INL/EXT-09-16815

Safeguards Envelope Progress FY 09

Duc Cao
Nicole Schonenbach
Richard Metcalf
Robert Bean

September 2009

The INL is a U.S. Department of Energy National Laboratory
operated by Battelle Energy Alliance

Idaho National Laboratory

INL/EXT-09-16815

Safeguards Envelope Progress FY 09

Duc Cao
Nicole Schonenbach
Richard Metcalf
Robert Bean

September 2009

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Office of Nuclear Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

Document Cover Sheet

	Report on Multi-Attribute Correlation Analysis
Name/Title of Deliverable/Milestone	INL/EXT-09-16815 "Safeguards Envelope Progress FY09"
Work Package Title and Number	C.30.40.10.01.02 Advanced Control and Integration
Work Package WBS Number	C.30.40.10.01 Advanced Safeguards and Security Development
Responsible Work Package Manager	Robert Bean / Robert Bean 9/28/09 (Name/Signature)

Date Submitted September 30, 2009

Quality Rigor Level for Deliverable/Milestone ☒ Rigor Level 3 ☐ Rigor Level 2 ☐ Rigor Level 1 ☐ Nuclear Data

This deliverable was prepared in accordance with Idaho National Laboratory's
(Participant's Name)

QA program which meets the requirements of

☒ DOE Order 414.1 ☐ NQA-1-2000

This Deliverable was subjected to:

☒ Independent Technical Review ☐ Peer Review

Independent Technical Review (ITR)	**Peer Review (PR)**
Review Documentation Provided	**Review Documentation Provided**
☐ Signed ITR Report or,	☐ Signed PR Report or,
☐ Signed ITR Concurrence Sheet or,	☐ Signed PR Concurrence Sheet or,
☒ Signature of ITR Reviewer(s) below	☐ Signature of PR Reviewer(s) below

Name and Signature of Peer Reviewers(s)/Independent Technical Reviewer(s)

Dale Kotter Dale K Kotter 9/28/09

Safeguards Envelope Progress FY09

INL/EXT-09-16815
Revision 0

September 2009

Approved by

Robert Bean

September 28, 2009
Date

ABSTRACT

The Safeguards Envelope is a strategy to determine a set of specific operating parameters within which nuclear facilities may operate to maximize safeguards effectiveness without sacrificing safety or plant efficiency. This paper details advanced statistical techniques that will be applied to real plant process monitoring (PM) data from the Idaho Chemical Processing Plant (ICPP). As a result of the U.S. having no operating nuclear chemical reprocessing plants, there has been a strong interest in obtaining process monitoring data from the ICPP. The ICPP was shut down in 1996 and a recent effort has been made to retrieve the PM data from storage in a data mining effort. In a simulation based on this data, multi-tank and multi-attribute correlations were tested against synthetic diversion scenarios. Kernel regression smoothing was used to fit a curve to the historical data, and multivariable, residual analysis and cumulative sum techniques set parameters for operating conditions. Diversion scenarios were created and tested, showing improved results when compared with a previous study utilizing only one-variable Z-testing[7].

FOREWORD

The work described herein supported by the U.S. Department of Energy under DOE Idaho Operations Office Contract DE-AC07-05ID14517.

ACKNOWLEDGEMENTS

Neil Liester of Idaho Cleanup Project has been a huge help in retrieving the raw data from the old ICPP. Duc and Nikki would like to thank Ernie Laible at INTEC for his assistance with obtaining metadata documents from storage. Leroy Lewis of Idaho National Laboratory has provided a wealth of knowledge, helping to decipher the meaning of much of the data. Thanks also to Ray Miles and Kim Ho of Idaho National Laboratory for working to put together the database which currently holds the raw data. Humberto E. Garcia of INL and J. Wesley Hines of the University of Tennessee Knoxville were both consultants for methods of statistical analyses, and Yi Cao was the author of the Kernel Regression function. The authors would also like to thank Richard Metcalf and Aaron Bevill of Idaho National Laboratory/Texas A&M University for their foundational work in the summer 2008.

CONTENTS

ACRONYMS

AAKR	Auto associative kernel regression
DOE	Department of Energy
FAR	False Alarm Rate
IAEA	International Atomic Energy Agency
INL	Idaho National Laboratory
ICPP	Idaho Chemical Processing Plant
MAUA	Multi-Attribute Utility Analysis
MBP	material balance period
MC&A	material control and accountability
MUF	material unaccounted for
PM	process monitoring
PND	probability of nondetection
PP	physical protection
SNM	special nuclear material
SQ	significant quantity

Safeguards Envelope Progress FY09

INTRODUCTION

The benefits of nuclear power are again allowing the nuclear industry to refocus efforts on developing new technologies and processes to further the safe use of this reemerging power source. Specifically, it is clear that modern nuclear facilities will require enhanced safeguards, but for nuclear energy to remain competitive in a free market, these enhanced safeguards cannot significantly increase cost. Safeguards have not historically been integrated into the design process, or even integrated fully into the operation of facilities. This has resulted in external, regulatory requirements adding synthetic costs to nuclear facilities because the industry has not embraced a systems engineering approach to safeguards. While the design phase systems engineering approach to safeguards would be Safeguards-by-Design, the operating-phase systems engineering approach is the creation of a Safeguards Envelope. The Safeguards Envelope Program is currently working on a project to increase the security within nuclear facilities, using the Idaho Chemical Processing Plant as an example case.

