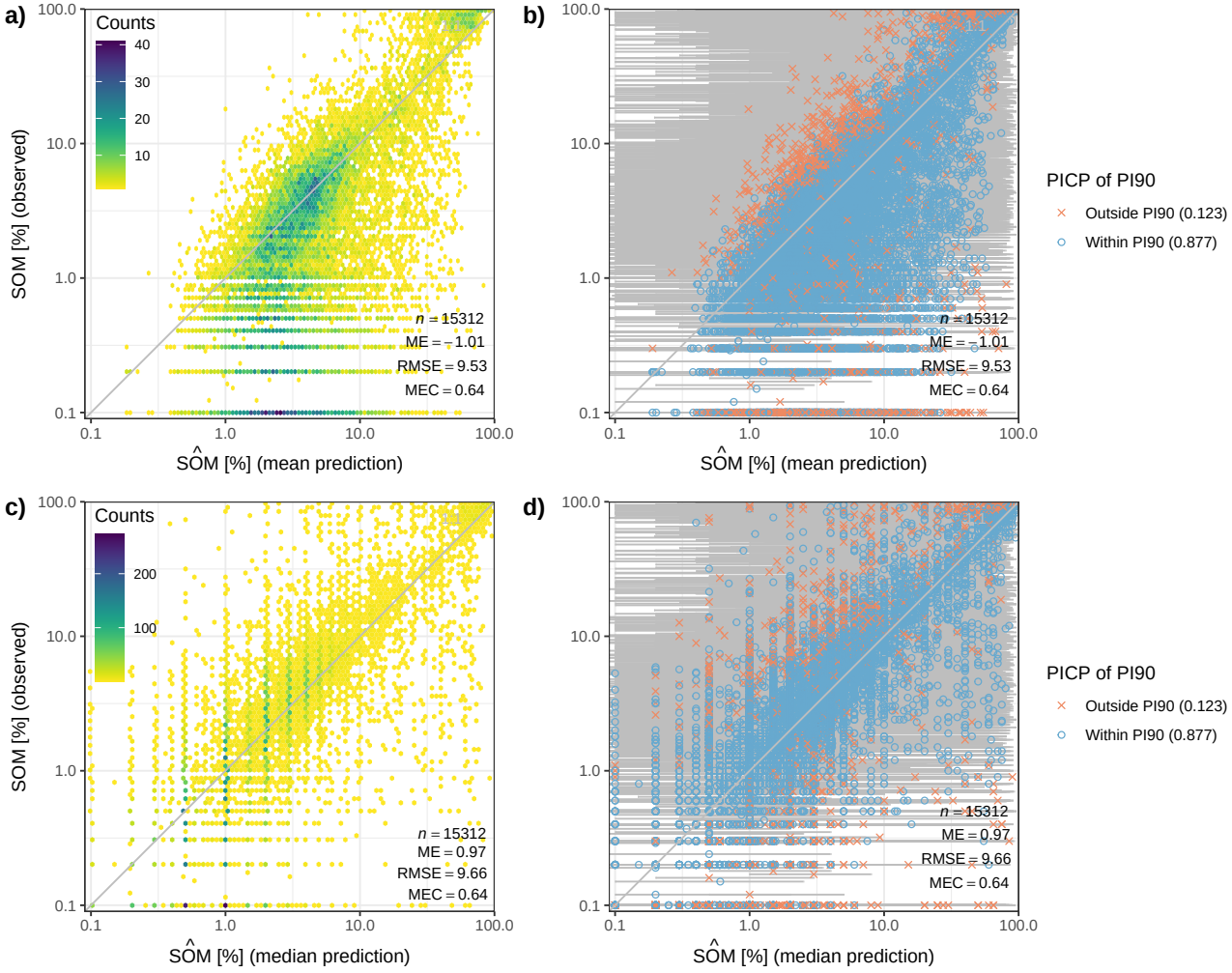

Figure S54. Predicted mean (a & b) and median (c & d) SOM [%] on the x-axis vs. measured SOM content [%] on the y-axis (log-scale). Accuracy plots and metrics (ME, RMSE and MEC) were computed using 10-fold cross-validation of PFB laboratory measurements between 1953-2022 (Sect. 2.6). Plots a & c emphasize point density whereas plots b & d visualize prediction uncertainty (PI90 as error bars) and the PICP90 in the figure legends.

S5.3 Variable importance

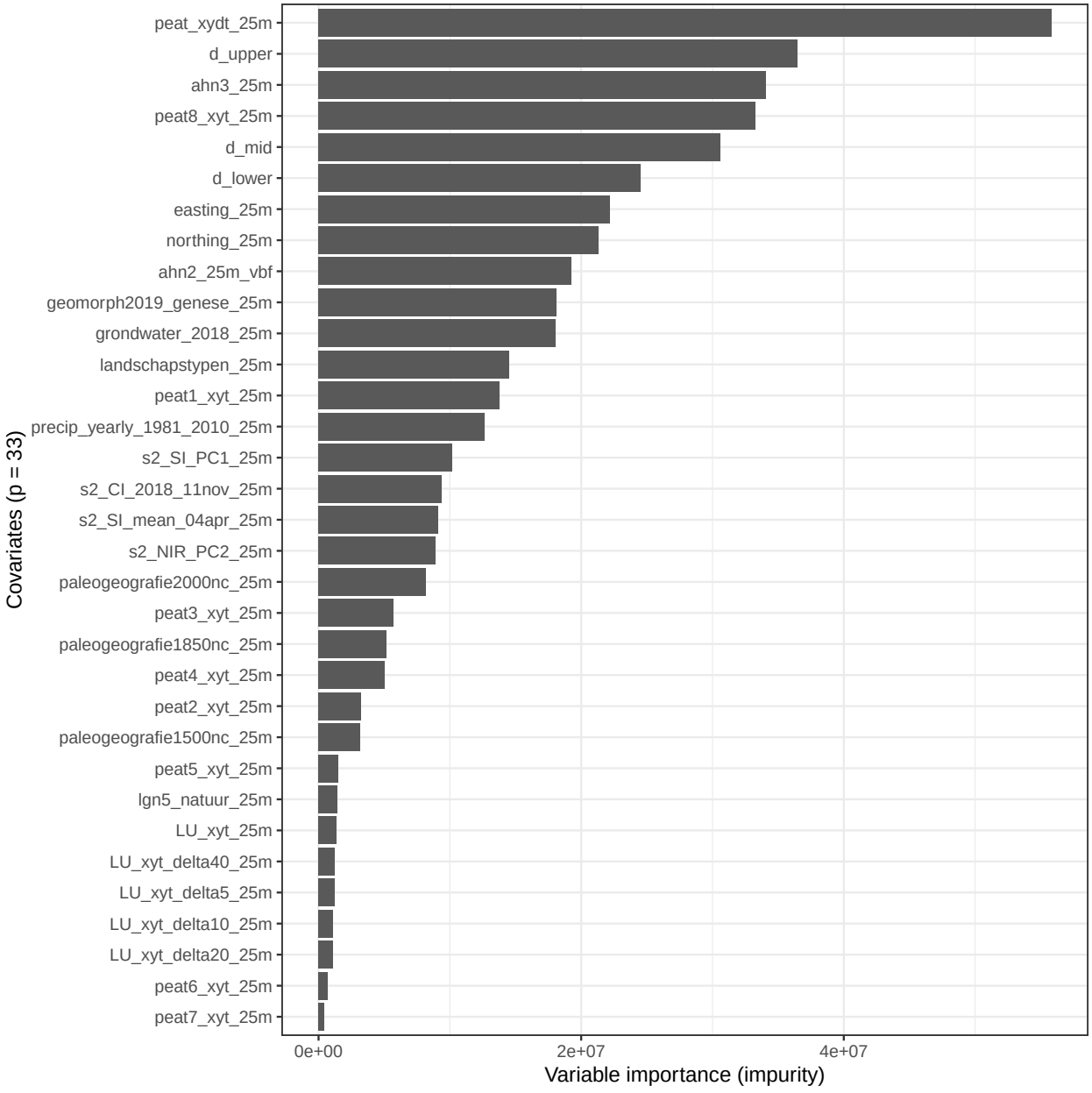

Figure S55. Variable importance for predicting SOM [%], assessed using the impurity method (Breiman, 2002) (Sect. 2.4). "d_upper", "d_mid" and "d_lower" denote the upper, midpoint and lower boundary of a sampled soil horizon during calibration and target depth layer during prediction (Sect. 2.2, Table 5). 2D+T and 3D+T dynamic covariates are denoted with "xyt" and "xydt", respectively (Helfenstein et al., 2024, Fig. S3 & Table S1). Variable importance using impurity favors covariates with more distinct values and is biased against categorical covariates because these have a finite number of binary splits due to limited number of classes (Sandri and Zuccolotto, 2008, 2010).

S6 pH

S6.1 Prediction maps

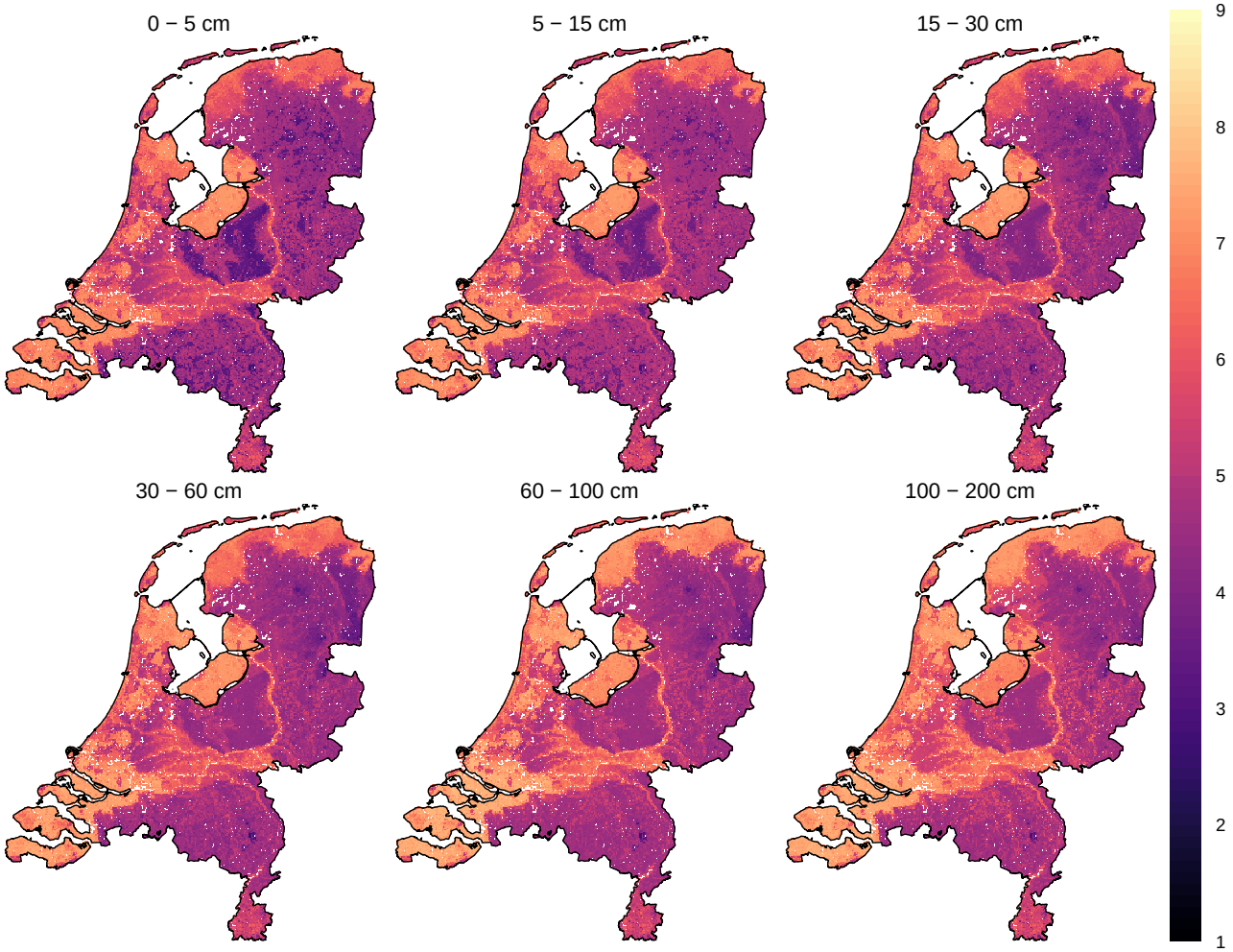

Figure S56. Predicted mean pH [KCl] for every GSM depth layer.

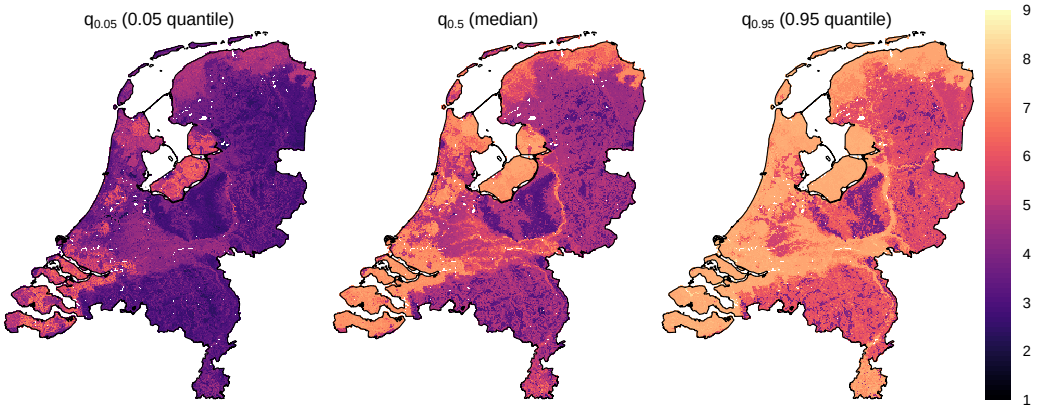

Figure S57. Predicted 5th, 50th (median) and 95th quantile for pH [KCl] from 0 - 5 cm depth.

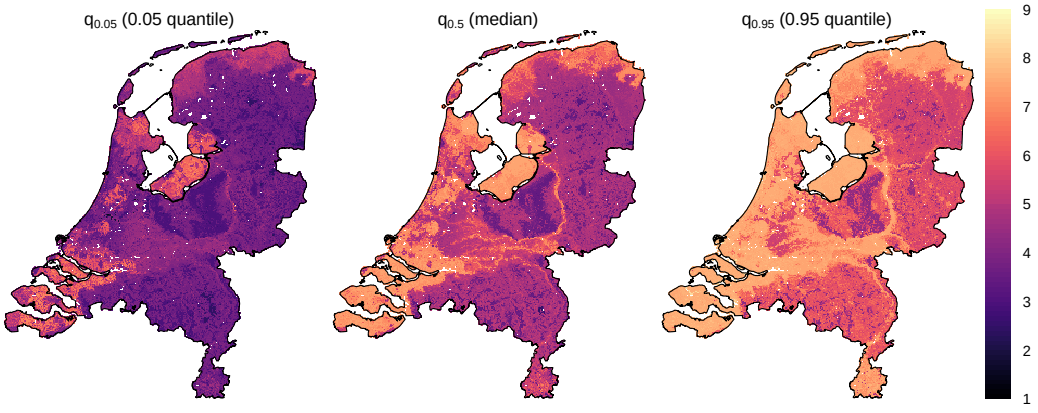

Figure S58. Predicted 5th, 50th (median) and 95th quantile for pH [KCl] from 5 - 15 cm depth.

