SOIL HETEROGENEITY AND SOIL TESTING

TH. J. FERRARI

Agricultural Experiment Station and Institute for Soil Research T.N.O., Groningen

and

F. H. B. VERMEULEN

Laboratory for Soil and Crop Testing, Oosterbeek

SUMMARY

(a) The information obtained from soil testing possesses a relative value, partly because the result of analysis is no more than an estimate of the condition of the soil. The degree of reliability of the examination is closely connected with the magnitude of the error of analysis and of the sampling error. For the rest, the result only represents the state of affairs at one particular moment.

(b) The error of analysis at the Laboratory for Soil and Crop Testing is, generally speaking, small, and does not play a great part compared with other errors (in the broadest sense of the word). An error of 4% can be expected (duplicate determination).

(c) In various respects the fields are not homogeneous. The investigator is faced by the fact that only one sample can be taken from a field, which sample has to be representative of the whole field. The standard deviation, influenced by analysis and sampling, averages 10% in the case of the phosphate, potassium and magnesium determinations, and 5% in the case of the humus and clay determinations. The error in the pH value is approximately 0.1 per unit. There is little difference between grassland and arable land, or between the various types of soil, in these respects. The size of the field has, on the average, no effect on the magnitude of the sampling error.

(d) Owing to the great influence of the sampling error on the accuracy of soil testing, sampling must be carried out with the utmost care; sampling procedure has to satisfy standard requirements (Cline, 1944).

(e) If it is desired to compare the results of two or more samplings of the same field, taken with a certain interval of time between them, more or less great changes in the situation, under the influence of climate, farming operations, etc., will have to be taken into consideration.

(f) In spite of all measures that may be applied, results will always possess a certain degree of unreliability. Accordingly, the agricultural adviser should never regard the figures produced by soil testing as being absolutely infallible.

(g) Advice to the effect that examination of the soil of a field should be repeated, for instance every five years, acquires more weight when considered in connection with this relativity of the value of the results. The longer a field is under observation, the greater will be the significance and value of soil testing. In this way, an insight will be gained into the reliability of soil testing in general via the results obtained in the particular field concerned.

INTRODUCTION

Every year more than 100,000 soil samples are sent by Dutch farmers to the Laboratory for Soil and Crop Testing at Oosterbeek, in which about 1,400,000 analyses are carried out annually. Some of the samples submitted are from fields the soil of which has been analyzed before. In certain cases

1) Received for publication July 23, 1955.
the differences between the earlier analyses and the later ones are so great that they cannot be explained by the manuring treatment applied. And these differences have sometimes shaken confidence in soil testing as a science — a reaction which has also been experienced in other countries, notably Scotland (Robertson and Simpson, 1954).

This prompted the Agricultural Experiment Station and Institute for Soil Research, and the Laboratory for Soil and Crop Testing, to undertake an investigation into the influence of soil heterogeneity on the accuracy of the results of soil testing. Previous investigations in this sphere in the Netherlands (and also in other countries) have not produced enough data to enable a correct idea to be obtained concerning this. The investigations have generally been carried out in a small number of fields, and the area sampled there has been small by comparison with the average size of Dutch fields (Visser, 1937; Bruin, 1938; Harmsen, 1939–1940; Schuffelen, Hudig and Wtewaall, 1944–1945; Castenmiller, 1950; Boskma and Sluijsmans, 1953).

**Accuracy of soil testing**

General procedure in carrying out a soil test is as follows. The field concerned is sampled by means of a number of borings. The bulk sample is then examined for certain properties in the laboratory. There are various reasons why the result fails exactly to represent the actual condition of things. They are connected partly with sampling, partly with analysis (in the broadest sense of the word).

The fact that no soil sample is a perfect reflection of the field sampled inevitably implies that the estimate of the condition of the field is subject to a sampling error of $s_F$. In reality this error is greater, because sampling may be performed in different periods in the year, and in different years; it is possible for soil characteristics to alter during the year (van der Pauw, 1948; Ferrari and Vermeulen, 1955). However, from the point of view of economical procedure, it is to be desired that the soil samples should be taken, as far as is possible, throughout the whole year.

