https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

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.

Correspondence to: Zachary P. Simpson (zps@ecos.au.dk)

Abstract. Methods for soil test phosphorus (STP) differ globally and even within countries. While soil P tests often correlate, the relationship is often specific to a region or study, making any conversion equation difficult to transfer between contexts. Agronomic recommendations, P transport models, syntheses, and other works lack a strong basis for converting between STP. We fulfil this need for converting between two common STP measurements – Olsen P and Mehlich-3 P – through models which account explicitly for soil properties. Hypothesizing that soil properties at the sample level govern how STP values relate, we built models using a combined dataset of ca. 900 soils across the conterminous US and Canada, spanning 10 soil orders (USDA), 3 to 89% clay, and pH 4 to 9. Model complexity ranged from conventional 'region-level' regressions (no soil data) to models adapting to many facets of soil P chemistry, presenting users several viable options suitable for their data context. Depending on data availability, the user could convert between Olsen P and Mehlich-3 P with half the error or less of conventional 'region-level' regressions. This reduction in conversion error impacts agronomic recommendations, environmental risk assessments (e.g., P index), and calibration of P transport models (e.g., SWAT+). While it remains best to simply measure the STP of interest, the conversion models here should prove useful in many contexts where that is not feasible. We provide the models in ready-to-use formats, depending on the covariates available to the user and whether the user wants to apply the equations in a spreadsheet or within model code.

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

1 Introduction

Testing soils for likely crop response to phosphorus (P) fertilizer is more than a century old (e.g., Dyer, 1894). In that time, a multitude of tests for soil P (STP) have been proposed, ranging from extractions of minutes to hours, acid to alkaline solution pH, buffered vs. unbuffered, presence of chelators, and more. Each STP method may have been developed for a particular region, soil type, or crop type, which was reflected in the design of the operational methods. These contrasting STP methods create persistent difficulties in agronomic recommendations and research, such as in comparing crop fertility responses or in guiding P management strategies across regions. For example, coordinating nutrient management plans is key for several trans-boundary catchments – e.g., the Lake Erie basin (Macrae et al., 2021), the Chesapeake Bay watershed (Kleinman et al., 2019), and Baltic Sea basin (HELCOM, 2021) – but made more difficult since these areas employ three or more different STP methods (Eriksson et al., 2013; Jordan-Meille et al., 2012; Lyons et al., 2023). Additionally, STP values are increasingly used outside the agronomic contexts for which they were developed to model the risk or quantity of P transport from agricultural fields to aquatic environments. These empirical models parameterize P transport as a function of one particular STP, meaning that soils measured using a different STP method are often excluded.

Naturally, researchers and commercial laboratories have derived equations to convert between STPs. One increasingly widespread use of STP conversions is in soil testing. In the US, for example, a common scenario is the use of commercial labs outside the region, where the resulting soil report may give an STP (e.g., Mehlich-3 P; Mehlich, 1984) not conforming to the local fertilizer recommendations, thus necessitating conversion to the appropriate STP (e.g., Olsen P; Olsen et al., 1954). A similar trend seems to apply for other regions, such as in Europe (Mattila and Rajala, 2022). Steinfurth et al. (2021) reviewed many of these conversion equations for eight common STP methods with an emphasis on converting to Olsen P. The conversion equations took the form of a (usually linear) regression of one STP on another and varied considerably between studies. In their review, the authors highlight the limitations of conversion factors in general as well as the importance of matching the conversion equation to the soil in question to avoid serious conversion error. That is, for a study soil, one should avoid applying a conversion equation derived from a region with too dissimilar soils, such as from contrasting soil orders. The consequences of conversion errors are easy to imagine: over- or under-applying P fertilizer to a cropping system, extreme over- or under-estimation of P losses, and syntheses muddied by datasets using different STP.

Each of the region-level regressions reviewed by Steinfurth et al. (2021) and those used elsewhere suffer from similar faults. Often, soil properties at the sample level were not included in the conversions, thus overlooking substantial variance *within* a region. Users from outside regions are left with little option but to guess which region-level equation might suit their soils best. In a synthesis of soil P fertility recommendations across much of Europe, where 10 STP methods are in use, Jordan-Meille et al. (2012) acknowledged that "the choice of these conversion methods affects our results considerably." In a global analysis of STP drawdown as a P sustainability strategy, McDowell et al. (2025) had to convert 36% of observations to Olsen P, using one conversion equation per soil test and likely explaining much of why 'Country' was the most important predictor of drawdown rate. For P transport models, soil P data often serve as inputs only when expressed on a specific basis; e.g., the Soil and Water Assessment Tool (SWAT+) may initialize soil P content if given soil P extracted via anion exchange membrane/resin (the original, but rarely measured, basis) or if given the much more common Mehlich-3 P, assuming that yet another soil P conversion from Mehlich-3 P to anion exchange basis is tolerable (Vadas and White, 2010). A similar model initialization context is in Muntwyler et al. (2023), where available STP data were converted once or even twice to get to the model basis. These authors were all aware of the uncertainty introduced with these coarse STP conversions. Yet, users are largely unaware of the incompatibility in soil P measurements and of model sensitivity to STP conversions. Errors due to converting STP are largely unaddressed and likely hamper several areas of agronomic and environmental research focused on soil P.

In this work, we focus on conversions between Mehlich-3 P (P_{M3}) and Olsen P (P_{Ols}), the two most frequent STP methods used agronomically in the US (Lyons et al., 2023), Canada (typically Olsen, e.g., Grant and Flaten, 2019), several European jurisdictions (Jordan-Meille et al., 2012), New Zealand (Olsen; McDowell et al., 2020) and elsewhere. Improving conversions between P_{M3} and P_{Ols} is timely due to recent interest in agronomic and environmental guidance for STP protocols, emphasis on legacy P in catchments, and continued modelling efforts which rely on the ability to compare STP data on a common basis. To address this need, we utilized a large dataset spanning multiple soil types and P application histories to develop conversions for P_{M3} and P_{Ols} that emphasize generalizable and predictive relationships using readily-available soil properties. We hypothesized that (1) STP conversions differ substantially between regions; (2) STP conversions vary systematically across readily-available soil/sediment properties (e.g., texture, organic matter, pH); and (3) conversions between different STP are largely governed by P lability chemistry.

2 Methods

2.1 Site Descriptions and Sampling in Legacy P Project

Details of the United States Department of Agriculture (USDA) Legacy Phosphorus project (henceforth, ‘Legacy P’), soil and sediment sampling and analyses are in Simpson et al. (2025). Briefly, seven sites (Figure 1) were selected from the USDA Conservation Effects Assessment Project Watersheds Assessment Studies Network, representing wide variation in USDA soil orders (Alfisols, Aridisols, Inceptisols, Mollisols, Ultisols, and Vertisols), climate (mean annual precipitation: 274 – 1450 mm; mean annual temperature: 7.6 – 18.0 °C), cropping systems, fertilizer sources, and P application histories. Soils were sampled in the spring of 2022 and followed a standard protocol across sites. Soils were collected from the 0-5 and 5-15 cm depths while sediments in headwater streams or ditches were collected from the 0-2 to 0-5 cm depth. In total, 622 samples were collected, ranging from 64 at Snake River to 100 at W. Lake Erie (~90 per site). Throughout the text, we simplify to ‘soils’ even though our analyses include sediments as well (see also Simpson et al., 2025).

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

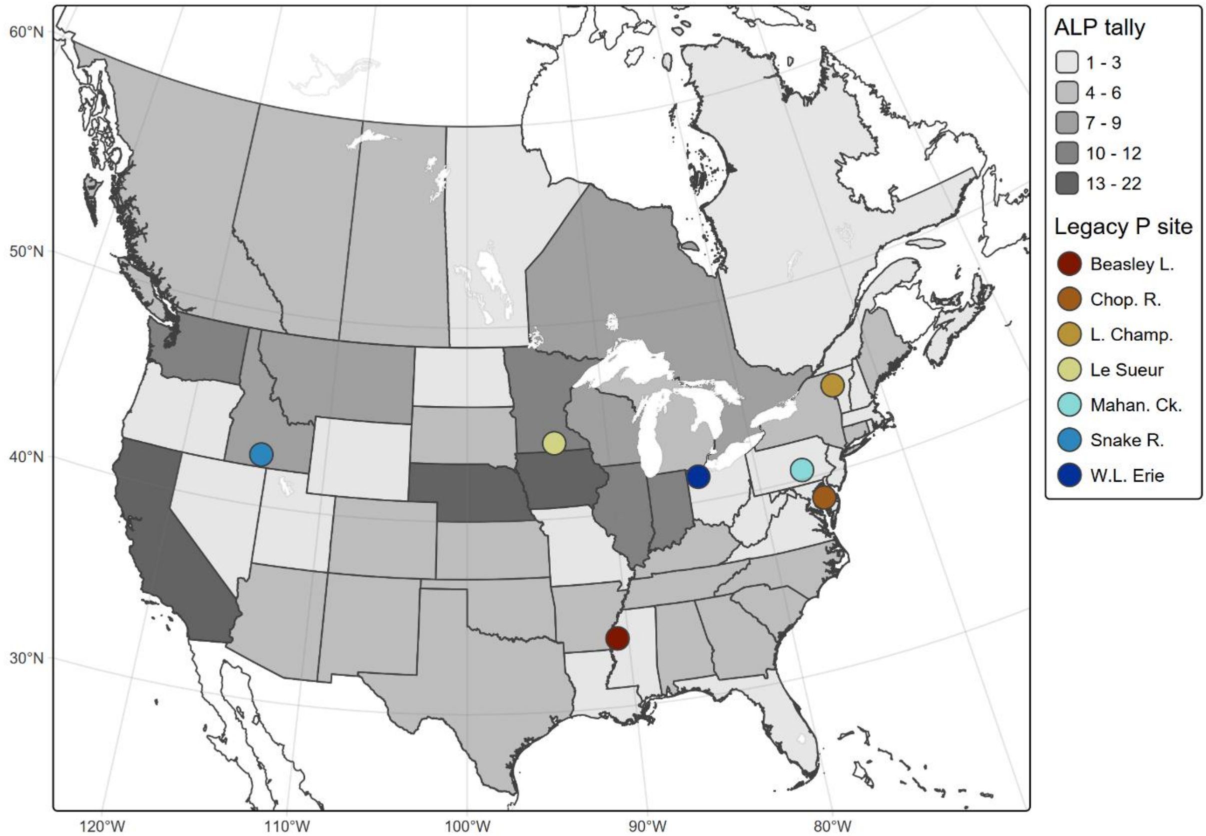


Figure 1: Map of soil sources (n = 897). Soils in the Agriculture Laboratory Proficiency (ALP) program (n = 275) come from the United States, including all 48 conterminous states (excluding Alaska and Hawaii), and from eight provinces in Canada. States/provinces are shaded according to the number of samples in the ALP dataset; blank areas indicate no data. Additionally, soils come from the seven study regions in the United States Department of Agriculture Legacy P project. Points are indicative of the study watershed(s) locations but are not exact. Legacy P sites each had 64 to 100 samples (total n = 622).

2.2 Soil analyses in Legacy P Project

Analyses and corresponding methods are provided in Table 1 while QA/QC and further details are provided in Simpson et al. (2025). Both Mehlich-3 and Olsen extractions were analyzed via inductively-coupled plasma optical emission spectroscopy (ICP-OES) (Kovar and Pierzynski, 2009). As Mehlich-3 emphasizes multiple elements (e.g., cations, trace metals), ICP-OES analysis is nearly universally used by laboratories. More variation exists among laboratories for P_{Ols} analyses, with colorimetry traditionally being prevalent, but ICP-OES has increasingly been adopted (Eriksson et al., 2013; Kovar and Pierzynski, 2009). We note this since ICP-OES measures total P in solution compared to colorimetry which only measures reactive P (Shwiekh et al., 2013). Some previous work suggests negligible to minor (ca. 10%) differences between the two analyses for Olsen extracts (Adesanwo et al., 2013; Wanke et al., 2023). For the Legacy P soils, we found in a pilot analysis of the two analyses

(32 samples and 3 reference soils) a median absolute deviation between the two P_{Ols} of 4.0 mg kg^{-1} (cf. standard deviation in P_{Ols} via ICP-OES of 21.0 mg kg^{-1}) and little consistent evidence for the bias from potential molybdate-unreactive P. Therefore, we consider the P_{Ols} values in the Legacy P dataset as approximately equivalent to those measured via colorimetric analysis; in line with common laboratory practice, we do not utilize other Olsen-extractable elements measured via ICP-OES.

Table 1: Select characteristics of the samples in the USDA Legacy P dataset (see also Simpson et al., 2025) and the Agriculture Laboratory Proficiency (ALP) dataset. For ALP, n = 275 soils are summarized for the available properties. For the USDA Legacy P, both soils (n = 560) and sediments (n = 62) are combined; additional P sorption and related variables are reported (not available for ALP). For EPC_0, labile P, S_{max}, and Bache-Williams index, 30 samples from the Snake River site were missing from analysis. SD is standard deviation.

