Interfaces
Vol. 40, No. 3, May–June 2010, pp. 208–221
ISSN 0092-2102 | EISSN 1526-551X | 10 | 4003 | 0208

informs®
DOI 10.1287/inte.1090.0488

Kaizen and Stochastic Networks Support the Investigation of Aircraft Failures

Izack Cohen
Faculty of Industrial Engineering and Management, Technion, Technion City, Haifa 32000, Israel, izik.cohen@yahoo.com

Investigating the causes of aircraft failures and preventing their reoccurrence are crucial to achieving and maintaining a high flight safety level; technical failure-analysis teams usually perform these functions. We developed and applied a dual-phased process to improve the investigative procedures that these teams use. In the first phase we used a Kaizen method to reconstruct the investigation process. In the second phase we created a simulation model of the resulting stochastic processing network to evaluate alternative configurations. The results indicated a significant improvement in throughput time and investigation quality. In addition, this unique improvement process could be adapted for use by the many organizations that concurrently run several types of jobs (or projects) in a stochastic and dynamic environment.

Key words: engineering; applications: networks, graphs; queues; simulation.
History: This paper was refereed. Published online in *Articles in Advance* April 7, 2010.

Following an aircraft safety incident, returning the fleet to flying is crucial to the aircraft operators and manufacturers. Technical failure-analysis teams must determine the failure's cause and act appropriately to prevent additional failures before decision makers and managers can decide that resuming flights is safe. These teams, which routinely operate under time pressure, perform a critical role in achieving, sustaining, and improving flight safety. They work in dynamic, stochastic environments in which they deal with randomly occurring investigations, managers who need reliable information quickly to support the decisions they must make, and scarce resources (e.g., workers, equipment, budget, and materials). In addition, the first stages of an investigation are usually characterized by a high degree of uncertainty about the failure's cause. Such a scenario often results in an environment in which more than half the time used by the investigative process is typically spent in queues waiting for resources, parts, information, or the decisions of third parties (Table 3 of Cohen et al. 2004).

We developed a dual-phased process that integrates a Kaizen project and stochastic processing networks to improve the performance of an air force failure-analysis team. In the first phase, we used a Kaizen method to reconstruct the investigation process. Kaizen is a technique of Japanese origin that is designed to rapidly improve business processes within one week. In the second phase, we created a simulation model of the resulting stochastic processing network to evaluate alternative configurations.

We combined the merits and neutralized the deficiencies of both approaches; the Kaizen method is very effective in reconstructing and implementing new processes; however, its short duration and intensive nature prevent in-depth analyses of complex environments. A stochastic processing networks' approach can provide the missing in-depth analysis; however, it is time consuming if one wants to explore various possibilities to improve a process. We address this problem by using a Kaizen project as the initial phase of the process.

In this paper, we discuss the failure-analysis and team-management processes, describe the process that we developed (i.e., our improvement framework) and its main components (i.e., Kaizen project and stochastic processing networks), detail the application process and model construction, and discuss our results and conclusions.

Aircraft Failure-Analysis Introduction

Failure analysis is a complex topic that involves many disciplines; these include mechanics, physics, metallurgy, chemistry, corrosion, manufacturing processes, stress analysis and its associated numerical techniques (e.g., finite and boundary element methods), design analysis, and fracture mechanics (e.g., environmentally induced cracking, nondestructive evaluation, and probabilistic risk evaluation, as de Castro and Fernandes 2004 describe). Because of the variety of problems that failure analysis must address and the diverse approaches to solving these problems, Thomas (1989) suggests that a single general approach cannot encompass all areas of interest. We focus on technical aircraft failure investigations that answer the following questions: What happened? How did it happen? Why did it happen? What should be done to prevent such a failure in the short, medium, and long terms?

Our failure-analysis methodology encompasses the following components:

(1) A demand for a new investigation and collection of available information and evidence;

(2) An initial investigation that includes documentation and photography, visual (macroscopic and microscopic) inspection, and characterization of materials;

(3) An initial report that comprises possible scenarios, initial recommendations, and a proposed plan of action;

(4) An investigation using laboratories, experiments, analyses, and field studies; and

(5) A final report with conclusions and recommendations.

The air force team responsible for these activities is comprised of a commander (the principal investigator), investigators, and laboratory technicians. The equipment available enables the team to prepare and analyze evidence; examples of the equipment used are cleaning and cutting machines, stereoscopes for macroscopic examinations, and a scanning electron microscope (SEM) for microscopic examinations. The team receives laboratory, engineering, and analyses services from organizations specializing in materials engineering (metallurgy and nonmetallic materials), mechanical testing and experiments, mechanical engineering and analyses (e.g., finite element models), and nondestructive inspections (NDT).

An incoming investigation is classified according to its priority. It is placed in a queue until a set of predetermined standard requirements (e.g., a collection of failure-related data and evidence and receiving of formal approval to investigate) has been met. The commander then assigns it to an investigator based on two considerations—matching the investigation to the skills of the investigator (primary) and balancing the team load (secondary). For example, a complicated investigation must be assigned to an experienced investigator. Each investigator has several concurrent investigations that might be at different stages in their life cycles. The team's main goals are to conduct high-quality investigations and minimize throughput time. Consequently, an investigator's performance is evaluated based on throughput time and the quality of completed investigations, which the principal investigator evaluates. All investigators strive to perform their work efficiently; they report to the commander at biweekly progress meetings.

The improvement initiatives arose because of the lengthy completion times of investigations (the average time was 130 days in 2005) and initial reports (the average time was 55 days in 2005) and a need to improve the investigation quality (i.e., provide a definitive answer to the "Why did it happen?" question).

High-priority investigations usually receive essential resources immediately; therefore, we focused on improving standard-priority investigations, which represent approximately 60 percent of investigations. Our objectives were to reduce the average completion time of the investigations, to reduce the average submission time of the initial reports, and to improve the quality of the investigations.

