bibtex ✓

Defining Ranges of Soil Characteristics[1]

via "tolerance intervals"

I. J. Jansen and R. W. Arnold[2]

ABSTRACT

A method is developed for using quantitative sample data to characterize mappable bodies of soil. Inclusions are separated into two kinds. The first consists of included bodies of soil that are recognized as separate entities, but not mappable because of scale limitations. The second consists of scattered areas having extreme expression of some significant property but not recognized as separate geographic entities.

Inclusions of the first kind are sampled separately and treated as separate entities. The location and setting of these inclusions can be described in relation to the dominant soil body. A range of expression for each measurable soil property of interest is determined separately for each soil entity. The ranges are derived from the sample data and are designed to include a fixed portion of the total population of values rather than to be all inclusive. Soil areas having extreme expression of some particular property (falling outside of the defined range) are inclusions of the second kind. The location and setting of those soil areas cannot be readily described but their extent is controlled.

The tolerance interval procedure is used to calculate range limits from sample data when a normal distribution can be assumed. The resulting range can be said to include a predetermined portion of the total population of values with a specified degree of confidence. When the distribution is believed to deviate markedly from normal, the range limits are set subjectively by observing a frequency distribution of the sample data. Ranges set by the tolerance interval procedure were generally somewhat wider than those set subjectively, because the tolerance interval procedure allows for the imperfect relationship between the sample and the population and enables one to make confidence statements about the ranges defined.

Additional Index Words: **characterizing soils, inclusions, mapping unit purity, soil variability, tolerance interval, soil mapping.**

RUDEFORTH (1969) suggested a quantitative soil survey method involving the following steps: (i) soil characteristics are observed and described in a pattern unrelated to any delineations, (ii) readily identifiable "land units" are delineated on a map, and (iii) data collected in step one are used to characterize the land units delineated in step two. The land units envisioned by Rudeforth apparently were mapping units delineated by landscape interpretation. Wilding et al. (1964) characterized some Ohio soil series by statistical interpretation of available morphological, physical, and chemical data. They used several parameters to express central tendencies and limits of variation.

Mostly- Blank

This paper presents and applies a method for using quantitative sample data to characterize soils. The method is useful for establishing new mapping unit concepts and for evaluating presently established ones. Soils are characterized by defining a range of expression for each property. The range is designed to include a controlled portion of the total variation rather than to be all inclusive. The basic soil entities for which ranges are defined are conceptual soil landscape units. They are defined primarily in terms of landscape parameters and soil genesis theory. Soil mapping units would in most instances include more than one soil landscape unit.

portion of the range

There is some divergence in the way ranges of soil characteristics are interpreted and used. A soil taxonomist uses the ranges of characteristics of a class to determine which pedons are included in and which are excluded from that class. To him a range of characteristics establishes class limits, and his primary consideration in defining ranges of characteristics is creating an effective taxonomic system.

A soil survey user uses the ranges of soil characteristics defined by soil taxonomists to evaluate the properties of identifiable soil bodies. To him a range of characteristics is most useful if it is designed to give the maximum amount of information about the properties of soil bodies that he can locate on the landscape.

Two views of ranges

The mapping units of soil maps are commonly characterized by assigning the names of one or more soil series to each mapping unit. The characteristics of each soil series are presented in written descriptions and in tables. The mapping units always have inclusions of soil other than the named soil or soils. Inclusions are not considered to be a

[1]Contribution from Dep. of Agronomy, Cornell Univ., Ithaca, N.Y. 14850. Agronomy paper no. 1115. Presented in part before Div. S-5, Soil Sci. Soc. Am. in Chicago, Ill., 14 Nov. 1974. Received 21 April 1975. Approved 19 Sept. 1975.

[2]Former Graduate Assistant at Cornell Univ., now Assistant Professor of Pedology at the Univ. of Illinois, Urbana and Associate Professor of Soil Sci. Cornell Univ., Ithaca, N.Y., respectively.

✓

2

serious problem unless they are highly contrasting or excessive in extent. The method proposed here is intended to improve control over inclusions and to more effectively characterize identifiable bodies of soil for land users and planners.

Consider arbitrarily separating mapping unit inclusions into two different kinds. The first consists of included bodies of soil that are recognized as being genetically different. These included bodies are identifiable geographic bodies. Their location within delineated soil bodies can be described so that soil survey users could feasibly identify and locate them even though they are not delineated separately. They could be mapped out separately at some larger scale but are too intricately intermingled with the dominant soil body to be separated at the chosen mapping scale. A geographic pattern is recognized, but can't be mapped because of scale limitations. These included bodies are considered to be part of a separate soil landscape unit and are to be characterized separately. An example of this kind of inclusion would be an area of soils too small to delineate separately, having a concave surface, developed in erosional sediments, and located along a small upland drainage way which passes through a delineated area where the dominant soils have a convex surface and have developed in glacial till. The soils of such an included area would be expected to differ considerably in the expression of several properties from the dominant soil body.

