A new approach to the analysis of complex biological systems

A new approach to the analysis of complex biological systems

S I Neuymin[1,2], O Yu Olshvang[2] and S I Katsman[3]

[1] Russian Academy of Sciences, Ural Branch, Institute Botanic Garden, 202a, 8-Marta str, Ekaterinburg, 620144, Russian Federation
[2] Ural State Medical University of the Ministry of Health of the Russian Federation, Ekaterinburg, Russian Federation
[3] Ural Federal University, Ekaterinburg, Russian Federation

E-mail: sergneu@mail.ru

Abstract. A new approach to the analysis of biological systems according to the complexes of related elements is suggested. Rules for integral transformation of sets of attributes in biosystems are developed. For the first time, the resulting indicator of the system of contraction mappings (Ri SCM) that we developed was used for the analysis of biological systems. A comprehensive study of biosystems is largely determined by the algorithm for converting the totality of sequences of its elements. The structural and mathematical analysis of the complex of related elements which was used reflects the homologous variability of biological systems, based on which the classification of combinations of attributes is carried out.

1. Model of the resulting indicator of the system of contraction mappings

A comprehensive study of biosystems is largely determined by the algorithm for converting the totality of sequences of its elements. A sequence, one of the basic concepts of mathematics, is formed from elements of any nature, represented by natural numbers. A sequence is any correlation that makes up with each natural number "n" some element an of the set M [1].

The rules for the integral transformation of sequences have been developed that allow to create a “panoramic”, complex vision (study) of the “appearance” and identify the variability of plants - representatives of this population (ecotype, etc.). In our opinion, it is possible to disclose an integrated analysis scheme based on the use of the suggested algorithm of the resulting indicator of the system of contraction mappings” [2, 4, 5]. N.I. Vavilov's algebraic formula was used for designating variations within the attribute $\{a_1, a_2, a_3, a_4\ldots\}$, $\{b_1, b_2, b_3, b_4\ldots\}$, $\{b_1, b_2, b_3, b_4\ldots\}$ etc.

The flowchart for obtaining the resulting indicator of the system of contraction mappings is shown in figure 1. It contains: I-matrix of source data *K*ij. II- Matrix of dynamic contraction of source data of *S*ij system element sequences. (1):

$$\begin{array}{ccccccc} a_1\circ a_2 & a_2\circ a_1 & a_3\circ a_1 & a_4\circ a_1 & \ldots a_{k-1}\circ a_1 & a_k\circ a_1 \\ a_1\circ a_3 & a_2\circ a_3 & a_3\circ a_2 & a_4\circ a_2 & \ldots a_{k-1}\circ a_2 & a_k\circ a_2 \\ a_1\circ a_4 & a_2\circ a_4 & a_3\circ a_4 & a_4\circ a_3 & \ldots a_{k-1}\circ a_3 & a_k\circ a_3 \\ \ldots & \ldots & \ldots & \ldots & & \\ a_1\circ a_k & a_2\circ a_k & a_3\circ a_k & a_4\circ a_k & \ldots a_{k-1}\circ a_k & a_k\circ a_{3-1} \end{array} \quad (1)$$

where ○ is the operation of "contraction";

$a_1 \circ a_2$, $a_1 \circ a_3$, $a_1 \circ a_4$, ..., $a_1 \circ a_k$, ..., $a_k \circ a_1$, $a_k \circ a_2$, $a_k \circ a_3$,…, $a_k \circ a_{k-1}$ – results of sequential interaction of elements with all subsequent elements depending on their actual location in the biosystem. Accordingly, their number depends on the number of elements analyzed. $a_1 \circ a_2$ $a_2 \circ a_1$ $a_3 \circ a_1$ $a_4 \circ a_1$ … $a_k \circ a_1$ – combinations of the first level of interaction of elements; $a_1 \circ a_3$ $a_2 \circ a_3$ $a_3 \circ a_2$ $a_4 \circ a_2$… $a_k \circ a_2$ – combinations of the second level of interaction of elements; $a_1 \circ a_4$ $a_2 \circ a_4$ $a_3 \circ a_4$ $a_4 \circ a_3$… $a_k \circ a_3$ – combinations of the third level of interaction of elements, etc.

