Equilibrium and Extreme Principles in Discovering Unknown Relationships from Big Data
Part 2: Non-statistical Mathematical Methods in Project Management

By Pavel Barseghyan, PhD

Abstract

The pure statistical methods of Big Data mining are too data dependent, which makes them error prone to a large extent. The semi-statistical methods of Advanced Data Analytics are to some extent better, because they use data grouping/structuring before applying the statistical methods. For this reason the Advanced Data Analytics can be considered as partially data dependent and less error prone.

In order to avoid the disadvantages of pure statistical and semi-statistical methods, non-statistical methods should be used instead.

The goal of non-statistical methodologies in Big Data analysis is to derive functional relationships between the parameters of the system under study in a pure analytical way, and then apply them for data analysis and interpretation.

Non-statistical ways of quantitative description such as state equations or variational principles are common in theoretical physics and other sciences.

The second part of this paper is devoted to demonstrating how the new analytical way of Big Data Analysis can be used in the project management area. In particular it shows how this new non-statistical, analytical methodology can be used to discover unknown relationships between project parameters.

Key words: Big Data Analytics, Non-statistical methods, State equation, Variational principles, Equilibrium of projects, Functional relationships between project parameters.

Introduction

The causal relationships that lie behind the data and facts are typically the logical basis for the processing and interpretation of statistical data collected in specific areas of human activity. These causal relationships have a central role for the expert management of processes that are based on intuition and experience of humans.

If the data is processed by purely statistical methods, without structuring the data, the results are usually highly dependent on the specific data. Sometimes the results can even be almost meaningless because that direct dependency on data.

The likelihood that the patterns and relationships extracted from unstructured data by using pure statistical methods will be qualitatively adequate is negligible. In that case, if these results are qualitatively inadequate and hence they do not correctly reflect the trends contained in the data, than it is meaningless to even talk about their quantitative adequacy.

Qualitative or behavioral adequacy of patterns, extracted from the data can be improved by structuring the data, or splitting data points into groups based on sound principles or local hypotheses.

Increasing the degree of generality and logical validity of these principles and hypotheses improves the probability of retrieving qualitatively and behaviorally adequate relationships from data.

The gradual improvement of understanding and interpretation of causal relationships contained in the data, and the transition from the level of using intuition, experience and statistical perception of information to the level of systematization of knowledge in the form of mathematical models and quantitative theories, makes it possible to give a completely new, non-statistical interpretation of data.

In various fields of science such a non-statistical approach makes it possible constructing quantitative theories reflecting the deterministic structure of the causality in the form of mathematical equations.

During the 17th and 18th centuries this approach has enabled to see deterministic core lying behind a huge variety of random and chaotic nature of mechanical processes, and to describe it in the form of laws and corresponding equations of mechanics. This would have never been possible to implement on the basis of statistical interpretation of data on the mechanical processes.

Currently, the main challenge for all areas that deal with Big Data is the huge variety of forms of data, and its chaotic and random character. These factors make it extremely difficult to identify the causal relationships hidden in the data, and to describe the deterministic structure of these relationships in the form of mathematical equations.

The current state of these modern areas dealing with Big Data resembles the empirical stage of the mechanics. However, after its empirical stage mechanics took the way of extracting universal laws of deterministic nature contained in data, and not the way of their statistical processing.

This is a universal path of development that should be followed by all the areas dealing with interpretation of large data, such as marketing, project management, development of new drugs, and many others.

Non-statistical ways of Big Data analysis and interpretation

Let us briefly consider some of the methods and means of describing phenomena and processes that are widely used in various areas of quantitative science.

The main disadvantage of statistical or semi statistical methods of solving problems associated with processing of Big Data, is that they do not provide the qualitative or behavioral adequacy of the description of the investigated phenomena and processes.

The application of non-statistical methods for analyzing Big Data primarily aims to ensure the adequacy of the qualitative or behavioral description of these phenomena and processes.

The three main methods of non-statistical data analysis that are being used in classical sciences are:

1. Equations of state or macro balance conditions of equilibrium of the phenomena and processes that consolidate the parameters of investigated objects in a single expression, such as the classical equation of state of an ideal gas in physics [1].

2. Quantitative description of the phenomena and processes based on the micro balance conditions of equilibrium that usually take the form of differential equations, such as the heat transfer equation [2] and the equations of hydrodynamics [3].

