THE UNIVERSAL METHODOLOGY OF PROCESS IDENTIFICATION

Dobromir Bonacin¹ and Zdravko Carev²

¹ Faculty of Education, University of Travnik, Bosnia & Herzegovina
² Kaštela, Croatia

Abstract
By means of the model of simulating the development of pseudo-objects in a finite space the sub-segments of the totally defined process were successfully recognized and described. These sub-segments show the domination in comparison to other sub-processes in particular periods. On the basis of the positions of certain sub-segments, but also on the basis of their interactions, the bearers of the pieces of information that designate the process were identified and in terms of parameters determined as: Communicativeness, Expansion, Level of Organization, Coherence, Stableness and Harmony. By means of the algorithm that places the polar taxons cumulatively into the 'Indly' position, the completely consistent and uninterrupted series of related points were obtained that described important phases of each process: 1) release of initial energy, 2) variations induced by energy, 3) collection and amassment of materials, 4) cognition or learning, 5) external integration of resources and 6) preparation for the transfer into a new process or to a new level. This paper offers the regularities, that is, the universal parameters and phases of a process that are transparent and easily applicable in many situations and in any field.

Keywords: process, identification, taxons, time, development

Introduction
In the broadest sense, processes imply the events, regardless of the ways in which they are defined, connected with some objects or entities within a time interval set in a particular way. This implicitly means that there will be no process if there are no visible changes on objects monitored throughout a particular transformation. This is the reason why some previously set parameters – the variables by means of which these objects are measured – are used to describe and to monitor the transformation of objects. Under the classical cybernetics definitions, we are talking about compound of methods that guide us to some type of status definition and regulation, as shown by Andrews (2002), Novaković (1990), and Gordon (1978). Considering those methodological principles the existence of entire group of procedures for system analysis and system identification with final aim is evident, when processes are primarily too complex or stochastic to process analysis. Some basic examples are presented in works of Bartolomew (1973), Bartlett (1978), Campbell (1958), and Kecman (1990). In the same wide sense, a measurement implies any operation that, in congruence with a complete and accurate set of rules, makes it possible to describe a sign or a number that relates to a particular characteristic to an object which is a member of a homogeneous set of objects, so that any two objects that differ in this characteristic may be differentiated from one another according to this characteristic, and that any two objects that are identical according to this characteristic may be considered to be identical. Owing to methodology and computer development, it is possible to create projects with multivariate methods that include a large number of parameters to control events or processes as shown by Ogata (1967), Šurina (1991) and Astrom (1989).

Thus, the set of values of some variables designating a set of objects is defined. If these variables are applied in two or more consecutive time points, then it is possible to compare the data between such two points. If the results in these two points display the differences that are not the result of measurement errors, it may be correctly said that a process in the interval between the two time points affected the objects and induced the changes registered by a measurement. That was presented by Bonacin (2002), Hotelling (1933), Jardine & Sibson (1971), Momirović at al. (1987), Ristanović at al. (1986), Rosenbrock (1967) and Zadeh & Desoer (1963). By reasonably assuming that generally at least one permanent system of events may affect the objects, and consequently the results displayed as differences, the issue of defining the process is decomposition of a composite process in its parts that can be described in a particular phases as presented by Bonacin & Carev (2002) or Momirović at al. (1987). Likewise, by assuming these systems are generally of events, the elements of the composite need does not contain the phases that overlap in time, the issue of process identification apparently comes down to determining the existence, intensity, onset, duration and completion of a particular part of the sub-segmented process as shown by Bonacin & Carev (2002). Assuming that it is possible to collect the multivariate data in the space that extends over some variables and that it is possible to monitor the objects whose transformations we are interested in, through numerous consecutive time points, then the identification of any processes comes down to detecting those sub-processes that commence their dominant agency in one of these n time points which is methodologically proved by Bonacin & Carev (2002) or Momirović at al. (1987).
Likewise, in some particular point the agency of a sub-process decreases and creases to be dominant for the entire process, thus making way for another sub-process to begin. Process identification is simply the detection of the position of the time point within the total number of points when a sub-process becomes dominant. Many processes are too complex, and seemingly rarely allow too large a number of time points through which the performance of a set of objects is followed. However, all processes characterized by a set of acquired parameter values allow such an approach. Lately, the number of such processes in many scientific areas is increasingly high, for example, computer simulations, in medicine and diagnostics when entities are continuously engaged in the analyses of processes on specialized devices such as monitors or treadmills implying the analysis of ventilation-related issues, in real-time process monitoring, in tele-controlled analysis, in data analysis on the basis of different video and stimulation devices, etc. It is, therefore, possible to define such algorithms and such models of data synthesis that provide reliable process identification in technical sciences, but also elsewhere as shown by Bonacin & Carev (2002).