		USDA Legacy P		**ALP**	
Variable	Unit	Median (Mean ± SD)	Min-Max	Median (Mean ± SD)	Min-Max
Olsen P	mg kg^{-1}	30.0 (33.1 ± 20.4)	1.56 - 136	16.5 (22.3 ± 20.3)	2.10 – 134
Mehlich-3 P	mg kg^{-1}	49.8 (92.4 ± 99.8)	1.31 – 628	47.2 (71.2 ± 89.1)	1.25 – 737
pH	S.U.	6.78 (6.85 ± 0.823)	4.83 – 8.73	6.41 (6.49 ± 1.05)	4.05 – 9.02
Clay	g kg^{-1}	345 (345 ± 158)	32.2 – 890	177 (192 ± 105)	30.7 – 599
Sand	g kg^{-1}	235 (273 ± 152)	0 – 917	447 (464 ± 214)	95 – 942
TOC	g kg^{-1}	14.7 (18.5 ± 12.2)	2.13 – 95.7	15.2 (19.4 ± 21.4)	2.47 – 280
TIC	g kg^{-1}	0 (3.30 ± 6.35)	0 – 41.6	0.0998 (1.92 ± 5.98)	0 – 66.7
DPS_{ox}	mol mol^{-1}	0.133 (0.183 ± 0.130)	0.031 – 0.667		
DPS_{M3}	mol mol^{-1}	0.0688 (0.276 ± 0.715)	0.00288 – 6.74		
Total P	mg kg^{-1}	622 (658 ± 280)	142 – 1700		
EPC_0	mg L^{-1}	0.113 (0.449 ± 1.09)	0.0057 – 15.6		
24 h P desorption via AEM (Labile P)	mg kg^{-1}	59.4 (71.6 ± 55.6)	1.20 – 325		
Langmuir S_{max}	mg kg^{-1}	331 (395 ± 245)	111 – 2320		
Bache-Williams Index	(mg P 100 g^{-1}) per log(µmol P L^{-1})	3.99 (4.87 ± 3.42)	1.35 – 38.8		

2.3 ALP Soil Dataset and Analyses

To expand the range of soil characteristics covered in this analysis, we also included soils (n = 275) analyzed in the Agriculture Laboratory Proficiency (ALP) Program (see Figure 1; https://collaborative-testing.com/program-1.php). Briefly, the ALP program collects agricultural soils across North America; here, 10 USDA soil orders were represented, with Alfisols, Entisols, Inceptisols, Mollisols, and Ultisols having >20 samples, while fewer samples came from Andisols, Aridisols, Histosols, Spodosols, and Vertisols. Soils were pulverized to pass < 0.8 mm sieve, homogenized, and shipped to >120 soil testing laboratories for analysis of 122 physical and chemical properties. We use the consensus median proficiency values for each soil property. Most of the soil physical and chemical properties covered in the Legacy P dataset were also reported in the ALP dataset; oxalate extractions and lability-related variables are notable exceptions (Table 1). Methods of analysis for soils in the ALP dataset largely reflect methods of analysis for soils in the Legacy P dataset except for Olsen P, which was measured by colorimetry (Miller et al., 2013).

2.4 Statistical Analyses

2.4.1 Region-level bivariate slope via Passing-Bablok regression

Data was analyzed in multiple steps to establish relationships between P_{M3} and P_{Ols} and improve conversion between the two. Following the approaches in previous literature, initial tests considered the bivariate relationship between P_{M3} and P_{Ols} (i.e., slope) at the region/dataset level as well as for the entire combined dataset. Since neither STP measurement is 'dependent' on the other, it is arbitrary to regress one on the other; indeed, slopes can change somewhat depending on the orientation. Instead, we used equivariant Passing-Bablok regression (Dufey, 2020) to find the line that minimizes error for both variables via the 'mcr' package (Potapov et al., 2024). The resulting slope is insensitive to differences in scale, accounts for errors in both measurements, and is robust to outliers; we present the resulting line as passing through the origin. In preliminary analyses, we found that fits were consistent for each site regardless of including sediments or not; thus, we present all sample data from the Legacy P dataset.

2.4.2 Models using the ratio of Olsen P to Mehlich-3 P

The conversion slope from the above method provides one value (slope) for a given dataset, e.g., *a study region*. Few studies have attempted to incorporate additional *sample-level* soil characteristics in such conversion models. For example, Shwiekh et al. (2015) presented several regressions for multiple STPs as a function of other STP, total C, pH, and other properties. However, the regressors there represented the effect on the *dependent* STP only, holding all else equal; in reality, a difference in a soil property, all else equal, likely influences multiple STP simultaneously but to varying degrees. Further, reversing the regression equation to estimate the original predictor STP is awkward as the original coefficients were minimizing the variance in the original dependent STP, not the predictor STP.

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

Because the conversion slope is effectively a ratio between the two STP values, we developed predictive models of this ratio also at the individual *sample level* via generalized additive models (GAMs; Wood, 2017), utilizing additional covariates for the relationship:

$$R_i \sim \text{Gamma}(\mu_i, \phi), \tag{1}$$

$$\log(\mu_i) = \beta_0 + \sum_j f_j(x_{j,i}), \tag{2}$$

where R_i is the observed ratio of P_{Ols} to P_{M3} for the ith sample, considered here to be Gamma-distributed since the ratio is a positive value, ratios are naturally left-skewed, and the variance is often scale-dependent (i.e., increases with larger magnitudes of STP); the conditional mean of this Gamma distribution is μ_i, whose log is a function of an intercept (β_0) and the sum of j potentially nonlinear functions (f_j) of covariates $x_{j,i}$ (which may be paired to capture interactive effects); and φ is the dispersion parameter. The f_j are assumed to be smooth and are penalized towards linear functions via cubic regression smoothing splines with shrinkage; interactions between multiple covariates (e.g., both Fe_{M3} and Al_{M3}) are represented with tensor product smooths also with shrinkage bases. We used the 'mgcv' package (Wood, 2017) to fit GAMs of R via restricted maximum likelihood. Note that, unlike models of one specific STP, models of R may be easily compared across contrasting sets of predictors, e.g., when using P_{Ols} vs. P_{M3} as a predictor.

We built models of R in five manners, varying in utility (exploratory to applied) and in which predictors ($x_{j,i}$) were used. The first set is 'region-level', while the remaining four are 'sample-level' by utilizing data specific to the sample:

1. Region-level intercepts – either an intercept for the entire combined dataset ('overall') or an intercept per each study region. These are akin to the Passing-Bablok slopes where each study region/dataset yields one conversion slope and thus provide reference models to compare with sample-level models.
2. Sample-level: Basic predictors – clay concentration, pH, TOC, and TIC, but no P data, were used. Study region was included but only as a varying intercept to account for the non-independence within study regions.
3. Sample-level: Basic + STP predictors – Either P_{Ols} or P_{M3} extraction data were available, but not both. However, there were some distinctions for either soil test.
    a. Mehlich-3 – models expanded to include Al_{M3}, Ca_{M3}, Fe_{M3}, Mg_{M3}, and Mn_{M3} as predictors. Interactions between Al_{M3}/Fe_{M3} and Ca_{M3}/Mg_{M3} were considered.
    b. Olsen – only P_{Ols} was considered for models available to potential users. However, we also tested the Bache-Williams index (BWI; Legacy P data only; Bache and Williams, 1971) as a potential readily-measurable property describing the P buffer capacity.

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

4. Sample-level: Lability models – these models leveraged much of the information in the Legacy P dataset that otherwise is not commonly measured or readily available, including predictors such as: EPC_0, labile P, oxalate extraction, and bicarbonate-dithionite extraction. Emphasis was given to the central role of P lability (the quantity-intensity-capacity relationship for P on soil surfaces) in STP measurements (Holford, 1980; Olsen et al., 1954; Simpson et al., 2025). STP variables were excluded as we tested whether P lability was sufficient to translate between P_{M3} and P_{Ols}.
5. Sample-level: Generalized Linear Models (GLMs) – using the same predictors as for the models in (3) but with the added constraint that all predictor effects (f_j) must be linear on the link scale (constraining the GAM to be a GLM). Models were still fit by the 'mgcv' package via restricted maximum likelihood. Interaction smooths were replaced with ordinary (linear) interaction terms. The rationale is that while (3) could be widely used, GLMs are more easily understood and readily implemented in, e.g., transport model code. We focused the most performative models in (3) to reconsider as GLMs.

For each model, we used visual diagnostics (e.g., deviance residual plots) to check model assumptions, aided by the 'gratia' package (Simpson, 2025). We compared models in their predictive capacities via Akaike Information Criterion (AIC; Burnham and Anderson, 2004), percent bias, root mean square log error (RMSLE), and median absolute deviation (MAD); the latter three statistics were computed with 10-fold cross-validation. As a severe test when comparing a chosen model to reference equations in Steinfurth et al. (2021), we used leave-one-site-out cross-validation, where either one Legacy P site or the ALP data were omitted when refitting the model and the fitted model predicted the left-out sample.
Associated R code and a spreadsheet version of our final conversion equations are available at Figshare (https://doi.org/10.6084/m9.figshare.32133124).

3 Results

3.1 Soil Physical and Chemical Properties

Soils in the study varied greatly in their physical and chemical properties (Table 1). Notably, Legacy P and ALP covered similar extents of soil P, pH, TOC, and TIC, but ALP included more samples with high sand content than did Legacy P (medians: 447 vs. 235 g kg^{-1}) while the opposite was true for clay (177 vs. 345 g kg^{-1}). Across all data, sample pH ranged from acidic (4.05) to highly alkaline (9.02); most soils had 10 to 25 g kg^{-1} TOC, but did vary from 2.1 to 280 g kg^{-1}; P_{M3} ranged

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

from 1.25 to 737 mg kg^{-1} and P_{Ols} from 1.56 to 136 g kg^{-1}, with medians of 49.4 and 26.2 mg kg^{-1}, respectively. Observed *R* was highly variable, having a median of 0.441, inter-quartile range of 0.278 to 0.720, and overall range of 0.096 to 3.68.

3.2 Bivariate relationships between Mehlich-3 P and Olsen P

Simple bivariate relationships between P_{M3} and P_{Ols} (Figure 2) depended greatly on the study region (Legacy P) or dataset (Legacy P, ALP, or both). Using all data, the overall slope was 0.277 (i.e., suggesting 0.277 units of P_{Ols} for every unit of P_{M3}). At the region level, Passing-Bablok regression slopes ranged greatly from 0.193 (Mahan. Ck.) to 0.754 (Beasley L.) – clearly, a conversion slope from one region does not port to another. Some regions grouped more closely together including the alkaline soils (Aridisols) from Snake R. (0.318) to the neutral samples (Mollisols and Inceptisols) of the Le Suer watershed (0.619). Notably, conversion slopes at the region level varied beyond the extents of the previously reported minimum (0.27) and maximum (0.63) in the literature (Steinfurth et al., 2021).

