First Inter-laboratory Comparison Report
of the Regional Soil Laboratory Network for Asia
SEALNET
First Inter-laboratory Comparison Report of the Regional Soil Laboratory Network for Asia SEALNET

by
Nopmanee Suvannang, Chairperson of GLOSOLAN
Christian Hartmann, GLOSOLAN working group (IRD researcher)

Reviewers
Philip Moody, Australasian Soil and Plant Analysis Council (ASPAC)
Robert Dehayr, Australasian Soil and Plant Analysis Council (ASPAC)

Editors
Lucrezia Caon, Global Soil Partnership, FAO
Fiona Bottiglieri, Global Soil Partnership, FAO
Matteo Sala, Global Soil Partnership, FAO
Isabelle Verbeke, Global Soil Partnership, FAO

Food and Agriculture Organization of the United Nations
Rome, 2019
# Contents

Tables IV  
Figures V  
Preface VII  
Acknowledgements VIII  
Definitions and terminology IX  
Executive summary XI  
1. Introduction 1  
2. Testing material: preparation and sending  
   2.1 Sample preparation 2  
   2.2 Composition of the set of samples sent to each laboratory 2  
   2.3 Sample distribution 2  
   2.4 Results submission 2  
3. Statistical evaluation 3  
   3.1 General principles 3  
   3.2 Accuracy assessment using ‘consensus value’ 3  
   3.3 Using standard descriptive statistics for the inter-laboratory comparison 4  
   3.4 Using ‘robust’ descriptive statistics for the inter-laboratory comparison 6  
   3.5 Laboratory precision 6  
4. Report on inter-laboratory comparison for SEALNET 7  
   4.1 Quality control chart (Figures 4 to 10) 7  
   4.2 z score of each analytical result (Figure 11 to 17) 16  
   4.3 Mean z score for each laboratory and analytical parameter (Figures 18 to 24) 24  
   4.4 Histograms of distribution and fit with a normal model (Figure 25 to 31) 32  
5. Report on laboratory precision 42  
   5.1 Precision estimated from single cv values (Figure 33) 42  
   5.2 Precision estimated from mean-cv (Figure 34) 44  
6. General discussion 46  
   6.1 Analysing the lab performance in terms of accuracy and precision 46  
   6.2 Analysing the performance and homogeneity of SEALNET 47  
Conclusions and recommendations 49  
References 50  
Appendix 1. List of participating laboratories 51
Tables

Table 1. Agreed method endorsed during the first meeting of laboratories’ managers in Bogor (Indonesia) in 2017 (*agreed method that was recommended).

Table 2. Number of data analysed for each parameter and each soil type.

Table 3. The z-score interpretation

Table 4. Summary of the performance of the participating laboratories which fall into unsatisfactory

Table 5. For each parameter and soil type, the number of results received (n), the median and MADe of these results; the number of outliers, the mean (consensus value), the standard deviation (sd) and the coefficient of variation (cv) calculated after excluding the outliers from the dataset.

Figures

Figure 1. Illustration of the concept of accuracy: on the left side all results are close to the centre of the target where the ‘true value’ is located, so they are called ‘accurate’; on the right side, all results are far from the true value and they are considered as having low accuracy (orange area) or being ‘inaccurate’ (surrounding white area).

Figure 2. Graphical presentation of the identification of outliers (in this example there is only one outlier; the blue dot on the top).

Figure 3. Illustration of the concept of precision, i.e. being able to hit the target on the same position, whatever the position.

Figure 4. pH value determined by 1:2.5 soil: water suspension. The solid line corresponds the consensus value, the dotted lines to +/- 2 and +/- 3 sd.

Figure 5. Organic carbon content determined by the Walkley & Black method (OC_WB). The solid line corresponds the consensus value and the dotted lines to +/- 2 and +/- 3 sd.

Figure 6. Organic carbon content determined by the combustion method (OC_Comb). The solid line corresponds the consensus value and the dotted lines to +/- 2 and +/- 3 sd.

Figure 7. Available P content determined by the Olsen method (P_Olsen). The solid line corresponds the consensus value and the dotted lines to +/- 2 and +/- 3 sd.

Figure 8. Available P content determined by the Bray 1 method (P_Bray.1). The solid line corresponds the consensus value and the dotted lines to +/- 2 and +/- 3 sd.

Figure 9. Available P content determined by the Bray 2 method (P_Bray.2). The solid line corresponds the consensus value and the dotted lines to +/- 2 and +/- 3 sd.

Figure 10. Exchangeable K content determined using ammonium acetate as an extractant (K_exch). The solid line corresponds the consensus value and the dotted lines to +/- 2 and +/- 3 sd.

Figure 11. The z score for each single pH determined on a 1:2.5 soil:water suspension. The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.

Figure 12. The z score for each single organic carbon content determined using the Walkley and Black method (OC_WB). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.

Figure 13. The z score for each single organic carbon content determined using combustion method (OC_Comb). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.

Figure 14. The z score for each single available P content determined using the Olsen method (P_Olsen). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.

Figure 15. The z score for each single available P content determined using the Bray 1 method (P_Bray.1). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.

Figure 16. The z score for each single available P content determined using the Bray 2 method (P_Bray.2). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.
Figure 17. The z score for each single exchangeable K content determined using ammonium acetate as an extractant (K_exch). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.

Figure 18. The mean z score for the replicates of pH measures determined on a 1:2.5 soil: water suspension, the coloured lines correspond to the limit of warning (green) and unsatisfactory (red) results.

Figure 19. The mean z score calculated from the replicates of organic carbon content determined using the Walkley and Black method (OC_WB). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.

Figure 20. The mean z score calculated from the replicates of organic carbon content determined using combustion method (OC_Comb). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.

Figure 21. The mean z score calculated from the replicates of available P content determined using the Olsen method (P_Olsen). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.

Figure 22. The mean z score calculated from the replicates of available P content determined using the Bray 1 method (P_Bray.1). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.

Figure 23. The mean z score calculated from the replicates of available P content determined using the Bray 2 method (P_Bray.2). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.

Figure 24. The mean z score calculated from the replicates of exchangeable K content determined using ammonium acetate as an extractant (K_exch). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.

Figure 25. Histograms of the distribution of all pH measures, determined on 1:2.5 soil: water suspension, overlaid by a model of normal distribution (bell curve).

Figure 26. Histograms presenting the distribution of all organic carbon content, determined using the Walkley and Black method (OC_WB), overlaid by a model of normal distribution (bell curve).

Figure 27. Histograms presenting the distribution of all organic carbon content, determined using the combustion method (OC_Comb), overlaid by a model of normal distribution (bell curve).

Figure 28. Histograms presenting the distribution of all available P content, determined using the Olsen method (P_Olsen), overlaid by a model of normal distribution (bell curve).

Figure 29. Histograms presenting the distribution of all available P content, determined using the Bray 1 method (P_Bray.1) overlaid by a model of normal distribution (bell curve).

Figure 30. Histograms presenting the distribution of all available P content, determined using the Bray 2 method (P_Bray.2) overlaid by a model of normal distribution (bell curve).

Figure 31. Histograms presenting the distribution of all exchangeable K content determined using ammonium acetate as an extractant (K_exch) overlaid by a model of normal distribution (bell curve).

Figure 32. Precision of each analytical method estimated from the coefficient of variation (cv) between replicates from the same soil (955, 970, KI, LB from left to right). Note that a low cv indicates a high precision and vice versa.

Figure 33. Laboratories' precision calculated from variations between analytical results obtained on the replicates of the same soil samples (3 rep. for 955 and 970, 4 rep. for KI and LB). Note that a low cv indicates a high precision and vice versa.

Figure 34. Laboratories' mean precision calculated from the four coefficients of variation presented on Figure 33; error bars represent the standard deviation around the mean. Note that a low cv indicates a high precision and vice versa; short error bars indicates similar precision for the different soil types and vice versa.
Preface

The Global Soil Laboratory Network (GLOSOLAN) was formally established under the framework of the Global Soil Partnership (GSP) in November 2017, when its first meeting took place at FAO Headquarters in Rome, Italy. GLOSOLAN’s objectives are: (1) to strengthen the performance of laboratories through use of standardized methods and protocols, and (2) to harmonize soil analysis methods so that soil information is comparable and interpretable across laboratories, countries and regions. In this context, GLOSOLAN plans to develop open access Standard Operating Procedures and manuals on good laboratory practices, execute regional and global proficiency testing, and increase the overall performance of laboratories through the organization of training sessions. By April 2019, over 220 laboratories from all continents were registered in GLOSOLAN.

The potential of GLOSOLAN is enormous. By increasing laboratory performance, the Network will support decision making at field and policy levels; support countries in reporting on the Sustainable Development Goals (SDGs) and on other international commitments; contribute to the development of international standards and indicators; contribute to the establishment of the Global Soil Information System (GLOSSIS), which is another priority activity of the GSP. GLOSOLAN will also contribute to the development of harmonized methods for the assessment and monitoring of degraded lands, the impact of climate change on lands, and other threats to soil functions, as identified in the Status of the World Soil Resources report. The Network has the potential to improve the connection between soil chemistry, physics and biology; contribute to and improve soil classification and description; assist companies manufacturing laboratory equipment in improving their products; expand the opportunities for technical and scientific cooperation; strengthen the capability of extension services; identify research needs; and increase investments in soil related research.

