Supplemental Information for:

Unified Conversion Equations between Olsen and Mehlich-3 Soil Phosphorus Tests

Zachary P. Simpson [1,*], Christopher W. Rogers [2,*], Joshua D. Mott [3,4], Natalja Čerkasova [5,6], Kyle R. Elkin [7], Robert O. Miller [8], Carl H. Bolster [9], Michael J. White [10], P.J.A. Kleinman [11]

[1] Department of Ecoscience, Aarhus University, Aarhus, Denmark
[2] USDA-ARS Northwest Irrigation and Soils Research Laboratory, Kimberly, ID, USA
[3] Southern Piedmont Agricultural Research and Extension Center, Virginia Tech, Blackstone, VA, USA
[4] School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, USA
[5] Texas A&M AgriLife Research, Temple, TX, USA
[6] Klaipėda University, Marine Research Institute, Klaipėda, Lithuania
[7] Agricultural Analytical Services Laboratory, The Pennsylvania State University, PA, USA
[8] Agricultural Laboratory Proficiency Program, Collaborative Testing Services Inc., Windsor, CO, USA
[9] USDA-ARS Food Animal Environmental System Research Unit, Bowling Green, KY, USA
[10] USDA-ARS Grassland Soil and Water Research Laboratory, Temple, TX, USA
[11] USDA-ARS Soil Management and Sugarbeet Research Unit, Fort Collins, CO, USA

[*] These authors contributed equally to this work.

Contents:

Statistical models for ratio of Olsen P to Mehlich-3 P

Table S1. Models for the conversion of Mehlich-3 P to Olsen P. Each model uses different sources of information to predict *R* (ratio of Olsen P to Mehlich-3 P); those marked with * are focused in Figure 3 of main text and with # in Figure S2. The five categories of models described in the main text are indicated here, just that the GLM category is lumped with the respective STP models. All models assume the Gamma distribution with a log link. The model terms describe in shorthand the predictors for *R* (on the log link scale): an overall intercept (except for first model, not shown) is included in each model; 'site' on its own is a fixed effect for site (i.e., 8 intercepts for combined dataset); '(1 | site)' refers to a varying intercept for site; s() and te() refer to 'smoothing spline' and 'tensor product smooth' functions of the variables; ti() is a 'tensor product interaction' term and is used here when decomposing te() terms into marginal vs. interactive effects. All continuous predictors were log-transformed beforehand except for pH (no transform) and TIC (log1p transformed). Notes briefly indicate changes in the model structure, usually relative to the prior model. Akaike information criterion (AIC) is for the given dataset; other statistics are based on 10-fold cross-validation (CV). The root mean square log error (RMSLE) is expressed both as the log error value and parenthetically as (relative) percent error in *R*. Due to changes in datasets, AIC and CV statistics should only be compared row-wise between models for a given dataset. DPS_{ox}: degree of P saturation via oxalate extraction; DCS_{ox}: molar ratio of TOC to (Al_{ox} + Fe_{ox}); MAD: median absolute deviation.

			ALP and Legacy P data				Only Legacy P data			
Model	Model terms	Notes	AIC	% bias	MAD	RMSLE	AIC	% bias	MAD	RMSLE
Region-level intercept(s)										
***Overall**	*R* ~ 1	Overall mean	596	0.79	0.284	0.716 (104.5%)	622	1.6	0.35	0.779 (117.8%)
***Site-based**	*R* ~ site	Mean *R* by site	-304	0.71	0.115	0.418 (52%)	-1	1.36	0.141	0.424 (52.7%)
Sample-level: Basic predictors										
No STP: 1	*R* ~ (1 | site) + s(clay) + s(TOC) + s(pH) + s(TIC)	Only basic properties	-509	0.16	0.105	0.384 (46.8%)	-195	2.75	0.123	0.383 (46.6%)
***No STP: 2**	*R* ~ (1 | site) + s(clay) + s(TOC) + te(pH, TIC)	Only basic properties	-519	0.07	0.107	0.383 (46.7%)	-215	1.49	0.12	0.371 (44.9%)
Sample level: Basic predictors + Mehlich-3										
With M3: 1	*R* ~ (1 | site) + s(clay) + s(TOC) + te(pH, TIC) + s(P_{M3})	As above, now including STP (Mehlich-3)	-1211	-0.42	0.068	0.266 (30.4%)	-970	0.7	0.057	0.197 (21.8%)
With M3: 2	*R* ~ (1 | site) + s(clay) + s(TOC) + te(pH, TIC) + s(P_{M3}) + s(Fe_{M3})	Introduce Fe	-1341	-0.47	0.058	0.251 (28.6%)	-1109	0.63	0.051	0.179 (19.6%)
With M3: 3	*R* ~ (1 | site) + s(clay) + s(TOC) + te(pH, TIC) + s(P_{M3}) + s(Fe_{M3}) + s(Al_{M3})	Introduce Al	-1381	-0.51	0.058	0.245 (27.8%)	-1118	0.71	0.051	0.178 (19.5%)
With M3: 4	*R* ~ (1 | site) + s(clay) + s(TOC) + te(pH, TIC) + s(P_{M3}) + te(Fe_{M3}, Al_{M3})	Introduce Fe:Al interaction	-1388	-0.6	0.058	0.245 (27.8%)	-1169	0.66	0.049	0.171 (18.6%)