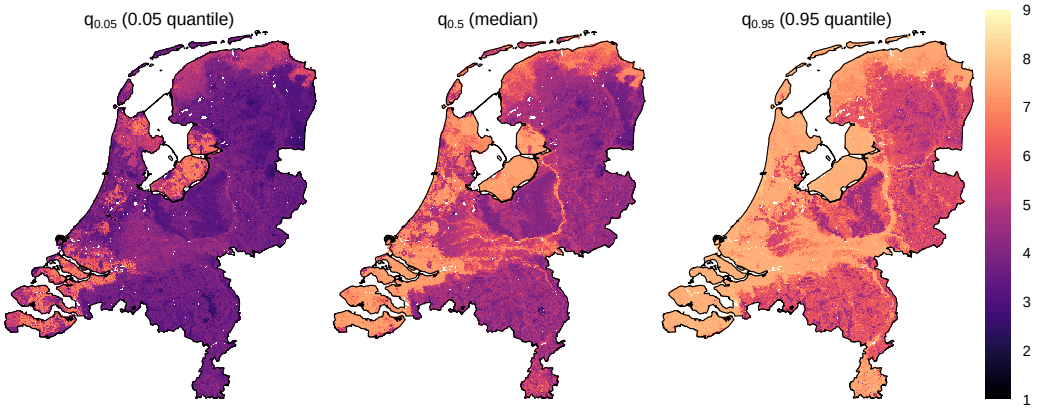

Figure S59. Predicted 5th, 50th (median) and 95th quantile for pH [KCl] from 15 - 30 cm depth.

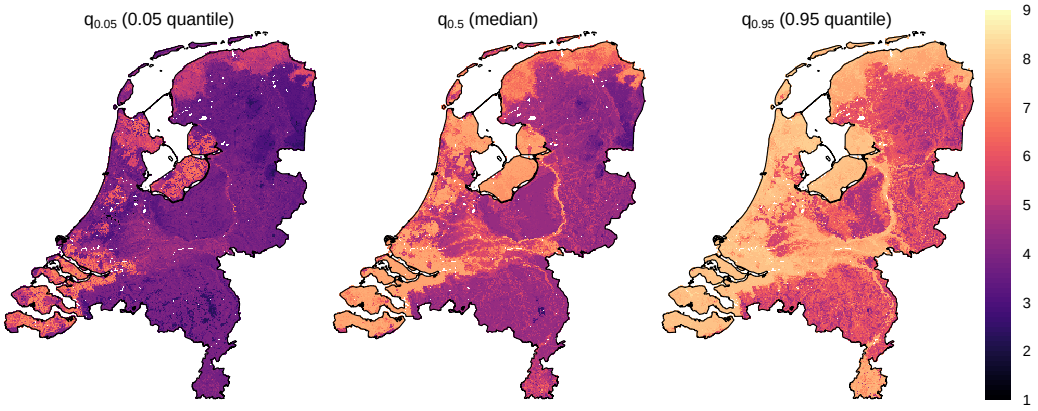

Figure S60. Predicted 5th, 50th (median) and 95th quantile for pH [KCl] from 30 - 60 cm depth.

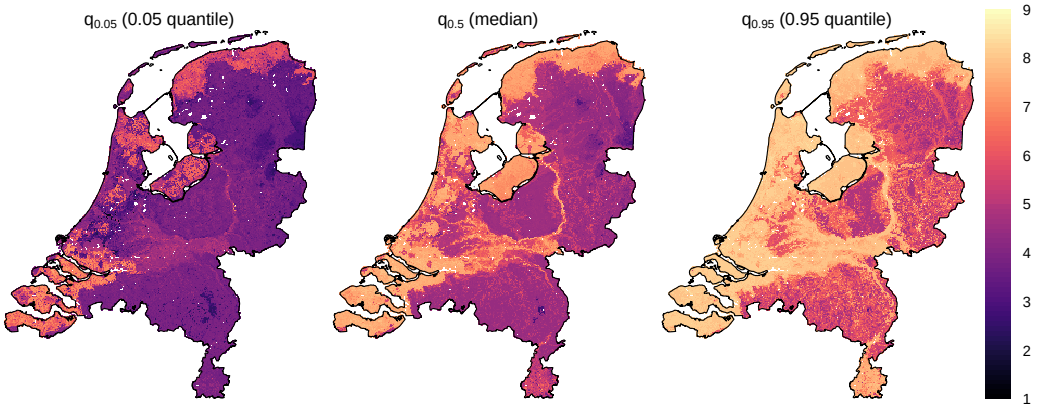

Figure S61. Predicted 5th, 50th (median) and 95th quantile for pH [KCl] from 60 - 100 cm depth.

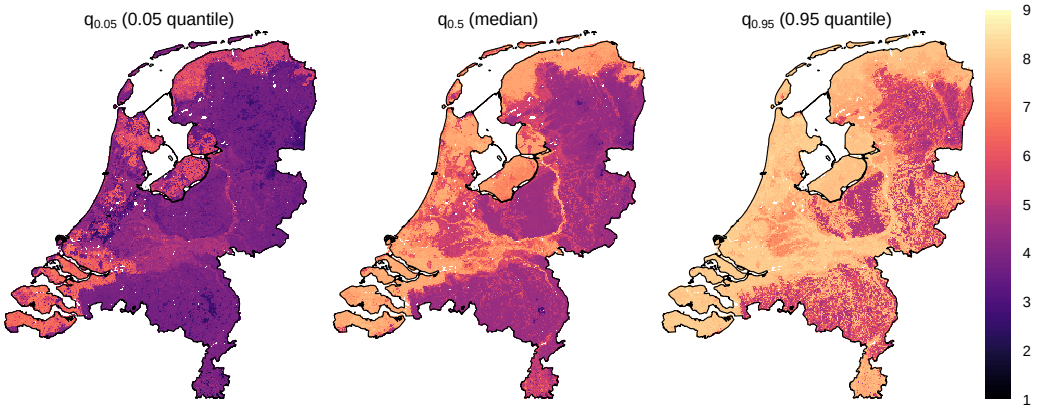

Figure S62. Predicted 5th, 50th (median) and 95th quantile for pH [KCl] from 100 - 200 cm depth.

S6.2 Accuracy assessment

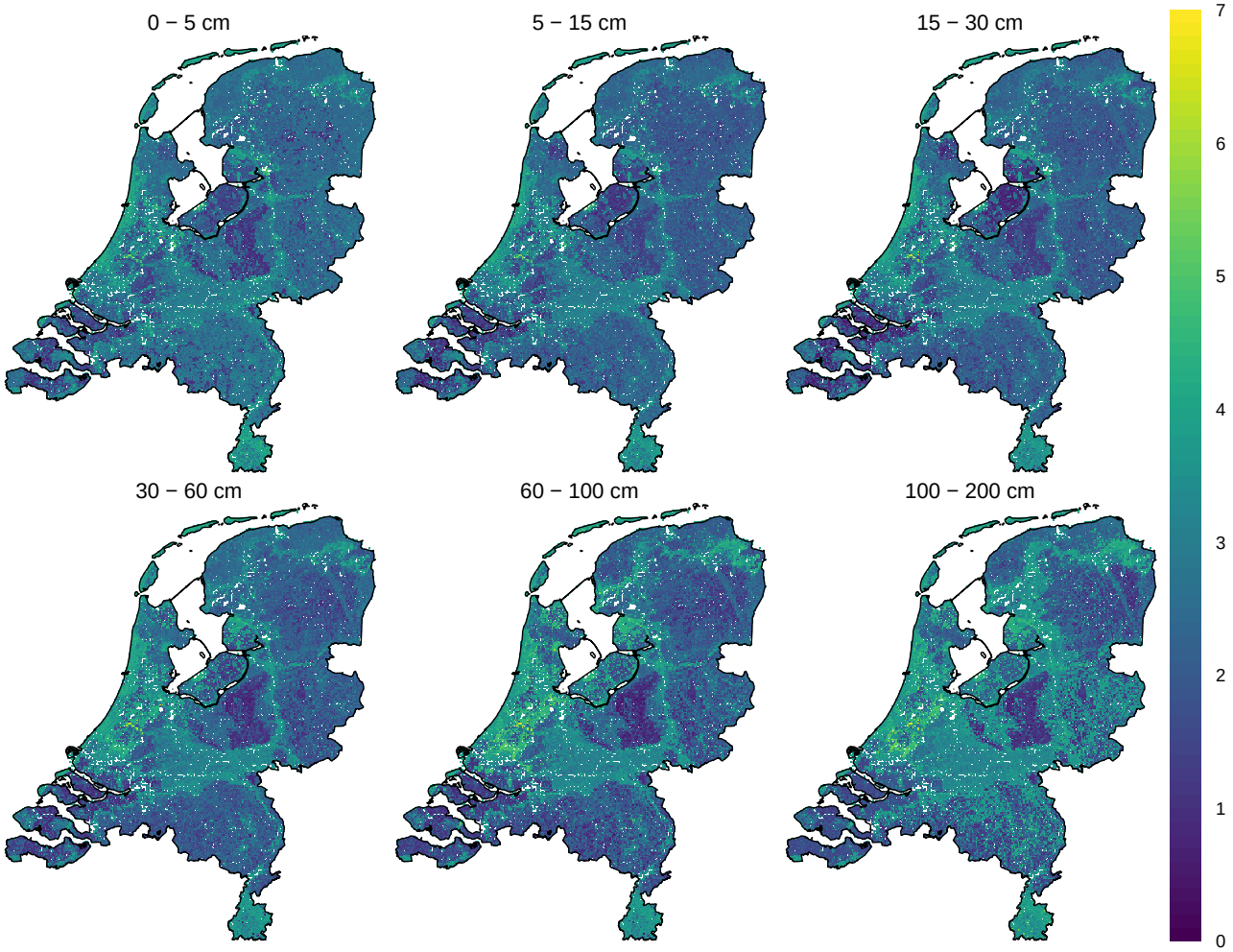

Figure S63. Maps of the PI90 as a measure of prediction uncertainty for pH [KCl] for every GSM depth layer.

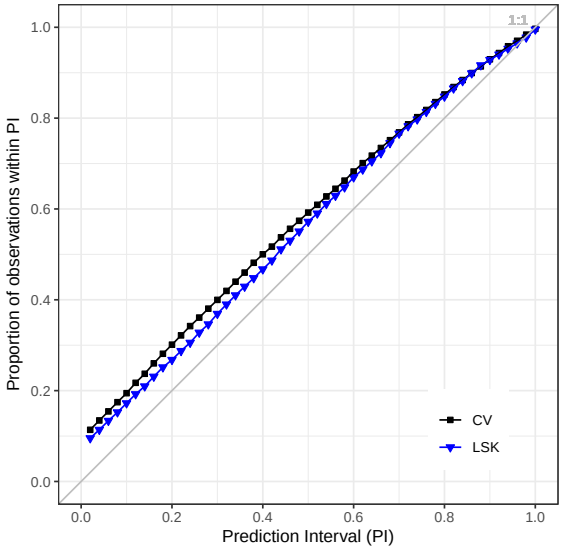

Figure S64. Prediction interval coverage probability (PICP) for prediction intervals between 0.02 and 1 of data used for different statistical validation strategies: 10-fold cross-validation of PFB laboratory measurements (CV) and LSK laboratory measurements (Sect. 2.6). The closer the points are to the 1:1 line, the more accurate the prediction uncertainty. Lines connecting the points do not represent actual data and are only for visual guidance.

Table S6. Model accuracy metrics of mean and median pH [KCl] predictions using design-based inference of LSK laboratory measurements (Sect. 2.6). ME = mean error; RMSE = root mean squared error; MEC = model efficiency coefficient; PICP90 = prediction interval coverage probability of the PI90. The lower and upper 97.5% confidence limits of the accuracy metrics were computed using design-based inference according to de Gruijter et al. (2006) and Sect. 2.6.2 of Helfenstein et al. (2022). Together, the lower and upper 97.5% confidence limits give the 95% confidence interval (CI95).

Prediction	Depth (cm)	n	**ME**	CI95 of ME	**RMSE**	CI95 of RMSE	**MEC**	CI95 of MEC	**PICP90**
	0 - 15	1250	**0.14**	0.09, 0.18	**0.74**	0.71, 0.77	**0.71**	0.68, 0.73	**0.92**
	15 - 30	245	**0.16**	0.06, 0.26	**0.68**	0.57, 0.77	**0.91**	0.9, 0.92	**0.9**
Mean	30 - 60	1322	**0.16**	0.12, 0.2	**0.73**	0.7, 0.77	**0.73**	0.7, 0.75	**0.93**
	60 - 100	1248	**0.09**	0.05, 0.13	**0.73**	0.69, 0.77	**0.74**	0.71, 0.77	**0.94**
	100 - 200	852	**0.05**	-0.01, 0.1	**0.81**	0.74, 0.87	**0.77**	0.74, 0.8	**0.92**
	0 - 15	1250	**0.13**	0.09, 0.17	**0.73**	0.69, 0.77	**0.71**	0.68, 0.74	**0.92**
	15 - 30	245	**0.15**	0.05, 0.26	**0.65**	0.52, 0.77	**0.92**	0.91, 0.93	**0.9**
Median	30 - 60	1322	**0.15**	0.11, 0.19	**0.74**	0.7, 0.77	**0.72**	0.7, 0.75	**0.93**
	60 - 100	1248	**0.08**	0.04, 0.12	**0.73**	0.68, 0.78	**0.74**	0.71, 0.77	**0.94**
	100 - 200	852	**0.05**	-0.01, 0.11	**0.82**	0.74, 0.89	**0.76**	0.72, 0.8	**0.92**

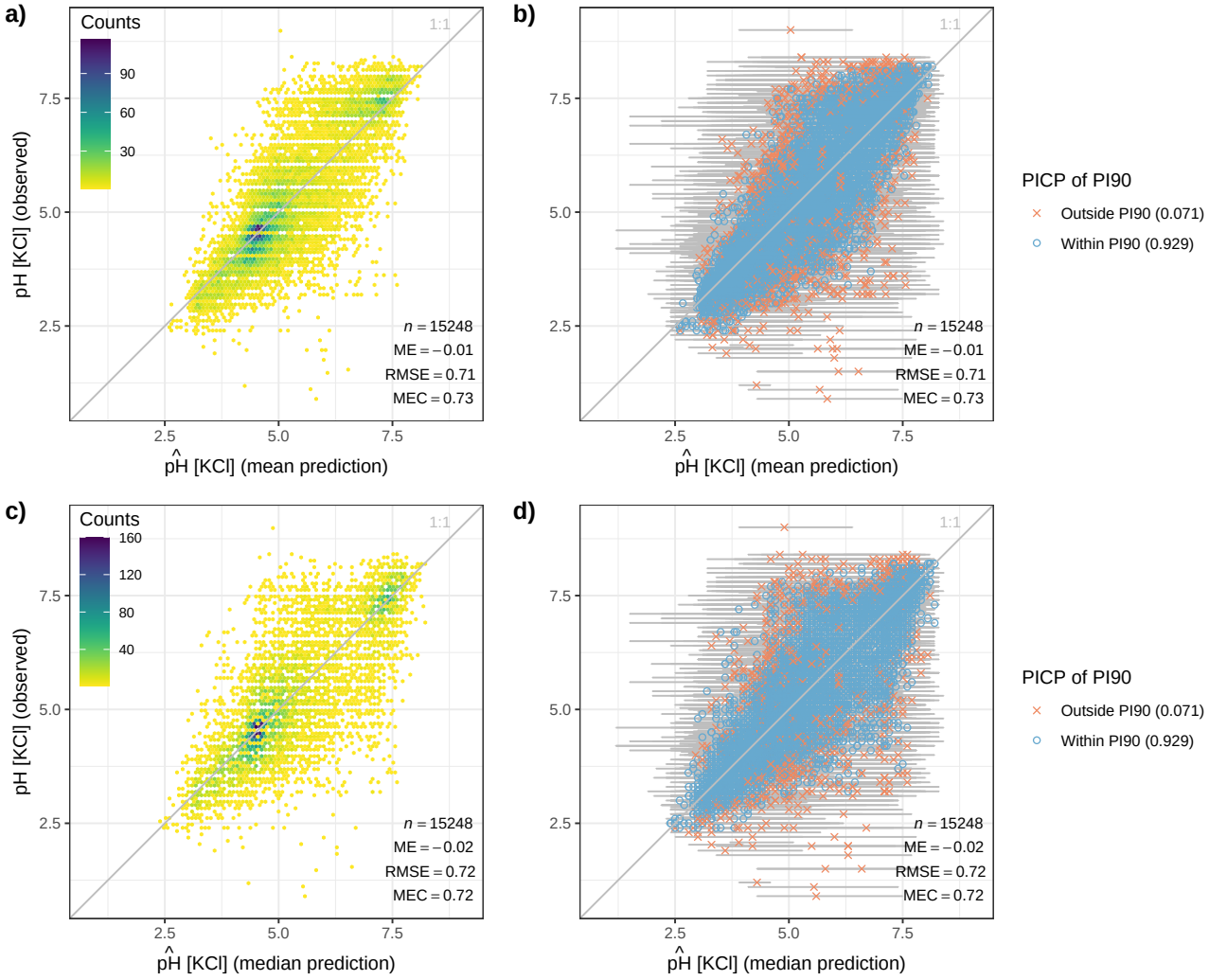

Figure S65. Predicted mean (a & b) and median (c & d) pH [KCl] on the x-axis vs. measured pH [KCl] on the y-axis. Accuracy plots and metrics (ME, RMSE and MEC) were computed using 10-fold cross-validation of PFB laboratory measurements (Sect. 2.6). Plots a & c emphasize point density whereas plots b & d visualize prediction uncertainty (PI90 as error bars) and the PICP90 in the figure legends.

