ISSN 1064-2293, Eurasian Soil Science, 2012, Vol. 45, No. 3, pp. 221–230. © Pleiades Publishing, Ltd., 2012.
Original Russian Text © V.A. Rozhkov, 2012, published in Pochvovedenie, 2012, No. 3, pp. 259–269.

Classiology and Soil Classification

V. A. Rozhkov

Dokuchaev Soil Science Institute, Russian Academy of Agricultural Sciences, per. Pyzhevskii 7, Moscow, 119017 Russia

E-mail: rva39@mail.ru

Received November 22, 2010

Abstract—Classiology can be defined as a science studying the principles and rules of classification of objects of any nature. The development of the theory of classification and the particular methods for classifying objects are the main challenges of classiology; to a certain extent, they are close to the challenges of pattern recognition. The methodology of classiology integrates a wide range of methods and approaches: from expert judgment to formal logic, multivariate statistics, and informatics. Soil classification assumes generalization of available data and practical experience, formalization of our notions about soils, and their representation in the form of an information system. As an information system, soil classification is designed to predict the maximum number of a soil's properties from the position of this soil in the classification space. The existing soil classification systems do not completely satisfy the principles of classiology. The violation of logical basis, poor structuring, low integrity, and inadequate level of formalization make these systems verbal schemes rather than classification systems *sensu stricto*. The concept of classification as listing (enumeration) of objects makes it possible to introduce the notion of the information base of classification. For soil objects, this is the database of soil indices (properties) that might be applied for generating target-oriented soil classification system. Mathematical methods enlarge the prognostic capacity of classification systems; they can be applied to assess the quality of these systems and to recognize new soil objects to be included in the existing systems. The application of particular principles and rules of classiology for soil classification purposes is discussed in this paper.

DOI: 10.1134/S106422931203009X

INTRODUCTION

The development of soil classification systems remains one of the major challenges of pedology. The theory of classification is also being developed by other sciences. The importance of this work independently from the particular field of science is generally recognized: "Whatever the diversity of objects in group "A" and whatever the differences between them before they are classified, a universal theory of classification should be developed for them," [3]. The theory and methodology of classification may be the same for different fields of science. Soil science can also apply the achievements in the general theory of classification and contribute to them on the basis of its own experience. Theoretical speculations are based on the compression of diverse factual materials into easily "observable" structures (taxa) [15]. In this sense, classification is a necessary prerequisite for the development of the theory; at the same time, it may be the result of theoretical considerations. This dual role of classification has always been recognized by Russian pedologists. Afanas'ev [2] noted that each classification system could be considered a philosophical system of pedology expressed in logical schemes and symbols and reflecting the general credo of science and the achievements at the particular stage of its development.

Classiology can be defined as a science about the theory of classification and its applications [23]. This is a specific field of science that has its own societies and working groups in many countries. The importance of this field has been recognized by the international Committee on Data for Science and Technology (CODATA) (the IUSS is a member of CODATA). Attempts to organize the classiology society were undertaken in Russia. Two special conferences on the development of the theory of classification were organized in 1978 in Borok (Yaroslavl oblast) and in 1982 in Novosibirsk. These conferences attracted specialists in different fields of science (philosophers, mathematicians, biologists, etc.) interested in methodological aspects of classification. However, no official society of "classiologists" was organized.

Theoretical problems of classification remain the object of heated discussions between enthusiasts participating in the "classiological" movement. A general theory of classification has yet to be developed.

The same is true with respect to the theory of soil classification. The new Russian substantive-genetic soil classification system [9] has not found universal support among Russian soil scientists because of various subjective and objective reasons. The WRB system is also criticized. It should be noted that, at present, we have no special studies devoted to the comparative analysis of soil classification systems developed by different authors; attempts to find correlations between soil names in different classification systems do not disclose differences in the principles of these classification systems. In fact, most of these systems do not satisfy the requirements of logic; they lack strict formal definitions of the objects classified, well-defined space of soil characteristics, and inner integrity of their structure. The level of formalization of classification decisions and of the particular classification procedures is insufficient for the application of strict mathematical methods. In other words, these classification systems resemble arbitrary schemes and lists of objects rather than classification systems *sensu stricto*.

As noted more than 50 years ago, the major challenge is "to develop the scientific principles of soil classification rather than the classification system itself" [6, pp. 78–79]. The systems may change in the future, but their principles should be preserved intact. Discussions on the principles of classification have a long history. Since the time of Aristotle (384 BC) and Adanson (1727–1806), the logic of classification has been in the focus of interest. In particular, the ideas of J.S. Mill[1] [16] have been widely applied in pedology for developing the concepts of soil classification systems [2, 4, 5, 26], but they have not resulted in formulation of the rules of construction of these systems. These rules are particularly important for the development of "natural" or basic classification systems as opposed to applied systems [31–34]. It is generally recognized that the rules of logic are essential for a basic soil classification system to withstand the test of time [10, 25–28].

In this paper, some aspects of the theory of classification are discussed with the aim to develop the basis for constructing classification algorithms.

GENERAL NOTIONS

The central idea of any classification system is to formalize the descriptions of particular objects and to create "the space of characteristics" reflecting the purpose of classification. In essence, classification systems represent information systems generalizing the diversity of the properties of classified objects and representing them in a convenient "observable" form [22].

J.S. Mill [16] gave a comprehensive definition to the notion of classification that has been applied and cited by pedologists, including Afanas'ev [2], Fridland [26, 27], and many others. He wrote: ...{Classification, thus regarded, is a contrivance for the best possible ordering of the ideas of objects in our minds; for causing the ideas to accompany or succeed one another in such a way as shall give us the greatest command over our knowledge already acquired, and lead most directly to the acquisition of more. The general problem of Classification, in reference to these purposes, may be stated as follows: To provide that things shall be thought of in such groups, and those groups in such an order, as will best conduce to the remembrance and to the ascertainment of their laws.