GLOSOLAN operates at the regional level through its Regional Soil Laboratory Networks (RESOLANs) and at the national level through National Reference Laboratories identified by the GSP national focal points. These National Reference Laboratories are tasked to establish National Soil Laboratory Networks in order to transfer GLOSOLAN knowledge to the other national laboratories that can spontaneously register in the network. The first regional network linked to GLOSOLAN was established in Asia in November 2017. The network was named after the already existing South-East Asian Laboratory Network (SEALNET), which was launched in 2014 by Mrs. Nopmanee Suvannang, at that time, Head of the Soil Analysis Laboratory and researcher from the Land Development Department of Thailand, and Dr. Christian Hartmann, researcher from the Institut de Recherche pour le Développement – IRD, France.

During their first meeting, managers from 18 National Reference Laboratories in Asia decided to maintain the name “SEALNET” for their regional network and elected Dr. Jamyang (the laboratory manager from The Soil and Plant Analytical Laboratory - SPAL, Bhutan) as Chair and Dr. Gina P. Nilo (laboratory manager from the Bureau of Soils and Water Management - BSWM, Philippine) as vice-Chair. They agreed on the SEALNET work plan for the year 2018, which included the conduction of an independent assessment of the technical performance of SEALNET laboratories through an inter-laboratory comparison.

This exercise was co-funded by the Global Soil Partnership (GSP) of the Food and Agriculture Organization of the United Nations (FAO), L’Institut de Recherche pour le Développement (IRD, France), and the Land Development Department (LDD, Ministry of Agriculture and Cooperatives, Thailand). I wish to express the gratitude of the GSP and FAO to all partners involved and to Mrs. Nopmanee Suvannang and Dr. Christian Hartmann, who led this initiative with professionalism and on a voluntary basis. Our gratitude also goes to the laboratory managers who analysed the samples and provided data, and to the external reviewers who helped to ensure the high quality of the analysis. It is our hope that the results and conclusions of this report will assist SEALNET laboratories in improving their performance, and inspire other RESOLANs and laboratories to join GLOSOLAN.

Mr. Eduardo Mansur
Director Land and Water Division
Acknowledgements

The authors wish to express their gratitude to all reviewers, whom provided constructive comments and criticisms. A special thanks goes to Dr. Philip MOODY and Mr. Robert DEHAYR from the Australasian Soil and Plant Analysis Council (ASPAC). Our gratitude also goes to the Land Development Department, Ministry of Agriculture and Cooperatives of Thailand, the Institut de Recherche pour le Développement (IRD, France) and the Global Soil Partnership of FAO for financially supporting the execution of this inter-laboratory comparison. Ultimately, the authors wish to thank the Indonesian Soil Research Institute (ISRI) for hosting the First Regional Soil Laboratory Network for Asia (SEALNET) meeting and overall allowing this exercise to be conducted.
Definitions and terminology

ACCURACY
The closeness of agreement between a test result and the accepted reference value. Note: The term ‘accuracy’, when applied to a set of test results, involves a combination of random components and a common systematic error or bias component.

A quantity referring to the differences between the mean of a set of results or an individual result and the value which is accepted as true or correct value for the quantity measured [EURACHEM Guide, 1998].

ASSIGNED VALUE
Best available estimate of the true value [UNODC, 2009].

CERTIFIED REFERENCE MATERIAL (CRM):
Reference material one or more of whose property values are certified by a technical procedure, accompanied by or traceable to a certificate or other documentation which is issued by a certifying body [UNODC, 2009].

CONSENSUS VALUE
Value produced by a group of experts or referee laboratories using the best possible methods. It is an estimate of the true value [UNODC, 2009].

ERROR (OF MEASUREMENT)
The value of a result minus the true value [EURACHEM Guide, 1998].

INTERNAL QUALITY CONTROL
Set of procedures undertaken by a laboratory for continuous monitoring of operations and results in order to decide whether the results are reliable enough to be released. Quality control of analytical data primarily monitors the batchwise trueness of results on quality control materials, and precision on independent replicate analysis of test materials [UNODC, 2009].

OUTLIERS
Outliers are extreme values, so far separated from the other values that it suggests they are (i) coming from a different population or (ii) resulting of an error in measurement or in transcription [EURACHEM Guide, 1998].

PRECISION
The closeness of agreement between independent test results obtained under stipulated conditions. Note: Precision depends only on the distribution of random errors and does not relate to the true value or specified value. The measure of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results. ‘Independent test results’ means results obtained in a manner not influenced by any previous result on the same or similar test object. Quantitative measures of precision depend critically on the stipulated conditions. Repeatability and Reproducibility are particular sets of extreme conditions [ISO Guide 35]. ‘A measure for the reproducibility of measurements within a set that is of the scatter or dispersion of a set about its central value’ [EURACHEM Guide, 1998].

PROFICIENCY TESTING (also called ‘External QC’ or ‘inter laboratory comparison’)
A periodic assessment of the performance of individual laboratories and groups of laboratories that is achieved by the distribution by an independent testing body of typical materials for unsupervised analysis by the participants [EURACHEM Guide, 1998].

QUALITY CONTROL
A set of activities or techniques whose purpose is to ensure that all quality requirements are being met. Simply put, it is examining “control” materials of known substances along with patient samples to monitor the accuracy and precision of the complete examination process [UNODC, 2009].

REFERENCE MATERIAL (RM)
Reference material, one or more of whose property; are certified by a technical procedure, accompanied by, or traceable to, a certificate, or other documentation, which is issued by a certifying body [UNODC, 2009].
STANDARD DEVIATION
This is a measure of how values are dispersed about a mean in a distribution of values: The standard deviation’s for the whole population of ‘n’ values is given by:

\[ \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}} \]

In practice we usually analyse a sample and not the whole population. The standard deviation’s for the sample is given by:

\[ s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}} \]

[EURACHEM Guide, 1998]

UNCERTAINTY (OF MEASUREMENT) i.e. MEASUREMENT UNCERTAINTY:
‘Parameter associated with the result of a measurement that characterises the dispersion of the values that could reasonably be attributed to the measurand. Note: The parameter may be, for example, a standard deviation (or a given multiple of it), or the width of a confidence interval. Uncertainty of measurement comprises, in general, many components. Some of these components may be evaluated from the statistical distribution of the results of a series of measurements and can be characterised by experimental standard deviations. The other components which can also be characterised by standard deviations, are evaluated from assumed probability distributions based on experience or other information. It is understood that the result of the measurement is the best estimate of the value of the measurand and that all components of uncertainty, including those arising from systematic effects, such as components associated with corrections and reference standards, contribute to the dispersion’ [EURACHEM Guide, 1998].

Z score
Standardized measure of performance, calculated using the participant result, assigned value and the standard deviation for proficiency assessment [WHO, 2016].
Executive summary

The first proficiency test of SEALNET was organized in 2018 with the purpose of assessing the performance and inter-/intra-laboratory variability of 16 National Reference Laboratories in Asia. Testing soil samples were prepared with the financial support of the International Joint Laboratory ‘Impact of Rapid Land Use Change on Soil Ecosystem Services (LMI-LUSES)’ from the Institut de Recherche pour le Développement (IRD, France) and the Land Development Department (LDD, Thailand), and shipped to participating laboratories by FAO’s Global Soil Partnership (GSP).

Each laboratory received replicates of four soil types. In total, each laboratory was asked to analyse 14 soil samples for soil pH (in soil water ratio: 1:2.5), organic carbon (using Walkley & Black and/or dry combustion), available phosphorus (using Olsen and/or Bray 1 and/or Bray 2 methods) and Exchangeable K (using NH₄OAc method). Laboratories were not informed about the nature of the samples, whose coding was also randomised at the purpose of preventing laboratories to compare each others results.

The GSP collected and anonymized laboratories’ results before sending them to Ms. Nopmanee Suvannang and Mr. Christian Hartmann for the statistical analysis. In order to estimate the laboratory’s accuracy, the consensus value was calculated after identifying and excluding outliers in the dataset. The variability between laboratories was estimated from the standard deviation around the consensus value while the performance of each laboratory was estimated by calculating the commonly used z-score. Robust statistic parameters (median, MADₑ) were calculated for information. Intra-laboratory variability (precision) was estimated by calculating the coefficient of variation around the mean value of the replicates coming from a given soil type for each laboratory.

This report presents the results of the analysis using different figures to help laboratory managers and other non-specialist readers to perceive the different aspects of (i) the laboratory performance evaluation, (ii) the way to identify the technical problems in case of poor performances and (iii) suggesting which solutions can be proposed to improve the analytical performances. Overall, the variability around the consensus value in nearly all laboratories was variable depending on the soil characteristic: it was low for soil pH, medium and questionable for organic carbon (OC), and generally high and often unsatisfactory for available P and exchangeable K. Because the laboratory’s precision (intra-laboratory variability) was also variable depending on the soil characteristic, poor laboratory performances were not related to differences in the Standard Operating Procedures (SOPs) used. Otherwise, variability in precision could be related to (i) the lack of quality control inside the laboratories, in particular the absence of internal control samples, and (ii) the lack of sufficient initial and ongoing professional training of the staff. The same observations apply to the soil organic carbon results obtained by dry combustion. Even though this method generally provides better results than the oxidation method, the low performance of one National Reference Laboratory confirmed that staff training and qualification is necessary to get high performance.

In conclusion, it is recommended that (i) all laboratories should implement good laboratory practices and quality control programmes, regardless to the adoption of SEALNET/GLOSOLAN SOPs; (ii) laboratories with a high performance should be identified for the purpose of providing training to laboratories in need, and (iii) inter-laboratory comparisons should be organized on a regular basis (at least twice a year) for the purpose of monitoring laboratories’ performance and of measuring the impact of staff training and SOPs implementation.
1. Introduction

Southeast Asia Laboratory Network (SEALNET) is the regional network of laboratories from the countries of the Asian Soil Partnership (ASP), a regional section of the GSP. The objective of Pillar 5 of the GSP and of SEALNET is to help soil laboratories produce analytical results that can be compared, wherever the soil sample was analysed inside the Region.