The Idaho Chemical Processing Plant (ICPP) was a spent nuclear fuel reprocessing plant at the Idaho National Laboratory (INL) site which operated from 1953 through 1996. In 1980, a state-of-the-art process monitoring system was installed which measured temperature and pressure information at regular time intervals. This data is very valuable to process monitoring (PM) research, which is continually attempting to increase the ability to detect ever smaller diversions of special nuclear material (SNM) as well as increase the material balance period (MBP).

Two factors determine the optimum MBP. The false alarm rate (FAR) is the rate or percent of alarms which falsely declare a diversion scenario is taking place. The probability of nondetection (PND) is simply 1 minus the probability that a diversion will be detected. For a given set of parameters, decreasing the FAR usually requires relaxing the operational constraints and thresholds, but at the same time can increase the PND as it raises the possibility of hiding an abnormality. Thus, optimizing the MBP is also a problem of optimizing the FAR and PND.

Different statistical tests, however, provide different optimal FAR and PND. In this study, kernel regression analysis is applied to a declared ‘event’ from ICPP PM data to create a best fit curve. A trial data set is simulated from the ICPP data consisting of a ‘normal set’ and a ‘diversion set.’ Residual analysis and cumulative sum techniques are applied to determine optimum bounds for acceptable operating conditions based upon resultant FAR, PND, and MBP.

THEORY OF SAFEGUARDS ENVELOPE

Definition of Safeguards

Safeguards are put into place to protect nuclear material from proliferation or other non-declared purposes, and are vital for securing the future of nuclear energy domestically and globally. This principle defines the need for fields such as nuclear nonproliferation, which guards against the theft or diversion of SNM. SNM is tracked through a nuclear facility, and that facility is responsible for ensuring that no more than one significant quantity (SQ) is unaccounted for in a given timeframe, ranging form one year to as little as 30 days. Depending upon the size of the plant, this can be an enormous and seemingly impossible task.

Definition of the Safeguards Envelope

Safeguards envelopes are operational spaces designed similarly to the idea of a safety envelope[1]. For years, the concept of an area of operation that does not needlessly endanger the public, personnel, or equipment of a nuclear facility has been a major component of nuclear facility design. This safety envelope methodology can just as easily be applied to safeguards, such that an operating space can be constructed that does not needlessly risk proliferation activities, while also not overburdening the operator with regulatory costs. The goal is simply to define a set of operational parameters which increase the probability of detecting a diversion of nuclear material[1] and apply them to operating and new nuclear facilities to make safeguards a point of optimization for operations instead of a fixed, ad hoc procedure. The most effective way to develop these parameters is to use real nuclear plant process monitoring data and perform statistical analyses and modeling methods.

Previous Work

Much effort has already gone into research to determining these parameters. Tom Burr of Los Alamos National Laboratory (LANL) has done several studies on various types of statistical analyses as well as multivariate correlation. While he has looked into performing some of these analyses over the transient modes of a plant (filling and emptying a tank or tank to tank transfers), much of his research was done over static states in which nuclear material was not moving at all[2, 3, 4, 5, 6].

Richard Metcalf and Aaron Bevill of Idaho National Laboratory/Texas A&M University retrieved some of the process monitoring data from ICPP and performed a basic Z-test with a simulated diversion to demonstrate the usefulness of this data[7].

PROBLEM STATEMENT

ICPP Facility

During its 43 years of operation, process monitoring was of great importance in the day to day operations of the Idaho Chemical Processing Plant. When the state-of-the art level/density scanner was introduced as part of the process monitoring system, the accuracy of the data improved drastically, allowing a greater confidence in nuclear materials control and accountability (MC&A). With this more reliable data, statistical analysis methods can be more effectively utilized to detect diversions of special nuclear material (SNM), and determine optimum operating parameters for both materials accountability and operator impact.

Diversion Detection

Process monitoring techniques and analysis methods are a primary focus in attempts to increase the ability to detect a diversion. The goal is to be able to detect as small of a diversion as possible without raising the false alarm rate (FAR) or the probability of nondetection (PND) above a reasonable level. If a FAR is too high it is not cost effective, for every alarm must be investigated which would be intrusive on the plant operator's other duties. With a high PND the issues are obvious; it is unacceptable to rely on a system for nuclear security when it fails to detect diversions. By utilizing advanced statistical analysis techniques, one can determine a balance of optimum working parameters and also obtain a better material balance period (MBP). In this study, MBP is used as the figure of merit because it can easily accept the FAR and PND into a single metric. This single figure of merit allows for a single optimization point rather than two, but these are both available for more detailed or plant specific studies.