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

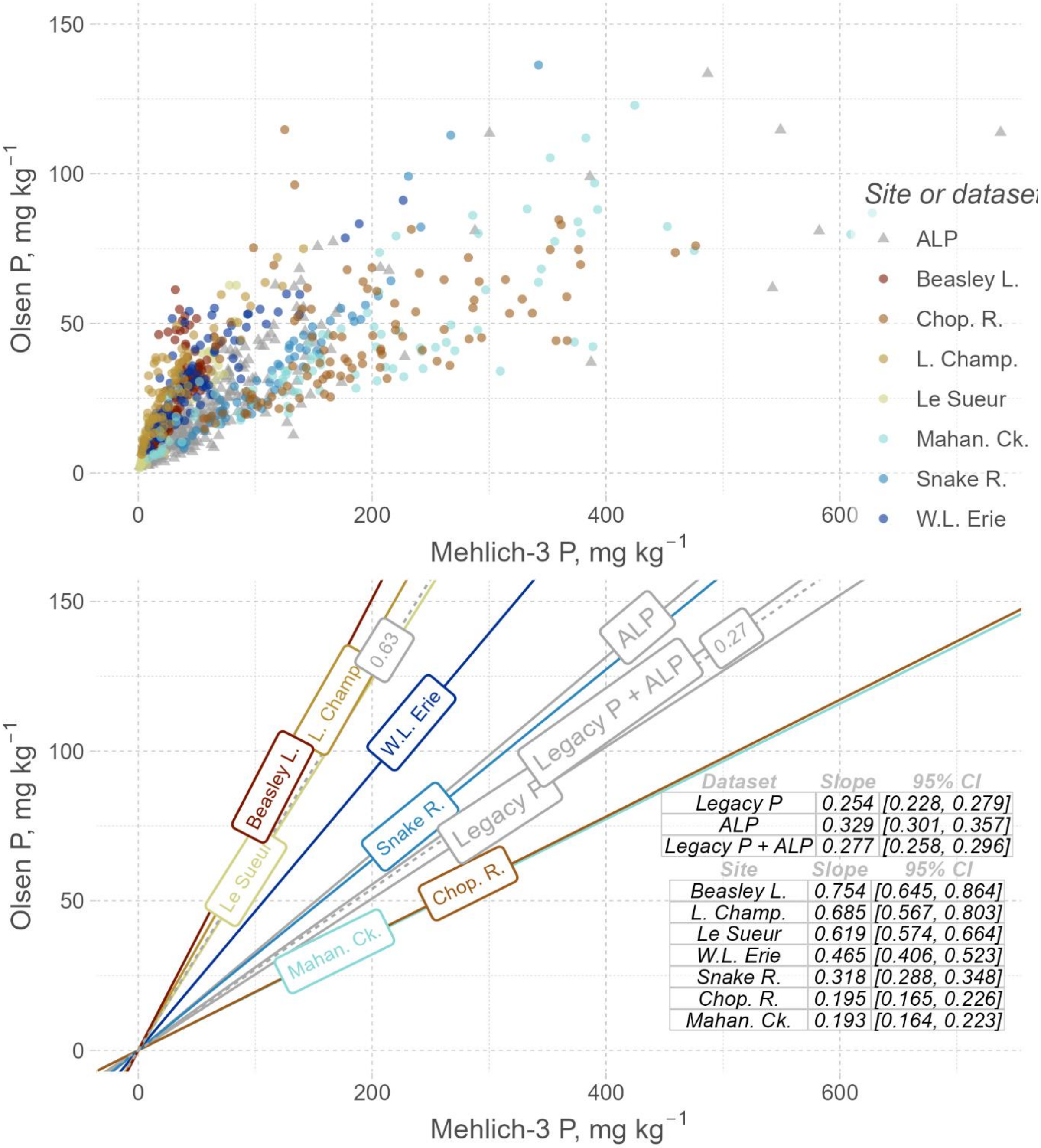


Figure 2: Relationship between Olsen P and Mehlich-3 P across combined ALP (gray) and Legacy P (colored) datasets. (Top) individual data at the site/dataset level are summarized to a conversion slope (bottom) via Passing-Bablok regression (inset table gives the conversion slope and its 95% confidence interval). For reference, the previously reported maximum (0.63) and minimum (0.27) conversions found in the review by Steinfurth et al. (2021) are shown in bottom panel.

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

3.3 Predictive Models for the Olsen to Mehlich-3 Conversion Ratio

As an alternative to region-level conversion slopes, we fit models of the ratio of P_{Ols} to P_{M3} (R) according to five categories of predictor sets as detailed in the Methods (Table S1 for full results). The best models within each category differed markedly in their performance (Figure 3). The region-level intercepts model captured similar information as region-specific bivariate slopes, where model intercepts of R for regions resembled the respective slopes (notable exceptions were L. Champ., W.L. Erie, and Le Sueur, whose predicted R were larger than their respective bivariate slope). Aside from the overall intercept model, the region-level intercepts model was the worst-performing model, where average errors (via RMSLE) in R were ±52%. The region-based model has limited utility outside the seven study sites in Legacy P but serves as a useful model benchmark. Modeling R instead with sample-specific predictors, the remaining four categories of models showed large gains in performance (e.g., AIC lower by 200 to 1300 compared to the region-based model; see also Figure S3). We expect most, if not all, potential users will have observations of either P_{M3} or P_{Ols} depending on which is common in their region; this is useful considering that the magnitude of soil P was consistently the largest predictor of R and, conversely, the model with only basic predictors (TOC, TIC, pH, and clay) still had considerable error (47%). The best-performing model was for Mehlich-3 (Figure S1), which explained 91% of the deviance and had average error of 26%. This model had its strongest predictor in P_{M3}, but by utilizing texture, organic matter, pH, carbonates, and Mehlich-3 extractable metals, the model was generalizable: prediction error on the R scale (MAD) was only 0.057 – half that of the region-level intercepts model. The Olsen model, while only leveraging P_{Ols}, was still effective in predicting R. The Olsen model covered 84% deviance and improved average errors (44%) over the region-level models.

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

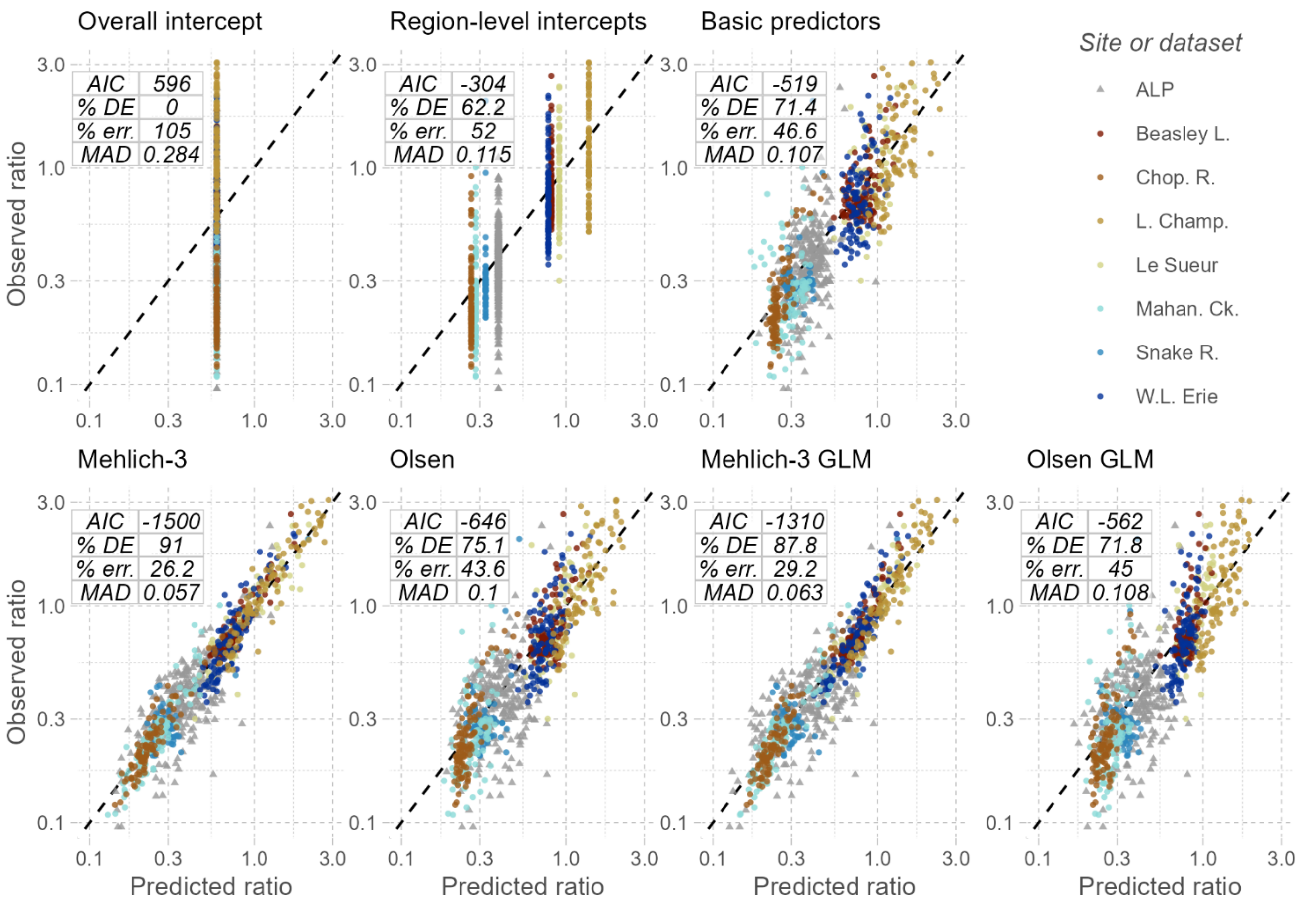


Figure 3: Plots of observed vs. predicted conversion ratios (Olsen P to Mehlich-3 P) for 7 different predictive models. Table S1 gives more detail for each model; all were fit as GAMs, but only 'basic predictors', 'Mehlich-3', and 'Olsen' models included smooth nonlinear terms. The GLMs used the same predictors as the respective GAMs. The inset table gives model performance statistics: Akaike Information Criterion (AIC), percent deviance explained (% DE), percent relative error via root mean square log error (% err.), and median absolute deviance (MAD; on ratio scale). The latter two error statistics were via 10-fold cross-validation. Percent bias is not shown but was negligible (at most, +/- 1.2%). The dashed line is the 1:1. Note use of log scales.

Observing P_{M3} data, rather than P_{Ols}, gives a natural advantage in predicting *R*. Much more of the variance in *R* is due to the variance in P_{M3}: coefficients of variation across the study sites varied as high as 0.97 for P_{M3} while for P_{Ols} the greatest value was 0.69. Additionally, while Olsen extracts are typically only analyzed for P, the multiple Mehlich-3 extractable elements added useful predictors of *R*: the best Mehlich-3 model leveraged Fe_{M3}, Al_{M3}, Ca_{M3}, Mg_{M3}, and Mn_{M3}. As these predictors may speak to soil P buffer capacity, we considered an Olsen model including the BWI as a predictor (Table S1; Legacy P data only): MAD (0.076) and average error (30%) all improved considerably, closing some of the gap to the Mehlich-3 models.

To investigate soil P chemical mechanisms behind STP conversions, we considered 'lability' models (Legacy P data only). Using only basic soil properties, EPC_0, P_{AEM}, BWI, and oxalate-extractable P, Fe, and Al, the best lability model performed as well as the best Mehlich-3 model: 92% deviance explained and 24% average error (Figure S2). The predictors for P intensity

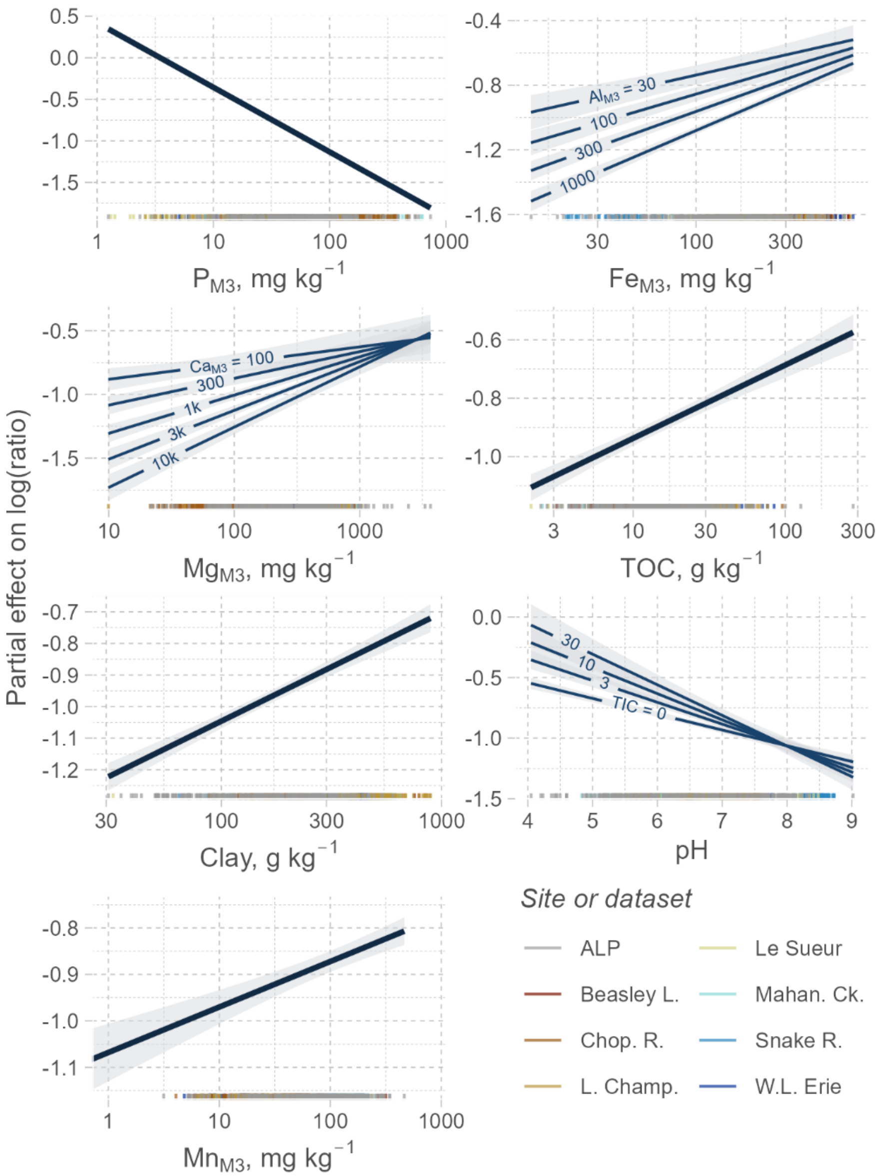
https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

(EPC_0) highlighted in Simpson et al. (2025) also provided a good basis for predicting *R*, while the addition of oxalate-extractable P provided further benefit. In this lability model, effects for P_{AEM} and P_{ox} behaved similarly to those for P_{M3} or P_{Ols} in sample-level models with basic predictors and STP, with steep declines in *R* with greater P concentration. Oxalate-extractable Al and Fe were less sensitive as predictors for *R* compared to their Mehlich-3 counterparts likely due to the inclusion of BWI as the P buffer capacity variable; each of these variables indicated positive effects on *R*.