Classification thus considered, differs from classification in the wider sense, in having reference to real objects exclusively, and not to all that are imaginable: its object being the due coordination in our minds of those things only, with the properties of which we have actually occasion to make ourselves acquainted. But, on the other hand, it embraces *all* really existing objects. We can not constitute any one class properly, except in reference to a general division of the whole of nature; we can not determine the group in which any one object can most conveniently be placed, without taking into consideration all the varieties of existing objects, all at least which have any degree of affinity with it...

There is no property of objects which may not be taken, if we please, as the foundation for a classification or mental grouping of those objects; and in our first attempts we are likely to select for that purpose properties which are simple, easily conceived, and perceptible on a first view, without any previous process of thought...

The ends of scientific classification are best answered, when the objects are formed into groups respecting which a greater number of general propositions can be made, and those propositions more important, than could be made respecting any other groups into which the same things could be distributed. The properties, therefore, according to which objects are classified, should, if possible, be those which are causes of many other properties; or, at any rate, which are sure marks of them. Causes are preferable, both as being the surest and most direct of marks, and as being themselves the properties on which it is of most use that our attention should be strongly fixed...

The phrase Natural Classification seems most peculiarly appropriate to such arrangements as correspond, in the groups which they form, to the spontaneous tendencies of the mind, by placing together the objects most similar in their general aspect; in opposition to those technical systems which, arranging things according to their agreement in some circumstance arbitrarily selected, often throw into the same group objects which in the general aggregate of their properties present no resemblance, and into different and remote groups, others which have the closest similarity. It is one of the most valid recommendations of any classification to the character of a scientific one, that it shall be a natural classification in this sense also; for the test of its scientific character is the number and importance of the properties which can be asserted in common of all objects included in a group; and prop-

[1] John Stuart Mill (1806–1873) was a British philosopher and political economist. In 1864, his fundamental monograph—*A System of Logic*—was published, in which a wide range of the problems of scientific investigation were discussed. His ideas are of great value for modern information technologies. Russian translation of this book (the edition of 1864) appeared in 1914 and has not been republished later.

erties on which the general aspect of the things depends are, if only on that ground, important, as well as, in most cases, numerous. But, though a strong recommendation, this circumstance is not a *sine qua non*; since the most obvious properties of things may be of trifling importance compared with others that are not obvious... Our natural groups, therefore, must often be founded not on the obvious but on the unobvious properties of things, when these are of greater importance.

...the classification of objects should follow those of their properties which indicate not only the most numerous, but also the most important peculiarities. What is here meant by importance? It has reference to the particular end in view; and the same objects, therefore, may admit with propriety of several different classifications. Each science or art forms its classification of things according to the properties which fall within its special cognizance, or of which it must take account in order to accomplish its peculiar practical end...

...when we are studying objects not for any special practical end, but for the sake of extending our knowledge of the whole of their properties and relations, we must consider as the most important attributes those which contribute most, either by themselves or by their effects, to render the things like one another, and unlike other things; which give to the class composed of them the most marked individuality; which fill, as it were, the largest space in their existence, and would most impress the attention of a spectator who knew all their properties but was not specially interested in any... Natural groups... are given by Type, not by Definition...

The class is steadily fixed, though not precisely limited; it is given, though not circumscribed; it is determined, not by a boundary-line without, but by a central point within; not by what it strictly excludes, but by what it eminently includes; by an example, not by a precept; in short, instead of a Definition we have a Type for our director.

A Type is an example of any class, for instance a species of a genus, which is considered as eminently possessing the character of the class...

The problem is, to find a few definite characters which point to the multitude of indefinite ones.

Our conception of the class continues to be grounded on the characters; and the class might be defined, those things which *either* possess that set of characters, *or* resemble the things that do so, more than they resemble any thing else...

The class, therefore, is constituted by the possession of *all* the characters which are universal, and most of those which admit of exceptions...

Not only, therefore, are natural groups, no less than any artificial classes, determined by characters; they are constituted in contemplation of, and by reason of, characters. But it is in contemplation not of those characters only which are rigorously common to all the objects included in the group, but of the entire body of characters, all of which are found in most of those objects, and most of them in all. And hence our conception of the class, the image in our minds which is representative of it, is that of a specimen complete in all the characters; most naturally a specimen which, by possessing them all in the greatest degree in which they are ever found, is the best fitted to exhibit clearly, and in a marked manner, what they are...

A Nomenclature in science is, as we have said, a system of the names of Kinds...

...true principles of rational classification... are applicable to all cases in which mankind are called upon to bring the various parts of any extensive subject into mental co-ordination. They are as much to the point when objects are to be classed for purposes of art or business, as for those of science...} [16] (English citations are given according to the edition of 1882).

Many ideas discussed below are borrowed from *A System of Logic* published in 1864.

Dokuchaev argued that a valid natural soil classification should embrace all the soils (all Russian soils).

Classification is a separation of a set of objects into homogeneous subsets (classes), each of which includes at least one object; these subsets should have no common objects. The procedure of classification is the attribution of a given object to one of the known classes, i.e., the identification of the object, or its "recognition."

The theory of classification is a system of propositions making it possible to construct and transform classification systems, rationalize them, and explain them [3]. *A class* is a subset of objects that have some common character (or a combination of characters) distinguishing them from other objects and classes in the initial set of objects. Simple classes have only one discriminating character; complex classes are distinguished on the basis of a set of characters[2]. The notion of *a character* includes both the name of some property of the objects and the value of this property. The values of characters may be specified according to various scales allowing their treatment with mathematical methods [20]. The scales may be either logical (qualitative) or arithmetic (quantitative). Logical scales are separated into nominal and binary scales (qualitative unranked scales) and ordinal (qualitative ranked) scales. Arithmetic scales are subdivided into the absolute scales and the scales of intervals, ratios, or differences. Characters or their combinations used to discriminate between the objects form the basis of classification. A studied object is described with the help of these characters. Soil classification systems may be built on the basis of different soil characters, e.g., genetic horizons, elementary pedogenic processes, proper soil

[2] These groups of classes correspond to mono- and polythetic types of classification systems.

properties, etc. The resulting classification systems will differ in their goals and in their foundations.

RESULTS

The following kinds of classification systems are applied in the soil science.

1. *Natural and artificial* (applied, technical).