But the sampling error is not the only source of inaccuracy; for the analysis itself is liable to a certain error $s_a$. As far as the Netherlands is concerned, this also includes the error arising from the fact that analyses are carried out at four different regional laboratories. Vermeulen (1953) has already made some statements about the accuracy of routine analysis in these laboratories.

The final result of a soil test therefore has a total error $s_T$, composed of both the above errors.

It is obvious that the importance of soil testing to practical agriculture stands or falls by the possibility of accurately ascertaining the condition of the field under examination. Research workers therefore concentrated immediately on this aspect of the subject; hence, systematic investigations of it date from as far back as the beginning of the century (inter alia, Robinson and Lloyd, 1915–1916). Accordingly, investigators are generally familiar with the errors mentioned. It is, however, striking that, in many soil tests, no attempt is made to find out what errors have caused the deviation present. This means that statistical treatment of results is lacking; and it is therefore im-
possible to get an impression of the significance of the various errors via an objective yardstick (see, inter alia, BEAR and MCCLURE, 1920; HÄHNE, 1926; HORTON and STINSON, 1938–1939; RIEHM, 1943; ROBERTSON and SIMPSON, 1954). The same holds good of the various handbooks dealing with the examination of soil. The writers of these works confine themselves to discussing the manner in which soil samples have to be taken, and leave out of consideration the questions of the precision of the method of soil examination (sampling and analysis) (WIEGNER, 1922; PIPER, 1944; Mitscherlich, 1954).

Character of the Heterogeneity of the Soil

The literature of the subject states that soil may be highly inhomogeneous, even when the field to be sampled looks homogeneous enough. No useful purpose will be served by detailed discussion here of the treatment of this point by the authorities on the subject. For this, we would refer to the following, among others: ROBINSON and LLOYD (1915–1916), WAYNICK (1918), CHAMINADE (1933), FRANCK (1935), YOUDEN and MEHLICH (1937–1938), BRUIN (1938), RIEHM (1941), BUNDGAARD and LARSEN (1944), CLINE (1944), SCHUFFELEN, HUDIG and WITTEWAALL (1944–1945), REED and RIGNEY (1947), DOWNES and BECKWITH (1951), RAUPACH (1951), TISDALE (1951), KELLER, MAGEE and DOW (1952), ROBERTSON and SIMPSON (1954). We shall content ourselves with mentioning some general aspects only.

One of the most striking facts emerging from all these investigations is that the heterogeneity of the soil, even over small areas, may be very great. SCHUFFELEN, HUDIG and WITTEWAALL (1944–1945) took borings at intervals of 10 cm from plots one metre square in some fields of clay and sandy soils, and examined each boring separately for contents of K, Ca, Mg and PO₄ (MORGAN-VENEMA technique). The outcome of determination of the K content

![Variation in the K status within a (small) area of one square metre. The K statuses are expressed as percentages of the average.](image_url)
of a sandy soil can be seen in Fig. 1. Results have been expressed as percentages of the average. The figures show that, at intervals of 10 cm, the differences may exceed 100%. The sampling error amounts here to about 40% per sample (the error of analysis has been taken into consideration).

Differences at greater distances are often more systematic in character. By way of example we quote the results, as yet unpublished, of an examination carried out by van der Pauw. A pasture on sandy soil was sampled intensively by taking a boring at every 50 cm, each of which borings was analyzed separately. In Fig. 2 the results of analysis have been plotted against the location of the sample concerned in the field. The systematic trend has been indicated by a line. If we regard the spread round the line as having been caused by the above-mentioned distribution at short intervals, and by the error of analysis, it will be seen that great differences in level exist. (See also van der Pauw and Ris, 1955).