**Materials and Methods**

To illustrate the identification process for the purpose of this paper, the data about the development model of pseudo-objects in a finite two-dimensional space ranging from a completely empty space to the complex phenomena occurring in it were mostly generated and simulated on a computer. First of all, it was 9x9 area defined as a space where everything occurs (there were several different models, e.g., 7x7, 8x8, 10x10..., and all of them derive same results). The simplest entity was white, empty surface of 9x9 points, explaining that there was nothing. Then the authors generated few hundred entities, and the computer, using random generator function, generates more entities in that space, with different structure, from simplest point to more complex figures. Finally, the computer made a semi-random choice and chooses final 700 objects from that bulk. This number of 700 entities was not chosen randomly, because that number allows any correlation, factor or taxon saturation etc. that is larger than 0.10 to be significant at probability level of 0.01. It was done several times with randomly chosen different sets of 700 entities, and the results were always the same. Acquisitioned collection of data for 700 entities was simulated and monitored with a larger number of variables that was eventually reduced to 14 acquisitioned variables and 4 arbitrary variables. This reduction was made by classical factor model with oblique rotations defined by Momirović et al. (1987) and programmed by Bonacin (2002), so only variables with significant saturations of any factor were included in further model. Those 14 variables were: number of points for information receiving (BRPR), number of points which cannot receive direct information from outside area restricted by skeleton (BRSK), total number of points (BRTO), most distance free extern point (NSET), most distance free point in general (NSTO), maximal number of steps for information transmitting in the worst case (IMXV), minimal number of steps for information transmitting in the best case (IMNV), number of points for information emission (BREM), number of connections (BRSP), total number of lines (BRLU), number of points in "prison" (BRVR), total number of direct relations between points (BREL) and total number of free points rounded (ZAPO).

Figure 1. Some examples of primary entities

Arbitrary 4 variables were: simplicity (JEDS), reproducibility (REPR), regularity (PRAV) and symmetry (SIME). Each of those four arbitrary variables were estimated by 3 independent judges, and final estimate result was generated by their common measure, by projecting their estimates on the first factor as common measuring subject generated by factor analysis of principal components founded by Hotelling (1933) and programmed by Bonacin (2002). It is very interested that several variables showed distribution that is different than normal, but all mechanisms of higher level (taxons) show absolute normal distribution. That fact was established by standard Kolmogorov-Smirnov testing like Momirović (1987) proposed and programmed by Bonacin (2002). To achieve an accurate identification, these 18 variable data were taxonomized according to the Momirović’s (1987) model of polar taxons until the general and ultimate taxon was derived. The procedure first generates 6, then 3 taxons of higher level, then two. Finally it was one global taxon derived. Taxonomic procedure was chosen because it most efficiently describes the objects (entities), and if we want to understand processes it is obviously that we have to maintain the transformation of our data in object’s space, not in the space of variables which is frequently the case.