2. *According to their contents:*

—genetic;

—factor-based;

—interpretative;

—substantive: (1) according to conservative features and

(2) according to dynamic features (regimes);

—utilitarian (economic uses of soils); and

—combined (factor-genetic).

3. *According to their structure:*

—hierarchical and

—ordinate.

4. *According to their basis:*

—monothetic (based on a single character) and

—polythetic (based on a set of characters).

5. *Numerical* classification applies multivariate statistical methods, cluster analysis, and pattern recognition methods.

Natural classification describes classes of objects really existing in nature in such a way that a position of a given object in the classification system points to all the characters of this object. According to Lyubishchev [13], the closer the classification system to this ideal situation, the less artificial this system. Natural classification systems are the systems, in which the number of characters (properties) of the object in functional dependence on the position of this object in the classification system gains its maximum. Such classifications have the greatest prognostic value. As noted by Voronin [3], they are also stable in time and are recognized by many scientists (the latter is important, because verification of the real existence of all the taxa included in such systems is a difficult task requiring considerable resources). At present, we have limited possibilities to create such natural classification systems for the soils. The theory of classification lags behind the demands of time, and the number of tasks to be solved rapidly increases [6]. Often, we deal with incorrect formulations of the goals of classification. For example, it is incorrect to integrate data on the genesis, morphology, composition, and fertility of soils in a single classification. It is impossible to create a universal classification addressing many different and, often, contradictory issues simultaneously.

Sibirtsev [24] distinguished between three types of soil classifications: (a) natural-historic classifications that are based on the natural soil properties, (b) technical or applied classifications, and (c) economic classifications. Natural-historic classifications were further subdivided into geologic-petrographic, chemical, physical, and combined. Technical (applied) classifications were subdivided into those applicable for farming, forestry, or other particular uses (e.g., for hygienic purposes). He argued that the goals of classification should be clearly formulated. The goals formulated by Sibirtsev are still topical. Dokuchaev argued that both "internal" and "external" characters are important for the creation of the natural scientific classification of soils.

The choice of particular characters should be based on the genetic approach to soils [12]. The major challenge of pedology is to distinguish between "old and new, inherited and newborn" features, properties, and processes and to identify them in soils [7]. The problem is to discriminate between soil taxa and archetypes on the basis of the system of chosen characters. The values of these characters depend on the goal of classification. In some cases, a single character may be more useful than 20 characters, provided that the latter are not intended for the preset task [32].

The definition of the *object of soil classification* is not a trivial problem. Indeed, the notion of soil is still indefinite, the genetic concepts applied in soil science are often hypothetical, the soils are heterogeneous both in the vertical and horizontal directions. their boundaries with other natural objects often have a continual character (i.e., it is not easy to separate soil from other objects), and inner soil boundaries (between different genetic horizons) also often have a continual character. Moreover, soil properties are dynamic (change with time); relict and modern and conservative and labile properties can be distinguished. These difficulties have an objective character.

There are also subjective difficulties related to our intention to integrate substantive, functional, and factorial aspects into a single classification; often, various utilitarian aspects are also included. In other words, we attempt to integrate quite different goals into a single system.

From my point of view, the object (unit) of soil classification system may be specified in the following way:

(1) Field and/or analytic description of a soil profile (which actually means the description of the pedon as a 3D body);

(2) Values of soil properties determined at standard depths;

(3) The vector of changes in the soil properties (these properties should be selected with respect to the particular goal of classification); and

(4) On the basis of this information, we can use parameters of equations approximating distributions of the particular property (or of their entire set) in the soil profile.

The choice of the properties to be analyzed is dictated by the goals of a given classification. The importance of this choice is illustrated by Fig. 1, on which