With M3: 5	*R* ~ (1 | site) + s(clay) + s(TOC) + te(pH, TIC) + s(P_{M3}) + te(Fe_{M3}, Al_{M3}) + s(Ca_{M3})	Introduce Ca	-1386	-0.62	0.06	0.248 (28.1%)	-1215	0.95	0.046	0.171 (18.6%)
With M3: 6	*R* ~ (1 | site) + s(clay) + s(TOC) + te(pH, TIC) + s(P_{M3}) + te(Fe_{M3}, Al_{M3}) + s(Ca_{M3}) + s(Mg_{M3})	Introduce Mg	-1466	-0.74	0.058	0.242 (27.4%)	-1287	0.36	0.045	0.163 (17.7%)
With M3: 7	*R* ~ (1 | site) + s(clay) + s(TOC) + te(pH, TIC) + s(P_{M3}) + te(Fe_{M3}, Al_{M3}) + te(Ca_{M3}, Mg_{M3})	Introduce Ca:Mg interaction	-1470	-0.96	0.057	0.238 (26.9%)	-1281	0.89	0.044	0.164 (17.8%)
With M3: 8	*R* ~ (1 | site) + s(clay) + s(TOC) + te(pH, TIC) + s(P_{M3}) + te(Fe_{M3}, Al_{M3}) + s(Ca_{M3}) + s(Mg_{M3}) + s(Mn_{M3})	Add Mn, drop Ca:Mg	-1478	-0.77	0.057	0.238 (26.8%)	-1304	0.31	0.042	0.162 (17.6%)
***With M3: 9**	*R* ~ (1 | site) + s(clay) + s(TOC) + te(pH, TIC) + s(P_{M3}) + te(Fe_{M3}, Al_{M3}) + te(Ca_{M3}, Mg_{M3}) + s(Mn_{M3})	Add Ca:Mg back in	-1499	-0.92	0.057	0.233 (26.2%)	-1298	0.64	0.044	0.16 (17.4%)
***With M3: GLM**	*R* ~ (1 | site) + clay + TOC + pH*TIC + P_{M3} + Fe_{M3} * Al_{M3} + Ca_{M3} * Mg_{M3} + Mn_{M3}	Linear version of M3 model 9	-1311	-1.21	0.063	0.256 (29.2%)	-1166	0.73	0.048	0.172 (18.8%)
Sample level: Basic predictors + Olsen										
***With Olsen: 1**	*R* ~ (1 | site) + s(clay) + s(TOC) + te(pH, TIC) + s(P_{Ols})	Basic properties and STP (Olsen)	-646	-0.54	0.1	0.362 (43.6%)	-455	1.03	0.091	0.3 (35%)
***With Olsen: GLM**	*R* ~ (1 | site) + clay + TOC + pH * TIC + P_{Ols}	Linear version of Olsen model 1	-562	-0.5	0.108	0.371 (45%)	-319	0.78	0.112	0.331 (39.3%)

With Olsen: + BWI	R ~ (1 $\mid$ site) + s(clay) + s(TOC) + te(pH, TIC) + s(P_{Ols}) + s(BWI)	Bonus Olsen model including Bache-Williams index as P buffer predictor					-632	0.18	0.076	0.263 (30.1%)
				Sample level: Lability						
P lability: 1	R ~ (1 $\mid$ site) + s(DPS_{ox}) + s(clay) + te(pH, TIC) + s(DCS_{ox})	No STP information; emphasis on oxalate extraction					-627	1.15	0.084	0.261 (29.9%)
P lability: 2	R ~ (1 $\mid$ site) + s(EPC_0) + s(P_{AEM}) + s(BWI) + te(pH, TIC) + s(DCS_{ox})	The basic 'lability' model (intensity, quantity, buffer capacity) based on Simpson et al. (2025)					-706	-0.15	0.081	0.253 (28.8%)
P lability: 3	R ~ (1 $\mid$ site) + s(EPC_0) + s(P_{AEM}) + s(BWI) + te(pH, TIC) + s(TOC) + s(DCS_{ox})	Add in TOC					-769	0.53	0.077	0.239 (27%)
P lability: 4	R ~ (1 $\mid$ site) + s(EPC_0) + s(P_{AEM}) + s(BWI) + te(pH, TIC) + s(TOC) + s(DCS_{ox}) + s($(Fe{:}P)_{BD}$)	Include reductively-soluble Fe:P					-778	0.46	0.075	0.237 (26.7%)
P lability: 5	R ~ (1 $\mid$ site) + s(EPC_0) + s(P_{AEM}) + s(BWI) + te(pH, TIC) + s(TOC) + s(DCS_{ox}) + s($(Fe{:}P)_{BD}$) + s(DPS_{ox})	Re-introduce DPS_{ox}					-859	0.39	0.071	0.224 (25.1%)
P lability: 6	R ~ (1 $\mid$ site) + s(P_{AEM}) + s(BWI) + te(pH, TIC) + s(TOC) + s(DCS_{ox}) + s($(Fe{:}P)_{BD}$) + s(DPS_{ox})	Drop EPC_0 (intensity) as other predictors essentially lead to same information (see model in Simpson et al. 2025)					-857	0.34	0.068	0.222 (24.9%)

P lability: 7	*R* ~ (1 | site) + te(P_{AEM}, BWI) + te(pH, TIC) + s(TOC) + s(DCS_{ox}) + s($(Fe:P)_{BD}$) + s(DPS_{ox})	Let labile P and buffer capacity interact					-855	0.39	0.069	0.223 (25%)
P lability: 8	*R* ~ (1 | site) + s(P_{AEM}) + s(BWI) + te(pH, TIC) + s(TOC) + s(DCS_{ox}) + s($(Fe:P)_{BD}$) + s($(Al+Fe)_{ox}$)	Above interaction not helpful; drop DPS_{ox} and instead just focus the $(Fe+Al)_{ox}$ denominator					-783	0.29	0.07	0.236 (26.6%)
P lability: 9	*R* ~ (1 | site) + s(P_{AEM}) + s(BWI) + te(pH, TIC) + s(TOC) + s(DCS_{ox}) + s($(Fe:P)_{BD}$) + s($(Al+Fe)_{ox}$) + s(P_{ox})	P_{ox} appears important, likely connection to M3 chemistry; introduce as separate term (outside of DPS_{ox})					-865	0.29	0.069	0.222 (24.8%)
P lability: 10	*R* ~ (1 | site) + s(P_{AEM}) + s(BWI) + te(pH, TIC) + s(TOC) + s(DCS_{ox}) + s($(Fe:P)_{BD}$) + te($(Al+Fe)_{ox}$, P_{ox})	Akin to DPS_{ox}, let the numerator and denominator terms interact.					-877	0.22	0.065	0.219 (24.5%)
P lability: 11	*R* ~ (1 | site) + s(P_{AEM}) + s(BWI) + te(pH, TIC) + ti(TOC) + ti($(Al+Fe)_{ox}$) + ti(TOC, $(Al+Fe)_{ox}$) + ti(P_{ox}) + ti($(Al+Fe)_{ox}$, P_{ox}) + s($(Fe:P)_{BD}$)	Refine the oxalate terms – turn into marginal and interaction terms; drop $(Fe:P)_{BD}$ as it seems weak					-859	0.38	0.067	0.229 (25.8%)
# P lability: 12	*R* ~ (1 | site) + s(P_{AEM}) + s(BWI) + te(pH, TIC) + te(TOC, $(Al+Fe)_{ox}$) + s(P_{ox}) + s($(Fe:P)_{BD}$)	As above, but drop the less impactful interaction between P_{ox} and $(Al+Fe)_{ox}$; simplify ti() terms back to te() and s()					-864	0.42	0.069	0.224 (25.1%)