The Improvement Framework

We developed a hierarchical, dual-phased improvement process. The first phase was designed to evaluate the existing air force failure-analysis methodology and improve it by reconstructing the investigation process. In the second stage, we developed a dynamic, stochastic processing network model. Its aim was to support decision making and examine additional adjustments that could be made to the team's operating policies. In the remainder of this section, we present the Kaizen framework and elaborate on our approach to developing the model.

What Is Kaizen?

Kaizen is the Japanese word for improvement. Many companies use Kaizen projects (or events) to improve their business process performance (Bradley and Willett 2004). Unlike other gradual improvement programs, Kaizen is designed to achieve a marked improvement within one week. A Kaizen project might impose significant changes such as relocating workers, rearranging equipment, or introducing different operational methods and processes. Direct benefits often include lower throughput times, higher quality, lower costs, and better product or service levels. The improvement process involves using various tools and methods to characterize the problem, a formal root cause analysis, and the use of industrial engineering and methods of operations management. Kaizen is popular in manufacturing, service-business process, education, and organizational design settings (Bradley and Willett 2004, Brunet and New 2003, Swank 2003, Emiliani 2005, Berger 1997).

In this section, we review the Kaizen methodology steps as they apply to our work.

Step 1. Identify the problem and define the project.

Before a Kaizen project can be initiated, a specific problem must be defined (e.g., the excessively long throughput time of a standard investigation). Employee participation is required. Therefore, a Kaizen team must include members who are familiar with the problem and its complex processes. We assembled a team of investigators, customers (mainly from headquarters), suppliers (e.g., workshop technicians and a technician who develops the relevant experiments), and an objective external participant. We then gathered information about the problem, including throughput times, quality data, costs, and demand, and used this information to define quantitative goals and evaluate improvement alternatives.

Step 2. Prepare process maps.

During this stage, the team maps the existing work processes (i.e., the steps from the arrival of an investigation to its completion). These maps are very specific; each component is evaluated based on its value to the customer. If possible, the evaluation is quantitative; otherwise, it is qualitative, i.e., high, medium, or low value. Asking the customers if they would be willing to pay for a specific part of the process often helps them to evaluate the real value of that part. The other attributes (i.e., throughput time, cost, required equipment, etc.) of each component are then evaluated. This allows team members to become familiar with the overall process flow and its specific components. It creates a common understanding of the differentiated value of the process parts, enables the team to identify and exploit improvement opportunities, and reduces the number of process components that are deemed less important or that contain waste.

Our analysis of the process maps indicated that internal processes (e.g., site visit coordination, approval process of a report distribution, waiting for part cutting, etc.) were less important to our customer. Whenever possible, we reduced these processes or improved them as Step 3 describes.

Step 3. Identify improvement opportunities.

In this step, the Kaizen team generates ideas to improve the process by examining the mapped process and related data. The team then asks questions and writes suggestions for improvements. It thoroughly examines each idea, discusses its applicability for improvement, and identifies possible effects. At this stage, the team also screens ideas. For example, one idea suggested was to move a cutting machine to the failure-analysis team facilities and to train team members to operate it. The cutting-workshop technician instinctively resented this initiative (“You are taking my work”). The congenial atmosphere within the team permitted a discussion that demonstrated both the significance of the process improvement and its benefit to all stakeholders (e.g., the improvement would reduce the number of unexpected, after-hours calls for technicians). The discussion ended with a unanimous agreement to implement this initiative, which simplifies the first stages of an investigation when speed is critical. Either the investigator or laboratory technician could cut the parts, thus reducing bureaucracy and saving time.

At this stage, management usually grants the Kaizen charter that sets the project goals (e.g., to reduce an investigation’s average throughput time to 90 days).

Step 4. Improve the process.

In this stage, this step the team defines an improved process and simulates its flow. This sometimes leads to detecting process bottlenecks that must be solved; the team must estimate and assess the performance measures and match them to its charter. In complex processes, time constraints prevent in-depth simulation; therefore, the process flow and its dynamic and stochastic nature would not be realized.

The team documents the actions and resources that it needs to implement the improved process. The Kaizen policy requires the use of existing resources as much as possible (using a Kaizen project to manipulate management to obtain budgets and resources is unacceptable).

Step 5. Prepare to implement the process.

In this step, the team prepares to commence work according to the new process. The key principle is to apply all the changes promptly. At this stage, we taught the new investigation process to all involved people who were not already part of the Kaizen team, moved the cutting equipment, and began to use the improved process on new investigations.

Step 6. Gain management approval.

The final step is to present the Kaizen project results to management for its approval. Management offers its guidelines and directions about significant gaps that need resources. In our case, management was enthusiastic about the potential improvement and approved the new process, which was implemented immediately.

Stochastic Processing Networks

We modeled the investigation process using a stochastic processing network based on the works of Cohen et al. (2004) and Adler et al. (1996). Adler et al. validated the model using an existing research and development organization and showed that the model accurately simulated its performance. We considered investigations as projects because they were unique in their operational requirements and differed in activity durations. Still, they could be modeled as a group because they shared common characteristics such as the precedent relationships among the activities that comprise an investigation (e.g., measurements and nondestructive inspections of failed parts always precede macroscopic and microscopic inspections, which involve the parts-cutting procedures and irreversible chemical preparation procedures). Moreover, some activities were common to all investigations; for each such activity, the historical activity times should conform to a common distribution function. Therefore, the failure-analysis team operated in a stochastic environment of multiple, concurrent investigations that competed for the same set of scarce resources (e.g., investigators and laboratory equipment).

When several activities of an investigation can start processing simultaneously, we refer to the phenomenon as a fork. Conversely, when an activity cannot begin until the precursory activities have been completed, we call it a join. We refer to the duration required to complete an activity as its processing time and the intervals between successive investigation releases as interarrival times.

We use network diagrams to illustrate the activities' precedence requirements. The network is stochastic because interarrival times, processing times, and precedence requirements are subject to random (stochastic) variability. When an investigation arrives at a node (resource), its processing begins immediately or it is placed on a resource queue to await processing. These queues are managed according to priority rules and are subject to resource availability. Waiting also occurs in synchronization queues in which activities are delayed because of precedence constraints.