For the GSP, the Asian Region consists of 24 countries, each of which having to appoint its own reference laboratory when participating in GSP and ASP activities. At the first SEALNET meeting organised in November 2017 in Bogor (Indonesia), 18 countries sent at least one representative of their national reference laboratory (see meeting report on http://www.fao.org/3/i9063EN/i9063en.pdf). During discussions, it became clear that, for the main soil characteristics, most of the laboratories used analytical methods based on the same chemical and physical principles, but many differences were observed concerning the details of the analytical procedures.

To estimate if the results coming from different laboratories could reasonably be compared, it appeared necessary to evaluate the impact of these differences in analytical procedures on the final analytical result. Thus it was decided to organise an inter-laboratory comparison by sending the same soil sub-samples to all reference laboratories and letting them analyse these samples according to agreed method (Table 1) following individual laboratory procedures for determination of soil pH, organic carbon (OC), available P and exchangeable K in order to assess their performance and comparability on the basis of a statistical analysis of their results.

Preliminary results of the statistical and performance analyses were presented in November 2018 during the second SEALNET meeting in Bhopal (India) and during the second GLOSOLAN meeting in Rome (Italy).

The current report is presenting the details of:
- the procedures used to prepare and send the test soil materials;
- the procedure of statistical analysis;
- the figures presenting the performance of the laboratories concerning accuracy and precision;
- comments on the interpretation of the performance addressed to the laboratory managers but also to the stakeholders that use the results for decision making;
- a conclusion on the performance and comparability of the analytical results and finally some recommendations for the future inter-laboratory comparisons.

### Table 1. Agreed method endorsed during the first meeting of laboratories’ managers in Bogor (Indonesia) in 2017 (* agreed method that was recommended).

<table>
<thead>
<tr>
<th>Soil testing parameter</th>
<th>Method</th>
<th>Noted</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH in water 1:2.5</td>
<td>Adjust the soil :water to 1:2.5 and follow your regular SOP</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>OC</td>
<td>Walkley &amp; Black*</td>
<td>Follow your regular SOP and report which method that you have used</td>
<td>percent</td>
</tr>
<tr>
<td></td>
<td>Dry combustion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avail P</td>
<td>Olsen P*</td>
<td>Follow your regular SOP and report which method that you have used</td>
<td>mg/kg</td>
</tr>
<tr>
<td></td>
<td>Bray 1 P</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bray 2 P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exch K</td>
<td>NH₄OAc*</td>
<td>Used your regular SOP</td>
<td>mg/kg or cmolc/kg</td>
</tr>
</tbody>
</table>
2. Testing material: preparation and sending

Four soil samples coming from different locations and having consequently different characteristics were selected by the Central Laboratory of the Land Development Department (LDD, Ministry of Agriculture and Cooperatives, Thailand) and prepared with the technical and financial support of the LMI-LUSES (http://www.luses.ird.fr), IRD (France). The code names of the soil samples were: '955', '970', 'LB', 'KI'.

2.1 Sample preparation

Approximately 50 kg of each soil were air dried (temperature < 40°C), machine ground, and sieved through a 0.5 mm sieve. The samples were then homogenized by hand, subsamples and packed in zip lock plastic bags (200 plastic bags of approximately 250 g for each soil). All bags were stored at room temperature (20 to 25°C) and protected from light.

The homogeneity of the subsamples was tested by randomly sampling 10 percent of them, following the International Harmonized Protocol for Proficiency Testing of Chemical Analytical Laboratories, AOAC (Thomson et al., 2006). These materials were analysed in duplicate for four parameters: pH water (soil solution ratio of 1:1 w/v), organic carbon (Walkley & Black method), available P (Bray 2 and Olsen methods) and Exch K (NH₄OAc extraction). The homogeneity was tested by one way analysis of variance (ANOVA, single factor) ISO Guide 35, at 95 percent of confidence level by calculating an F-statistic. No significant difference could be observed for within and between packages' standard deviation using the F-test (F-calculated < F-critical at 5 percent).

For each soil type, 80 plastic bags were randomly selected again and mixed together. After careful homogenisation, a fraction splitter was used to prepare subsamples of 30 g each. These subsamples were put in a small zip locked plastic bags and kept at room temperature and protected from light, so that each laboratory got its samples ordered and labelled differently from the other laboratories.

2.2 Composition of the set of samples sent to each laboratory

The objective of this technical report on inter-laboratory comparison or PT was aimed to assess the accuracy of the participating laboratories, and selected four samples with characteristics corresponding to those of soils commonly analysed by the participants. Moreover, the precision of each laboratory was also assessed from the replication of the same soil: 3 replicates for '955' and for '970' and 4 replicates for 'KI' and for 'LB'.

Total of 14 samples were prepared for each laboratory (3 x '955' + 3 x '970' + 4 x 'KI' + 4 x 'LB'). In the set of 14 samples, the soil types and replicates were randomly distributed; randomisation was different for each set so that each laboratory did not have its samples in the same order.

It is noteworthy that the laboratories did not know: (i) how many different soil types and how many replicates were prepared for each set; and (ii) that the order of the samples was not the same for all the laboratories.

Well-packed parcels containing the set of 14 samples, each in zip locked plastic bags, and instructions on how to present the analytical results, were prepared, accompanied by all custom documents and provided to the FAO regional office in Bangkok (Thailand) in March 2018 which supported by sending samples to the participating laboratories.

2.3 Sample distribution

In March 2018, FAO sent one parcel to each reference laboratory of the 18 countries that participated to the Bogor meeting. Each lab was requested to inform FAO of the parcel's delivery after checking that the content of the package was complete and in good condition.

2.4 Results submission

The results had originally to be sent on 15 July 2018, but the deadline was extended to 15 October 2018.

Each participating laboratory sent its results by email to the GSP Secretariat at FAO headquarters.
The statistical evaluation of an inter-laboratory comparison can be made according to different procedures depending on (a) the number of participants, (b) the type of material that is analysed & the type of analyses, (c) the various analytical skills of the participants and (d) the final objective of the comparison.

Table 2 indicates, for each soil type and for each parameter, the number of analytical results that we used for the statistical analysis. In most of the cases, this number was high enough (>15) to permit the use of common statistical tests.

The determination of the values of the four parameters are based on chemical reactions that are common for soil analysis laboratories, and these determinations should not present any technical problems for the participants. On the other hand, different surveys and discussions conducted during SEALNET meetings, showed that: (i) the 16 laboratories have very different skills concerning the use of quality control procedures; and (ii) the staff have very different levels of education and training for chemical analysis and soil analysis. Moreover, not all 16 laboratories were participating in national or international laboratory inter-comparisons; consequently the staff of laboratories not yet involved in such comparisons are probably not familiar with statistical analysis in general, and with statistical analysis of inter-laboratory comparison in particular. As one of the main goals of SEALNET is to get all laboratories involved, including the less advanced ones, the providers tried to keep statistical analysis as simple as possible (i.e., using concepts and parameters that are most commonly used in soil science) and hopefully sufficiently clear to be understood by all participants.

The ultimate objective of the statistical analysis was to provide information concerning precision and accuracy of the ASP ‘national reference laboratories’ that could be useful for the laboratory managers, and also for other stakeholders and government agents that are not familiar with soil analysis. This reinforced motivation to keep the information accessible even to people not familiar with inter-laboratory comparison, and to use mainly classical statistical analysis to assess laboratory performance.

The statistical analysis and figures were made using R, a statistical and graphical software which is free (https://www.r-project.org/).
certified reference material provided by companies;
analytical results provided by one or several accredited laboratories;
consensus value calculated from analytical results provided by expert laboratories; and
consensus value calculated from analytical results provided by laboratories participating to a proficiency test.

In this report, the assigned value consists of the consensus value calculated from the data provided by all participating laboratories.

There are limitations in using certified reference material CRM in this SEALNET first inter-laboratory comparison, one of the important reasons is that CRMs is expensive due to the costly process of producing them. Thus, it is necessary for SEALNET to produce its own reference samples, this would make it possible to distribute its own samples to the network at a reasonable cost but still high quality.

3.3. Using standard descriptive statistics for the inter-laboratory comparison

With standard descriptive statistics, it was possible to calculate a consensus value simply by using the arithmetic mean of the data provided by the participants. But the arithmetic mean can be strongly affected by only a small number of extreme values that are not related to random variations (transcription mistakes, biased results, etc.). The mean has to be calculated after removing outliers from the data set. Consequently, the way to identify outliers becomes of particular importance as it could have a strong influence on the arithmetic mean used as a consensus value.

Outliers can be detected in different ways in relation to the type of data set, the objectives, etc. In this report, a basic test based on the interquartile range (IQR) was decided for the excluding of outliers. IQR is a measure of the statistical dispersion of a data set that is equal to the difference between the upper and lower quartiles (Figure 2):

\[ \text{IQR} = Q_3 - Q_1. \]

with \( Q_1 \) and \( Q_3 \) being the value of the first and third quartile, respectively.

