A new approach to the analysis of complex biological systems

S I Neuymin\textsuperscript{1,2}, O Yu Olshvang\textsuperscript{2} and S I Katsman\textsuperscript{3}

\textsuperscript{1} Russian Academy of Sciences, Ural Branch, Institute Botanic Garden, 202a, 8-Martistr, Ekaterinburg, 620144, Russian Federation
\textsuperscript{2} Ural State Medical University of the Ministry of Health of the Russian Federation, Ekaterinburg, Russian Federation
\textsuperscript{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 \cite{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” \cite{2, 4, 5}. N.I. Vavilov’s algebraic formula was used for designating variations within the attribute \{a\(_1\), a\(_2\), a\(_3\), a\(_4\),…\}, \{b\(_1\), b\(_2\), b\(_3\), b\(_4\),…\}, \{b\(_1\), b\(_2\), b\(_3\), b\(_4\),…\} 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 Kij. II- Matrix of dynamic contraction of source data of Sij system element sequences. (1):

\[
\begin{align*}
& a_1 \circ a_2 \ a_2 \circ a_1 \ a_3 \circ a_1 \ a_4 \circ a_1 \ldots a_k \circ a_1 \ a_k \circ a_1 \\
& a_1 \circ a_3 \ a_2 \circ a_3 \ a_3 \circ a_2 \ a_4 \circ a_2 \ldots a_k \circ a_2 \ a_k \circ a_2 \\
& a_1 \circ a_4 \ a_2 \circ a_4 \ a_3 \circ a_4 \ a_4 \circ a_3 \ldots a_k \circ a_3 \ a_k \circ a_3 \\
& \ldots \ldots \ldots \ldots \\
& a_1 \circ a_k \ a_2 \circ a_k \ a_3 \circ a_k \ a_4 \circ a_k \ldots a_k \circ a_k \ a_k \circ a_{k-1} \\
\end{align*}
\]
where $\circ$ is the operation of "contraction";

\[
\begin{align*}
\text{for } a_1 \circ a_2, a_1 \circ a_3, a_1 \circ a_4, \ldots, a_1 \circ a_k, a_k \circ a_2, a_k \circ a_3, \ldots, a_k \circ a_k - \text{results of sequential interaction of elements with all subsequent elements depending on their actual location in the biosystem. Accordingly, their number depends on the number of elements analyzed.}
\end{align*}
\]

\[
\begin{align*}
\text{combinations of the first level of interaction of elements;}
\end{align*}
\]

\[
\begin{align*}
\text{combinations of the second level of interaction of elements;}
\end{align*}
\]

\[
\begin{align*}
\text{combinations of the third level of interaction of elements, etc.}
\end{align*}
\]

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 $S_i (I)$. Accordingly, the sums for the variants of combinations (first – third) are written as: 

\[
\begin{align*}
(a_1 \circ a_2) + (a_1 \circ a_4) + (a_{k-1} \circ a_2) + (a_{k-1} \circ a_4) + \ldots + (a_{k-1} \circ a_k) = Q_1, (a_1 \circ a_3) + (a_{k-1} \circ a_3) + (a_{k-1} \circ a_2) + \ldots + (a_{k-1} \circ a_1) = Q_2, (a_1 \circ a_2) + (a_1 \circ a_3) + (a_{k-1} \circ a_2) + \ldots + (a_{k-1} \circ a_1) = Q_{k-1}.
\end{align*}
\]

Introduce the system base value «$B$» and express $P_1, P_2, P_3, \ldots, P_{k-1}$ using the formula: 

\[
\begin{align*}
P_i = B \times \frac{Q_i}{Q_1 + Q_2 + \ldots + Q_{k-1}}.
\end{align*}
\]

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 $S_j$ is constructed, just the values $P_1', P_2', P_3', \ldots, P_{k-1}'$ 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 $S_f$ is as follows (2).

\[
\begin{align*}
P_1' \circ P_2' & \quad P_3' \circ P_1' \quad P_3' \circ P_1' \\
P_1' \circ P_3' & \quad P_2' \circ P_3' \quad P_3' \circ P_2' \\
\end{align*}
\]

\[
\text{Figure 1. The flowchart "the resulting indicator of the system of contraction mappings".}
\]
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) - \( S_{ji} \) (figure 1). The sum of combinations of the first level is written as follows: \((P_1 \circ P_2) + (P_2 \circ P_1) = Q_1\), 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\textsuperscript{a} | Elements |
|-----------|------------|-----------------|----------|
| \( X \)   | \( a \)    | \( Sa \)         | \( (a_1, a_2, a_3, \ldots a_m) \) |
|           | \( b \)    | \( Sb \)         | \( (b_1, b_2, b_3, \ldots b_m) \) |
|           | \( c \)    | \( Sc \)         | \( (c_1, c_2, c_3, \ldots c_m) \) |
|           | (\ldots)   | (\ldots)        | (\ldots) |

\( \text{SRE}^a \) - sequence of related elements.

