


# A comparison of soil tests for available phosphorus in long-term field experiments in Europe

S. NAWARA<sup>a</sup> , T. VAN DAEL<sup>a</sup>, R. MERCKX<sup>a</sup>, F. AMERY<sup>b</sup>, A. ELSSEN<sup>c</sup>, W. ODEURS<sup>c</sup>, H. VANDENDRIESSCHE<sup>c,d</sup>, S. MCGRATH<sup>e</sup>, C. ROISIN<sup>f</sup>, C. JOUANY<sup>g</sup>, S. PELLERIN<sup>h</sup>, P. DENOROY<sup>h</sup>, B. EICHLER-LÖBERMANN<sup>i</sup>, G. BÖRJESSON<sup>j</sup>, P. GOOS<sup>k,l</sup>, W. AKKERMANS<sup>k</sup> & E. SMOLDERS<sup>a</sup>

<sup>a</sup>Division of Soil and Water Management, Department of Earth and Environmental Sciences, KU Leuven, Kasteelpark Arenberg 20 bus 2459, 3001 Leuven, Belgium, <sup>b</sup>Institute for Agricultural and Fisheries Research, Plant Sciences Unit - Crop Husbandry and Environment, Burgemeester Van Gansberghelaan 109, 9820, Merelbeke, Belgium, <sup>c</sup>Soil Service of Belgium, Willem De Croylaan 48, 3001, Heverlee, Belgium, <sup>d</sup>Division of Crop Biotechnics, Department of Biosystems, KU Leuven, Willem de Croylaan 42, 3001 Heverlee, Belgium, <sup>e</sup>Rothamsted Research, Department of Sustainable and Grassland Systems, West Common, Harpenden AL5 2JQ, UK, <sup>f</sup>Walloon Center of Agricultural Research, Rue du Bordia 4, 5030 Gembloux, Belgium, <sup>g</sup>Université de Toulouse, INPT, INP-PURPAN, INRA, 24 chemin de Borde rouge, 31320, Castanet Tolosan, France, <sup>h</sup>ISPA, Bordeaux Sciences Agro, INRA, 71 avenue Edouard Bourlaux, CS 20032, 33882 Villenave d'Ornon CEDEX, France, <sup>i</sup>Faculty of Agricultural and Environmental Sciences, University of Rostock, Justus-von-Liebig-Weg 6, 18051, Rostock, Germany, <sup>j</sup>Department of Soil and Environment, Swedish University of Agricultural Sciences, Lennart Hjelm's väg 9, Uppsala, Sweden, <sup>k</sup>Division of Mechatronica Biostatistic and Sensor, Department of Biosystems, KU Leuven, Kasteelpark Arenberg 30, bus 2456, 3001 Leuven, Belgium, and <sup>l</sup>Faculty of Applied Economics, Department of Engineering Management, University of Antwerp, Prinsstraat 13, B-2000, Antwerpen, Belgium

## Summary

Most soil tests for available phosphorus (P) perform rather poorly in predicting crop response. This study was set up to compare different established soil tests in their capacity to predict crop response across contrasting types of soil. Soil samples from long-term field experiments, the oldest >100 years old, were collected in five European countries. The total number of soil samples ( $n = 218$ ), which differed in cropping and P treatment, and originated from 11 different soil types, were analysed with five tests: ammonium oxalate (Ox), ammonium lactate (AL), Olsen P, 0.01 M CaCl<sub>2</sub> and the diffusive gradient in thin film (DGT). The first three tests denote available P quantity (Q), whereas the last two indicate P intensity (I) of the soil solution. All five tests were positively related to the crop yield data ( $n = 317$ ). The Q-tests generally outperformed I-tests when evaluated with goodness of fit in Mitscherlich models, but critical P values of the I-tests varied the least among different types of soil. No test was clearly superior to the others, except for the oxalate extraction, which was generally poor. The combination of Q- and I-tests performed slightly better for predicting crop yield than any single soil P test. This Q + I analysis explains why recent successes with I-tests (e.g. DGT) were found for soils with larger P sorption than for those in the present study. This systematic evaluation of soil tests using a unique compilation of established field trials provides critical soil P values that are valid across Europe.

## Highlights

- We compared soil P tests for predicting crop response across contrasting soil types.
- No test was clearly superior to the others except for the oxalate extraction, which was generally poor.
- This study suggests that intensity tests do not perform markedly better than quantity tests.
- The evaluation of soil P tests on this unique dataset provided critical soil P values across Europe.

Correspondence: S. Nawara. E-mail: sophie.nawara@kuleuven.be

Received 16 January 2017; revised version accepted 29 August 2017

## Introduction

Phosphorus (P) is an essential element for agricultural production. In Europe, for many decades more P was added to agricultural soil than was taken up by the plants, leading to an accumulation of soil P (Sattari *et al.*, 2012). The increased environmental concerns and global P crisis in 2007–2008 (Cordell *et al.*, 2009) have renewed interest in more efficient use of P and have led to reduced P fertilizer applications on agricultural soil. Across the EU-28, the total average soil P balance is now approaching a steady state, and with negative balances in some regions there is a need to re-evaluate the quantification of soil available P because most soil P tests were calibrated at times when soil P stocks were increasing and residual P was still relatively young.

Soil available P refers to P present in the soil solution and P that is reversibly adsorbed on the solid phase or present in labile precipitates. The former is the directly available pool and its concentration is denoted as the P intensity (I) (Beckett & White, 1964). The latter is the pool that buffers the soil solution P and is denoted as the P quantity (Q). The combination of Q and I is termed the potentially available pool. Mechanistic models of nutrient uptake suggest that P uptake is generally more sensitive to I than to Q when evaluated in a parameter sensitivity analysis (Barber, 1984). This is because of the central role of P diffusion in P uptake. The Q–I concepts disregard the complexity of rhizosphere composition that alters the fate of P near the root surface and they do not account for the role of P mineralization in P bioavailability.

In Europe, more than 10 different methods are routinely used to quantify the plant available P in soil, all of them rooted in local, historical and scientific backgrounds (Neyroud & Lischer, 2003; Jordan-Meille *et al.*, 2012). None of these methods extract the total plant available P, but are used as an indicator of the extent of bioavailable P. Traditional soil P tests are based on chemical extractions (e.g. alkaline extraction with NaHCO<sub>3</sub> (Olsen) and acid extraction with ammonium oxalate) that release large fractions of soil P often associated with operational definitions, but generally denoting the potentially available P pool. Chemical soil extractions offer practical advantages, and they can evaluate the availability of P on an empirical basis for field trials (Jordan-Meille *et al.*, 2012). However, chemical extractions are often questioned because they do not mimic the entire process that controls plant nutrient bioavailability, and they extract large amounts of P that are not available for plant uptake because the pH and salt concentration of the extracts are not relevant to the root environment. As an alternative to chemical extractions, 'sink methods' (e.g. resin membranes and the diffusive gradient in thin film technique (DGT)) have been introduced to quantify plant available P based on conceptual grounds (i.e. imitating the root as a zero sink for P under P-deficient conditions). Sink methods are applicable to all types of soil and are conceptually more attractive than chemical extractions because they do not involve strong chemical alteration or change of soil pH (Van Raij, 1998).

All soil tests are based on a model for the plant nutrient uptake process and must be evaluated in terms of their predictive power. Pot

experiments with European soils showed that the Olsen P extraction was superior to the CaCl<sub>2</sub> extraction when examined for their ranges of critical P values (Sánchez-Alcalá *et al.*, 2015) or for their relation between dry matter production and soil tests (Humphreys *et al.*, 2001). Bomans *et al.* (2005) reported that no single chemical extraction method can be considered clearly superior to others in all circumstances for European soil types. The performance of DGT was examined on tropical soils in a pot trial by Six *et al.* (2013) and in a range of field trials on Australian soils by Mason *et al.* (2010) and Speirs *et al.* (2013). These studies concluded that on these strongly buffering tropical soils DGT was superior to other soil P tests for predicting plant growth. The DGT has not yet been compared with other Q- or I-tests on agricultural long-term field trials in Europe.

This study was set up to compare different established soil P tests systematically in their capacity to predict crop response to soil P on contrasting types of soil in Europe, and to identify the relative importance of Q and I for crop response. This study was based on long-term field experiments, each with their own historical background in which significant and large responses to P have been established for years or decades. The soil samples were taken from arable fields with various crops and soil properties. The plant available soil P was measured by five established soil P tests, including DGT, on all soil samples. The soil tests were first compared in terms of analytical precision and then evaluated for their capacity to explain crop yields across soils and crops.