A result \( x_i \) was considered as an outlier if

\[ x_i < Q_1 - (1.5 \times \text{IQR}) \quad \text{or} \quad x_i > Q_3 + (1.5 \times \text{IQR}), \]

As our samples did not yet have an assigned value, we had to calculate a consensus value from the data provided by the participants. The way the consensus value was calculated is of particular importance (as all the results were compared to that value).
The outliers were identified and removed from the data set, and the mean of the new data set became the ‘consensus value’ (i.e., the centre of the ‘target’ to which all laboratories should get as close as possible).

\[ X = \frac{(x_1 + x_2 + \ldots + x_{ns})}{ns} = \frac{\Sigma xi}{ns} \]

- \( X \) is the average of the analytical results of the new data (i.e., EXCLUDING the outliers)
- \( \Sigma \) represents the sum of all the analytical results \( x_i \) to \( x_{ns} \) (i.e., EXCLUDING the outliers)
- \( xi \) is an individual value
- \( ns \) is the number of the results in the new data set (EXCLUDING the outliers)

The standard deviation (sd) was used to estimate the dispersion of analytical results around the consensus value:

\[ sd = \left( \frac{\Sigma (x_i - X)^2}{ns} \right)^{1/2} \]

Once the consensus value and the dispersion (sd) around this consensus were calculated, it was necessary to define a tolerance range (values within the range being considered as satisfactory and values outside the range being considered as unsatisfactory).

Results were considered as:

- satisfactory if they were inside the interval of \( X \pm 2sd \);
- questionable if they were in the interval of \( X +/ -2sd \) and \( X +/ -3sd \)
- unsatisfactory if they were outside the interval of \( +/ -3sd \).

This rule is based on the assumption that the data (excluding the outliers) are normally distributed. Thus ‘satisfactory’ indicates that a given result fits within random deviation from the consensus value as it is within 95 percent of the data of a normally distributed population.

A result is regarded as questionable when the probability of it belonging to the population is < 5 percent (such an analytical result should occur only one time in 20).

A result is incorrect if its distance from the consensus value is > 3 sd.

It is noteworthy that when the consensus value is close to the detection limit, sometimes the data are not normally distributed; there is an asymmetry with few values below the consensus value, and many more above the consensus value. In this case the calculation of \( X - 2sd \) or \( X - 3sd \) can result in negative values, which have no meaning from a mathematical point of view. Even when this problem was observed, in this report, we kept the same calculation and the same way to present the critical values on the figure to keep the statistical analysis similar for the entire report.

The performance of each laboratory was estimated by calculating the z-score that represents the difference between a reported result and the consensus value, divided by the standard deviation:

\[ z = \frac{x_i - X}{sd} \]

\( x_i \) is the value determined by the laboratory;
\( X \) is the consensus value of the parameter (calculated excluding outliers);
\( sd \) is the standard deviation around the mean (calculated excluding outliers).

The interpretation of the z-score (Table 3) is based on the concept that normally distributed analytical results lie within two standard deviations with a probability of 95 percent and within three standard deviations with a probability of 99.7 percent.

Z-score values close to zero indicate results that are similar to the consensus value.

- \(|z| \leq 2\): regarded as satisfactory as it indicates normal or random deviation between the consensus value and the laboratory’s results.
- If \( 2 < |z| \leq 3\): regarded as questionable (it is considered as a ‘warning signal’ as such scores should occur with 5 percent probability). If such a z-score is obtained, the laboratory should firstly review its procedures to detect any possible systematic error.
- If \( |z| > 3\): regarded as unsatisfactory, the result not being correct.

\[^1\] The bars on each side of ‘z’ indicate the ‘absolute value’, i.e. ignores the minus sign.
Table 3. The z-score interpretation

<table>
<thead>
<tr>
<th>Ranges</th>
<th>Evaluation results</th>
</tr>
</thead>
<tbody>
<tr>
<td>(</td>
<td>z</td>
</tr>
<tr>
<td>(2 &lt;</td>
<td>z</td>
</tr>
<tr>
<td>(</td>
<td>z</td>
</tr>
</tbody>
</table>

The z-score was calculated for each single result, and finally calculated a mean z-score that is the average of the results for the three (soil sample 955 and 970) or 4 (‘KI’ & ‘LB’) replicates of the same soil. The mean-z-score provides a global view on the laboratory accuracy for a given parameter.

3.4. Using ‘robust’ descriptive statistics for the inter-laboratory comparison

It can sometimes be difficult to identify outliers and take them out of the statistical analysis. Consequently, specific statistical tests have been developed that are not sensitive to outliers (this is why they have been called ‘robust’, as they provide the same results even if the dataset contains several extreme values). To determine the consensus value, robust statistics use the median; and to determine the dispersion of results around the consensus value, robust statistics use the median absolute deviation (MAD).

The value of the MAD can be compared to the value of the standard deviation, by multiplying the MAD by a constant \(k\) which depends on the distribution. For normally distributed data \(k\) is 1.4826, thus the dispersion is estimated by 1.4826 x MAD.

In this report, the standard statistics (mean, standard deviation, z-score) have been chosen. Median and MAD are also presented for information.

3.5. Laboratory precision

Precision is the ability to provide the same result after repeating the analysis; in other words it is the ability to have little variation between two analytical results (or having small distances between two hits on the target, whatever the position of the first hit) (Figure 3).

Measurement of precision was made possible by providing several replicates for the same soil sample, without the laboratory being informed of the existence of the replicates, thus without having the possibility of selecting the most probable value of the replicates. Moreover, for the different soils, we did not provide the same number of replicates, making it more difficult for the laboratories to guess the existence of replicates when looking at the list of their results. Finally, the randomisation of the 14 samples was different for each laboratory, making nearly impossible for different laboratories to compare their results and to guess the existence of replicates before sending their results.

Figure 3. Illustration of the concept of precision, i.e. being able to hit the target on the same position, whatever the position
The mean was estimated by the standard deviation around the mean. To make easier precision comparison between the different parameters, the results were also expressed as a coefficient of variation:

\[ \text{cv (percent)} = \frac{100 \times \text{sd}}{\text{mean}}. \]

It is important to note that the highest precision corresponds to the lowest cv.

4. Report on inter-laboratory comparison for SEALNET

4.1. Quality control chart (Figures 4 to 10)

The primary tool for statistical control is the control chart that presents for each analytical parameter the consensus value and the upper and lower control limit (+/-2 and +/-3 standard deviations around the consensus value).

The next figures present the quality control charts for each of the seven analytical methods.

Along the X axis, the laboratory anonymous code was plotted (from 101 to 116) and along the Y axis the value of the parameter that was analysed. The results that were obtained for the three replicates of samples '955' and '970' and the four replicates of the samples 'KI' and 'LB' were plotted for each laboratory.

Important:

- There were no missing values, i.e. when a laboratory analysed a soil sample, they always analysed the three (955 & 970) or four replicates (KI & LB) that we provided. But when a laboratory had high precision (getting similar results for the different replicates we provided), the symbols are more or less superimposed and it becomes impossible to see single analytical results on the quality control chart which provides a visual estimation of the precision.
- In this report we decided to have the same scale for the Y axis for all the figures, although our soil samples had either high or low contents of all elements to analyse. To be able to compare the control limits of the four samples, even if not totally correct on a mathematical point of view. Moreover, if we had adapted the scale of the figure to the control limits, all figures would have looked very similar (with a random distribution of most of the results between the control limits).

Soil pH

All 16 laboratories performed this analysis. The quality charts for all four samples display similar performance with standard deviations around 0.2 and 0.3 units. Most of the laboratories had all of the results inside the limits. Only labs 11 and 116 had some results outside the limits, and only lab 104 had all the results outside of the limits for samples '955' and 'KI' that were the most sandy ones. Note that the dispersion of results was higher for the two coarse-textured soils (955' and 'KI') compared to the soils with finer texture (970' and 'LB').

Organic carbon by oxidation (Walkley & Black method)

Fifteen laboratories provided results for this analysis. Samples '955' and 'KI' had a very low organic carbon content with consensus values around 0.3 percent, while 'LB' and '970' had much higher contents: 2.6 and 3.3 percent respectively. When OC content was low ('955' and 'KI'), the statistical dispersion of the result was also low (0.1 percent); in contrast, with increasing content, we observed a significant increase in the dispersion of results (0.4 percent & 0.6 percent for 'LB' and '970', respectively).

Note that for some of the four soil samples, several laboratories provided results outside the control limits (101, 103, 109, 111 and 115).

Organic carbon by combustion

Only five laboratories provided results according to this method, but we decided to use the same procedure of statistical analysis to make the comparison with other results easier, even though we knew that it was important to be careful with the interpretation of the statistical analysis.

2   +/- 2SD was between approximately +/- 0.4 and 0/- 0.6 units around the mean  and +/- 3SD was between +/- 0.6 and 0/- 0.9 units around the mean.

3  +/- 2SD was between approximately +/- 0.4 and 0/- 0.6 units around the mean and +/- 3SD was between +/- 0.6 and 0/- 0.9 units around the mean.
Compared to the previous method, the OC content appears to be slightly higher. For ‘955’ and ‘KI’ that have a low carbon content we obtained a consensus value around 0.4 percent with a standard deviation around 0.1 percent.

For ‘LB’ and ‘970’, the consensus values were 3.0 and 5.4 percent, respectively (compared to 2.6 and 3.3 percent respectively with the oxidation method).

Concerning the sample ‘970’, it was in fact not possible to detect outliers because of the large dispersion of results. This explains the large interval of the control limit and the high consensus value. For the sample ‘LB’, 4 of 5 laboratories had very close values with quite narrow control limits, making it possible to detect the results of laboratory 112 as outliers.

**Available P – Olsen method**
Fourteen laboratories of 16 provided results for this method.

Samples ‘955’, ‘KI’ and ‘LB’ had low contents (4.1, 4.4, 10.6 mg/kg respectively) while the consensus value for ‘970’ was 134.7 mg/kg.