The Application

Based on our dual-phased framework, we conducted a Kaizen project and then developed and ran a simulation of a stochastic processing network model to compare and evaluate the impact of various alternatives for resource allocation and control. Our main objective was to reduce the throughput time of standard investigations (i.e., the average time from the investigation's start to writing the initial report and the average time from the investigation's start to writing the final report).

In addition, we attempted to increase the quality of the investigations to prevent future failures. First, we had to find a way to define a good-quality investigation: is it only an investigation that prevents a future failure? Our answer to that question was no; the investigation process and its final conclusions can also determine the investigation's quality. Moreover, through no fault of the investigator, a failure can occur after an investigation that had been deemed good because of its high-quality conclusions. Therefore, we

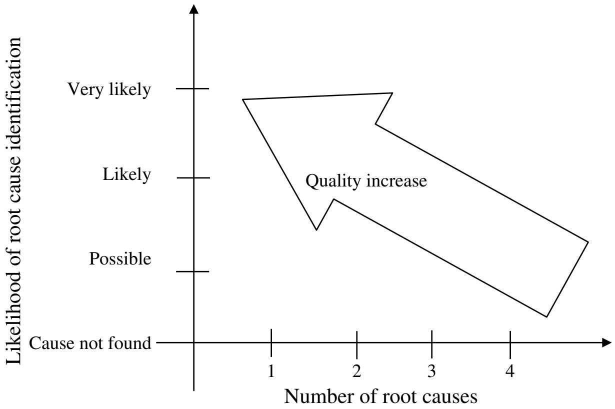

Figure 1: The graph illustrates a rating scheme to determine the quality of the investigations.

based the quality of an investigation on its final conclusions. Our aim was to focus future treatment upon the failure's root causes; a good investigation finds a few root causes. We developed a rating system for the quality of the investigations (Figure 1). In an ideal situation, an investigation indicates a single root cause of the failure with a high degree of certainty; thus, it pinpoints the treatment and prevention of future failures. The cure for a specific failure type and fleet treatment becomes more complex and labor intensive as the number of causes increases, the likelihood of finding the root causes decreases, or a combination of both factors.

The Kaizen Project Application

During the Kaizen charter meeting, we set the following performance objectives:

(1) Average throughput time: 90 days (it had previously been 130 days),

(2) Average throughput time until initial report: 14 days (it had previously been 55 days), and

(3) The percentage of investigations (with up to two root causes and likelihood levels of likely or very likely): 50 percent (it had previously been 26 percent).

The unique amalgamation of available information, the Kaizen framework, and the skillful team members from different disciplines enabled us to improve the investigation process. This process addresses the lack of managerial skills and experience of some investigators. In the previous investigation process, the performance measures were almost solely dependent upon the investigator's managerial and professional skills, thus often leading to poor performance. Analyses of investigation throughput time indicated that more than half the time was spent in queues waiting for resources, parts, data, or third-party decisions; this estimation is consistent with previous research of multiproject stochastic environments (Cohen et al. 2004). The new process (Figure 2) was designed to rectify such situations by using built-in decision points (e.g., nodes A, H, M, and N). Accurate decisions at these points should have improved the investigation's quality, reduced throughput time, and truncated delays that stalled the progress of the investigation. Most of the managerial burden was transferred from the investigator to a professional promoter, a new position created by the Kaizen team; the promoter collected the required information, coordinated meetings, and ensured that the investigator had the necessary information available. In the original process, a lab technician or an investigator had performed this function (as a secondary function). The lab technicians and investigators used to prioritize their "primary" tasks (e.g., polishing a specimen) over those of their secondary functions (i.e., the functions transferred to the promoter). Clearly, this did not reduce the investigation's throughput time.

In this section, we describe an example of the new process using the node references in Figure 2. The left main wheel of a fighter aircraft exploded during a flight takeoff; however, the crew managed to keep the plane on the runway. The air force headquarters officer initiated an investigation. The acceptance procedure (node A), which defined the investigation's goals and closure deadlines and mapped information and evidence gaps, was applied. These gaps had been identified previously as a major cause of delays in ongoing investigations. For all aircraft accidents, a site visit proximate to the mishap is crucial for a comprehensive understanding of details and the development of the events that led to the failure; the promoter coordinated the visit. The investigator inspected the runway and the aircraft and interviewed the air and technical crews (node B). The investigator took photographs at the site and collected evidence such as wheel parts. The evidence (i.e., wheel parts) and several new and used wheels