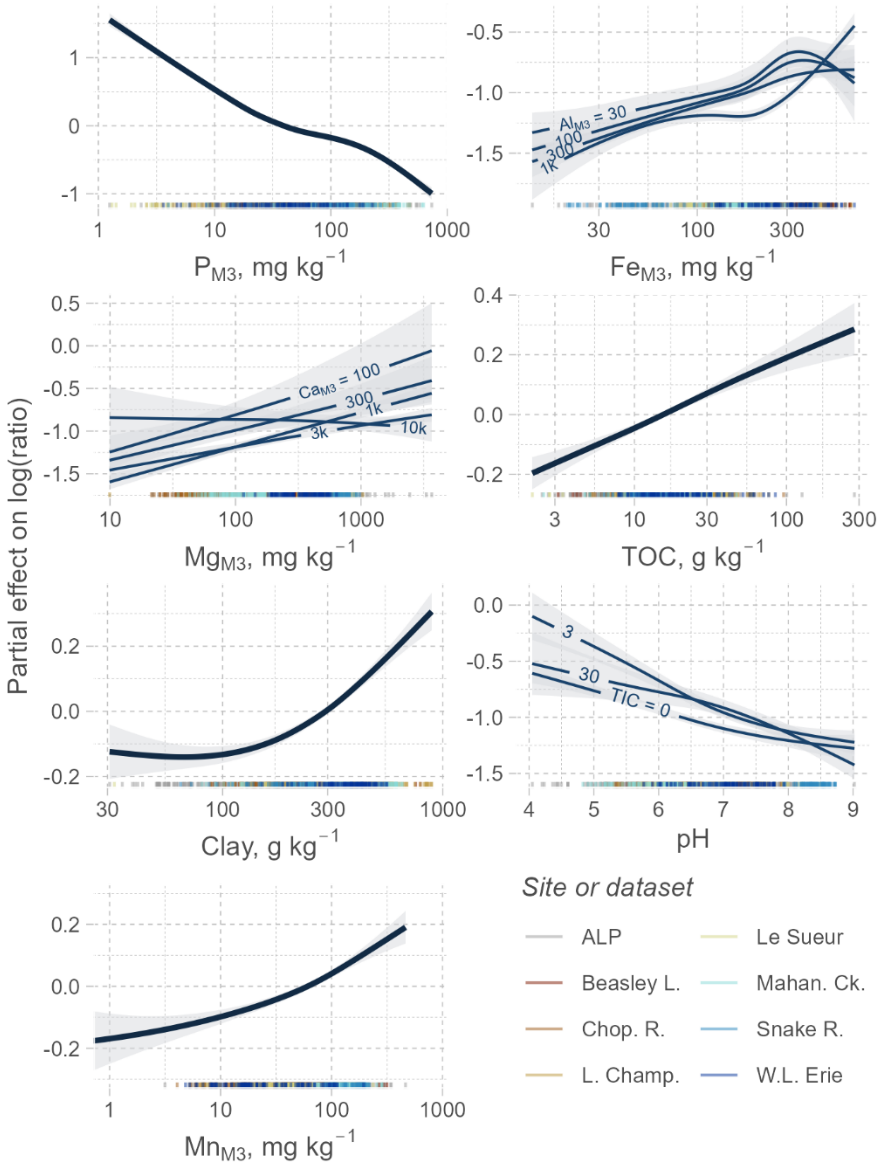

Figure S1. The partial effects for the Mehlich-3 GAM of the ratio of P_{Ols} to P_{M3}, *R*. Each panel shows the predictor's partial effect (and its 95% credible interval as the shaded ribbon) on *R* (holding all other predictors at their medians). For interaction terms, one of the two predictors is split into representative levels with separate curves drawn. Note that all predictor units are as shown in Table 1 (main text). The colored rug plot along the bottom indicates the marginal distribution of the *x*-axis predictor across the sites (Legacy P) or dataset (ALP). Partial effects for interactions are shown across the full *x*-axis though they may not be realistic (e.g., TIC >0 g kg^{-1} when pH <6).