**Fig. 2. Systematic change, over a considerable distance, in P number, K content and P citr, on sand grassland. The distance is given in 0.5 metres.**

It is to be expected that heterogeneity of the soil, and consequently the sampling error, will increase with the size of the field. Data concerning this, however, are scanty. Youden and Mehlich (1937—1938) and Downes and Beckwith (1951) did, in fact, find an increase in case of the pH. Reed and Rigney (1947), on the other hand, state that, as regards potassium, heterogeneity is the decisive factor in a small area. We shall revert to this again later.

The conclusion can also be drawn that the degree of heterogeneity of the soil need not necessarily be the same for all soil characteristics (see, for instance, Robertson and Simpson, 1954).
SAMPLING ERROR AS A MEASURE OF THE HETEROGENEITY OF THE SOIL

In studying the problem of sampling technique, it is customary to regard the field to be sampled as a collection of "sampling units", differing from each other in horizontal (and vertical) direction. The deviation present is designated by \( s_h = \sqrt{\frac{1}{n-1} \sum (h_i - \bar{h})^2} \); in which \( h_i \) is the result of analysis of a sampling unit, \( \bar{h} \) the average result, and \( n \) the number of sampling units concerned in the examination. The standard error, or standard deviation of the mean \( s_{\bar{h}} \) is a measure of the reproducibility of the bulk sample, which is composed of \( n \) sampling units; \( s_{\bar{h}} \) is calculated from \( s_h = \frac{s_h}{\sqrt{n}} \).

In following this procedure we have assumed the chemical analysis to be free from error. The relation of the sampling error to the error of analysis is such that the order of magnitude of the sampling error is not appreciably altered by ignoring the error of analysis. On the basis of data from the literature of the subject Cline (1944) concludes that the error of analysis is from one third to one sixth as large as the other errors. He decides that "the limit of accuracy generally is determined by the sample, not by the analysis". Vermeulen (1953) gives a survey of the magnitude of the errors of analysis in the practical service laboratories in the Netherlands. He arrives at the same conclusion.

We shall not consider further the demands made on correct sampling in connection with the above. For these we would refer to the various statistical handbooks (e.g., Youden, 1951), and also to the article by Cline (1944), in which the principles of soil sampling are discussed in detail.

From the literature (consulted), it is difficult to obtain a statement as to the average error to be expected in carrying out soil tests.

In the first place, statistical treatment of results is often lacking, which, of course, automatically implies that no sampling error can be stated.

Another drawback is that most of the investigations in question relate to areas ranging from small to very small. The results obtained in these cases cannot then be applied as they stand to sampling for practical purposes.

A further great disadvantage arises from the fact that particular investigations have generally been directed only to one or two fields — a procedure which makes it very dangerous to draw general conclusions from the results. Exceptions to this are formed by the investigations of Keller, Magee and Dow (1952), which related to 132 plots.

From the facts that the plots differ somewhat as regards homogeneity, and the degree of inhomogeneity cannot be judged visually, it follows that the ideal would be for the sampling error invariably to be stated whenever the results of examining a plot of land are reported. This would necessarily involve the individual analysis of a number of sampling units from each plot.

It is obvious that practical and financial reasons make it impossible to apply this method. Accordingly, in practice only a bulk sample is taken, from which it may be concluded that the degree of reliability of the soil examination cannot be stated in respect of each individual plot of ground. It is therefore necessary to make do with a statement of the average degree of accuracy.
after which it is still possible to strive for more precise determination of the various plots according to field size, soil type, etc.

It will be best to take as many sampling units as possible, because the sampling error will then be small and the degree of accuracy great.

The answer to the question as to how many sampling units have to be taken depends on the heterogeneity of the soil, the degree of accuracy desired, and practical possibility. It will be clear that advice which fails to take these sources of error into consideration will not be acceptable (see, for example, Rieim, 1943). The instructions current in the Netherlands prescribe 30–40 samples. In this case the sampling error is reduced by about 80%. If larger numbers of sampling units are taken, the relation of effect to costs becomes much more unfavourable.