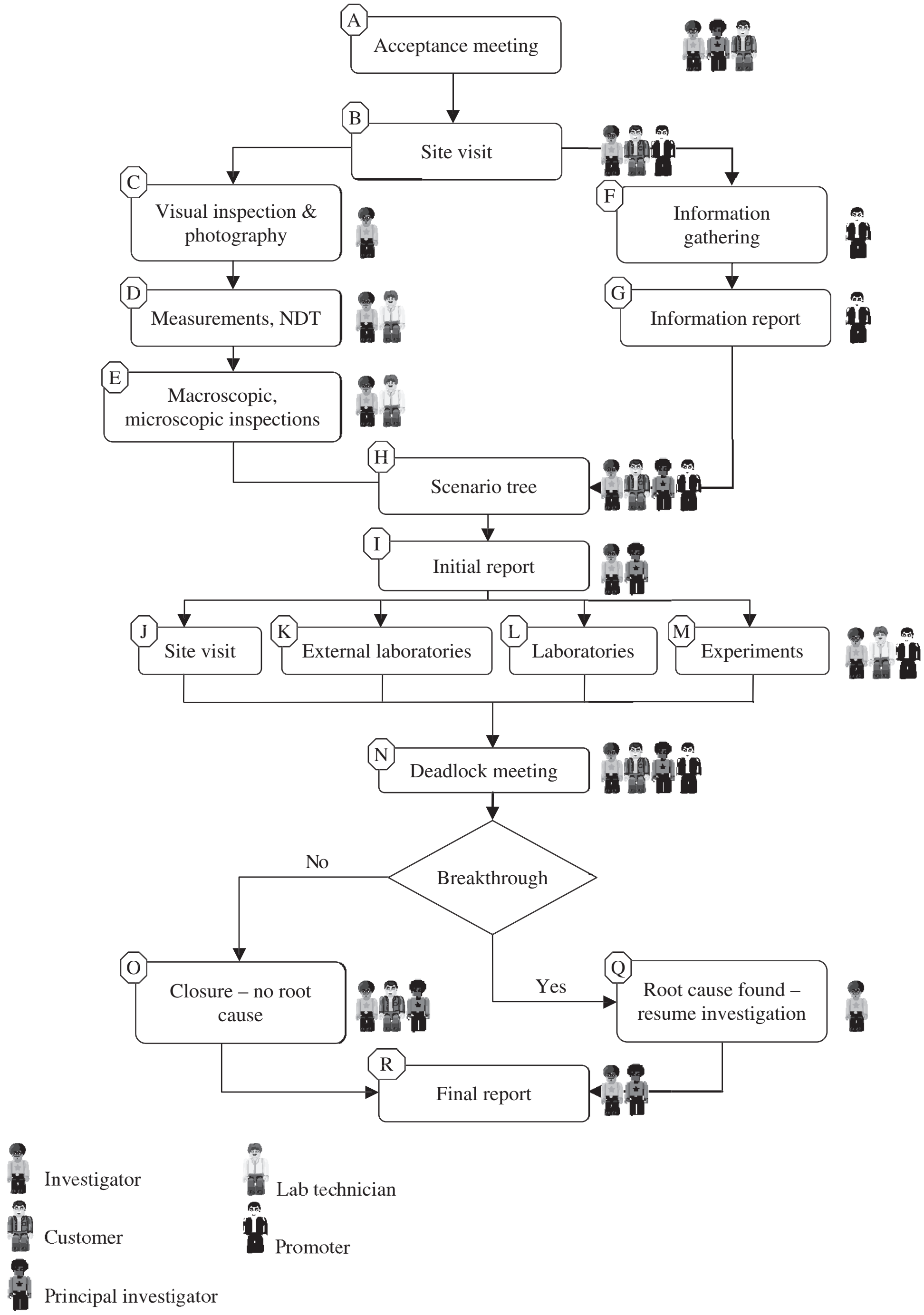

Figure 2: The schematic illustrates the new investigation process.

were brought to the laboratories for comparison and examination (nodes C, D, and E). Simultaneously, the promoter collected and analyzed information, e.g., technical instructions, maintenance manuals, and maintenance-history data records of the failed aircraft, which would be important to the investigator in formulating an assessment (nodes F and G). The aircraft's database and other operational information were analyzed to find abnormal values or patterns. In the original process, concurrent work on nodes C, D, and E and nodes F and G was not possible because the investigators or lab technicians also performed the promoter's tasks and thus were tied up with their primary work (nodes C, D, and E).

A scenario tree technique (node H) was used to focus the investigation on the most probable chain of events that led to the failure. Kaplan et al. (2005) give a detailed description of the scenario tree technique. An example of such a scenario could be that the wheel hit a foreign object on the runway, causing an initial crack that propagated rapidly because of the aircraft's weight, creating an overload of the wheel material and a total failure. However, all probable scenarios had to be considered and either be proven or refuted.

The initial report (node I) summarizes current views, recommendations, and future investigation directions. The report usually enables the customer to make preliminary decisions, e.g., return the fleet to flight but limit the aircraft's weight, inspect the wheels for damage, or keep the runways clear of foreign objects.

The investigation then follows the directions defined in the preliminary report (nodes J to M); for example, the wheel underwent thorough laboratory inspections that supported the hypothesis that the initial damage was caused by a foreign object and was propagated until failure. A series of experiments designed to repeat the progress of the failure was then conducted on wheels from different manufacturers. The investigator inspected the conditions and maintenance procedures at the various runways.

A deadlock meeting (node N) was scheduled for the shortest of the following times—60 days after the start of the investigation (preset as 75 percent of the desired investigation throughput time of 90 days) or when the investigator felt that the investigation could be finalized. The final report (node R) was then written.

The Simulation Model

The construction of the model was based on the stochastic processing network methodology (Figure 3). We assumed that release times between successive investigations followed an exponential distribution with a mean time that we calculated based on historical data over a two-year period. This assumption seems reasonable because each investigation is independent of all others. Moreover, some empirical evidence has suggested that the exponential distribution provides a reasonable fit to project activity durations. We assumed that the processing times of resources for specific activities (e.g., the duration of the promoter's coordination activities prior to a site visit, the duration of a site visit, etc.) were exponentially distributed. We did not allow multitasking or preemption, and we assumed that each resource unit processed a single activity until completion; only then would it switch to another activity. However, a higher-priority investigation must sometimes preempt an ongoing investigation. Nonetheless, the assumption was plausible because we modeled only standard-priority investigations and accordingly adjusted resource capacities, release times, etc. Investigations that entered the processing network had to be completed. The priority rule for investigations that wait in resource queues is first-come, first-served. Based on our familiarity with the process, we defined the principal investigator, investigators, laboratory technicians, promoters, and the SEM as scarce resources.

Our experimental tool was a simulation model written in C and run using a Microsoft Visual Studio C++ 2005 compiler. The simulation runs started with a warm-up period equivalent to about four years of investigations; these several hundred investigations were sufficient to ensure a steady state. We discarded the transient period. The model included 47 activities and a maximum of 20 resource units, which we allocated among the resource types. We defined the maximum work in process (WIP) as 800 investigations, a number large enough to allow long simulation runs, and replicated each simulation 50 times.

The notations and formulas that we use for the model are presented in the appendix.