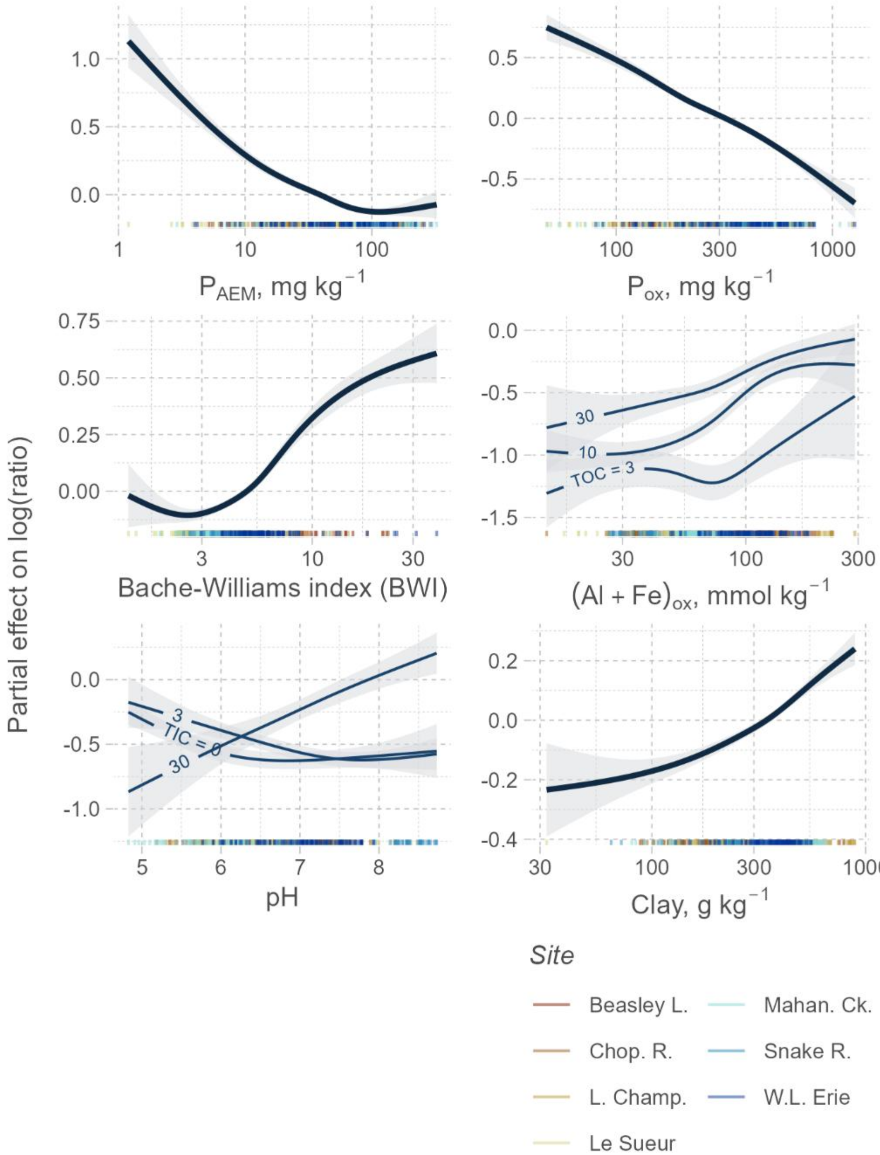

Figure S2. The partial effects for the 'lability' GAM (model number 12; see Table S1) of the ratio of P_{Ols} to P_{M3}, *R*. Each panel shows the predictor's partial effect (and its 95% credible interval as the shaded ribbon) on *R*, holding all other predictors at their medians. For interaction terms, one of the two predictors is split into representative levels with separate curves drawn. Note that no soil test P data (any result from the Olsen and Mehlich-3 extractions) were used and that ALP data were not useable here. All predictor units are as shown in Table 1; BWI has units of (mg P 100 g^{-1}) per log(µmol P L^{-1}). The colored rug plot along the bottom indicates the marginal distribution of the *x*-axis predictor across the sites. Partial effects for interactions are shown across the full *x*-axis though they may not be realistic (e.g., TIC >0 g kg^{-1} when pH <6).

