PREPRINT – Nimmo, J.R., 2004, Aggregation: Physical Aspects, *in* Hillel, D., ed., Encyclopedia of Soils in the Environment: London, Academic Press.

Aggregation: Physical Aspects

J. R. Nimmo, U.S. Geological Survey, Menlo Park, CA 94025, USA

Key Words: soil structure, aggregation, soil strength, fractals, erosion, soil hydraulic properties, tillage, soil compaction.

A soil aggregate is "a group of primary soil particles that cohere to each other more strongly than to other surrounding particles." Soil aggregates form through the combined action of aggregation and fragmentation processes. That is, attractive and disruptive forces act on the particles in the soil to cause greater cohesion among some particles, and groups of particles, than others. Most soils break up naturally into some form of aggregates, as pictured in Figure 1. Important physical aspects of aggregates include their size, density, stability, structure, and their effect on the transport of fluids, solutes, colloids, and heat.

Figure 1. Soil with aggregates partially separated, in a tray. Photograph from the Historic Russian Soil Collection of the St. Petersburg Academy of Forestry, provided by Jennifer Harden.

The analysis of soil aggregation is important in a variety of applications. Aggregation is a major influence on the growth and effectiveness of roots. Aggregate stability and size information may be used to evaluate or predict the effects of various agricultural techniques, such as tillage or addition of organic matter. Aggregate analysis is often used in experiments where various tillage methods are applied and then evaluated by examining the stable aggregates that result. Because of their direct relation to cohesive forces, aggregate size and stability are important to the understanding of soil erosion and surface sealing. Analysis of dry aggregates is logically related to wind erosion effects while wet analysis may be more appropriate to evaluate or predict erosion due to rainfall impact and runoff. The stability of wet aggregates can be related to surface seal development and field infiltration, as water-stable cohesion among particles may lead to restriction of water entry and formation of surface seals. Through these erosion and sealing effects, as well as the relation between aggregation and structural features such as macropores, aggregate analysis may increase the understanding of most aspects of soil water behavior, including runoff, infiltration, and redistribution, as well as soil aeration. Increasingly, aggregate properties are used in models that predict soil hydraulic properties, including water retention and unsaturated hydraulic conductivity.

Closely related terms include ped, clod, and crumb. A ped is an aggregated unit representative of the innate structural classification of the soil. It has a characteristic shape related to structural designations such as prismatic, columnar, and blocky. The term clod applies to an aggregate separated from the bulk soil by artificial means such as digging or plowing. "Crumb" is an archaic term referring to an aggregate less than about 5 mm in diameter.

Forces on soil particles

The strength of interparticle cohesion depends on a variety of soil physical, chemical, and biological influences. Some of the most important of these are air-water surface tension, intermolecular attractive forces between water and solids, cementation by precipitated solutes, entanglement by roots and fungal hyphae, and various chemical phenomena. The forces of soil cohesion depend strongly on water content and other conditions.

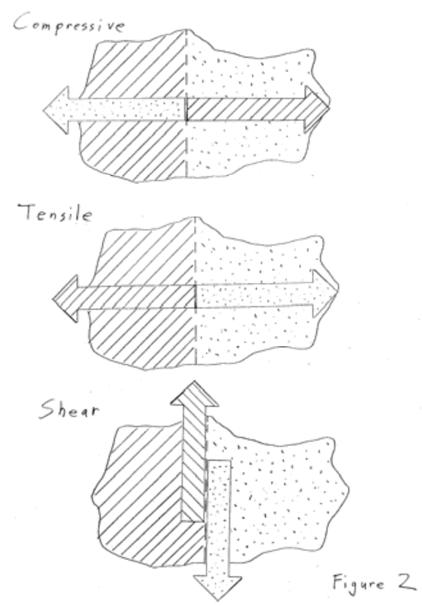

Figure 2. Types of stresses on an aggregate. Stresses are defined with respect to a selected plane, shown as the dashed line in this cross-sectional diagram. The two types of shading are arbitrary, merely designating portions of the aggregate on the two sides of the plane. Each arrow indicates the force acting on the portion of the aggregate with the same type of shading, at the selected plane.