**Size of the Sampling Error in Soil Tests for Practical Agriculture**

The investigation into accuracy of sampling, carried out by the Agricultural Experiment Station and the Laboratory for Soil and Crop Testing, related to about 1000 plots of ground, scattered throughout the entire Netherlands. Each plot was sampled in duplicate. The bulk samples (40 sampling units drawn from grassland, 30 from arable land) originated from various types of soil. From the differences between the duplicates the $s_n$ could be calculated from $s_n = \sqrt{\frac{\sum d^2}{2n}}$, $n$ being the number of differences.

The object of the investigation was to obtain an answer to the following questions:

- How big is the average sampling error of the various determinations?
- Does soil management (arable land, pasture) play a part in this?
- Is the degree of accuracy obtained in a large field less, on the average, than that obtained in a small field?
- Has the type of soil any influence on results?

One of the most striking results of this investigation was the discovery that the size of the field concerned has, on the average, no effect on the magnitude of the sampling error. This conclusion holds good only of fields with an area of between 0.3 and 2.5 hectares. It will therefore be seen that, over short distances, heterogeneity is more or less the decisive factor. And it is obvious that this conclusion is of great significance to soil tests carried out for practical purposes. The area of most fields in the Netherlands lies between these limits.

Table 1 gives a summary of the results obtained. In this table, the sampling error has been expressed as a percentage of the factor under examination; only as regards the pH value is the error given in absolute figures. Furthermore, the errors have been arranged in various groups. For practical reasons, the sampling error and error of analysis are combined in the one figure; after all, the advisory expert is concerned with the total influence of the two. Moreover, as we have seen, the share of the error of analysis in this is small. As regards the chemical analysis, an error of 4% (duplicate determination) can be taken into account.

Before embarking on a discussion of the data in the table, it should be pointed out that the sampling error generally increases as the values of the
Table 1. Average sampling errors in grassland and arable land when 40 and 30 sampling units, respectively, were taken (the figures in italics relate to a number of samples less than 20).

<table>
<thead>
<tr>
<th></th>
<th>pH absolute</th>
<th>K content</th>
<th>P number</th>
<th>P cit</th>
<th>MgO</th>
<th>Humus</th>
<th>Humus (loss on ignition)</th>
<th>% clay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grassland and arable land</td>
<td>0.09</td>
<td>12</td>
<td>13</td>
<td>11</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Grassland</td>
<td>0.09</td>
<td>11</td>
<td>14</td>
<td>11</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Arable land</td>
<td>0.10</td>
<td>8</td>
<td>11</td>
<td>10</td>
<td>12</td>
<td>6</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

grassland

<table>
<thead>
<tr>
<th></th>
<th>as %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diluvial sandy soil</td>
<td>0.09</td>
</tr>
<tr>
<td>Young sea clay</td>
<td>0.08</td>
</tr>
<tr>
<td>Old sea clay</td>
<td>0.07</td>
</tr>
<tr>
<td>River clay</td>
<td>0.11</td>
</tr>
<tr>
<td>Reclaimed peat subsoils</td>
<td>0.07</td>
</tr>
<tr>
<td>Peat soils</td>
<td>0.11</td>
</tr>
</tbody>
</table>

arable land

<table>
<thead>
<tr>
<th></th>
<th>as %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diluvial sandy soil</td>
<td>0.09</td>
</tr>
<tr>
<td>Young sea clay</td>
<td>0.07</td>
</tr>
<tr>
<td>Old sea clay</td>
<td>0.07</td>
</tr>
<tr>
<td>River clay</td>
<td>0.13</td>
</tr>
<tr>
<td>Reclaimed peat subsoils</td>
<td>0.07</td>
</tr>
<tr>
<td>Peat soils</td>
<td>0.22</td>
</tr>
<tr>
<td>Loess soils</td>
<td>0.13</td>
</tr>
</tbody>
</table>

$\% = P$ soluble in 1% citric acid.