S6.3 Variable importance

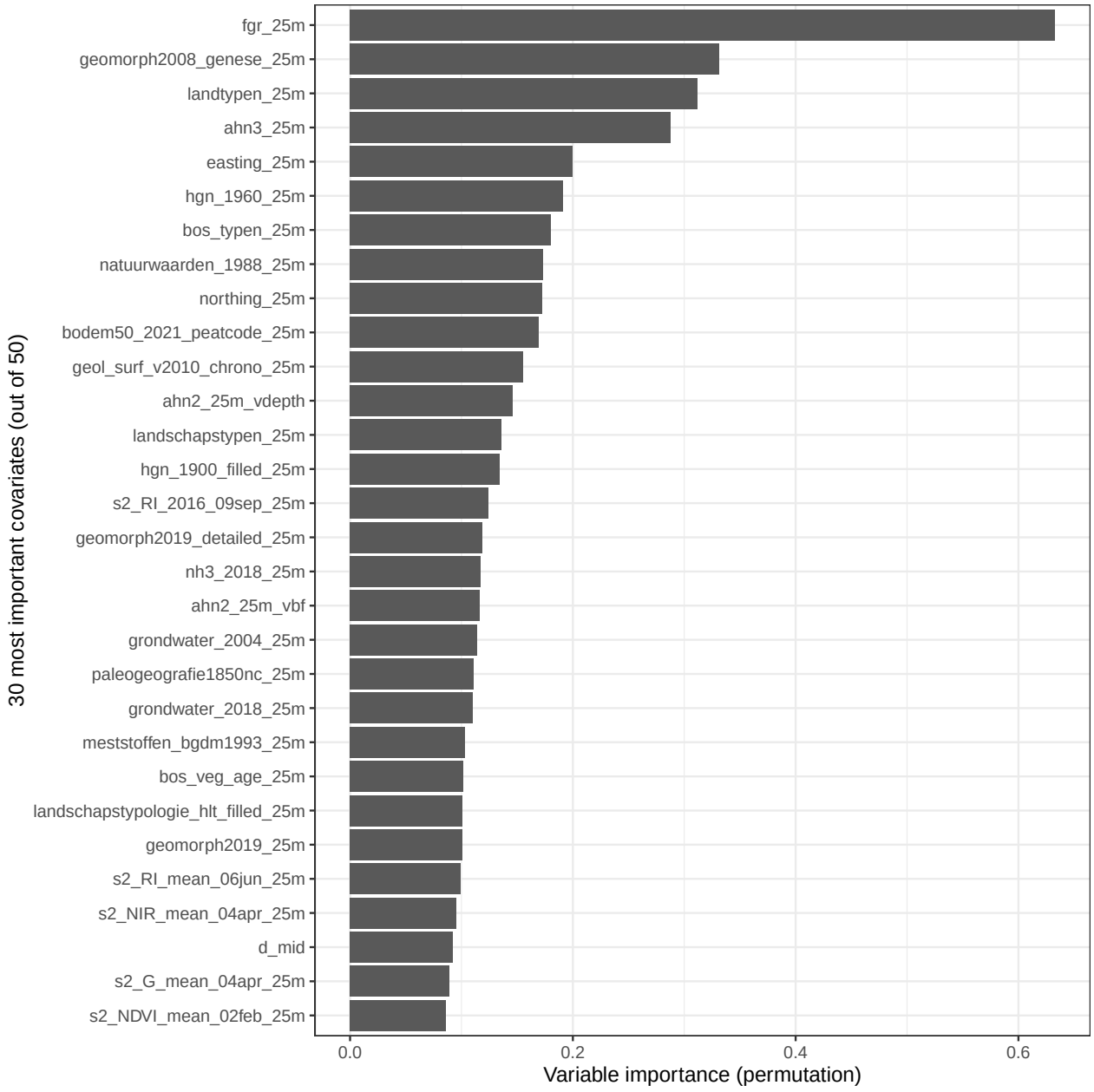

Figure S66. Variable importance for predicting pH, assessed using the permutation method (Breiman, 2002) (Sect. 2.4). "d_upper", "d_mid" and "d_lower" denote the upper, midpoint and lower boundary of a sampled soil horizon during calibration and target depth layer during prediction (Sect. 2.2, Table 5).

S7 Total N (N_{tot})

S7.1 Prediction maps

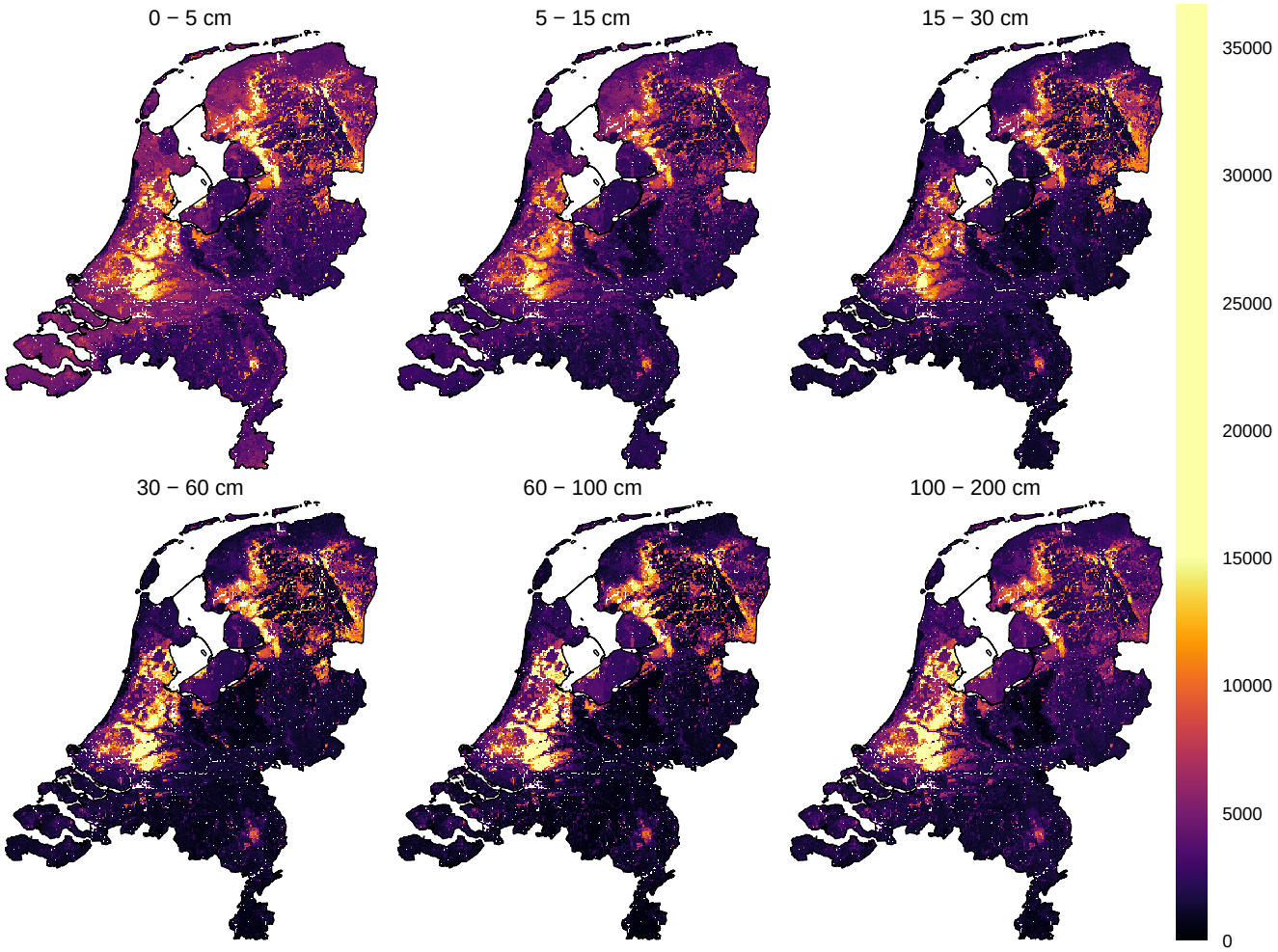

Figure S67. Predicted mean N_{tot} [mg/kg] for every GSM depth layer.

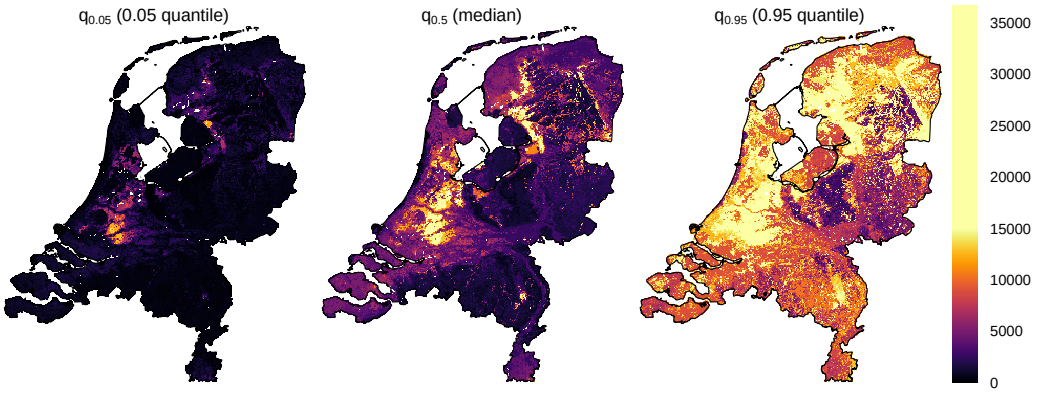

Figure S68. Predicted 5th, 50th (median) and 95th quantile for N_{tot} [mg/kg] from 0 - 5 cm depth.

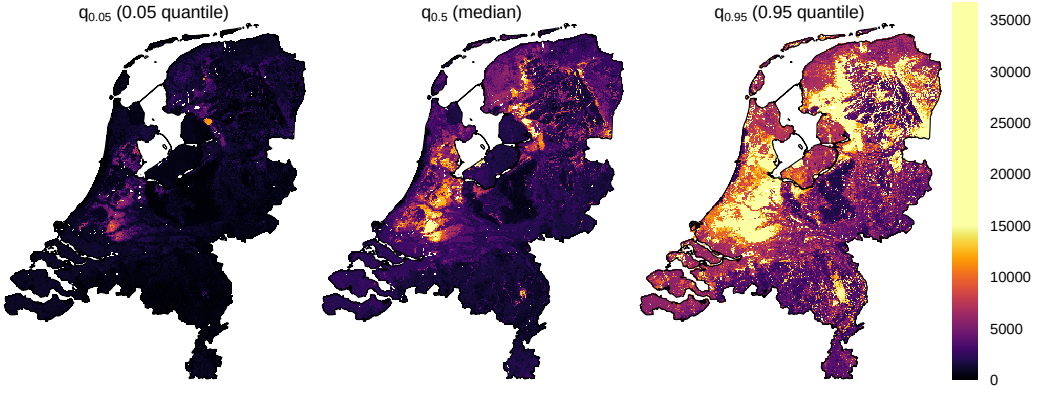

Figure S69. Predicted 5th, 50th (median) and 95th quantile for N_{tot} [mg/kg] from 5 - 15 cm depth.

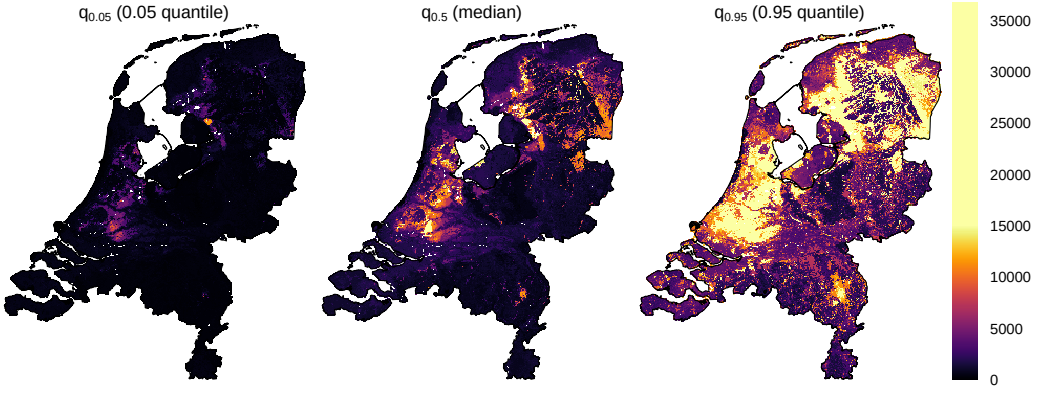

Figure S70. Predicted 5th, 50th (median) and 95th quantile for N_{tot} [mg/kg] from 15 - 30 cm depth.

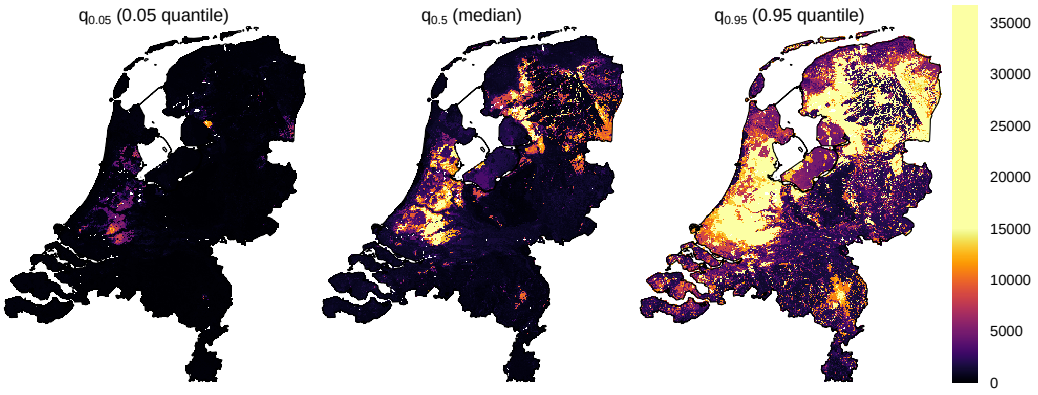

Figure S71. Predicted 5th, 50th (median) and 95th quantile for N_{tot} [mg/kg] from 30 - 60 cm depth.