Fundamentally, the forces of aggregation can be considered as stresses, that is, force per unit area acting on a given cross-sectional plane within the aggregate. These categorize as compressive, tensile, and shear stresses (Figure 2). Compressive stresses push particles closer together, as for example by the weight of soil above a given horizontal plane. Tensile stresses pull apart, like forces from soil shrinkage. Shear stresses act along a plane parallel to the direction of force, as in an aggregate at the edge of a zone of compaction. Tensile and shear stresses tend to disrupt aggregates. Compressive stresses tend to consolidate aggregates, except that when they are uneven across a plane, they lead to shear stresses that disrupt.

Several influences act to hold soil particles together. Water in the soil does this through surface tension, and additionally through the attractiveness of water molecules for soil solids and for each other. Dissolved ions are important, especially in terms of the electric double layer. The tendency of soil particles to have a negative surface charge means that water close to them is rich in positive ions, which in turn attracts other particles, in a process of flocculation. Because clay particles are especially sensitive to flocculating influences, higher clay content of a soil generally makes for more aggregation. Chemicals that precipitate or otherwise turn into cementing agents also enhance aggregation. Typical cementing substances include calcium carbonate, humus, and oxides or silicon, iron, and aluminum. Various other chemicals, especially certain organic compounds such as polysaccharides, attract soil particles. Some organic materials exert forces through surface tension or electrical charge; others, like roots and fungal hyphae, adhere to soil as part of their natural function. Because aggregation in general is favorable to plant growth, an evolved characteristic of plants is that they generate decay products that promote aggregation. A major part of this influence is to promote aggregation. Bacteria and other microorganisms contribute similarly to aggregation. Organic material artificially added to the soil for the purpose of increasing aggregation is usually far less effective, per unit mass, than organic material naturally present. Not just the type and quantity of organic compounds but also their microscale distribution are critically important. Typically particles within an aggregate may be held together by a sort of glue made up of water, clay, and organic materials (Figure 3).

There is a similar variety of mechanisms that pull soil particles apart, discouraging aggregation, either directly or by a decrease of attractive force. Some of the most common are associated with the addition of water. The breakup of aggregates that results from this, especially from sudden immersion, is called slaking. Increased water content can dissolve cementing precipitates and can decrease flocculation while the resultant dilution weakens the effects of electric double layers. As water infiltrates an aggregate, the expansion of trapped air, as well as the release of adsorbed air from newly wetted surfaces, can generate substantial disruptive force. Other disruptive mechanisms include the expansion of water upon freezing, impacts of rain or falling objects, and vibrations—either natural, or artificial like ultrasound or jostling on a sieve. Mechanisms associated mainly with compressive force disrupt by the generation of shear stresses. Examples include foot or wheel traffic, which is always to some degree uneven across the land surface, and gravity acting on an uneven mass distribution of soil or on an aggregate unevenly supported from below.

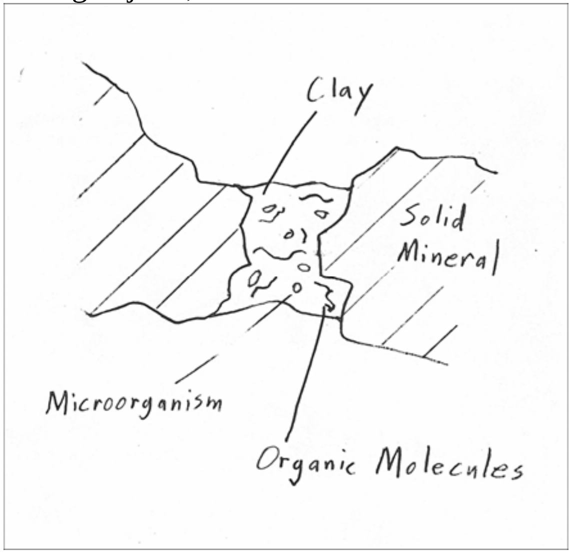

Figure 3. Diagram of the microscopic region between solid grains within an aggregate.