Factor concerned increase. In the case of very low values a minimum is reached, as a result of which the percentual error in this range is greater than the value stated. A general view of the situation is given in Fig. 3, in which the top line represents the connection between the magnitude of the sampling error (errors of sampling and analysis) and the potassium content. The figure therefore indicates that, in examining arable land with a potassium content of 30, an average error of 2.5 can be taken into account. The bottom line indicates the share of the error of analysis in this. We would point out that both these errors have to be squared if a correct comparison is to be made.

From the table it will be seen that, on an average, the total error in grassland is just as great as that in arable land. The average potassium content of grassland is higher than that of arable land.

A distinction must also be made between the errors in the chemical determination of phosphate, potassium and magnesium contents and those in the determination of organic matter and clay. The error in the latter case lies in the neighbourhood of 5% ; the error in the first case is about twice as great, and amounts to 10%. *Robinson and Lloyd* (1915—1916) also state that such a distinction is to be made.
This result therefore means that, in a plot with a potassium content which, if infallibly determined, would be 50, on an average one of three samplings will yield a result less than 45 or greater than 55. Occasionally, even a very great deviation will be encountered. Big differences are sometimes found on resampling and re-testing a field, either with or without an interval of time between the samplings. These differences may therefore largely be due to the fact that the soil in the fields is heterogeneous.

The table also shows that the differences in homogeneity between the various types of soil are usually small. It would appear that loess and river clay soils are less homogeneous in some respects than are other types of soil. The data concerning the peat soils are not very reliable, owing to their scantiness.

BIBLIOGRAPHY


CASTENMILLER, G. M. C.: Rapport van een onderzoek naar de mogelijkheid om in grondmonsters voor aardappelmoedsonderzoek op betrouwbare wijze de pH van de grond te bepalen. Report, Agricultural Experiment Station and Institute for Soil Research (T.N.O.), Groningen, 1950.


Hétérogénéité et analyse de sol

Résumé et conclusions

La valeur des analyses de sols dépend hautement de la possibilité d’obtenir une mesure de l’état de fertilité d’une parcelle au moyen d’un échantillon de sol et les déterminations chimiques qui y sont faites au laboratoire. Le résultat analytique est donc chargé d’une faute d’échantillonnage et d’une faute d’analyse. Généralement, la littérature sur ce problème ne distingue pas assez le part de ces différentes fautes. Les recherches sur l’ampleur de la faute d’échantillonnage ne se rapportent souvent que sur quelques cas et sur des surfaces restreintes.

Le Laboratoire d’analyse des sols et des végétaux et la Station Expérimentale d’Agronomie et Institut pour le Sol T.N.O. ont entamé des recherches sur l’ampleur de la faute d’échantillonnage au moyen d’à peu près 1000 parcelles dispersées sur toute l’étendue du territoire des Pays-Bas. L’échantillonnage s’est fait en double. Les conclusions suivantes purent être tirées :

a) Les fautes d’analyses sont généralement restreints et de peu d’importance par rapport aux autres fautes (dans le sens le plus large du terme). Elles mesurent à peu près 4 %.

b) Les parcelles sont hétérogènes à différents points de vues, tandisque l’échantillon de sol pris doit être représentatif pour la parcelle entière. Sous l’influence de l’échantillonnage et d’analyse la déviation standard est en moyenne de 10 % pour les déterminations de l’acide phosphorique, de la potasse et du magnésium et de 5 % pour celles de l’humus et de l’argile. La déviation standard pour le pH est d’à peu près 0.1 unité. Entre les terres cultivables et les prairies, ainsi qu’entre les différents types de sols il n’y a que peu de différence.

c) L’hétérogénéité à courte distance est le facteur déterminant de la faute d’échantillonnage. Il ne fut pas possible de démontrer une influence de la grandeur de la parcelle (0.3–2.5 ha).