## Materials and methods

### Soil samples

Soil samples were collected from different long-term field experiments in Europe (Tables 1, 2) based on the hypothesis that differences in growth response in the field were related only to differences in plant available P resulting from different P fertilization regimes. The fertilizer history of each experiment is described in detail in Appendix S1, Supporting Information. Some soil samples were taken freshly from sites and dried, whereas others have been archived for decades. Archiving might increase the fixation of P in the soil and affect the amount of P measured as plant available P. This will have the largest effect on the intensity tests.

A total of 218 soil samples from the plough layer were taken from 11 different sites in five different countries (the towns, countries and coordinates of the sites are given in Table 1). The 218 soil samples do not represent an equal number from each soil type, and they varied in treatment and number of sampling years per site (Table 1). For each soil sample, the corresponding yield data were provided. For some sites, yield data of the sampling year were supplemented with those of the adjacent years to include more crops; the assumption was that soil P test values vary little over adjacent years. This augmented the number of data points included in this study from 218 to 317. In total, six crops were included in the analysis: wheat (*Triticum aestivum* L.), barley (*Hordeum vulgare* L.), potato (*Solanum tuberosum* L.), sugar beet (*Beta vulgaris* L.), maize (*Zea mays* L.) and flax (*Linum usitatissimum* L.). Summer

**Table 1** Overview of the long-term P-response trials in Europe

Location	Trials <i>n</i> <sup>a</sup>	Soil texture <sup>b</sup>	Start	Number of rates of P including control	Sampling year <sup>c</sup>	Crop	Olsen P in P <sub>0</sub> / mg kg <sup>-1</sup>	Relative yield at P <sub>0</sub> / %	Soil samples <i>n</i>
Gembloux (Belgium) 50°34'57"N, 4°41'12"E	5	Loam (Luvisol with some Regosol)	1967	3	2015 2014 2013 2012 2011	Wheat Flax Barley Wheat Sugar beet	16 16 16 16 16	95 49 91 98 88	18
Ath (Belgium) 50°36'53"N, 3°46'3"E	8	Loam (Abp)	1965	5	1982 1985 1980 1981 1983 1984 1986 1987	Wheat Wheat Barley potato Barley Sugar beet Barley Sugar beet	27 27 27 27 27 27 27 27	95 96 90 58 95 93 84 98	9
Rostock (Germany) 54°3'41"N, 12°5'6"E	2	Loamy sand (Stagnic Cambisol)	1998	No compost: 2 + Compost: 2	2015	Maize	27 35	90 106	16
Carcarès Sainte Croix (France) 43°52'N, 0°44'W	1	Sand (Arenosol)	1972	6	2004	Maize	21	92	24
Pierroton (France) 44°44'N, 0°46'E	1	Sand (Podzol)	1995	5	2015	Maize	5.8	61	20
Toulouse (France) 43°31'48"N, 1°30'20"E	1	Loamy clay – clay (Luvisol)	1968	4	2006	(Durum) Wheat <sup>d</sup>	4.7	71	16
Peldon <sup>e</sup> (UK) 51°48'29"N, 0°52'43"E	1	Silty clay loam (Eutric Vertic Stagnosol)	2009	18	2011	Wheat	9.7	69	18
Great Carlton <sup>e</sup> (UK) 53°20'47"N, 0°05'56"E	1	Fine loam (Eutric Albic Luvic Stagnosol)	2009	18	2011	Wheat	9.1	81	18
Exhaustion Land (UK) <sup>f</sup> 51°48'46"N, -0°22'29"E	10	Sandy loam (Chromic Luvisol)	1856	4 rates of P on 5 plots with different NPK history	1999 2008	Wheat Wheat	4.2 4.8	15–69 27–69	2 × 20
Saxmundham (UK) <sup>g</sup> 52°13'19"N, 1°28'5"E	2	Sandy clay loam (Eutric Luvic Stagnosol)	1899	18	1978 1982	Wheat Wheat	5.4 4.6	47 63	2 × 18
Lanna (Sweden) 58°20'52"N, 13°7'35"E	1	Silty clay (Aquic Haplocryept)	1936	3	2013	Wheat	8.1	73	3
Total number									
11	33					6			218

A description of the experiments is given in the Supporting Information. Relative yield at P<sub>0</sub> is calculated as the ratio of yield at P<sub>0</sub> to that at the largest rate of P (P<sub>max</sub>). Olsen P for P<sub>0</sub> (measured in this study) is the mean value of all P<sub>0</sub> plots, unless otherwise stated in the Supporting Information.

<sup>a</sup>Number of crop–year combinations or plots with other treatments and fertilizer histories per location.

<sup>b</sup>Soil classification according to the World Reference Base for Soil Resources, except for the trial of Ath (Belgium), which has a soil classification of Abp according to the Belgian classification system, and the soil of Lanna (Sweden), which is classified as a Aquic Haplocryept according to the USDA soil taxonomy.

<sup>c</sup>Year of paired soil and crop data; years in italics refer to year of crop sampling for which soil samples were not available and replaced by soil samples of the closest year given in regular font.

<sup>d</sup>Durum wheat was analysed together with the other wheat data as one crop.

<sup>e</sup>Most of the 18 plots at this location were fertilized with P to a variable extent to obtain soil P values (Olsen) between 10 and 25 mg P kg<sup>-1</sup> soil. No treatment replicates exist at this location, and Olsen P in P<sub>0</sub> is that of the plot with the smallest soil P value.

<sup>f</sup>Four P rates (P<sub>0</sub>, P<sub>1</sub>, P<sub>2</sub> and P<sub>3</sub>) starting in 1986 on five different plots with different fertilization history; the P treatments were suspended in 1992. Starting from 2000, P was again applied yearly as maintenance dressing (i.e. equivalent to crop P offtake) to all +P treatments. The P<sub>0</sub> Olsen is the smallest P value among all treatments and the relative yields in the control are given as a range for the five plots.

<sup>g</sup>Sampling years differ in N fertilization: 1978 (120 kg N ha<sup>-1</sup>) and 1982 (160 kg N ha<sup>-1</sup>). Olsen P and relative yield are given for the plot with the smallest soil P value, because this location does not have treatment replicates.

**Table 2** Soil characteristics of the long-term P-response experiments in Europe

Location	Fe-Ox / mg kg <sup>-1</sup>	Al-Ox / mg kg <sup>-1</sup>	pH <sup>a</sup>	Soil texture	Clay / %	Sand / %	Silt / %	Gravel / %	OM / %	Reference
Gembloux <sup>b</sup>	2511	742	5.9	Loam	18	4	78	–	1.7 <sup>c</sup>	Personal communication
Ath	2239	515	6.5	Loam	–	–	–	–	–	–
Rostock	2008	637	5.8	Loamy sand	–	–	–	–	2.5	Requejo & Eichler-Löbermann, 2014
Carcarès Sainte Croix	1025	441	4.9	Sand	6	81	14	0	1.8	Pellerin <i>et al.</i> , 2000
Pierroton	270	547	4.7	Sand	5	93	2	0	2.2	Denoroy <i>et al.</i> , 2013; personal communication
Toulouse	1374	656	6.9	Loamy clay–clay	27	35	38	–	0.9	Colomb <i>et al.</i> , 2007
Peldon	3517	659	6.6	Silty clay loam	31	10	59	–	3.9	Knight <i>et al.</i> , 2014; personal communication
Great Carlton	2391	598	6.0	Fine loam	30	44	26	–	1.6	Knight <i>et al.</i> , 2014; personal communication
Exhaustion Land	3006	1038	7.1	Sandy loam	20	28	52	–	–	Personal communication
Saxmundham	1615	631	7.1	Sandy clay loam	25	50	25	–	–	Personal communication
Lanna	3718	1097	5.9	Silty clay	47	6	46	1.3	4.4	Bergström & Shirmohammadi, 1999

<sup>a</sup>Mean pH of soil samples measured in 0.01 M CaCl<sub>2</sub>; solid:liquid ratio = 1:10.

<sup>b</sup>Particle-size distribution determined on two plots out of the 54 from the experiment; an average value is reported.

<sup>c</sup>% C converted to % OM by multiplying by a factor of 1.6.

Oxalate extractable iron (Fe-Ox), oxalate extractable aluminium (Al-Ox) and pH (measured in 10<sup>-2</sup> M CaCl<sub>2</sub>) were calculated as the average value of all soil samples at a location (measured in this study). Other soil characteristics (particle-size distribution and organic carbon (OC)) were obtained from data reported earlier in the literature to illustrate the varying soil characteristics, but therefore do not correspond to the year of the soil sample analysed in this study.

and spring crops were not analysed separately; soft wheat and durum wheat (Toulouse, France) were both categorized as wheat.

The 317 data points were divided into 33 different trials. A trial is defined as a year–crop combination, a different NPK fertilizer history preceding the different P treatments and, at one location, different organic matter treatments in addition to the P treatments (Table 1). In all trials, crop growth was recorded at different rates of P denoted as: P0 for no P input, P1 for a P input equivalent to annual export, P2 for double P1, and so on. The largest rate of P application will be denoted as Pmax. Other macro- and micro-nutrients had been applied in sufficient amounts. In NPK factorial experiments, only rates of P at adequate rates of N and K were selected.