Find the sums of variants for combinations built depending on the level of interaction of the elements of the system using the matrix of the dynamic contraction of the source data S_{ij} (I). Accordingly, the sums for the variants of combinations (first – third) are written as: $(a_1 \circ a_2) + (a_2 \circ a_1) + (a_3 \circ a_1) + (a_4 \circ a_1) + \ldots + (a_k \circ a_1) = Q_1$, $(a_1 \circ a_3) + (a_2 \circ a_3) + (a_3 \circ a_2) + (a_4 \circ a_2) + \ldots + (a_k \circ a_2) = Q_2$, $(a_1 \circ a_4) + (a_2 \circ a_4) + (a_3 \circ a_4) + (a_4 \circ a_3) + \ldots + (a_k \circ a_3) = Q_3$, $\ldots (a_1 \circ a_k) + (a_2 \circ a_k) + (a_3 \circ a_k) + (a_4 \circ a_k) + \ldots + (a_k \circ a_{k-1}) = Q_{k-1}$.

Introduce the system base value «B» and express P_1, P_2, P_3, … P_{k-1} using the formula: $P_i = B \times \frac{Q_i}{Q_1 + Q_2 + \ldots + Q_{k-1}}$. Then, $P_1 + P_2 + P_3 + \ldots + P_{k-1} = B$. The base value is set for the analyzing the system as a whole. The sum of the elements included in the sequence of elements is always equal to the base value "B".

Subsequently, a new dynamic contraction matrix S_{ij} is constructed, just the values P'_1, P'_2, P'_3, …, P'_{k-1} act as elements of the systemic interaction of attributes. Thus, at each step, the matrix columns are reduced by one. The process continues until the dynamic contraction matrix S_{ij} is as follows (2).

$$\begin{vmatrix} P'_1 \circ P'_2 & P'_2 \circ P'_1 & P'_3 \circ P'_1 \\ P'_1 \circ P'_3 & P'_2 \circ P'_3 & P'_3 \circ P'_2 \end{vmatrix} \quad (2)$$

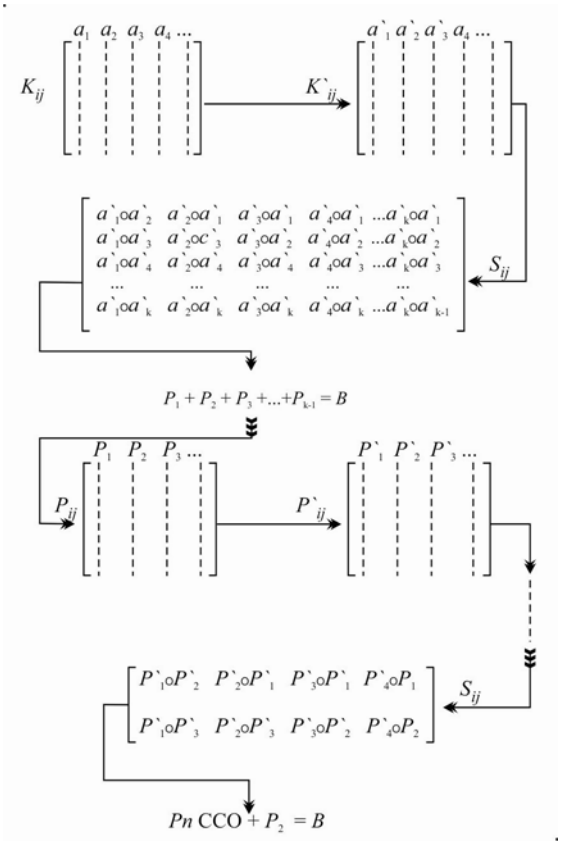

Figure 1. The flowchart "the resulting indicator of the system of contraction mappings".

The sums of combinations are calculated again according to the levels of interaction of the elements of the dynamic data contraction matrix (not the source data, but those obtained at the last stage) - S'_{ij} (figure 1). The sum of combinations of the first level is written as follows: $(P'_1 \circ P'_2) + (P'_2 \circ P'_1) + (P'_3 \circ P'_1) = Q_1$, the sum of combinations of the second level: $(P'_1 \circ P'_3) + (P'_2 \circ P'_3) + (P'_3 \circ P'_2) = Q_2$. Using the base value «*B*», $P_1 = B \times \frac{Q_1}{Q_1 + Q_2}$, $P_2 = B \times \frac{Q_2}{Q_1 + Q_2}$ is calculated.