Model Validation and Design of Alternative Scenarios

Model validation was necessary to establish its reliability. We simulated the model with existing resources following the Kaizen project and compared the results with actual (i.e., historical) data (case 1). We labeled our simulated model as the base case (case 2); it

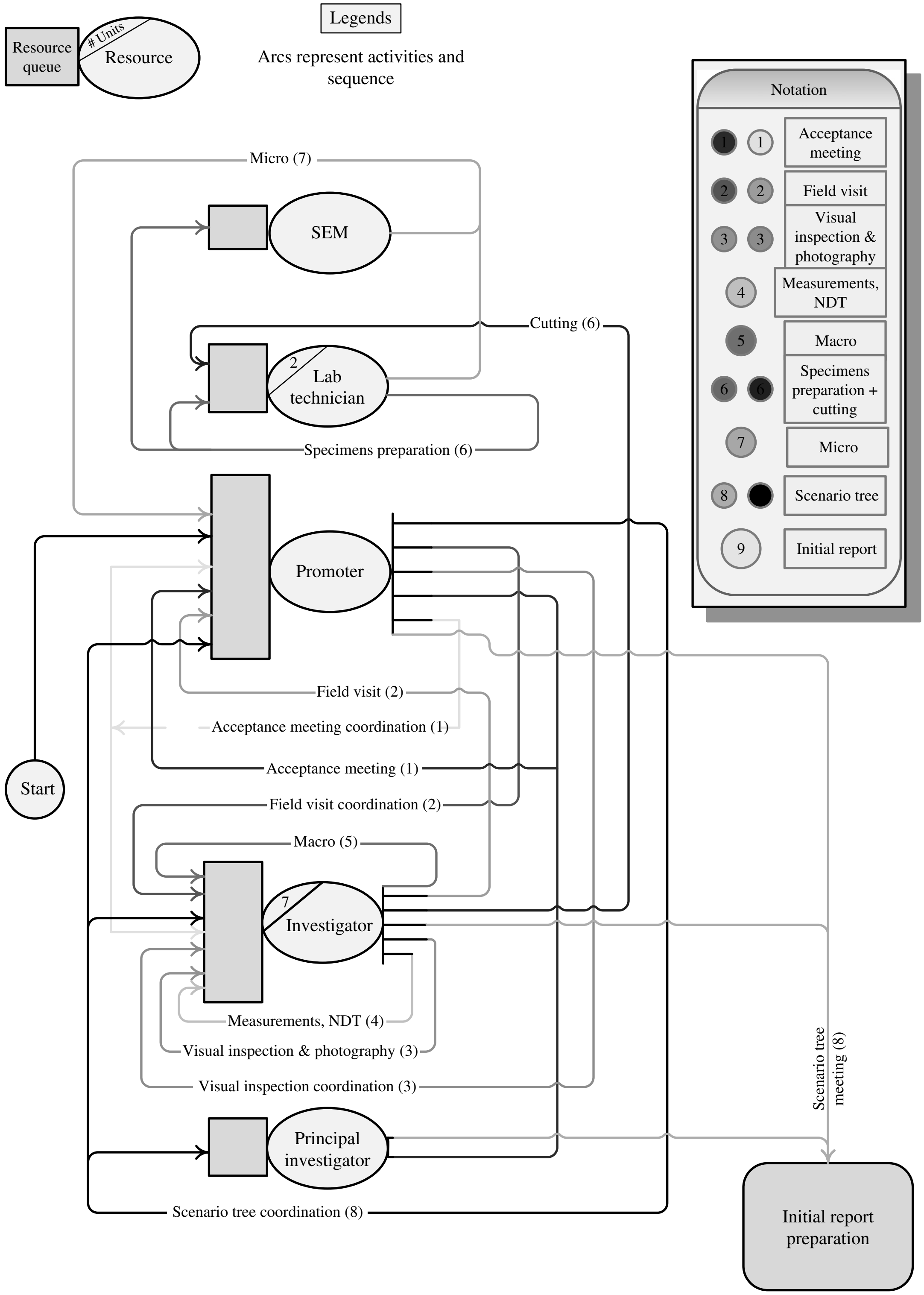

Figure 3(a): The flowchart shows the stochastic processing network model of investigations (until the initial report).