The second kind of inclusion consists of statistical extremes in the expression of soil properties within bodies of soil that are conceptually homogeneous. No coherent geographic pattern is recognized. They are not characterized separately, but their extent is controlled on an individual property basis. These areas are not considered to be separate entities. They can not or will not be located or singled out for separate treatment. The defined ranges of characteristics are intended primarily to give information about the properties of a soil entity rather than to create the soil entity by defining its boundary conditions.

ranges are for informative purposes

METHODS

This method to define range limits is intended to yield the maximum amount of information about the character of identifiable soil bodies. Limits are set by sampling the set of soil bodies to be characterized and analyzing the sample data. Inclusions of the first type are excluded by intentionally avoiding those areas when the sample is drawn; they are conceptually different, hence should be sampled separately. Inclusions of the second type are determined by using a statistical procedure that will set limits to include only a fixed portion of the total population of values.

not class definition or membership

Table 1—Tolerance factors (values of K) as related to sample size (n), confidence level (γ), and portion (P) of the population to be included in the tolerance interval. These values are rounded to two decimal places from values complied by Eisenhart et al. (1947).

	$\gamma = 0.75$		$\gamma = 0.90$	
n	$P = 0.75$	$P = 0.90$	$P = 0.75$	$P = 0.90$
5	1.83	2.60	2.45	3.49
10	1.49	2.13	1.78	2.54
15	1.40	1.99	1.59	2.28
20	1.35	1.93	1.51	2.15
25	1.32	1.88	1.45	2.08
50	1.26	1.79	1.34	1.92
75	1.23	1.76	1.30	1.86
100	1.22	1.74	1.28	1.82
120	1.21	1.73	1.26	1.80

The method was applied, using data from a small study tract near Ithaca, New York. A grid was superimposed on a tract 64 m^2 to divide it into cells, each cell being 4 m on each side. A soil core (soil section) from the center of each cell was described and sampled. Particle size distribution of the subsoil and organic carbon of the surface soil were determined in the laboratory for the core representing each cell. All data were punched on cards for manipulation by computer.

Two conceptual soil entities (soil landscape units) were defined for the study tract in terms of landscape parameters and soil genesis theory. Current soil series limits and other taxonomic constraints were not a factor. One soil landscape unit includes areas that have a convex surface form, are higher in elevation than those of the other unit, and have soils developed mostly in lacustrine sediments. The other soil landscape unit consists of lower concave areas and has soils developed mostly in a mantle of post-glacial erosional sediments overlying the lacustrine sediments. Soils of the high convex soil landscape unit are primarily Glossoboric Hapludalfs and those of the low concave unit are primarily Acric Ochraqualfs. Each of the grid cells was considered to be entirely within one or the other of the landscape units. The cells were grouped according to the soil landscape unit within which they were located. There were 120 cells in the low, concave unit and 136 cells in the high convex unit.

Tolerance limits

The tolerance interval procedure was the basic statistical procedure used to set ranges of soil characteristics to characterize the soils of each soil landscape unit. A tolerance interval is an interval which includes a fixed portion of a population of values with a specified degree of confidence (Wine, 1964). Tolerance limits are the boundary values which define a tolerance interval. Tolerance intervals should not be confused with confidence intervals. The former makes a confidence statement about the portion of the population that is included in the interval. The latter makes a confidence statement about whether or not the true value of some parameter is within the interval.

typically distribution-specific

Everson et al. (1965) present a detailed discussion of the tolerance interval concept. Briefly, for a normal population with a true mean (μ) and a true standard deviation (σ) upper and lower tolerance limits are calculated with the formulas $\mu + K\sigma$ and $\mu - K\sigma$, respectively. The value K is a constant that varies in size according to the portion (P) of the population to be included within the limits.

see notation

In application of the procedure, μ and σ must be estimated by $\overline{X}$ and s, respectively. Since these parameters are estimated, a confidence statement must be made about the tolerance interval calculated. The K values used with the estimated parameters are dependent on the sample size (n) and the confidence level (γ) chosen as well as on the portion of the population to be included. Table 1 provides tolerance factors for a few sample sizes. A more extensive table of tolerance factors has been compiled by Eisenhart et al. (1947, p. 102–107).

Using data from Table 2 and a tolerance factor from Table 1, the following calculations set tolerance limits ($\gamma = 0.90$, $P = 0.75$, and $n = 120$) for surface soil organic matter content of the low, concave soil landscape unit:

$$\overline{X} \pm Ks = 5.96 \pm (1.26)(0.67) \begin{cases} 6.8 \\ 5.1 \end{cases}$$

" 90% conf. That 75% samples in interval "

Table 2—The mean and standard deviation for each of four characteristics of soils of the two soil landscape units studied.