For the three samples with low P content, the dispersion between results provided by the different laboratories was also low (sd around 5 mg/kg). For ‘970’ the dispersion was higher, resulting in a standard deviation of 99 mg/kg, and thus control limits that could not be presented on the figure.

**Available P – Bray 1 method**
Only seven laboratories of 16 provided results for this method.

According to this method, samples ‘955’, ‘KI’ and ‘LB’ had low content (1.9, 3.4, 7.3 mg/kg respectively) and the consensus value for ‘970’ was 103.2 mg/kg. As for the previous method, dispersion around the consensus value was low when P content was low, whereas sample ‘970’ had an extremely large dispersion.

**Available P – Bray 2 method**
Only three laboratories provided results according to this method.

The comments are the same as for the two previous methods; we can only indicate that surprisingly with the Bray 2 methods, sample ‘LB’ seems to have a much higher consensus value: 60.7 mg/kg (compared to 10.6 and 7.3 mg/kg for Olsen and Bray 1 methods respectively).

**Exchangeable K (Ammonium Acetate method)**
All the 16 laboratories provided results for this parameter. The four samples had various contents of exchangeable K: 20.3, 76.1, 161.3 and 367.5 mg/kg for ‘KI’, ‘955’, ‘970’ and ‘LB’ respectively. The standard deviation increased with increasing content: 15.4, 34.3, 73.7 and 181.6 mg/kg for ‘KI’, ‘955’, ‘970’ and ‘LB’ respectively. Most of the laboratories have all, or most, of their results inside the control limits (only lab 115 had all its results outside of the limits for sample 955 and 970 and lab 108 for KI sample).
Figure 4. pH value determined by 1:2.5 soil: water suspension. The solid line corresponds the consensus value, the dotted lines to +/- 2 and +/- 3 sd.
Figure 5. Organic carbon content determined by the Walkley & Black method (OC_WB). The solid line corresponds the consensus value and the dotted lines to +/- 2 and +/- 3 sd.
Figure 6. Organic carbon content determined by the combustion method (OC_Comb). The solid line corresponds the consensus value and the dotted lines to +/- 2 and +/- 3 sd.
Figure 7. Available P content determined by the Olsen method (P_Olsen). The solid line corresponds the consensus value and the dotted lines to +/- 2 and +/- 3 sd.
Figure 8. Available P content determined by the Bray 1 method (P_{Bray.1}). The solid line corresponds the consensus value and the dotted lines to +/- 2 and +/- 3 sd.
Figure 9. Available P content determined by the Bray 2 method (P_{Bray.2}). The solid line corresponds the consensus value and the dotted lines to +/- 2 and +/- 3 sd.
Figure 10. Exchangeable K content determined using ammonium acetate as an extractant (K_exch). The solid line corresponds the consensus value and the dotted lines to +/- 2 and +/- 3 sd.
4.2. z score of each analytical result
(Figure 11 to 17)

The quality control chart is used on a daily basis inside a given laboratory. Using the z score is more common for the evaluation and inter-comparison of laboratories. Consequently, we have calculated the z score for the results provided by the participating laboratories.

The benefits of this type of data visualization are:
- to have a score for each result and to compare the score of different replicates made by a given laboratory;
- to compare the score of different soil types which have very different consensus values with very different control limits.

The figures below present the performance z score for each of the 7 analytical methods. The laboratory anonymous code (from 101 to 116) have been plotted along the X axis and along the Y axis, the z-score of each results (they will be results of three replicates for samples ‘955’ and ‘970’ and of four replicates for the samples ‘KI’ and ‘LB’).
Figure 11. The z score for each single pH determined on a 1:2.5 soil:water suspension. The coloured lines correspond to the limit of questionnable (green) and unsatisfactory (red) results.
Figure 12. The z score for each single organic carbon content determined using the Walkley and Black method (OC_WB). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.
Figure 13. The z score for each single organic carbon content determined using combustion method (OC_Comb). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.
Figure 14. The z score for each single available P content determined using the Olsen method (P Olsen). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.
Figure 15. The z score for each single available P content determined using the Bray 1 method ($P_{\text{Bray.1}}$). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.
Figure 16. The z score for each single available P content determined using the Bray 2 method (P_Bray.2). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.
Figure 17. The z score for each single exchangeable K content determined using ammonium acetate as an extractant (K_exch). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.
4.3. Mean z score for each laboratory and analytical parameter (Figures 18 to 24)

The previous results were used to calculate a mean z score for each parameter and each laboratory. The results were ordered from the laboratory having the lowest to the highest z score (along the Y axis).

As the consensus value was calculated using the data provided by the laboratories, by definition, most of the laboratories fit inside the control limits. This section will present only some short comments about the laboratories that are outside of the control limits. These comments do not constitute a complete discussion but consider a few hypotheses that can be put forward.

**Soil pH**

The participating SEALNET laboratories demonstrated high precision for pH water 1:2.5 testing. Only laboratory 104 was outside of the limits (in particular for the acidic soils). Concerning the laboratory 111: its average precision was correct (z < 2 or 3) but we observed a large standard deviation around the mean Z score for the samples KI and LB (figure 18). In the 4 replicates analysed for each of those 2 samples, only one was incorrect (figure 11). It seems possible that a transcription problem may have occurred because if the extreme results were reversed from one soil sample to the other sample, there would be no excessive standard deviation for the mean z score.

**Organic carbon by oxidation (Walkley & Black method)**

Firstly, it is noteworthy that results outside the limits comprise only z-scores > 3 (and never < 3). The laboratories that are concerned are laboratories 101 and 115 (for soil '955' and 'KI' that have low OC content), and laboratory 103 and 109 (for soil '970' and 'LB' that have high OC content).

Laboratory 101 also has a z-score outside of the range for the 'KI' sample, but more importantly, it has again large standard deviations for KI and LB samples that seem again may come from transcription problems.

**Organic carbon by combustion**

Only laboratory 112 is outside the limits, while the others seem to be in the same range of results.

**Olsen-P method**

The results outside the limits consist only in z-score > 3 (and never < 3). Laboratory 115 is outside of the limits for all the soil types, and laboratory 109 is out only for soil 'LB'. Laboratories 104 and 116 present large standard deviation of the z score.

**Bray 1-P**

Only laboratory 113 is outside of range for only soil 'LB', the results show the large variations for laboratory 111, 112 and 116.

**Bray 2-P**

Due to the low number of laboratories (three) using this method, the statistical analysis was not reported here.

**Exchangeable K (Ammonium Acetate method)**

Again we observe that the results outside the limits comprise only z-scores > 3 (and never < 3). Laboratory 115 is outside of the limit for all the soil types, and laboratory 101 for two soil types ('955' and 'KI') that have low and medium content in K. Laboratory 104 and 108 are outside of range only for the KI soil sample, i.e. with the lowest content in K. Large standard deviation was observed for laboratory 101 for all four soil samples, and also for laboratories 104 and 115.

Table 4 summaries the performance of the participating laboratories which fall into Unsatisfactory according to the z score interpretation from Table 3. This summary was extracted from all data that fall in the Z score > 3 (Figure 11-17).

<table>
<thead>
<tr>
<th>Soil types</th>
<th>pH</th>
<th>OC_WB</th>
<th>Olsen-P</th>
<th>K_avail</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>lab code</td>
<td>n</td>
<td>lab code</td>
</tr>
<tr>
<td>955</td>
<td>2</td>
<td>(104-106)</td>
<td>2</td>
<td>(103-115)</td>
</tr>
<tr>
<td>970</td>
<td>0</td>
<td>2</td>
<td>(103-109)</td>
<td>1</td>
</tr>
<tr>
<td>LB</td>
<td>1</td>
<td>(111)</td>
<td>3</td>
<td>(103-109-111)</td>
</tr>
</tbody>
</table>

Table 4. Summary of the performance of the participating laboratories which fall into unsatisfactory
Figure 18. The mean z score for the replicates of pH measures determined on a 1:2.5 soil: water suspension, the coloured lines correspond to the limit of warning (green) and unsatisfactory (red) results.
Figure 19. The mean z score calculated from the replicates of organic carbon content determined using the Walkley and Black method (OC_WB). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.
Figure 20. The mean z score calculated from the replicates of organic carbon content determined using combustion method (OC Comb). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.
Figure 21. The mean z score calculated from the replicates of available P content determined using the Olsen method (P Olsen). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.
Figure 22. The mean z score calculated from the replicates of available P content determined using the Bray 1 method (P\_Bray.1). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.
Figure 23. The mean z score calculated from the replicates of available P content determined using the Bray 2 method (P_Bray.2). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.
Figure 24. The mean z score calculated from the replicates of exchangeable K content determined using ammonium acetate as an extractant (K_exch). The coloured lines correspond to the limit of questionable (green) and unsatisfactory (red) results.
4.4. Histograms of distribution and fit with a normal model (Figure 25 to 31)

To obtain a visual representation of the distribution of each dataset, a histogram was plotted to present all the data provided by participating laboratories (including outliers) then overlaid by the model of normal curve fitting the dataset. It is important to note that the model was calculated using the mean $X$ and standard deviation $sd$ of the dataset after excluding the outliers. A narrow and high normal curve corresponds to a low dispersion (small $sd$) around the mean value (i.e. the consensus value); on the contrary, a large and low normal curve, corresponds to a high dispersion (high $sd$) of analytical results around the consensus value.

Along the $X$ axis, the value of the characteristic that was analysed was plotted, and the density of probability of occurrence that can be considered as a proxy of the frequency of occurrence of each analytical value appears along the $Y$ axis.