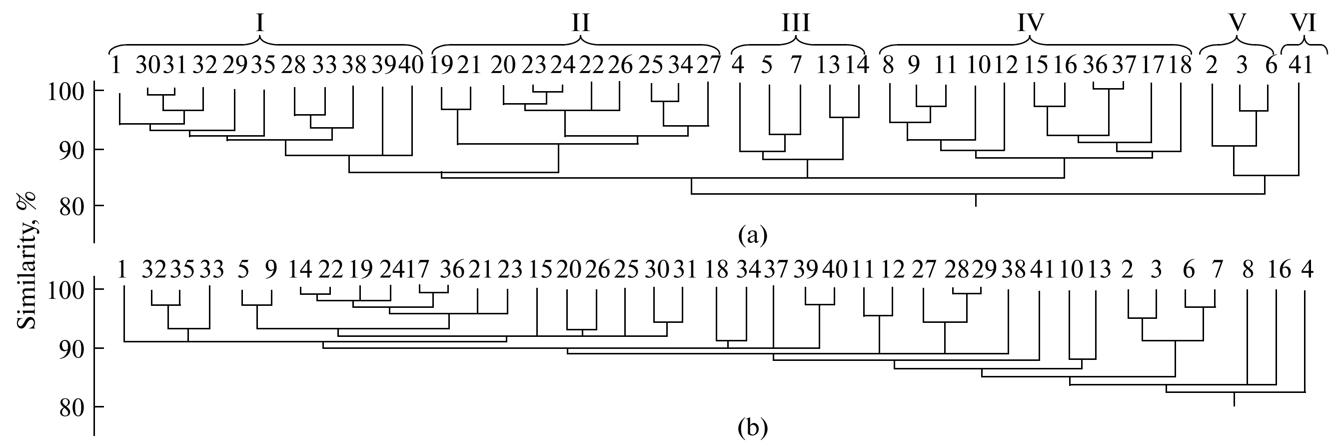

Fig. 1. Dendrograms of soil types according to their (a) process-based and (b) profile-based codes (according to [19]). I—Tundra soils: (*1*) arctic, (*2*) gley tundra, and (*3*) podzolic-gley tundra. II—Taiga soils: (*4*) podburs, (*5*) palevye (pale) taiga, (*6*) gley taiga, (*7*) mucky gley taiga, (*8*) gley-podzolic, (*9*) podzolic, (*10*) podzols, (*11*) soddy-podzolic, (*12*) peat podzolic-gley, and (*13*) rendzinas. III—Forest and forest-steppe soils: (*14*) soddy calcareous, (*15*) pseudopodzolic, (*16*) pseudopodzols, (*17*) brown forest, (*18*) podzolized brown forest, (*19*) gray forest, and (*20*) soddy gley. IV—Steppe soils: (*21*) podzolized chernozems, (*22*) leached chernozems, (*23*) typical chernozems, (*24*) ordinary chernozems, (*25*) southern chernozems, (*26*) meadow-chernozemic, and (*27*) dark chestnut. V—Desert soils: (*28*) light chestnut, (*29*) brown semidesert, (*30*) gray-brown desert, (*31*) light sierozems, (*32*) typical sierozems, and (*33*) meadow sierozems. VI—Subtropical and salt-affected soils: (*34*) cinnamonic, (*35*) gray cinnamonic, (*36*) yellow-brown, (*37*) krasnozems and zheltozems (red and yellow subtropical), (*38*) solonchaks, (*39*) solonetzes, (*40*) solodic, and (*41*) swampy.

dendrograms of soils are constructed on the basis of process-based (genetic) and profile-based (soil horizons) soil codes.[3] It can be seen that the process-base dendrogram more or less adequately reflects the traditional zonal (factor-based) distribution of the soils, whereas the profile-based dendrogram does not show any zonal regularities. This is the result of ambiguous definitions (and designations) of soil horizons (as applied in this case). For example, the humus-accumulative A1 horizon is the same for soddy-podzolic soils and chernozems, though, certainly, these horizons differ in their natural characteristics. The same is true with respect to other horizons. Thus, the definition (description) of a soil taxon through a combination of the indices of soil horizons does not always lead to the correct identification of this taxon.

As follows from this example, the choice of the characters (properties) of soils to be used as the basis of soil classification is very important. Excessive number of low-informative characters can distort the general structure of classification system, including the particular classes of objects, the relationships between them, the prognostic capacity of the classification, and the identification of new objects. The analysis of “informativity” of the applied characteristics should be performed. This is also important in order to develop a compact classification system. It is evident that classification systems in which the number of classes is close to the number of classified objects are of little scientific value. In turn, the number of classes depends on the number of characters and grades of their particular values. This is illustrated by Fig. 2, which explains the idea of the classification of enumeration and the dependence of the classification system on the number of selected characters and their grades [3].

This is an abstract example. However, it can be easily specified. Let the first level of this classification be represented by three grades of the humus content (1—low, 2—moderate, and 3—high). The second level is distinguished on the basis of three grades of soil acidity (1—low, 2—moderate, and 3—high). At the third level, only two grades of base saturation are used: 1—saturated and 2—nonsaturated. The choice of the particular set of characters and their grades dictates

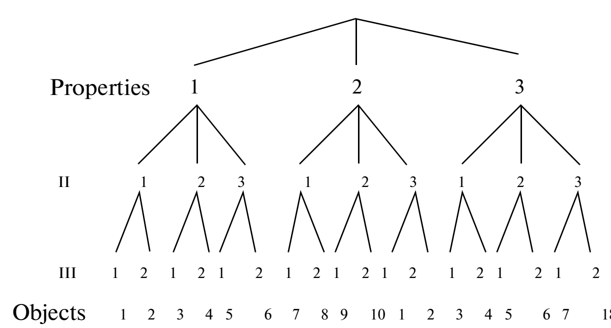

Fig. 2. Classification as a listing of objects.

[3] The algorithms of the development of these dendrograms and the methods of their interpretation were discussed in detail in [4].

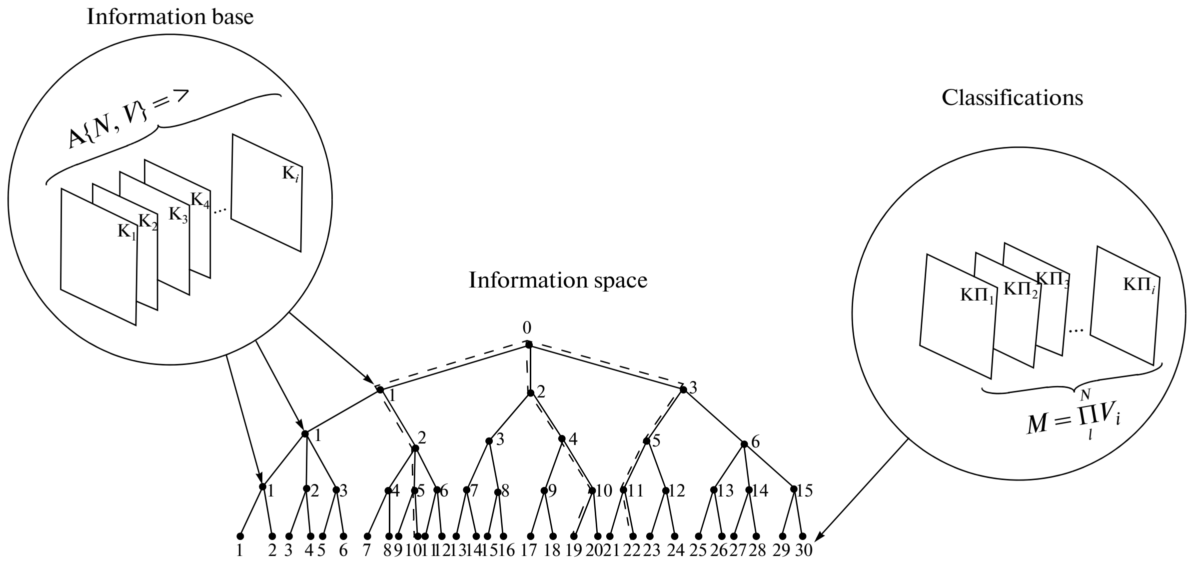

Fig. 3. Information base of classification. Dashed line shows a branch of the particular classification.

the potential number of classes equal to the number of combinations of the characters and their grades. In the given example, only 18 combinations are possible. However, some of them do not exist in nature (e.g., there can be no strongly acid soils saturated with bases). Thus, the enumeration of classes in this classification is redundant. However, this redundancy does not complicate the organization and analysis of the data; it ensures the completeness of the classification system. If we change the taxonomic level of the particular criterion (e.g., if we consider base saturation at the first level and the humus content at the third level, the number of classes will remain the same. It can be increased, if we add an additional character and its grades. Thus, we may apply the fourth character with three grades (low, moderate, and high) at the fourth level, and the number of classes will be $18 \times 3 = 54$.