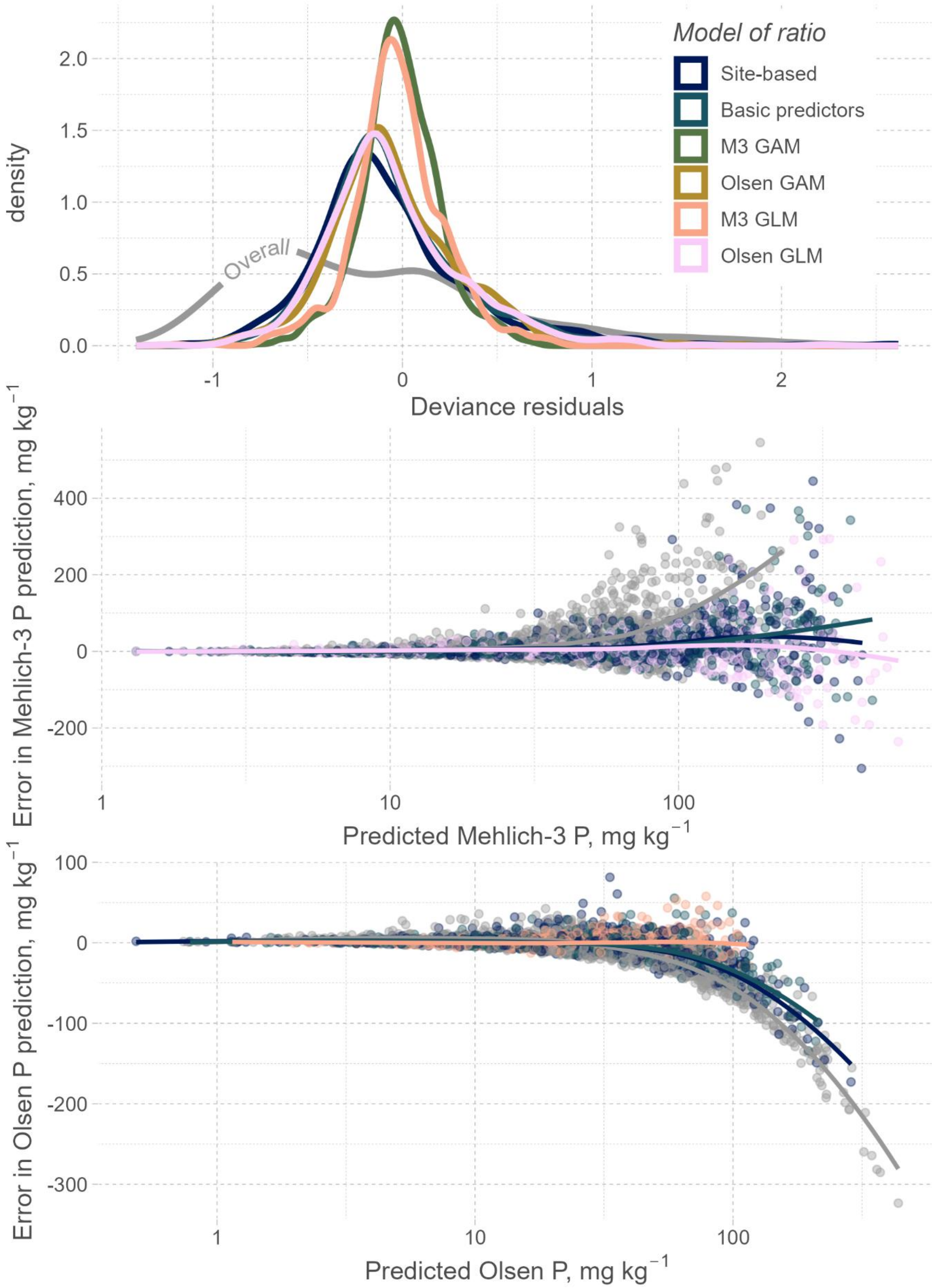

Figure S3. Residual diagnostic plots for the conversion ratio (*R*) models highlighted in Figure 3 (main text). Top panel shows deviance residual distributions; ideally, these should have mean zero, appear somewhat symmetrical, and not stretch out too widely. The other two panels show prediction errors in Mehlich-3 P (middle) and Olsen P (bottom) as functions of the predicted concentration. I.e., *R* is predicted and then subsequently used to derive the other soil test P concentration. For clarity, only the GLM versions of our preferred models are shown. Note that only the GLMs maintain a residual distribution centered around zero as soil test P magnitude increases. Lines in bottom two panels are LOESS smoothers for visual clarity.

Validation exercise with literature equations

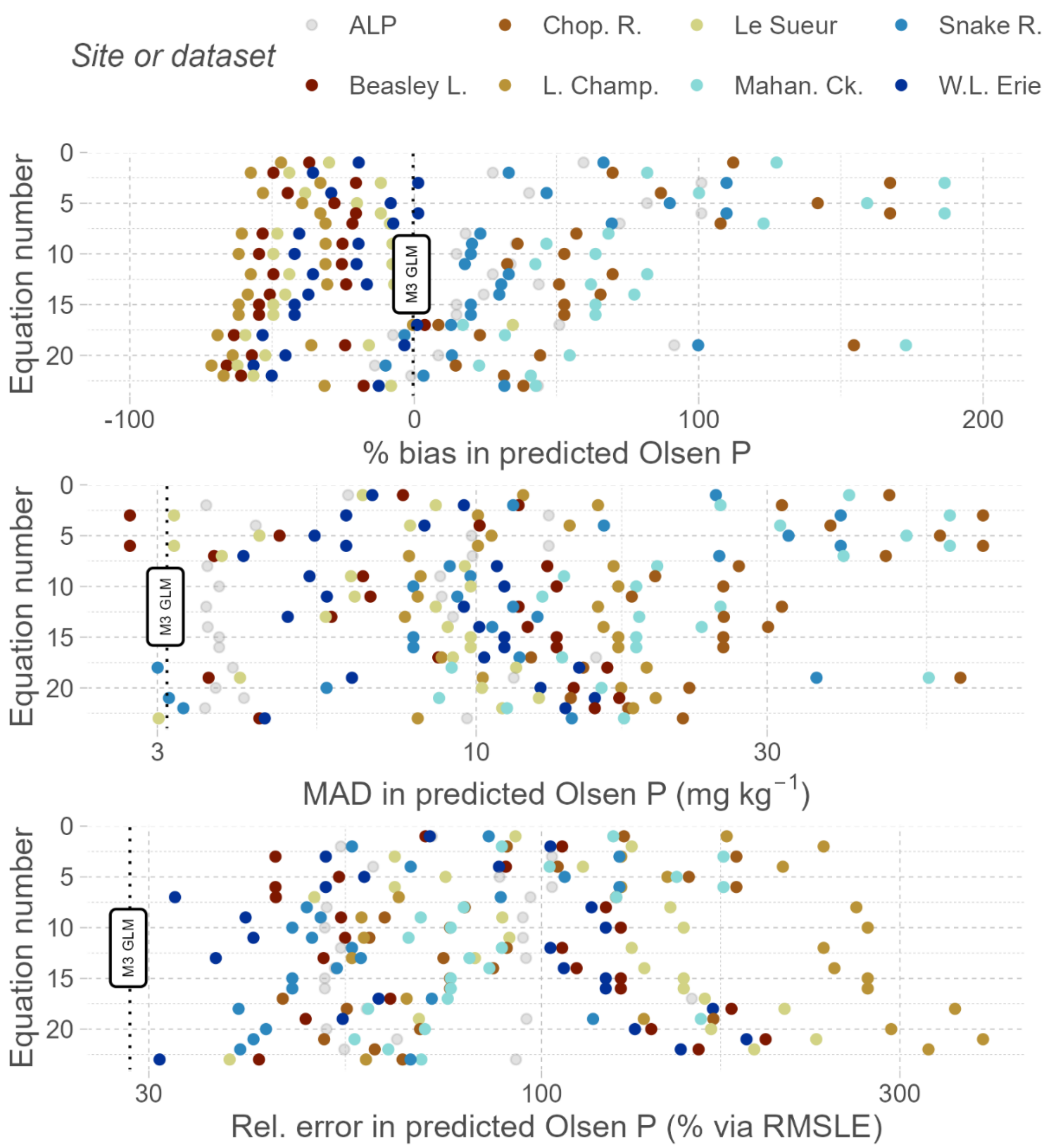

Figure S4. Validation errors when predicting Olsen P concentration given Mehlich-3 P concentration via a published regression relationship based only on correlation between Olsen P and Mehlich-3 P. For this, we used the equations reviewed by Steinfurth et al. (2021); the order in their Table 9 is the 'equation number' here. When applied to any subset of the ALP and Legacy P dataset, the potential error is usually substantially worse than, for example, the Mehlich-3 GLM developed in the current study (vertical dotted line). Note that panels for MAD and relative error are shown on log-scale.

