Unifying framework for six sigma and process control

The advances presented will improve quality and productivity

P. B. DESHPANDE and **R. Z. TANTALEAN**, Six Sigma and Advanced Controls, Inc., Louisville, Kentucky

This article presents advances in six sigma and process control and illustrates how six-sigma concepts can be embedded in process control applications and vice versa. Thirty-five percent of US businesses are reported to have embraced six sigma.[1] The concepts presented could motivate many of the rest to embrace it as well. Doing so will further boost quality, productivity and competitive position. The advances reported may be used within process industries where the major impact factors are seen to be influencing both the mean and the standard deviation of response variables. Potentially, such applications exist across a wide array of diverse industries including polymerization, biological and biochemical processes, among others.

Background. The work presented was inspired by a handwritten page from the book, *Jack: Straight from the Gut* (Fig. 1).[2] Taking "Order to Delivery Time" as an illustration, Dr. Welch made a convincing case for how focusing on improving the *mean* performance (performance on average) is insufficient to compete in today's global marketplace. One must focus on reducing *variability*, that is, the mean must be moved in a favorable direction and the standard deviation reduced. When this is achieved, all the benefits of six sigma accrue. To learn how, consider the relationship between the response variable, x, deemed to be normaly distributed, and the standard normal variable, z, for an application with a double-sided specification:

$$z_1 = \frac{USL - \mu}{\sigma} \qquad (1a)$$

$$z_2 = \frac{LSL - \mu}{\sigma} \qquad (1b)$$

Here, μ is the response variable mean, σ its standard deviation, and *USL* and *LSL* are the upper and lower specification limits, respectively. The total area under the normal curve being one, the area between z_1 and z_2 represents good product while the area outside these limits represents defective product. To reduce the defective area, the values of z_1 and z_2 must be increased. The specifications *USL* and *LSL* are driven by customer needs and are not amenable to adjustments for other reasons. To increase the "good" area, therefore, the denominator of Eq. 1 must be reduced and the numerator increased. That is, the standard deviation of the response variable must be reduced and its mean moved in a favorable direction (increased, reduced or brought to target, depending on the specific application under scrutiny). Fig. 2 shows the dramatic influence of increasing z on defect reduction.

I (first author) have stared at the handwritten page for several years ever since I acquired the book. His example implied that major impact factors exist that influence both the mean and the

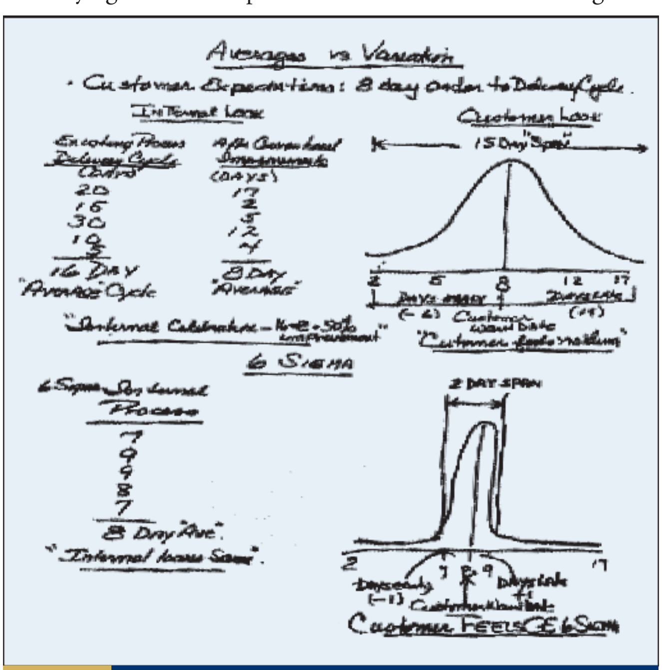

FIG. 1 Mean performance versus reducing variability with permission of Jack and Suzy Welch from *Jack: Straight from the Gut*.[2]

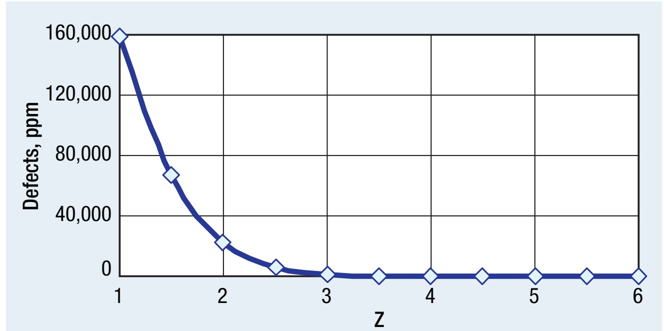

FIG. 2 Defects versus *z* (single-sided specification) for *z* = 4.5, defects are 3.45 per million opportunities.

standard deviation of response variables. Indeed, GE six-sigma professionals showed that these factors do exist for both moving the mean to target and reducing the standard deviation to +1 day.