We assume that each of the attributes is characterized by its sequence of related elements (its members, or shares) (SRE): \( Sa, Sb, Sc, \ldots \), where \( Sy = (y_1, y_2, \ldots y_m) \), and \( y = y_1 + y_2 + \ldots + y_m \). (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, \( Sy = (y_1, y_2, y_3, y_4, \ldots, y_m) \):

\((y_1, y_2, y_3, y_4, \ldots, y_m) \rightarrow P_{nx}(y_1)\)

\((y_2, y_3, y_4, y_5, y_6, y_7, \ldots, y_{m-1}) \rightarrow P_{nx}(y_2)\)

\((y_3, y_4, y_5, y_6, y_7, \ldots, y_{m-1}, y_m) \rightarrow P_{nx}(y_3)\)

\(\ldots\)

\((y_{m-1}, y_m, y_1, y_2, y_3, y_4, \ldots, y_{m-2}) \rightarrow P_{nx}(y_{m-1})\)

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

$$\tilde{S}_X^\ast = (Pn_x(y_1), Pn_x(y_2), \ldots, Pn_x(y_m)).$$

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 |
|-----------|--------|------------------------------------------------|
| $X$       | $\tilde{S}_{Xa}$ | $Pn_x(a_1)$, $Pn_x(a_2)$, $Pn_x(a_3)$, $\ldots$, $Pn_x(a_m)$ |
|           | $\tilde{S}_{Xb}$ | $Pn_x(b_1)$, $Pn_x(b_2)$, $Pn_x(b_3)$, $\ldots$, $Pn_x(b_m)$ |
|           | $\tilde{S}_{Xc}$ | $Pn_x(c_1)$, $Pn_x(c_2)$, $Pn_x(c_3)$, $\ldots$, $Pn_x(c_m)$ |
|           | $\ldots$ | $\ldots$ |

SRE$^a$ - sequence of related elements.

We consider the coincidence in the compared sequences of the values of Ri SCM ($R_{i1}$, $R_{i2}$, $R_{i3}$ $\ldots$, $R_{im}$) 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 $\tilde{S}_X^\ast \sim \tilde{S}_X^{\prime}\ast$.

All the obtained values of $R_{i}(y_i)$ 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 $- p_i$:

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

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: $p_{i1}$, $p_{i2}$, $p_{i3}$, $\ldots$, $p_{il}$. 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 $\mu$.

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

The indicator $\mu$ 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 ($p_{i1} = 1/l$, $p_{i2} = 1/l$, $\ldots$, $p_{il} = 1/l$). 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:

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

The fraction of the rare frequencies Ri SCM $h$ is calculated using the following formula:
To calculate the static (selective) error $S_b$, the following formula is used:

$$S_b = \sqrt{\frac{h(1-h)}{N}}$$

(7)

According to L.A. Zhivotovsky [6], the indicator $h$ gives new information, in comparison with $\mu$, on the nature of the positional diversity of frequencies $Ri$ SCM. While $\mu$ 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: $Sa = (a_1, a_2, a_3)$, $Sb = (b_1, b_2, b_3)$, $Sc = (c_1, c_2, c_3)$, 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 \(^{\vee}\) | Biosystem | Numerical equivalents of $Ri$ SCM |
|-----------|-------------|------------------|-----------|------------------|
|          | $\hat{S}a$  | $P_n_1(a_1)$     | $P_n_2(a_2)$ | $P_n_3(a_3)$ | $P_n_1$ | $P_n_2$ | $P_n_3$ |
| $A$       | $\hat{S}b$  | $P_n_1(b_1)$     | $P_n_2(b_2)$ | $P_n_3(b_3)$ | 318     | 636     | 546     |
| $B$       | $\hat{S}c$  | $P_n_1(c_1)$     | $P_n_2(c_2)$ | $P_n_3(c_3)$ | 500     | 429     | 571     |
| $C$       | $\hat{S}a$  | $P_n_1(c_1)$     | $P_n_2(c_2)$ | $P_n_3(c_3)$ | 500     | 500     | 500     |
| $\hat{S}b$  | $P_n_1(b_1)$     | $P_n_2(b_2)$ | $P_n_3(b_3)$ | $P_n_1$ | $P_n_2$ | $P_n_3$ |

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        |
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.

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 2.** Reference position of frequencies Ri SCM.

**Figure 3.** Morphogenetic differences of biosystems.
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 $R_i$ SCM.

So, for example, $\tilde{S}_{Aa} = \tilde{S}_{Bb}$, i.e. attribute $a$ of biosystem $A$ is homologous (in terms of the value of $R_i$ SCM) to attribute $b$ of biosystem $B$. However, the model biosystem $C$ has a unique $\tilde{S}_{Cb}$ 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 $R_i$ SCM (figure 2).

3. Conclusion
The proposed resulting indicator of the system of contraction mappings ($R_i$ SCM) largely predetermined the options for converting sequences of attributes of a particular discrete system. $R_i$ 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