Aggregates become less stable with increase in size. This generalization applies within a given soil, and should not be confused with the idea that soils forming larger aggregates have greater aggregate stability. One reason large aggregates are less stable is simply that interparticle forces vary, and the bigger the aggregate, the greater likelihood that it contains a planelike region of low tensile strength where it breaks in response to stress. Similarly, the bigger the aggregate, the greater likelihood that it contains an expanding root or other agent that breaks it apart. Some disruptive stresses increase as the size of an aggregate increases, for example its weight increases with size. In general, attractive forces (cementing, intermolecular attraction, etc.) are predominantly short-range, whereas disruptive forces (which are mechanically transmitted through the soil fabric) are predominantly longer-range. Thus for a given soil the balance of forces within an aggregate increasingly favors disruption as larger aggregates are considered. This fundamental linkage between aggregate size and aggregate stability is a crucial factor in virtually any assessment of soil aggregation.

Whether a given sample of soil tends toward relatively large or relatively small aggregates depends chiefly on the interparticle attractive forces, which depend largely on texture and organic matter content. Attractive forces vary more from soil to soil than do disruptive effects like gravity and surface traffic. Therefore they dominate the issue of how the attractive and repulsive forces balance out.

The soil's characteristic response to shear stresses—the extent to which it undergoes plastic as opposed to brittle deformation—is also important. Both interparticle attraction and plasticity depend strongly on soil texture and the composition of the soil-plant-water system.

The aggregates of a fine-textured soil with much organic matter are likely to be larger than those of a sandy soil. In fact the net interparticle attractive force in a sand could easily be so small that the characteristic aggregate size it would indicate is smaller than the individual soil particles, so in fact aggregates do not occur.

Aggregate Physical Properties and their Measurement

Basic properties

Some physical properties of aggregates can be determined directly, especially where it is possible to physically isolate individual aggregates. Aggregate size and shape can be determined optically, by comparison with a ruled grid or by analysis of digital images. Because aggregates have irregular shape, their size cannot be indicated by a single linear dimension. Thus a choice must be made whether to indicate size by greatest dimension, average dimension, diameter of an equivalent sphere, or some other measure. The volume and bulk density of an aggregate can be measured by the clod method, using the aggregate's weight in air and in a liquid of known density, after coating it to prevent liquid intrusion. Alternatively, to reduce or eliminate the need for coating, a fine granular material of known bulk density may be used instead of a liquid. The strength of an aggregate can be measured, at least operationally, by breaking it with a known mechanical force applied by impact or by gradual increase in magnitude.

In most applications, attributes such as size, density, and strength are important as characterizations of the bulk soil, rather than of particular aggregates. Then if a method employing measurements on individual aggregates is chosen, it must be applied to enough aggregates to establish adequate confidence in representative property values computed by statistical techniques. Alternatively, many methods are available that can be applied to aggregates in bulk. Some of these methods are described in the next section.

Soil within an aggregate may be more homogeneous than within a greater volume of soil, but like any body of soil, it is not perfectly homogeneous. At one extreme, it might have a monolithic character not readily subdivided into units larger than individual particles. Alternatively, an aggregate may comprise smaller aggregates held by greater forces within themselves than between each other. In this way, each subdivision of aggregates may comprise a smaller subdivision of aggregates, down to a limit as the subdivided aggregates approach the size of particles. This sort of structure has led to proposed fractal models for the structure of individual aggregates. Discussions below, on the density dependence of aggregate size, and on mathematical representations of aggregate size distribution, explore this issue further.