3. Quantitative descriptions derived from variational principles, such as classical mechanics equations derived from the principle of least action [4, 5].

It must be emphasized that all of these approaches and means for quantitative description are closely related to the real world and strongly rely on data, facts and their generalizations coming from specific areas of human activity.

The purpose of this paper is to discuss only the issues and problems related to the use of the equation of state for the analysis and interpretation of data in the field of project management.

The meaning and use of the equations of state

Complex systems are characterized by many parameters, whose values in the state of equilibrium may vary within certain limits. This means that any system may be in equilibrium with many combinations of the numerical values of its parameters.

Each of these combinations of parameters represents a certain state of the system, and any changes of these parameter values transfer the system from one state to another.
At equilibrium of the system there is a balance between the values of system parameters that is quantitatively expressed in the form of the equation of state.

There are numerous examples of such equations of state in physics, including the equation of state of ideal gases, van der Waals equation for real gases, and many others [1].
Equations of state are powerful means that allow describing various trajectories along which systems in equilibrium move from one state to another in their parametric space.

The equation of state describes various transitions between different states of the system, but it is not able to find a specific path of transition in the parameter space without specific additional conditions.

The additional condition selects a single path among an infinite number of possible trajectories described by the equation of state. This additional condition may have the form of constancy of certain parameters, or extreme conditions, and so on.

Different combinations of additional conditions with the equation of state of a system are able to describe quantitatively the different behavioral manifestations of the system.

For example, in thermodynamics the combination of the state equation of ideal gases with the principle of conservation of energy (first law of thermodynamics) and the condition of absence of heat exchange with the environment leads to the equation of the adiabatic process [6].

Capabilities of the equations of state

Approach based on the equations of state for a long time has been used in various fields of science to describe the macro behavior of systems, and is suitable for quantitative reflection of the equilibrium state of any system.

This approach, which could be an integral part of the theory of systems, makes it possible to derive analytically functional relationships of a fundamental nature between the parameters of systems.

The main point of this approach is that the equation of state consolidates in a single mathematical expression all system-level parameters of investigated objects, together with the nonlinear functional relationships between these parameters in an implicit form.

For specific areas of knowledge, the fundamental dependencies that are derived from the equations of state create opportunities for estimations and forecasts of system parameters as well as for analyzing and interpreting the corresponding data.

In particular, when having incomplete data on specific systems, functional relationships of the fundamental nature allow to fill this gap by using the theory and then to do “what if” analysis.

To illustrate the above general provisions and in order to make them more specific, let us consider the equation of state of a project work in combination with some goal functions, such as minimum of the total project effort, or minimum risk of not finishing the project on time. Our goal will be to derive the functional relationships between project parameters analytically [7, 8].

Equilibrium of human work: State equation of projects

The general laws of human activity, according to which the work of people will be successful if there is a balance or equilibrium between the complexity of the work on the one hand, and the capacity or productivity of work performers on the other hand, are valid for project works as well [7, 9].

The state equation of the project, obtained on the basis of its equilibrium considerations has the following form [7]

$$N * T * P = W\ , \qquad (1)$$

where N - is the size of the project team [Persons],
P - is the team productivity [Complexity Units/Person Week],
T - is the duration of the project [Weeks],
W - is the complexity of the project [Complexity Units].

Project management experience and intuition of people tell that the state equation of projects (1) implicitly reflects many causal dependencies and corresponding functional relationships between the system-level parameters of the projects.

But the equation of state alone is insufficient to represent these hidden relationships explicitly. For this purpose we also need additional restrictions in the form of extreme conditions that enable choosing from the infinite number of trajectories contained in the equation of state only those that meet a specific additional condition.

In terms of successful project management, in most cases we should be interested in those trajectories contained in the state equation (1) that correspond to the extreme condition of the minimum total project effort.

The joint solution of the state equation of projects and the condition of minimum total effort

One of the main objective functions in the project planning and management is minimizing project cost, which is closely related with the requirement of minimizing total project effort E. The joint representation of the minimum project effort requirement with the equation of state of the project will have the following form:

$$\left.\begin{array}{ll} N * T * P = W & \text{(2-1)} \\ \min\ E & \text{(2-2)} \end{array}\right\} \qquad (2)$$

Let us consider a project with certain complexity W and project duration T, which is in a normal state of equilibrium and has a total effort value of E_0 and a number of performers equal to N_0. We know that there is a certain correspondence between the project complexity W and total project effort E, as well as between the total effort E and the number of performers N. Consequently, if under certain circumstances there is a change in the complexity of the project W or its duration T, the result will be a new equilibrium with some new values of total project effort E_1 and number of performers N_1.