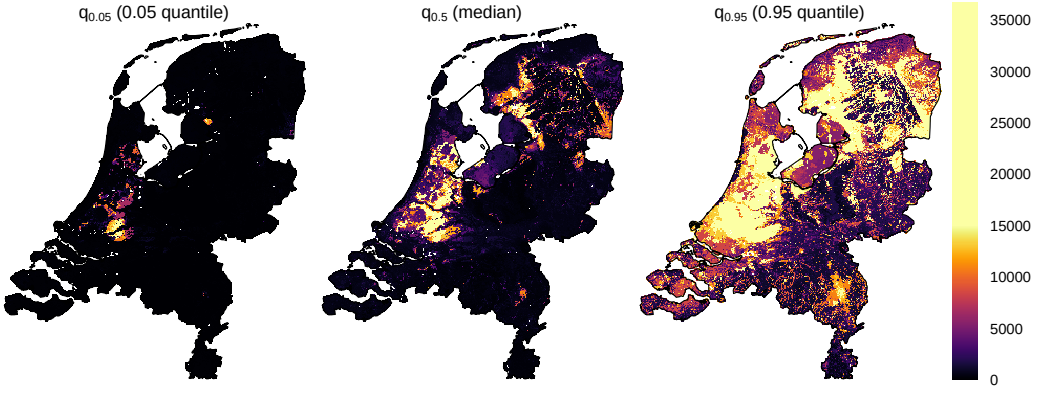

Figure S72. Predicted 5th, 50th (median) and 95th quantile for N_{tot} [mg/kg] from 60 - 100 cm depth.

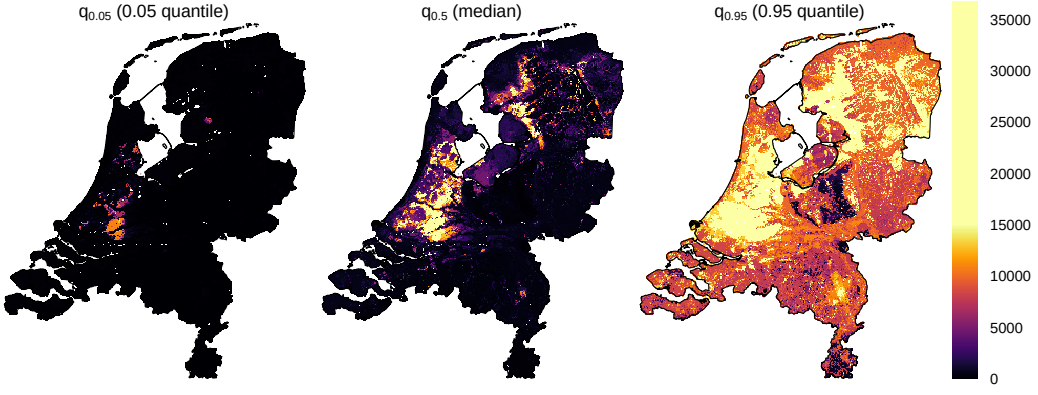

Figure S73. Predicted 5th, 50th (median) and 95th quantile for N_{tot} [mg/kg] from 100 - 200 cm depth.

S7.2 Accuracy assessment

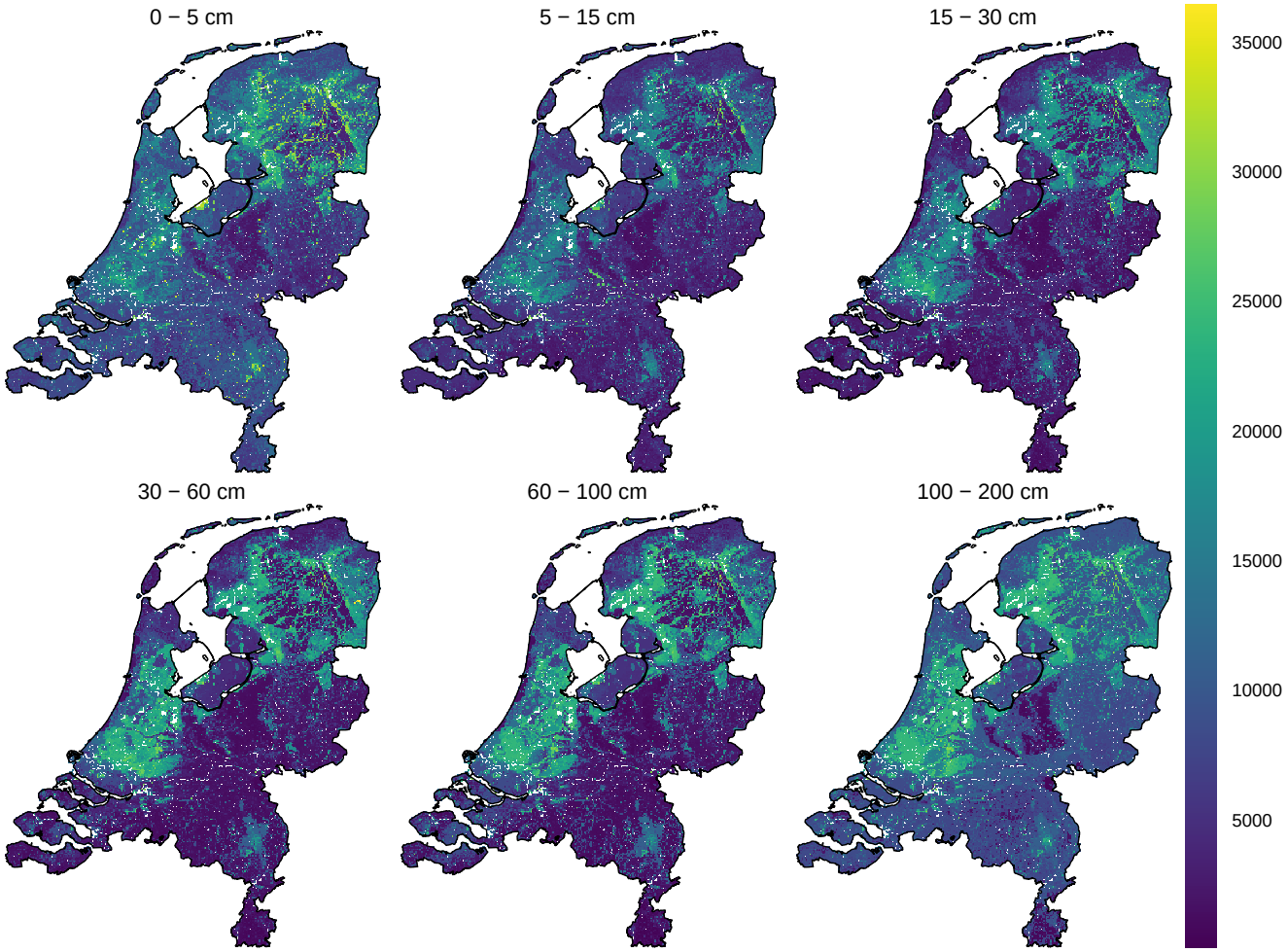

Figure S74. Maps of the PI90 as a measure of prediction uncertainty for N_{tot} [mg/kg] for every GSM depth layer.

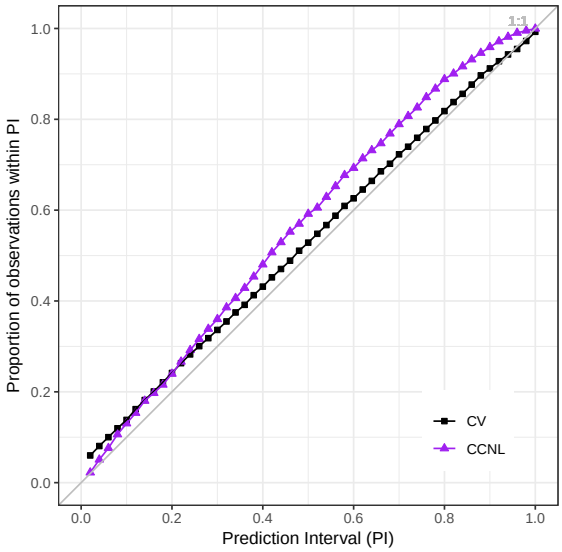

Figure S75. Prediction interval coverage probability (PICP) for prediction intervals between 0.02 and 1 of data used for different statistical validation strategies: 10-fold cross-validation of PFB laboratory measurements (CV) and CCNL laboratory measurements (Sect. 2.6). The closer the points are to the 1:1 line, the more accurate the prediction uncertainty. Lines connecting the points do not represent actual data and are only for visual guidance.

Table S7. Model accuracy metrics of mean and median N_{tot} [mg/kg] predictions using design-based inference of CCNL laboratory measurements (Sect. 2.6). ME = mean error; RMSE = root mean squared error; MEC = model efficiency coefficient; PICP90 = prediction interval coverage probability of the PI90. The lower and upper 97.5% confidence limits of the accuracy metrics were computed using design-based inference according to de Gruijter et al. (2006) and Sect. 2.6.2 of Helfenstein et al. (2022). Together, the lower and upper 97.5% confidence limits give the 95% confidence interval (CI95). For the 100 - 200 cm depth layer, metrics are based on 10-fold cross-validation of PFB laboratory measurements and thus without CI95 because laboratory measurements were not available from CCNL below 100 cm (Sect. 2.1.2 & 2.6).

Prediction	Depth (cm)	n	**ME**	CI95 of ME	**RMSE**	CI95 of RMSE	**MEC**	CI95 of MEC	**PICP90**
	0 - 15	1144	**-511**	-612, -410	**2165**	2001, 2318	**0.44**	0.34, 0.52	**0.95**
	15 - 30	1144	**-511**	-612, -410	**2165**	2001, 2318	**0.44**	0.34, 0.52	**0.95**
Mean	30 - 60	1140	**-1407**	-1524, -1290	**2981**	2796, 3156	**0.27**	0.13, 0.39	**0.97**
	60 - 100	1140	**-1408**	-1525, -1290	**2981**	2796, 3156	**0.27**	0.13, 0.39	**0.97**
	100 - 200	149	**-221**	-	**3530**	-	**0.67**	-	**0.89**
	0 - 15	1144	**120**	22, 218	**1994**	1809, 2164	**0.52**	0.44, 0.59	**0.95**
	15 - 30	1144	**120**	22, 218	**1994**	1809, 2164	**0.52**	0.44, 0.59	**0.95**
Median	30 - 60	1140	**-609**	-732, -485	**2678**	2446, 2892	**0.41**	0.29, 0.52	**0.97**
	60 - 100	1140	**-609**	-732, -486	**2678**	2446, 2892	**0.41**	0.29, 0.52	**0.97**
	100 - 200	149	**408**	-	**3788**	-	**0.62**	-	**0.89**

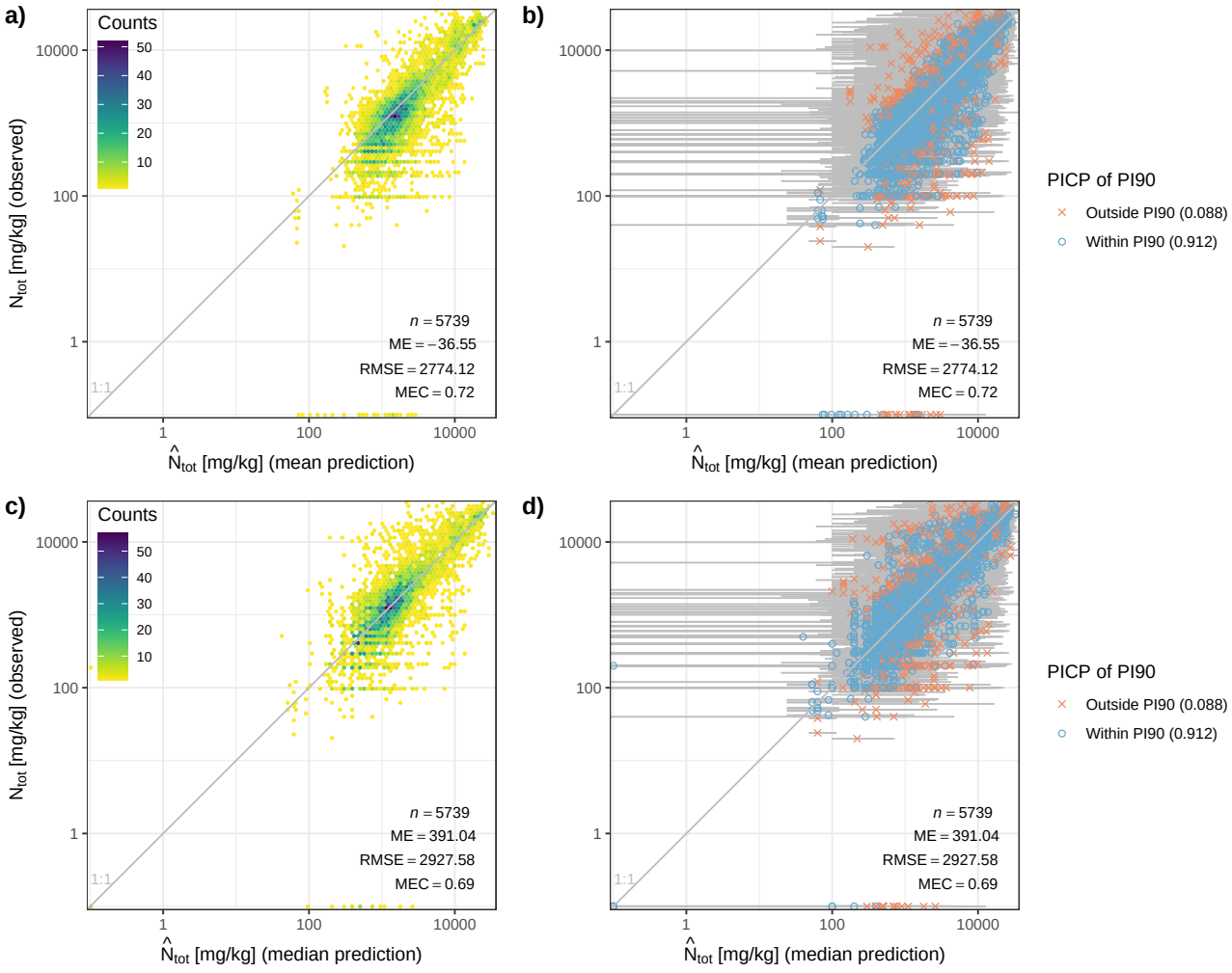

Figure S76. Predicted mean (a & b) and median (c & d) N_{tot} [mg/kg] on the x-axis vs. measured N_{tot} [mg/kg] on the y-axis (log-scale). Accuracy plots and metrics (ME, RMSE and MEC) were computed using 10-fold cross-validation of PFB laboratory measurements (Sect. 2.6). Plots a & c emphasize point density whereas plots b & d visualize prediction uncertainty (PI90 as error bars) and the PICP90 in the figure legends.