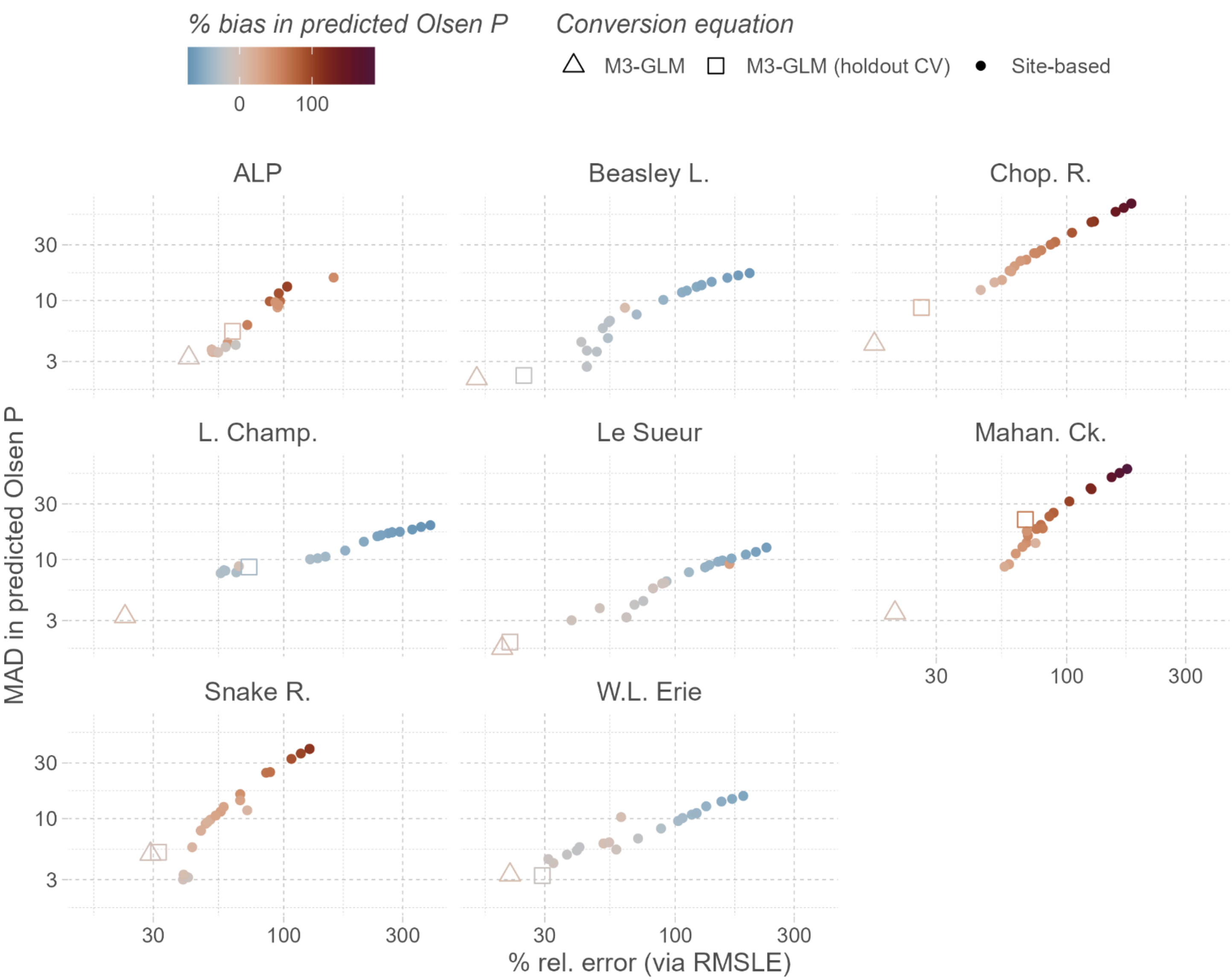

Figure S5. Validation errors when predicting Olsen P concentration given Mehlich-3 P concentration via a published regression relationship based only on correlation between Olsen P and Mehlich-3 P (sourced from Steinfurth et al. (2021)); see Figure S4 for the same statistics grouped by specific equation. In comparison, the Mehlich-3 GLM developed in the present study is shown based on applying the final equation to each site (triangle) and, for a much stricter test, based on leave-one-site-out cross-validation (square). The latter is displayed for the site held out.