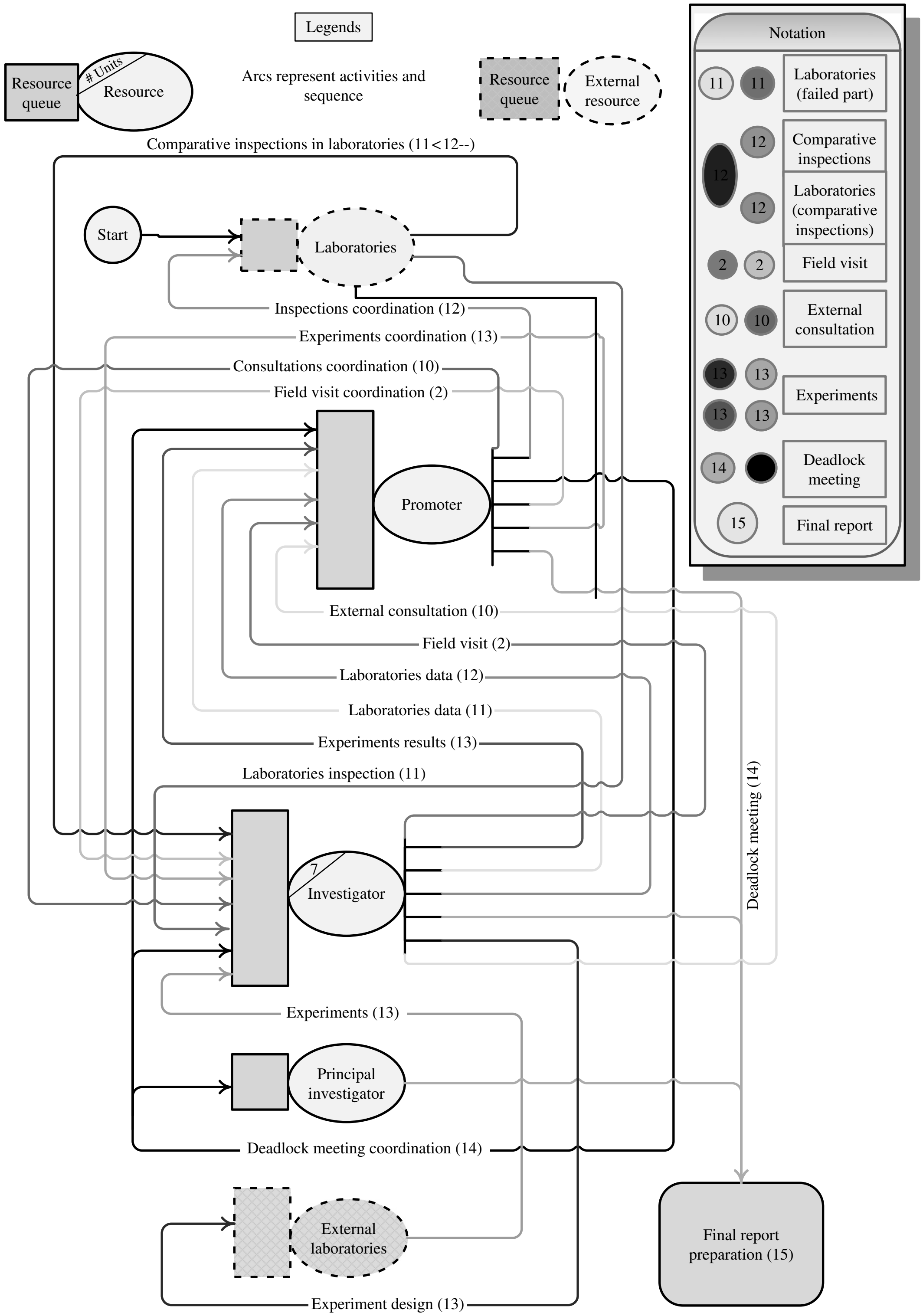

Figure 3(b): The flowchart shows the stochastic processing network model of investigations (after the initial report).

included one principal investigator, seven investigators, two lab technicians, one promoter, and one SEM. Although the original process did not use a promoter, the investigators and laboratory technicians performed the function as part of their responsibilities. In our base case, we included one promoter to perform these tasks. This was reasonable because one of our targets was to compare and evaluate the impact of different resource allocation and control methodologies in the new process. In the absence of documented information, the validation was based on an expert's opinion. Simulated results showed compatibility with the real results; for example, the simulated average throughput time to complete the initial report was 42 days compared with 55 days taken for real results—a 23 percent difference. The average simulated throughput time of an investigation was 130 days compared with 129 for real results—a 0.8 percent difference. The average number of incoming and completed simulated investigations was 74 and 68 per year, respectively, compared with 74 and 70, respectively, using real data. Simulation results indicated resource utilization levels of 97, 73, 55, 52, and 48 percent for the promoter, SEM, principal investigator, lab technicians, and investigators, respectively. When we analyzed the average time volume (work hours) that each resource spent on an individual investigation, we found that the investigator spent 52 percent (of the work hours) and the promoter spent 14 percent. Initially, when we compared the resources' time investment to their utilization levels, this surprised us. However, because seven investigators and one promoter processed several concurrent investigations, the results become self-explanatory. The investigation's throughput-time analysis showed that 45 percent was spent in resource queues. This was consistent with previous research and manifested the potential for reducing the amount of time spent in queues (Cohen et al. 2004).

Our simulation compared the impact of alternative scenarios and control methodologies. First, we added investigators, promoters, laboratory technicians, and one SEM to analyze the case in which resources were virtually unlimited (case 3). We then applied a semiclosed control, i.e., a constant number of projects in process. According to this control, new investigations are started based on a predetermined number of investigations in process (NPIP) (Anavi-Isakow and Golany 2003). A new investigation is allowed to enter and is assigned to an investigator who has fewer than the stipulated concurrent NPIP. If all investigators have been assigned NPIP investigations, the investigation waits in an external queue until it can be processed. Thus, NPIP is defined as the maximum number of investigations that an investigator can process concurrently. Multiplying NPIP by the number of investigators defines the maximum number of investigations in process.

In case 4, we reduced the WIP by setting NPIP to 5 (35 concurrent investigations were allowed compared with 49 in case 2). In case 5, we added an additional promoter to our base case to demonstrate the effect of increasing the overall number of resources. In case 6, we added an additional promoter to our base case and simultaneously decreased the number of laboratory technicians to one. Finally, in case 7, we added an additional promoter to the base case and simultaneously decreased the number of investigators to six. In cases 6 and 7, our objective was to explore different resource allocations.

Simulation Results and Insights

In this section, we describe the results of our simulation and discuss some insights that we gained. The results are summarized in Table 1. The notations we use in Table 1 are as follows:

x_1, x_2, x_3, x_4, and x_5 represent the number of principal investigators, investigators, promoters, lab technicians, and SEM, respectively.

NPIP is the maximal number of concurrent investigations assigned to an investigator.

T_{mean} and $T_{\mathrm{ir,\,mean}}$ are the average investigation throughput time and the average time to complete the initial report, respectively.

ρ_1, ρ_2, ρ_3, ρ_4, and ρ_5 represent the utilization of the principal investigator, investigators, promoters, lab technicians, and SEM, respectively.

WT represents the portion of time that an investigation waits to be serviced.

Increasing the number of resources (case 3) changed the investigations' throughput time to 77 days; this compares with 130 as in the base case. The time taken until the completion of the initial report was 22 days versus the original 42 days, thus reducing this time by 48 percent. Resources were virtually unlimited

Case no.	x_1, x_2, x_3, x_4, x_5	*NPIP*	T_{mean} (σ)	$T_{\text{ir, mean}}$	$\rho_1, \rho_2, \rho_3, \rho_4, \rho_5$	*WT* (%)
1	(1, 7, 1, 2, 1)	7	129	55	Unknown	Unknown
2	(1, 7, 1, 2, 1)	7	130 (30)	42	(0.55, 0.52, 0.97, 0.48, 0.73)	45
3	(1, 9, 3, 3, 3)	—	77 (28)	22	(0.60, 0.45, 0.35, 0.29, 0.28)	8
4	(1, 7, 1, 2, 1)	5	99* (27)	27	(0.51, 0.49, 0.92, 0.47, 0.72)	28
5	(1, 7, 2, 2, 1)	7	87 (29)	22	(0.59, 0.58, 0.52, 0.51, 0.79)	19
6	(1, 7, 2, 1, 1)	7	130 (30)	43	(0.53, 0.50, 0.46, 0.99, 0.74)	46
7	(1, 6, 2, 2, 1)	8	88 (29)	22	(0.57, 0.62, 0.49, 0.49, 0.74)	18

Table 1: The table summarizes the results for the seven cases that were evaluated.
***Long and unrealistic waiting time before initiating the investigation.**

because investigations wait in queues only 8 percent of their throughput time compared with 45 percent in the base case. This case illustrates the optimum results that we could have expected.