S7.3 Variable importance

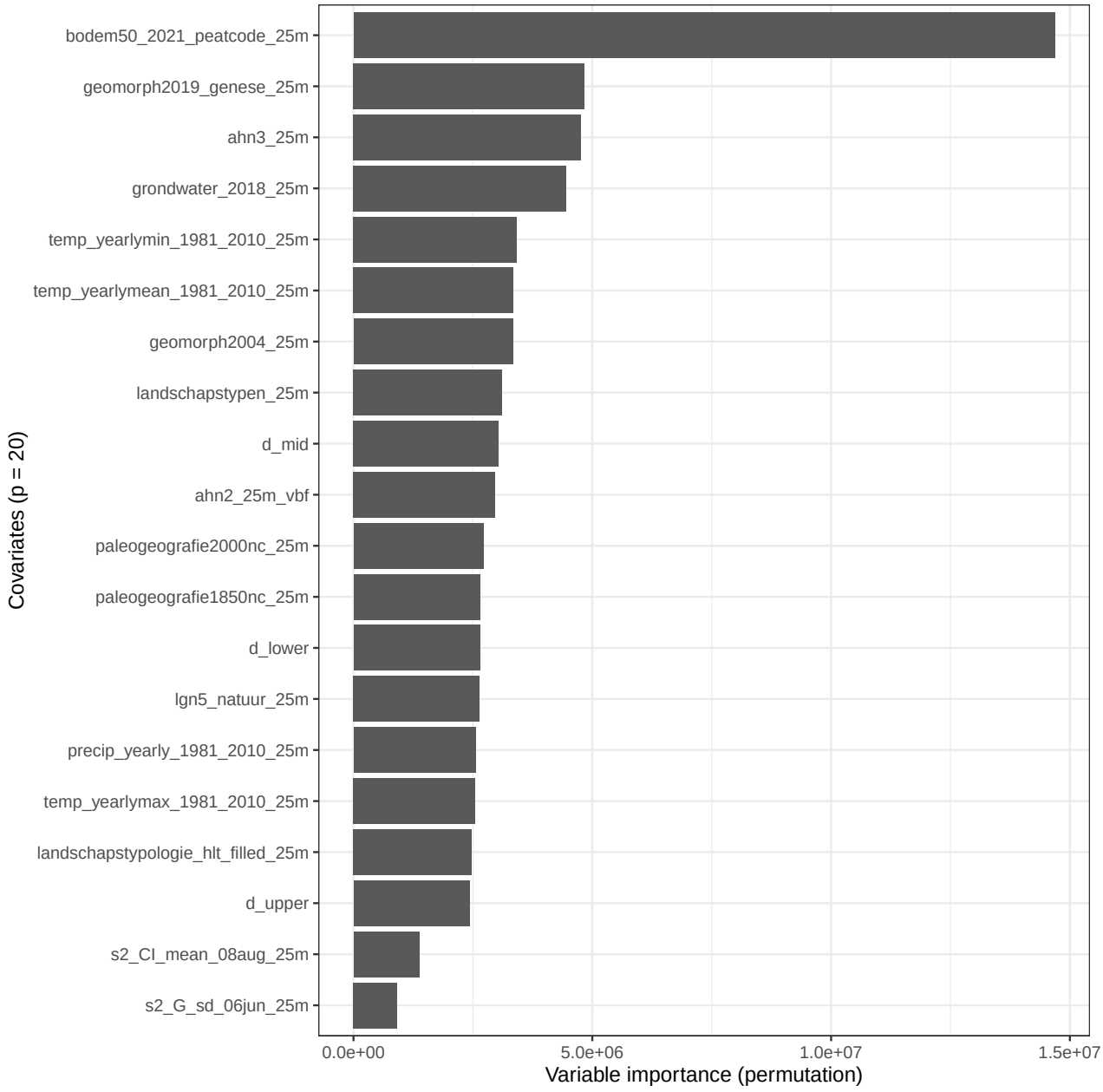

Figure S77. Variable importance for predicting N_{tot}, assessed using the permutation method (Breiman, 2002) (Sect. 2.4). "d_upper", "d_mid" and "d_lower" denote the upper, midpoint and lower boundary of a sampled soil horizon during calibration and target depth layer during prediction (Sect. 2.2, Table 5).

S8 Oxalate-extractable P (P_{ox})

S8.1 Prediction maps

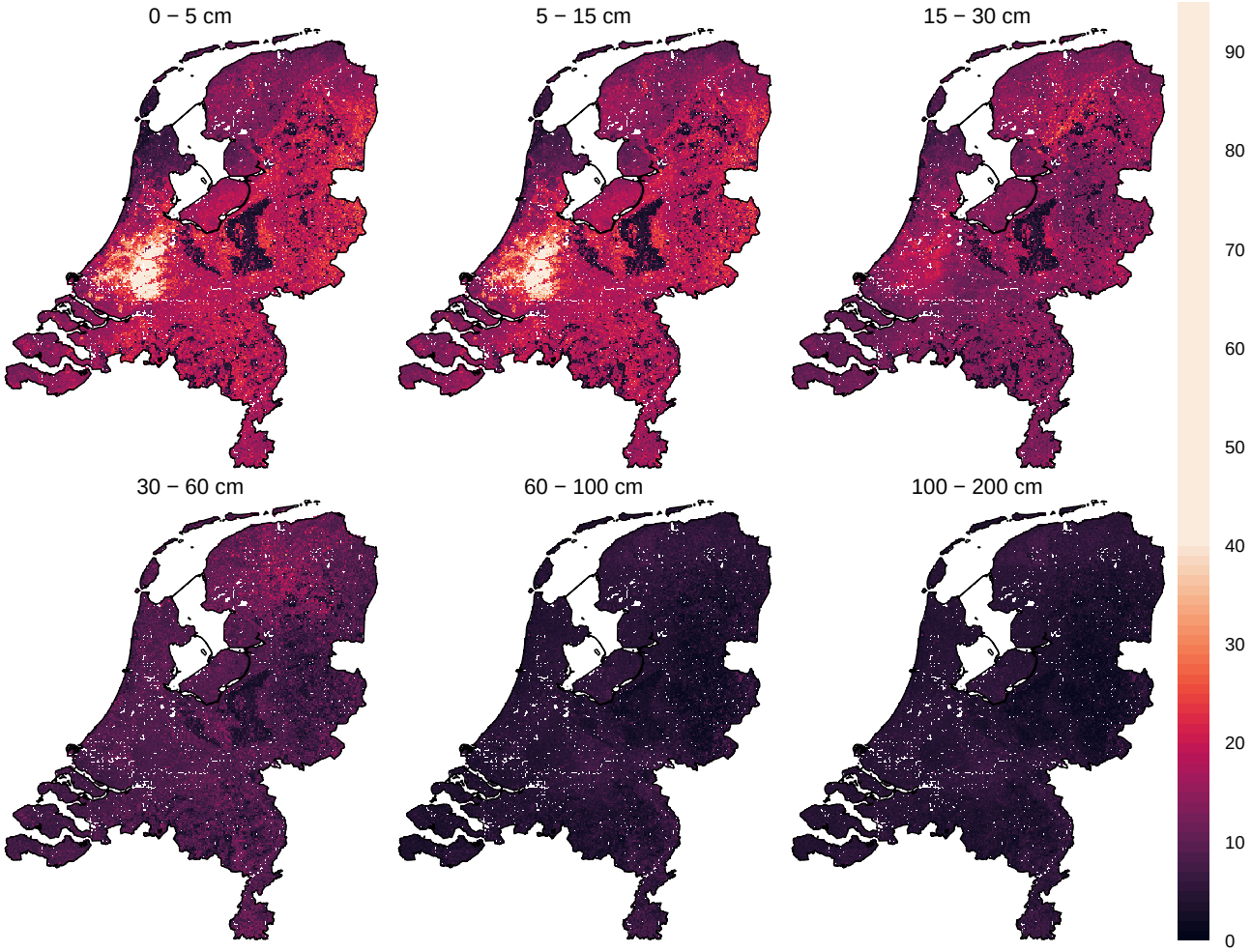

Figure S78. Predicted mean P_{ox} [mmol/kg] for every GSM depth layer.

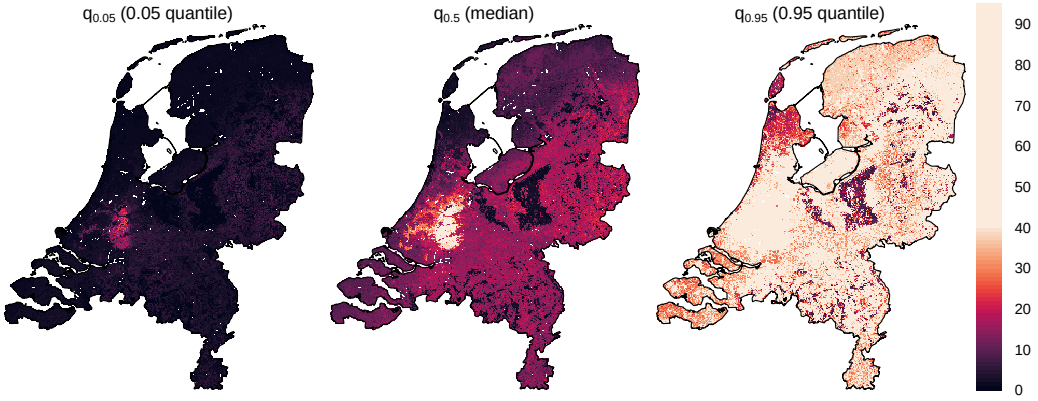

Figure S79. Predicted 5th, 50th (median) and 95th quantile for P_{ox} [mmol/kg] from 0 - 5 cm depth.

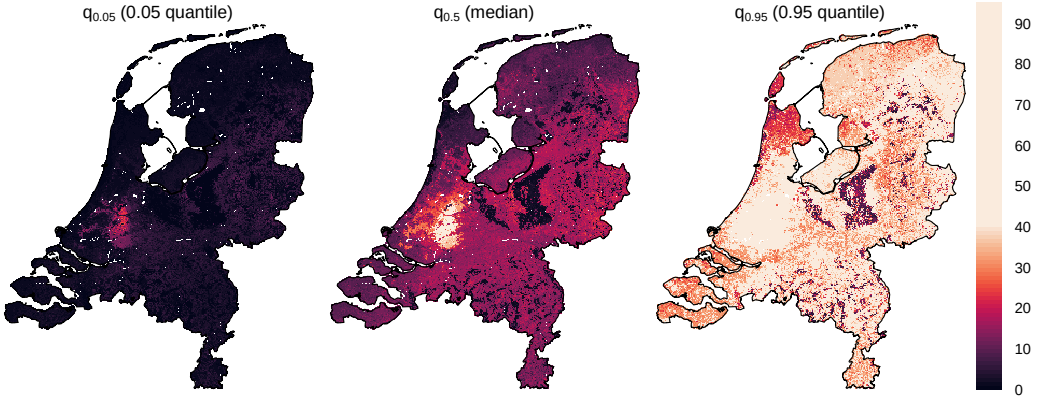

Figure S80. Predicted 5th, 50th (median) and 95th quantile for P_{ox} [mmol/kg] from 5 - 15 cm depth.

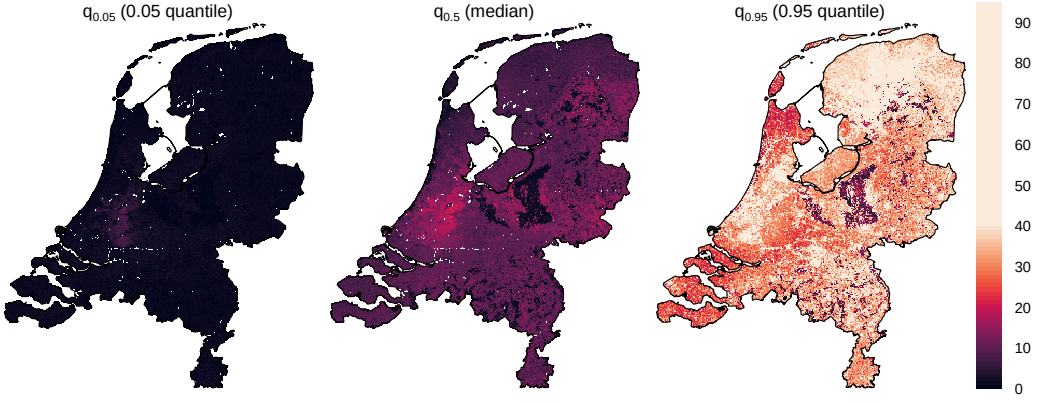

Figure S81. Predicted 5th, 50th (median) and 95th quantile for P_{ox} [mmol/kg] from 15 - 30 cm depth.

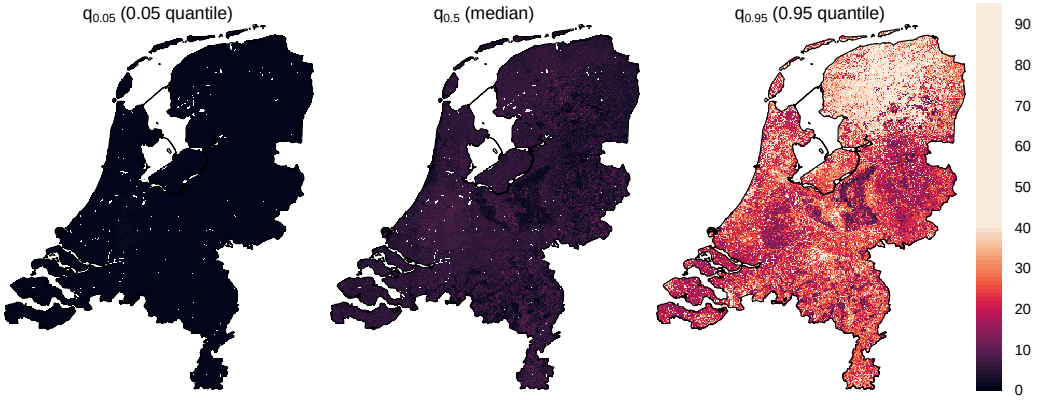

Figure S82. Predicted 5th, 50th (median) and 95th quantile for P_{ox} [mmol/kg] from 30 - 60 cm depth.

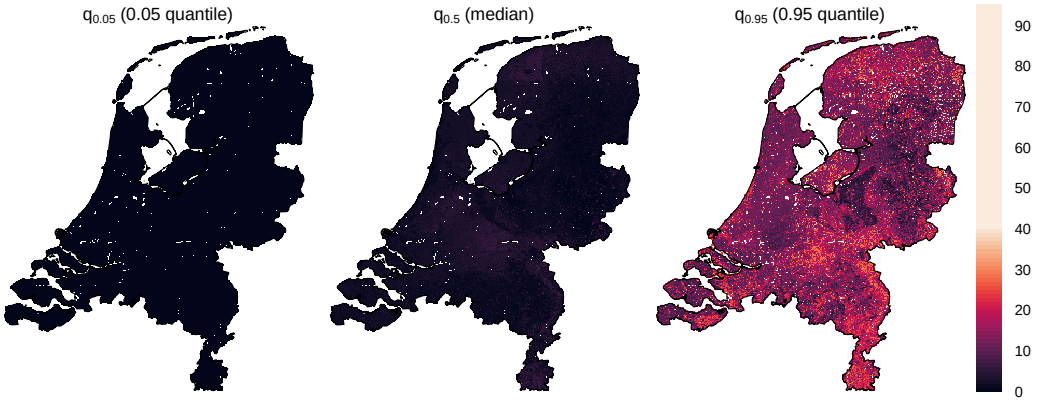

Figure S83. Predicted 5th, 50th (median) and 95th quantile for P_{ox} [mmol/kg] from 60 - 100 cm depth.