	Low, concave unit		High, convex unit	
Characteristic	$\overline{X}$	s	$\overline{X}$	s
Surface soil thickness	33.03	11.41	25.66	5.29
Surface soil organic matter content	5.96	0.67	4.76	0.63
Percent sand in the subsoil	14.40	12.91	4.74	5.53
Percent clay in the subsoil	31.37	7.91	35.64	4.83

Values outside real ranges are possible

distributional assumptions

This indicates that 75% or more of the population of soil sections in the low, concave soil landscape unit should be expected to have a surface soil organic matter content between 5.1 and 6.8%. If a large number of samples were drawn to determine 75% tolerance limits at the 90% confidence level, 9 out of 10 should be expected to yield limits that do include 75% or more of the population.

It is necessary to assume a normal distribution for the tolerance interval procedure to be valid. For most soils it is reasonable to assume a normal distribution for many properties. Yet for some soil properties the distribution may not be expected to be normal.

A computer program was written that plotted a frequency distribution and calculated $\bar{X}$ and s for each property evaluated within each soil landscape unit. In instances where it is reasonable to assume a normal distribution, $\bar{X}$ and s can be used to calculate tolerance limits. For characteristics where there is reason to believe that the distribution deviates markedly from normal, limits for the range of that characteristic can be chosen subjectively by observing the frequency distribution. By the subjective procedure limits are selected to bracket the mode revealed by the frequency distribution, excluding the desired number of extreme values. If one wanted the range to include about 90% of the soil landscape unit, limits would be chosen to include about 90% of the observations. This subjective procedure yields only an approximation. There is no assurance that 90% of the soil landscape unit will in fact fall within the range. The tolerance interval procedure presented above does provide assurance at the chosen confidence level that the range includes the desired portion of the soil landscape unit. In this particular study both procedures were used to set ranges for four soil characteristics.

use of mode

RESULTS AND DISCUSSION

Ranges set by the tolerance interval procedure were generally wider than those set subjectively (Table 3). This is as expected because the tolerance interval procedure allows for the fact that the data are from a sample and enables one to make confidence statements about the range defined for the population as a whole. No confidence statements can be made about the subjective ranges.

It should be emphasized that all statements about the portion of a soil landscape unit that falls within a tolerance interval are on an individual property basis. The portion of the soil landscape unit that should be expected to be within all ranges defined would be much smaller and would depend on the number of properties evaluated and the degree of independence of those properties from each other.

data-space constraints

The negative lower limits for the subsoil sand content tolerance intervals in Table 3 can be attributed to a distribution of sand content values that is severely skewed and truncated at zero. The subjective limits would be more appropriate in such instances for characterizing a soil landscape unit.

A soil landscape unit concept is developed here as the basic unit to be characterized because the objective is to define ranges to characterize identifiable soil bodies in a meaningful way. Characterizing soil landscape units rather than series frees one from the constraints built into the taxonomic system so that ranges of characteristics can be fitted to identifiable bodies on the landscape. Series names could be assigned to the soil landscape unit much as is now done for mapping units. Series ranges would be of interest to the soil taxonomist, but soil survey users would use the ranges defined for soil landscape units to determine the properties of soil bodies that they could identify with the help of the soil map and the written text.

The emphasis in soil mapping should be to separate, either cartographically on the map or verbally in the text, landscape units that are as uniform in significant soil properties as is feasible considering the given landscape and survey intensity. By the method proposed in this paper, the emphasis in characterizing soils for soil survey users is on defining ranges of characteristics to fit those soil landscape units.

* This procedure is also potentially useful for developing and refining series concepts. It is highly unlikely that mutually exclusive classes could be defined by this procedure alone and constraints imposed by the taxonomic system must be considered. However, not all properties have constraints imposed by the taxonomy, and two classes only need to differ in the expression of one property to be mutually exclusive. Hence, even when some range limits are arbitrarily imposed to achieve some taxonomic objective, the ranges for other characteristics of that soil could be set by analysis of sample data.

The sampling pattern must be random in relation to the distribution of soil properties within the conceptual soil entities being characterized. Existing data from typifying pedons would not be suitable for this purpose, because such pedons are selected on the basis of their profile characteristics.

critical! non-independent samples ⇓ bias ⇓ lower effective DF

The grid pattern used in this study was selected because one of the objectives of the project was to search for alternative geographic patterns within a small segment of landscape. That portion of the project will be reported on separately. A simple grid sampling pattern is appropriate for

Table 3—Alternative sets of range limits determined by the tolerance interval procedure and by subjective evaluation of frequency distributions (P = portion of the total population to be included in the interval).