Reducing the maximum number of concurrent investigations from 49 (base case) to 35 (case 4) enabled a throughput time of 99 days. This was in accordance with Little's law, which states that reducing the amount of WIP, although keeping the same throughput rate, results in lower throughput times. However, the average waiting time of an incoming investigation until the acceptance meeting was 21 days, an unacceptable number of days. The promoter utilization was 92 percent, a relatively high number. We investigated the option of increasing the overall resource number by adding an additional resource unit (i.e., another promoter) in case 5; performance improved in comparison with case 4 and the base case. The throughput time was 87 days, and the time until the completion of the initial report was 22 days; this was similar to the result we achieved using unlimited resources. The promoters' utilization decreased to 52 percent (compared with 92 percent in case 4) and the portion of time that an investigation spent in queues was 19 percent. If increasing the number of resources was unreasonable, it would have been logical to eliminate the less-utilized resource type, the lab technician (51 percent). Simulating the system with only one lab technician (case 6) led to an investigation throughput time of 130 days and a mean time of 43 days until the completion of initial reports. This did not meet our objectives. Clearly, the single lab technician became the bottleneck resource with a utilization of 99 percent. Queueing theory has shown that a utilization of close to 100 percent leads to poor performance and long waiting times in queues (i.e., network congestion), which was 46 percent in this case. In case 7, we decreased the number of investigators to six and increased NPIP to eight to maintain the approximate number of concurrent investigations (48) similar to the base case (49). Our results showed that on average an investigation took 88 days to complete and the initial report took on average 22 days. Resource utilization ranged from a high of 74 percent for the SEM to a low of 49 percent for the promoters and lab technicians.

Simulation results showed that completing the initial report took an average of 22 days—even under optimum conditions. Conversely, a total throughput time of 90 days for an investigation was a realistic goal. Although cost considerations were not initially defined as a constraint, avoiding unnecessary expenses was desirable. Adding more resources increased the costs; however, changing their mix (e.g., adding a promoter instead of an investigator) was insignificant because of small differences between the salaries. When we kept the overall number of resources at 12, we found it advisable to replace one investigator with a promoter (case 7).

Conclusions

We improved the failure-analysis team's processes and focus in keeping with the insights that we gained from both the Kaizen project and the simulation model. The new investigation process followed the process we describe in Figure 2. We established a promoter function, gave it a resource, and prioritized the promoter's work. When that did not suffice, we enlisted an investigator or lab technician to help. We estimated that the number of staff hours invested in the promoter's job each month equaled approximately two positions. Six months after implementation, the average investigation's throughput time was 92.5 days, and the initial report was completed within 31 days on average. The quality of investigations (measured by the percentage of investigations with up to two root causes and with likelihood levels of likely or very likely) was 68 percent (it was 26 percent before our improvement process). We believe that a combination of the new quality-rating scheme (Figure 1) and the improved investigation process kept the team more focused and resulted in such a significant improvement. Results stayed relatively stable after a year (90 days throughput time, 36 days to the initial report, and a quality measure of 55 percent).

It is important to note that personnel changes are frequent in a military environment, especially among compulsory service soldiers and officers. During the year, the failure-analysis team changed, several urgent and time-consuming investigations were required, causing ongoing investigations to be deferred, and two investigators (one of whom was senior) left their positions and were subsequently replaced. Therefore, the improvement is all the more impressive. The average improvement rates after 6 and then after 12 months were (1) throughput time improved by 30 percent, thus meeting our goal; (2) the time to complete the initial report decreased by 20 percent; and (3) the investigation's quality measure improved by approximately 136 percent.

Summary

We have presented a unique hierarchical, dual-phased improvement framework in which we integrated a Kaizen project and stochastic networks modeling. We found that they complemented each other and proved very useful in improving the performance of an air force failure-analysis team that investigates aircraft failures.

The intensive effort that the Kaizen project participants made enabled a thorough examination of the investigation process. The team members mapped the process, identified improvement opportunities and inefficient procedures, and constructed a new process that differed greatly from the previous one. For example, they restructured it to include built-in decision points for treating problematic investigations, relocated machines and responsibilities, and defined new performance criteria. The new process uses the common industrial engineering principles of limiting WIP and balancing resource utilization.

We demonstrated that controlling the number of investigations in process was a simple and effective management policy. Its implementation required minimal effort and prevented high resource utilization and network congestion.

As this paper documents, our improvement framework applies to failure analysis in a military environment; however, it can also be applied to job shops, service organizations, and multiproject environments, such as organizations that process software maintenance projects, product development in the chemical industry, and maintenance or retrofit projects in aerospace companies (Leung 2002, Adler et al. 1996, Gemmill and Edwards 1999).

We conclude with two observations that provide insights for those who plan a similar improvement effort.

(1) A Kaizen project is an efficient tool for achieving change in a short time. Its insights should be applied immediately because they usually produce better processes and improve the results. Nevertheless, in complex systems, it is necessary to make a trade-off between the short-term effort and the ability to make a thorough analysis. To achieve more significant results, we recommend performing steady-state modeling and analysis.

(2) A trade-off exists between the loading of a system and its steady-state performance. In loaded systems, the time-based performance decreases with the load. It is a managerial decision to allocate resources and set the loading level in such a way that the system operates efficiently (i.e., resources are not underutilized) and the overall performance is satisfactory. In heavily loaded systems, it might be wise to limit the amount of WIP. Although this comes at the price of delaying incoming work, the reward is superior performance.