The example Dr. Welch used to make his point was clearly a transactional problem. I was telling myself, if major impact factors existed in transactional applications that influenced the standard deviations of response variables, not just their means, they must also exist in manufacturing applications. However, I could not understand how. That is, how could the repetitive application of the same values of known major impact factors say, pH, temperature and time in an illustrative reacting system, give rise to varying values of the response variable (process outcome), say yield? We understand that the values of the response variable would not be identical time after time because of common-cause variability. Thus, if a number of experiments at the same factor levels were conducted, the response variables would not be identical due to common-cause variability arising from uncontrollable factors, but that is different from saying that the major impact factors influence the standard deviation response variables, not just their means.

To continue with the explanation, flows are routinely used in process industries as major impact factors. If air-top pressure on a control valve was chosen as a major impact factor, then the associated response variable would or could be different for the same value of the air-top pressure because of line supply-pressure changes. However, we also know how to overcome this lacuna by installing a fast control loop to maintain flow more or less constant regardless of changes in the supply pressure. Now, if flow were chosen as the major impact factor, we should get the similar values of the response variable for identical values of flow subject, of course, to common-cause variability considerations. Therefore, I was back to square one. Recently, the answer suddenly struck me. I have gathered sufficient courage to state it as a natural law:

If identical values of major impact factors lead to substantially different values of the response variable(s) over and beyond common-cause variability considerations, then the population of response variable(s) must be heterogeneous or deemed to be heterogeneous.

Heterogeneity of response variable populations can occur for two reasons: (1) There are uncertainties in the inputs (impact factors) that manifest themselves as heterogeneity of the response variable population and (2) the population of response variables is heterogeneous due to imperfect mixing in manufacturing applications. In either case, if a stratified random sample of size n representative of a heterogeneous response variable population is drawn and the mean and the standard deviation of the sample are computed, the major impact factors will be seen to be influencing both the mean and the standard deviation of the response variable. No mathematical proof of this law can be given, at least I do not know how. However, the way to disprove a natural law is to find evidence to the contrary. I look forward to reading about reader reactions to this natural law and to learning about any evidence that refutes it.

We cite two examples for further clarification of issues (1) and (2), one in nonmanufacturing and the other in manufacturing. The nonmanufacturing example is intended to shed light on the uncertainty issue (1). My associate, Mark Goldstein, a Certified Six Sigma Master Black Belt, uses a catapult experiment to demonstrate how the settings of major impact factors may be optimized to reduce variability. Here, several teams conduct experiments to show that identical values of major impact factors (e.g., pin location, launch angle and hook position) result in different values of the response variable (distance in feet the projectile travels). He also shows that the results of different teams are different owing to common-cause variability considerations. Mark then uses the data collected from designed experiments to develop two regression equations, one relating the response variable mean and the other, its standard deviation to the major impact factors. He then uses the regression models in an optimization algorithm to compute the best settings to apply for maximizing the mean and minimizing the standard deviation of the response variable. He concludes the experiment by applying the optimized settings to the catapult and demonstrates that the variability has been reduced. In this application, though it is not obvious that the population of response variables is heterogeneous, we can only say that the response variable is *deemed* to have come from a heterogeneous population.

Here, the major impact factors (pin location, launch angle and hook position) appear to be deterministic. However, certain causes may manifest themselves as uncertainties in these major impact factors rendering them stochastic. Some examples of such uncertainties are: The pin over which the rubber band passes can rotate. The viscoelastic properties of the rubber band can change over time. The duration for which the experimenter holds the launch position before releasing the projectile is not the same from one launch to the next, and so on. Because of these and other such uncertainties, the projectile may travel varying distances for the same values of major impact factors from one launch to the next.