#### Methods of soil P testing

Five tests were used: phosphorus extraction with ammonium oxalate (Ox) (Schwertmann 1964), extraction with ammonium lactate and acetate at pH 3.75 (AL) (Egnér *et al.*, 1960), extraction with 0.5 M NaHCO<sub>3</sub> at pH 8.5 (Olsen) (Olsen *et al.*, 1954), extraction with 0.01 M CaCl<sub>2</sub> (CaCl<sub>2</sub>) (Houba *et al.*, 2000) and the diffusive gradient in thin film technique (DGT) (Zhang *et al.*, 1998; Degryse *et al.*, 2009). Abbreviations for the measured concentrations of P and testing methods are summarized in Table 3. More detailed information about the analytical procedures is provided in Appendix S1, Supporting Information. It is generally assumed that Ox, AL and Olsen measure the Q+I from the soil, which is the potentially available P pool (further denoted as Q-tests), whereas CaCl<sub>2</sub> and DGT are indices of the P intensity (I) of the soil. The

oxalate extraction was included as an extreme Q-test, whereas the other tests are used frequently in scientific studies or as official procedures to predict crop responses to soil P (Jordan-Meille *et al.*, 2012; Six *et al.*, 2013; Speirs *et al.*, 2013).

All soil P tests on the 218 soil samples of the field experiments were analysed in the laboratory of the first author. The tests were executed over several weeks and each batch of samples included two blank samples, a duplicate sample and an internal control sample (in duplicate) for quality control. A batch of samples was re-analysed when the results of both duplicates of the internal control sample deviated by more than 10% of the long-term average of that sample result. Prior to the entire study, the analytical precision of all soil P tests was evaluated with eight soil samples by a ‘round robin test’ in three laboratories, each using the same methodology but with shakers, centrifuges and ICP-OES instruments that varied. Preliminary experiments to optimize the Olsen extraction are described in Appendix S1, Supporting Information.

#### Data analysis

The yields at the largest rates of P (Pmax) varied among crop species, locations and years. To enable a uniform comparison between the different trials, crop yield data were standardized by converting them to a yield index, namely relative yield (RY, %), which is yield relative to the corresponding yield observed at Pmax for a given trial (Equation (1)). The latter is referred to hereafter as the maximal yield (100% RY) and was calculated as the average yield of the replicates of Pmax, unless specified otherwise in the

**Table 3** Summary of the soil testing methodologies

Soil P test								
Full name	Abbreviation	P measured	Extracting solution	Solid:liquid ratio / g:ml	Extraction time / minutes	Phase separation	P analysis	Reference
Ammonium oxalate extraction	Ox	P-Ox	0.2 M (COONH <sub>4</sub> ) <sub>2</sub> 0.14 M (COOH) <sub>2</sub> ± pH 3	1:20	120	2.5 µm filtration	ICP-OES	Schwertmann, 1964
Ammonium lactate and acetate extraction	AL	P-AL	0.1 M NH <sub>4</sub> -lactate 0.4 M acetic acid pH 3.75	1:20	240	Centrifugation <sup>a</sup>	ICP-OES	Egnér <i>et al.</i> , 1960; Goetelen <i>et al.</i> , 2011
Sodium hydrogen carbonate extraction	Olsen	P-Olsen	0.5 M NaHCO <sub>3</sub> pH 8.5	1:20	30	Centrifugation <sup>a</sup>	Colorimetry	Olsen <i>et al.</i> , 1954; Watanabe & Olsen, 1965
Calcium chloride extraction	CaCl <sub>2</sub>	P-CaCl <sub>2</sub>	0.01 M CaCl <sub>2</sub>	1:10	120	Centrifugation <sup>a</sup>	ICP-MS	Houba <i>et al.</i> , 2000
Diffusive gradient in thin film technique	DGT	P-DGT	Ferrihydrite-based binding gel	Saturation	2–48 hours	Diffusion through a 0.45-µm filter membrane	ICP-MS	Zhang <i>et al.</i> , 1998, Mason <i>et al.</i> , 2005

<sup>a</sup>10 minutes at 1830 g

description of the long-term field experiments in Appendix S1, Supporting Information:

$$\text{RY (\%)} = \frac{\text{yield}}{\text{maximal yield}} \times 100. \quad (1)$$

Data analysis was carried out with the statistical program JMP Pro 12.2.0 (SAS, Cary, NC, USA). The data included all treatments from P0 to Pmax. The RY was plotted against the soil available P determined for each of the five soil P tests (Ptest). A Mitscherlich model was fitted by nonlinear regression using the data (Equation (2)):

$$\text{RY (\%)} = b0 + (100 - b0) \times (1 - \exp(-b1 \times \text{Ptest})), \quad (2)$$

where  $b0$  is the RY at  $\text{Ptest}=0$  and  $b1$  is the slope of the model. The maximum yield (100%) is approached asymptotically with increasing P concentration; at  $\text{Ptest}=1/b1$ , the RY is  $b0 + 0.63 \times (100 - b0)$ . The underlying assumption in this model is that the yields at the largest rate of P in each trial (RY = 100, by definition) are at the plateau. The White test ( $P > 0.05$ ) (White, 1980) confirmed that the residuals from the regression model were homoscedastic for each soil P test. The analysis was first performed for the complete dataset with all crops (317 data points, 33 trials). The  $R^2$  of the fit was determined by plotting the predicted RY against the observed RY. Critical soil P values ( $P_{\text{crit}}$ ) (i.e. the soil P value at an RY of 95%; the latter is the arbitrarily defined cut-off indicating P deficiency in this study) were determined for each soil P test using an inverse prediction of Equation (2).

In addition, the effect of the qualitative variable crop on  $P_{\text{crit}}$  was studied by adding dummy variables (0 and 1) to the model. Assuming that the effect of the crop on the intercept ( $b0$ ) is negligible, the crop affects only the slope ( $b1$ ) of the model, which

can be coded with dummy variables (Equation (3)). The resulting model is:

$$\begin{aligned} \text{RY (\%)} = & b0 + (100 - b0) \times (1 - \exp(-b1_{\text{av}} \\ & \times (1 + b1_w \times d_w + b1_b \times d_b + b1_p \times d_p + b1_s \times d_s \\ & + b1_m \times d_m + b1_f \times d_f) \times \text{Ptest})), \end{aligned} \quad (3)$$

where  $b0$  is the intercept,  $b1_{\text{av}}$  is the average slope,  $b1_x$  is the slope specific for crop  $x$ ,  $d_x$  is the dummy variable specific for crop  $x$ ,  $w$  is wheat,  $b$  is barley,  $p$  is potato,  $s$  is sugar beet,  $m$  is maize and  $f$  is flax. This approach enabled the overall  $P_{\text{crit}}$  value per crop to be determined for each soil test, except for potato and flax, which were cultivated in one trial only. Moreover,  $P_{\text{crit}}$  was determined for each trial separately. This was successful for the trials with an average RY < 95% at P0 (26 trials out of 33). For seven trials with an RY ≥ 95% at P0, the Mitscherlich curves could not be fitted (slope not significant); therefore,  $P_{\text{crit}}$  was not detected and these trials were considered to be non-responsive because of a large extractable P concentration at P0.

For several purposes in the data analysis, coefficients of variation ( $\text{CV}_x$ : standard deviation/ $\bar{X}$ ) were determined. The variation of soil P values per test ( $\text{CV}_{\text{test}}$ ) indicates the distinctive potential of a test (i.e. the potential to distinguish between soil samples). This inherent variability is essential for the evaluation of the soil P tests.

#### Evaluation of the single soil P tests

The single soil P tests were evaluated by three parameters. Parameters 2 and 3 below, the uncertainties, were developed by the authors to correct for the inherent variability of the soil P tests ( $\text{CV}_{\text{test}}$ ):

- 1 The  $R^2$ -value of the relation between the observed and predicted RY.
- 2 The statistical uncertainty (SU) measures the uncertainty on the general  $P_{crit}$  value. Soil P tests with a large  $CV_{test}$  have larger confidence limits around  $P_{crit}$ . To correct for this inherent variability, the SU is calculated, which is defined as the relative width of the 95% confidence interval of  $P_{crit}$  divided by  $CV_{test}$ .
- 3 The crop-specific relative uncertainty (RU) is a measure of the variation in  $P_{crit}$  per crop ( $CV_{Pcrit}$ ) among different trials after fitting a Mitscherlich model for each individual field trial. This RU is defined as the ratio of  $CV_{Pcrit}$  and  $CV_{test}$ . This calculation was limited to the wheat data because this crop had the largest number of trials with an  $RY < 95\%$  at P0 for which a trial-specific critical P value could be determined ( $n = 16$ ). Trial-specific critical P values for other crops could be determined for a maximum of three trials per crop only.