DATA DESCRIPTION

The raw data obtained from ICPP contains over six gigabytes of information and spans from October of 1986 through April 1996. Over this 10 year period, data was recorded for each tank, centrifuge, evaporator, column, valve, jet, and air lift. Some of this data, for example for a valve, simply consists of either a 1 or a 0 to denote whether it was open or closed. The centrifuges only had speed indicators, and dissolution tanks only contain information for off-gas control signal or charge soot hydraulics. The most effectively analyzed process monitoring data consists of that found in the accountability tanks, feed tanks, and sample pots.

The raw measured data associated with these tanks leads to the ability to calculate specific gravity, volume, and level of material in the tank. These values are calculated from the highly accurate LR and DR dip tube measurements. LR and DR are each pressure differences between a given dip tube (D or L) and a reference vapor pressure (R)[8] illustrated in Figure 1 below.

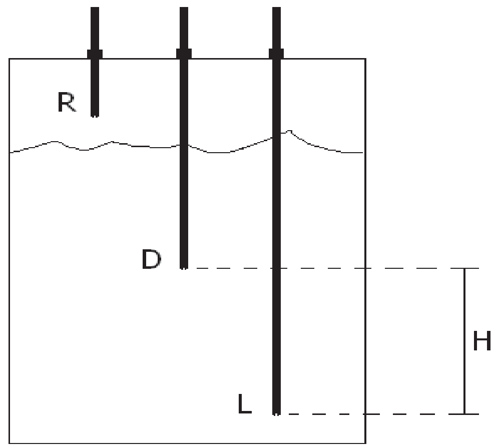


Figure 1. Crude diagram of the dip tube set-up within a tank. R measures reference vapor pressure, D measures pressure at one depth in the tank, L measures the pressure at a set height above the bottom of the tank, and H is a known distance between L and D.

Density is calculated using the equation below, where g is the acceleration due to gravity and H is a known distance between D and L. Volume and level are calculated with additional constants and are not described in detail here.

$$\rho = \frac{LR - DR}{gH} \tag{1}$$

Although the pressure measurement error associated with the dip tubes is extremely small (level calculations can reach less than 0.015%[8]), the purely raw data measurements LR, DR, and TT (Temperature) are used in the program with added noise to simulate more realistic data relative to most nuclear facilities.

STATISTICAL ANALYSIS

In this study, kernel regression was used to create a best fit function to the data received from ICPP. Kernel regression is a state estimation technique which is considered a nonparametric technique, for unlike linear regression, it does not assume a fundamental distribution in the data[9]. At each observed data point, a Kernel, or weighted function, is centered, and the Kernel assigns a weight to each position based on its proximity to each data point[9]. A more detailed discussion of Kernel regression can be found in Appendix A.

The proposed algorithm compares historical and trial data sets and tests the ability to detect a diversion by looking at two items: degree of residual randomness and deviation from the mean. To determine the effectiveness of the statistical tests, we perform a Markov Monte Carlo simulation and run 500,000+ trials as a simple method for finding out the resultant FAR and PND values.

ANALYSIS

Theory

In reality, data always has noise, and due to this noise detecting small diversions is often difficult. To an approximation, we can assume that all measurements take the following form:

$$y_{measured}(t) = y_{true}(t) + \varepsilon_{calibration} + \varepsilon_{measurement} \qquad (2)$$

where $\varepsilon_{calibration}$ is the calibration error and $\varepsilon_{measurement}$ is the measurement error. Calibration error is due to the non-perfect tuning of the measurement device and is usually a static additive error. The error however is randomly distributed from one device to another. The more familiar measurement error is that which arises from small fluctuations within the control volume (e.g. miniscule temperature fluctuations, or small movement) and is known to be normally distributed. As Eq. (2) shows, both errors mask what the true value actually is and can hamper any type of verification process. Indeed, both can also be averaged assuming enough data exists to do so. Unfortunately, that is not the case in most scenarios, including our ICPP data. This is the realm in which statistical tests find their application as they look to the overall data trends to discover any abnormalities. Before tests are created, diversion behavior must first be understood.

Material diversions affect two components of measurement data: residual randomness and deviation from the mean or “expected” value. A residual is defined as the difference between the measured value and the true value where $y_{true}(t)$ would be an exact analytical value.

$$y_{residual}(t) = y_{measured}(t) - y_{true}(t) = \varepsilon_{calibration} + \varepsilon_{measurement} \qquad (3)$$

As Eq. (3) shows, a measurement residual should be nothing more than a time series of errors with a random distribution and mean of zero. In a diversion case however, the residual would take on an entirely different behavior. First, it is important to understand that abnormal data can be seen as normal data with an added deviation where *diverted(t)* is the nuclear quantity taken as a function of time as shown in Eq. (4) below.

$$y_{abnormal}(t) = y_{true}(t) + \varepsilon_{calibration} + \varepsilon_{measurement} - \text{diverted}(t) \qquad (4)$$

If the residual of this curve was computed with respect to the true values of a normal curve, illustrated in Eq. (5) below, then it becomes obvious that the residual of an abnormal data curve is just a normal residual, such as Eq. (3), but with an added non-random and/or non-zero mean function.

$$y_{abnormal}(t) - y_{true}(t) = \varepsilon_{calibration} + \varepsilon_{measurement} - \text{diverted}(t) \qquad (5)$$

In other words, to determine whether or not a tank has been tapped, one simply needs to look at the residual of its data; if the residual has neither a purely random distribution nor a zero mean, then assume that a diversion has taken place. These tests can be performed with hypothesis Z-testing, standard deviation calculation, or cumulative sum examinations.