In fact, none of the existing soil classifications ensures such definiteness regarding the number of classes and the form of their representation. As a rule, this problem is solved in an inverse way based on our notions about "central soil images" (soil types). Known central images of soil types are included in the classification structure; if they do not fit it, a new set of characteristics allowing us to separate the given soil from other soils is added. The validity of this set for separation of other soils is not tested. In some cases, exceptions from general principles of separation of soil taxa are made for the given classes.

Thus, the problem of classification is largely related to the construction of a system of informative soil characteristics.

In soil science, the idea of classification based on the enumeration of the characters seems to be very promising. To have such a classification, we need to develop an *information base of classification* (IBC) [22, 29, 30, 35]. The idea of a classification of enumeration is illustrated in Fig. 3.

To create an information base for such a classification, we have to choose soil characters pertaining to the goals of any possible classification and to fix the grades of their values. In Fig. 3, the system of such characters is denoted as the *information space* of the classification. In essence, it is designed in the same way as a classification of enumeration.

All the known soil characteristics can be used as for this purpose. In particular, this approach was tested earlier [1]. However, it cannot be considered an optimum approach, because specific target-oriented classifications should be considered separately from general natural classifications designed for their use by the scientific community.

As was shown earlier, the number of classes in a classification of enumeration is a product of the number of grades in the selected characters. If we reduce the number of characters or the number of grades, we shall have a different classification with a different number of classes. Thus, we should select the characters pertaining to the goals of particular classification from their total set in the information space. In this case, the number of classes will be reasonably limited. At the same time, the IBC should contain a vast body of information allowing us to generate soil classifications for different purposes. Target-oriented classifications built on some subset of the total set of characters in the IBC represent particular branches of the classification tree. This approach opens good possibilities for establishing correlations between different classifications on the quantitative basis (provided that all these classifications have the same information space). Some examples of this approach are discussed below.

Thus, the information base of soil classification can be defined as a formalized fixed system of soil characteristics encompassing the total body on knowledge on the soil properties and the conditions of soil formation. In contrast to databases in automated information systems that contain data on the particular soil pits, the IBC should contain information on the grades (intervals) of the characters used to describe soils and to separate them into classes. These grades should be expressed in the form of certain scales (classifiers). The IBC should be a complete, multiple, and unified (in a broader sense, formalized) system. Its completeness means that we have a sufficient number of characters to generate different target-oriented classifications. The multiple character of the IBC means that the characters should reflect different aspects of our knowledge about soils (e.g., it should reflect chemical and physical soil properties, pedogenetic processes, etc.). The unified nature of the IBC means that the particular soil properties should be presented in uniform scales (nominal, ordinal, or various arithmetic). The formalized nature of the IBC denotes the level of its organization, i.e., the description of the relationships between different soil properties in the form of a relational structure of characters and their grades.

Geometrically, the IBC (or its particular part) compose the information space (IS) of classification. If we have a hierarchical system of relationships between the particular soil characteristics, we may develop the classification of enumeration, i.e., to distinguish between soil classes that can be separated on the basis of a given set of characteristics. The more complete the IBC and the larger the dimensions of IS based on it, the larger the number and diversity of soil classes that can be obtained. While developing a particular classification, we use only some part of the IBC and the corresponding IS. The development of the complete (exhaustive) IBC (and IS) is a serious challenge. In contrast to “central images of soils” that are used in expert classification systems, we deal elements of the IBC with quite definite information volumes integrated into the system via quantitatively measurable relationships.

The use if the IBC concept allows us to construct various classifications based on the chosen subsets of characters. The diagnostics of the separated soil groups (classes) can be optimized. On this basis, target-oriented and information-saturated classification systems can be produced. To develop the IBC and to apply it for soil classification purposes, a logical MERON software system has been developed.

Ultimately, the IBC represents a relational information system in the particular field of knowledge that integrates hierarchical and ordinate (network-type) structures of classification.

Upon the development of a genetic classification, its characteristics of division should include the characters reflecting soil genesis; these characters should generalized and formalized, which is a difficult challenge. If we develop a classification of potential soil uses, its characteristics of division should include the characters directly related to soil use. Not all the characters included in the IBC should be used simultaneously for the development of a particular classification. Their choice depends on the goals of this classification. We should bear in mind that the soils of different geneses may have similar fertility and vice versa. In fact, we deal with different objects of classification, if we consider soils from the genetic and utilitarian points of view. Thus, a degraded chernozem (agrochernozem) may be less fertile than a well-cultivated podzol (agropodzol). There are no grounds to mix these objects in a unified classification, because such a classification will contain an excessively large number of classes, and its interpretation and practical application will be too difficult. One of the purposes of classification is to compress information in the particular field of knowledge, which is only possible, if the particular goals of this classification are strictly specified and its structure corresponds to these goals. As shown earlier, the addition of new characters, as well as the reduction of the number of characters, lead to the appearance of a new classification. Thus, to develop a classification, we should specify its goals and determine necessary and sufficient characters to be used for the division of objects. It is probable that the absence of a clear formulation of the goals of classification is one of the reasons for the absence of a universally recognized soil classification. The existing systems remain too schematic.

There are three approaches to the choice of a combination of characters to be used in a given classification. The first approach suggests that the information base should include the entire diversity of characters (soil attributes) used by different researchers. This approach has partly been realized in soil science [1, 28]. According to the second approach, we only select those characters that have been named by most of the experts. The third approach suggests the application of the methods of multivariate statistics to determine the most informative characters [19]. In the soil-ecological tasks, this method made it possible to reduce the initial lists of characters by 46–80% (http://lem.edu.mhost.ru/doc/presentations/Rozhkov.pdf).

The graphic representation of a fragment of a well-known factor-based soil classification [8] is shown in Fig. 4.