**Polar taxons algorithm**

Let the $Z = (z_{ij}) ; i=1,...,n; j=1,...,m$ be standardized matrix of collected data with description of some set of entities $E = \{e_i ; i=1,...,n\}$ with some variable set $V = \{v_j ; i=1,...,m\}$.
Let the $R = 1/n Z^T Z$ is variable correlation matrix. Structure definition of matrix $R$ is

$$R = \sum_{p=1}^{q} \lambda_p x_p x_p^T + \sum_{p=q+1}^{m} \lambda_p x_p x_p^T,$$

where $q = \text{num}(\lambda_p) \geq 1.0$, and where $\lambda_p$; $p=1, ..., m$ are eigen values of matrix $R$, and $x$ are eigenvectors of matrix $R$ normalised to satisfaction

$$x_p^T x_p = 1 \forall p, p=1, ..., m.$$

First we organise the first $q$ eigen values $\lambda_p$ in diagonal matrix $L = (\lambda_p)$ and the first $q$ eigenvectors in matrix $X = (x_p)$. Un-standardized principal components derived from data matrix $Z$ will be vectors $K = Z X$. Then we apply polar taxons algorithm with $T$ orthonormal matrix defined as

$$C = (g_p) = KT$$

in order to maximise Kaiser Varimax function.

Covariances of taxonomic dimensions derived by these transformations are elements of matrix;

$$C = G^T G / n = T^T L^T T.$$

If $D = \text{diag} (C)$ is one diagonal matrix of taxonomic dimensions variances, correlations of taxonomic dimensions will be

$$M = D^{-1} C D^{-1}.$$

We standardize taxonomic variables with operation $L = G D^{-1}$.

Structure of taxonomic dimensions will be in matrix $F = Z' L / n = X L T D^{-1}$, and their coordinates in oblique space defined with taxonomic vectors will be columns of matrix $A = F M^{-1} = X T D$.

Obviously, if initial objects data described by 18 variables indicates the simplest space representation, then the general and ultimate taxon represents the final solution in the defined space. It is clear by intuition, and is easy to prove that the final solution is something to which our objects (entities) are converging in defined space, according with their characteristics described with variables of lower level. Owing to large number of entities, it is easy to prove that global representation is ensured and that it is almost irrelevant if there were 700 or 7000 entities, which is firmly verified by Central Limit Theorem.

Model of polar taxons produces one bipolar characteristic for each taxon so the recognition of those characteristics is connected with: a) recognition of variables which define taxons in variable space (Table 1.), and b) recognition of typical entities with maximal projections on extreme sides of taxon (Figure 2.). Consequently, six easily interpretable taxons were obtained: 1. communicativeness (+ passivity, - activity), 2. expansion (+ material, - spatial), 3. level of organization (+ complexity, - simplicity), 4. coherence (+ dispersion, - compactness), 5. stableness (+ stability, - variability) and 6. harmony (+ order, - chaos).

In that moment, the “time continuum” was defined by sorted taxonomic data according with general taxon, although the authors rather call it “development continuum” because of the higher position of entity which this taxon represents; higher level of development. Although that continuum implies 700 points these 6 taxons were monitored for the purpose of an easier identification of sub-processes, that is, through a set of time points dominated by a particular sub-process.