Unfortunately, detection with the above methodology is difficult for two reasons: not knowing $y_{true}(t)$, and having sparse data. Computing the most accurate residuals requires knowing before-hand what $y_{true}(t)$ is, which is technically impossible. In fact, knowing it would imply perfect measurements and make this entire statistical process pointless. However, what is known is the historical data, which tells what the measurement "ought" to be. With that, it becomes feasible to make good *approximations* of $y_{true}(t)$, especially with good fitting techniques. One must take caution, for approximations can be too uncertain if the base data is too sparse. Even the tests themselves can be misleading if not enough information is present. Again, advanced statistics become useful. Numerous techniques have evolved which take advantage of sparse data and create reliable models to work with (e.g. Principle Component Analysis, Least-Squares Fit, Student's T-testing). With both reliable historical data and advanced statistics, it becomes very possible to distinguish abnormal behavior from normal operating conditions.

Numerical theory aside, some important notes must be mentioned about the testing scenario. The setting involves a reprocessing facility tank filling and flushing a (assumed) homogenous nuclear material solution. Measurements of the solution's level, density, and temperature (LDT) are taken every four minutes and assumed to have a form similar to that of Eq. (2), but with an assumed zero calibration error. An artificial diversion is introduced in the same way as Eq. (4) and involves gradually taking 0.5% of the tank (in terms of level) in a linear fashion until flushing is complete. The exact start and stop times of the tank fill and flush is assumed known at all times.

Methodology

The basic premise of the algorithm is to take a trial data set and statistically compare it with a historical set by residual analysis. Before beginning residual construction, the data must first be collected and processed in the appropriate manner. A fully transient state was first isolated within the data logs and its LDT information was extracted into a data array with MATLAB. Because the state-of-the-art measurement systems ICPP had at the time, there was very little error within the data itself. Therefore, it was assumed that this information represented the "true" values, henceforth called the true curve, with which to build our simulated, noisier measurements.

In order to fully test the algorithm capabilities, a total of three curves were created: a historical curve, a normal trial curve, and a diversion trial curve. The latter two were meant to test the FAR and PND respectively. To build the "historical" curve, Gaussian noise with a standard deviation of 0.2% was added to the true set to simulate measurement error and labeled accordingly. To create a normal trial curve, henceforth called the normal curve, true curve values were again taken and similar noise was added. Creating the diversion trial curve, henceforth called the diversion curve, followed a similar process, but this time with a linearly increasing diversion function that peaked with a value of 0.5%.

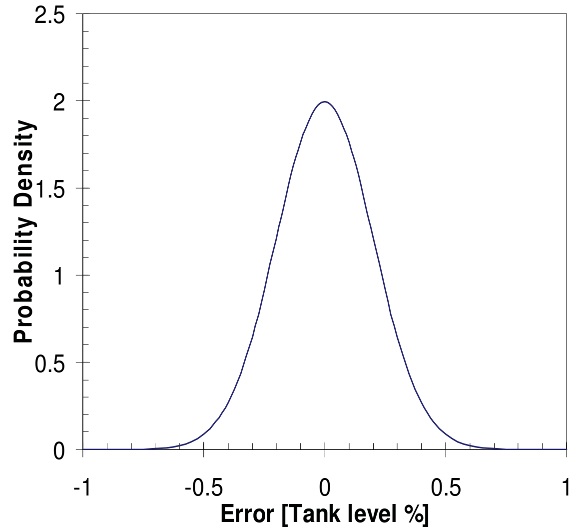


Figure 2. Probability density of measurement error. In this case, it is for the tank level.

Once the three simulated curves were created, Kernel regression was performed on the historical set to later approximate residuals. Kernel regression is a powerful state estimation technique designed to fit an approximate curve to noisy data. Unlike most familiar regression techniques, kernel regression is non-parametric and does not actually make any initial assumptions about the shape of the curve. Instead, it applies a Gaussian weight function centered at each data point and gives each neighboring point a contribution that is proportional to their distance.