The density of aggregates and how it correlates with aggregate size can provide evidence concerning the nature of soil structure, and data for predicting or correlating with other soil properties. Figure 4 shows an example of such data for soils of three different textures. For the two finer-textured soils, the smaller aggregates have greater density, which more closely approximates the particle density of soil minerals and indicates a tighter, more compact, and probably more stable structure. The material labeled "sand", (actually "quartz sand with pebbles" in the original reference) has an aggregate density that depends little on size and differs little from the particle density of pure quartz, about 2.65 g/cm^3. This indicates essentially no aggregation, except possibly for some small aggregates (about 0.1 mm in diameter) of the smallest particles.

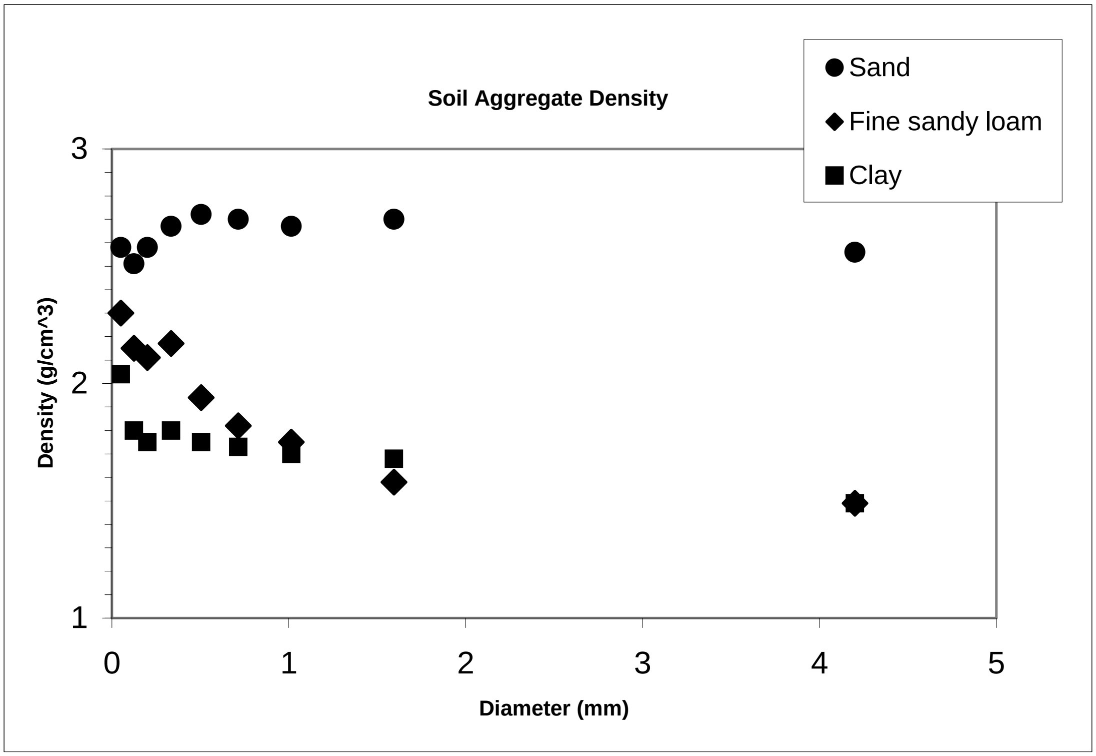

Figure 4. Measured aggregate density as a function of mean aggregate size for soils of three textures. The sample labeled “Sand” contained pebbles as well as quartz sand. Data are from Chepil, W.S., 1950, Methods of estimating apparent density of discrete soil grains and aggregates, Soil Science, v. 70, p. 351-362.