A summary of the main characteristics of the statistical analysis is included on the right side of each figure where we have reported the main results of ‘common’ and ‘robust’ descriptive statistics:

- $n$ = number of results provided by the laboratories,
- $median$ = median value using all the set of $n$ results,
- $MAD$ = median absolute deviation using all the $n$ results, which was multiplied by a constant $k = 1.4826$ to be comparable with the standard deviation ($sd$)
- $outlier$ = number of outliers (they were not used to calculate mean & $sd$)
- $mean$ = average/mean without outliers (= assigned value)
- $sd$ = standard deviation calculated without outliers (= standard uncertainty)
- $cv percent$ = coefficient of variation (100 * $sd / mean$
  or 100 * $\sigma / \sigma$)

**Soil pH**

On the $X$ axis we have used the same scale for all four soil types, ranging from pH 5 to pH 8. The acidic soils (955 and KI) have most of the results on the left side and the alkaline soils (970 and LB) have most of the results on the right side, and in all figures it is easy to observe the outliers.

**Organic carbon by oxidation (Walkley & Black method)**

On the $X$ axis we have used the same scale for all four soil types, ranging from 0 to 6 percent of carbon. A large dispersion of the results can be observed for the KI sample.

**Organic carbon by combustion**

On the $X$ axis we have used the same scale for all four soil types, ranging from 0 to 8 percent of carbon. A large dispersion of the results can be observed for the 970 sample.

**Olsen-P method**

To have a better picture of the dispersion of the results, along the $X$ axis, different scales have been used depending on the soil type: 0 to 50 mg/kg for 955 and KI, 0 to 150 mg/kg for LB and 0 to 800 mg/kg for 970. This figure shows the large dispersion existing for all soil types.

**Bray 1-P & Bray 2-P**

To have a better picture of the dispersion of the results, along the $X$ axis, different scales have been used depending on the soil type. A very large dispersion can be observed for the KI and 970 sample when analysed for Bray 1-P and Bray 2-P methods, and for LB when using the Bray 2-P method.

**Exchangeable K (Ammonium Acetate method)**

To have a better picture of the dispersion of the results, along the $X$ axis, different scales have been used depending on the soil type. A very large dispersion can be observed for the 970 and LB sample.
Figure 25. Histograms of the distribution of all pH measures, determined on 1:2.5 soil: water suspension, overlaid by a model of normal distribution (bell curve).
Figure 26. Histograms presenting the distribution of all organic carbon content, determined using the Walkley and Black method (OC_WB), overlaid by a model of normal distribution (bell curve).

<table>
<thead>
<tr>
<th>Method</th>
<th>Sample Size</th>
<th>Median</th>
<th>MAD</th>
<th>Outliers</th>
<th>Mean</th>
<th>SD</th>
<th>CV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OC_WB</td>
<td>45</td>
<td>0.34</td>
<td>0.12</td>
<td>6</td>
<td>0.33</td>
<td>0.11</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>3.34</td>
<td>0.86</td>
<td>6</td>
<td>3.25</td>
<td>0.61</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.35</td>
<td>0.12</td>
<td>9</td>
<td>0.33</td>
<td>0.1</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>2.66</td>
<td>0.42</td>
<td>9</td>
<td>2.63</td>
<td>0.36</td>
<td>14</td>
</tr>
</tbody>
</table>
Figure 27: Histograms presenting the distribution of all organic carbon content, determined using the combustion method (OC_Comb), overlaid by a model of normal distribution (bell curve).
Figure 28. Histograms presenting the distribution of all available P content, determined using the Olsen method (P_Olsen), overlaid by a model of normal distribution (bell curve).
Figure 29. Histograms presenting the distribution of all available P content, determined using the Bray 1 method (P_Bray.1) overlaid by a model of normal distribution (bell curve).
Figure 30. Histograms presenting the distribution of all available P content, determined using the Bray 2 method ($P_{Bray.2}$) overlaid by a model of normal distribution (bell curve).
Figure 31. Histograms presenting the distribution of all exchangeable K content determined using ammonium acetate as an extractant (K_exch) overlaid by a model of normal distribution (bell curve).
The descriptive statistics for all parameters are summarised in Table 5.

This table shows that in SEALNET's case, using robust statistics was not necessary as they provided nearly the same information; the mean is close to the median and sd is close to MAD.

<table>
<thead>
<tr>
<th>Soil testing parameter</th>
<th>Soil name</th>
<th>n</th>
<th>Median</th>
<th>MADe</th>
<th>Outliers</th>
<th>mean</th>
<th>sd</th>
<th>cv (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>955</td>
<td>48</td>
<td>5.2</td>
<td>0.3</td>
<td>4</td>
<td>5.2</td>
<td>0.3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>970</td>
<td>48</td>
<td>7.2</td>
<td>0.2</td>
<td>0</td>
<td>7.3</td>
<td>0.2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>KI</td>
<td>64</td>
<td>5.4</td>
<td>0.3</td>
<td>6</td>
<td>5.5</td>
<td>0.3</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>LB</td>
<td>64</td>
<td>7.8</td>
<td>0.3</td>
<td>2</td>
<td>7.8</td>
<td>0.3</td>
<td>3</td>
</tr>
<tr>
<td>OC_WB</td>
<td>955</td>
<td>45</td>
<td>0.3</td>
<td>0.1</td>
<td>6</td>
<td>0.3</td>
<td>0.1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>970</td>
<td>45</td>
<td>3.3</td>
<td>0.9</td>
<td>6</td>
<td>3.3</td>
<td>0.6</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>KI</td>
<td>60</td>
<td>0.4</td>
<td>0.1</td>
<td>9</td>
<td>0.3</td>
<td>0.1</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>LB</td>
<td>60</td>
<td>2.7</td>
<td>0.4</td>
<td>9</td>
<td>2.6</td>
<td>0.4</td>
<td>14</td>
</tr>
<tr>
<td>OC_(Comb)</td>
<td>955</td>
<td>15</td>
<td>0.5</td>
<td>0.2</td>
<td>3</td>
<td>0.4</td>
<td>0.1</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>970</td>
<td>15</td>
<td>5.4</td>
<td>1.2</td>
<td>0</td>
<td>5.4</td>
<td>1.1</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>KI</td>
<td>20</td>
<td>0.4</td>
<td>0.2</td>
<td>1</td>
<td>0.5</td>
<td>0.1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>LB</td>
<td>20</td>
<td>3.2</td>
<td>0.5</td>
<td>4</td>
<td>3.0</td>
<td>0.2</td>
<td>8</td>
</tr>
<tr>
<td>POlsen</td>
<td>955</td>
<td>33</td>
<td>2.4</td>
<td>2.8</td>
<td>4</td>
<td>4.1</td>
<td>4.8</td>
<td>117</td>
</tr>
<tr>
<td></td>
<td>970</td>
<td>42</td>
<td>150.7</td>
<td>149.3</td>
<td>3</td>
<td>134.6</td>
<td>99.1</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>KI</td>
<td>45</td>
<td>2.3</td>
<td>2.1</td>
<td>5</td>
<td>4.4</td>
<td>4.4</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>LB</td>
<td>56</td>
<td>10.5</td>
<td>6.6</td>
<td>9</td>
<td>10.6</td>
<td>5.5</td>
<td>52</td>
</tr>
<tr>
<td>P Bray.1</td>
<td>955</td>
<td>20</td>
<td>2.0</td>
<td>0.7</td>
<td>2</td>
<td>1.9</td>
<td>0.6</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>970</td>
<td>15</td>
<td>119.9</td>
<td>43.5</td>
<td>0</td>
<td>103.2</td>
<td>42.3</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>KI</td>
<td>28</td>
<td>4.0</td>
<td>1.2</td>
<td>4</td>
<td>3.4</td>
<td>1.0</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>LB</td>
<td>19</td>
<td>8.9</td>
<td>5.8</td>
<td>4</td>
<td>7.3</td>
<td>3.8</td>
<td>53</td>
</tr>
<tr>
<td>P Bray.2</td>
<td>955</td>
<td>9</td>
<td>1.9</td>
<td>1.3</td>
<td>0</td>
<td>2.1</td>
<td>0.8</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>970</td>
<td>9</td>
<td>138.9</td>
<td>7.2</td>
<td>0</td>
<td>123.0</td>
<td>27.0</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>KI</td>
<td>12</td>
<td>3.0</td>
<td>0.4</td>
<td>0</td>
<td>3.0</td>
<td>0.7</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>LB</td>
<td>12</td>
<td>68.7</td>
<td>12.2</td>
<td>0</td>
<td>60.7</td>
<td>15.4</td>
<td>25</td>
</tr>
<tr>
<td>Exch K (NH₄OAc)</td>
<td>955</td>
<td>48</td>
<td>82.9</td>
<td>30.7</td>
<td>6</td>
<td>76.1</td>
<td>34.3</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>970</td>
<td>48</td>
<td>186.6</td>
<td>78.3</td>
<td>5</td>
<td>161.3</td>
<td>73.7</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>KI</td>
<td>62</td>
<td>16.0</td>
<td>10.2</td>
<td>8</td>
<td>20.3</td>
<td>15.4</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>LB</td>
<td>64</td>
<td>400.9</td>
<td>172.6</td>
<td>3</td>
<td>367.5</td>
<td>181.6</td>
<td>49</td>
</tr>
</tbody>
</table>
Moreover, large coefficients of variation (cv) were generally observed (except for pH) depending on the soil parameter. To make the cv easier to visualise and interpret, we present them as a bar plot (Figure 32). This figure shows that:

- soil pH has the lowest cv as mentioned; but it is important to note that pH corresponds to a log scale: a difference of one unit indicates a change of one order of magnitude in the [H+] concentration. Moreover, the range of soil pH is quite limited between approximately 4 to 9; thus a difference of one pH unit for a soil at pH=4 would lead to a cv of 25 percent, and at pH=9 to a cv of 11 percent. Consequently, the cv of pH measurement remains low by definition and it is not a relevant indicator to identify the variation of pH analysis between laboratories; for pH it is more relevant to look at the absolute range of variation.
- for OC Walkley & Black and combustion, the cv was generally <20 percent (ranging from approximately 15-30 percent). It is noteworthy that the cv’s were similar for OC_Walkley & Black (chemical oxidation) and OC_Combustion. This shows that **having high cost and high sensitivity instruments does not guarantee good results.**
  - for Bray1-P, Bray 2-P and exchangeable K, the cv’s were generally between 20 and 30 percent, but for exchangeable K of the KI sample, the cv went up to >70 percent.
  - Olsen-P has the highest cv, ranging from 50 to 117 percent.