After conversion through the system base value *"B"* we define:

$$Ri \text{ SCM} = P_1$$

2. Structural and mathematical analysis of systems of related elements

Systems biology needs a comprehensive review of the entire set of complex processes of interaction of attributes in continuous variability. In this regard, the resulting indicator of the system of contraction mappings (*Ri* SCM) suggested above can be used.

Consider the procedure for the sequential preparation of data for calculation and their transformation using the Ri SCM algorithm (table 1). Let *A, B, C,* ... be a set (biosystems, representatives of one species, or subspecies, or population, etc.); *a, b, c,* ... - attributes according to which the plants are described. Denote the number of these attributes by *n*.

Table 1. Example of an algebraic expression for source data.

Biosystem	Attributes	SRE$^{\backslash a}$	Elements
X	*a*	*Sa*	$(a_1, a_2, a_3, \dots a_m)$
	b	*Sb*	$(b_1, b_2, b_3, \dots b_m)$
	c	*Sc*	$(c_1, c_2, c_3, \dots c_m)$
	…	…	…

SRE $^{\backslash a}$ - sequence of related elements.

We assume that each of the attributes is characterized by its sequence of related elements (its members, or shares) (SRE): *Sa, Sb, Sc,*…, where $Sy = (y_1, y_2, \dots y_m)$, and $y = y_1 + y_2 + \dots + y_m$. (For simplicity, we assume that the sequence according to the attribute «*a*» – *Sa,* , the sequence according to the attribute «*b*» – *Sb,* , the sequence according to the attribute «*c*» – *Sc*, … consist of the same number of elements).

The sequences of related elements are unified using the resulting indicator of the system of contraction mappings (*Ri* SCM). The essence of unification is a spiral permutation of the numerical values of elements according to the scheme below:

Scheme (*X, y*).

Let *X* – biosystem, *y* – attribute, $Sy = (y_1, y_2, y_3, y_4, \dots, y_m)$:

$$(y_1, y_2, y_3, y_4, \dots, y_m) \rightarrow Pn_X(y_1)$$
$$(y_2, y_3, y_4, \dots, y_m, y_1) \rightarrow Pn_X(y_2)$$
$$(y_3, y_4, \dots, y_m, y_1, y_2) \rightarrow Pn_X(y_3)$$
$$\dots\dots\dots\dots \quad \dots\dots..$$
$$(y_m, y_1, y_2, y_3, y_4, \dots, y_{m-1}) \rightarrow Pn_X(y_m)$$

The result of unification is new sequences that represent the original elements:

$$\tilde{S}_x{}^y = (Pn_X(y_1), Pn_X(y_2), \ldots, Pn_X(y_m)).$$

Unification is carried out according to all attributes.

Table 2 shows the result of the unification of the elements using Algorithm 1. Unification is performed for all biosystems.

Table 2. Obtaining the resulting indicator of the system of contraction mappings by using Algorithm 1.

Biosystem	SRE\a	Algebraic formulas of *Ri* SCM, in some sequences				
		Pn_1	Pn_2	Pn_3	…	Pn_m
X	$\tilde{S}_Xa$	$Pn_X(a_1)$	$Pn_X(a_2)$	$Pn_X(a_3)$	⋯	$Pn_X(a_m)$
	$\tilde{S}_Xb$	$Pn_X(b_1)$	$Pn_X(b_2)$	$Pn_X(b_3)$	⋯	$Pn_X(b_m)$
	$\tilde{S}_Xc$	$Pn_X(c_1)$	$Pn_X(c_2)$	$Pn_X(c_3)$	⋯	$Pn_X(c_m)$
	…	…	…	…	…	…

SRE \a - sequence of related elements.

We consider the coincidence in the compared sequences of the values of *Ri* SCM (Ri_1, Ri_2, Ri_3 … Ri_m) as an indication of homologous variability. In other words, we believe that the attribute *v* of the biosystem X_1 is homologous (according to the calculation of *Ri* SCM) to the attribute *v* of the biosystem X_2 if $\tilde{S}X_1{}^y = \tilde{S}X_2{}^v$.