Overall, this classification “has niches” for 216 factor-controlled soil types. Not all of them have their soil names; missing names (which means that such soils have not been described) are indicated by question marks. These question marks are important, as they point to the gaps in our knowledge. Moreover, of we analyze the geography of soils as shown in this clas-

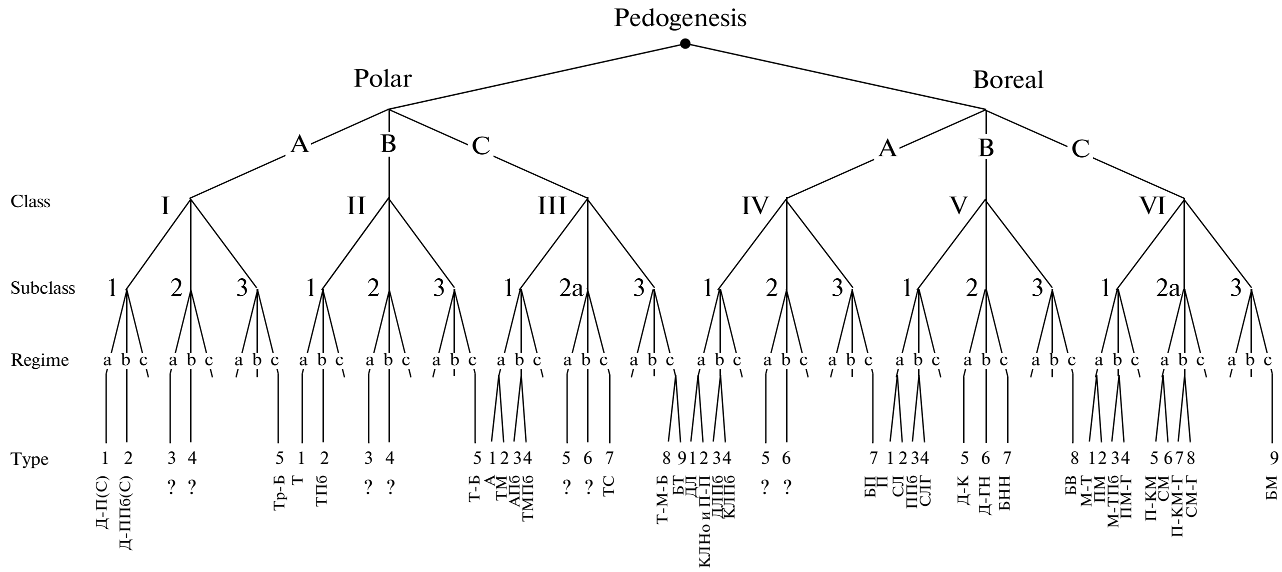

Fig. 4. Fragment of the *Classification of Soils of the Soviet Union* (according to Ivanova and Rozov [8]). The Polar and Boreal groups of soils. Pedogenetic sectors: A—oceanic, B—continental, and C—ultracontinental. Soil classes: (I) soddy-mucky subarctic soils, (II) tundra soils (beyond the permafrost area), (III) tundra soils (permafrost area), (IV) soddy soils and acid nonpodzolized and surface-podzolized forest soils, (V) taiga-forest podzolized and podzolic soils, (VI) permafrost-affected taiga soils. Soil subclasses: (1) biogenic, (2) bio-lithogenic, (2a) bio-lithogenic and halomorphic, and (3) biohydrogenic. Soil regimes: (a) automorphic, (b) semihydromorphic, and (c) hydromorphic. Soil types: Д-П(С)—soddy-mucky (subarctic), Д-ППб(С)—soddy-mucky of subarctic mires, Тр-Б—forb swampy, Т—tundra, ТПб—tundra boggy, Т-Б—tundra bog, А—arctic, ТМ—tundra permafrost-affected, АПб—arctic boggy, ТМПб—tundra boggy permafrost-affected, ТС—tundra solonchaks (marsh soils), Т-М-Б—tundra bog permafrost-affected, БТ—peat mounds, ДЛ—soddy forest, КЛНо, П-П—acid nonpodzolized and surface-podzolic forest, ДЛПб—soddy boggy forest, КЛПб—acid boggy forest, БП—peat soils of transitional bogs, П—podzolic, СЛ—gray forest, ППб—boggy podzolic, СЛГ—gleyed gray forest, Д-К—soddy calcareous, Д-ГН—saturated soddy-gley, БНН—saturated peat soils of low moors, БВ—peat soils of high moors, М-Т—taiga permafrost-affected, ПМ—palevye permafrost-affected, М-ТПб—boggy taiga permafrost-affected, ПМ-Г—gleyed palevye permafrost-affected, П-КМ—mucky calcareous permafrost-affected, СМ—solodic permafrost-affected, П-КМ-Г—gleyed mucky calcareous permafrost-affected, СМ-Г—gleyed solodic permafrost-affected, and БМ—peat bog permafrost-affected soils.

sification system, we can make certain predictions of changes in the character of soils in relation to their evolution under the impact of expected climate changes. We can say that Glinka's expectations about the predictive power of the factor-based soil classification—"Tell me what are the local conditions, and I shall tell you what kind of soil should be in this place" [4, p. 14]—have been realized.

In this example, a formal approach to the development of classification has been applied. Four characters used for the division of the entire set of soil objects at the four hierarchical levels (I–IV) have been selected, and their grades have been specified. The plot shows all the potential factor-ecological niches of soil formation, though not all of them are "filled" with known soil types. The resulting classification is a classification of enumeration. A fundamental importance of the principles used in the design of this system can hardly be overestimated. The structure and logic of the developed system are explicitly defined. The goals of the classification are reflected through the system of chosen characters. It should be stressed that the choice of these characters is a key element of the classification. An eclectic choice of many characters would not result in the creation of a "multi-target" classification; instead, the indefiniteness of classification decisions will increase.