It can be observed that when the concentration of the tested soil parameter was low, the cv was high (for example the OC content in samples 955 and KI). This is to be expected because as the analyte concentration in the soil gets closer to the detection limit, even small variations have a large impact on the final coefficient of variation. However, for Olsen-P, very high cv was observed for samples having high as well as low values. This indicates that there were large variations in results across laboratories for all soils.

![Coeff. of variation inter-laboratories for different analytical parameters & soil types](image)

Figure 32. Precision of each analytical method estimated from the coefficient of variation (cv) between replicates from the same soil (955, 970, KI, LB from left to right). Note that a low cv indicates a high precision and vice versa.
5. Report on laboratory precision

The Figure 33 presents the precision of each of the 16 participating laboratories for pH, OC-Walkley & Black, OC-combustion, Olsen-P and Exch K.

The precision represents the variability in the results they obtained when re-analysing the same sample in the same conditions. We provided three replicates for soils ‘955’ & ‘970’, and four replicates for soils ‘KI’ & ‘LB’. We expressed this variability as a coefficient of variation (cv in percent) around the mean value between the replicates that were calculated for each soil and each laboratory.

Note that:

- the lowest cv corresponds to the highest precision.
- precision is not related to accuracy: a laboratory can have a high precision even if it has a low accuracy.
- on the other hand, accuracy of the different replicates, can depend on the precision: if a laboratory has a low precision (i.e. high variability between replicates), some of its results will perhaps be accurate.

5.1 Precision estimated from single cv values (Figure 33)

For soil pH, the results showed that cv are low (except for lab 111); this means that variability between replicates is low and thus the precision is good.

For OC-Walkley & Black, approximately half of the laboratories have coefficients of variation <20 percent, i.e. these laboratories were able to re-analyse the same sample and provide results with the standard deviation between replicates representing less than 20 percent of the mean value. The other half of the laboratories returned cv>20 percent for at least one soil type and the highest cv that observed was around 60 percent.

For the OC-combustion method, only five laboratories made this analysis; four of them had cv<20 percent, and one lab had at least one cv>20 percent.

For Olsen-P it was found that: (i) most of the laboratories had high cv; and (ii) surprisingly the cv was not similar for the four different soil samples they analysed (some soil samples had high cvs when analysed by one lab but low cvs when analysed by another lab).

“For Exchangeable K it was found that (i) more than 50 percent of the participating laboratories had cv >20 percent and this happen to all soil types”
Figure 33. Laboratories' precision calculated from variations between analytical results obtained on the replicates of the same soil samples (3 rep. for 955 and 970, 4 rep. for KI and LB). Note that a low cv indicates a high precision and vice versa.
5.2 Precision estimated from mean-cv (Figure 34)

We also used these results to calculate a mean cv (for the four soil types) and standard deviation around the mean (Figure 34)

Soil pH: all laboratories (except one) have a very low cv (<10 percent), and this means a high precision. But we have to remember again that pH is a log scale and that a change of one unit corresponds to a change of one order of magnitude in [H+].

OC Walkley & Black: Eight laboratories have a mean cv<10 percent and this precision was the same whatever the soil type. Six laboratories have 10 percent<cv<20 percent with generally different precision for the different soil types (except lab 101 that presented the same precision for all soil types). One laboratory (111) had a mean cv>40 percent and large variation between soils (meaning that the cv of individual soils ranged from low to high).

OC combustion: Only five laboratories provided the data for this method, and as mentioned previously, four had low and similar cvs and one had high and variable cvs depending on the soil analysed.

Olsen-P: Only three laboratories had a low and homogeneous cv, whereas nine laboratories had cv>20 percent, and up to 80 percent for lab 104.

Bray 1-P and Bray 2-P (data not presented) also had large and similar variations.

K-exchangeable: About 11 laboratories had low and stable cv’s <20 percent, whereas the other labs had higher cv’s going up to 100 percent (lab 101, 104)
Figure 34. Laboratories’ mean precision calculated from the four coefficients of variation presented on Figure 33; error bars represent the standard deviation around the mean. Note that a low cv indicates a high precision and vice versa; short error bars indicates similar precision for the different soil types and vice versa.
6. General discussion

The SEALNET first inter-laboratory comparison had two objectives:

- Firstly to provide each laboratory with a tool to learn about their performance in terms of precision (are they able to replicate the same analysis and get similar results each time?) and accuracy (are they able to provide results that are similar to the consensus value?). In the case of poor performance, all possible reasons should be investigated: (i) to identify the most likely cause; and (ii) to enable action to be taken to prevent it happening again. As several participating laboratories are not used to analysing their results in such a way, in this discussion some examples of diagnosis and suggestions of action will be provided as examples.

- Secondly to provide a picture of the GSP reference laboratories is the Asia region in order to: (i) make a diagnosis of the current quality of the results provided at the regional scale and estimate if results provided by these laboratories can be put together to build a relevant picture of the soil resource at regional scale; (ii) identify the strengths and weaknesses of the SEALNET network; and (iii) make recommendations of actions for SEALNET to improve the quality of results in order to fit with the needs of soil management at the scale of the Asian Region.

6.1. Analysing the lab performance in terms of accuracy and precision

For the laboratory manager, the mean z-score is very useful because the results are presented in a graphical way that is easy to understand. The results can be interpreted to identify the possible problems and can be easily compared to the results of other laboratories. Moreover, the mean z-score can also be used to monitor the evolution of laboratories performances over time.

In this part of report, some poor results will be presented as examples to show different ways of identifying the possible causes.

The five measured parameters have to be considered separately:

**Soil pH** (Figure 18): Most of the laboratories are inside the satisfactory limits (z<2). Only laboratory 104 has two instances (for the soils 955 and KI) of producing outlier e-results (i.e., z>3). The scale for this figure (from z=-5 to +5) does not allow the exact value to be determined; but in Figure 25, the laboratory has provided values of pH>7 when the consensus value was around pH 5. It is unlikely that a mistake was made during the analysis and no calculation has to be made to report the results. It therefore seems most likely that a mistake was made when copying the results from a result logbook to a computer, or from a computer to the result sheet sent to FAO.

It is important to note that part of the challenge when participating in an inter-laboratory comparison is to correctly perform calculations and result transcription. It is a pity that correct analytical results are not sent to the PT provider, not only for the laboratory itself, but also for the network, as the number of results used to calculate the consensus value is decreased due to this mistake.

**OC Walkley & Black** (Figure 19): Laboratory 111 has provided accurate and precise results for the samples 955 and 970, but for the samples KI, the z-score was >3 with very high variation around the results. For sample LB, the mean z score was close to 0 but there was large variation around the mean. Again, it seems unlikely that such a mistake came from the analytical process as two soil samples were analysed very accurately. Again, it seems likely a mistake was made when copying the results (in Figure 5, a result of OC around 4 percent was provided for KI when the consensus value was 0.3 percent, and a result of 0.3 was provided for LB when the consensus value was around 3 percent.

Laboratory 113 had quite a low z score that was always inside the satisfactory limit. We can observe that all the z-scores are negative, suggesting that the distribution is not random, but it might have a bias; so it would be useful to check the different
steps and chemicals used in the lab to understand
the origin of the bias, and make the necessary
correction.

**OC combustion** (Figure 20): Five laboratories
provided results and only one (lab 112) had serious
problems with a z-score > 3 for samples 955
and LB. In Figure 6, the results were far from
the consensus value and from Figure 34 the
precision of this laboratory was not as good as in
the other laboratories using this same method.
Laboratory 112 also measured OC Walkley & Black
and obtained z-score close to 0, meaning they
had high accuracy. In addition, Figure 34 shows
they also had very good precision. Consequently
we observe that laboratory 112 had better results
using chemical oxidation than using a combustion
method with its associated high cost. This
suggests that expensive equipment does not
guarantee high quality results. In this case, the
competency of the staff who are managing the
instrument needs to be assessed to ensure that
they are operating the instrument correctly and
according to the method SOP.

**Olsen-P** (Figure 21): Two samples had very low
values close to detection limits, which explains
why some laboratories could not provide results.
Laboratory 104 had very high variability in results
in the warning limit (soil 955) and satisfactory limit
(soil LB) but these results had very high variations
(large error bars). This large variation results from
low precision (when analysing the same sample
several times, large variations were observed).
However, for samples 970 and KI the accuracy
and precision were quite good. Different reasons
can explain this lack of precision, and in this case,
perhaps the samples were not analysed at the
same time or by the same persons. Checking
the instrument operator logbook and sample
batch registration and reporting would allow this
possible source of variation to be investigated.