3.4 Conversion Equations and Comparison to Literature

Considering the sample-level Olsen or Mehlich-3 models, it is convenient to convert from a GAM to a GLM (nonlinear smooths requiring more effort to incorporate elsewhere), i.e., constraining the predictors to be linear on the log scale, though there will likely be performance losses for either soil test due to less model flexibility (Fig. 3). Still, the GLM versions of the respective best GAM fits were both comparable to their originals. For the Mehlich-3 GLM (Figure 4), average error increased to 29% (from 26%) while for Olsen it increased to 45% (from 44%). The cross-validation MAD for both GLMs (0.063 and 0.108, respectively) suggested that typical errors were not much larger than for the respective GAMs. The model parameters for either soil test are given in Table 2; these may be readily incorporated by users wishing to convert between either STP, and we give readily-adaptable R code and a spreadsheet for this purpose.

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026



270 **Figure 4: The partial effects for the GLM version of the best-performing Mehlich-3 GAM of the ratio of P_{Ols} to P_{M3}, *R*. (All terms from the GAM were kept but constrained to be linear on the log-scale.) 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. The colored rug plot along the bottom indicates the marginal distribution of the *x*-axis predictor across the sites (Legacy P) or**
275 **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).**

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

Table 2: Parameter estimates and their 95% confidence intervals for the generalized linear models of the ratio of Olsen P to Mehlich-3 P (R), assuming that either (left) Mehlich-3 data or (right) Olsen P data are available. Each parameter is for the log(mean), therefore predictions should be the exponentiation of the sum of all terms. I.e., predicted ratio = exp(1.71 + 0.275 × log(clay) + … + 0.0425 × log(Mn_{M3})). With this ratio, one may then convert between P_{M3} or P_{Ols} with e.g. $P_{Ols} = P_{M3} \times R$. Predictor units are g kg^{-1} for clay, TOC, and TIC; S.U. for pH; and mg kg^{-1} for all Mehlich-3 extractable elements and Olsen P. Natural log is used except for TIC, where 'log1p' means log(1 + TIC). Interaction predictors are simply the product of the two covariates. Note that all terms from the best performing GAM were kept here though some coefficients are small and/or uncertain; some terms are small/uncertain but are important in interaction terms (e.g., Mg_{M3}).

Predictor	*Mehlich-3 GLM*		*Olsen GLM*	
	Coefficient	95% C.I.	Coefficient	95% C.I.
Intercept	1.71	0.64, 2.78	0.021	-0.594, 0.635
log(clay)	0.149	0.104, 0.194	0.185	0.123, 0.247
log(TOC)	0.109	0.071, 0.147	0.148	0.093, 0.203
pH	-0.130	-0.166, -0.094	-0.207	-0.254, -0.160
log1p(TIC)	0.284	0.064, 0.505	-0.278	-0.614, 0.058
pH × log1p(TIC)	-0.0356	-0.0659, -0.0053	0.0455	0.0011, 0.0900
log(P_{M3})	-0.338	-0.359, -0.316		
log(P_{Ols})			-0.228	-0.272, -0.184
log(Fe_{M3})	0.0142	-0.149, 0.177		
log(Al_{M3})	-0.233	-0.362, -0.103		
log(Fe_{M3}) × log(Al_{M3})	0.0293	0.0023, 0.0563		
log(Ca_{M3})	-0.259	-0.355, -0.164		
log(Mg_{M3})	-0.0950	-0.264, 0.0735		
log(Ca_{M3}) × log(Mg_{M3})	0.0326	0.0109, 0.0544		
log(Mn_{M3})	0.0425	0.0181, 0.0669		

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

What gain in performance can a user expect with the equations developed here? To answer this question, we compared the Mehlich-3 GLM in Table 2 to the equations reviewed by Steinfurth et al. (2021), which relied on the conventional region-level correlation between P_{Ols} and P_{M3}. Further, we used leave-one-site-out cross-validation for the Mehlich-3 GLM as a severe test to compare against the literature equations. Depending on the reference equation used, errors in either P_{Ols} values (Figures S4 and S5) or in the ratio *R* (Figures S6 and S7) were severe. For region-level literature equations, relative errors often exceeded 100% while that for the Mehlich-3 GLM was ca. 30%. For the ALP, L. Champ., and Mahan. Ck. subsets, a few region-level literature equations had comparable performance compared to the Mehlich-3 GLM when holding out that subset's data; in all cases, however, the final calibrated Mehlich-3 GLM was the most predictive. If one chose the *best* region-level literature equation for their individual soil sample, P_{Ols} predictions would still substantially improve if using the Mehlich-3 GLM: average relative errors typically decreased by ~50% while absolute bias decreased by 75% (Figure S8).

4 Discussion

4.1 Toward broadly applicable soil test P conversions

Translating between different STP values is not straightforward, but is frequently required to integrate datasets generated with differing STP methods. Examples include multi-region or multi-study syntheses or meta-analyses (McDowell et al., 2025; Muntwyler et al., 2023), agronomic comparisons across regions (e.g., scrutinizing and harmonizing fertility recommendations across borders) (Jordan-Meille et al., 2012; Lyons et al., 2021; Steinfurth et al., 2022), and for P-transport modeling (Muntwyler et al., 2024; Vadas and White, 2010). This is often the case in any such study at regional, national, or global scales. Here, we demonstrated a relatively simple method for converting between Olsen and Mehlich-3 P values which, in comparison to the conventional use of a region-level regression, cuts errors in half. The performance gain of the introduced method ('sample-level' modeling) over the conventional method ('region-level' modeling) is largely due to incorporating key soil predictors varying at the sample level, as suggested by Steinfurth et al. (2021). In contrast, region-level models for the two STP values, though a good first step, are inefficient for two connected reasons: (1) often, they must average over considerable variation in soil properties (e.g., texture, pH) even at the site level; and (2) structurally, these models improve little from more observations. This latter point is important as often considerable work goes into sampling and analyzing soils across a context only to produce a single regression of one STP to another, such as for the ~2700 samples in Culman et al. (2020). Such datasets could be reanalyzed with sample-level models, like those developed here, to substantially improve conversion relationships.
The use of region-level models can lead to severe errors – at best around 50%, but often >100% – whether using the most ideal reference equation (Figure S8) or even data from the site itself (Table S1). In contrast, the GLM provided in Table 2 leveraged soil clay concentration, organic and inorganic carbon, pH, and Mehlich-3 extractable elements to yield relative errors ~25%. While we believe there to be room for improvement, particularly for the more difficult conversion from P_{Ols} to P_{M3}, the models we offer should apply across many soil orders and P exposures.

The cross-validation error metrics reported here are, for simplicity, a summary statistic, e.g. MAD (Table S1). These metrics deal with *average* out-of-sample prediction error which may only show small gains for a better model (Gelman et al., 2014); the full distribution of prediction errors are also important when comparing these models, especially towards the extremes in soil P concentrations. For instance, region-level regressions might have acceptable bias *overall*, but they suffer markedly when predicting either P_{Ols} above 30 mg kg^{-1} or P_{M3} above 50 mg kg^{-1} (Figure S3), with extreme negative and positive biases, respectively. In contrast, a major benefit of the sample-level STP ratio models developed here is that bias remains near zero across a wide range in STP and, further, prediction error distributions are more constrained and avoid extreme errors. This is also why RMSLE, which is sensitive to extreme errors, decreased dramatically for the GLMs compared to the reference region-level equations while MAD did not (Figure S5). In short, sample-level models are more reliable in general but especially so when dealing with more extreme STP contexts.

Another benefit of modeling STP conversions as a ratio at a sample- rather than region-level is that it captures the wide variation in the relationships between soil tests with readily-available soil properties, obviating the need to match equations based on the study region or soil type. We were surprised that region-level slopes of P_{Ols} to P_{M3} across Legacy P sites in the US (Figure 2) varied even more than previously reported in the literature (Steinfurth et al., 2021). Applying region-level conversions from literature to sites like Beasley L. or Mahan. Ck. would be at best heavily biased, but at worst could give errors exceeding 100% (Figures S4; S6). A region-level slope is already an inefficient means for converting between STPs, but it also generalizes poorly to settings where data on contrasting soil P metrics are unavailable. Modeling the ratio, *R*, between other soil P metrics or pools may prove an effective strategy in more contexts where soil P data are mismatched.

4.2 Chemical basis for conversion equations

The two strongest sets of predictors for *R* were those pertaining to the soil P quantity and to the soil's P buffer capacity, though predictors related to chelatable metal oxides and other mechanisms were also important. In our models, *R* decreased with greater soil P quantity (P_{M3}, P_{Ols}, or labile P) and increased with greater P buffer capacity (e.g., clay concentration, Bache-Williams index). These relationships speak to the contrasting chemistry of the two soil tests, indicating opportunities to further improve conversions but also limitations of soil test conversions.

Since most STP methods were developed to extract some proportion of so-called 'plant available' P (Bray, 1948; Sims et al., 2000), they were often based on the various chemical mechanisms crops employ (or benefit from) to acquire P (Mehlich, 1984; Olsen et al., 1954; Truog, 1930), such as the release of protons, organic anions (especially carboxylates), and other ligands or chelating agents (Lambers, 2022; Nguyen et al., 2024; White and Hammond, 2008). As protonation (via strong acids and/or low pH buffers) is very effective at mobilizing much of the inorganic P in mineral soils, tests with greater acidity tend to extract greater quantities of P (Mattila and Rajala, 2022; Messiga et al., 2014; Wuenscher et al., 2015). Therefore, typically $R < 1$ since $P_{M3} > P_{Ols}$ (cf. 13% of soils in our study, usually with <30 mg P_{Ols} kg^{-1} and/or >40% clay, had $R > 1$). This disparity between P_{M3} and P_{Ols} grows with increased soil P. Therefore, one of the strongest advantages of the sample-level models of *R* here is that predictions change with magnitude of soil P quantity, unlike for the region-level regressions that assume a static *R*

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

through a fixed slope term. Similar relationships might apply to conversions between Olsen and acidic extractions, e.g., Bray-Kurtz, ammonium-lactate, or oxalate.

While Olsen et al. (1954) and Mehlich (1984) were seemingly less concerned with (or aware of) the P lability dynamic in soils (Beckett and White, 1964; Holford, 1997), much of the P extracted in their tests *is* dictated by the lability of soil phosphate, where greater soil P intensity entails greater P concentration extracted into either solution (Holford, 1980; Koopmans et al., 2002; McDowell et al., 2001; Pierzynski et al., 2005). Some authors refer to certain soil tests as more related to either 'quantity' or 'intensity'; we posit that it is more accurate to say STP is mostly a blend of all three aspects of soil P lability (Simpson et al., 2025), even though STP in some conditions appears to relate more strongly with either soil P intensity or quantity. The different relationships between soil P lability and either P_{Ols} or P_{M3} are clearly displayed in the 'lability' GAM in Figure S2. Using data independent from the soil tests, *R* decreased nonlinearly with labile P (quantity) yet increased with Bache-Williams index (buffer capacity); we further note that all the information that encoded EPC_0 (intensity) across the Legacy P dataset (see Fig. 5 in Simpson et al., 2025) is also present in this GAM, hence EPC_0 was not needed as a predictor. The decrease in *R* with greater soil P quantity reflects how P_{M3} is more sensitive to P loading than is P_{Ols}. The positive effect for buffer capacity, *conditional on P quantity*, deserves more discussion.

Positive effects for P buffer capacity parameters on *R* were evident for the GLMs (Table 2; Fig. 4), notwithstanding interaction effects, through the contributions of clay, Ca_{M3}, Mg_{M3}, Al_{M3}, Fe_{M3}, and Mn_{M3} concentrations (the latter three likely indicating greater concentrations of metal (hydr)oxides). Similarly for the 'lability' GAM (Fig. S2), in addition to Bache-Williams index, more clay and oxalate-extractable Fe + Al predicted greater *R*. This relationship is not immediately clear and requires consideration of both extractions in detail. All aspects of P lability are in reference to conditions that attempt to maintain the chemical integrity of the soil surface (Larsen, 1967), such as by using an appropriate ionic background. In contrast, Olsen (0.5 M $NaHCO_3$, pH 8.5) and Mehlich-3 (containing high concentrations of acetic and nitric acids, NH_4NO_3, NH_4F, and EDTA; pH 2.5) are more aggressive and mobilize much more soil P into solution on short timescales, particularly inorganic fractions, through several mechanisms (Barrow and Shaw, 1976; Cade-Menun et al., 2018; Mendez et al., 2020). Independent of soil P quantity, greater soil P buffer capacity implies greater amounts of inorganic P stored in less labile forms or positions (van Doorn et al., 2024; Holford, 1997); indeed, this idea underlies the three-pool model for soil inorganic P (Jones et al., 1984; Sharpley et al., 1984) in ecohydrological models such as SWAT+ and many others (Pferdmenges et al., 2020; Radcliffe et al., 2015). Both STP extracting solutions are effectively dissolving more of the continuum of inorganic P than just the labile pool, but the critical difference is that this reaction is *kinetically constrained*. On the sub-hour timescale, when the amount of P desorbed per time is greatest (Lookman et al., 1995), the difference in desorption between 5 minutes (Mehlich-3) and 30 minutes (Olsen) is large and might explain the discrepancy in *R*. To rephrase as a new hypothesis: greater soil P buffer capacity implies greater amounts of inorganic P beyond the 'labile' threshold, which both soil test solutions will eventually extract; while Mehlich-3 is the more chemically aggressive solution, and will extract much more P into solution than Olsen for equivalent times, it has much less time to do so; therefore, *given a certain amount of labile P*, the derivative in STP with respect to buffer capacity is greater for Olsen than for Mehlich-3, and thus *R* increases with more buffer capacity.