And the change in the number of performers from N_0 into N_1 should be implemented in accordance with the requirement of minimum total effort $\min E$. In other words, the number of performers N should be changed from N_0 into N_1 in such a manner, so that in the new conditions again the project could be finished with spending the minimum possible effort.

Such a change in the project parameter space can be carried out along with the trajectory that corresponds to the rapid change, or the gradient of the number of performers N.
This means that the formulation of the problem of the analytical derivation of functional relationships between project parameters (2) can take the following form:

$$\left.\begin{array}{lr} N*T*P=W & \text{(3-1)} \\ GradN & \text{(3-2)} \end{array}\right\} \qquad (3)$$

Thus, the gradient of the number of performers is a curve in the project parameter space, whose equation can be obtained from the joint solution of the equation of state (3-1) and the condition (3-2).

The projections of the gradient curve on the coordinate planes (P,N), (P,T) and (T,N), respectively, will represent the functional relationships between the above mentioned pairs of parameters of the project.

From a mathematical point of view the solution of this problem leads to the following system of differential equations [8].

$$\frac{dW}{dP}=-\frac{P}{W}\ , \qquad (4)$$

$$\frac{dW}{dT}=-\frac{T}{W}\ , \qquad (5)$$

$$\frac{dP}{dT}=\frac{T}{P}\ . \qquad (6)$$

Assuming that the boundary conditions for the equations (4), (5) and (6) are the values W_0, N_0, P_0 and T_0, their solutions will have the following form

Table 1

Formula #	Function	Relationship
1	$N = f(T)$	$N = \frac{1}{T}\sqrt{\frac{W_0^{\;2} + T_0^{\;2} - T^2}{T^2 - T_0^{\;2} + P_0^2}}$
2	$N = f(E)$	$N = \frac{E^2}{\sqrt{(T_0^{\;2} - P_0^{\;2})E^2 + T_0^{\;2} + W_0^{\;2}}}$
3	$N = f(P)$	$N = \frac{1}{P}\sqrt{\frac{W_0^{\;2} + P_0^{\;2} - P^2}{P^2 - P_0^{\;2} + T_0^{\;2}}}$
4	$T = f(N)$	$T = \sqrt{\frac{N^2(T_0^{\;2} - P_0^{\;2}) + 1 + \sqrt{(-P_0^{\;2}N^2 + T_0^{\;2}N^2 + 1)^2 + 4N^2(W_0^{\;2} + T_0^{\;2})}}{2N^2}}$
5	$T = f(E)$	$T = \frac{\sqrt{W_0^{\;2} + T_0^{\;2} - E^2(P_0^{\;2} - T_0^{\;2})}}{E}$
6	$T = f(P)$	$T(p) = \sqrt{P^2 - P_0^{\;2} + T_0^{\;2}}$
7	$E = f(N)$	$E = \frac{1}{\sqrt{2}}\sqrt{N^2(T_0^{\;2} - P_0^{\;2}) + 1 + \sqrt{((-P_0^{\;2}N^2 + T_0^{\;2})N^2 + 1)^2 + 4(W_0^{\;2} + T_0^{\;2})}}$
8	$E = f(T)$	$E = \sqrt{\frac{W_0^{\;2} + T_0^{\;2} - T^2}{T^2 - T_0^{\;2} + P_0^{\;2}}}$
9	$E = f(P)$	$E = \frac{1}{P}\sqrt{W_0^{\;2} + P_0^{\;2} - P^2}$
10	$P = f(N)$	$P = \frac{1}{\sqrt{2}}\sqrt{(P_0^{\;2} - T_0^{\;2}) + \sqrt{(P_0^{\;2} - T_0^{\;2})^2 + \frac{4(W_0^{\;2} + P_0^{\;2})}{N^2}}}$
11	$P = f(T)$	$P = \sqrt{T^2 - T_0^{\;2} + P_0^{\;2}}$
12	$P = f(E)$	$P = \sqrt{\frac{W_0^{\;2} + P_0^{\;2}}{E^2 - 1}}$

$$W^2 - W_0^{\;2} = -p^2 + p_0^{\;2}\,, \qquad (7)$$

$$W^2 - W_0^{\;2} = -T^2 + T_0^{\;2}\,, \qquad (8)$$

$$p^2 - p_0^{\;2} = T^2 - T_0^{\;2}\,. \qquad (9)$$

The joint solution of these equations allows finding functional relationships between the parameters of the project. These relationships are shown in Table 1.