d) Vue la grande importance de la faute d’échantillonnage sur la sécurité des analyses de sols, il s’avère nécessaire de procéder très soigneusement à l’échantillonnage ; l’échantillonnage doit satisfaire aux exigences posées (CLINE, 1944).

e) Si l’on veut comparer les résultats de deux ou plusieurs échantillonnages d’une même parcelle prélevés à des dates différents, il faudra tenir compte des changements plus ou moins grands intervenus sous l’influence du climat, de la mode d’exploitation, etc.

f) Malgré toutes ces mesures il restera un certain degré d’insécurité. Le conseiller ne devra donc pas voir les chiffres d’analyses dans un sens trop absolu.

g) A cause de cette valeur relative des analyses de sols, le conseil de faire répéter les analyses de sols d’une même parcelle, par exemple une fois tous les cinq ans, gagne en valeur. La valeur et l’importance des analyses de sols croissent à mesure que cette parcelle a été sous contrôle plus longtemps. Il est ainsi possible également de connaître la sécurité de l’analyse du sol pour la parcelle en question.

Bodenungleichmässigkeit und Bodenuntersuchung

Zusammenfassung und Schlussfolgerungen


Das Betriebslaboratorium für Boden- und Gewächsuntersuchungen und der Landwirtschaft­liche Versuchsanstalt und Institut für Bodenforschung T.N.O. untersuchten die Grösse der Probe­entnahmefehler bei ungefähr 1000 zweimal bemusterten Parzellen aus allen Teilen der Niederlande. Die folgenden Schlussfolgerungen wurden gezogen :

a) Der Analysefehler ist im Allgemeinen gering und spielt im Vergleich mit anderen Fehler (im weitesten Sinne des Wortes) keine grosse Rolle. Sie beträgt ungefähr 4 %.

b) Die Parzellen sind in verschiedenen Hinsichten unregelmässig. Gewöhnlich kann nur
eine Probe entnommen werden und diese soll dann repräsentativ sein für die ganze Parzelle. Unter Einfluss der Probe-entnahmefehler und der Analysefehler beträgt die Standardabweichung im Mittel 10 % für die Phosphat-, Kali- und Magnesiabestimmungen und 5 % für die Humus und Tonbestimmungen. Die Standardabweichung für die pH-bestimmungen beträgt ungefähr 0.1 Einheit. Zwischen Grünland und Ackerböden und zwischen den verschiedenen Bodentypen gibt es nur geringe Differenzen.

c) Die Unregelmäßigkeit über kurze Abstände in einer Parzelle bestimmt die Grösze der Probe-entnahmefehler. Es konnte kein Einfluss der Parzellegrösze (0.3–2.5 ha) festgestellt werden.

d) Die grosse Bedeutung der Probe-entnahmefehler für die Zuverlässigkeit der Bodenuntersuchungen bedingt eine sorgfältige Probe-entnahme; die Probe-entnahme soll den gestellten Anforderungen entsprechen (CLINE, 1944).

e) Indem zwei oder mehrere Proben einer selben Parzelle verglichen werden müssen, wenn zwischen diesen Probe-entnahmen eine gewisse Periode vergangen ist, dann soll man mit mehr oder weniger grossen, durch Klima, Betriebsführung, usw. bedingten Aenderungen Rechnung halten.

f) Ungeachtet dieser Maßnahmen bleibt eine grosse Unzuverlässigkeit übrig. Der Berater soll daher die Analyseresultate der Bodenuntersuchung nicht all zu reell auffassen müssen.

g) Des relativen Wertes der Bodenuntersuchungen wegen, bekommt der Rat z.B. jede fünf Jahre eine Parzelle aufs Neue untersuchen zu lassen, eine gewisse Bedeutung. Desto länger eine Parzelle unter Kontrolle ist, desto gröszer die Bedeutung und der Wert der Bodenuntersuchung wird um die betreffende Parzelle kennen zu lernen.