All the obtained values of $Ri_X(y_i)$ for all biosystems *X* and attributes *y* are arranged in a single variational series $Z = (z_1, z_2, z_3 \ldots z_k)$ in the form of an ascending scale of values.

In each biosystem *X* we determine how many times all values of the variation series *Z* are encountered; we get the sequence $(n_1, n_2, n_3, \ldots, n_k)$, and then the frequencies of the variational series are calculated from the values of *Ri* SCM – p_i:

$$p_i = \frac{n_i}{n_1 + n_2 + n_3 + \cdots + n_k} \quad (3)$$

In formula 3 – $p_1 + p_2 + p_3 + \ldots + p_k = 1$

We plot a chart where the values of the variation series *Z* are plotted vertically, and the analyzed biosystems *A, B, C, ...* are located horizontally. In each biosystem, opposite each value z_i of the variation series *Z*, the relevant frequency p_i is indicated.

We analyze the positional variability of the frequencies p_i in accordance with the algorithm for converting intrapopulation diversity [6], which is adapted for the objectives of our study.

Consider the principle of frequency analysis *Ri* SCM. Suppose that within the spectrum *Z* of the values of *Ri* SCM in a separate biosystem, *l* of specific nonzero values (morphs) were found: p_{i1}, p_{i2}, p_{i3}…, p_{il}. Then $p_{i1} + p_{i2} + p_{i3} + \ldots + p_{il} = 1$. Let *N* be the sample size, i.e. the number of elements in the SRE multiplied by the number of attributes in the biosystem. The diversity within the same frequency spectrum *Ri* SCM will be evaluated in a particular biosystem using the indicator μ.

$$\mu = (\sqrt{p}_{i1} + \sqrt{p}_{i2} + \ldots + \sqrt{p}_{il})^2 \quad (4)$$

The indicator μ provides an estimate of the diversity of the frequencies of the *Ri* SCM, the number of morphs, in our case, the number of identical values of *Ri* SCM. Its maximum possible value is *l* - with the coincidence of all the values of *Ri* SCM ($p_{i1} = 1/l, p_{i2} = 1/l, \ldots \ldots p_{il} = 1/l$). In case of uneven frequency distribution of morphs $\mu < l$. Under the monomorphism $l = 1$. To calculate the static (selective) error $S\mu$, the following formula is used:

$$S_\mu = \sqrt{\frac{\mu\,(l-\mu)}{N}} \quad (5)$$

The fraction of the rare frequencies *Ri* SCM *h* is calculated using the following formula:

$$h = 1 - \mu/l \tag{6}$$

To calculate the static (selective) error S_h, the following formula is used:

$$S_h = \sqrt{\frac{h(1-h)}{N}} \tag{7}$$

According to L.A. Zhivotovsky [6], the indicator *h* gives new information, in comparison with μ, on the nature of the positional diversity of frequencies *Ri* SCM. While μ estimates the degree of frequency diversity *Ri* SCM, the indicator *h* estimates the structure of this diversity.

To illustrate the above reasoning and constructions, we consider a specific example.

Figure 2 shows the elements obtained by analyzing the sequences of related elements: $\boldsymbol{S}a = (a_1, a_2, a_3)$, $\boldsymbol{S}b = (b_1, b_2, b_3)$, $\boldsymbol{S}c = (c_1, c_2, c_3)$, related to biosystems *A, B, C.* All elements reflect differences between themselves, which significantly complicates their comprehensive study.

Using the resulting indicator of the system of contraction mappings, the sequences of related elements of the attributes are unified according to the above Scheme (X, y). The results of this unification are shown in table 3.

Table 3. Obtaining the resulting indicator of the system of contraction mappings by using Algorithm 1.