Size and Stability

Fundamentals

The most common way to characterize the size distribution or stability of the soil’s aggregates is with measurements on a coherent volume of soil containing a representative number of aggregates, but the characterization is complicated by the interrelationships of aggregate properties. Especially important is that aggregate size and net cohesive force are conceptually inseparable. Measured sizes depend on the disruptive force applied to separate the aggregates, so force and size cannot be measured independently. For example, size determination by sieving cannot be done without the disruptive force of collisions between the aggregates and the sieve. Methods tend to be called stability methods or size methods depending on which of these gets more emphasis. This difference may be in the technique itself, for example in a size method that relies on a specified disruptive force, or a stability method that relies on the effect of force on a given size of aggregates. Alternatively, the difference may be in the interpretation, for example in the presentation of sieving-derived data as a distribution function to indicate the relative abundance of aggregate sizes, or as a single index (such as an average aggregate size) that is considered to indicate stability.

Ideally, the size and strength of aggregates would be defined on a fundamental physical basis, in the way that hydraulic conductivity can be defined in terms of flux and potential gradients. In that case, any given measurement technique would provide an approximation to the defined ideal. Improved methods would produce results that are increasingly close approximations of the ideal. For aggregates, the definition would have to encompass both size and force, but the difficulty of quantifying the force prevents development of such a definition. Research on disruptive and cohesive forces may eventually solve this problem. Present characterizations have to rely on operational definitions that endorse the result of a particular procedure with a particular apparatus, so we must accept that the method cannot be separated from the definition.

In choosing a method for obtaining aggregate size and stability information, one can focus on either stability or size distribution, and on either wet or dry aggregates. The needs of the application should guide this choice. Erosion applications, for example, usually relate more directly to stability, while hydraulic and gas transport properties may relate more directly to the size distribution. The choice of wet or dry aggregates for measuring may depend on which condition most resembles the field situation, or on such considerations as reproducibility or consistency with other measurements.

Commonly used methods

There are three realms of variables in the procedures to define aggregate characteristics: the disrupting force or energy applied, the distribution of aggregates and particles, and the conditions of testing.

The forces applied to fragment or separate aggregates of the main bulk of soil are fundamentally artificial, though in some ways they resemble forces in the natural setting. It is impossible to deliberately exert forces that exactly oppose the microscopic forces of cohesion. Most practical methods rely on the variable and poorly known forces in a combination of sieving, grinding, or vibration. Some methods use other aggregate-disruptive phenomena, such as the forces involved when liquid is introduced into relatively dry soil.

Some methods quantify an aspect of the applied force or energy. For example, the rupture-threshold approach considers one aggregate at a time, squeezing the aggregate between parallel plates while measuring both the applied force and the linear displacement. The drop-shatter method considers a known mass of soil dropped from a known height onto a hard surface; the difference in potential energy associated with the distance of fall serves as an index of the energy applied to break apart the aggregated soil. Stability methods usually produce their own stability index as a result of the specified procedures and data analyses, for example the fraction of soil weight that comprises stable aggregates. Stability interpretations may also be derived from aggregate size distributions, usually by mathematically converting the tabular data or parameterized distribution formula to an average or other simple index. The mathematical representation of size distribution thus serves beyond its role as a convenience, as the link between size distribution and stability. The basic idea is that the presence of bigger aggregates implies greater stability. The most widely used index for this purpose is the mean weight diameter, defined as the sum of the weighted mean diameters of all size classes, the weighting factor of each class being its proportion of the total sample weight. Ideally this would be determined from integration of the cumulative abundance of aggregates as a function of diameter. The geometric mean diameter can also serve as an aggregate size index, though in recent literature it appears less than the mean weight diameter.

Miscellaneous methods

Many techniques involve deliberate wetting or immersion of the sample. Wet compared to dry measurements on aggregates effectively measure different physical properties of the soil. It is not only the degree of wetness that is important, but also the means by which water has been applied. Wetting the soil in a vacuum, for example, reduces the disruptive forces associated with trapped air and thus results in larger aggregates.