## Results

### Soil properties

The 11 experiments are on substantially different types of soil in Europe; they cover a wide range of texture (sand–clay), pH, Fe-Ox, Al-Ox and Ca (Table 2). The pH ranges from 4.4 to 7.4 (measured in  $CaCl_2$ ), which corresponds to the 10th to 95th percentile of the pH values measured in the GEMAS dataset on 2108 European agricultural soils (Reimann *et al.*, 2014). The range of oxalate extractable Fe, Al and calcium carbonate ( $CaCO_3$ ) concentrations was compared with that of a subset of 500 samples of the GEMAS dataset (Janik *et al.*, 2015). Oxalate extractable Fe in our samples ranges from 150 to 3884  $mg\ kg^{-1}$ , which covers the 5th to 65th percentile of the European Fe-Ox data, and Al-Ox ranges between 192 and 1202  $mg\ kg^{-1}$ , which covers the 5th to the 50th percentile of the European Al-Ox data. The largest concentration of Fe-Ox + Al-Ox (on a molar basis) in the dataset studied occurred in Lanna (Sweden) and the smallest concentration in Pierroton (France) and Carcarès Sainte Croix (France). Data for  $CaCO_3$  concentration were available only for the heterogeneous soil of Toulouse, which is crossed by a calcareous band. Concentrations of  $CaCO_3$  in this field varied from 0 to 2% in the sampling year (2006), which applies to 65% of the soil samples in the European dataset. Therefore, the soil samples used in this study cover a wide range of soil properties, but samples with large P buffer capacities, such as soil with large Ca, Fe and Al concentrations, were somewhat under-represented.

The target soil P index described in the UK Fertiliser Manual (RB209) (Defra, 2010) is 16–25  $mg\ l^{-1}$  (measured as P-Olsen), which can be converted to 12–19  $mg\ kg^{-1}$  when an average bulk density of 1.3  $kg\ l^{-1}$  is assumed. The P-Olsen value at P0 was smallest ( $< 6\ mg\ P\ kg^{-1}$  soil) for the trials in Pierroton, Toulouse, Exhaustion Land and Saxmundham, followed by the trials in Lanna, Great Carlton and Peldon (8–10  $mg\ P\ kg^{-1}$  soil) (Table 1). These P values indicate a suboptimal amount of P because they were well below

**Table 4** The analytical precision of the soil tests in a ‘round robin test’ with eight samples in three laboratories that used similar methodologies

Soil P test	CV / %		
	Minimum	Maximum	Mean
Ox	2	20	9
AL	2	15	7
Olsen	7	28	14
$CaCl_2$	5	15	11
DGT	5	18	10

The variance is expressed as the coefficient of variation (CV) of the mean test result for each soil sample among the laboratories. Ox, ammonium oxalate; AL, ammonium lactate and acetate; DGT, diffusive gradient in thin film.

the target soil index. Their corresponding relative yields at P0 varied between 15 and 81%. The trials at Gembloux (16  $mg\ kg^{-1}$ ) and Carcarès Sainte Croix (21  $mg\ kg^{-1}$ ) have slightly larger concentrations of P at P0; these values were in the recommended range from the UK Fertiliser Manual. The trials of Rostock and Ath have adequate soil P concentrations at P0 according to this Fertiliser Manual. This is also reflected by the large relative yields at P0 ( $\geq 84\%$  except for potato).

### Soil P tests

The  $CV_{roundrobin}$  for the five soil P tests among the mean laboratory values reported was largest for Olsen (14%) and smallest for AL (7%) (Table 4).

Mean soil P values for each soil test across all samples are presented in Table 5. The mean ratios of extracted soil P values expressed relative to P-Olsen decreased in the order P-Ox 8.1 > P-AL 3.2 > P-Olsen 1 > P- $CaCl_2$  0.096. For DGT, no such ratio can be derived unequivocally because the quantity of soil from which the DGT samples P varies with soil properties and is a theoretical value. Assuming that for the European soils studied, mean P sampling depth during DGT analysis was 2 mm (Degryse *et al.*, 2009), and the average contact time was 24 hours, then the estimated mean P-DGT value of 3.8  $mg\ kg^{-1}$  is slightly larger than the mean P- $CaCl_2$  value of 3.0  $mg\ kg^{-1}$ , and the value is 0.12 relative to P-Olsen.

The soil P values per test vary widely among field experiments and treatments, and the relative range expressed by the  $CV_{test}$  is strikingly different among soil tests; some soil tests have more potential to produce different results for different soil samples. The  $CV_{test}$  decreased in the order of DGT >  $CaCl_2$  > Olsen > AL > Ox (Table 5, left), which is the reverse of the extracted quantities. This means that DGT gives the largest distinctive potential and P-Ox the smallest. The Pearson correlation coefficients among the soil tests were strongest between  $CaCl_2$  and DGT ( $r = 0.98$ , both I-tests), and between AL and Olsen ( $r = 0.96$ , both Q-tests) (Table S1, Supporting Information). The smallest coefficients were observed between the oxalate extraction and intensity tests (poor Q–I correlation).

**Table 5** Results of soil P tests from soils used in the data analysis (left) and the critical P values ( $P_{crit}$ ) defined as the soil P value associated with 95% relative yield, modelled with the Mitscherlich curve on data for all crops, locations and years (right)

Soil P test	Soil P test values in soil collection			Critical value $P_{crit}$ (at 95% relative yield)			
	Mean	Range	$CV_{test} / \% = A$	Mean $P_{crit}$	95% confidence interval		Statistical uncertainty (SU) (= B/A) / %
					Range	Width of interval relative to $P_{crit} / \% = B$	
Ox / mg kg <sup>-1</sup>	250	23–700	45	310	260–360	32	71
AL / mg kg <sup>-1</sup>	100	6.2–320	69	73	64–81	23	33
Olsen / mg kg <sup>-1</sup>	31	3.3–100	70	19	17–22	26	37
CaCl <sub>2</sub> / mg kg <sup>-1</sup>	3.0	0.4–19	120	1.7	1.5–1.9	24	20
DGT / µg l <sup>-1</sup>	200	3.6–1800	190	33	27–38	33	<u>17</u>

The relative range of soil P values is expressed by  $CV_{test}$ . The relative width of the 95% confidence interval of the critical value of a test (~uncertainty) depends on  $CV_{test}$  (~variability). The statistical uncertainty accounts for this inherent variability and is smallest for DGT (underlined). Ox, ammonium oxalate; AL, ammonium lactate and acetate; DGT, diffusive gradient in thin film.

### Relative yield as a function of soil P tests

The response of crop yield to each soil P test was modelled with the Mitscherlich equation (Figure 1). Relative yield increased with increasing P values. To represent the crop response better at the smaller values, the soil P test values are plotted on a log<sub>10</sub> scale (Figure S1, Supporting Information). The Olsen ( $R^2 = 0.49$ ) and AL ( $R^2 = 0.46$ ) extractions explained the crop response best. On a normal scale, CaCl<sub>2</sub> and DGT appear to fit adequately, but on the log<sub>10</sub> scale the data are more scattered at the smaller concentrations, which explains the smaller  $R^2$  values (Figure 1).

The mean critical soil P value ( $P_{crit}$ ) and the width of the 95% confidence interval relative to  $P_{crit}$  for each soil P test are listed in Table 5. This relative width was largest for DGT and smallest for Olsen. However, it can be argued that this relative width must be contrasted with the inherent variability of the soil P test. The statistical uncertainty (SU, see Evaluation) accounts for this and was smallest for DGT followed by CaCl<sub>2</sub> (Table 5, right).

The overall  $P_{crit}$  value per crop is given in Table 6, and  $P_{crit}$  values per trial are given in Table S2 in the Supporting Information. For some crops and soil P tests, there was a significant effect of crop ( $P < 0.05$ ) on the Mitscherlich slope and, therefore, on the  $P_{crit}$  values. In general, crop-specific  $P_{crit}$  values decreased in the following order: potato > flax > barley ~ sugar beet ~ wheat > maize. The overall crop-specific  $P_{crit}$  values should be treated with caution. For example, the overall crop-specific critical P-Olsen for maize was 18 mg kg<sup>-1</sup>, which is smaller than the mean  $P_{crit}$  of the individual trials (31, 30 and 19 mg kg<sup>-1</sup>; one additional set was unresponsive, Table S2, Supporting Information). This difference is explained by the statistical analysis. One trial only had a pronounced response at P0 and outweighs the Mitscherlich curve (Equation (3)), and therefore determined the overall crop-specific critical P value.