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

While soil P quantity and buffer capacity appeared to be the most significant factors explaining the variation in *R*, chelatable P and metal oxides were also predictive. This is relevant as many of the sorption sites for P in soils such as clay minerals, metal oxides, and carbonates can be chelated thus mobilizing the bound P into solution (Golterman and Booman, 1988; Jan et al., 2013). While the Olsen extraction contains no such chelating power, Mehlich-3 contains acetate and EDTA, both strong chelating agents. For this reason, Mehlich-3 extractable metals correlate strongly with oxalate-extractable metals (Kleinman and Sharpley, 2002). The oxalate extraction is noteworthy since chelation is the principal reason for its high P extraction efficiency (Borggaard, 1992; Jan et al., 2013; Schwertmann, 1991). For example, in the Legacy P dataset, ca. 40 to 80% of total P was extracted with oxalate; hence, much of the inorganic P in these soils can mobilize with chelation. This chelation-based aspect of the Mehlich-3 extraction is difficult to predict with only P_{Ols} and basic soil properties, but the opposite is true when predicting P_{Ols} given Mehlich-3 data, likely explaining much of the difference in predictive performance between the Mehlich-3 and Olsen models (Table S1). Indeed, including oxalate-extractable P, Fe, and/or Al substantially improved predictive performance in the 'lability'-based models (Table S1). Unless oxalate or similar extraction data (Rogers et al., 2019) are more available, there will remain a gap in our ability to convert toward P_{M3} or similar soil test P featuring chelators from soil tests without such chelators. Therefore, the conversion toward P_{Ols} is the more reliable direction.

4.3 Agronomic utility of soil test P conversions

Our results make clear the fundamental challenges in reconciling different STP across regions: (1) the same STP value for any method but for two different soils does not imply the same P availability to either plants or solution and, relatedly, (2) the conversion ratio between different STP for one soil is not inherently applicable for another soil. For several reasons, different STP and different crop calibrations are used for agronomic decisions across many regions, including different states in the US (Lyons et al., 2023) and countries/regions in Europe (Jordan-Meille et al., 2012). Agronomists in these regions typically recommend fertilizers based on local crop correlation and calibration trials with the locally-preferred STP, resulting in a wide range in STP thresholds (Steinfurth et al., 2022). Thus, introducing alternative STP – for example, through commercial soil testing laboratories – risks losing agronomic relevance. We show that conversion via region-level equations, compared to the sample-level models we propose, can lead to gross errors with major agronomic consequences.

The primary benefit of sample-level conversion models is their ability to account for nuances in soil contexts. Figure 5 illustrates the conversion from P_{M3} to P_{Ols} for all 897 soils studied here when assuming various P_{M3} values. Given P_{M3}, each region-level equation reviewed in Steinfurth et al. (2022) predicts precisely *one* P_{Ols} value. In contrast, the Mehlich-3 GLM presented here gives 897 unique conversions tailored to the soil properties. This flexibility matters greatly when considering agronomic thresholds (i.e., the STP at which crop response to added P ceases to occur); for comparison, the average critical P_{Ols} for many regions and crops is ca. 15 mg P_{Ols} kg^{-1} (Steinfurth et al., 2022). Such thresholds map to precisely one P_{M3} value when converting with region-level equations, thus, if given P_{M3}, only one judgment would be made regardless of soil context. In reality, agronomic decisions are heavily soil-dependent. Further, the Mehlich-3 GLM even accommodates some rare extremes. For example, assuming 100 mg P_{M3} kg^{-1}, all reference equations predicted >15 mg P_{Ols} kg^{-1}, yet the Mehlich-3 GLM

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

predicted <15 mg P_{Ols} kg^{-1} for three soils. These three soils were all very sandy, from dry or arid climates, and derived from eolian parent material (e.g., sand dune) or similar. Such soils are extremely poor in their P buffer capacity, thus the very low P_{Ols} values are sensible. At the other extreme, the sample-level model accommodated soils with very high P buffer capacity, e.g., clay- and Al-rich soils from the L. Champ. site, predicting nearly double the P_{Ols} value compared to literature equations. Overall, the most agronomically problematic scenarios are when STP values are very low, where recommended P-application rates increase rapidly per unit yield, and where the total crop P-demand is high. To illustrate, we will discuss two widely grown crops with contrasting P demands – barley and potato – with the P recommendations from the western US state of Idaho, where P_{Ols} is the current recommend STP (Figure 6). Using the Idaho recommendations for these crops (Spackman et al., 2023; Stark et al., 2004), P fertilizer would be applied when STP is at or below either 15 (barley) or 20 (potato) mg P_{Ols} kg^{-1}. Simulating these decisions for all 897 soils in our dataset, but with P_{M3} assumed to be 10, 30, or 50 mg kg^{-1}, Figure 6 shows dramatic variation in the recommended P-application depending on the equation used. Given a P_{M3} value, any one region-level equation leads to a *single* recommendation, no matter the other soil properties; meanwhile, the sample-level model (Mehlich-3 GLM) had 3 to 5 (the maximum) recommended rates across the soils depending on scenario. Several region-level equations led to the most extreme recommendation for all soils while the Mehlich-3 GLM led to that recommendation for a minority of soils. These disagreements are just as apparent in the STP range most likely to elicit crop response. A clear example is for 10 P_{M3} kg^{-1} for a potato crop: half of the equations suggested the maximum rate (156 kg P ha^{-1}) for every soil, but the GLM reserved this extreme for just 7% of soils.

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

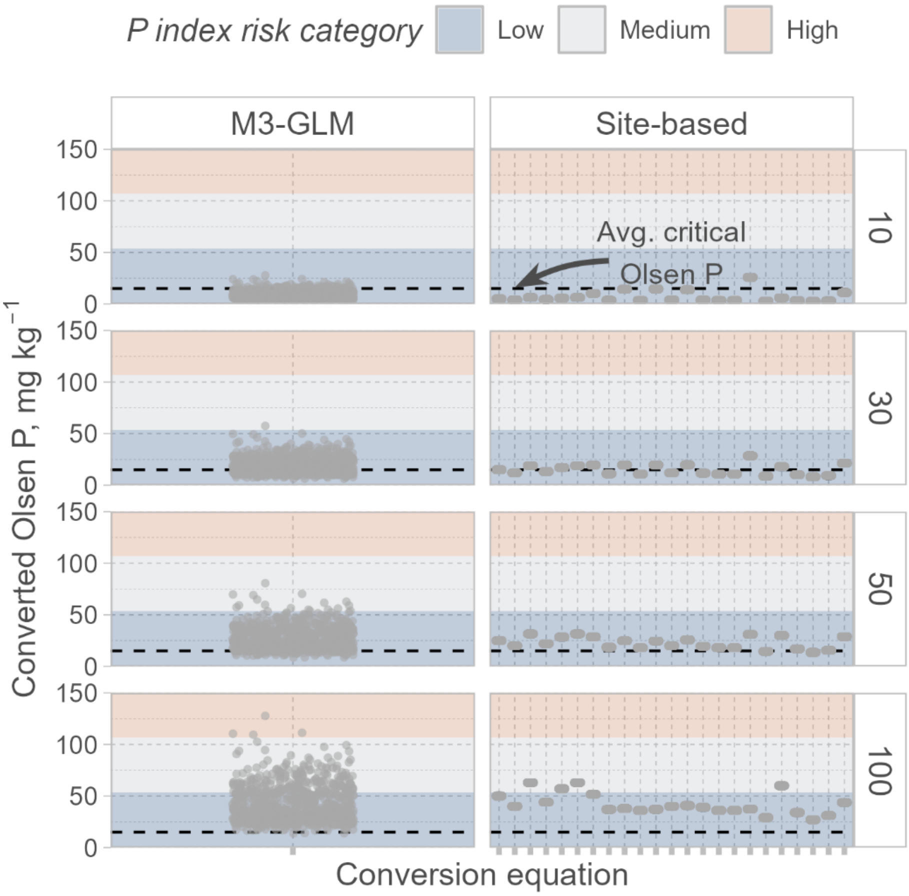


Figure 5: Assuming different values of Mehlich-3 P concentration (across rows; 10, 30, 50, or 100 mg P kg^{-1}) but with all other observed soils properties held fixed (n = 897; ALP + Legacy P), the predicted Olsen P concentration via (left) this study's best performing Mehlich-3 GLM or via (right) one of the 23 region-level equation as reviewed in Steinfurth et al. (2021; ordered 1-23 according to their Table 9). The dashed line at 15 mg kg^{-1} Olsen P is the average critical (agronomic) value across multiple crops (Steinfurth et al., 2022) where average relative yields fall below 95% (for comparison only). Note that under M3-GLM, only one equation is shown, but given more width to visualize the data; for all the region-level equations, only one unique prediction is available per given Mehlich-3 P concentration. The background shades indicate the P loss risk category according to Idaho's P index (Leytem et al., 2017) for a sensitive scenario (again, for reference only).

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

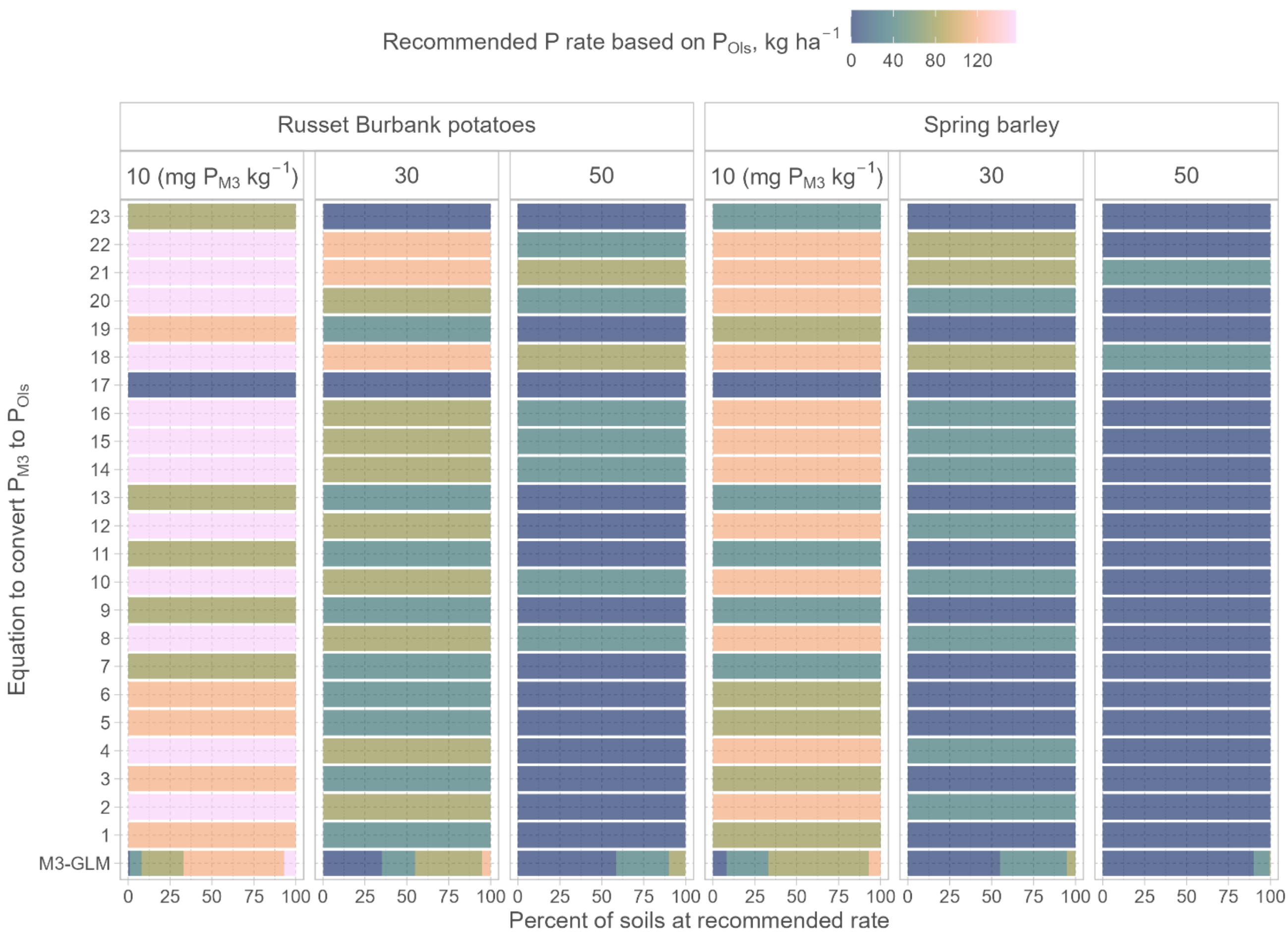


Figure 6: Using the equations and all soil properties as in Figure 5 for given Mehlich-3 P concentrations of 10, 30, or 50 mg kg^{-1}, what fertilizer P application might be recommended for a potato or barley crop in Idaho? The fertility recommendations are based on Olsen P concentration (assuming 0% free lime for simplicity). Being the only equation to leverage sample-level information, the Mehlich-3 GLM (this study) leads to a variety of P fertilizer recommendations depending on context. In contrast, all 23 of the reference (region-level) equations give one conversion of Mehlich-3 P and thus one fertilizer recommendation for all 897 soils, no matter the soil properties.