Fast wetting with no vacuum involves immersion of air-dried aggregates in water for a period of time before beginning the mechanical sieving process. This type of wetting causes disintegration and slaking, which may be undesirable. High-vacuum fast wetting involves de-airing aggregates in a vacuum chamber under high vacuum, then instantaneously wetting them inside the chamber. It generally produces minimal disruption. Slow aerosol wetting, in which samples on screens are wet by vapor from below, produces little disintegration. Stabilities are higher and more reproducible with this type of wetting than with vacuum wetting. Wetting by slow wicking with or without vacuum allows aggregates to draw water in from moist filter paper. Used instead of water, organic solvents such as methanol may reduce aggregate disintegration by slaking, and may better preserve aggregate structure in drying.

A method based on differences in water retention curves for fast-wetted and slow-wetted aggregates from replicate soil samples, is similar to the older "high-energy moisture characteristic" method.

Another use of the rate-of-wetting phenomenon is by measuring soil water retention curves for beds of fast-wetted and slow-wetted aggregates. The less the stability, the greater will be the difference between the curves for the fast-wetted and slow-wetted samples. The resulting index is comparable among different soils but not in connection with other stability indices.

Ultrasonic dispersion can supply the disruptive force to associate with aggregate stability. The energy level that achieves a plateau in the quantity of aggregates remaining intact serves as an index of stability.

Stability is sometimes considered operationally in terms of the fraction of sample weight remaining after a prescribed sieving operation. Other methods measure the energy needed to break aggregates by crushing with parallel plates, as described above in connection with quantification of the energy of rupture. The results for a significant number of aggregates need to be reduced to a statistical representation indicative of the properties of the bulk sample. The energy required per increase in aggregate surface area can serve this purpose, as can a distribution function that indicates the probability of failure for a given applied rupture energy.

Issues of general importance

The size distribution and stability of aggregates depend on numerous factors besides the soil type. An obvious consideration is spatial variability, expected to be substantial as it is for most other soil properties. Aggregate stability can increase with storage time of the sample. It also can increase with increasing salt content of the water, and is likely to decrease with temperature.

Soils with concretions (assemblages of primary particles that cannot be broken apart by the disaggregation processes of the chosen method) must be analyzed with respect to the application. In some cases the concretions may be treated as indivisible particles because they are stable under normal cultivation practices, or in other cases as stable aggregates because they usually have porosity, internal surface area, and substantial exchange capacity.

Some soils, especially from humid regions, may be nearly 100% stable in terms of the fraction remaining after prescribed sieving. Greater disruptive force, achievable by increasing the duration and amplitude of sieving, or by a more disruptive wetting technique, enables the detection of differences among highly stable soils, though with the drawback of precluding comparison with results obtained by more standard procedures on less stable soils. The use of multiple methods increases the likelihood that comparisons will be possible among diverse soils.