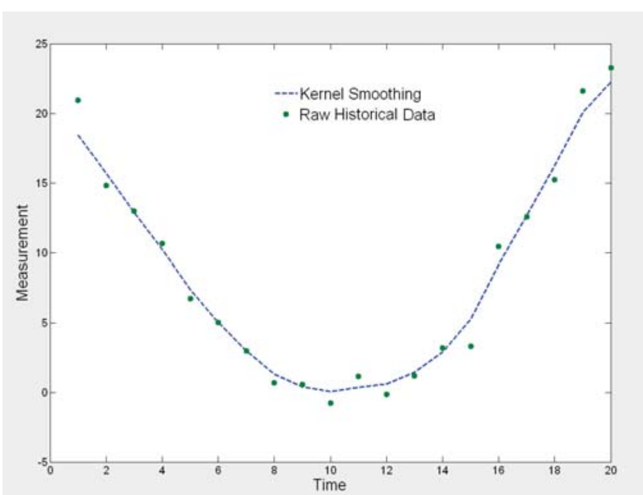


Figure 3. Kernel Smoothing on fake historical data. The example above is a noisy quadratic.

The degree of fitting is also a user-set parameter, called the kernel bandwidth. Too low of a value connects the dots poorly, while one too high will "over fit" and produce large errors. This is one of the parameters that can be optimized in the algorithm for best performance.

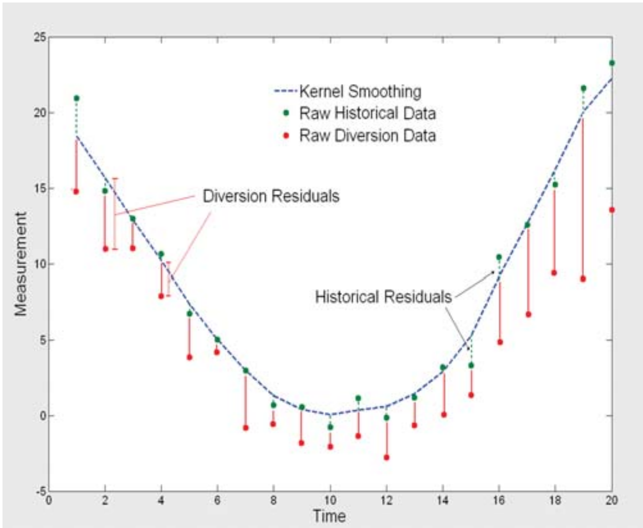


Figure 4. Diagram showing residual analysis with historical data and diversion data.

Once the kernel smoothed historical curve is obtained, the difference between that curve and the two trial curves (normal and diversion curve) give each trial curve their respective residual approximations. This is done by simply subtracting the raw data from the kernel smoothed curve for both the historical and trial case.

With the computed residuals, hypothesis Z-testing is then used to test for the criteria of randomness. To be considered normal, the residuals must have a mean of zero and a standard deviation similar to that of the historical residual (historical data minus kernel smoothed data). The resultant probability reveals how well the trial residual follows the stated constraints and can be compared to a threshold for judgment.

In addition to the Z-test, a cumulative sum threshold test was also implemented in order to measure mean and deviation behaviors in a way that the Z-test does not.

To this end, the residual vector components were each taken to the 1.5 power in order to better distinguish outliers (an L 1.5 norm). Then, a summation of the residual was taken and compared to a threshold to determine abnormal trends. Both these tests were used in OR fashion (if *either* test dismissed a case as a diversion, that result must be recorded).

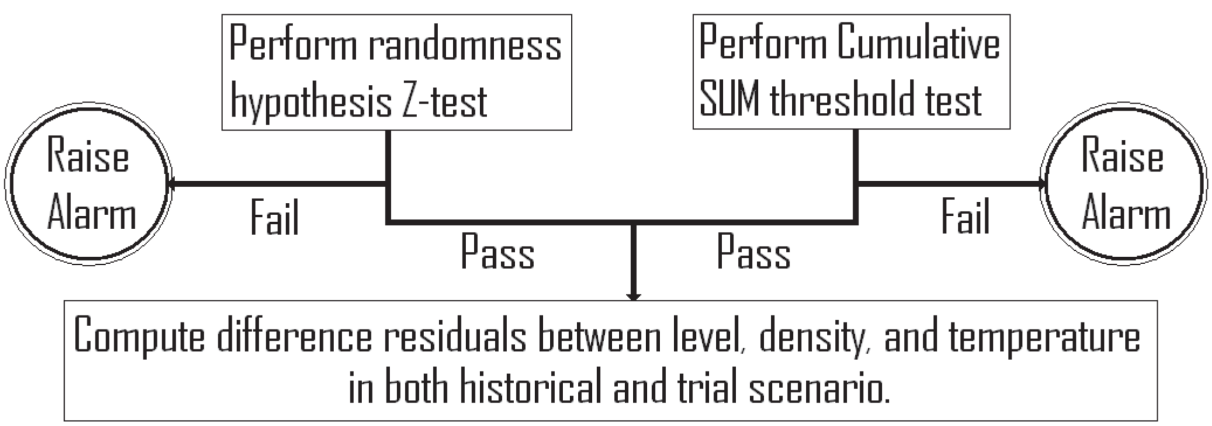


Figure 5. Flow-chart of the test priorities and the order in which they are executed, as well as how their results are interpreted.