Large batch or semibatch polymerization reactors are a good example of heterogeneity due to mixing issues in the manufacturing sector. Here, one or two samples are drawn from the sample port in the reactor vessel at the end of the batch cycle to infer product quality and the information is used in many cases to make changes in the major impact factors for the following batch in the belief that this will improve quality. Unfortunately, the population of response variable being heterogeneous because of mixing issues, the response variables may vary substantially from one location to another in the reactor. Therefore, had samples been taken from a different location in the reactor vessel, they would have suggested widely different changes in the major impact factors for the following batch.

To ascertain whether the variability in a response variable is due to common causes or uncertainties in the major impact factors or mixing issues, the suggested approach is to examine the regression equation relating the standard deviation of the response variable to the major impact factors. High-correlation coefficients (R^2) and favorable p-values are suggestive of uncertainties/missing issues while poor-correlation coefficients and unfavorable p-values are suggestive of common-cause variability. Such situations are bound to exist in batch and continuous process applications as well.

What is the big deal with this natural law anyway? Well, in the past 35 years as a chemical engineering educator and industry consultant, my students and I developed a number of control laws aimed at achieving perfection. However, we never concerned ourselves (neither did anyone else to my knowledge) with the possibility that perhaps both the mean and the standard deviation of the response variable(s) can be improved.

Based on foregoing ideas, we believe we have advanced the state-of-the-art of six sigma and advanced control of static (batch) and continuous processes that should serve as a unifying framework for six sigma and process control.

Advances in process control of static systems. This advance pertains to static systems such as batch polymerization reactors. Leffew[3] successfully tested constrained model-predictive control (CMPC) on a semibatch polymerization reactor for improved control of with in the batch operations but the focus here is on batch-to-batch operations. For clarity, let us assume that a standard recipe-driven operating strategy is used. Thus, the focus in the proposed strategy is on major impact factors (e.g., initiator, modifier, chain transfer agent concentrations, etc.) to use from one batch to the next to achieve the best value of the mean of the response variable and the least possible standard deviation. The quality attribute of the product at the end of the batch cycle is the response variable. In theory, there could be more than one response variable.

The suggested approach is to conduct classical designed experiments in which the major impact factors are varied and the product quality ascertained at the end of each batch. The magnitude of changes in the factors must be large enough to produce sufficiently large changes in the response variable. Since the population of response variables is heterogeneous, a stratified sampling plan must be used. The number of repetitions and replicates needed will depend on the heterogeneity of the population. This strategy will require a mechanism to draw random samples from various locations in the vessel and, therefore, could pose practical difficulties in some applications and will obviously incur additional expense. The financial benefits from improved product consistency, productivity and competitive position would have to justify the additional cost of multiple measurements involving repetitions and replicates. There is incentive here for companies to figure out how to make such measurements with instrumentation.

Once the experiments are devised and conducted, they will produce data that, when analyzed with standard statistical design of experiments software packages, will lead to two predictive equations per response variable: one for the mean and the other for the standard deviation, both as a function of the factors to be optimized. The terms involving products of major impact factors need to be retained in the regression model to account for the possible presence of interaction among them, something that chemical engineers typically do not do because it renders the model nonlinear. Quadratic terms could be included if warranted.

With the regression model at hand, one may proceed to compute the optimal values of the factors to apply to the process to achieve the best value of the mean of the response variable and lowest value of the standard deviation. For an illustrative system with two factors and one response variable, the procedure is as follows: The regression equations selected for the illustrative example are of the form:

$$\bar{y} = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 \qquad (2a)$$

$$s = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2 \qquad (2b)$$

The major impact factors, x_1 and x_2, are in coded form. The terms are defined under Nomenclature at the end of the article. The goal of optimization is to find the values of x_1 and x_2 such that a user-defined objective function is minimized. The complete optimization problem takes on the form:

$$Min\ J = [mega\{C_1(V_1^U + V_1^L) + C_2(V_2^U + V_2^L)\} + C_3 x_1 + C_4 x_2] \qquad (3)$$

subject to the following constraints:

$$a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 - \bar{y} = a_0 - d^1 \qquad (4a)$$

$$b_1 x_1 + b_2 x_2 + b_3 x_1 x_2 - s = -b_0 - d^2 \qquad (4b)$$

$$\bar{y} + S_1^u - V_1^u = \bar{y}^u \qquad (4c)$$

$$\bar{y} - S_1^L + V_1^L = \bar{y}^L \qquad (4d)$$

$$s + S_2^u - V_2^u = s^u \qquad (4e)$$

$$s - S_1^L + V_1^L = s^L \qquad (4f)$$

$$x_1 \geq -1 \qquad (4g)$$

$$x_1 \leq 1 \qquad (4h)$$

$$x_2 \geq -1 \qquad (4i)$$

$$x_2 \leq 1 \qquad (4j)$$

An examination of Eqs. 4a–4j reveals that the terms on the right sides involve upper and lower bounds on $\bar{y}$ and s, the biases a_0 and b_0 and the terms d^1 and d^2.