	P = 90%		P = 75%	
	Low-concave landscape unit	High-convex landscape unit	Low-concave landscape unit	High-convex landscape unit
Surface soil thickness (cm)				
Tolerance limits*	12–54	16–35	19–47	19–32
Subjective limits†	20–50	18–31	23–39	20–28
Surface soil organic matter content (%)				
Tolerance limits*	4.8–7.2	3.6–5.9	5.1–6.8	4.0–5.5
Subjective limits†	5.0–7.0	4.0–6.0	5.2–6.5	4.0–5.1
Sand in the subsoil (%)				
Tolerance limits*	–8.8–37.6	–5.2–14.7	–1.8–30.6	–2.2–11.7
Subjective limits†	1.5–35.5	0.0–10.5	1.5–22.5	1.5– 5.0
Clay in the subsoil (%)				
Tolerance limits*	17.1–45.6	26.9–44.3	21.4–41.3	29.6–41.7
Subjective limits†	17.5–44.5	28.0–43.5	21.5–38.5	30.0–41.0

* Confidence level = 0.90, two-sided. † No confidence statements possible.

application of the tolerance interval procedure but does have some disadvantages. Some soil landscape units may have more observations than needed and others may have too few. Also, the ranges defined can be applied with confidence only to the area covered by the grid.

A stratified sampling structure that is randomized at each level, such as that used by McCormack and Wilding (1969), should be considered when the objective is to characterize soil landscape units that are distributed over a large geographic area. Such a pattern would not be independent of map delineations as specified by Rudeforth (1969), hence the data could not be used effectively to characterize the mapping units of a later or different map. Fewer total observations would be needed with a stratified sampling design, however. The actual number of observations needed for each unit depends on inherent soil variability and the level of precision desired.

The decision concerning what constitutes an inclusion of the first kind, and thus can be sampled separately, is an arbitrary one. Care must be exercised not to use this as just a device to discard a few extreme values. Generally, observations should not be excluded only because of extreme expression of one or more properties. Inclusions of the first kind should be identifiable geographic bodies that can be characterized separately and whose setting can be described in relation to the dominant soil. They should be areas believed to differ in the historical expression of one or more genetic factors. They might be considered genetic inclusions.

The pedon would be the appropriate sampling unit. In practice, a preliminary study should be made to determine whether or not there is significant variation between soil sections with pedons. If significant lateral variation is observed within pedons one random soil section should not be considered to adequately represent a pedon. Each observation (sampling unit) should then be a composite or mean from more than one soil section within a pedon, because considering each soil section to be representative of a pedon would result in a biased estimate of the standard deviation for some soil properties and resulting tolerance intervals would be too wide.

In this approach to defining property ranges, realism is given priority. Interpretive considerations might make it desirable in some instances to have range limits somewhat different from those which result from applying the tolerance interval procedure. Changing the range limits arbitrarily without changing the geographic soil entities mapped or described in the text would only be deceptive. A better alternative would be to conduct a soil landscape study to determine whether a different set of geographic soil entities (soil landscape units) having the property differences desired could be consistently separated, either cartographically on the map or verbally in the text. If it proves to be feasible to define a more significant set of soil landscape units, the new units could then be characterized by the tolerance interval procedure.

Calculated values are useful starting point

ACKNOWLEDGMENTS

The authors wish to acknowledge the contribution of Dr. F. B. Cady, professor of statistics and biometrics at Cornell University, who suggested the tolerance interval procedure for this application.

LITERATURE CITED

1. Eisenhart, C. M., M. W. Hastay, and W. A. Wallis. 1947. Techniques of statistical analysis. McGraw-Hill, New York. 473 p.
2. Everson, L. E., C. S. Shih, and F. B. Cady. 1965. A comparison of the uniform blowing and hand methods for the purity analysis of *Poa pratensis* seed. Proc. Int. Seed Test. Assoc. 30:493–511.
3. McCormack, D. E., and L. P. Wilding. 1969. Variation of soil properties within mapping units of soils with contrasting substrata in northwestern Ohio. Soil Sci. Soc. Am. Proc. 33:587–593.
4. Rudeforth, C. C. 1969. Quantitative soil surveying. Welsh Soils Discussion Group. 10:42–48.
5. Wilding, L. P., George M. Schafer, and R. B. Jones. 1964. Morley and Blount Soils: A statistical summary of certain physical and chemical properties of some selected profiles from Ohio. Soil Sci. Soc. Am. Proc. 28:674–679.
6. Wine, L. R. 1964. Statistics for scientists and engineers. Prentice-Hall, Englewood Cliffs, N.J. 671 p.