Qualitative analysis and adequacy of the derived expressions

Consider the behavior of some of the resulting functional dependencies between the parameters of projects from Table 1 [8].

1. Team productivity P vs. team size N ($P = f(N)$).

A graph of this function constructed in accordance with formula 10 of Table 1 is shown in Fig.1 and is a family of curves for various combinations of the parameters P_0, T_0 and W_0.

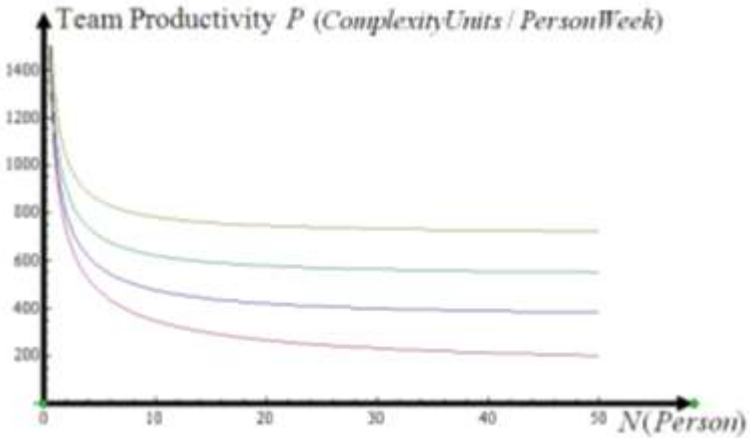

Fig.1 Team productivity vs. team size for different constant values of project parameters

As expected, the productivity of the team P is a slowly falling function of the team size N. This slow decline in productivity with the increase of team size is due to increased interaction and contacts between people.

2. Project total effort E vs. project duration T ($E = f(T)$).

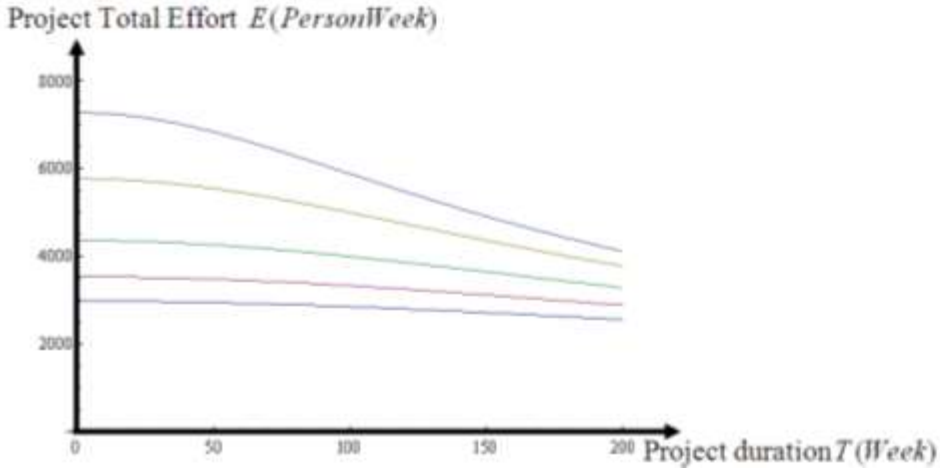

Fig.2 Project total effort vs. project duration for different constant values of project parameters

The graph of this function constructed according to formula 8 of Table 1 is shown in Fig.2. Again, it represents a family of curves for various combinations of the parameters P_0, T_0 and W_0. Decrease of the total effort E of the project as a function of its duration T is explained by the fact that at the constant complexity of the project, an increase in the duration of work is accompanied by a decrease in the size N of the team. In turn, reducing the size of the team increases its productivity, which naturally reduces the effort required to complete the project work.

3. Team size N vs. project duration T ($N = f(T)$).

The graph of this function constructed according to formula 1 of Table 1 is shown in Fig.3, and again it is a family of curves for various combinations of the parameters P_0, T_0, and W_0.