Our model has several limitations. Quantitatively capturing the impact of improvements, such as moving equipment, using a professional promoter, or designing built-in decision points, is difficult. We designed the simulation to compare different resource allocations and loading and control methodologies; it did both satisfactorily. It is worth applying common sense changes that are likely to improve the process, e.g., equipment relocation; with time, the effect of such improvements on the process throughput can be measured quantitatively. The model did not consider factors such as the experience of the investigators or the external resources needed to carry out the investigations. Although our results were satisfactory, developing a more accurate model that considers these factors and could help to quantify them would be useful.

Our model also assumes that activity durations follow an exponential distribution. Some empirical evidence and the application's results have shown that assuming an exponential distribution is reasonable for project activities. However, the model can easily be adapted to other distributions, including empirical ones. Our observations and insights in this section will be valid for any distribution selected.

We believe that the simulation tool will be used in the future for examining process improvements, the impact of resource allocations, and the impact of adaptations in activity duration.

Appendix. Notations and Formulas

We defined x as a vector representing the resource allocation $x_1, \ldots, x_5$, where $x_k \in \{1, 2, 3, \ldots\}$, $\forall k = 1, \ldots, 5$. We needed to determine the allocation of all the resources where x_1 represented the number of principal investigators and x_2, x_3, x_4, and x_5 represented the number of investigators, promoters, lab technicians, and SEM, respectively.

As Cohen et al. (2004) describe, we considered n simulation replications; the length of each was m. We defined the duration of the ith investigation from the jth replication as Y_{ij} $(i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n)$. The average investigation throughput time for replication j was $I_j = \sum_{i=l+1}^{m} Y_{ij}/(m-l)$, where l represented the length of the warm-up period. Because replications were independent of each other, we assumed that each T_j was an independent random variable with an expectation $E(T_j)$ that was approximately equal to the true steady-state average. We then computed the overall average throughput time as $T_{\text{mean}} = \sum_{j=1}^{n} T_j/n$ and defined the mean standard deviation as $\sigma = \sum_{j=1}^{n} \sqrt{\sum_{i=l+1}^{m} (Y_{ij} - T_j)^2/(m-l-1)/n}$. Similarly, we defined $T_{\text{ir, mean}}$ as the average throughput time until the initial report. Resource utilization (i.e., the loading of a resource type k) was $\rho_k = \sum_{j=1}^{n} (\sum_{i=l+1}^{m} T_{ijk}/U_j)/(n * N_k)$, where T_{ijk} was the processing time of investigation i by resource k in replication j, N_k was the number of k-type resource units, and U_j represented the steady-state duration of replication j. An investigation was either being serviced or was waiting in queue to receive service. We defined the portion of this waiting time as WT.

Acknowledgments

Many people were involved in the work described in this paper. I particularly acknowledge the contributions of Rafael Kimchi, David Kazir, Jacob Shmerler, Ofer Levy, Avy Shtub, and Avishai Mandelbaum.

References

Adler, P. S., A. Mandelbaum, V. Nguyen, E. Schwerer. 1996. Getting the most out of your product development process. *Harvard Bus. Rev.* (March–April) 134–152.

Anavi-Isakow, S., B. Golany. 2003. Managing multi-project environments through constant work-in-process. *Internat. J. Project Management* **21**(1) 9–18.

Berger, A. 1997. Continuous improvement and Kaizen: Standardization and organizational designs. *Integrated Manufacturing Systems* **8**(2) 110–117.

Bradley, J. R., J. Willett. 2004. Cornell students participate in Lord Corporation's Kaizen projects. *Interfaces* **34**(6) 451–459.

Brunet, A. P., S. New. 2003. Kaizen in Japan: An empirical study. *Internat. J. Oper. Production Management* **23**(12) 1426–1446.

Cohen, I., A. Mandelbaum, A. Shtub. 2004. Multi-project scheduling and control: A process-based comparative study of the critical chain methodology and some alternatives. *Project Management J.* **35**(2) 39–50.

de Castro, P. M. S. T., A. A. Fernandes. 2004. Methodologies for failure analysis: A critical survey. *Materials Design* **25**(2) 117–123.

Emiliani, M. L. 2005. Using Kaizen to improve graduate business school degree programs. *Quality Assurance Ed.* **13**(1) 37–52.

Gemmill, D. D., M. L. Edwards. 1999. Improving resource-constrained project schedules with look-ahead techniques. *Project Management J.* **30**(3) 44–55.

Kaplan, S., S. Visnepolschi, B. Zlotin, A. Zusman. 2005. *New Tools for Failure and Risk Analysis*. Ideation International, Farmington Hills, MI.

Leung, H. K. N. 2002. Estimating maintenance effort by analogy. *Empirical Software Engrg.* **7**(2) 157–175.

Swank, C. K. 2003. The lean service machine. *Harvard Bus. Rev.* **81**(10) 123–129.

Thomas, J. M. 1989. Industrial experience with structural failure. T. Wierzbicki, N. Jones, eds. *Structural Failure*. John Wiley & Sons, New York, 511–545.

The colonel heading the Resources, Budgets & Quality Department writes: "Several years ago we became acquainted with the Kaizen methodology. We applied it to many projects and found it very useful. While most of our improvement processes are aimed at reducing the costs, this one is unique and goes beyond that.

"Air forces realize that safety and cost-effectiveness are of paramount importance; they are therefore constantly striving to improve and up-grade these aspects of air travel. In this regard we implemented the Kaizen technique, and found that it is ideal for addressing and treating flight safety and all the related ramifications, and that it enables the Air Force to embark on and complete its missions with a significant degree of certainty. Thus the resultant improvement is more meaningful than one that aims at 'just' reducing costs.

"The work described in the article achieved a significant improvement in our technical failure investigations' process. This means that we are safer and more effective.

"The other unique finding through the research is that our improvement methodology has been enriched beyond Kaizen by introducing Stochastic Processing Networks modeling. We shall therefore use these tools to bring about improvements in the future."