Moving into the multi-variate tests, a difference residual comparison between, say, the level curve was compared against that of the density and temperature curves. This was done by computing the residuals for each individual curve with the same methodology as the one-variable scenario, and then taking the residual difference between curves. This was done to compute an “effective” difference between the two curves in terms of a normal distribution. Having these residuals, the same Z-test and cumulative sum threshold tests from before were used, but this time a stricter criterion was used to decide final judgment on a trial case. Because multi-variate calculations add more noise than a single, both the comparison between, say, a trial level curve and a density curve as well as a temperature must agree in result (there must be a majority rule). This result then was also used in areas where the one-variable tests were unable to detect anything abnormal. This way, the baseline PND could only be decreased at a cost of a small sacrifice of the FAR.

Finally a *transient* operated at half speed is simulated in order to grab more data points. This was primarily to show that with more data points, and assuming easy diversion detection at steady-states, the FAR and PND values can be markedly decreased and the MBP should increase. To simulate half-speed at the transients, linear interpolation was done with the true curve between data points and applying the algorithm to the ‘new’ data to obtain new FAR, PND, and MBP values.

Results

The results of the tests were largely satisfactory and showed especially large improvement when considering half-speed operation at the transients. Assuming a baseline of 8 days, an increase in MBP as high as 50% was observed for three variables at half speed. For an explanation of the MBP derivation, see Reference 7, Appendix A. With three variables, FAR was improved from 1.47% to 0.79% and PND had a drastic improvement from 10.36% to an impressive 1.30% when changing to half speed.

Of particular interest is the change from analyzing only one variable with just Z-testing, to analyzing one variable with higher statistics, to analyzing three variables with higher statistics. The FAR for the developed algorithm improved from only Z-test results for one variable, but when three variables were analyzed together, FAR was actually worse. PND and MBP on the other hand, improved for each test. Overall, while FAR was worse with three variables, multivariate analysis appears to be more beneficial, for PND is lower and MBP is able to be longer, benefiting plant operations. A summary of the results are tabulated below.

TABLE 1

Tabulated Results Including Previous Work.

Data Type	Z-test	Kernel Reg.	CUMSUM	Speed	FAR	PND	MBP
One-Var.	YES	NO	NO	Normal	1.39%	23.20%	8.57
One-Var.	YES	NO	NO	Half-speed	2.91%	1.41%	10.85
One-Var.	YES	YES	YES	Normal	1.15%	12.39%	9.05
One-Var.	YES	YES	YES	Half-speed	0.44%	1.85%	11.5
Three-Var.	YES	YES	YES	Normal	1.47%	10.36%	9.15
Three-Var.	YES	YES	YES	Half-speed	0.79%	1.30%	12.03

More work, however, still needs to be performed as there are multiple variables within the algorithm that can be further optimized. Additionally, even more tests could be added such as tank-tank correlations and neural network optimization to achieve even better results. Some of the tests currently used could also be modified such that they offer less FAR, higher signal-to-noise ratios, or require fewer computations. Nevertheless, it is clear that a combination of both half speed operation at the transients and better statistical tests can not only improve security, but also productivity.

EXPECTED CHALLENGES

Determining Equipment Failure versus Diversion

An alarm is raised because the plant begins operating outside the accepted parameters. However, the abnormality can be due to something as devious as material diversion, or as innocent as equipment failure. Since equipment does wear out and eventually fail, it could affect the process monitoring system and its ability to detect diversions effectively, raising the FAR[7]. A code was developed in 1997 to address this specific problem, called IGENPRO, which was designed based on fuzzy logic and PM techniques[10]. IGENPRO attempted to effectively estimate when a component might fail within the plant, and could be used to develop a more proactive maintenance schedule, rather than waiting until things failed completely. This system or possibly a more advanced code could alleviate the issue of increased FAR due to equipment failure.

Tight Boundaries for Each Variable

Three variables are being analyzed in this study: level, density and temperature. To analyze each variable independently, one must define bounds, or parameters for the variable, outside of which the process cannot operate. Individually these parameters work well. However, when correlating multiple variables for a given level and density, as an example, one would expect temperature to be within a more narrow range, and could thus raise an alarm more easily, raising the FAR. The desired scenario is that which achieves as low FAR as possible for each variable without significantly increasing PND. Further investigation to identify the proper optimization is needed.

FUTURE WORK

Multi-Attribute Utility Analysis

Multi-Attribute Utility Analysis, (MAUA) is a decision analysis technique which utilizes several methods to analyze several factors or variables, and come up with a single decision[7]. This tool will be applied to two safeguards areas of focus: proliferation resistance (PR) and diversion detection/physical protection (PP).

MAUA can be applied to a plant process to determine a lower threshold which the PR level is not allowed to go below during plant operation in both steady state and transient state conditions. This could help to determine areas of the plant which need additional safeguards measures to raise PR[7].