The conversion from P_{M3} to P_{Ols} is particularly relevant for farmers and agronomists. A recent trend in soil testing appears for samples sent to (frequently far-flung) commercial laboratories, where there is often a preference for multi-elemental extracts such as Mehlich-3 largely due to the reduced cost of a multi-nutrient extractant compared to multiple tests for various nutrients (Kaiser, 2024; Mattila and Rajala, 2022; Miller et al., 2013; Rogers et al., 2019). This is worrying considering the complex chemistry of the Mehlich-3 extract, as discussed above, and the lack of crop calibration to P_{M3} for the soils in question. While conversions from Mehlich-3 data to P_{Ols} were the most robust here, 29% relative error may still obscure local P fertilizer recommendations.

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

Fertilizer decisions are critical for achieving sustainable nutrient management – in both the environmental and agronomic senses. STP measurements and subsequent fertilizer recommendations from external labs can vary widely, entailing significant costs to the farmer, sometimes unnecessarily (Baker et al., 2026; Jordan-Meille et al., 2012; Liuzza et al., 2020). Critical STP values based on crop calibration trials are already uncertain due to difficulty in reconciling different soils, climates, and crop characteristics as well as relatively arbitrary model selection choices (Cate Jr. and Nelson, 1971; Fryer et al., 2019; Pearce et al., 2022). Compounding measurement error and uncertain crop calibration values with >50% relative error due to region-level conversions can render STP data useless for practical purposes. Refining conversions between P_{M3} and P_{Ols} is a relatively small but important step to improving fertilizer recommendations and guiding P fertilizer policy across regions.

4.4 Environmental utility of soil test P conversions

Soil P measurements most germane to hydrological transport of P, such as P intensity and labile P, are infrequently made. So, it is pragmatic to instead use, if possible, other widely-measured soil P data, such as STP. If we are to leverage these data as inputs for P transport models and risk assessments (e.g., P index), we need flexibility in converting soil P metrics toward the appropriate basis with minimal error. The conversion equations developed here fulfill some of this need and should prove useful for such environmental applications.

A clear example is in P indices and related agricultural P assessment (Liu et al., 2025; Osmond et al., 2025), where STP data are often the main variable defining the 'supply' component for P transport risk. Thus, STP conversion errors can skew risk assessments. This is especially true when critical threshold values of STP are used categorically to define risk or permittable actions. Near such critical values, small changes in STP values have great impacts. Returning to Figure 5, the converted P_{Ols} are graded according to Idaho's P index (Leytem et al., 2017), which requires P_{Ols}. When using region-level equations, no nuance is allowed and so all soils are given *one* risk class. At 100 mg P_{M3} kg^{-1}, most region-level equations assigned all soils the 'low' risk category while four assigned all soils 'medium'; the Mehlich-3 GLM, in contrast, assigned most soils 'low' risk but still identified 23% of the soils as 'medium' risk and a further 0.4% as 'high'.

A similar pattern will apply to P transport models that incorporate STP data. For example, many P transport models rely on the 3-pool soil inorganic P model (Das et al., 2019; Jones et al., 1984; Neumann et al., 2021), such as SWAT+ (Vadas and White, 2010). The first and perhaps most critical inorganic P pool is currently termed 'solution P', though originally Jones et al. (1984) called it 'labile P' (the same meaning we use in this work). In SWAT+ and other models utilizing this 3-pool model, labile P also partly defines the other two inorganic P pools, essentially dictating much of the total P content of the soil and any particulate P losses to local waters. Given large observed variability in soil labile P concentrations (e.g., 3 orders of magnitude in Simpson et al., 2025) and its central role in the simulation of P transport processes, labile P is critical input. Yet labile P is seldom measured, leaving the vast majority of model users to either assume an unideal default or base labile P on data they do have, which is likely STP. For the latter, labile P is estimated based on a relationship with the more frequently measured – for the eastern US – P_{M3} (Muenich et al., 2016; Vadas and White, 2010; White et al., 2010). Leaving aside the issue of converting any STP to labile P (another conversion likely fraught with error), this initialization for soil labile P poses problems for many

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

other regions globally that measure other STP for their agronomic and environmental purposes (Jordan-Meille et al., 2012; Lyons et al., 2023). As a result, modelers routinely face the choice of (i) using conversions based on dissimilar soils and thus likely entailing large errors (Figure S6), (ii) tuning other parameters to compensate, or (iii) ignoring the mismatch altogether.

Other catchment and regional models face similar challenges. In Balt-HYPE (Baltic Sea basin – Hydrological Predictions for the Environment; Arheimer et al., 2012), soil P pools are set using land-use-based coefficients and national datasets. However, different countries contribute STP data based on different extraction methods, and the Balt-HYPE documentation does not fully explain how those are reconciled. INCA-P (Integrated Catchment Model of P dynamics; Jackson-Blake et al., 2016) differs by initializing soil P pools based not on STP but on total P (i.e., digestion), relying on a soil P variable far less frequently measured and divorced from soil P lability. These examples reinforce the point that there is no consistent, transparent treatment of STP across models, even though STP is often the only soil P information available at scale.

All models simulating soil P dynamics but given no or erroneous soil P inputs will likely misrepresent the relevant processes, thus yielding either poor predictions or fair predictions for the wrong reasons. The soil labile P pool, particularly in agricultural catchments, dwarfs annual P fluxes (Simpson et al., 2025) and its size is central to simulating catchment P dynamics. In SWAT+, if the labile P pool is wrong, calibration to observed in-stream P loads can still achieve good fit by adjusting parameters that govern in-stream processes to compensate, while the internal P pools and edge-of-field P loads remain unrealistic. While the calibrated model seems to exhibit good performance, the parameterization and internal state of the model is unreasonable. Any scenario analyses, such as predicting how quickly legacy P will decline under reduced nutrient inputs or how long a lag to expect in improved water quality, will likely be unreasonable as well. The STP conversions proposed here alleviate some of this uncertainty for some of the most widely used models and may further guide additional model-relevant STP conversions.

4.5 Limitations and future work

We echo the caution of Steinfurth et al. (2021): "A routine use of conversions, e.g., to reduce costs in laboratories, cannot be recommended." Any conversion propagates uncertainty to downstream calculations. On average, relative error due to conversion with our GLMs is 29% (given Mehlich-3 data) or 45% (given P_{Ols}); while not >100% like many other region-level models, this error is still considerable depending on the case. Users should be critical when using our conversion equations, or any others, and they may consult our hold-one-site-out cross-validation (Figure S8) as a strenuous test of our models' generalizability.

In certain contexts, the models here may not suit well. First, our dataset did not include soils from the Gelisols or Oxisols orders, with the latter having extremely high P buffer capacity (Barbosa et al., 2022; Boitt et al., 2018) and providing an important agricultural soil for tropical regions. Second, soils with very heavy manure additions may have distinct soil P relationships from many of those used here (Condron et al., 2005; Rubæk et al., 1999; Schmieder et al., 2018). Not only is more of the total P present as organic P, but soil tests vary dramatically in their extraction of various organic P compounds or inorganic P complexed with organic matter. Alkaline extracts such as 0.5 M $NaHCO_3$ (Olsen, Colwell) will remove the more

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

labile organic P fraction (Condron et al., 2005; Hedley et al., 1982), but Mehlich-3 can also extract large amounts of certain organic P such as phytate-P in soils amended with poultry litter (Cade-Menun et al., 2018). We suspect some of this behavior explains the peculiar relationship at the Mahan. Ck. site (Figures S5; S8) where specific soils received considerable cattle and/or swine manure. Further improvement in conversion of P_{Ols} to P_{M3} would likely be improved by including other soil extractant data (analogous to Fe_{M3}, Mg_{M3}, etc., such as ammonium-acetate extraction) that are used in alkaline soils (Miller et al., 2013; Rogers et al., 2019). Lastly, while we built models with observed basic soil properties, such as clay and TOC concentrations, some users may need to instead use geospatial products (e.g., Poggio et al., 2021) or similar to estimate these properties; however, such estimates can introduce more uncertainty in soil P conversions.

We believe the method outlined here – modeling the conversion ratio itself at sample-level, rather than either of the STP – has potential for further development. More extensive datasets and from additional regions can remediate possible blind spots in our dataset. Alternative models may build upon the success we found here, where we focused on models that could readily port to other tools (without, say, storing a random forest object somewhere to be fed input data). The variety of alternative predictors we tested (Table S1) suggest several soil properties, though less commonly measured, can be helpful for converting between STP and related metrics: the P buffer capacity (e.g., Bache-Williams index) and related aspects of P lability (labile P via anion exchange membranes/resins; EPC_0), and oxalate-extractable Al, Fe, and P.

5 Conclusions

By extending the idea of a 'conversion slope' between two STP measurements at a region level to a model of their ratio at the sample level, we developed robust predictive conversion equations using readily-available variables applicable across a wide variety of soils. Compared to prior STP conversions based on region-level regressions, our models can cut average conversion errors in half and avoid the most egregious errors. Despite this improvement, different STP remain, to some degree, irreconcilable due to the chemistry involved, making direct measurements nonpareil. Still, for Mehlich-3 and Olsen specifically, reasonable conversion efficiencies are possible when accounting for the soil P lability properties and chelatable Fe, Al, and P. This is in line with the chemical mechanisms present in each extraction; similar investigations for other STP may lead to more universal conversions between multiple tests. Specific to Olsen and Mehlich-3, it is fortunate that the conversion GLM developed here works particularly well when given Mehlich-3 data, as the conversion towards P_{Ols} could be the more frequently encountered conversion in the future. For additional soil P metrics not evaluated here, we expect similar utility in conversions is possible by moving beyond region-level correlations between soil tests. This may further improve works such as large-scale maps of soil P, bridging agronomic efforts between neighboring states/regions/countries, and meta-analyses desiring a consistent soil P basis.

When direct measurements are infeasible and conversions necessary, providers and users of soil P data should be informed and careful, as conversion errors are surprisingly impactful. Reducing STP conversion error helps with both environmental and agronomic decision-making; however, soil monitoring programs, field trials, analytical capacities, accessible soil

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

databases, and more must be bolstered to provide the soil P data necessary to answer questions critical to sustainable P management.

Code, data, or code and data availability

R code for the analyses, along with a spreadsheet version for the conversions, is given at Figshare: https://doi.org/10.6084/m9.figshare.32133124.

Data from the USDA Legacy P project was previously published (Simpson et al. 2025) and made available at Ag Data Commons: https://doi.org/10.15482/USDA.ADC/25892602.v1

Data from the ALP is published at: https://collaborative-testing.com/program-1.php

Author contributions

CWR and ZPS conceived the study; ZPS, CWR, JDM, and NC conceptualized the study; ZPS and JDM designed methodology with input from CWR; ZPS performed analyses; KRE and ROM curated data and provided technical advice on lab analyses; ZPS and CWR wrote the manuscript with contributions from all co-authors.

Competing interests

The authors declare that they have no conflict of interest.

Disclaimer

Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

USDA is an equal opportunity employer and provider.

Acknowledgements

We thank all the landowners and field and laboratory staff involved in making the combined ALP & USDA Legacy P dataset possible. The USDA Legacy P dataset was made possible through the USDA Conservation Effects Assessment Project.

Financial support

Support came from the USDA-NRCS Conservation Effects Assessment Project (Agreement NRC21IRA0010879).

Review statement

The review statement will be added by Copernicus Publications listing the handling editor as well as all contributing referees according to their status anonymous or identified.

References

Adesanwo, O. O., Ige, D. V., Thibault, L., Flaten, D., and Akinremi, W.: Comparison of Colorimetric and ICP Methods of Phosphorus Determination in Soil Extracts, Communications in Soil Science and Plant Analysis, 2013.