The IBC concept opens new possibilities for a quantitative assessment of correlation between classification systems developed by different authors. Thus, a common information base was developed for the *Classification of Soils of the Soviet Union*, the FAO legend to the *Soil Map of the World*, and *Soil Taxonomy* [20]. On this basis, a dendrogram of similarity between soil taxa from these classification systems was composed (Fig. 5). It is based on 64 characters (soil horizons, some soil properties, and conditions of soil formation). Soils at the higher taxonomic levels of these classification systems were described by these 64 indices in terms "present," "absent", and "may be." The obtained picture is rather ambiguous and requires special analysis. It is probable that the characters used for the soils' division and the compared taxonomic levels should be optimized. However, the possibility of a clear visualization of different classification systems is important; this method can be used to compare classification decisions.

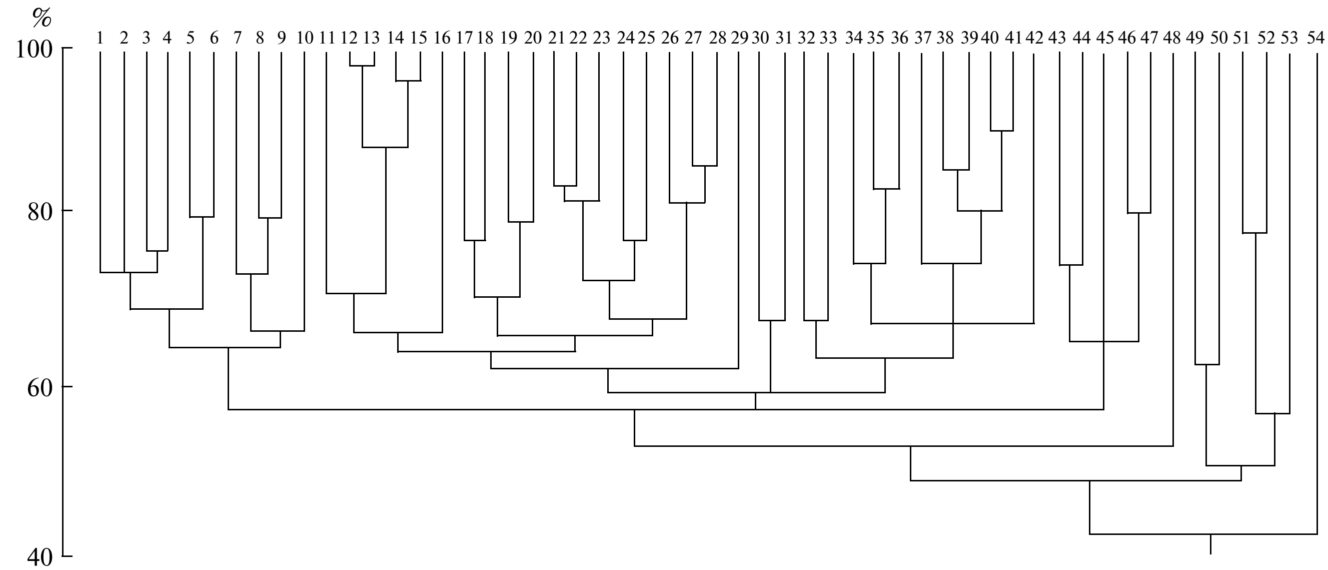

Fig. 5. Dendrogram of similarity between soil taxa used in the *Russian Soil Classification System*, the FAO *Soil Map of the World* legend, and *Soil Taxonomy*. Soils: (1) gleyzems, (2) soddy organo-accumulative, (3) Fluvisols (J), (4) Gleysols (G), (5) lithozems, (6) Lithosols (I), (7) fersiallitic, (8) Ultisols, (9) Oxisols, (10) Planosols (W), (11) cryozems, (12) peat soils, (13) residual peat soils, (14) Histosols, (15) Histosols (O), (16) weakly developed soils, (17) Al–Fe-humus soils, (18) Podzols (P), (19) Spodosols, (20) Podzoluvisols (D), (21) crusty soils, (22) Rankers (U), (23) Rendzinas (E), (24) Arenosols (Q), (25) Regosols (R), (26) Ferralsols (F), (27) Nitosols (N), (28) Acrisols (A), (29) Entisols, (30) texture-differentiated soils, (31) Cambisols (B), (32) metamorphic soils, (33) Luvisols (L), (34) humus-accumulative soils, (35) Gleyzems (M), (36) Phaeozems (H), (37) low-humus soils, (38) Chernozems (C), (39) Kastanozems (K), (40) Xerosols (X), (41) Yermosols (Y), (42) Mollisols, (43) halomorphic, (44) Solonchaks (Z), (45) Solonetz (S), (46) Vertisols, (47) Vertisols (V), (48) Alfisols, (49) alkaline clay-differentiated, (50) аллювиальные, (51) volcanic, (52) Andosols (T), (53) Inceptisols, and (54) Aridisols.

Earlier, it was proposed that an analogous information base could be developed as the basis of the WRB system [29, 30, 35]. From my point of view, the WRB system has great cognitive and pragmatic values as a tool to correlate between different national soil classification systems [11, 17, 18]. However, the WRB should be realized with the use of modern information technologies with inclusion of expert systems.[4] The possibility of application of expert systems toward the problems of soil classification was shown in [21, 35].

CONCLUSIONS

The use of some of the methods from the rich methodological arsenal of classiology makes it possible to unify the principles of the development of soil classification systems. To achieve success, the particular goals of a given soil classification should be strictly formulated and the rules of logic should be strictly followed. The ideas of the classification of enumeration and of the information base of classification together with the use of quantitative methods to assess information value of the particular soil characteristics seem to be promising for the development of the basic classification of soils. Such a classification should generalize classification ideas of different scientists and scientific schools with the maximum objectivity, which should ensure its broad recognition. The approaches toward the development of soil classification discussed in this paper may also serve as guidelines for the development of various target-oriented soil classification systems.

[4] An expert system is an intellectual information system consisting of declarative (the knowledge base) and procedural (rules of logic deduction and algorithms of the inference engine) parts and supported by special software.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project no. 11-04-1123a.