**Bray 1-P** (Figure 22): Only seven laboratories
provided the data for this method. The results from
laboratory 116 presented large variations between
replicates, which indicates a lack of precision. It is
noteworthy that laboratory 116 also used Olsen-P
method with much better precision. A detailed
analysis of this situation should be made by the lab
manager to understand why different precisions
were obtained when using different methods.

**Exchangeable K** (Figure 24): Several laboratories
had high z-scores, indicating low accuracy due to
analytical results that were too high. Laboratory
115 obtained high z-scores for all four soil samples.
This systemic error could result from using
analytical processes that were perhaps different
from the other laboratories. This aspect needs
to be investigated by comparing the lab’s SOP
for exchangeable K with the SOPs of the other
laboratories. However, it is also noteworthy that
a very high variability also existed between the
replicates, i.e. a low precision that cannot be
explained by the SOP.

**As a conclusion**, we observe two main problems:

- **the existence of very high mean z-scores**
  that could come from transcription errors
  (copying and reporting results that are not the
  ones obtained during the analysis). Note that
  we did not observe very low mean-z-scores;
  this is consistent for the soils with low analyte
  concentrations as there is a limit at ‘0’, but in
  soils with high analyte concentrations it is
  technically feasible to observe such extremely
  low z-scores. And if the z-score are not
  randomly distributed around the average
  value, it indicates the possible existence of a
  bias that should be identified.

- **the existence of large variability**
  between replicate analyses, i.e. low
  laboratory precision. This situation suggests
  that even if laboratories have standard
  operating procedures: (i) there are variations
  in the way the analyses are carried out; and (ii)
  the quality control procedures in place are not
  sufficient to detect and correct this problem.

6.2. Analysing the performance and
homogeneity of SEALNET

Even if figures such as Figures 25 to 31 are not
commonly presented in inter-laboratory reports,
they are the most useful tools for making a
diagnosis of the global performance.
Outliers: We have observed that, by definition, there is a large gap between the outliers and the samples inside the population. This means that outliers result not from small changes of procedures, but from large mistakes. In the previous section we have already mentioned the possible problem of faulty transcriptions. Another characteristic of the outliers is that most of them correspond to excessively high values, not to excessively low values. When all the faulty results are not randomly dispersed but are located on only one side of the correct data set, this suggests the existence of a systemic bias. With the limited information we have about the analytical conditions, it is not possible to suggest the cause(s) of this bias.

Large dispersion occurs between laboratories for Olsen-P (Figure. 28) with cvs of 52, 74, 100 & 117 percent for sample LB, 970, KI and 955 respectively, and exchangeable K (Figure. 31) with cvs of 49, 46, 76 and 45 percent. This can result from variations in procedures being used by the various laboratories, and makes it very difficult to compare results provided by different laboratories. Indeed, for available P, we can observe that this high variability is also observed in the laboratory precision. This means that even when individual laboratories replicate an analysis with their own operating procedures, they are not able to produce similar results. The large number of laboratories with this problem of precision, not only for available P but also for exchangeable K, confirms that the large variability between results comes not only from the existence of different SOPs, but from the absence of effective quality control procedures.

One could argue that using instruments with high technology would help solve this problem of precision. However, the observation has been made for laboratory 112 that low precision for OC was obtained for the combustion method, but high precision with the oxidation method. This demonstrates that having high cost and high sensitivity instruments does not guarantee reproducible results.

Conclusions and recommendations

The objective of GLOSOLAN and of SEALNET is to help soil laboratories to produce analytical results that can be compared, wherever the soil sample is analysed inside the Region. Results for the same analysis from different laboratories can only be confidently compared if the results provided by the different laboratories demonstrate a acceptable variation around the consensus value. To answer the question of comparability, we have to examine separately the four parameters we tested:

- for soil pH we have a low variability and results from different laboratories can be compared;
- for organic carbon, the variability is higher making the comparison more difficult; and
- for available phosphorus and exchangeable potassium, the variability is so large, that comparison of results across laboratories does not seem possible. For extractable phosphorus, even replicate results inside a single laboratory are often extremely variable.

To improve the situation, the adoption of standard operation procedures (SOPs) by all laboratories will not be sufficient. At the same time as discussion goes forward to adopt harmonised SOPs, it is also necessary to implement good laboratory practices and processes as well as the use of internal quality control and procedures to identify, and take action to rectify, bad scores.

The benefit of this implementation can be easily benchmarked and monitored using the same procedures as above and observing a decrease in coefficient of variation.

Moreover, it seems necessary to identify the laboratories having the best performance (in particular the best precision) to replicate the analysis on some soil samples to obtain ‘assigned values’ for these samples. It will then be necessary to define in advance a range of ‘fit for purpose’ indicators for assessment of laboratory QA/QC.
References


Appendix 1.
List of participating laboratories

<table>
<thead>
<tr>
<th>Country</th>
<th>Final address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bangladesh</td>
<td>Ms. Begum Samia Sultana Central Laboratory (CL), Soil Resource Development Institute (SRDI) Ministry of Agriculture, ‘Mrittika Bhaban’, Krishi Khamar Sarak, Dhaka-1215</td>
</tr>
<tr>
<td>Bhutan</td>
<td>Mr. Jamyang The Soil and Plant Analytical Laboratory (SPAL) National Soil Services Centre, Simtokha P.O.Box 907 Thimphu</td>
</tr>
<tr>
<td>Cambodia</td>
<td>Mr. Sun Sarak Soil and fertilizer Laboratory Address #54B/49F, Street 395-656 Toek Laak3, Tuol Ko Phnom Penh</td>
</tr>
<tr>
<td>India</td>
<td>Mr Ashok Kumar Patra ICAR-Indian Institute of Soil Science, Nabi Bagh, Berasia Road, Bhopal - 462038 Madhya Pradesh</td>
</tr>
<tr>
<td>Indonesia</td>
<td>Ms. Linca Anggria &amp; Ms. Lenita Herawati Indonesian Soil Research Institute Kampus Penelitian Pertanian Cimanggu Jl. Tentara Pelajar No. 12 Bogor 16114</td>
</tr>
<tr>
<td>Japan</td>
<td>Dr. Yuji Maejima 3-1-3 Kannondai, Tsukuba, Ibaraki 305-8604 Institute for Agro-Environmental Sciences, NARO</td>
</tr>
<tr>
<td>Lao PDR</td>
<td>Mr. Xaysatith Souliyavongsa Soil laboratory Unit. Department of agriculture land management (DALaM) Ministry of agriculture and forestry Nongviengkham Village Xaitany district, Vientiane capital</td>
</tr>
<tr>
<td>Malaysia</td>
<td>Mr. Abd Razak Bin Abu Samah Analytical Service Section Soil Resource Management and Conservation Department of Agriculture Ikm No: M/2586/5277/08 Jalan Sultan Salahuddin 50632 Kuala Lumpur</td>
</tr>
<tr>
<td>Myanmar</td>
<td>Ms. Su Su Win Soil And Plant Analysis Laboratory Department of Agricultural Research Yezin, Nay Pyi Taw</td>
</tr>
<tr>
<td>Nepal</td>
<td>Mr. Janardan Khadka, PhD Soil Management Directorate Ministry of Agriculture Development Hariharbhaban, Lalitpur</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country</th>
<th>Final address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pakistan</td>
<td>Mr. Arshad Ali Land Resources Research Institute (NARC) National Agricultural Research Centre, Park Road, Islamabad-Pakistan 44000 Islamabad</td>
</tr>
<tr>
<td>Philippines</td>
<td>Ms. Gina P. Nilo Bureau Of Soils And Water Management Laboratory Services Division Elliptical Road Corner Visayas Ave. Diliiman, Quezon City</td>
</tr>
<tr>
<td>Republic of Korea</td>
<td>Dr. Chang Hoon Lee 166, Nongsengmyeong-ro, Iseo-myeon, Wanjungu, Jeollabuk-do 55365, Republic of Korea Soil and Fertilizer Devisio(room 615), Department of Agricultural Science</td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>Mr. Priyantha Weerasinghe Horticultural Crops Research and Development Institute Gannorouwa Peradeniya 20400</td>
</tr>
<tr>
<td>Thailand</td>
<td>Mr. Rattanachart Chuaybuidda Office of Science for Land Development Land Development Department 2003/61 Phahonyothin Rd. Land Yao, Chatuchak, Bangkok 10900</td>
</tr>
<tr>
<td>Vietnam</td>
<td>Mr. Do Duy Phai Central Analytical Laboratory - Soils and Fertilizers Research Institute (SFRI) Le Van Hien, Duc Thang, Bac Tu Liem Ha Noi</td>
</tr>
</tbody>
</table>

**Excused countries:**

- **Mongolia**, who participated in the inter-laboratory comparison exercise but was unable to submit their results by the provided deadline.
- **China**, who was willing to participate in the inter-laboratory comparison exercise but could not import soil samples.
The Global Soil Partnership (GSP) is a globally recognized mechanism established in 2012. Our mission is to position soils in the Global Agenda through collective action. Our key objectives are to promote Sustainable Soil Management (SSM) and improve soil governance to guarantee healthy and productive soils, and support the provision of essential ecosystem services towards food security and improved nutrition, climate change adaptation and mitigation, and sustainable development.

GLOSOLAN is a Global Soil Laboratory Network which aims to harmonize soil analysis methods and data so that soil information is comparable and interpretable across laboratories, countries and regions. Established in 2017, it facilitates networking and capacity development through cooperation and information sharing between soil laboratories with different levels of experience. Joining GLOSOLAN is a unique opportunity to invest in quality soil laboratory data for a sustainable and food secure world.

Thanks to the financial support of