The bounds are user specifications and because the process regression model is known, the biases can be calculated from Eqs. 1a and 1b. The terms d^1 and d^2 are feedback signals, calculated by sampling the response variable at the end of each batch, computing its average and standard deviation, and subtracting from them the values $\tilde{\bar{y}}$ and $\tilde{s}$ predicted from the respective regression equation. Chemical engineers will recognize this strategy as CMPC commonly used in continuous process applications. A block diagram of the CMPC strategy is shown in Fig. 3.

The linear objective function suggested in Eq. 3 is believed to be sufficient. The goal of optimization is to compute the best possible values of $\bar{y}$ and the least possible value of s such that the optimization index, J, is minimized. Due to the large size of mega (for example, 10^6), the optimizer will first attempt to eliminate the violation variables in Eq. 3. If it can do so, only then will it focus on minimizing the costs as specified by the user. In other words, the bounds on $\bar{y}$ and s are treated as soft constraints while the bounds on x_1 and x_2 are treated as hard constraints and are therefore never violated. The terms C_1 and C_2 allow for relative weighting of the response variables while C_3 and C_4 allow for relative cost minimization of factors. Maximization may be achieved by making the signs of the cost coefficients negative. The user may specify the bounds on $\bar{y}$ and s or may specify targets. For the latter, the upper and lower bounds are set equal to the target. The number of response variables and the number of major impact factors will determine whether cost minimization (or profit maximization) is theoretically possible. This optimization problem is not amenable to linear programming due to the

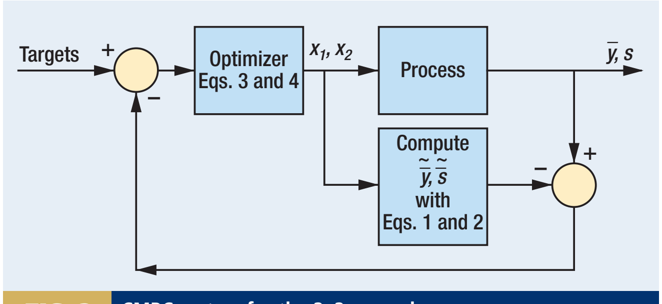

FIG. 3 CMPC system for the 2x2 example.

presence of the product x_1x_2. This means that it may not always be possible to guarantee the global optimum. A local optimal solution should be sufficient.

Statistical packages have their own algorithms for solving optimization problems such as this. Here we describe a generic procedure for solving this constrained optimization problem.

1. Assume trial values of x_1 and x_2 within allowable ranges.
2. Compute $\bar{y}$ and s from Eqs. 4a and 4b, respectively.
3. Compute the violation and slack variables from Eqs. 4c–4f.
4. Compute J. If J has reached the minimum, end. If not, return to step 1 with new trial values of the factors and repeat.

Successful optimization strategies for this problem will ensure that the trial values of the major impact factors progressively move toward the optimum from one iteration to the next. The optimization surface in this problem and in problems like this is unusual owing to the presence of the product of the factors x_1 and x_2. This poses difficulties to the optimization algorithm in its ability to find the global optimum. A local optimum solution is believed to be sufficient.

Advances in six sigma. The foregoing ideas and concepts are an advance in the state-of-the-art of six sigma as well. The advance belongs to the "Improve Phase" of six-sigma investigations. In this phase, six-sigma practitioners traditionally utilize the regression models developed in the "Analyze Phase" to find the optimal values of major impact factors to obtain the best values of the response variable averages and the least values of their respective standard deviations. When this is done, defects reduce and the benefits of six sigma accrue. In the "Control Phase", the response variables means and standard deviations are monitored typically with control charts to detect the arrival of new assignable causes. These assignable causes are worked on to return the response variables to their respective states of natural variability. The foregoing concepts suggest that the Improve Phase should not be a one-time calculation in those applications where the major impact factors *can be adjusted*. When the major impact factors cannot be manipulated, the feedback signals d^1 and d^2 are zero. But when they can be manipulated, formulating it as a CMPC problem and using exactly the same procedure will lead to improved performance.