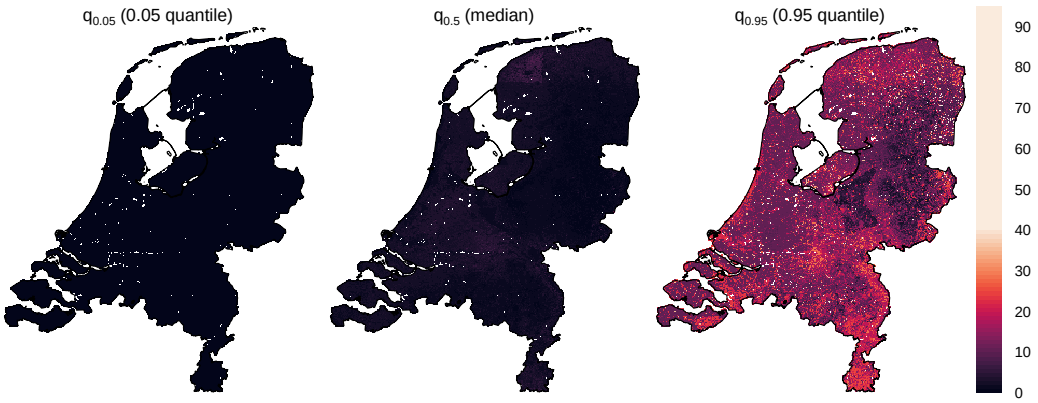

Figure S84. Predicted 5th, 50th (median) and 95th quantile for P_{ox} [mmol/kg] from 100 - 200 cm depth.

S8.2 Accuracy assessment

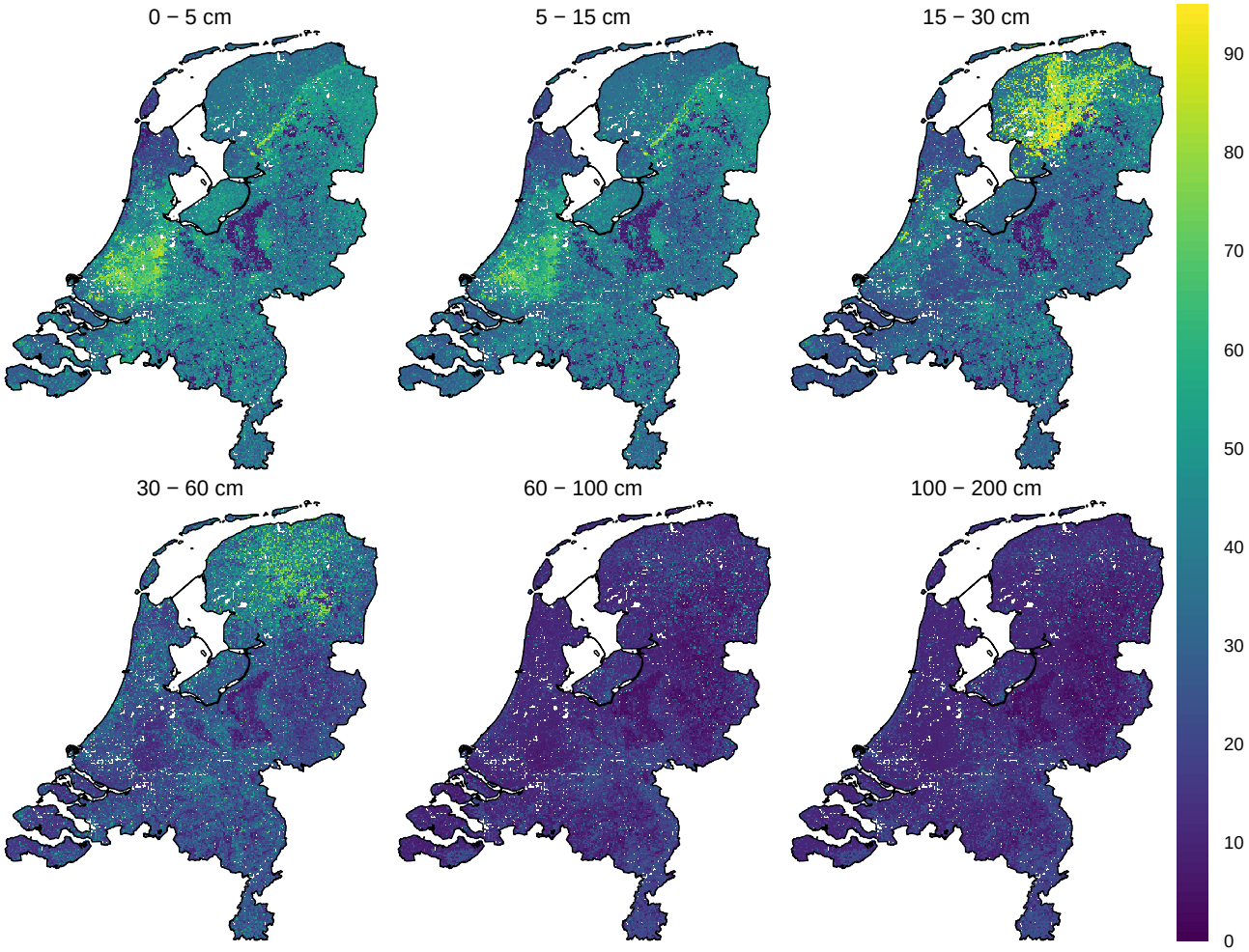

Figure S85. Maps of the PI90 as a measure of prediction uncertainty for P_{ox} [mmol/kg] for every GSM depth layer.

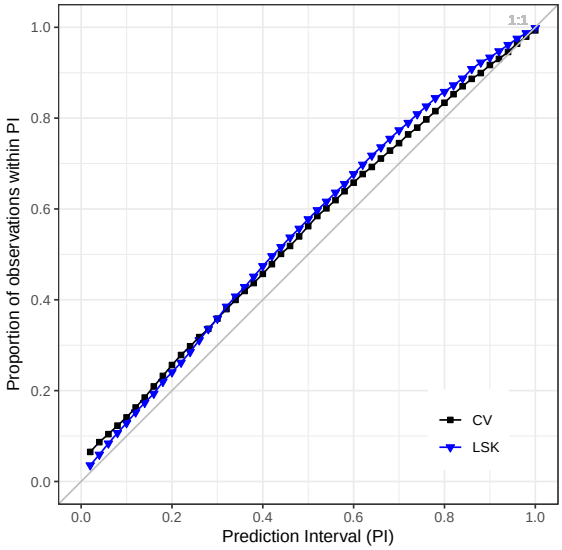

Figure S86. Prediction interval coverage probability (PICP) for prediction intervals between 0.02 and 1 of data used for different statistical validation strategies: 10-fold cross-validation of PFB laboratory measurements (CV) and LSK laboratory measurements (Sect. 2.6). The closer the points are to the 1:1 line, the more accurate the prediction uncertainty. Lines connecting the points do not represent actual data and are only for visual guidance.

Table S8. Model accuracy metrics of mean and median P_{ox} [mmol/kg] predictions using design-based inference of LSK laboratory measurements (Sect. 2.6). ME = mean error; RMSE = root mean squared error; MEC = model efficiency coefficient; PICP90 = prediction interval coverage probability of the PI90. The lower and upper 97.5% confidence limits of the accuracy metrics were computed using design-based inference according to de Gruijter et al. (2006) and Sect. 2.6.2 of Helfenstein et al. (2022). Together, the lower and upper 97.5% confidence limits give the 95% confidence interval (CI95).

Prediction	Depth (cm)	n	**ME**	CI95 of ME	**RMSE**	CI95 of RMSE	**MEC**	CI95 of MEC	**PICP90**
	0 - 15	1341	**-1.27**	-1.73, -0.8	**8.71**	8.09, 9.29	**0.25**	0.19, 0.3	**0.94**
	15 - 30	279	**-2.92**	-3.87, -1.96	**7.29**	6.13, 8.29	**0.17**	0.03, 0.28	**0.94**
Mean	30 - 60	1422	**-2.2**	-2.58, -1.82	**7.7**	6.94, 8.4	**-0.11**	-0.21, -0.05	**0.94**
	60 - 100	1342	**-0.19**	-0.52, 0.15	**6.03**	4.78, 7.07	**0.04**	0.01, 0.07	**0.92**
	100 - 200	948	**-0.86**	-1.15, -0.57	**4.16**	3.6, 4.65	**0.04**	-0.03, 0.09	**0.93**
	0 - 15	1341	**1.65**	1.17, 2.12	**8.97**	8.23, 9.66	**0.2**	0.14, 0.26	**0.94**
	15 - 30	279	**0.16**	-0.77, 1.09	**6.31**	4.7, 7.59	**0.38**	0.3, 0.46	**0.94**
Median	30 - 60	1422	**1.23**	0.88, 1.59	**7.12**	6.15, 7.97	**0.05**	0.02, 0.09	**0.94**
	60 - 100	1342	**1.71**	1.39, 2.04	**6.13**	4.83, 7.2	**0**	-0.02, 0.04	**0.92**
	100 - 200	948	**0.86**	0.58, 1.14	**4.01**	3.31, 4.59	**0.11**	0.07, 0.15	**0.93**

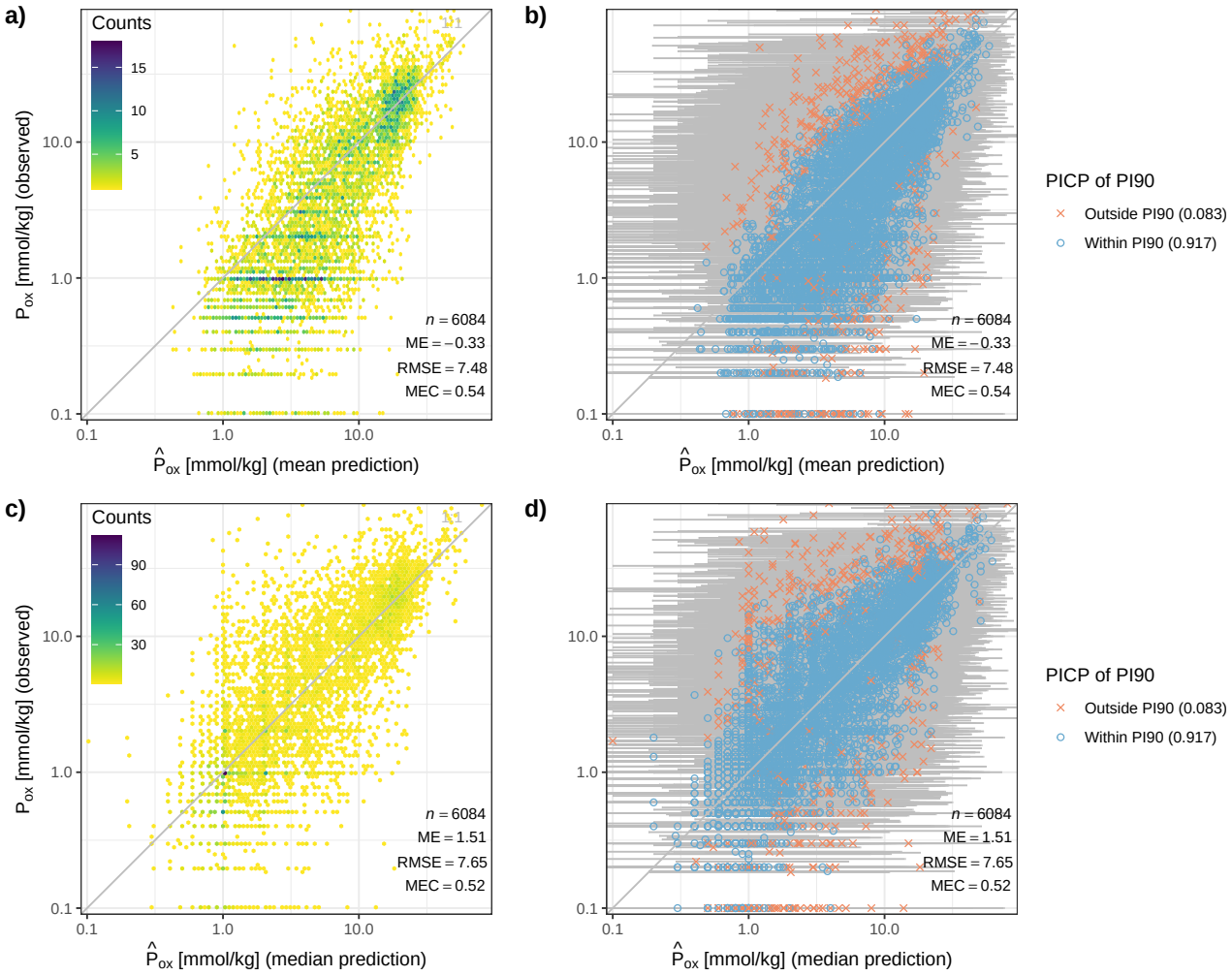

Figure S87. Predicted mean (a & b) and median (c & d) P_{ox} [mmol/kg] on the x-axis vs. measured P_{ox} [mmol/kg] on the y-axis (log-scale). Accuracy plots and metrics (ME, RMSE and MEC) were computed using 10-fold cross-validation of PFB laboratory measurements (Sect. 2.6). Plots a & c emphasize point density whereas plots b & d visualize prediction uncertainty (PI90 as error bars) and the PICP90 in the figure legends.

S8.3 Variable importance

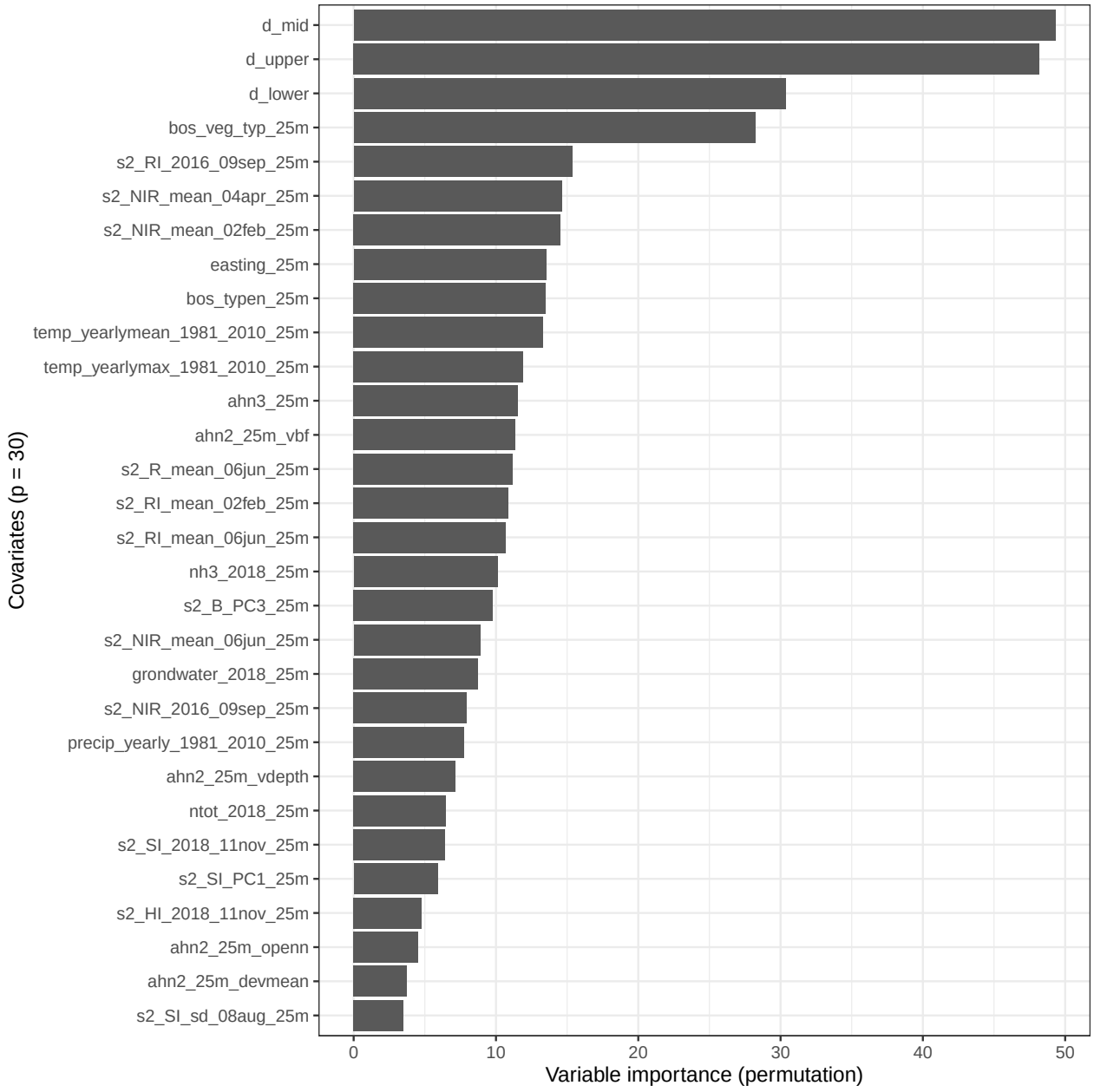

Figure S88. Variable importance for predicting P_{ox}, assessed using the permutation method (Breiman, 2002) (Sect. 2.4). "d_upper", "d_mid" and "d_lower" denote the upper, midpoint and lower boundary of a sampled soil horizon during calibration and target depth layer during prediction (Sect. 2.2, Table 5).