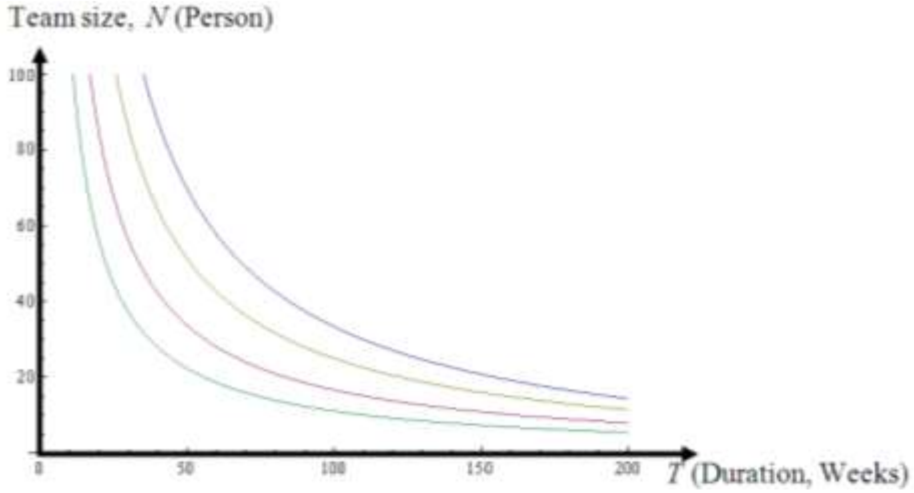

Fig.3 Team size vs. project duration for different constant values of project parameters

This function deviates from the hyperbolic dependence because the total effort of the project is also a function of the variable team productivity.

Interpretation of the obtained results from the point of view of project change management

Considering that the known values of project parameters N_0, P_0, T_0, $E_0 = N_0T_0$ and $W_0 = N_0T_0P_0$ correspond to some particular option of a project that's being planned, let's analyze the use of the obtained analytical solutions for estimating the parameters of other options of that project.

For example, suppose the original plan was to finish the project with W_0 complexity in $T_0 = 3$ years, but now it became necessary to finish it sooner, say in $T_1 = 2$ years. It is clear that this will require a new, higher number of performers N_1, which in turn will change the team productivity value from P_0 into P_1, and ultimately will require a new total project effort of E_1 as opposed to the originally required value of E_0. To estimate the new values of these three parameters we can use the formulas 1, 8 and 11 from Table1.

Another example: Suppose there is a need to change the project team size. In that case one can estimate the new values of project duration T_1, total effort E_1, and the resulting team productivity P_1 with the help of formulas 4, 7 and 10 of Table1.

Interpretation of project data using the obtained analytical expressions

The resulting mathematical relationships can be efficiently used for the analysis and interpretation of new project data. This is done by the same methods as it is being done in the areas of knowledge that have reliable quantitative methods of fundamental nature, such as mechanics or electrodynamics.

The main feature and most important advantage of this methodology is that it does not require having large amounts of data for adequate decision making, since the underlying solid theory allows meaningful interpretation of scattered sparse data and even single data points.
More details related to the analysis and interpretation of project data using this methodology can be found in [10].

Conclusions

1. Non-statistical (deterministic) methods of fundamental nature have been crucial in the development of physics and other quantitative sciences, and over time, these same techniques will win a dominant position in the modern fields that are currently facing problems of dealing with Big Data.
2. The main purpose of non-statistical methods is the description of the deterministic skeleton of the investigated processes, and to reflect the essence of existing causal relationships in them.
3. In case of the need to reflect random components of the investigated processes, the developed mathematical models of deterministic nature can be complemented by probabilistic means.
4. Short-term business goals should not deflect the development of the methods of Big Data analysis from the non-statistical (deterministic) fundamental ways, and lead it towards developing statistical and structured statistical ways by means of using modern computing technologies.
5. The development of deterministic methods in specific areas is not an easy task and requires a certain amount of time, but even imperfect methods of fundamental nature can play the role of a guide for semi statistical techniques of the modern data analytics, which could be crucial for business.

6. Different areas with Big Data problems can have their own equations of state that are the reflection of internal balance or equilibrium of the processes under study for each specific area.
7. In fact, the equation of state is the deterministic mathematical model of Big Data for each specific area.
8. The equation of state implicitly contains all the existing functional relationships between parameters of the phenomena and processes in specific areas, which in parallel are contained in Big Data too.
9. For disclosure of these functional relationships the equations of state alone are not sufficient. There is a need for specific additional conditions in order to analytically derive the existing functional relationships between parameters.
10. In this paper nonlinear functional relationships between project parameters are derived analytically.
11. The relationships obtained can be used for both planning and execution of projects, and for analyzing and interpreting data.