Using MAUA for diversion detection is an aspect which will be explored, but not as a current main focus. Several different diversions would be simulated using a Markov Monte Carlo method, with varying weights and paths. Compared with more general algorithms, MAUA would be able to optimize a set of parameters to specific diversion types, which could raise detection rates and lower FAR and PND. However to apply this method in real time would require a detailed knowledge of the diversion paths' significance which is rarely the case in real-life operations.

Automated Optimization

While applying kernel regression analysis to process monitoring data improved the ability to detect a diversion, lowered the PND, and also resulted in a lower FAR than the student Z-testing performed by Metcalf and Bevill, further optimizations can be made through automated optimization techniques such as neural network modeling and auto associative kernel regression (AAKR) modeling. One such option is the use of the Process Equipment Monitoring (PEM) MATLAB toolbox developed by J. Wesley Hines at the University of Tennessee Knoxville.

Scalability Studies

Once working parameters have been formed for nuclear facilities, we must determine whether these conditions can be applied to a facility and allow it to operate efficiently. Some areas of concern are how things may scale, operational impact, and cost. Much of the data monitoring methods should scale linearly, according to the size of the plant. The volume of data obtained from the process, and the length of the material balance period dictate the impact on plant operators. For a large process monitoring system, more data points will be collected, which will increase the amount of data which must be tracked and analyzed. A longer MBP will allow the plant to operate longer between material accounting inventories, which could save the plant extraneous costs. Further study is needed to identify scaling issues that must be addressed as the Safeguards Envelope concept is applied to larger facilities.

APPENDIX A

Kernel Regression

Problem

A scientist has collected some measurements, and they seem to have no discernable trend that he can describe (it is clearly non-linear). However, he wants to accurately interpolate his data such that he can well approximate a true description of his measurements. The scientist cannot use simple fitting methods such as least square fit or polynomial regression because he has no clue as to the correct shape of the data and does not want to assume incorrectly.

Solution

A non-parametric fitting technique is needed that does not make any initial assumptions as to the shape of the data, yet still reliably creates a best fit curve. Kernel Regression is a good candidate for this type of problem.

How it works

With a given data set, a kernel (or weight function) is centered at each data point and at each point is used to evaluate the weight of its neighbors for local fitting.

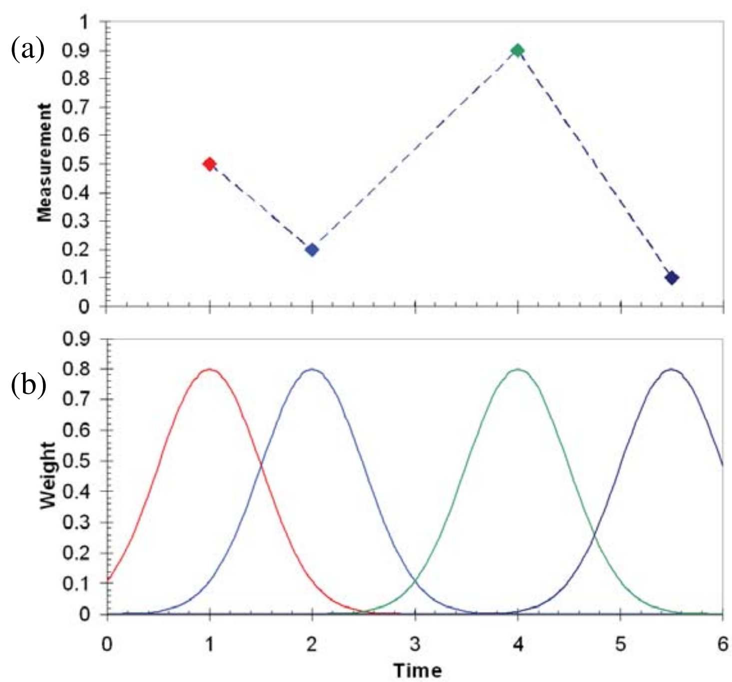


Figure A-1a, A-1b. Plot of a) measurement points which have no linear relationship and

b) associated Gaussian weight functions for their respective data points.

In reality, there exist many different kernel functions (e.g. square, quartic, cosine), but the Gaussian remains the most popular. The Gaussian kernel function is as follows:

$$K(x) = \frac{e^{\frac{-(x-X)^2}{2a^2}}}{\sqrt{2\pi}a} \tag{1}$$

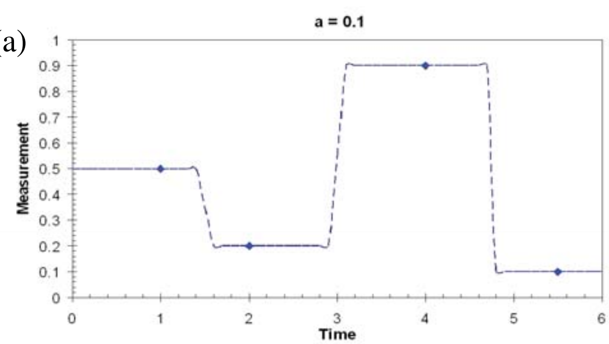
Where X represents the x-value of the measurement point, x represents the x-value of the interpolated point, and a represents the kernel bandwidth. More will be explained about the kernel bandwidth later, but for now assume it to be any value.