Biosystem	SRE**	Calculation formula *			Biosystem	Numerical equivalents of *Ri* SCM		
		Pn_1	Pn_2	Pn_3		Pn_1	Pn_2	Pn_3
A	$\tilde{S}a$	$Pn_A(a_1)$	$Pn_A(a_2)$	$Pn_A(a_3)$	A	318	636	546
	$\tilde{S}b$	$Pn_A(b_1)$	$Pn_A(b_2)$	$Pn_A(b_3)$		500	375	625
	$\tilde{S}c$	$Pn_A(c_1)$	$Pn_A(c_2)$	$Pn_A(c_3)$		500	429	571
B	$\tilde{S}a$	$Pn_B(a_1)$	$Pn_B(a_2)$	$Pn_B(a_3)$	B	500	375	625
	$\tilde{S}b$	$Pn_B(b_1)$	$Pn_B(b_2)$	$Pn_B(b_3)$		318	636	546
	$\tilde{S}c$	$Pn_B(c_1)$	$Pn_B(c_2)$	$Pn_B(c_3)$		500	500	500
C	$\tilde{S}a$	$Pn_C(a_1)$	$Pn_C(a_2)$	$Pn_C(a_3)$	C	500	429	571
	$\tilde{S}b$	$Pn_C(b_1)$	$Pn_C(b_2)$	$Pn_C(b_3)$		288	678	534
	$\tilde{S}c$	$Pn_C(c_1)$	$Pn_C(c_2)$	$Pn_C(c_3)$		500	500	500

In table 4 the values of *Ri* SCM are arranged from the minimum value to the maximum from bottom to top. On the contrary, for each value of *Ri* SCM, the frequency of occurrence of this value in each biosystem is given.

Table 4. The values of the resulting indicator of the system of contraction mappings in the biosystems "*A", "B", "C"* and their occurrence in each of them.

Values of *Ri* SCM	Frequency of *Ri* SCM		
	Biosystem «*A*»	Biosystem «*B*»	Biosystem «*C*»
678	0	0	0.111111
636	0.111111	0.111111	0
625	0.111111	0.111111	0
571	0.111111	0	0.111111
545	0.111111	0.111111	0
534	0	0	0.111111
500	0.222222	0.444444	0.444444
429	0.111111	0	0.111111
375	0.111111	0.111111	0
318	0.111111	0.111111	0

288	0	0	0.111111
N	9	9	9
l	8	6	6
μ	7.87	5.44	5.44
Sμ	0.342	0.580	0.580
h	0.017	0.093	0.093
Sh	0.043	0.097	0.097

Figure 2 graphically shows the data of the reference position of the *Ri* SCM. The biosystems "*A*" and *"B*" are homologous in six, the biosystems "*A"* and "*C*" are homologous in three frequency positions. Three model biosystems are compared using the frequencies of *Ri* SCM (figure 2). In Figure 3, a dendrogram of morphogenetic distances between model biosystems is plotted based on *Jaccard Ri* SCM frequencies.

The biosystem *C* (Mgd = 0.35) has the greatest distance in comparison with the model biosystems *A* and *B*. Nevertheless, all model biosystems differ significantly from each other, although they have homologous variants of compounds.

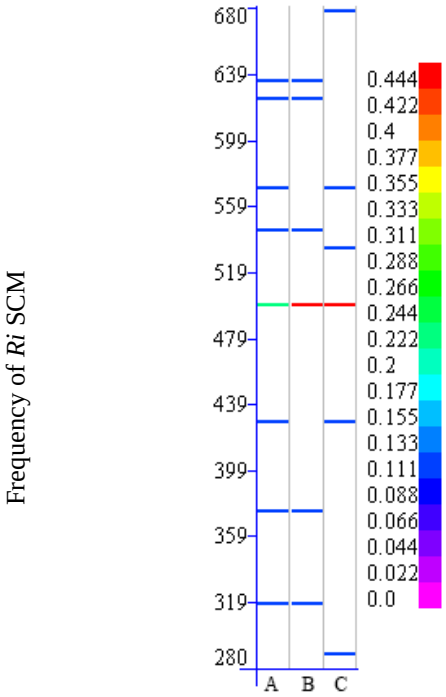

Figure 2. Reference position of frequencies Ri SCM.

where, A, B, C are the analyzed biosystems.