A robust soil P test can be defined as one for which  $P_{crit}$  varies the least among trials (i.e. among soil types and years of the trial). This variation is expressed by the relative uncertainty (RU, see Evaluation) (Table 7). The smallest variation in  $P_{crit}$  values among soils for the wheat data ( $n = 16$ ) was observed for P-CaCl<sub>2</sub>. Table S3 in the Supporting Information gives the RU values for other crops.

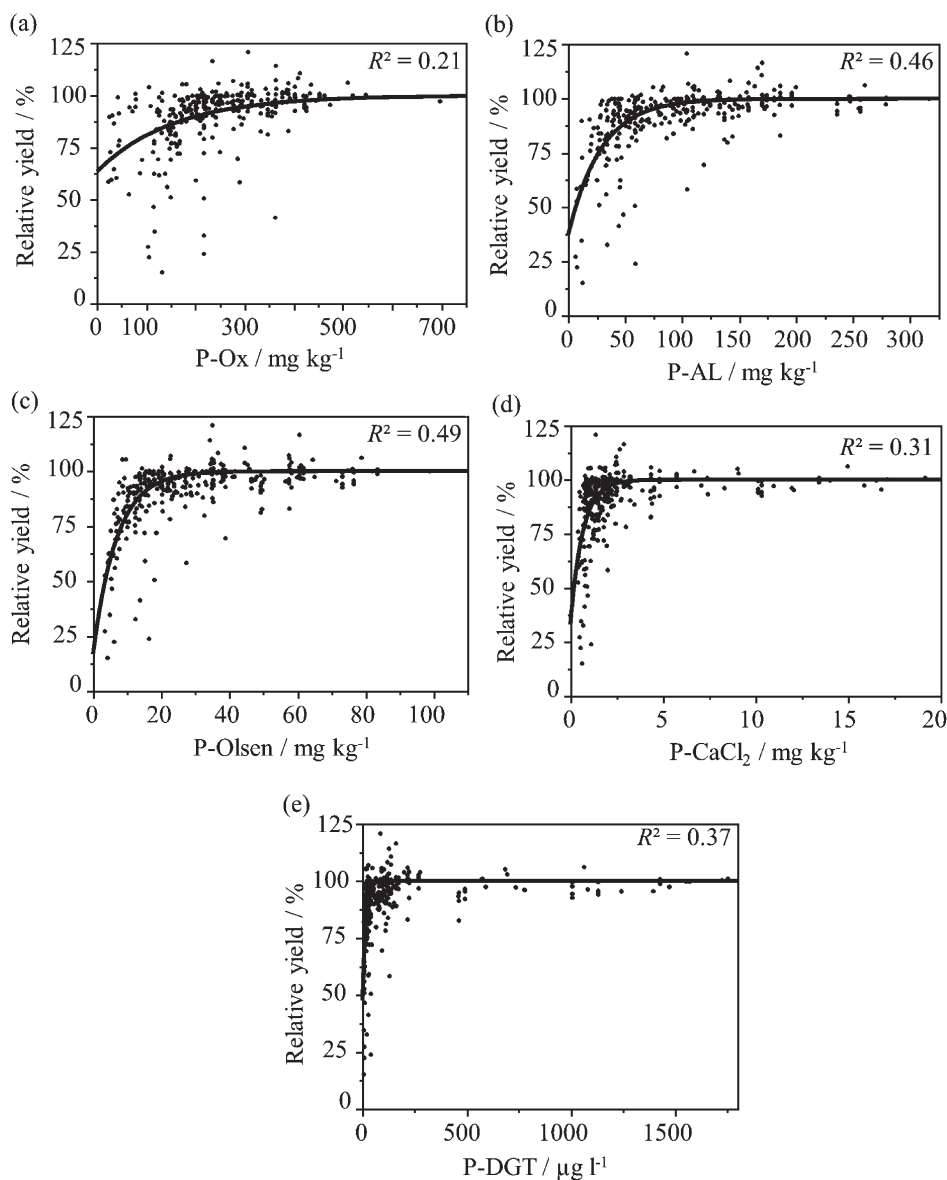
## Discussion

### The crop dependence of $P_{crit}$

Results from the long-term field experiments suggested that wheat was the least responsive crop to P, whereas potato and flax were the most responsive to P measured in the plough layer (Table 1). The latter is confirmed in this study by the  $P_{crit}$  values determined for each crop (Table 6). For all soil P tests, potato needs larger soil P values to obtain an RY of 95%. Potato plants have a shallower root system; therefore, their uptake efficiency of nutrients is small, especially at small concentrations of P (Dechassa *et al.*, 2003). This contrasts with deep rooting crops that can use P from deeper soil layers, which leads to a different sensitivity to topsoil P availability. In the present study, maize was the least responsive to P (smallest  $P_{crit}$  values). It should be noted that crop-specific  $P_{crit}$  values might not have much practical relevance for agricultural fields with crop rotation because an adequate value of P for the whole rotation needs to be maintained.

### Soil P tests

The largest amount of P was extracted by Ox, which is a only fraction of the total P in soil. Total soil P was not analysed, but other studies of mineral soils have shown that the ammonium oxalate test extracts more than half of total soil P: about 50% in Neyroud & Lischer (2003), about 70% in Stutter *et al.* (2015) and up to 90% in Koopmans *et al.* (2004). Previous isotope dilution studies have shown that the oxalate test extracts fractions of soil P that are not plant available (Six *et al.*, 2012) because of the dissolution of amorphous Fe and Al oxides. The AL and Olsen extractions are milder procedures and correlate strongly ( $r = 0.96$ ) with each other. This is remarkable given their contrasting pH values of extraction (acid versus alkaline) and with about three-fold more P-AL than P-Olsen extracted. Otabbong *et al.* (2009) showed that P-AL can be inter-converted to P-Olsen, and *vice versa*, when pH and clay content are included. Data here confirm that pH and texture class contribute significantly ( $P < 0.0001$ ) to the conversion of AL to Olsen data, but



**Figure 1** Relative yield as a function of soil P values measured by the different soil P tests: (a) P-Ox, (b) P-AL, (c) P-Olsen, (d) P-CaCl<sub>2</sub>, (e) P-DGT. The points are individual observations; lines are predicted with the Mitscherlich model. The same plots with  $x$ -axes on a  $\log_{10}$  scale are presented in the Supporting Information (Figure S1). Ox, ammonium oxalate; AL, ammonium lactate and acetate; DGT, diffusive gradient in thin film.

the proportion of variance explained (%) was small and a constant conversion factor of 3.2 can be advocated (details not shown).

More than 90% of soil solution P concentrations typically range from 0.02 to 0.05 mg P l<sup>-1</sup> (Barber, 1984). At a gravimetric water content of 30%, this is equivalent to 0.006–0.015 mg P kg<sup>-1</sup> soil only. The 0.01 M CaCl<sub>2</sub> extracts a larger fraction than this, but merely as the result of the larger liquid:solid ratio (101 kg<sup>-1</sup>) than in soil (0.31 kg<sup>-1</sup>) (i.e. the larger buffer power of P in the soil replenishes the liquid P). The DGT and CaCl<sub>2</sub> tests correlate remarkably strongly ( $r = 0.98$ ), confirming earlier analyses (Moody *et al.*, 2013). At large P concentrations, the absolute concentrations in the CaCl<sub>2</sub> soil extract were close to those measured by DGT. A

log–log regression showed that the slope of P-CaCl<sub>2</sub> (in mg P l<sup>-1</sup>) to P-DGT (in mg P l<sup>-1</sup>) was 0.89 (i.e. the data were near the 1:1 line when limited to data > 0.1 mg P l<sup>-1</sup> of P-DGT). However, at smaller P-DGT concentrations, the P-CaCl<sub>2</sub> data were up to 10 times as large as P-DGT. This might result from an incomplete desorption of soil P during the small exposure times between the DGT and the soil or, conversely, to an extraction of colloidal P in the CaCl<sub>2</sub> extract of soil with small P concentrations, overestimating true ionic PO<sub>4</sub>.

Comparison of the P values of I-based tests with those of Q-based tests indicates the importance of sorption for the mobility of P in soil, which is largely influenced by soil properties. For example,



**Table 6** Crop-dependent critical P values ( $P_{\text{crit}}$ ) and mean values for all crops per test among locations and years combined

Soil P test	Critical P value (at 95% relative yield)						
	Mean	Wheat <sup>a</sup>	Flax <sup>b</sup>	Potato <sup>b</sup>	Sugar beet <sup>a</sup>	Barley <sup>a</sup>	Maize <sup>a</sup>
Ox / mg kg <sup>-1</sup>	310	310	340	<b>400</b>	280	310	<b>120</b>
AL / mg kg <sup>-1</sup>	73	69	140	<b>200</b>	77	75	<b>46</b>
Olsen / mg kg <sup>-1</sup>	19	<b>18</b>	<b>40</b>	<b>76</b>	23	22	18
CaCl <sub>2</sub> / mg kg <sup>-1</sup>	1.7	1.7	2.3	<b>7.8</b>	1.2	1.2	<b>1.2</b>
DGT / µg l <sup>-1</sup>	33	<b>28</b>	<b>134</b>	<b>940</b>	44	38	32
N <sub>obs</sub>	317	176	18	4	28	31	60
N <sub>trial</sub>	33	23	1	1	2	3	3

<sup>a</sup>Significant differences were determined from the 95% confidence interval of the crop-specific dummy variables.