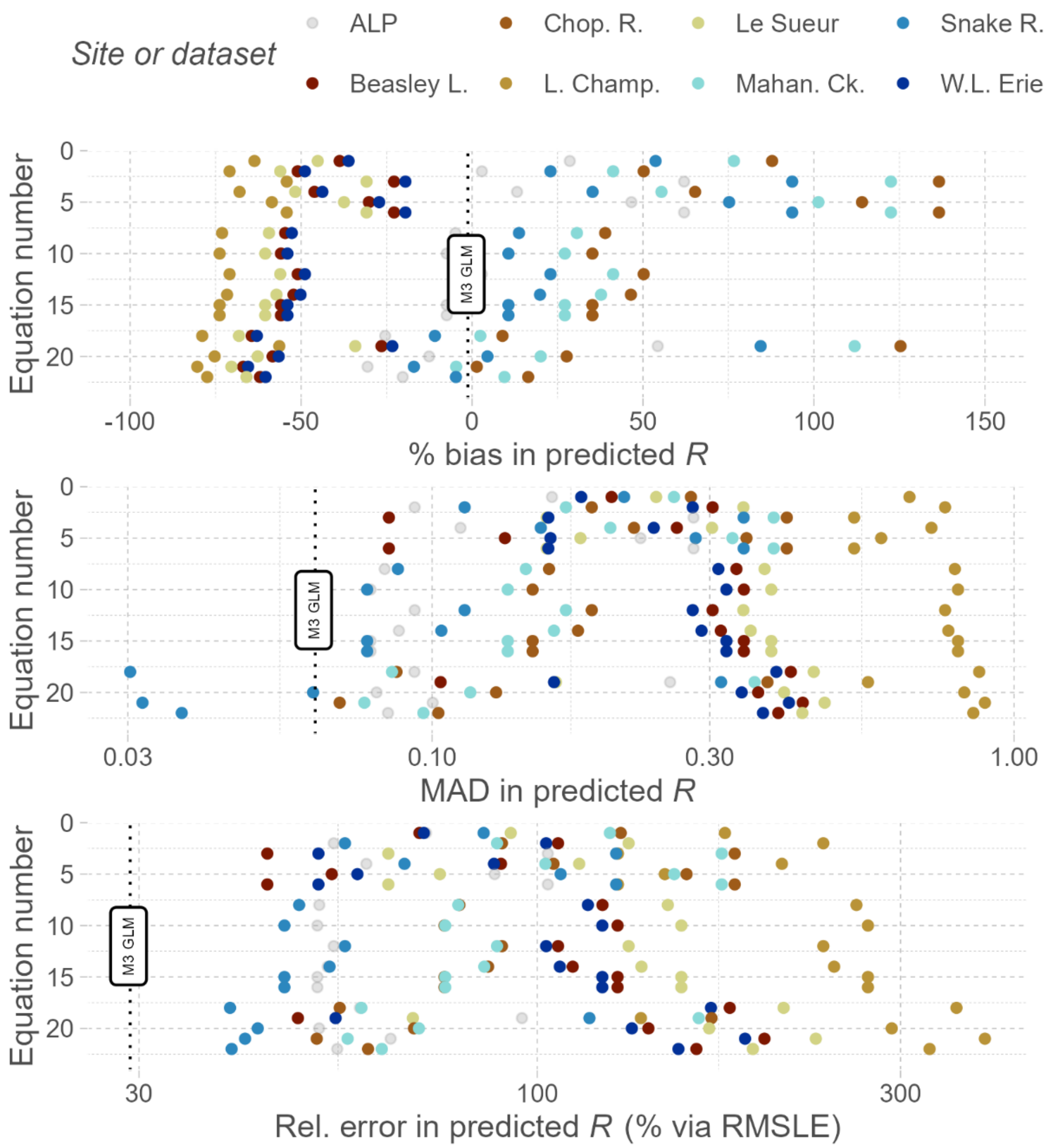

Figure S6. Validation errors for the ratio of Olsen P to Mehlich-3 P (R) via a published regression relationship based only on correlation between Olsen P and Mehlich-3 P. For this, we used the equations reviewed by Steinfurth et al. (2021); the order in their Table 9 is the 'equation number' here. Note that for those equations, the slope coefficient is the ratio used here (ignoring equations with nonzero intercept or nonlinear dependence on Mehlich-3 P). When applied to any subset of the ALP and Legacy P dataset, the potential error is usually substantially worse than, for example, the Mehlich-3 GLM developed in the current study (k-fold cross-validation values given with vertical dotted line). Note that panels for MAD and relative error are shown on log-scale.