S9 Cation exchange capacity (CEC)

S9.1 Prediction maps

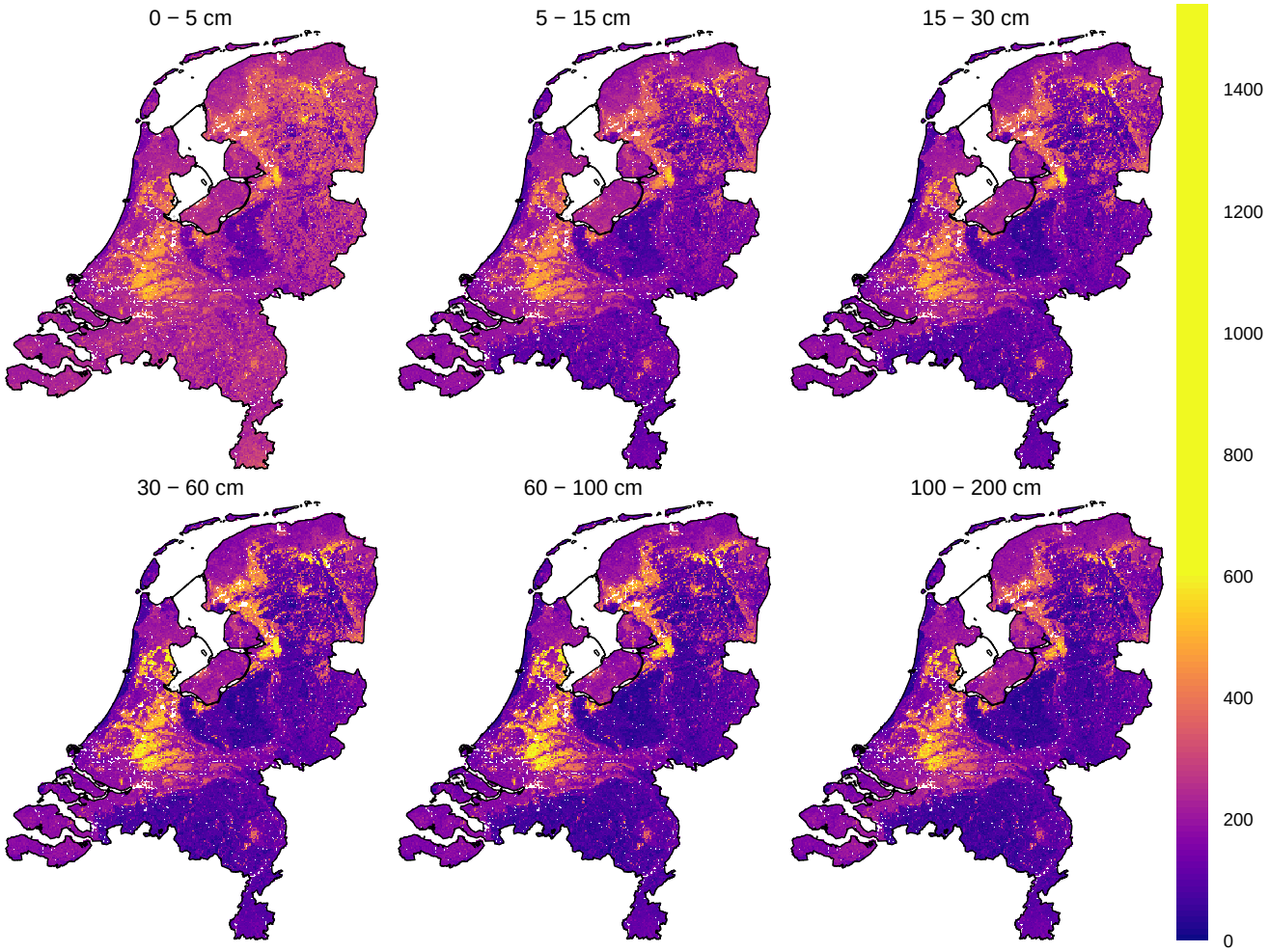

Figure S89. Predicted mean CEC [mmol(c)/kg] for every GSM depth layer.

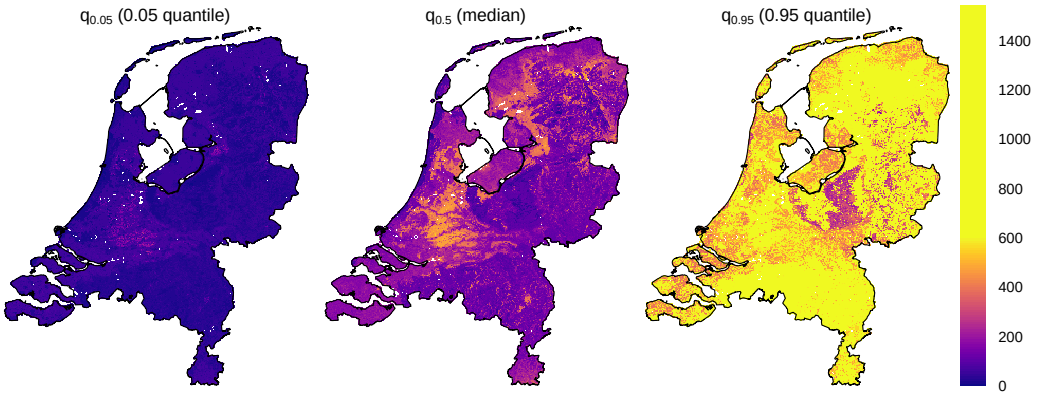

Figure S90. Predicted 5th, 50th (median) and 95th quantile for CEC [mmol(c)/kg] from 0 - 5 cm depth.

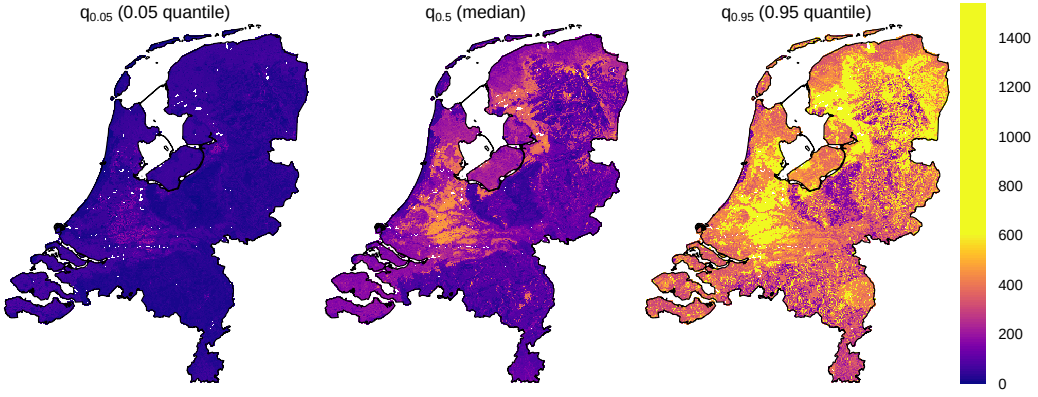

Figure S91. Predicted 5th, 50th (median) and 95th quantile for CEC [mmol(c)/kg] from 5 - 15 cm depth.

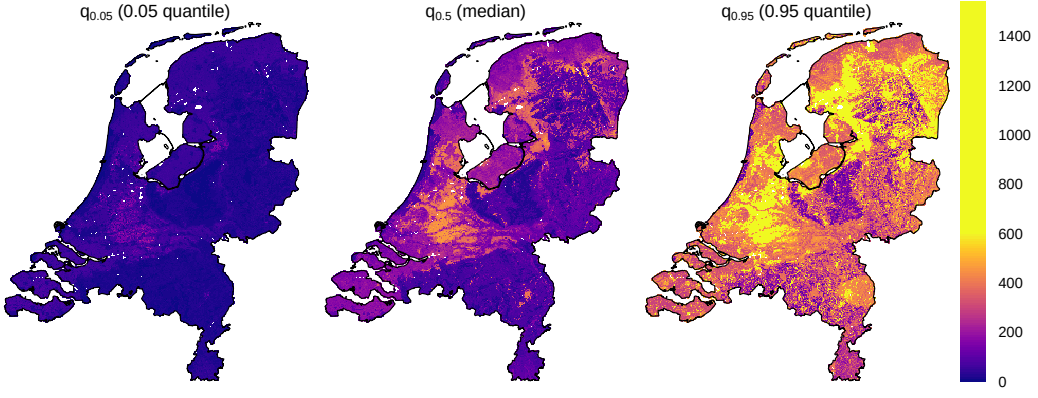

Figure S92. Predicted 5th, 50th (median) and 95th quantile for CEC [mmol(c)/kg] from 15 - 30 cm depth.

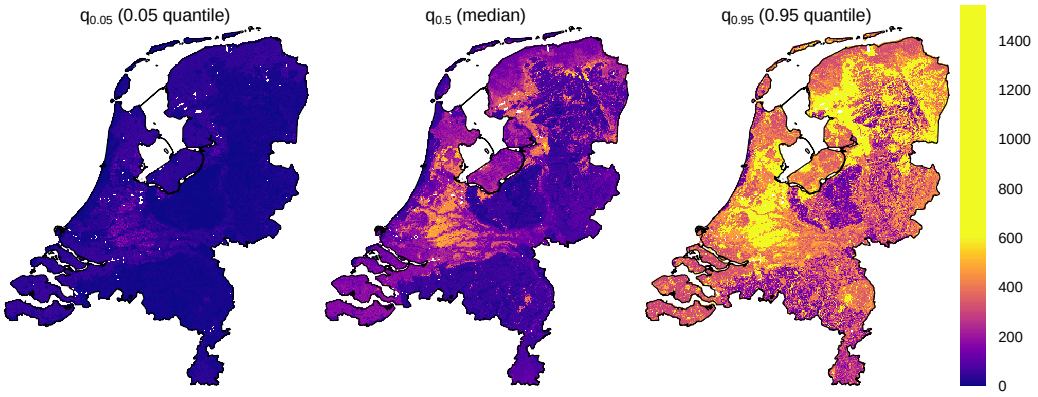

Figure S93. Predicted 5th, 50th (median) and 95th quantile for CEC [mmol(c)/kg] from 30 - 60 cm depth.

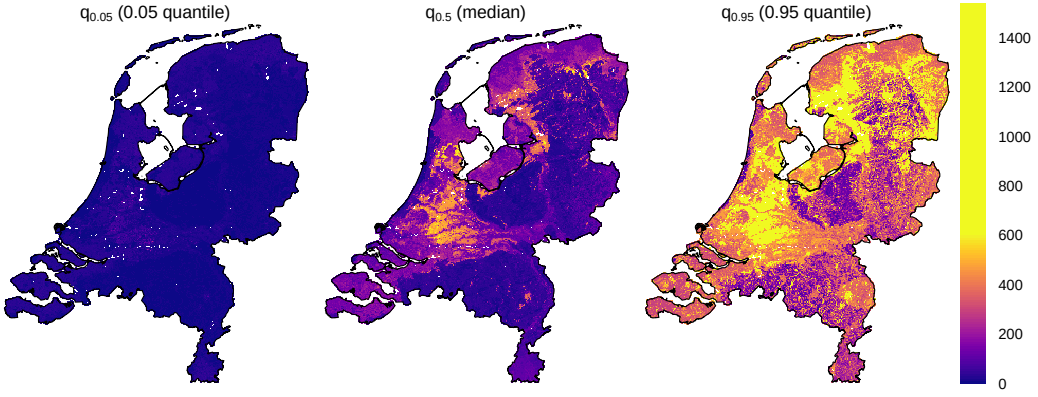

Figure S94. Predicted 5th, 50th (median) and 95th quantile for CEC [mmol(c)/kg] from 60 - 100 cm depth.

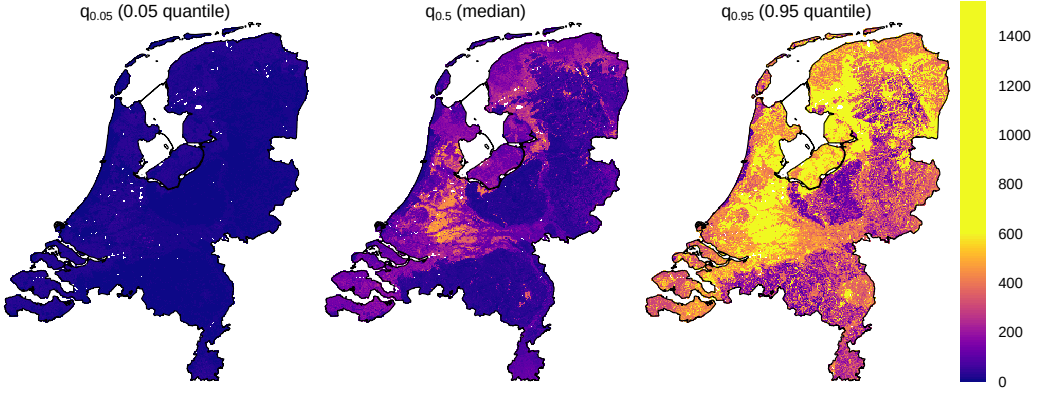

Figure S95. Predicted 5th, 50th (median) and 95th quantile for CEC [mmol(c)/kg] from 100 - 200 cm depth.