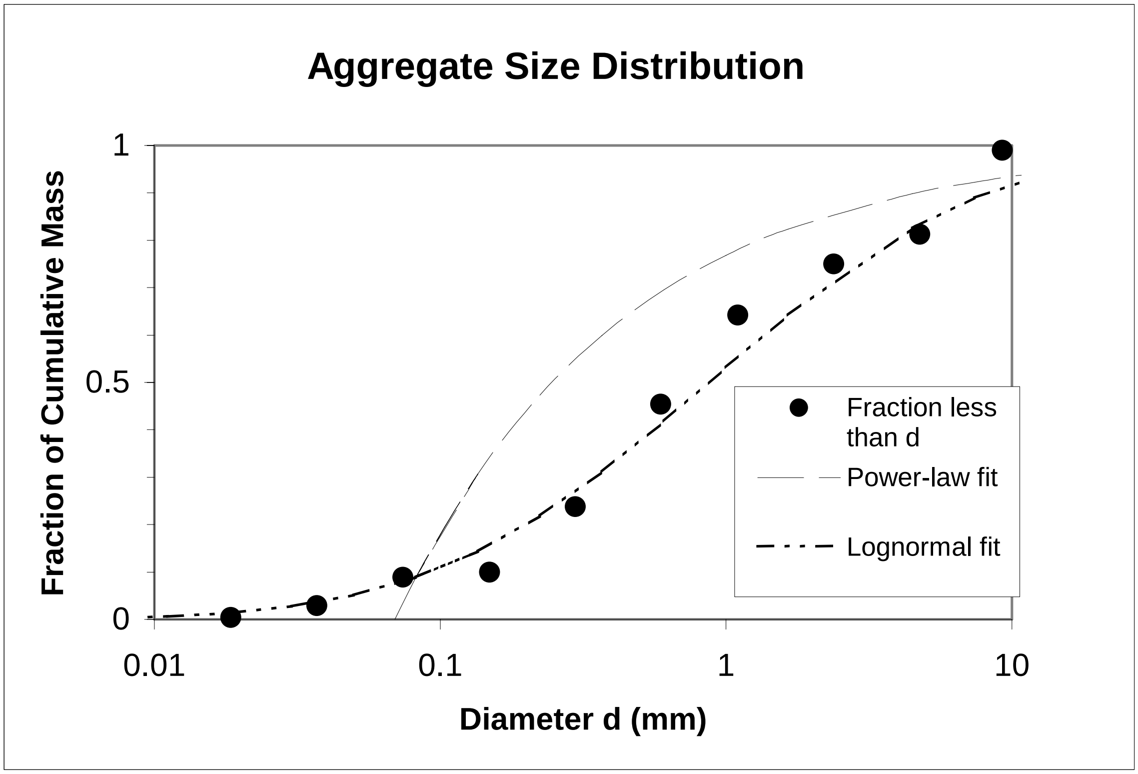

Figure 5. Measured and fitted aggregate size distribution for Sharpsburg silty clay loam. Data are from Wittmuss, H.D., and Mazurak, A.P., 1958, Physical and chemical properties of soil aggregates in a brunizem soil: Soil Science Society of America Proceedings, v. 22, p. 1-5.

Representation and interpretation

To represent size distributions, the fraction of material at particular values of effective aggregate diameter can be graphed directly or cumulatively, as in Figure 5. For convenience in representation or for further mathematical development, these data can be fitted to a specific mathematical form. Of various mathematical functions that have been used to fit the aggregate data, the lognormal distribution is one of the most useful and reasonably fits data from a variety of soils. Being a normal (Gaussian) distribution on a log scale, this distribution is skewed toward the small-diameter end of the range covered. It also has appropriate tapering-off of abundance at both the small- and large-diameter extremes. The lognormal representation has also been used in some of the recent hydraulic property models that are based on aggregate properties

Fractal interpretations have been applied to both aggregate stability and size distribution. A fractal characterization is valid if each subunit of the system is structurally identical (at a reduced scale) to the whole system. This idea is attractive for aggregates since they are not made of primary soil particles on a fully equal basis. Larger aggregates may be thought of as being made of smaller aggregates that are more strongly bound internally than to each other. One attribute of fractal representation is that the cumulative number-size distribution can be represented as a power law; the cumulative abundance of objects greater than a given size is proportional to that size raised to some exponent. Like the lognormal model, a fractal model with an appropriate fractal dimension has a distribution skewed toward the small diameters. By fractal theory the power-law exponent is directly related to the mass fractal dimension, which may be known once the value of the exponent is established. Geometrically, the fractal dimension depends on the shape of the objects and the extent of fragmentation. Tests with the density, shape, and relative diameter variables represented fractally do not always show the degree of consistency over different scales that the most straightforward fractal models would predict. Natural aggregates may tend toward a monolithic internal structure, or otherwise to deviate from true fractal character. Another shortcoming of fractal models is that they are fundamentally unrealistic at extremes of the range. Even so, fractal models remain useful for relating disruptive force to aggregate size, and in general for the modeling of relationships between mechanical properties and other soil properties and conditions.