<sup>b</sup>Critical values determined without the crop-specific dummy variables. The values are the trial-specific critical soil P values because both flax and potato were cultivated in one trial only (see Table S2, Supporting Information). Significant differences of  $P_{\text{crit}}$  between potato or flax and the mean response were determined from the 95% confidence interval of both critical values.

Values in bold denote that the crop has a significantly different critical value compared with the mean response.

N<sub>obs</sub>, number of individual observations included per crop; N<sub>trial</sub>, number of trials included per crop. Ox, ammonium oxalate; AL, ammonium lactate and acetate; DGT, diffusive gradient in thin film.

**Table 7** Critical soil P values for wheat yield from 16 trials (different locations, fertilizer histories or years) and coefficient of variation (CV) among soils sorted by soil test

	Ox / mg kg <sup>-1</sup>	AL / mg kg <sup>-1</sup>	Olsen / mg kg <sup>-1</sup>	CaCl <sub>2</sub> / mg kg <sup>-1</sup>	DGT / µg l <sup>-1</sup>
$P_{\text{crit}}$ , range	150–290	27–120	8.3–46	0.8–2.7	6.1–110
$P_{\text{crit}}$ , CV / % (= C)	20	45	56	44	106
Soil tests, CV (Table 5) / % (= A)	45	69	70	120	190
$P_{\text{crit}}$ , relative uncertainty (= C/A × 100)	45	65	80	<u>37</u>	56

The relative uncertainty is calculated per soil P test as the ratio of CV  $P_{\text{crit}}$  for each soil test to the general CV of the test values (Table 5). The smallest value (most consistent value) is underlined. Detailed critical soil P values per trial are given in Table S2 in the Supporting Information. Results for crops other than wheat are listed in Table S3 in the Supporting Information because three or fewer trials only were included in the analysis. Ox, ammonium oxalate; AL, ammonium lactate and acetate; DGT, diffusive gradient in thin film.

the ratio of P-CaCl<sub>2</sub> (so-called mobile P) to P-Ox (closest to total P) indicates the fraction of (near) total P that is soluble. This fraction decreases by a factor of about 18 when the concentration of (Fe + Al)<sub>ox</sub> (on a molar basis) increases by a factor of about 10. This fraction (mobile P to total P) is largest in soil with a sandy texture. Similar, but weaker, trends occur for the ratio of P-Olsen to P-Ox, and the trends are even weaker when P-AL to P-Ox is evaluated (details not shown).

#### Single soil P tests indicating crop responses

The goal of this study was to compare the five soil tests for predicting crop response to soil P. This was evaluated by several criteria, excluding the practical aspects discussed below. First, the graphical analysis suggests that the oxalate extraction only ranks poorly, whereas all other soil tests show an association with crop response; no soil P test, however, stands out clearly as the best (Figures 1 and S1). This is confirmed by the goodness of fit of the Mitscherlich model; the largest  $R^2$  values are for the Olsen and AL extractions. Second, the statistical uncertainty of the 95% confidence interval of critical soil P values derived from the Mitscherlich curve was

smallest, therefore, most optimal for DGT and weakest for Ox (Table 5). Despite the variability in the samples, the 95% confidence interval was rather small for any soil P test (< 40% of the mean) because of the large number of data in the analysis. Moreover, the statistical uncertainties are not largely different among the tests (< factor 2, except for Ox). This means that this criterion is not relevant for the evaluation of the soil P tests because it has a similar value for all tests. In contrast, the third criterion for the evaluation of the soil P tests, the variability of  $P_{\text{crit}}$  per crop among different trials, is more informative (Table 7). Restricting the discussion to the wheat data only, the Olsen extraction appears to perform the worst. The  $P_{\text{crit}}$  values vary between 8 and 46 mg P kg<sup>-1</sup> depending on year and soil, and this range covers almost the entire spectrum of P-Olsen values in soils from this study. The CaCl<sub>2</sub> extraction is the most robust soil test in this respect because  $P_{\text{crit}}$  values vary within a factor of three and this is a small fraction of the spectrum of possible values (Table 7). Therefore, this soil test discriminates the deficient from adequate soil P fertility most consistently between years and soils. The same calculations were performed for other crops (Table S3), which might be of importance for deep rooting crops; they were not taken into account because a few trials only

were included. In general,  $P_{crit}$  depended more on soil location than on crop (data not shown). Finally, two additional criteria were analysed (Table S4, Supporting Information): the rate of success in predicting responsiveness correctly and the rate of failure to predict lack of P responsiveness. Differences between soil P tests were marginal, but the Q-tests performed best.

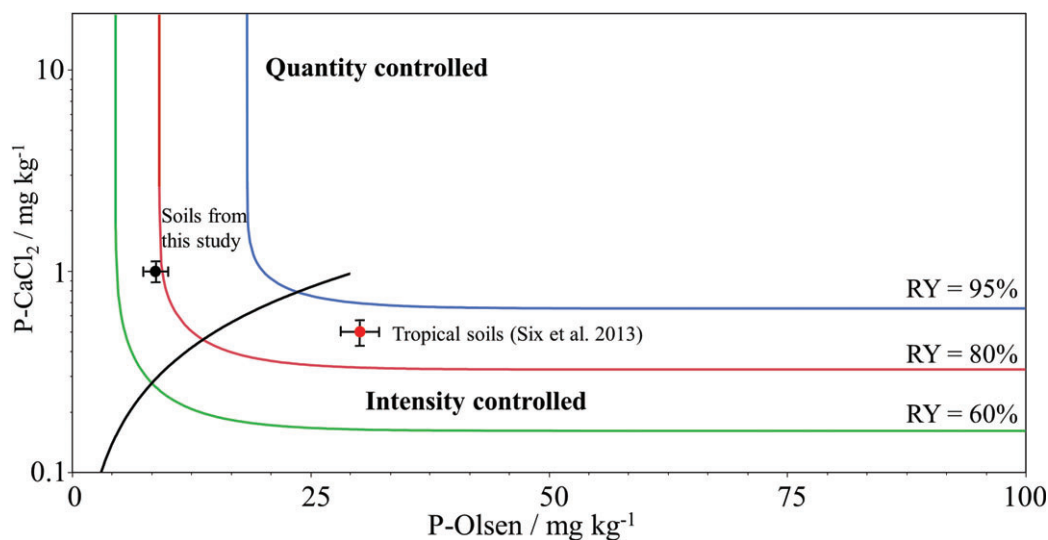
Overall, the oxalate extraction performed most poorly of all tests, which may be related to the lack of P mobility that affects bioavailability. Among the other four tests, the evaluation for predicting crop response does not provide a clear result. The Q-tests, AL and Olsen, are better than the I-tests in criterion 1 (fitting responsiveness), whereas the I-tests are better in criteria 2 and 3 (uncertainties surrounding the thresholds). It is important to note that some of the soils analysed in this study were from archived samples, and soils that strongly sorb P with large Fe, Al and Ca concentrations were under-represented. This might have influenced the results. Overall, none of the single tests evaluated is clearly superior to another. This is in contrast to recent studies on strongly P buffering tropical and Australian soils for which I-tests such as DGT outperformed traditional chemical extractions (Mason *et al.*, 2010; Six *et al.*, 2013; Speirs *et al.*, 2013). This suggests that I-tests become increasingly important for soil that sorbs P more strongly than the soils in this study. This can be predicted by soil mechanistic modelling (e.g. the Barber–Cushman model; Barber, 1984), which also shows that the combination of I and the P buffer capacity (PBC) explains P bioavailability most sensitively. Predictions of single soil test extractions have been improved previously by a combined model with Q- and I-tests (Van Rotterdam *et al.*, 2012) or by combining I or Q with the PBC (Shirvani *et al.*, 2005). The former is examined in the next section.