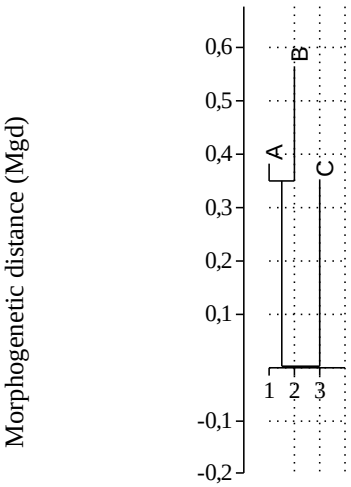

Figure 3. Morphogenetic differences of biosystems.

where, A, B, C are the analyzed biosystems.

To illustrate the work of *Ri* SCM, return to table 8. As a result of applying the unification procedure, characteristic homologous variants of the combination of attributes are obtained. In figure 4, arrows indicate homologous combinations of the quantities Ri SCM - *"B"*.

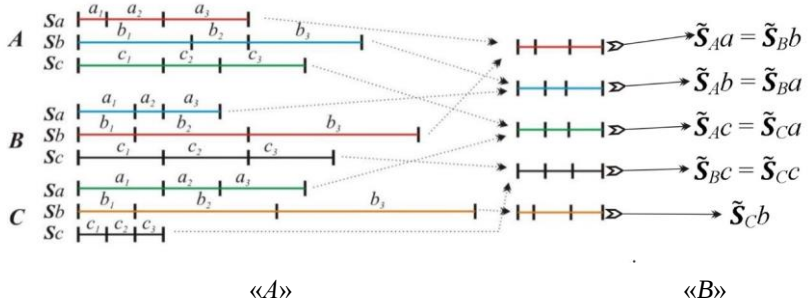

Figure 4. Variants of homologous variation of attributes.

where, *A, B, C* are biosystems of the same type; ***S*** are sequences of related elements; *a, b, c* are quantitative characteristics of these biosystems (attributes by which the biosystem is described); (a_1, a_2, a_3) sequences based on attribute *a*; (b_1, b_2, b_3) sequences based on attribute *b*; (c_1, c_2, c_3) sequences based on attribute *c;* $\tilde{S}$- a unified system of related elements; "*A"* - specific values of the observation results; "*B"* - homologous combinations of the values of *Ri* SCM.

So, for example, $\tilde{S}_A a = \tilde{S}_B b$, i.e. attribute *a* of biosystem *A* is homologous (in terms of the value of *Ri* SCM) to attribute *b* of biosystem *B.* However, the model biosystem *C* has a unique $\tilde{S}_C b$ sequence, which has no other homologous manifestations. Four variants of homologous variability and one variant with a specific sequence of compounds of attributes were identified (see *"B*") out of 9 sequences of attributes *"A"* in the presented model.

Variants of homologous variability also have their own specific positional characteristics of the values of the frequencies *Ri* SCM (figure 2).

3. Conclusion

The proposed resulting indicator of the system of contraction mappings (*Ri* SCM) largely predetermined the options for converting sequences of attributes of a particular discrete system. *Ri* SCM is the relative value obtained on the basis of the transformation of the matrix of information about the source data of that set of attributes according to which the standard equivalent segments are described.

Thus, the structural and mathematical analysis of the complex of related elements reflects the homologous variability of biosystems and made it possible to isolate and classify homologous combinations of attributes.

References

[1] Proskuryakov I V 1951 *Encyclopedia of Elementary Mathematics. Arithmetic* (State Publishing House of Technical and Theoretical Literature) **1** 77-252
[2] Neuymin S I 1991 *Ethiopian Wheat Variability System (Triticum aethiopicum Jakubz.)* (Abstract of the thesis. PhD of biol. sciences. L.) 19
[3] Neuymin S I A model of compression of the aggregate and epigenetic characteristics of plants in the form of an integral indicator
[4] Montile A A, Neuymin S I and Montile A I 2016 Software for processing a set of features of biological objects for the analysis of polymorphism and variability in biological systems *(Certificate of registration of software RUS 2016663118)*
[5] Neuymin S I and Katsman S I 2016 A mathematical model for analyzing a complex of related elements in biological systems *(Mathematical Biology and Bioinformatics: VI International Conference. Pushchino. Reports. ed. V.D. Lakhno)* 72-3
[6] Zhivotovskii L A 1979 The indicator of similarity of populations by polymorphic characters. *Journal of General Biology* **40(4)** 587-602