REFERENCES

1. *Automated Information Retrieval System POCHVA,* (VASKhNIL, Moscow, 1980), Part 1 [in Russian].
2. A. N. Afanas'ev, "Classification Problem in Pedology," in *Advances in Soil Science: Reports of Soviet Delegates to the 1st International Soil Science Congress in Washington* (Izd. Akad. Nauk SSSR, Moscow, 1927), pp. 49–108 [in Russian].
3. Yu. A. Voronin, *Theory of Classification and Its Applications* (Nauka, Novosibirsk, 1985) [in Russian].
4. K. D. Glinka, *Soil Science* (Izd. Novaya Derevnya, St. Petersburg, 1908) [in Russian].
5. G. V. Dobrovol'skii, "Some Methodological Problems of the Classification and Geography of Soils," Nauchn. Dokl. Vyssh. Shk., Biol. Nauki, No. 4, 31–38 (1970).

6. V. G. Zol'nikov, "On General Methodological Principles of the Genetic Classification of Soils," Pochvovedenie, No. 11, 70–79 (1955).
7. S. V. Zonn, "Modern Concepts of Podzolization and Pseudopodzolization and Their Manifestations in Soils," Pochvovedenie, No. 1, 142–151 (1978).
8. E. I. Ivanova and N. N. Rozov, "Classification of Soils of the Soviet Union," in *Reports of Soviet Soil Scientists to the VII International Soil Science Congress* (Izd. Akad. Nauk SSSR, Moscow, 1960), pp. 280–293 [in Russian].
9. *Classification and Diagnostic System of Russian Soils* (Oikumena, Smolensk, 2004) [in Russian].
10. V. L. Kozhara, "Problem of the Basic Classification of Landscapes and the Role of Geochemistry in Its Solution," in *Geochemistry of Landscapes and the Processes of Supergenesis* (Nauka, Moscow, 1973), pp. 132–160 [in Russian].
11. *Correlation of Soil Classifications*, (Karelsk. NTs Ross. Akad. Nauk, Petrozavodsk, 2005), p. 52 [in Russian].
12. Yu. A. Liverovskii, I. A. Sokolov, and V. O. Targulian, "On the Principles of Soil-Profile and Soil-Genetic Terminology," Pochvovedenie, No. 5, 114–121 (1973).
13. *A. A. Lyubishchev, 1890–1972* (Nauka, Leningrad, 1982) [in Russian].
14. S. V. Meien and Yu. A. Shreider, "Methodological Aspects of the Theory of Classification," Vopr. Filos., No. 12, 67–79 (1976).
15. S. V. Meien, B. S. Sokolov, and Yu. A. Shreider, "Classification and Nonclassical Biology," Vestn. Akad. Nauk SSSR, No. 10, 112–124 (1977).
16. J. St. Mill, *A System of Logic: Ratiocinative and Inductive,* Russian Translation (Izd. G.A. Lemana, Moscow, 1914) [in Russian].
17. *World Reference Base for Soil Resources: The Basis for International Classification and Correlation (Russian Edition),* (KMK, Moscow, 2007) [in Russian].
18. *Soil Nomenclature and Correlation,* (Karelsk. NTs Ross. Akad. Nauk, Petrozavodsk, 1999) [in Russian].
19. V. A. Rozhkov, *Soil Information Theory* (Agropromizdat, Moscow, 1989) [in Russian].
20. V. A. Rozhkov, "Algebra of the WRB (Formalization of the Concept," in Proc. All-Russia Conf. "Experimental Information in Soil Science: Theory and Methods of Standardization," (Izd. Mosk. Gos. Univ., Moscow, 2005), pp. 73–82 [in Russian].
21. V. A. Rozhkov and A. G. Zenin, "Knowledge Base of an Automated System of Soil Fertility Management," Tr. Pochv. Inst. im. V.V. Dokuchaeva (Moscow), pp. 86–91 (1985).
22. V. A. Rozhkov and V. S. Stolbovoi, "The Development of the Classification of Soils of the Soviet Union with the Use of an Automated Information System," Tr. Pochv. Inst. im. V.V. Dokuchaeva (Moscow), pp. 99–108 (1988).
23. S. S. Rozova, *Classification Problem in Modern Science* (Nauka, Sib. otd., Novosibirsk, 1986) [in Russian].
24. N. M. Sibirtsev, *Soil Science,* 3rd Ed. (St. Petersburg, 1914) [in Russian].
25. I. A. Sokolov, "On the Basic Classification of Soils," Pochvovedenie, No. 8, 113–123 (1978).
26. V. M. Fridland, "Some Problems of Soil Classification," Pochvovedenie, No. 7, 112–123 (1979).
27. V. M. Fridland, "Principal Features of the Profile-Genetic Component of the Basic Classification of Soils," Pochvovedenie, No. 6, 106–118 (1981).
28. V. M. Fridland, *Major Principles and Elements of the Basic Classification of Soils and the Program of Works for Its Development* (Izd. VASKhNIL, Moscow, 1982) [in Russian].
29. L. L. Shishov, V. A. Rozhkov, and V. S. Stolbovoi, "Information Base of Soil Classification," Pochvovedenie, No. 9, 9–20 (1985).
30. L. L. Shishov, V. A. Rozhkov, and V. S. Stolbovoi, "Information Reference System of Soil Classification," in *Advances in Soil Science. Reports of Soviet Soil Scientists to the XIII World Soil Science Congress* (Nauka, Moscow, 1986), pp. 191–197 [in Russian].
31. B. B. Avery, "General Soil Classification: Hierarchical and Co-Ordinate Systems," Trans. 9th Int. Congr. Soil Sci. (Adelaide, 1968), Vol. 4, 169–175.
32. W. L. Kubiena, "The Classification of Soils," J. Soil Sci. **9** (1), 9–19 (1958).
33. J. W. Muir, "The General Principal of Classification with Reference to Soils," J. Soil Sci. **13** (1), 22–30 (1962).
34. J. W. Muir, "A Natural System of Soil Classification," J. Soil Sci. **20** (1), 153–166 (1969).
35. V. A. Rojkov, V. S. Stolbovoy, B. V. Sheremet, A. G. Zenin, S. D. Orlov, "Calculation-Logical System of Maintaining the Soil Classification Information Base," in *Soil Classification* (Moscow, Nauka, 1990), pp. 27–34.