References

1. **Equation of state**, Wikipedia, http://en.wikipedia.org/wiki/Equation_of_state.
2. **Thermal conduction**, Wikipedia, http://en.wikipedia.org/wiki/Thermal_conduction.
3. **Fluid dynamics**, Wikipedia, http://en.wikipedia.org/wiki/Fluid_dynamics.
4. **Principle of least action**,Wikipedia, http://en.wikipedia.org/wiki/Principle_of_least_action.
5. **Variational principle**, http://en.wikipedia.org/wiki/Variational_principle.
6. **Adiabatic process**, Wikipedia, http://en.wikipedia.org/wiki/Adiabatic_process.
7. Pavel Barseghyan. (2009). **Principles of Top-Down Quantitative Analysis of Projects. Part 1: State Equation of Projects and Project Change Analysis**. PM World Today, May 2009 (Vol XI, Issue **V**). 16 pages. http://www.scribd.com/doc/113832236/Principles-of-Top-down-Analysis-of-Projects-Part-1-State-Equation-of-Projects-and-Project-Change-Analysis
8. Pavel Barseghyan (2009) **Principles of Top-Down Quantitative Analysis of Projects: Part 2: Analytical Derivation of Functional Relationships between Project Parameters without Project Data.** PM World Today, June 2009 (Vol. XI, Issue VI). 16 pages. http://www.scribd.com/doc/113911821/Principles-of-Top-down-Quantitative-

Analysis-of-Projects-Part-2-Analytical-Derivation-of-Functional-Relationships-Between-Project-Parameters-Without-P

9. Pavel Barseghyan (2013) **Equilibrium of Human Labor in Light of Supply and Demand Model from Microeconomic Theory**, *PM World Journal Volume 2, Issue 2 February 2013*.11 pages. http://pmworldjournal.net/wp-content/uploads/2013/01/pmwj7-feb2013-barseghyan-equilibrium-human-labor-featured-paper.pdf

10. Pavel Barseghyan (2010) "**Multi-parametric Models for Project Data Mining and Project Planning**". *PM World Today* – August 2010 (Vol. XII, Issue VIII). 13 pages. http://www.scribd.com/doc/114514650/Multi-Parametric-Models-for-Project-Data-Mining-and-Project-Planning

Acknowledgments

I would like to express my gratitude to Dr. Armen Vardanyan for his help in preparing the English version of both parts of this paper.

About the Author

Pavel Barseghyan, PhD

Yerevan, Armenia
Plano, Texas, USA

Dr. Pavel Barseghyan is a consultant in the field of quantitative project management, project data mining and organizational science. Has over 40 years' experience in academia, the electronics industry, the EDA industry and Project Management Research and tools development. During the period of 1999-2010 he was the Vice President of Research for Numetrics Management Systems. Prior to joining Numetrics, Dr. Barseghyan worked as an R&D manager at Infinite Technology Corp. in Texas. He was also a founder and the president of an EDA start-up company, DAN Technologies, Ltd. that focused on high-level chip design planning and RTL structural floor planning technologies. Before joining ITC, Dr. Barseghyan was head of the Electronic Design and CAD department at the State Engineering University of Armenia, focusing on development of the Theory of Massively Interconnected Systems and its applications to electronic design. During the period of 1975-1990, he was also a member of the University Educational Policy Commission for Electronic Design and CAD Direction in the Higher Education Ministry of the former USSR. Earlier in his career he was a senior researcher in Yerevan Research and Development Institute of Mathematical Machines (Armenia). He is an author of nine monographs and textbooks and more than 100 scientific articles in the area of quantitative project management, mathematical theory of human work, electronic design and EDA methodologies, and tools development. More than 10 Ph.D. degrees have been awarded under his supervision. Dr. Barseghyan holds an MS in Electrical Engineering (1967) and Ph.D. (1972) and Doctor of Technical Sciences (1990) in Computer Engineering from Yerevan Polytechnic Institute (Armenia). Pavel's publications can be found here: http://www.scribd.com/pbarseghyan and here: http://pavelbarseghyan.wordpress.com/. Pavel can be contacted at pavelbarseghyan@yahoo.com