Once applying the weight functions at each desired point, the interpolated y-value can be computed using the Nadaraya-Watson estimator:

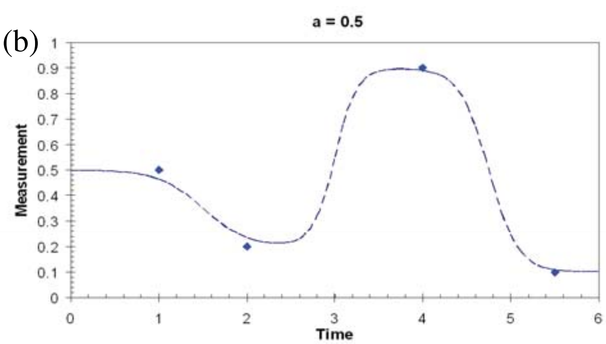
$$y_j = \frac{\sum_{i=1}^{n} Y_i K(x_j, X_i)}{\sum_{i=1}^{n} K(x_j, X_i)} \tag{2}$$

where i represents the ith measured point, j the jth interpolated point, Y_i the ith measurement, and y_j the jth weight, interpolated value. As the kernel bandwidth has yet to be chosen, here are the results for Figure A-1-a data at various bandwidth values.

(a)



(b)



(c)

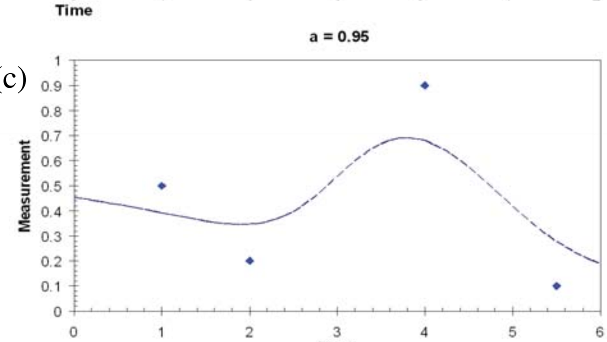


Figure A-2a, A-2b,A-2c. Plot of data from Figure A-1a with kernel smoothing at various kernel bandwidths. Higher values approach linear best-fit as shown by c).

As can be seen from Figure A-2, various kernel bandwidths give drastically different results. The kernel bandwidth is a user-set parameter that essentially controls the width of the weight function (or rather the "broadening"). Too low of a kernel bandwidth and each measurement point carries all the weight, resulting in just step interpolation such as in Figure A-2a. Too high of a value will "overfit" the data by giving every point nearly equal weight and will approach fitting a single line (linear best fit) to the entire data set. In order to find the best value of the kernel bandwidth, optimization is necessary. This usually requires some outside knowledge that can hint at which value is "right."

REFERENCES

[1] Metcalf, Richard; Ragusa, Jean; Bean, Robert; "Safeguards Envelope: the First Steps" 8th International Conference on Facility Operations, Safeguards Interface, Portland, OR March 30-April 4, 2008

[2] Burr, Tom L.; Wangen, Lawrence E.; Mullen, Mark F.; "Authentication of Reprocessing Plant Safeguards Data through Correlation Analysis"; LA-12923-MS (ISPO-374); (1995)

[3] Burr, Tom; Wangen, Larry; "Development and Evaluation of Methods for Safeguards Use of Solution Monitoring Data" LA-13185-MS, (LSPO-389), 1996

[4] Burr, Tom; Wangen, Larry; "Enhanced Safeguards Via Solution Monitoring" LA-13186-MS, (ISPO-390), 1996

[5] Burr Tom L.; Coulter C. Alton; Wangen, Larry E.; "Benchmark Data for a Large Reprocessing Plant for Evaluation of Advanced Data Analysis Algorithms and Safeguards System Design"; LA-13414-MS (ISPO-397); 1998

[6] Burr, Tom; Jones, James; Wangen, Larry. "Multivariate Diagnostics and Anomaly Detection for Nuclear Safeguards" LA-UR 94-2292

[7] Metcalf, Richard; Bevill, Aaron; Bean, Robert. "Safeguards Envelope Progress FY08" INL/EXT-08-14915; September 2008

[8] Boland, T.J.; Deveraux, H.R. "Precision Level/ Density Scanner: Advanced Safeguards Systems Development for Chemical Processing Plants"; Report: ENICO-1080; January 1982

[9] Teknomo, Kardi (2007) **Kernel Regression** http://people.revoledu.com/kardi/tutorial/regression/kernelregression/

[10] Morman, J.A.; Reifman, J.; Vitela, J.E.; Wei, T.Y.C.; Applequist, C.A.; Hippely, P.; Kuk, W.; Tsoukalas, L.H.; "IGENPRO Knowledge Based Digital System for Process Transient Diagnostics and Management"; CONF-9709106; 1997