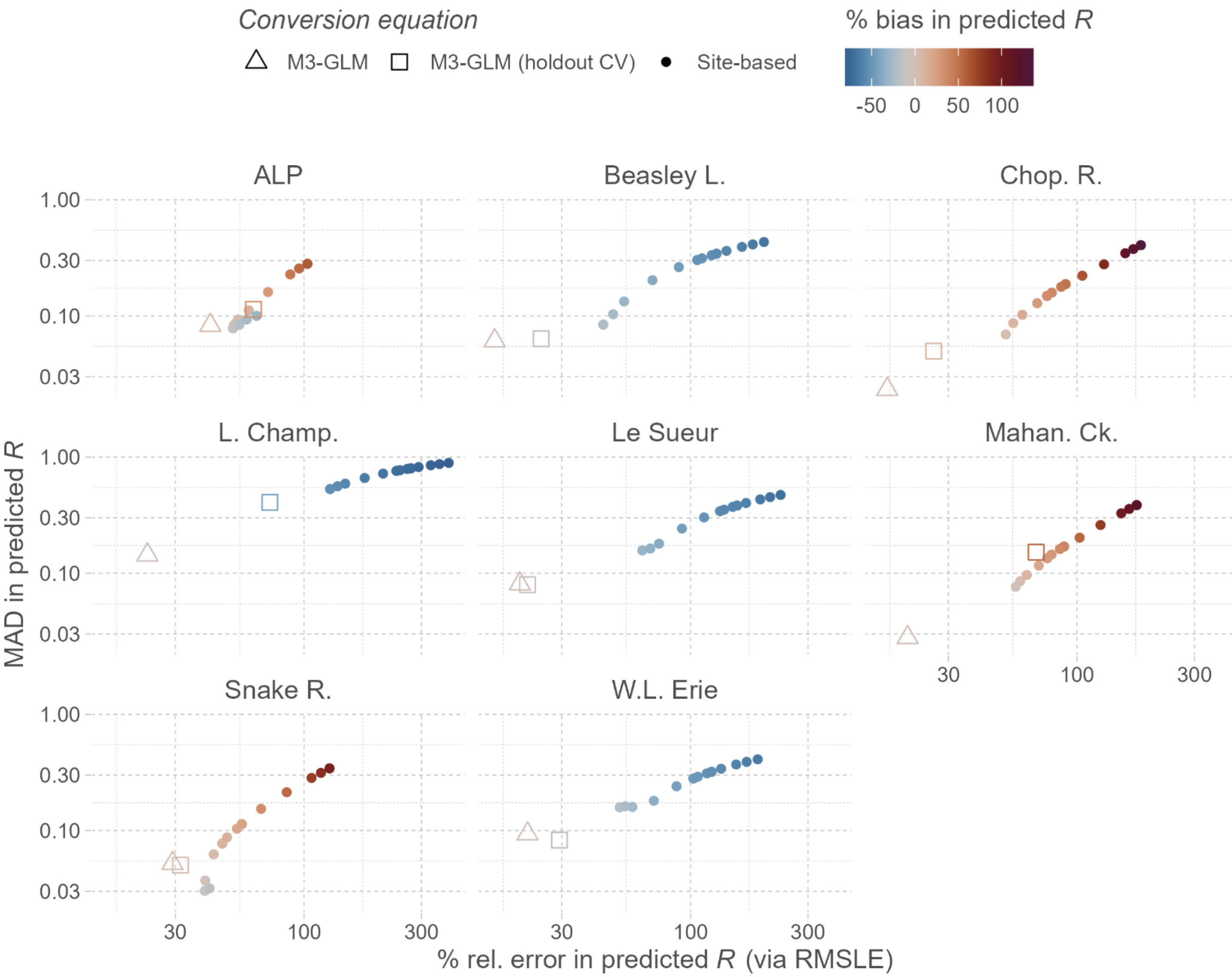

Figure S7. Same as Figure S5 but for the predicted ratio of P_{Ols} to P_{M3} (R); see also Figure S6.

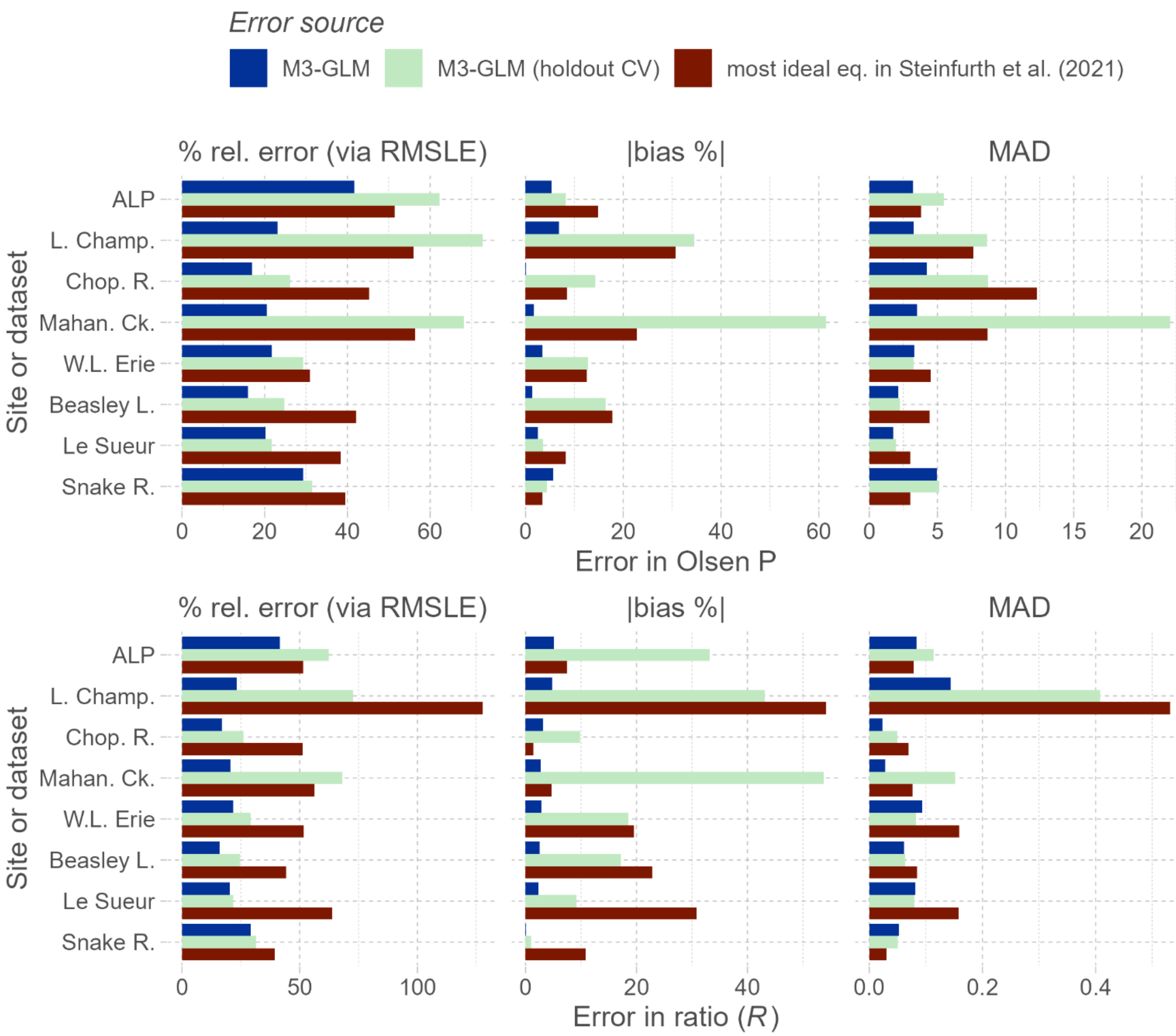

Figure S8. Comparison of validation errors across all sites (Legacy P) or datasets (ALP) in Olsen P concentrations (top) or the ratio of Olsen P to Mehlich-3 P (*R*; bottom) using either (1) the final Mehlich-3 GLM reported in main text, calibrated on all data; (2) the same GLM but calibrated without the subset indicated (as a hold-one-site-out cross-validation); or (3) the 'most ideal' equation in Table 9 of Steinfurth et al. (2021). For the latter, we calculated errors for all given reference equations and kept the best equation *per each sample* (see also Figures S4 and S6) – therefore this should be considered a very optimistic scenario. MAD has units of mg P kg^{-1} for top row and ratio of Olsen P to Mehlich-3 P in bottom row.

References

Steinfurth K, Hirte J, Morel C, Buczko U (2021) Conversion equations between Olsen-P and other methods used to assess plant available soil phosphorus in Europe – A review. Geoderma 401:115339. https://doi.org/10.1016/j.geoderma.2021.115339