<table>
<thead>
<tr>
<th>Taxonomic Variables</th>
<th>Tax1</th>
<th>Tax2</th>
<th>Tax3</th>
<th>Tax4</th>
<th>Tax5</th>
<th>Tax6</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRPR</td>
<td>0.53</td>
<td>0.12</td>
<td>0.44</td>
<td>0.09</td>
<td>0.33</td>
<td>-0.38</td>
</tr>
<tr>
<td>XXXX</td>
<td>0.49</td>
<td>0.40</td>
<td>-0.42</td>
<td>-0.26</td>
<td>-0.22</td>
<td>0.12</td>
</tr>
<tr>
<td>BRSK</td>
<td>0.72</td>
<td>0.30</td>
<td>-0.15</td>
<td>0.13</td>
<td>-0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>BRTQ</td>
<td>0.66</td>
<td>0.41</td>
<td>-0.08</td>
<td>-0.10</td>
<td>0.30</td>
<td>0.11</td>
</tr>
<tr>
<td>JEDS</td>
<td>-0.93</td>
<td>0.35</td>
<td>0.05</td>
<td>0.06</td>
<td>0.04</td>
<td>0.15</td>
</tr>
<tr>
<td>REPR</td>
<td>-0.89</td>
<td>0.48</td>
<td>0.10</td>
<td>0.04</td>
<td>-0.04</td>
<td>0.11</td>
</tr>
<tr>
<td>NSET</td>
<td>-0.57</td>
<td>-0.56</td>
<td>-0.09</td>
<td>0.08</td>
<td>0.27</td>
<td>0.30</td>
</tr>
<tr>
<td>NSTO</td>
<td>-0.60</td>
<td>-0.52</td>
<td>-0.13</td>
<td>0.06</td>
<td>0.19</td>
<td>0.38</td>
</tr>
<tr>
<td>PRAV</td>
<td>-0.68</td>
<td>0.83</td>
<td>0.09</td>
<td>0.21</td>
<td>-0.07</td>
<td>-0.04</td>
</tr>
<tr>
<td>SIME</td>
<td>-0.47</td>
<td>0.89</td>
<td>0.12</td>
<td>0.28</td>
<td>0.00</td>
<td>-0.10</td>
</tr>
<tr>
<td>IMXV</td>
<td>0.36</td>
<td>0.04</td>
<td>0.84</td>
<td>-0.11</td>
<td>-0.19</td>
<td>0.39</td>
</tr>
<tr>
<td>IMNV</td>
<td>0.35</td>
<td>0.16</td>
<td>0.71</td>
<td>-0.16</td>
<td>-0.30</td>
<td>0.48</td>
</tr>
<tr>
<td>BREM</td>
<td>-0.03</td>
<td>-0.26</td>
<td>0.68</td>
<td>0.12</td>
<td>0.12</td>
<td>-0.44</td>
</tr>
<tr>
<td>BRSP</td>
<td>0.46</td>
<td>-0.01</td>
<td>-0.13</td>
<td>0.75</td>
<td>0.09</td>
<td>0.23</td>
</tr>
<tr>
<td>BRUL</td>
<td>0.51</td>
<td>0.00</td>
<td>-0.06</td>
<td>0.77</td>
<td>0.05</td>
<td>0.18</td>
</tr>
<tr>
<td>BRVR</td>
<td>0.12</td>
<td>0.38</td>
<td>-0.03</td>
<td>-0.40</td>
<td>0.74</td>
<td>0.15</td>
</tr>
<tr>
<td>BREL</td>
<td>0.46</td>
<td>0.39</td>
<td>-0.16</td>
<td>-0.13</td>
<td>0.55</td>
<td>0.17</td>
</tr>
<tr>
<td>ZAPO</td>
<td>0.30</td>
<td>0.19</td>
<td>-0.31</td>
<td>-0.14</td>
<td>-0.80</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Figure 2. Typical entities with maximal projections on extreme sides of taxons (plus (+) and minus (-) signs represents the extreme characteristics of each taxon)
The results achieved by the objects in taxons obtained in this way were put into a system of cumulative data ranging from the simplest-initial (1) to the complex final point (700), simulating development of one object through 700 equidistant time points. The Indifg algorithm made by Momirović et al. (1987) was subsequently applied to analyse the data that were organised in this way. This algorithm decomposes the matrix of data into the left and right eigenvectors in order to identify related time points.

**Indifg algorithm**

Assuming that \( V = \{v_j; j = 1, \ldots, m\} \) is a set of quantitative variables, and \( T = \{t_i; i = 1, \ldots, n\} \) is a set of equidistant ordered time points, and \( e \) is an investigation object from some population \( P \). Operation \( P^* = T \otimes V \) is forming a matrix \( P^* = (p^*_ij); i=1,\ldots,n; j=1,\ldots,m \) as changes trajectories matrix of \( m \) variables at \( n \) time points connected with investigation object \( e \). Let \( \mu = (\mu_{ij}); i = 1 \ldots j = 1, \ldots, m \) is a result of \( m \) variables in the first time point (first row of matrix \( P^* \)), and \( \varphi = (\varphi_{ij}); i = 1, \ldots, n, \varphi_1 = 1.0 \) summ vector of \( n \) elements. Operation \( P = P^* - \varphi \mu^T \) is centering whole data matrix on the first time point value. By spectral decomposition of matrix