Advances in continuous process control applications. The concepts described here appear to have significant potential in continuous process industries. The continuous process is assumed to operate under the command of a DCS system. The approach to follow is to first determine the longest closed-loop settling time of the process. This is the sampling interval for making changes to the major impact factors. Then, a host of major impact factors is identified; typically, they are the setpoints of feedback controllers or CMPC strategies. The response variables are quality attributes of the product. With this information at hand, classical designed experiments may be carried out involving repetitions and replicates to identify the major impact factors and regression models relating the response variable means and their respective standard deviations to the major impact factors. Then, the foregoing optimization strategy may be implemented to achieve the best mean and lowest standard deviations of the response variables, the quality attributes.

Example. (Adapted from the Six Sigma Black Belt Participant-Guide of Air Academy Associates, Colorado Springs, Colorado). An experimenter conducts 2^3 full-factorial designed experiments on a process with three major impact factors and one dependent variable. The input–output data are analyzed with standard statistical software leading to the following regression equations:

$$\bar{y} = 1{,}051 + 163.8x_1 + 301.6x_2 + 75.07x_3 + 44.16x_1x_2 \quad (5a)$$

$$\tilde{s} = 35.08 - 0.3951x_1 + 1.08x_2 + 2.21x_3 + 2.047x_1x_3 + 2.685x_2x_3 \quad (5b)$$

The problem is to determine the major impact factors x_1, x_2 and x_3 to apply to the process such that $1{,}025 \leq \bar{y} \leq 1{,}050$ and $32 \leq s \leq 34$. Employing the optimization strategy described earlier, constrained optimization software gives the optimal values of the major impact factors according to: $x_1 = -0.751$, $x_2 = 0.644$ and $x_3 = -0.996$. In the absence of modeling errors and load disturbances, these factors will result in $\bar{y} = 1{,}026$ and $\tilde{s} = 33$ according to Eqs. 5a and 5b. The predicted values are within the specifications.

It is hoped that this article will motivate a number of companies to embrace the ideas presented and to find novel and higher-impact applications of the methodologies as suggested here. **HP**

ACKNOWLEDGMENTS

The authors thank Dr. Kenneth W. Leffew, DuPont Fellow, and business associate Mark Goldstein, Certified Master Black Belt, formerly of General Electric for their review and comments on this article.

NOMENCLATURE

A_0	Bias in the regression model, Eq. 1
a_1–a_3	Regression coefficients in the model
b_0	Bias in the s regression model
b_1–b_3	Regression coefficients in the s model, Eq. 2
C_1, C_2	Weights on the response variable mean
C_3, C_4	Cost coefficients on the major impact factors
d^1	$\bar{y}$ (*plant*) – $\bar{y}$ (*regression*)
d^2	s(*plant*) – s(*regression*)
J	Optimization index
S	Slack variables
s	Standard deviation of the response variable, y
V	Violation variables
xs	Major impact factors
$\bar{y}$	Response variable average

SUBSCRIPTS AND SUPERSCRIPTS

u	Upper limit
L	Lower limit

LITERATURE CITED AND FURTHER READING

1 Crocket, R. O. and McGregor, G., "Six Sigma Still Pays Off at Motorola, *Business Week*, December 4, 2006. p. 50.
2 Welch, J. and Byrne, J. A., *Jack: Straight from the Gut*, Warner Books, Inc., New York, 2003.
3 Leffew K. W., "Semi-batch Polymerization Process Control for Polymers Used in Microlithography," presented at Center Jacques Cartier Conference on Modeling, Monitoring and Control of Polymer Properties, Lyon, France, 2007.
4 Burden, A. C., Tantalean, R. Z. and Deshpande, P. B., "Control and Optimize Nonlinear Processes," *Chemical Engineering Progress*, 99, 2, February 2003. pp. 63–73.
5 Deming, W. E., "Quality, Productivity, and Competitive Position," MIT Center for Advanced Engineering Study, Cambridge, Maine, 1982.
6 Deshpande, P. B., "Six Sigma for Karma Capitalism," Six Sigma and Advanced Controls, Inc., Louisville, Kentucky, first quarter 2009.
7 Deshpande, P. B. Blog: www.on-a-quest-for-change.blogspot.com.
8 Deshpande, P. B., "A Small Step for Man: Zero to Infinity with Six Sigma,"