#### Combination of soil P tests

A Mitscherlich model can be proposed that uses both Q- and I-tests i.e. a non-linear response surface model based on two contrasting soil tests (Equation (4)):

$$RY (\%) = b_0 + (100 - b_0) \times (1 - \exp(-b_1 \times P_{test_1})) \times (1 - \exp(-b_2 \times P_{test_2})), \quad (4)$$

in which  $b_0$  is the intercept of the model, and  $b_1$  and  $b_2$  are parameters corresponding to  $P_{test_Q}$  and  $P_{test_I}$ , respectively. This model was fitted successfully on the Olsen and the  $CaCl_2$  extractions as indices for Q and I, respectively. Other Q–I combinations were tested, but the best fitting model was obtained for the combination discussed. Significant values of  $b_1$  and  $b_2$  were obtained (based on their 95% confidence intervals) and  $R^2$  of the combined model (0.50) is marginally larger than that of the P-Olsen (0.49) or P- $CaCl_2$  models (0.31). A response surface model is shown in 2-D in Figure 2, with the contour lines of predictions and with RY of 95, 80 and 60%. The black line represents a theoretical ratio of 30 for the P-Olsen to P- $CaCl_2$  ratio. Figure 2 illustrates that two markedly different zones affect RY. In the upper left part, a yield increase is possible only if P-Olsen increases; the yield is controlled by the quantity of P in the soil. Conversely, in the lower right part of the figure yield is controlled by I. A minor transition zone can be distinguished near the P-Olsen/P- $CaCl_2$  ratio of 30, where yield is controlled by both I and Q. Most soil P values of P-deficient soils (RY < 80%) from this study are in the Q-controlled part with a P-Olsen/P- $CaCl_2$  ratio  $\leq 20$ . The largest ratio in the soils of this



**Figure 2** Results of response-surface analysis of the relative yield data. The Mitscherlich response surface (Equation (4)) of the relative yield (RY) was fitted to a combination of results from two soil tests (Olsen and  $CaCl_2$  extraction) to give a model with  $R^2 = 0.50$ . Data are not shown and contour lines of relative yields are given as a function of two different soil P values. The black line represents a theoretical ratio of P-Olsen to P- $CaCl_2$  of 30. Below that line, P is more strongly sorbed and the crop response is controlled fully by P intensity (P- $CaCl_2$ ). The black circle is the average (error bars denote standard error) of the mean values of European soil samples from this study with RY < 80%. The red circle is the corresponding value of tropical soils (Six *et al.*, 2013).

study was for the loamy soil of Gembloux (Belgium), whereas the smallest was for the sandy soils of Carcarès Sainte Croix (France) (P-Olsen/P-CaCl<sub>2</sub> = 7). In contrast, that ratio in the tropical soils (RY < 80%) studied by Six *et al.* (2013) was about 60, which is well above the threshold of 30 and is in the I-controlled part of the graph. In the study by Six *et al.* (2013), DGT was markedly superior to Q-tests. This difference can be explained by soil properties: not only the smaller soil P values but also larger Al concentration, smaller organic matter content and lower pH. In summary, our analysis suggests that P bioavailability in European soils is affected more by the quantity than the intensity of P, which justifies the continued use of the Olsen and AL extractions as standard tests. The same conclusion was made when a combined model of P-AL and P-DGT was analysed.

### Practical aspects

Intensity tests are less popular in routine analysis than Q-tests because of smaller concentrations in the extracts, leading to detection limit issues (Van Raij, 1998). With detection limits of current ICP-OES near 0.01 mg P l<sup>-1</sup>, the I-tests are now within reach of commercial laboratories provided that procedural blank concentrations stay small. Despite the fact that the DGT technique is rather new, the consistency in inter-laboratory measurements was large (mean CV<sub>roundrobin</sub> = 10%, Table 4), which is better than for Olsen (mean CV<sub>roundrobin</sub> = 14%) but worse than for AL (mean CV<sub>roundrobin</sub> = 7%). Kleinman *et al.* (2001) observed a larger CV (22%) for the Olsen extraction than the CV<sub>roundrobin</sub> in this study when 24 samples were analysed by nine laboratories. If the Olsen test is applied to soil with more organic matter the results are often inconsistent because of background colour of the alkaline soil extract, which requires additional steps in the analyses. This analytical factor, together with the more labour-intensive colorimetric procedure than for low-cost ICP-OES, makes the Olsen test less desirable than, for example, the AL test, which predicted crop responses similarly. A Q–I combination only marginally improved the goodness of fit, suggesting that there is little added value in using two soil tests.

### Conclusion

Our study has shown that no single soil test used in commercial laboratories can provide a single critical value across field trials for P (i.e. there is considerable residual uncertainty). Among the established tests, none consistently outperformed the others, except that the oxalate extraction was inferior to all others. This accords with earlier studies on European soils. Overall, this study suggests that the I-tests, including DGT, do not perform markedly better than Q-tests because the retention of P in European soils is not as large as in strongly P-deficient and highly weathered soils. This unique database merged, for the first time, several key long-term trials on P in Europe, and the systematic evaluation of soil tests on this dataset provides critical values of soil P that are valid across Europe.

### Supporting Information

The following supporting information is available in the online version of this article:

#### Appendix S1.

Description of the long-term field trials.

Methods of soil P testing.

Optimization of the Olsen extraction: preliminary experiments.

Evaluation of the single soil P tests.

**Table S1.** Pearson correlation coefficients for all soil P tests.

**Table S2.** Critical P values sorted per crop for all trials.

**Table S3.** The range of critical soil P values observed in different trials (different locations or years) and its coefficient of variation (CV) among soils sorted by crop and soil tests.

**Table S4.** Success rate of predicting P fertilizer responsiveness correctly in soils with soil P test values below the critical values, and in contrast overall rate of failure to predict lack of P responsiveness in soils with soil P values above (or equal to) the critical values.

**Figure S1.** Relative yield as a function of soil P values measured by the five soil P tests on a log<sub>10</sub> scale.

**Figure S2.** Schematic representation of criteria used to evaluate adequacy of soil P tests to predict available P in soil.

### Acknowledgements

The authors thank the Flemish Land Agency for financial support (project APLM/2014/3). They also thank Charlotte Vermeiren, Joeri Plevoets and Ruben Warrinnier for their help in the laboratory, and Kristin Coorevits and Liske Versieren for the ICP measurements. Rothamsted Long Term Experiments are supported by the UK Biotechnology and Biological Sciences Research Council and the Lawes Agricultural Trust under the National Capabilities Programme grant and grant BBS/EC/00005197.

### References

- Barber, S.A. 1984. *Soil Nutrient Bioavailability: A Mechanistic Approach*. John Wiley & Sons, New York.
- Beckett, P.H.T. & White, R.E. 1964. Studies on the phosphate potentials of soils. Part III: the pool of labile inorganic phosphate. *Plant and Soil*, **21**, 253–282.
- Bergström, L.F. & Shirmohammadi, A. 1999. Areal extent of preferential flow with profile depth in sand and clay monoliths. *Journal of Soil Contamination*, **8**, 637–651.
- Bomans, E., Fransen, K., Gobin, A., Mertens, J., Michiels, P., Vandendriessche, H. *et al.* 2005. *Addressing Phosphorus Related Problems in Farm Practice. Final Report to the European Commission*. Soil Service of Belgium, Leuven. [WWW document]. URL <http://ec.europa.eu/environment/natres/pdf/phosphorus/AgriPhosphorusReport%20final.pdf> [accessed on 13 June 2017].
- Colomb, C., Debaeke, P., Jouany, C. & Nolot, J.M. 2007. Phosphorus management in low input stockless cropping systems: crop and soil responses to contrasting P regimes in a 36-years experiment in southern France. *European Journal of Agronomy*, **26**, 154–165.
- Cordell, D., Drangert, J.-O. & White, S. 2009. The story of phosphorus: global food security and food for thought. *Global Environmental Change*, **19**, 292–305.