\[
P = \sum_{j=1}^{m} Y_q X_q^T
\]

there are defined the changes trajectories generators. Those generators could be reduced to appropriate number by some criterion: a) \( k = \) number defined in advance, b) \( k = m \) (total number of variables, taxons, factors or similar initial data), c) \( k = \text{num} (\delta \leq \sum \delta \theta 1/m) \). Let we form \( \Delta = (\delta q) \) diagonal first \( k \) eigenvalues matrix with \( \delta q \geq \delta q + 1 \), and matrix \( Y = (y q) \) with first \( k \) left eigenvectors of data matrix \( P \), as well as matrix \( X = (x q) \) with first \( k \) right eigenvectors of data matrix \( P \). Parsimony transformation left eigenvectors of data matrix is performing because of detection of related time points by keeping the system of generators in orthogonal position. That was done by satisfying Kaiser Varimax function

\[
T = (tp) = Y W
\]

\[
ww^T = w^T w = I
\]

Matrix \( T \) represents taxons of time-point set, and in the \( W \) matrix there are cosines of directional time-taxonomic vectors in order with left vectors of trajectories matrix. By orthonormal transformation of right eigenvectors of data matrix there are defined the taxons of variables with varimax function maximization, and with a condition that this system is orthogonal to

\[
F = (ip) = Q
\]

\[
qq^T = q^T q = I
\]

Matrix \( F \) contains coordinates in orthogonal time taxonomic system, and the \( Q \) matrix contains angle cosines between right vectors of trajectories matrix and vectors that defines sets of time points.

The data for each such taxon, as well as for all taxons together were graphically presented. Smoothed representation line was generated by 5th order polynomial trend-line, although 4th order trend-line generates almost identical results. All methods were generated from algorithms and methods for multivariate data analysis and data signal processing proposed in segments by Bonacin (2002), Bonacin & Carev (2002), Carev (2000), Chabert at al. (2000), Chung (2002), Jardine & Sibson (1971), Karaman & Momirović (1984), Middleton & Goodwin (1990), Momirović at al. (1987), Monroe (1962)...
The presented Figures displays the indicators of increment in each individual sub-process, that is, in the effects that comprise the total analyzed process. Apparently, these six impacts (agencies) show the shapes of time distribution precisely in the way that suggests the domination of one out of the six agencies in a particular period.

At the beginning, it is evidently instability that is dominant, which is understandable because at the beginning of each process there are numerous reasons that lead to this instability. In the end, when the total process is stilled significant passivity is expressed, because the process itself went through all previous phases. These are definitely the two basic sub-processes that indirectly describe the commencement and the completion of a process. It is obvious that stabilization occurs after approximately 1/3 of the process (approximately at the point 234). Expansion may also be recognized within ultimate limits of the process, namely, in the first and in the second part of the process, although right out dominantly after the beginning. Simultaneously, the increase in the level of organization is mostly negative, which means that the structure of the entire process is mostly simple.

Simplicity turns into complexity at the end of the process. Stableness virtually acts according to the logarithmic function, but also according to the function that describes the object that was thrown off balance and that redresses this balance slowly. Harmony behaves in the same way, but formally near process finalization. Namely, immediately after the beginning of the process it takes the form of the chaos, but it stabilizes already from the middle of the process onwards. Finally, coherence (dispersion) is directly in congruence with the second (longer) phase of the total process. It is easy to recognize the initial phases of the process in which the initial uncontrolled release of energy of the process itself is reflected, approximately up to the point 10-12 (approx. 1.5% of the total), which is particularly noticeable in Figures 3-8. This describes the regularities of the initial impetus of the process. Somewhere around the point No. 110 (approx. 15.0% of the total) the chaos of the process is somewhat more expressed due to the variations that occur more between the point’s No. 10 and 110. Simultaneously, instability disappears and after about one third of the process the amassing of materials (or perceptions) is dominant and it is reflected as dispersion of the available material.