Arheimer, B., Dahné, J., Donnelly, C., Lindström, G., and Strömqvist, J.: Water and nutrient simulations using the HYPE model for Sweden vs. the Baltic Sea basin – influence of input-data quality and scale, Hydrology Research, 43, 315–329, https://doi.org/10.2166/nh.2012.010, 2012.

Bache, B. and Williams, E.: A phosphate sorption index for soil, Eur. J. Soil Sci., 22, 289–301, https://doi.org/10.1111/j.1365-2389.1971.tb01617.x, 1971.

Baker, M., Yost, M., Creech, J. E., Cardon, G., Gale, J., Price, S., Pace, M., Wytsalucy, R., Zesiger, C., Nelson, M., and Violett, R.: Comparison of various fertilizer recommendations for forage crops in the Western United States, Crop Science, 66, e70223, https://doi.org/10.1002/csc2.70223, 2026.

Barbosa, J. Z., Poggere, G., Mancini, M., Silva, S. H. G., Motta, A. C. V., Marques, J. J. G. de S. e M., and Curi, N.: National-scale spatial variations of soil phosphorus retention capacity in Brazil, Physics and Chemistry of the Earth, Parts A/B/C, 128, 103271, https://doi.org/10.1016/j.pce.2022.103271, 2022.

Barrow, N. J. and Shaw, T. C.: Sodium bicarbonate as an extractant for soil phosphate, II. Effect of varying the conditions of extraction on the amount of phosphate initially displaced and on the secondary adsorption, Geoderma, 16, 109–123, https://doi.org/10.1016/0016-7061(76)90034-3, 1976.

Beckett, P. H. T. and White, R. E.: Studies on the phosphate potentials of soils - Part III: The pool of labile inorganic phosphate, Plant Soil, 21, 253–282, https://doi.org/10.1007/BF01377744, 1964.

Boitt, G., Schmitt, D. E., Gatiboni, L. C., Wakelin, S. A., Black, A., Sacomori, W., Cassol, P. C., and Condron, L. M.: Fate of phosphorus applied to soil in pig slurry under cropping in southern Brazil, Geoderma, 321, 164–172, https://doi.org/10.1016/j.geoderma.2018.02.010, 2018.

Borggaard, O. K.: Dissolution of Poorly Crystalline Iron Oxides in Soils by EDTA and Oxalate, Zeitschrift für Pflanzenernährung und Bodenkunde, 155, 431–436, https://doi.org/10.1002/jpln.19921550513, 1992.

Bray, R. H.: Requirements for successful soil tests, Soil Science, 66, 83, 1948.

Burnham, K. P. and Anderson, D. R.: Multimodel Inference: Understanding AIC and BIC in Model Selection, Sociological Methods & Research, 33, 261–304, https://doi.org/10.1177/0049124104268644, 2004.

Cade-Menun, B. J., Elkin, K. R., Liu, C. W., Bryant, R. B., Kleinman, P. J. A., and Moore, P. A.: Characterizing the phosphorus forms extracted from soil by the Mehlich III soil test, Geochemical Transactions, 19, 7, https://doi.org/10.1186/s12932-018-0052-9, 2018.

Cate Jr., R. B. and Nelson, L. A.: A Simple Statistical Procedure for Partitioning Soil Test Correlation Data Into Two Classes, Soil Science Society of America Journal, 35, 658–660, https://doi.org/10.2136/sssaj1971.03615995003500040048x, 1971.

Condron, L. M., Turner, B. L., and Cade-Menun, B. J.: Chemistry and Dynamics of Soil Organic Phosphorus, in: Phosphorus: Agriculture and the Environment, John Wiley & Sons, Ltd, 87–121, https://doi.org/10.2134/agronmonogr46.c4, 2005.

Culman, S. W., Mann, M., Sharma, S., Saeed, M. T., Fulford, A. M., Lindsey, L. E., Brooker, A., Dayton, E., Warden, R., and Joern, B.: Calibration of Mehlich-3 with Bray P1 and Ammonium Acetate in the Tri-State Region of Ohio, Indiana and Michigan, Communications in Soil Science and Plant Analysis, 51, 86–97, https://doi.org/10.1080/00103624.2019.1695825, 2020.

Das, B., Huth, N., Probert, M., Condron, L., and Schmidt, S.: Soil Phosphorus Modeling for Modern Agriculture Requires Balance of Science and Practicality: A Perspective, Journal of Environmental Quality, 48, 1281–1294, https://doi.org/10.2134/jeq2019.05.0201, 2019.

van Doorn, M., van Rotterdam, D., Ros, G., Koopmans, G. F., Smolders, E., and de Vries, W.: The phosphorus saturation degree as a universal agronomic and environmental soil P test, Critical Reviews in Environmental Science and Technology, 54, 385–404, https://doi.org/10.1080/10643389.2023.2240211, 2024.

Dufey, F.: Derivation of Passing–Bablok regression from Kendall's tau, The International Journal of Biostatistics, 16, https://doi.org/10.1515/ijb-2019-0157, 2020.

Dyer, B.: XV.—On the analytical determination of probably available "mineral" plant food in soils, J. Chem. Soc., Trans., 65, 115–167, https://doi.org/10.1039/CT8946500115, 1894.

Eriksson, A. K., Ulén, B., Berzina, L., Iital, A., Janssons, V., Sileika, A. S., and Toomsoo, A.: Phosphorus in agricultural soils around the Baltic Sea – comparison of laboratory methods as indices for phosphorus leaching to waters, Soil Use and Management, 29, 5–14, https://doi.org/10.1111/j.1475-2743.2012.00402.x, 2013.

Fryer, M. S., Slaton, N. A., Roberts, T. L., and Ross, W. J.: Validation of Soil-Test-Based Phosphorus and Potassium Fertilizer Recommendations for Irrigated Soybean, Soil Science Society of America Journal, 83, 825–837, https://doi.org/10.2136/sssaj2019.02.0032, 2019.

Gelman, A., Hwang, J., and Vehtari, A.: Understanding predictive information criteria for Bayesian models, Stat Comput, 24, 997–1016, https://doi.org/10.1007/s11222-013-9416-2, 2014.

Golterman, H. L. and Booman, A.: Sequential extraction of iron-phosphate and calcium-phosphate from sediments by chelating agents, SIL Proceedings, 1922-2010, 23, 904–909, https://doi.org/10.1080/03680770.1987.11899738, 1988.

Grant, C. A. and Flaten, D. N.: 4R Management of Phosphorus Fertilizer in the Northern Great Plains, Journal of Environmental Quality, 48, 1356–1369, https://doi.org/10.2134/jeq2019.02.0061, 2019.

Hedley, M. J., Stewart, J. W. B., and Chauhan, B. S.: Changes in Inorganic and Organic Soil Phosphorus Fractions Induced by Cultivation Practices and by Laboratory Incubations, Soil Science Society of America Journal, 46, 970–976, https://doi.org/https://doi.org/10.2136/sssaj1982.03615995004600050017x, 1982.

HELCOM: HELCOM Baltic Sea Action Plan – 2021 update, HELCOM, 2021.

Holford, I. C. R.: Greenhouse Evaluation of Four Phosphorus Soil Tests in Relation to Phosphate Buffering and Labile Phosphate in Soils, Soil Science Society of America Journal, 44, 555–559, https://doi.org/10.2136/sssaj1980.03615995004400030024x, 1980.

Holford, I. C. R.: Soil phosphorus: its measurement, and its uptake by plants, Soil Res., 35, 227–240, https://doi.org/10.1071/s96047, 1997.

Jackson-Blake, L. A., Wade, A. J., Futter, M. N., Butterfield, D., Couture, R.-M., Cox, B. A., Crossman, J., Ekholm, P., Halliday, S. J., Jin, L., Lawrence, D. S. L., Lepistö, A., Lin, Y., Rankinen, K., and Whitehead, P. G.: The INtegrated CAtchment model of phosphorus dynamics (INCA-P): Description and demonstration of new model structure and equations, Environmental Modelling & Software, 83, 356–386, https://doi.org/10.1016/j.envsoft.2016.05.022, 2016.

Jan, J., Borovec, J., Kopáček, J., and Hejzlar, J.: What do results of common sequential fractionation and single-step extractions tell us about P binding with Fe and Al compounds in non-calcareous sediments?, Water Res., 47, 547–557, https://doi.org/10.1016/j.watres.2012.10.053, 2013.

Jones, C. A., Cole, C. V., Sharpley, A. N., and Williams, J. R.: A Simplified Soil and Plant Phosphorus Model: I. Documentation, Soil Science Society of America Journal, 48, 800–805, https://doi.org/10.2136/sssaj1984.03615995004800040020x, 1984.

Jordan-Meille, L., Rubæk, G. H., Ehlert, P. a. I., Genot, V., Hofman, G., Goulding, K., Recknagel, J., Provolo, G., and Barraclough, P.: An overview of fertilizer-P recommendations in Europe: soil testing, calibration and fertilizer recommendations, Soil Use and Management, 28, 419–435, https://doi.org/10.1111/j.1475-2743.2012.00453.x, 2012.

Kaiser, D.: What is the best soil test option for phosphorus?, Minnesota Crop News, 2024.

Kleinman, P. J. A. and Sharpley, A. N.: Estimating soil phosphorus sorption saturation from Mehlich-3 data, null, 33, 1825–1839, https://doi.org/10.1081/CSS-120004825, 2002.

Kleinman, P. J. A., Fanelli, R. M., Hirsch, R. M., Buda, A. R., Easton, Z. M., Wainger, L. A., Brosch, C., Lowenfish, M., Collick, A. S., Shirmohammadi, A., Boomer, K., Hubbart, J. A., Bryant, R. B., and Shenk, G. W.: Phosphorus and the Chesapeake Bay: Lingering Issues and Emerging Concerns for Agriculture, Journal of Environmental Quality, 48, 1191–1203, https://doi.org/10.2134/jeq2019.03.0112, 2019.

Koopmans, G. F., McDowell, R. W., Chardon, W. J., Oenema, O., and Dolfing, J.: Soil phosphorus quantity–intensity relationships to predict increased soil phosphorus loss to overland and subsurface flow, Chemosphere, 48, 679–687, https://doi.org/10.1016/S0045-6535(02)00146-7, 2002.

Kovar, J. L. and Pierzynski, G. M. (Eds.): Methods of Phosphorus Analysis for Soils, Sediments, Residuals, and Waters, Second Edition., Virginia Tech University, Virginia, USA, 2009.

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

Lambers, H.: Phosphorus Acquisition and Utilization in Plants, Annual Review of Plant Biology, 73, 17–42, https://doi.org/10.1146/annurev-arplant-102720-125738, 2022.

Larsen, S.: Soil Phosphorus, in: Advances in Agronomy, vol. 19, edited by: Norman, A. G., Academic Press, 151–210, https://doi.org/10.1016/S0065-2113(08)60735-X, 1967.

Leytem, A. B., Bjorneberg, D. L., and Tarkalson, D.: The Phosphorus Site Index: A Systematic Approach to Assess the Risk of Nonpoint Source Pollution of Idaho Waters by Agricultural Phosphorus, 2017.

Liu, J., Djodjic, F., Ulén, B., Aronsson, H., Bechmann, M., Bergström, L., Krogstad, T., and Kyllmar, K.: Toward better targeting of mitigation measures for reducing phosphorus losses from land to water: Andrew Sharpley's legacy in Norway and Sweden, Journal of Environmental Quality, 54, 777–787, https://doi.org/10.1002/jeq2.20558, 2025.

Liuzza, L. M., Bush, E. W., Tubana, B. S., and Gaston, L. A.: Determining Nutrient Recommendations for Agricultural Crops Based on Soil and Plant Tissue Analyzes between Different Analytical Laboratories, Communications in Soil Science and Plant Analysis, 51, 392–402, https://doi.org/10.1080/00103624.2019.1709488, 2020.

Lookman, Richard., Freese, Dirk., Merckx, Roel., Vlassak, Karel., and van Riemsdijk, W. H.: Long-Term Kinetics of Phosphate Release from Soil, Environ. Sci. Technol., 29, 1569–1575, https://doi.org/10.1021/es00006a020, 1995.

Lyons, S. E., Arthur, D. K., Slaton, N. A., Pearce, A. W., Spargo, J. T., Osmond, D. L., and Kleinman, P. J. A.: Development of a soil test correlation and calibration database for the USA, Agricultural & Environmental Letters, 6, e20058, https://doi.org/10.1002/ael2.20058, 2021.

Lyons, S. E., Clark, J. D., Osmond, D. L., Parvej, M. R., Pearce, A. W., Slaton, N. A., and Spargo, J. T.: Current status of US soil test phosphorus and potassium recommendations and analytical methods, Soil Science Society of America Journal, 87, 985–998, https://doi.org/10.1002/saj2.20536, 2023.

Macrae, M., Jarvie, H., Brouwer, R., Gunn, G., Reid, K., Joosse, P., King, K., Kleinman, P., Smith, D., Williams, M., and Zwonitzer, M.: One size does not fit all: Toward regional conservation practice guidance to reduce phosphorus loss risk in the Lake Erie watershed, Journal of Environmental Quality, 50, 529–546, https://doi.org/10.1002/jeq2.20218, 2021.