- Dechassa, N., Schenk, M.K., Claassen, N. & Steingrobe, B. 2003. Phosphorus efficiency of cabbage (*Brassica oleracea* L. var. capitata), carrot (*Daucus carota* L.), and potato (*Solanum tuberosum* L.) *Plant and Soil*, **250**, 215–224.
- Degryse, F., Smolders, E., Zhang, H. & Davison, W. 2009. Predicting availability of mineral elements to plants with the DGT technique: a review of experimental data and interpretation by modelling. *Environmental Chemistry*, **6**, 198–218.
- Denoroy, P., Mollier, A., Niollet, S., Gire, C., Barbot, C., Plénet, D. et al. 2013. Définir l'optimum agro-environnemental de fertilisation phosphatée grâce à un essai de longue durée (Defining the agro-environmental optimum of phosphate fertilisation through a long-term experiment). *Les 11èmes rencontres de la fertilisation raisonnée et de l'analyse du Comifer et du GEMAS se sont déroulées les, 20&21 November 2013, au Palais des congrès de Poitiers-Futuroscope*. [WWW document]. URL [http://www.comifer.asso.fr/images/pdf/11emes\\_rencontres/Interventions/Posters/13%20-%20Pascal%20DENOROY/Poster%20Pascal%20DENOROY.pdf](http://www.comifer.asso.fr/images/pdf/11emes_rencontres/Interventions/Posters/13%20-%20Pascal%20DENOROY/Poster%20Pascal%20DENOROY.pdf) [accessed on 12 July 2017].
- Department for Environment Food and Rural Affairs (Defra) 2010. *Fertiliser Manual (RB209)*. The Stationery Office, London.
- Egnér, H., Riehm, H. & Domingo, W.R. 1960. Untersuchungen über die chemische Bodenanalyse als Grundlage für die Beurteilung des Nährstoffzustandes der Böden. II. Chemische Extraktionsmethoden zur Phosphor und Kaliumbestimmung (Studies on chemical soil analysis as a basis for assessing the nutrient status of soils. II. Chemical extraction methods for the determination of phosphorus and potassium). *Kungliga Lantbrukshögskolans*, **26**, 199–215.
- Goetelen, L., Groep, A.N., Vanhoof, C. & Tirez, K. 2011. *Eindrapport: validatie meetmethodiek voor bepaling van plantbeschikbare P in ammoniumlactaat. (Final Report: Validation Methodology for the Determination of Plant Available P in Ammonium Lactate)*. Flemish Institute for Technological Research, Mol. [WWW document]. URL <https://esites.vito.be/sites/reflabos/onderzoeksrapporten/Online%20documenten/2011%20Rapport%20plantbeschikbaar%20P-2.pdf> [accessed on 13 June 2017].
- Houba, V.J.G., Temminghoff, E.J.M., Gaikhorst, G.A. & Van Vark, W. 2000. Soil analysis procedures using 0.01 M calcium chloride as extraction reagent. *Communications in Soil Science and Plant Analysis*, **31**, 1299–1396.
- Humphreys, J., Tunney, H. & Duggan, P. 2001. Comparison of extractable soil phosphorus with dry matter production and phosphorus uptake by perennial ryegrass in a pot experiment. *Irish Journal of Agricultural and Food Research*, **40**, 45–54.
- Janik, L.J., Forrester, S.T., Soriano-Disla, J.M., Kirby, J.K., McLaughlin, M.J. & Reimann, C. 2015. GEMAS: prediction of solid-solution phase partitioning coefficients (K<sub>d</sub>) for cationic metals in soils using mid-infrared diffuse reflectance spectroscopy. *Environmental Toxicology and Chemistry*, **34**, 224–234.
- Jordan-Meille, L., Rubæk, G.H., Ehlert, P.A.I., Genot, V., Hofman, G., Goulding, K. et al. 2012. An overview of fertilizer-P recommendations in Europe: soil testing, calibration and fertilizer recommendations. *Soil Use and Management*, **28**, 419–435.
- Kleinman, P.J.A., Sharpley, A.N., Gartley, K., Jarrell, W.M., Kuo, S., Menon, R.G. et al. 2001. Interlaboratory comparison of soil phosphorus extracted by various soil test methods. *Communications in Soil Science and Plant Analysis*, **32**, 2325–2345.
- Knight, S., Morris, N., Goulding, K., Johnston, J., Poulton, P. & Philpott, H. 2014. *Identification of Critical Soil Phosphate ( P ) Levels for Cereal and Oilseed Rape Crops on a Range of Soil Types*. Project Report No 529, Cereals and Oilseeds Division of the Agriculture and Horticulture Development Board (HGCA), Kenilworth. [WWW document]. URL <https://cereals.ahdb.org.uk/media/338174/pr529-final-project-report.pdf> [accessed on 13 June 2017].
- Koopmans, G.F., Chardon, W.J., Ehlert, P.A.I., Dolfig, J., Suurs, R.A.A., Oenema, O. et al. 2004. Phosphorus availability for plant uptake in a phosphorus-enriched noncalcareous sandy soil. *Journal of Environmental Quality*, **33**, 965–975.
- Mason, S., Hamon, R., Nolan, A., Zhang, H. & Davison, W. 2005. Performance of a mixed binding layer for measuring anions and cations in a single assay using the diffusive gradients in thin films technique. *Analytical Chemistry*, **77**, 6339–6346.
- Mason, S., McNeill, A., McLaughlin, M.J. & Zhang, H. 2010. Prediction of wheat response to an application of phosphorus under field conditions using diffusive gradients in thin-films (DGT) and extraction methods. *Plant and Soil*, **337**, 243–258.
- Moody, P.W., Speirs, S.D., Scott, B.J. & Mason, S.D. 2013. Soil phosphorus tests I: What soil phosphorus pools and processes do they measure? *Crop and Pasture Science*, **64**, 461–468.
- Neyroud, J.-A. & Lischer, P. 2003. Do different methods used to estimate soil phosphorus availability across Europe give comparable results? *Journal of Plant Nutrition and Soil Science*, **166**, 422–431.
- Olsen, S.R., Cole, C.V., Watanabe, F.S. & Dean, L.A. 1954. Estimation of available phosphorus in soils by extraction with sodium bicarbonate. *US Department of Agriculture Circular*, **939**, 1–19.
- Ottobong, E., Börling, K., Kätterer, T. & Mattsson, L. 2009. Compatibility of the ammonium lactate (AL) and sodium bicarbonate (Olsen) methods for determining available phosphorus in Swedish soils. *Acta Agriculturae Scandinavica Section B—Soil Plant Science*, **59**, 73–378.
- Pellerin, S., Mollier, A. & Plénet, D. 2000. Phosphorus deficiency affects the rate of emergence and number of maize adventitious nodal roots. *Agronomy Journal*, **92**, 690–697.
- Reimann, C., Birke, M., Demetriades, A., Filzmoser, P. & O'Connor, P. (eds) 2014. *Chemistry of Europe's Agricultural Soils. Part A: Methodology and Interpretation of the GEMAS Data Set*. Schweizerbart Science Publishers, Stuttgart.
- Requejo, M.I. & Eichler-Löbermann, B. 2014. Organic and inorganic phosphorus forms in soil as affected by long-term application of organic amendments. *Nutrient Cycling in Agroecosystems*, **100**, 245–255.
- Sánchez-Alcalá, I., del Campillo, M.C. & Torrent, J. 2015. Critical Olsen P and CaCl<sub>2</sub>-P levels as related to soil properties: results from micropot experiments. *Soil Use and Management*, **31**, 233–240.
- Sattari, S.Z., Bouwman, A.F., Giller, K.E. & Van Ittersum, M.K. 2012. Residual soil phosphorus as the missing piece in the global phosphorus crisis puzzle. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 6348–6353.
- Schwertmann, U. 1964. Differenzierung der Eisenoxide des Bodens durch Extraktion mit Ammoniumoxalatlösung (The differentiation of iron oxides in soils by extraction with ammonium oxalate solution). *Zeitschrift für Pflanzenernährung und Bodenkunde*, **105**, 194–202.
- Shirvani, M., Shariatmadari, H. & Kalbasi, M. 2005. Phosphorus buffering capacity indices as related to soil properties and plant uptake. *Journal of Plant Nutrition*, **28**, 537–550.
- Six, L., Pypers, P., Degryse, F., Smolders, E. & Merckx, R. 2012. The performance of DGT versus conventional soil phosphorus tests in tropical soils—An isotope dilution study. *Plant and Soil*, **359**, 267–279.
- Six, L., Smolders, E. & Merckx, R. 2013. The performance of DGT versus conventional soil phosphorus tests in tropical soils—maize and rice responses to P application. *Plant and Soil*, **366**, 49–66.

- Speirs, S.D., Scott, B.J., Moody, P.W. & Mason, S.D. 2013. Soil phosphorus tests II: a comparison of soil test-crop response relationships for different soil tests and wheat. *Crop and Pasture Science*, **64**, 469–479.
- Stutter, M.I., Shand, C.A., George, T.S., Blackwell, M.S.A., Dixon, L., Bol, R. *et al.* 2015. Land use and soil factors affecting accumulation of phosphorus species in temperate soils. *Geoderma*, **257–258**, 29–39.
- Van Raij, B. 1998. Bioavailable tests: alternatives to standard soil extractions. *Communications in Soil Science and Plant Analysis*, **29**, 1553–1570.
- Van Rotterdam, A.M.D., Bussink, D.W., Temminghoff, E.J.M. & Van Riemsdijk, W.J. 2012. Predicting the potential of soils to supply phosphorus by integrating soil chemical processes and standard soil tests. *Geoderma*, **189–190**, 617–626.
- Watanabe, F. & Olsen, S.R. 1965. Test of an ascorbic acid method for determining P in water and NaHCO<sub>3</sub> extracts from soil. *Soil Science Society of America Proceedings*, **29**, 677–678.
- White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.
- Zhang, H., Davison, W., Gadi, R. & Kobayahi, T. 1998. In situ measurement of dissolved phosphorus in natural waters using DGT. *Analytica Chimica Acta*, **370**, 29–38.