Six Sigma and Advanced Controls, Inc., Louisville, Kentucky, 2008.
[9] Deshpande, P. B. and Tantalean, R. Z., "Process Control and Optimization," Six Sigma and Advanced Controls, Inc., June 2007.
[10] Deshpande, P. B., "Six Sigma Enlightenment," *Business World*, October 4, 2004.
[11] Deshpande, P. B., Makker, S. L. and Goldstein, M., "Boost Competitiveness with Six Sigma," *Chemical Engineering Progress*, 95, 9, September 1999. pp. 65–70.
[12] Deshpande, P. B., "Emerging technologies and six sigma," *Hydrocarbon Processing*, April 1998.
[13] Deshpande, P. B., "Globalization, Economic Development, and Competitiveness: Opportunities and Challenges," R. N. Maddox Distinguished Lecture, University of Arkansas, Fayetteville, April 1998.
[14] Deshpande, P. B., Ramasamy, S. and Yerrapragada, S. S., "Neural Nets Improve Batch Quality," *Control Engineering*, April 1996. pp. 53–56.
[15] Deshpande, P. B., Bhalodia, M. A., Caldwell, J. A., and Yerrapragada, S. S., "Should You Use Constrained Model Predictive Control?," *Chemical Engineering Progress*, 91, 3, 1995. pp. 65-72.
[16] Deshpande, P. B., Hannula R. E., Bhalodia, M. and Hansen, C. W., "Achieve Total Quality Control of Continuous Processes," *Chem. Eng. Progress*, 89, 7, 1993.
[17] Deshpande, P. B., "Improve Quality Control On-line with PID Controllers," *Chem. Eng. Progress*, 88, 5, 1992.
[18] Doyle, J. C., "Analysis of Controls Systems with Structured Uncertainty," IEE Proceedings, Part D, No. 129, 1982. P. 242.
[19] Harry, M. J. and Lawson, J. R., *Six Sigma Productivity Analysis and Process Characterization*, Addison-Wesley, Reading, Maine, 1992.
[20] Krishnaswamy, P. R., Shukla, N. V., Deshpande, P. B. and M. N. Amrouni, "Reference System Decoupling for Multivariable Control," *Ind. Eng. Chem. Research*, 30, 4, 1991.
[21] Ramasamy, S., Deshpande, P. B., Tambe, S. S. and Kulkarni, B. D., "Robust Nonlinear Control with Neural Networks," *Proceedings of the Royal Society*, Series A, London, 449, June 1995. pp. 655–667.
[22] Schmidt, S. R. and Launsby, R. G., *Understanding Industrial Designed Experiments*, Air Academy Press, Colorado Springs, Colorado, 1998.

Pradeep B. Deshpande is professor emeritus and a former chair of the Department of Chemical Engineering at the University of Louisville. He is also a visiting professor of management, Gatton College of Business & Economics, University of Kentucky, and the founder president and chief executive officer of the Louisville-based Six Sigma and Advanced Controls, Inc. (SAC). He was among the first to introduce six sigma training in corporate India and in engineering and MBA programs here, in Greece and in India. Dr. Deshpande is an author or co-author of six textbooks and over 100 refereed technical papers and presentations. During his 30 years on the faculty at the University of Louisville, he supervised 20 PhD graduates and over 40 master's graduates. Dr. Deshpande is a fellow of ISA and a recipient of numerous awards for research and teaching. He is listed in *Who's Who in the World*, and has taught at the Indian Institutes of Technology in Kanpur and Madras, University of Bombay Department of Chemical Technology and has spent a year on sabbatical at India's National Chemical Laboratory in Pune.

Roberto Tantalean is a consultant with Six Sigma and Advanced Controls, Inc. He has served as assistant professor of chemical engineering at the Universidad de Trujillo for eight years. Dr. Tantalean obtained his bachelor's degree in chemical engineering from Universidad de Trujillo, a master's degree in computer science from Universidad de Cantabria in Spain, a master's degree and a PhD in chemical engineering, both from the University of Louisville, USA. He has developed and implemented a real-time communication facility for the constrained model-predictive controller software [ONLINE]→, and implemented an expert system for fault monitoring and abnormal situation management on an industrial-scale steam plant. Dr. Tantalean held an internship as a process engineer at Talara Refinery, Perú-Petro, the main Peruvian oil company. He has been part of the lecturer team in workshops in the areas of process control for Hindustan Petroleum Company in India and Six Sigma Training for the Private Universities Council in Kuwait. Dr. Tantalean is a consultant in the area of information technology for a US-based company He is a professor for a graduate school of engineering in a Peruvian private university in Software Engineering Quality.