This is the reason why the level of organization decreases significantly and will continue to decrease up to approximately 50% of the total duration of the process when the expressed increase of sources amassment is completed, and further arrangement in terms of complexity may proceed. Entirely, the listed issues describe the cognition or learning that is particularly accelerated towards the end of the process. Approximately at the half of the total process, expansion develops, that is, the external expansion of additional resources. Finally, the model becomes passive because it integrated everything that could be integrated so that it reaches the limits of the space and actually executes the preparations for opening a new space. These phases are very similar to the general phases of any process. As we can see, all these or similar parameters are sometimes partially mentioned in several publications, e.g., Akgunduz at al. (2002), Aneiros-Perez & Quintela-del-Rio (2002), Athans at al. (1974), Berlinet & Blau (2000), Chabert at al. (2000), Chen & Desoer (1967), Chung & Dey (2002), Fang (2001), Hirshman at al. (2002), Ikram & Zhou (2001), Liptser & Spokoiny (2000), Nandini & Debasis (2001), Rosenbrock (1967),..., but there is no system or coherent model that founded global parameters of process identification. That is what we present in this paper.

**Conclusion**

By the means of the model of simulating the development of pseudo-objects in a final space the sub-segments of the totally defined process were successfully recognized and described. These sub-segments show the domination in comparison to other sub-processes in particular development periods.
On the basis of the positions of certain sub-segments, but also on the basis of their interactions, the bearers of the pieces of information that designate the process were identified and in terms of parameters determined as: Communicativeness, Expansion, Level of Organization, Coherence, Stableness and Harmony. By means of the algorithm that places the polar taxons cumulatively into the Indifg position, the completely consistent and uninterrupted series of related points were obtained that described important phases of each process: 1) release of initial energy, 2) variations induced by energy, 3) collection and amassment of materials, 4) cognition or learning, 5) external integration of resources and 6) preparation for the transfer into a new process or to a new level. It may be concluded that all mentioned here, are the regularities, that is, the universal parameters and phases of a process that are transparent and easily applicable in many situations, and more attention should be paid to these issues regarding their essentials but also their utilization for the purpose of research in any field.

Figure 9. Trajectories of all sub-processes. Points 1-700 with intensity ranging from -.20 to +.20

Literature


**UNIVERZALNA METODOLOGIJA IDENTIFIKACIJE PROCESA**

**Sažetak**

Korištenjem modela simulacije razvoja pseudo-objekata u konačnom prostoru, uspješno su prepoznati i opisani subsegmenti cjelovito definiranog procesa. Ovi subsegmenti pokazuju dominaciju u usporedbi s ostalim subprocesima u pojedinom razdoblju. Na temelju pozicija pojedinih subsegmenata, ali i na temelju njihovih reakcija, granice dijelova informacija koje označavaju procese su identificirani i u terminima parametara određeni kao: Komunikativnost, Ekspanzija, Razina organizacije, Usklađenost, Stabilnost i Sklad.

Korištenjem algoritma koji kumulativno locira taksone u indifg poziciju, dobivene su potpuno dosljedne i neprekine serije taksona koje opisuju važne faze svakog procesa: 1) oslobađanje inicijalne energije, 2) varijacije inducirane energijom, 3) skupljanje materijala, 4) spoznavanje ili učenje, 5) eksternu integraciju resursa i 6) pripremu za prijelaz u novi process ili na novu razinu. Ovaj članak nudi pravila, odnosno, univerzalne parametre i faze procesa koji su transparentni i lako primjenjivi u mnogim situacijama i u bilo kojem području.

**Ključne riječi:** proces, identifikacija, taksoni, vrijeme, razvoj

Received: August 05, 2009.  
Accepted: November 22, 2009.  
Correspondence to:  
Asst.Prof. Dobromir Bonacin, PhD  
University of Travnik  
Faculty of Education  
72270 Travnik, Aleja konzula 5, Bosnia & Herzegovina  
Tel: 00 387 (0)62 90 74 30  
E-mail: dobromir.bonacin@st.t-com.hr

This paper was primary published in “Journal of Theoretics” Internet journal in 2001. It is a result of work within the framework of the research on “Process identification” by Bonacin, D. As the time passes, freshens of the article remains, so we decided to offer this significant work to wide auditorium.