S9.2 Accuracy assessment

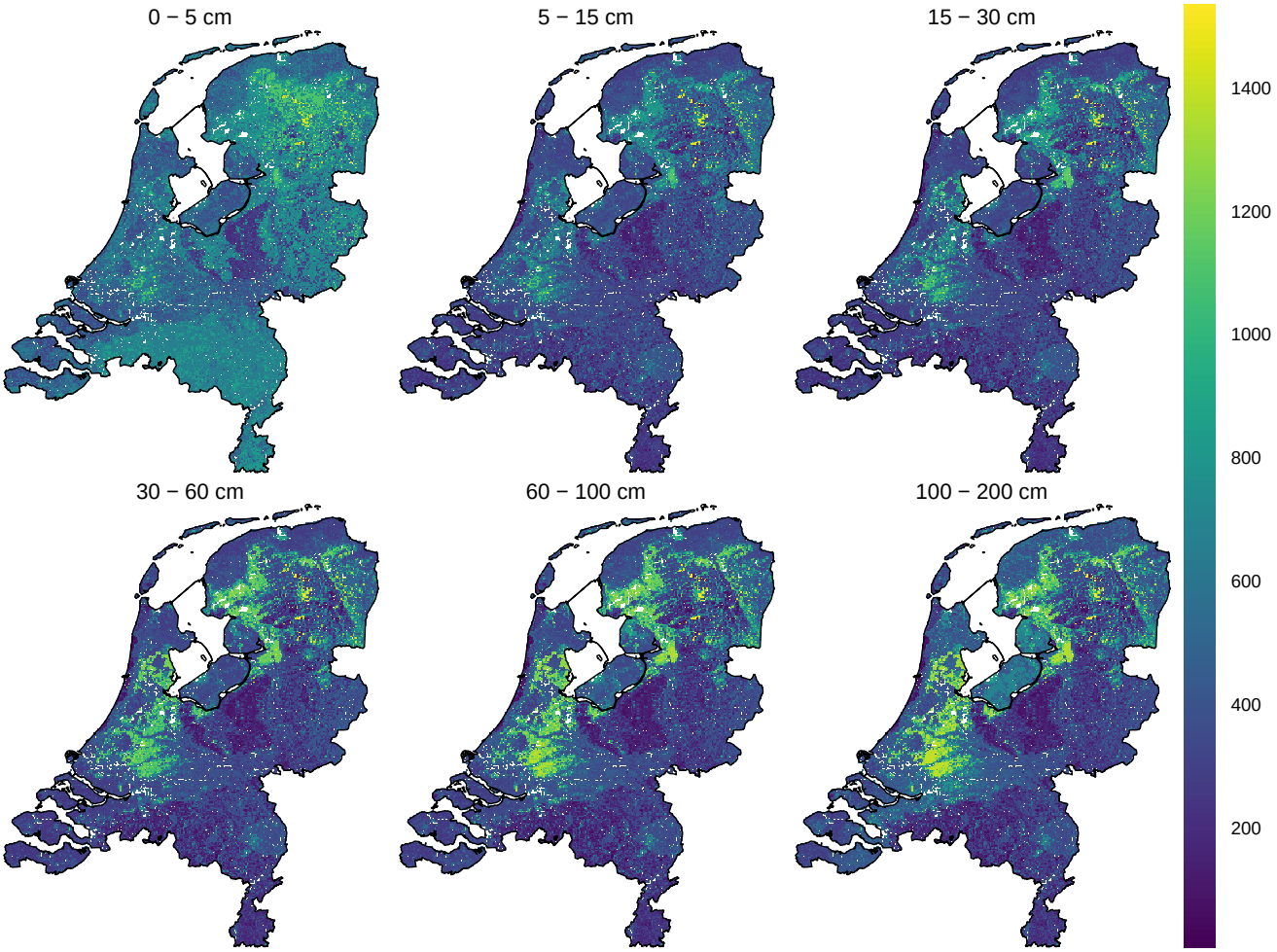

Figure S96. Maps of the PI90 as a measure of prediction uncertainty for CEC [mmol(c)/kg] for every GSM depth layer.

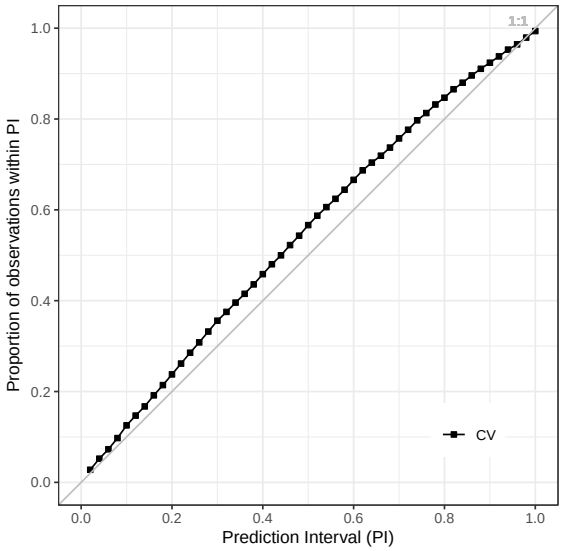

Figure S97. Prediction interval coverage probability (PICP) for prediction intervals between 0.02 and 1 of the PFB laboratory measurements used for 10-fold cross-validation (CV). The closer the points are to the 1:1 line, the more accurate the prediction uncertainty. Lines connecting the points do not represent actual data and are only for visual guidance.

Table S9. Model accuracy metrics of mean and median CEC [mmol(c)/kg] predictions using 10-fold cross-validation of PFB laboratory measurements (Sect. 2.6). Note that CEC was not measured in LSK or CCNL so design-based inference and computing 95 % confidence intervals (CI95) was not possible (Sect. 2.1.2). ME = mean error; RMSE = root mean squared error; MEC = model efficiency coefficient; PICP90 = prediction interval coverage probability of the PI90.

Prediction	Depth (cm)	n	**ME**	CI95 of ME	**RMSE**	CI95 of RMSE	**MEC**	CI95 of MEC	**PICP90**
	0 - 15	1489	**-5.12**	-	**114**	-	**0.59**	-	**0.93**
	15 - 30	502	**6.47**	-	**132**	-	**0.49**	-	**0.95**
Mean	30 - 60	824	**-0.25**	-	**142**	-	**0.47**	-	**0.92**
	60 - 100	728	**-3.69**	-	**156**	-	**0.38**	-	**0.91**
	100 - 200	269	**-24.02**	-	**123**	-	**0.16**	-	**0.9**
	0 - 15	1489	**22.51**	-	**119**	-	**0.56**	-	**0.93**
	15 - 30	502	**32.97**	-	**140**	-	**0.43**	-	**0.95**
Median	30 - 60	824	**29**	-	**149**	-	**0.42**	-	**0.92**
	60 - 100	728	**28.75**	-	**159**	-	**0.36**	-	**0.91**
	100 - 200	269	**13.51**	-	**115**	-	**0.26**	-	**0.9**

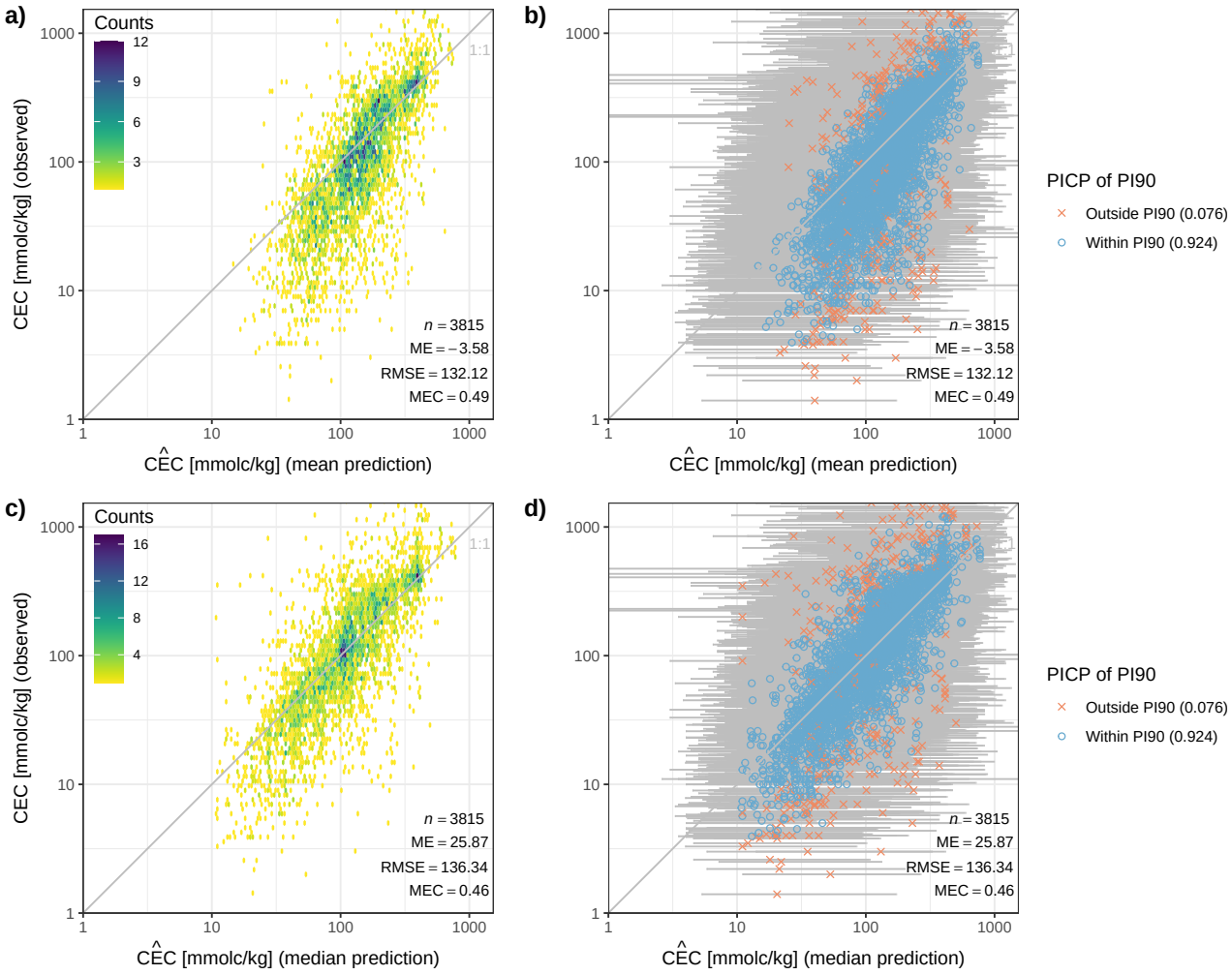

Figure S98. Predicted mean (a & b) and median (c & d) CEC [mmol(c)/kg] on the x-axis vs. measured CEC [mmol(c)/kg] on the y-axis (log-scale). Accuracy plots and metrics (ME, RMSE and MEC) were computed using 10-fold cross-validation of PFB laboratory measurements (Sect. 2.6). Plots a & c emphasize point density whereas plots b & d visualize prediction uncertainty (PI90 as error bars) and the PICP90 in the figure legends.

S9.3 Variable importance

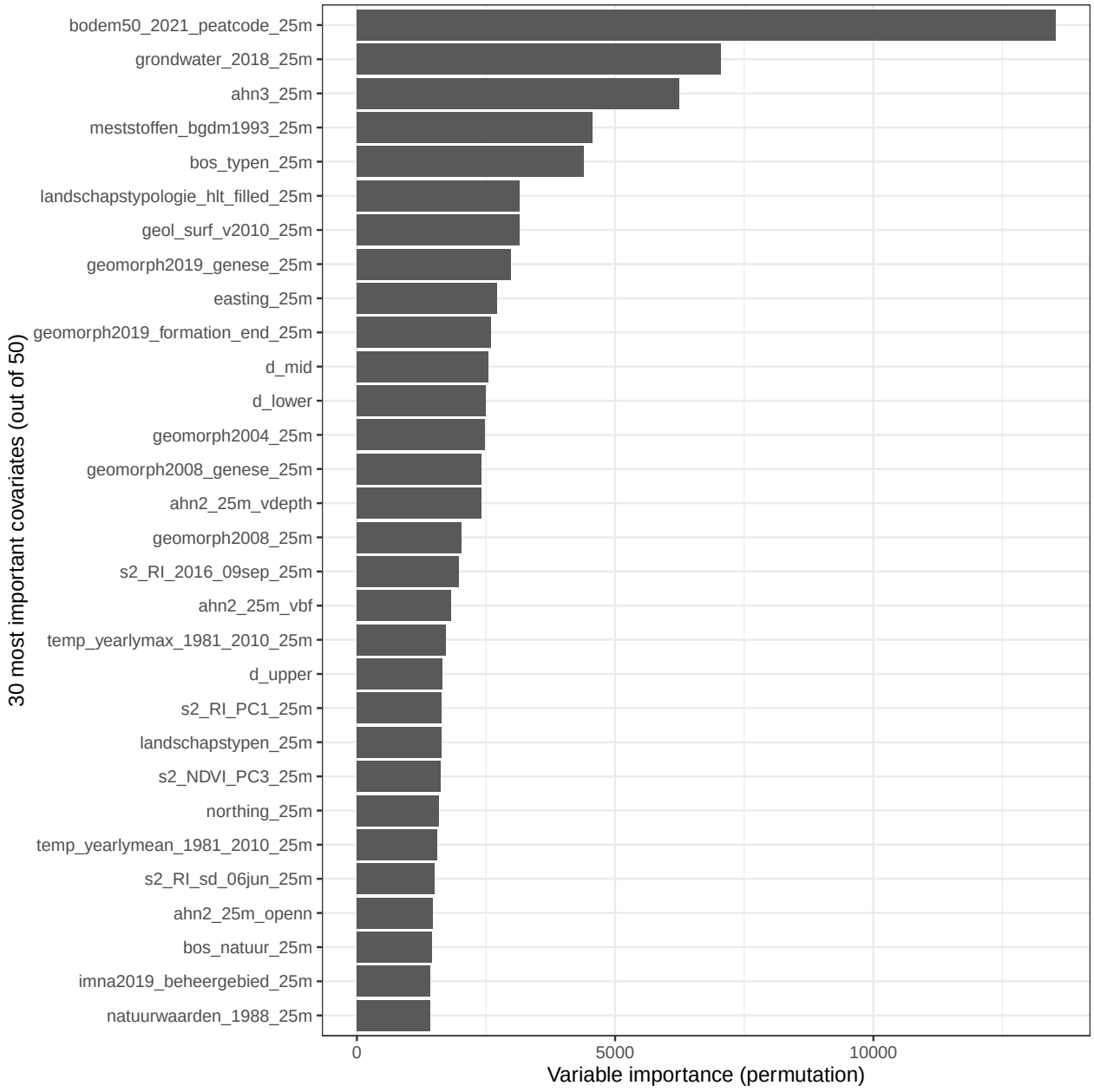

Figure S99. Variable importance for predicting CEC, assessed using the permutation method (Breiman, 2002) (Sect. 2.4). "d_upper", "d_mid" and "d_lower" denote the upper, midpoint and lower boundary of a sampled soil horizon during calibration and target depth layer during prediction (Sect. 2.2, Table 5).

References

Arrouays, D., McBratney, A., Minasny, B., Hempel, J., Heuvelink, G.B.M., MacMillan, R.A., Hartemink, A., Lagacherie, P., McKenzie, N., 2015. The GlobalSoilMap project specifications, in: Proceedings of the 1st GlobalSoilMap Conference, pp. 9–12. URL: https://www.isric.org/sites/default/files/GlobalSoilMap_specifications_december_2015_2.pdf, doi:https://doi.org/10.1201/b16500-4.

Breiman, L., 2002. Manual on Setting Up, Using, and Understanding Random Forests v3.1. Technical report ftp://ftp.stat.berkeley.edu/pub/users/breiman/Using_random_forests_v3.1.pdf. URL: ftp://ftp.stat.berkeley.edu/pub/users/breiman/Using_random_forests_v3.1.pdf.

de Gruijter, J.J., Brus, D., Bierkens, M., Knotters, M., 2006. Sampling for Natural Resource Monitoring. Springer, The Netherlands.

Helfenstein, A., Mulder, V.L., Heuvelink, G.B., Okx, J.P., 2022. Tier 4 maps of soil pH at 25 m resolution for the Netherlands. Geoderma 410, 115659. URL: https://linkinghub.elsevier.com/retrieve/pii/S0016706121007394, doi:https://doi.org/10.1016/j.geoderma.2021.115659.

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Hack-ten Broeke, M.J., 2024. Three-dimensional space and time mapping reveals soil organic matter decreases across anthropogenic landscapes in the Netherlands. Under review for Communications Earth and Environment.

Sandri, M., Zuccolotto, P., 2008. A Bias Correction Algorithm for the Gini Variable Importance Measure in Classification Trees. Journal of Computational and Graphical Statistics 17, 611–628. URL: https://doi.org/10.1198/106186008X344522, doi:https://doi.org/10.1198/106186008X344522. publisher: Taylor & Francis _eprint: https://doi.org/10.1198/106186008X344522.

Sandri, M., Zuccolotto, P., 2010. Analysis and correction of bias in Total Decrease in Node Impurity measures for tree-based algorithms. Statistics and Computing 20, 393–407. URL: https://doi.org/10.1007/s11222-009-9132-0, doi:https://doi.org/10.1007/s11222-009-9132-0.