Mattila, T. J. and Rajala, J.: Do different agronomic soil tests identify similar nutrient deficiencies?, Soil Use and Management, 38, 635–648, https://doi.org/10.1111/sum.12738, 2022.

McDowell, R., Sharpley, A., Brookes, P., and Poulton, P.: Relationship between soil test phosphorus and phosphorus release to solution, Soil Science, 166, 137, 2001.

McDowell, R., Dodd, R., Pletnyakov, P., and Noble, A.: The Ability to Reduce Soil Legacy Phosphorus at a Country Scale, Front. Environ. Sci., 8, https://doi.org/10.3389/fenvs.2020.00006, 2020.

McDowell, R. W., Simpson, Z. P., Doscher, C., Steinfurth, K., Mott, J. D., Margenot, A. J., Appelhans, S. C., Elledge, A. E., Thornton, C. M., Moore, P. A., Blackwell, M. S. A., Cade-Menun, B. J., Ros, M. B. H., Pavinato, P. S., Zavattaro, L., Soltangheisi, A., Zhang, T. Q., Haygarth, P. M., Burkitt, L., and Fenton, O.: Managing the reduction of soil phosphorus can prolong global reserves of fertilizer phosphorus and improve water quality, One Earth, 8, 101448, https://doi.org/10.1016/j.oneear.2025.101448, 2025.

Mehlich, A.: Mehlich 3 soil test extractant: A modification of Mehlich 2 extractant, Communications in Soil Science and Plant Analysis, 15, 1409–1416, https://doi.org/10.1080/00103628409367568, 1984.

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

Mendez, J. C., Hiemstra, T., and Koopmans, G. F.: Assessing the Reactive Surface Area of Soils and the Association of Soil Organic Carbon with Natural Oxide Nanoparticles Using Ferrihydrite as Proxy, Environ. Sci. Technol., 54, 11990–12000, https://doi.org/10.1021/acs.est.0c02163, 2020.

Messiga, A. J., Ba, Y., Ziadi, N., Bélanger, G., and Lafond, J.: Assessing the depletion of soil P following sequential extractions with Mehlich-3 and Olsen solutions, Archives of Agronomy and Soil Science, 60, 1445–1458, https://doi.org/10.1080/03650340.2014.884709, 2014.

Miller, R. O., Gavlak, R., and Horneck, D. (Eds.): Soil, Plant and Water Reference Methods for the Western Region, Western Coordinating Committee on Nutrient Management (WERA-103), 2013.

Muenich, R. L., Kalcic, M., and Scavia, D.: Evaluating the Impact of Legacy P and Agricultural Conservation Practices on Nutrient Loads from the Maumee River Watershed, Environ. Sci. Technol., 50, 8146–8154, https://doi.org/10.1021/acs.est.6b01421, 2016.

Muntwyler, A., Panagos, P., Morari, F., Berti, A., Jarosch, K. A., Mayer, J., and Lugato, E.: Modelling phosphorus dynamics in four European long-term experiments, Agricultural Systems, 206, 103595, https://doi.org/10.1016/j.agsy.2022.103595, 2023.

Muntwyler, A., Panagos, P., Pfister, S., and Lugato, E.: Assessing the phosphorus cycle in European agricultural soils: Looking beyond current national phosphorus budgets, Science of The Total Environment, 906, 167143, https://doi.org/10.1016/j.scitotenv.2023.167143, 2024.

Neumann, A., Dong, F., Shimoda, Y., Arnillas, C. A., Javed, A., Yang, C., Zamaria, S., Mandal, S., Wellen, C., Paredes, D., Feisthauer, N., Blukacz-Richards, E. A., Yerubandi, R. R., and Arhonditsis, G. B.: A review of the current state of process-based and data-driven modelling: guidelines for Lake Erie managers and watershed modellers, Environ. Rev., 29, 443–490, https://doi.org/10.1139/er-2020-0070, 2021.

Nguyen, P. V., McDowell, R. W., and Condron, L. M.: Impact of green manure crop species on rhizosphere soil phosphorus, Soil Res., 62, https://doi.org/10.1071/SR22257, 2024.

Olsen, S. R., Cole, C. V., Watanabe, F. S., and Dean, L. A.: Estimation of available phosphorus in soils by extraction with sodium bicarbonate, U.S. Department of Agriculture, USA, 1954.

Osmond, D. L., Kleinman, P. J. A., Coale, F., Nelson, N. O., Bolster, C. H., and McGrath, J.: A short history of the phosphorus index and Andrew Sharpley's contributions from inception through development and implementation, Journal of Environmental Quality, 54, 788–797, https://doi.org/10.1002/jeq2.20535, 2025.

Pearce, A. W., Slaton, N. A., Lyons, S. E., Bolster, C. H., Bruulsema, T. W., Grove, J. H., Jones, J. D., McGrath, J. M., Miguez, F. E., and Nelson, N. O.: Defining relative yield for soil test correlation and calibration trials in the Fertilizer Recommendation Support Tool, Soil Sci. Soc. Am. J., 86, 1338–1353, https://doi.org/10.1002/saj2.20450, 2022.

Pferdmenges, J., Breuer, L., Julich, S., and Kraft, P.: Review of soil phosphorus routines in ecosystem models, Environmental Modelling & Software, 126, 104639, https://doi.org/10.1016/j.envsoft.2020.104639, 2020.

Pierzynski, G. M., McDowell, R. W., and Thomas Sims, J.: Chemistry, Cycling, and Potential Movement of Inorganic Phosphorus in Soils, in: Phosphorus: Agriculture and the Environment, John Wiley & Sons, Ltd, 51–86, https://doi.org/10.2134/agronmonogr46.c3, 2005.

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, SOIL, 7, 217–240, https://doi.org/10.5194/soil-7-217-2021, 2021.

Potapov, S., Model, F., Schuetzenmeister, A., Manuilova, E., Dufey, F., and Raymaekers, J.: mcr: Method Comparison Regression, , https://doi.org/10.32614/CRAN.package.mcr, 2024.

Radcliffe, D. E., Reid, D. K., Blombäck, K., Bolster, C. H., Collick, A. S., Easton, Z. M., Francesconi, W., Fuka, D. R., Johnsson, H., King, K., Larsbo, M., Youssef, M. A., Mulkey, A. S., Nelson, N. O., Persson, K., Ramirez-Avila, J. J., Schmieder, F., and Smith, D. R.: Applicability of Models to Predict Phosphorus Losses in Drained Fields: A Review, Journal of Environmental Quality, 44, 614–628, https://doi.org/10.2134/jeq2014.05.0220, 2015.

Rogers, C. W., Dari, B., and Schroeder, K. L.: Comparison of Soil-Test Extractants for Potassium, Calcium, Magnesium, Sulfur, and Micronutrients in Idaho Soils, Agrosystems, Geosciences & Environment, 2, 190067, https://doi.org/10.2134/age2019.08.0067, 2019.

Rubæk, G. H., Guggenberger, G., Zech, W., and Christensen, B. T.: Organic Phosphorus in Soil Size Separates Characterized by Phosphorus-31 Nuclear Magnetic Resonance and Resin Extraction, Soil Science Society of America Journal, 63, 1123–1132, https://doi.org/10.2136/sssaj1999.6351123x, 1999.

Schmieder, F., Bergström, L., Riddle, M., Gustafsson, J.-P., Klysubun, W., Zehetner, F., Condron, L., and Kirchmann, H.: Phosphorus speciation in a long-term manure-amended soil profile – Evidence from wet chemical extraction, 31P-NMR and P K-edge XANES spectroscopy, Geoderma, 322, 19–27, https://doi.org/10.1016/j.geoderma.2018.01.026, 2018.

Schwertmann, U.: Solubility and dissolution of iron oxides, Plant Soil, 130, 1–25, https://doi.org/10.1007/BF00011851, 1991.

Sharpley, A. N., Jones, C. A., Gray, C., and Cole, C. V.: A Simplified Soil and Plant Phosphorus Model: II. Prediction of Labile, Organic, and Sorbed Phosphorus, Soil Science Society of America Journal, 48, 805–809, https://doi.org/10.2136/sssaj1984.03615995004800040021x, 1984.

Shwiekh, R., Kratz, S., Schick, J., Kammerer, H., Ahmed, S., and Schnug, E.: Determination of inorganic and organic P dissolved in water and Olsen extracts by inductively coupled plasma optical emission spectroscopy (ICO-OES) and colorimetry, Landbauforschung, 63, 303–306, https://doi.org/10.3220/LBF_2013_303-306, 2013.

Shwiekh, R., Schick, J., Kratz, S., Rückamp, D., and Schnug, E.: Comparative investigation and inter-calibration of different soil P tests, Journal für Kulturpflanzen, 67, 61–72, 2015.

Simpson, G. L.: gratia: Graceful ggplot-Based Graphics and Other Functions for GAMs Fitted using mgcv, 2025.

Simpson, Z. P., Mott, J., Elkin, K., Buda, A., Faulkner, J., Hapeman, C., McCarty, G., Foroughi, M., Hively, W. D., King, K., Osterholz, W., Penn, C., Williams, M., Witthaus, L., Locke, M., Pawlowski, E., Dalzell, B., Feyereisen, G., Dolph, C., Bjorneberg, D., Nouwakpo, K., Rogers, C. W., Scott, I., Bolster, C. H., Duriancik, L., and Kleinman, P. J. A.: Phosphorus lability across diverse agricultural contexts with legacy sources, Journal of Environmental Quality, 54, 851–869, https://doi.org/10.1002/jeq2.20632, 2025.

Sims, J. T., Edwards, A. C., Schoumans, O. F., and Simard, R. R.: Integrating Soil Phosphorus Testing into Environmentally Based Agricultural Management Practices, Journal of Environmental Quality, 29, 60–71, https://doi.org/10.2134/jeq2000.00472425002900010008x, 2000.

https://doi.org/10.5194/egusphere-2026-2518
Preprint. Discussion started: 20 May 2026

Spackman, J. A., Rogers, C. W., Marshall, J. M., Rashed, A., Morishita, D. W., and Adjesiwor, A.: 2023 Southern Idaho Spring Barley Quick Facts, University of Idaho, Moscow, ID, USA, 2023.

Stark, J., Westermann, D., and Hopkins, B.: Nutrient Management Guidelines for Russet Burbank Potatoes, University of Idaho, Moscow, ID, USA, 2004.

Steinfurth, K., Hirte, J., Morel, C., and Buczko, U.: 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, 2021.

Steinfurth, K., Börjesson, G., Denoroy, P., Eichler-Löbermann, B., Gans, W., Heyn, J., Hirte, J., Huyghebaert, B., Jouany, C., Koch, D., Merbach, I., Mokry, M., Mollier, A., Morel, C., Panten, K., Peiter, E., Poulton, P. R., Reitz, T., Rubæk, G. H., Spiegel, H., van Laak, M., von Tucher, S., and Buczko, U.: Thresholds of target phosphorus fertility classes in European fertilizer recommendations in relation to critical soil test phosphorus values derived from the analysis of 55 European long-term field experiments, Agriculture, Ecosystems & Environment, 332, 107926, https://doi.org/10.1016/j.agee.2022.107926, 2022.

Truog, E.: The Determination of the Readily Available Phosphorus of Soils, Agronomy Journal, 22, 874–882, https://doi.org/10.2134/agronj1930.00021962002200100008x, 1930.

Vadas, P. A. and White, M. J.: Validating Soil Phosphorus Routines in the SWAT Model, Trans. ASABE, 53, 1469–1476, https://doi.org/10.13031/2013.34897, 2010.

Wanke, D. J., Heichel, J., Zikeli, S., Müller, T., and Hartmann, T. E.: Comparison of soil phosphorus extraction methods regarding their suitability for organic farming systems, Journal of Plant Nutrition and Soil Science, 186, 599–608, https://doi.org/10.1002/jpln.202300129, 2023.

White, M. J., Storm, D. E., Busteed, P. R., Smolen, M. D., Zhang, H., and Fox, G. A.: A quantitative phosphorus loss assessment tool for agricultural fields, Environmental Modelling & Software, 25, 1121–1129, https://doi.org/10.1016/j.envsoft.2010.03.017, 2010.

White, P. J. and Hammond, J. P. (Eds.)De Kok, L. J., Hawkesford, M. J., and Stulen, I. (Eds.): The Ecophysiology of Plant-Phosphorus Interactions, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-1-4020-8435-5, 2008.

Wood, S. N.: Generalized Additive Models : An Introduction with R, Second Edition, Second., Chapman and Hall/CRC, https://doi.org/10.1201/9781315370279, 2017.

Wuenscher, R., Unterfrauner, H., Peticzka, R., and Zehetner, F.: A comparison of 14 soil phosphorus extraction methods applied to 50 agricultural soils from Central Europe, Plant, Soil and Environment, 61, 86–96, https://doi.org/10.17221/932/2014-PSE, 2015.