Conclusions

The concept of an aggregate arises simply because some particles in the soil adhere more strongly than others. Physical aspects of aggregation are fundamental to the character, function, and behavior of soil. They give insight, possibly even much-needed quantitative insight, into soil structure.

The physical characterization of aggregation requires crucial tradeoffs. Little can be quantified about aggregates without operational definitions and criteria. The choice between widely used, informally standardized methods and more novel ones often involves a substantial tradeoff between the need for consistency and the ultimate appropriateness of the method. The more standardized methods facilitate comparability but the quantitative indices they generate may not give the sort of aggregate characterization most pertinent to the application at hand. Reliance on operational definitions also makes it awkward to incorporate ongoing scientific advances in conceptualizations and techniques. Research that leads to standardization of the specified force and a fundamental physical definition would help in allowing aggregate-measurement technology to advance without loss of comparability. The difficulty of this undertaking parallels the difficulty of bringing the general concept of soil structure into an objective, quantitative realm, but the potential benefits of even partial success justify much effort.

Related Articles in ESE:

Aggregation: microbial aspects, Flocculation and dispersion, Structure of soils, Swelling and shrinking of soils, Fractal analysis of soils.

Further Reading:

Chepil, W.S., 1962, A compact rotary sieve and the importance of dry sieving in physical soil analysis: Soil Science Society of America Proceedings, v. 26, p. 4-6.

Currie, J.A., 1966, The volume and porosity of soil crumbs: Journal of Soil Science, v. 17, no. 1, p. 24-35.

Dickson, J.B., Rasiah, V., and Groenevelt, P.H., 1991, Comparison of four prewetting techniques in wet aggregate stability determination: Journal of Soil Science, v. 71, p. 67-72.

Hillel, D., 1980, Fundamentals of soil physics: San Diego, Academic Press, 413 p.

Holden, N.M., 1995, Temporal variation in ped shape in an old pasture soil: Catena, v. 24, p. 1-11.

Kosugi, K., and Hopmans, J.W., 1998, Scaling water retention curves for soils with lognormal pore-size distribution: Soil Science Society of America Journal, v. 62, p. 1496-1505.

Le Bissonnais, Y., 1996a, Aggregate stability and assessment of soil crustability and erodibility -- I. Theory and methodology: European Journal of Soil Science, v. 47, p. 425-437.

Le Bissonnais, Y., 1996b, Soil characteristics and aggregate stability, *in* Agassi, M., ed., Soil erosion, conservation, and rehabilitation: New York, Marcel Dekker, p. 41-60.

Matkin, E.A., and Smart, P., 1987, A comparison of tests of soil structural stability: Journal of Soil Science, v. 38, p. 123-135.

Nimmo, J.R., and Perkins, K.S., 2002, Aggregate stability and size distribution, *in* Dane, J.H., and Topp, G.C., eds., Methods of soil analysis, Part 4--Physical methods: Soil Science Society of America Book Series No. 5: Madison, Wisconsin, Soil Science Society of America, p. 317-328.

North, P.F., 1976, Towards an absolute measurement of soil structural stability using ultrasound: European Journal of Soil Science, v. 27, p. 451-459.

Oades, J.M., and Waters, A.G., 1991, Aggregate hierarchy in soils: Australian Journal of Soil Research, v. 29, p. 815-28.

Perfect, E., 1997, Fractal models for the fragmentation of rocks and soils: a review: Engineering Geology, v. 48, p. 185-198.

Rieu, M., and Sposito, G., 1991a, Fractal fragmentation, soil porosity, and soil water properties -- I. Theory: Soil Science Society of America Journal, v. 55, p. 1233-1238.

Rieu, M., and Sposito, G., 1991b, Fractal fragmentation, soil porosity, and soil water properties -- II. Applications: Soil Science Society of America Journal, v. 55, p. 1239-1244.